Identification of Target Genes of an Erythroid Transcription Factor Complex Containing SCL (TAL1)

Hau Ling Jim

Homerton College

Thesis submitted for the degree of

Doctor of Philosophy

University of Cambridge

2008

Table of Contents

Chapter 1 Introduction

Chapter 2 Materials and Methods

Chapter 3 Developing working siRNAs for members of the SCL erythroid complex in K562

Chapter 5 ChIP-on-chip analyses of the SCL erythroid complex

Abstract

Identification of Target Genes of an Erythroid Transcription Factor Complex Containing SCL (TAL1)

Haematopoiesis is the process whereby haematopoietic stem cells give rise to mature blood cell lineages. The SCL (TAL1) gene encodes a key transcription factor (TF) which is expressed in various blood lineages and is essential for haematopoietic development. It has been shown that the SCL protein forms a multi-protein complex during erythroid development with other TFs (GATA1, E2A, LDB1, and LMO2) which binds to a sequence-specific motif to regulate its target genes. Two complementary approaches were used here to identify novel target genes regulated by this TF complex during erythroid development.

In the first approach, short interfering RNAs (siRNAs) were transfected into the K562 cell line to knockdown transiently each of five TFs found in this complex. For these five TFs, a knockdown efficiency of at least 70% was confirmed at the mRNA and protein level within 48 hours after transfection. The biological consequences of these knockdowns were studied using Affymetrix GeneChips in order to identify gene expression changes of downstream targets. In the second approach, chromatin immunoprecipitation (ChIP) was performed for the five TFs of the complex in the K562 cell line and the resultant ChIP material was hybridised to a human transcription factor promoter microarray. A number of novel target genes for the SCL erythroid complex were identified and verified independently using both approaches. The data presented in this thesis revealed that members of the SCL-containing erythroid complex are involved in auto-regulation and regulate genes which have key roles in haematopoiesis or control chromatin structure and function. These findings demonstrate that the expression of this TF complex is tightly controlled and point to an important role for it in orchestrating fundamental biological processes which have profound effects on gene expression in erythroid development.

Chapter 1
Introduction

The human genome is regarded as the blueprint of life and the completion of its entire genomic sequence was a milestone in understanding the functions encoded in our genetic material. The genome contains of all of the coding and non-coding DNA sequences which control all of the functions within all cell types in our body. It is estimated that the human genome contains approximately 20,000 to 25,000 genes representing only 2% of genomic sequence [IHGSC (2004b) (Shabalina and Spiridonov, 2004)], while 98% of the genome is non-coding. The genes encode proteins controlling all of the various biological processes as well as ribosomal RNAs and proteins. The non-coding regions include maintenance elements, such as centromeres, telomeres and origins of replication which control DNA replication and repair, and elements such as promoters, enhancers/repressors, insulators, and regulatory RNAs (micro-RNAs) which regulate the spatial and temporal expression of coding genes.

Expression of eukaryotic genes is a tightly regulated process. It is crucial for genes to be expressed in the correct cell type to an appropriate level and at the correct time during cell differentiation and development in response to internal and external signals. Failure to regulate gene expression patterns can lead to serious consequences in genetic diseases. In the post-sequencing genomics era, with advances in both computational methods and genome-wide experimental approaches, it is important for us to study how different regulatory sequences and proteins interact to control gene expression, not only at a single gene locus, but globally across the genome within complex biological and transcriptional programmes. Understanding how gene expression is regulated is essential for us to fully delineate the function of our genome as well as to search for therapeutic remedies for genetic diseases.

1.1 Regulation of gene expression

Gene expression regulation can occur in different ways: during transcription, mRNA processing, and translation and at the level of protein stability. It is believed, however, that regulation occurs primarily at the transcriptional level. The transcriptional machinery of eukaryotes consists of two complimentary regulatory components: the *cis*-acting elements and the *trans*-acting elements.

The *cis*-acting elements are DNA sequences in the coding or non-coding regions of the genome. Epigenetic information can also be overlaid onto the *cis*-acting elements. This involves chromatin remodelling and modifications (histones or the DNA sequence itself) to create an accessible region

in the DNA for *trans*-factors to bind to initiate transcription. Conversely, some of these processes prevent *trans*-acting factors from binding to DNA by creating inaccessible chromatin environments.

The *trans*-acting elements are transcription factors or other DNA-binding proteins which recognise and bind to specific sequences in the *cis*-acting elements to initiate, enhance or suppress transcription. A transcription factor may regulate multiple genes or they may work in a combinatorial or complex manner to bind to the *cis*-regulatory elements at multiple transcription factor binding sites to generate a huge repertoire of unique and precise control patterns. It is estimated that the human genome encodes approximately 1800 transcription factors (Venter et al., 2001).

1.1.1 *Cis*-acting regulatory elements

Cis-regulatory DNA sequences include two distinct elements: promoters/proximal elements and the distal regulatory regions including enhancers, silencers or repressors, insulators and locus control regions (LCRs). These elements act in co-operation with one another to govern a co-ordinated expression pattern of a gene. They are summarised in Figure 1.1 and described in details below.

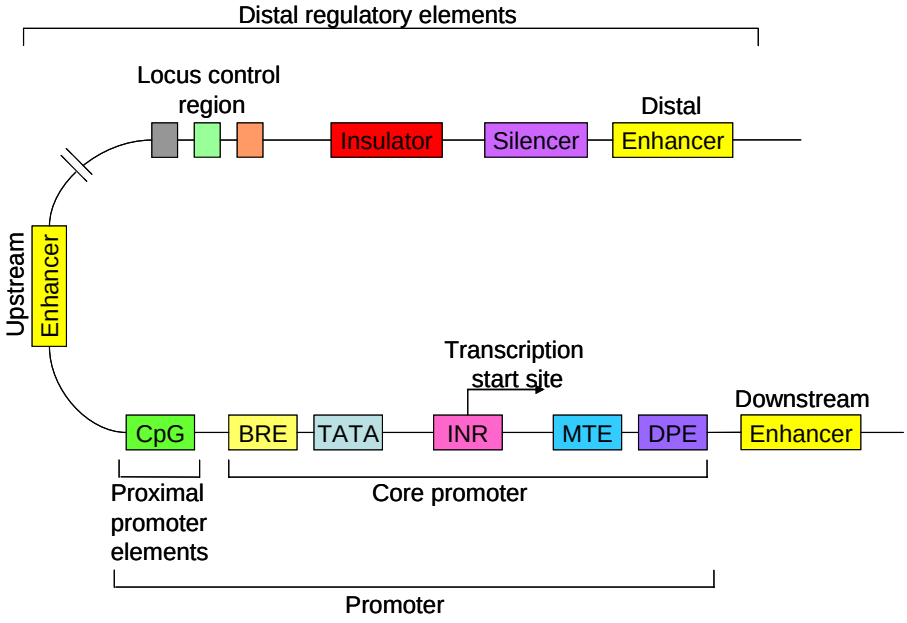


Figure 1.1. A schematic diagram of the types of *cis*-regulatory elements involved in regulation of gene expression. A typical promoter comprises a core promoter and proximal promoter elements such as CpG islands spanning about 1 kb around the transcription start site. The core promoter contains a TATA box (TATA), an initiator element (INR), a downstream promoter element (DPE), a motif ten element (MTE) and a TFIIB recognition element (BRE). Distal regulatory elements such as enhancers, silencers, locus control regions and insulators can be located upstream or downstream or even distant from the transcription start site. Various enhancers, silencers and locus control regions act together to activate or repress promoter activity while insulators prevent inappropriate regulation by regulatory signals from neighbouring genes.

1.1.1.1 Promoters

The RNA polymerase II (Pol II) promoter regions comprise the core promoter and the proximal promoter elements. Pol II promoters transcribe DNA to messenger RNA and small nuclear RNA (Section 1.1.2.1). The core promoter is located approximately 35 base pairs (bp) upstream or downstream of the transcription start site (TSS) and serves as the binding site of factors for assembly of the preinitiation complex (PIC). The core promoter contains a number of elements (Figure 1.2). The TATA box possesses the consensus sequence of TATAAAA located 25 to 30 bp upstream of the TSS. However, this consensus sequence may vary (Wong and Bateman, 1994; Zenzie-Gregory et al., 1993). Although the TATA box was believed to be a fundamental component of the core promoter, it was revealed that only 32% of the potential human core promoters contain the TATA box (Suzuki et al., 2001). The initiator element (INR) is located across the transcription start site (denoted as +1) from -3 to +5 having the consensus sequence of Py Py A(+1) N T/A Py Py. Downstream of the TSS, the downstream promoter element (DPE) functions in conjunction with the INR in TATA-less promoters and is located at +28 to +32 relative to the TSS and possesses the consensus sequence of A/G GA/T C/T G/A/C (Hahn, 2004; Smale and Kadonaga, 2003). Also located downstream of the TSS, the downstream core element (DCE) was first identified in the human β-globin promoter (Lee et al., 2005a). It is located at +10 to +45 relative to the TSS and acts distinct from the DPE. The motif ten element (MTE) is another newly defined element located at +18 to +27 relative to the TSS. It functions in a cooperative manner with the INR but independently from the TATA box and the DPE (Lim et al., 2004). All the core elements (TATA box, INR, DPE, DCE and MTE) initiate the recruitment of TFIID (Transcription factor IID) initiation complex to the promoter for transcription of gene to take place. Another core promoter element is the TFIIB recognition element (BRE) which is recognised by TFIIB instead of TFIID. It is located 3-6 bp upstream of the TATA box with the consensus sequence of G/C G/C G/A C G C C. BRE functions as a repressor of basal transcription whose repression is released upon the binding of activators. The existence of the core elements is not entirely universal (Gershenzon and Ioshikhes, 2005) and it is believed that other core elements may still remain to be discovered. Higher order structural properties of the DNA sequence are also involved in the recruitment of the PIC (Hahn, 2004).

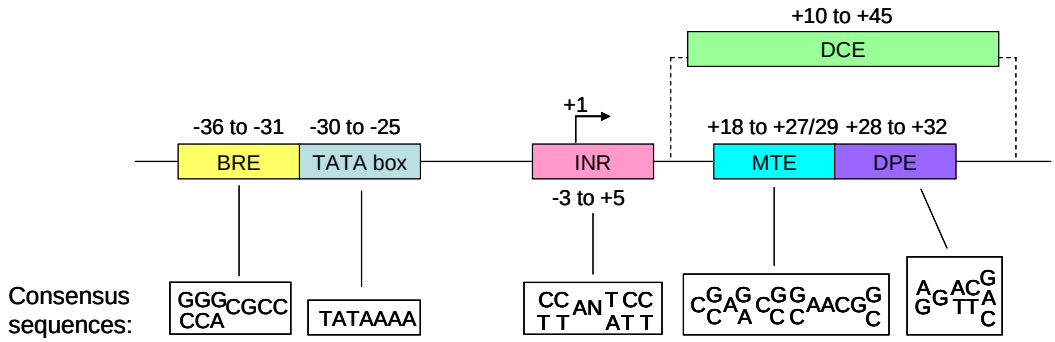


Figure 1.2. The RNA polymerase II core promoter. The locations of the core promoter elements TATA box, initiator element (INR), downstream promoter element (DPE), motif ten element (MTE), downstream core element (DCE) and TFIIB recognition element (BRE) corresponding to the transcription start site (+1) are shown. The consensus sequences of these elements are shown in the white boxes underneath each element. The joint arrow indicates the transcription start site (+1).

The proximal promoter elements refer to sequences upstream of the core promoter which can span up to a few hundred base pairs and can be involved in altering the rate of transcription. An example, of a proximal element is the CpG island which is 500 bp to 2 kilobase pairs (kb) in length and is highly GC rich (Smale and Kadonaga, 2003). They are associated with approximately 60% of human promoters. The core elements in the CpG islands have not been identified but CpG islands contain multiple binding sites for the transcription factor Sp1. CpG dinucleotides are the substrate of methylation by DNA methyltransferases and are normally underrepresented in the human genome as the methylated cytosine can undergo deamination to form thymine. However, CpG islands in the proximal promoters are not methylated in active genes. DNA methylation at CpG islands results in silencing of transcription and is implicated in epigenetic imprinting.

1.1.1.2 Enhancers

Enhancer elements increase the activities of promoters and thus facilitate the transcription of target genes in specific cell types during particular stages in development. Some promoters may be activated by a large repertoire of enhancers in different spatial and temporal environments or in response to different stimuli. An enhancer was first identified in the tumor virus SV40 and was found to increase transcriptional activities of heterologous genes in the host genome (Banerji et al., 1981). Soon after the discovery of the viral enhancer, the first endogenous enhancers in mouse and human were found to activate the immunoglobulin heavy chain gene in a tissue-specific manner (Banerji et al., 1983). A typical enhancer is approximately 50 bp to 1.5 kb in size and contains multiple transcription factor binding sites (TFBS) which are often conserved sequences with a certain degree of degeneracy which transcription factors recognise and bind. Different TFBS are arranged in a particular orientation to control the specificity of the enhancer. However, enhancer

elements *per se* are orientation and distance independent and can be located several kb upstream of the promoter, downstream of the promoter in intronic regions, or at/distal to the 3' end of the gene.

How does an enhancer mediate activation of its corresponding promoter? There currently are several models for its mode of action. Firstly, the proteins bound to enhancers and promoters may interact with each other by looping out the DNA sequence in between (Ptashne and Gann, 1997; Rippe et al., 1995; Vilar and Saiz, 2005). This results in the formation of a multi-protein complex for transcription to occur. Secondly, the enhancer and promoter may not come in contact with one another. Instead, the enhancer may direct the DNA element to localise into specific regions in the nucleus where high concentrations of transcription factors facilitate transcription (Lamond and Earnshaw, 1998). Alternatively, enhancers may act via supercoiling of DNA, nucleosome remodelling and altering chromatin structure to create an accessible structure for recruitment of regulatory proteins to initiate transcriptions (Freeman and Garrard, 1992). This will be discussed in more details in section 1.1.2.5. More recent studies have also demonstrated that RNA polymerase II (PolII) binds to distal enhancers and the PIC is assembled at the enhancer to promote formation of regulatory factor – promoter complexes for transcription (Louie et al., 2003; Spicuglia et al., 2002)

1.1.1.3 Silencers

In contrast to enhancers, silencers result in transcriptional repression rather than activation. Similar to enhancers, they are distance and orientation independent of gene structures. They can be located in the proximal promoter, as part of a distal enhancer, or occur independently in distal regions upstream or downstream of the gene they are regulating. Silencers are bound by repressor proteins to mediate repressions. These repressors may work independently, in cooperation with themselves (Harris et al., 2005) or other repressors (Sertil et al., 2003), or through the binding of a co-repressor (Chen and Evans, 1995).

There are two known mechanisms by which the association of repressors and silencers mediates transcriptional repression. The repressors may localise in the silencers preventing the access of an activator protein to their enhancers (Harris et al., 2005) or by preventing the binding of PolII or other basal transcription factors to the core promoter (Chen and Widom, 2005). Alternatively, the repressors may compete with activators for the same binding site to repress activation (Hoppe and Francone, 1998). Repressors may also recruit chromatin-remodelling enzymes or chromatin modifiers to create a chromatin structure which is unfavorable for the assembly of the transcriptional machinery (Heinzel et al., 1997).

1.1.1.4 Insulators

Insulators function in the genome to prevent genes from being incorrectly transcribed by the regulatory elements of the neighbouring genes. They are typically 500 bp to 3 kb in size. There are two main mechanisms for their function. Firstly, they may be present in the genome to block enhancer activity by inhibiting the interaction of promoters and enhancers (Zhao and Dean, 2004). Secondly, they may act by blocking the spread of repressive chromatin marks into regions containing transcriptionally active genes (West et al., 2002). Insulators are sometimes bound by trans-acting proteins to mediate their function. CTCF is one well-studied example which was found to bind to insulators at the β-globin locus (Bell et al., 1999) and to all known vertebrate insulators.

1.1.1.5 Locus control regions

The first locus control region (LCR) in mammals was discovered in the β-globin locus (Grosveld et al., 1987). LCRs are clusters of *cis*-regulatory elements such as enhancers, silencers and insulators where the collective action of these elements results in the overall control of gene expression. Similar to other *cis*-regulatory elements, LCRs can be located at upstream regions, downstream regions or within the introns of the gene they regulate. However, unlike normal enhancers or silencers, LCRs function in a copy number dependent manner and create an open chromatin structure for linked genes (Li et al., 2002).

1.1.1.6 Scaffold/Matrix attachment regions (S/MARs)

The eukaryotic genome is functionally compartmentalised into chromatin domains by attachment to nuclear matrixes or nuclear scaffolds which are protein-RNA structures within the nucleus. Such chromatin domains define gene transcriptional signatures and insulate the effects from adjacent genes. This is required for various biological functions to take place such as transcription and DNA replication. Scaffold/Matrix attachment regions (S/MARs) are DNA elements in the genome which mediate the attachment of chromatin loops to the nuclear matrix or nuclear scaffold. S/MARs are thus regarded as the borders of chromatin domains which range from 4 kb to 200 kb (Bode et al., 2003). S/MARs can function to insulate genes from any negative effects of the surroundings chromatin (Antes et al., 2001) or to increase transcription initiation rate even in the absence of an enhancer (Bode et al., 2000).

1.1.2 *Trans*-acting proteins involved in transcriptional regulation

In order for transcription to take place, various proteins and *trans*-acting elements are required for the assembly of the complete transcriptional machinery onto the various *cis*-acting elements. These proteins can be summarised as follows.

1.1.2.1 RNA polymerase

Transcription of genes from DNA to RNA is a three-step process involving initiation, elongation and termination. Initiation requires the association of RNA polymerase and general transcription factors to form a pre-initiation complex (PIC) at the promoter regions of genes. In eukaryotes, RNA polymerases are divided into three classes (RNA Pol I, II and III) according to the products they generate. RNA Pol I transcribes DNA to ribosomal RNAs (rRNAs) including the 28S, 18S and 6S subunits. RNA Pol II transcribes DNA to messenger RNA (mRNA) and small nuclear RNAs (snRNAs). RNA Pol III transcribes DNA to transfer RNA (tRNA) and 5S rRNA. The structure and transcriptional machinery of RNA Pol II is the most complicated among the three and the discussion below is focused on RNA Pol II.

The human RNA Pol II comprises 12 subunits, Rpb1 to Rpb12. Rpb1, 2, 3 and 11 have homologous counterparts in bacterial Pol whereas Rpb5, 6, 8, 10 and12 are common in all the three classes. Rpb4, 7 and 9 are unique components of RNA Pol II. Each of these subunits plays specific roles in transcription start site selection, alteration of elongation rate, interaction with activators and stability of RNA Pol (Lee and Young, 2000). Rpb1 contains a carboxyl-terminal repeat domain (CTD) which possesses repeats of the consensus sequence of Tyr-Ser-Pro-Thr-Ser-Pro-Ser. The CTDs are phosphorylated during the switch from initiation to elongation and this phosphorylation is facilitated by protein kinases (Dahmus, 1995).

1.1.2.2 Basal/General transcription factors

The formation of a PIC at the core promoter is a stepwise process which requires assembly of general transcription factors (GTFs) and RNA Pol II. The core promoter is first bound sequentially by TFIID, TFIIA, TFIIB, RNA pol II and TFIIF to form the PIC, followed by binding of TFIIE and TFIIH. This complex then unwinds 12-15 bp of DNA at the transcription start site of the promoter to create an open structure for the formation of mRNAs. Different GTFs interact with the promoter at different regions and have various functions (Table 1.1).

GTF	Functions
TFIID	• Contains two subunits, TATA-binding protein (TBP) and TBP-associated factor (TAF). TBP binds to the TATA box of the core promoter in an orientation-independent manner while TAF binds to INR and DPE and is required for promoter selection and transcriptional activation. • Eukaryotes also encode a TBP-related factor (TRF) which recognises other DNA sequences in TATA-less promoters for the initiation of transcription.
TFIIA	• A heterodimer which interacts with TBP and stabilises the TBP-DNA interaction. • Promotes binding of TFIID to DNA. • Involved in the activation of transcription by binding to activators.
TFIIB	• Interacts with TFIID and RNA Pol II and is required for transcription start site selection. Binds to BRE and downstream sequences of the TATA box. • A direct interacting partner of activators which may promote the recruitment of TFIIB to the promoter.
TFIIF	• A heterodimer containing two subunits, TFIIFβ and TFIIFα. TFIIFβ binds to either upstream or downstream of TATA box while TFIIFα binds to regions downstream of the TATA box. • Binds RNA Pol II tightly and is involved in avoiding non-specific DNA binding and stabilisation of the PIC.
TFIIE	• Binds to DNA sequences directly upstream of the transcription start site after the formation of the PIC. • Promotes the recruitment of TFIIH and stimulates the CTD kinase and helicase activities of TFIIH.
TFIIH	• Contains two subunits, a core subunit with helicase activities (XPD and XPB) and a kinase subunit (Cdk7). The helicase subunit is required in the unwinding of DNA to create an open structure while the kinase subunit phosphorylates the CTD in RNA pol II during the transition to elongation stage.

Table 1.1. General transcription factors and their functions.

1.1.2.3 Sequence-specific transcription factors

RNA polymerase and general transcription factors account for the basal activity of the transcriptional machinery. In order to fully turn on or off the transcription of a gene, sequence-specific transcription factors are required. They bind to *cis*-regulatory regions such as promoters, enhancers and silencers to exert their activation or repression functions. These transcription factors recognise and bind to transcription factor binding sites (TFBS) which are often conserved sequences with a certain degree of degeneracy. Some of the properties of sequence transcription factors are outlined below.

- Modular nature

A sequence-specific transcription factor may be composed of several modules: the DNA-binding module, the dimerisation module, the activation or repression module and the regulatory module. This multi-module property was first observed in the yeast GAL4 transcription factor where the GAL4 protein binds to LexA binding sites through the fusion to the DNA-binding protein LexA (Brent and Ptashne, 1985). Many families of DNA-binding modules have been identified. These include the helix-loop-helix motif which was first discovered in prokaryotes, the homeodomain, zinc finger motif, leucine zipper motif etc (Pabo and Sauer, 1992). Different types of activation modules are present. The activation module can be an acidic or negatively charged alpha helix

(Hope et al., 1988), glutamine or proline-rich regions (Courey and Tjian, 1988; Mermod et al., 1989) or hydrophobic beta sheets (Leuther et al., 1993). In some transcription factors, a regulatory module is required for its activity or sub-cellular localisation. These modules may be present on the same polypeptide or they can be distinct subunits which are detachable and function in a *trans* manner (Baeuerle and Baltimore, 1988).

- Recruitment of coactivators or corepressors

Some sequence-specific transcription factors require the recruitment of coactivators or corepressors by protein-protein interaction to carry out their functions. Some co-factors, such as TAF, act as bridging molecule to bring the sequence-specific transcription factors and the general transcriptional machinery together. Other co-factors, such as chromatin-remodelling factors or histone-modifying enzymes, are recruited to alter chromatin structure, thereby initiating the activation or repression effect.

- Combinatorial effects

Activation or repression by sequence-specific transcription factors is tightly controlled and specific so that transcription of their target genes is regulated in a temporal and/or spatial manner. However, the binding of a single transcription factor may not be sufficient to exert tight regulatory control on the gene of interest. In many cases, clusters of various transcription factor binding sites are located in the *cis*-regulatory element to generate a unique motif for a combination of sequence-specific transcription factors to bind. Such a cluster of transcription factors often function synergistically where the combined activation is greater than with any one factor working alone.

- Posttranslational modification

The activity of sequence-specific transcription factors can also be controlled by post-translational modifications. One example is the phosphorylation of the cyclic AMP response element binding protein (CREB). When this protein is phosphorylated by protein kinase A upon cyclic AMP stimulation, it is activated and initiates the transcriptional activation at the target promoter (Gonzalez and Montminy, 1989). Other examples of post-translational modification include acetylation of p53 which increases its DNA binding affinity (Gu and Roeder, 1997) and ubiquitylation of LexA-VP16 (Salghetti et al., 2001).

- Multiprotein families

One additional property of sequence-specific transcription factors, which provides even more unique and complex regulatory patterns, is that many are members of multiprotein families. Examples of transcription factor family includes the Sp family, the AP-1 family and the GATA

family. Family members are closely related and share the same or very similar DNA binding motifs. In spite of this, they play different roles in transcriptional activation or repression and control the expression of their own set of target genes at certain stage of differentiation or development or in certain cell types. The GATA family of transcription factors is one classical example. The GATA family includes GATA 1→6. They are divided into two sub-families: the haematopoietic sub-family GATA1, 2 and 3 and the non-haematopoietic sub-family GATA4, 5 and 6. GATA1, 2 and 3 are expressed in various haematopoietic and neuronal cell lineage to control lineage commitment and specification whereas GATA4, 5 and 6 are expressed in the heart and digestive organs controlling cardiac-specific gene expression and epithelial cell differentiation in the gut (Ferreira et al., 2005; Molkentin, 2000).

1.1.2.4 Coactivators/ Corepressors

Coactivators and corepressors are important regulators of gene expression although they appear to have no DNA-binding properties. Instead, to exert their function, they interact with other general or sequence-specific transcription factors, and can modify histones/DNA or remodel chromatin.

As mentioned preciously, TBP-associated factors (TAFs) are part of the TFIID complex. Although some TAFs may bind to promoter DNA directly, others may bind to activators and general transcription factors transmitting information between the two. Examples are TAFII40 and TAFII60 which act as bridges between the p53 activator and the initiation complex (Thut et al., 1995).

Mediators, another class of coactivators, first identified in yeast, are multisubunit complexes which activate transcription stimulating the phosphorylation of CTD of RNA pol II. They also interact with activators and transmit positive or negative signals to the promoter (Myers and Kornberg, 2000). Seven mediator subunits have been discovered in human so far.

Certain coactivators or corepressors act as docking molecules on activators or repressors. Instead of having intrinsic enzymatic activities, they recruit other necessary factors for binding to the initiation complex or chromatin remodelling factors for transcriptional activation or repression. One example is OCA-B, a coactivator of the activator octamer binding protein (OCT). It recruits some TAFs for the activation of immunoglobin genes in B cells (Wolstein et al., 2000). Another example is the nuclear receptor corepressor (NcoR) which recruits histone deacetylases (HDACs) to produce an inactive chromatin structure to repress expression of nuclear receptors (Privalsky, 2004).

Some co-factors may possess both activating and repressive functions. Friend of GATA1 (FOG-1), a cofactor of GATA1, can promote or inhibit transcription by directly recruiting histone acetyl transferases (HATs) or histone deacetylases (HDACs) at specific sites (Letting et al., 2004). FOG-1

also functions as a chromatin occupancy facilitator, a possibly new class of cofactor, where it facilitates the binding of GATA1 to sites originally bound by GATA2 (Pal et al., 2004).

1.1.2.5 Chromatin modifying factors

Epigenetic regulation by modification of chromatin plays a crucial role in regulating gene expression. Nucleosomes are the basic subunits of chromatin where DNA is packaged with histone proteins. The core histone proteins H2A, H2B, H3 and H4 bind to one another to form a protein octamer wrapping the DNA whereas the linker histone H1 binds to the outside of the nucleosome which stablisises the folding of the nucleosome. Nucleosomes have dynamic properties which are governed by a specific class of co-factors - chromatin modifying factors - which include chromatin remodelling complexes and histone modifying enzymes. These co-factors modify the structure of chromatin to facilitate or interfere with the recruitment of PICs and transcription factors to promoter regions or other regulatory elements.

A. Chromatin-remodelling complexes

There are at least five families of chromatin-remodelling complexes in eukaryotes: SWI/SNF, ISWI, NURD/Mi2, INO80 and SWRI families (Saha et al., 2006). All families contain an ATPase subunit where they use ATP-hydrolysis to modify chromatin structure and remodel nucleosomes. Other subunits in the complex may be involved in the modulation of ATPase activity and the targeting to specific regions of chromatin.

Two mechanisms by which chromatin-remodelling complexes function to modify chromatin structure to increase accessibilities of nucleosomal DNA have been described. The sliding of DNA with respect to the histone proteins is the most widely studied mechanism (Meersseman et al., 1992). The result is that the histone octamer is re-positioned to interact with different DNA elements instead of the original DNA elements. Another possible mechanism of nuclesomal re-positioning involves conformational changes (Lorch et al., 1999; Studitsky et al., 1994). Such conformation changes may result in the collapse of the altered nucleosome to a canonical nucleosome in contact with a different DNA segment. It is also possible that the histone proteins are released and interact with a new segment of DNA following conformation change.

B. Histone-modifying enzymes

Histone modifying enzymes promote the covalent modifications of the histone proteins. These covalent modifications include acetylation, methylation, phosphorylation, ubiquitylation, ADP ribosylation, sumoylation and isomerisation. Such modifications either affect the higher-order chromatin structure by disrupting histone-DNA interactions or recruit chromatin remodelling complexes and other proteins.

Histone acetylation was the first post-translational modification identified on histone proteins (Allfrey et al., 1964). This modification is characterised by the addition of an acetyl moiety to the ε-amino group of the lysine residue and is associated with transcriptional activation. Acetylation is a dynamic and reversible process controlled by histone acetyltransferases (HATs) and histone deacetylases (HDACs).

HATs are categorised into three families: the Gcn5-related N-acetyltransferase (GNAT) superfamily, the MYST family and the p300/CBP family. HATs do not bind DNA directly and they are usually components of multiprotein complexes which are recruited to promoter regions or other regulatory regions by interaction with DNA-bound activators (Utley et al., 1998). These complexes contain several subunits which carry out distinct functions including interaction with different types of transcription factors. In addition to the modular nature of HAT complexes, the combination of various components in the complexes also dictates the recognition site specificity (Grant et al., 1999).

While HATs confer transcriptional activation, HDACs correlate with repression. There are three classes of HDACs: class I, II and III. Class I includes HDAC 1, 2, 3 and 8 while class II includes HDAC 4, 5, 6, 7, 9 and 10. Members of these two classes share sequence similarities and require Zn^{2+} for their function. Class III HDACs are also called the Sir2 family and includes SIRT1-7. They have low amino acid sequence homology to class I and II and they are nicotinamide adenine dinucleotide (NAD^{+})-dependent. Similar to HATs, HDACs are found in protein complexes. For example, both HDAC 1 and 2 are found in the Sin3, NuRD and CoREST complexes which contain other subunits required for protein-protein interaction and chromatin remodelling.

Histone methylation, including the addition of methyl group in a mono-, di- or tri- manner, has been shown to occur at both lysine and arginine residues. Methylation of lysine on histone subunits H3 and H4 is catalysed by histone methyltransferases (HMTs) which share the common catalytic 130-amino-acid SET domain, except in the modifier Dot1 (Rea et al., 2000). Methylation of arginines is catalysed by the protein arginine methyltransferases (PRMT) family. Members of this family share a highly conserved core AdoMet binding region which forms the protein substrate binding cleft and has methyltransferase activity (Lee et al., 2005b).

The lysine demethylase (LSD1) demethylates H3K4 mono- or di-methylation by means of an amine oxidase reaction and mediates transcriptional repression (Shi et al., 2004). LSD1 associates with Co-REST, a transcriptional co-repressor, to demethylate nucleosomal substrates (Shi et al., 2005). LSD1, when present in an androgen receptor complex, also demethylates H3K9 methylation and activates transcription (Metzger et al., 2005). A distinct class of lysine demethylases is the Jumonji C (JmjC)-domain-containing family where the JmjC domain is the core catalytic domain. A number

of members have been identified for this class and they target different lysine residues. Reversal of arginine methylation involves deimination which is the process of converting a methyl-arginine to citrulline by the enzyme peptidylarginine deiminase 4 (PADI4) (Cuthbert et al., 2004; Wang et al., 2004b). This antagonises the effect of arginine methylation. However, only mono-methylated arginine residues have been demonstrated to undergo deimination.

Phosphorylation of histone H3 subunit at serine 10 has been shown to be associated with transcriptional activation of the immediate early genes in human such as *c-jun* (Mahadevan et al., 1991). MSK1/2 and RSK2 kinases mediate this phosphorylation function (Sassone-Corsi et al., 1999; Thomson et al., 1999).

Ubiquitylation of histone subunits involve the addition of a 76-amino-acid ubiquitin protein. H2A is ubiquitylated at a lysine residue by the Bmi/Ring1A-containing human Polycomb repressive complex 1-like complex (hPRC1L) which mediates transcriptional repression (Wang et al., 2004a). H2B lysine ubiquitylation is catalysed by RNF20/RNF40 and UbcH6 in human (Zhu et al., 2005) and by Rad6 and Bre-1 in yeast (Robzyk et al., 2000; Wood et al., 2003). De-ubiquitylation of H2B is carried out by Ubp8 in the SAGA or SILK complexes, while de-ubiquitylation of H2A requires 2A-DUB (Daniel et al., 2004; Zhu et al., 2007).

Other modifying enzymes are less well characterised. Mono-ADP-ribosyltransferases (MARTs) and poly-ADP-ribose polymerases (PARPs) mediate ADP ribosylation of histones (Hassa et al., 2006). FPR4 isomerises a proline residue on the H3 subunit (Nelson et al., 2006) which in turn regulates methylation of the proline residue. Sumoylation is characterised by the conjugation of the SUMO protein to its histone substrate substrate by Ubc9 and is linked to transcriptional repression by antagonising acetylation and ubiquitylation (Johnson, 2004).

1.1.3 The transcription factory

Co-ordinated gene expression patterns require a combined effort of various transcription factors and chromatin modifiers to direct gene expression at various loci at certain developmental time point or in response to external stimuli. The previous sections described the transcriptional machinery at the molecular level but indeed such co-ordinated gene expression regulation also requires chromosomal organisation in a three-dimensional space of the nucleus. It has been suggested that active genes are repartitioned into nuclear territories for transcription to take place (Chambeyron and Bickmore, 2004; Williams et al., 2006). Looping and intra- or inter-chromosomal interactions between regulatory loci or active genes have also been documented to provide integrated expression (Spilianakis et al., 2005; Zhao et al., 2006). A more recent study illustrated that these kinds of

chromosomal re-organisation require induction by ligands and is facilitated by co-activators and components of the chromatin-remodelling complexes (Nunez et al., 2008).

1.2 Transcriptional regulatory networks

Studying gene regulation at a single gene locus fails to give a full picture of global regulatory patterns - genes across the genome interact with proteins, through time and space, within the cell to control their expression. Transcriptional regulatory networks are the programmes of multiple interactions within cells including transcription factor-DNA interaction and other factors that modulate these interactions biochemically to control the expression of genes. Such networks are crucial in dictating cellular behaviours in response to specific signals or at different stages of development.

1.2.1 Properties of a transcription network

Transcriptional regulatory networks, like other biological networks, consist of nodes which are connected by edges (Figure 1.3). Nodes include various transcription regulator proteins such as transcription factors (TF), co-factors and chromatin regulators and various DNA elements such as promoters and enhancers. Edges are the physical interactions between regulator proteins (protein-protein interactions) and between regulator proteins and DNA elements (protein-DNA interaction).

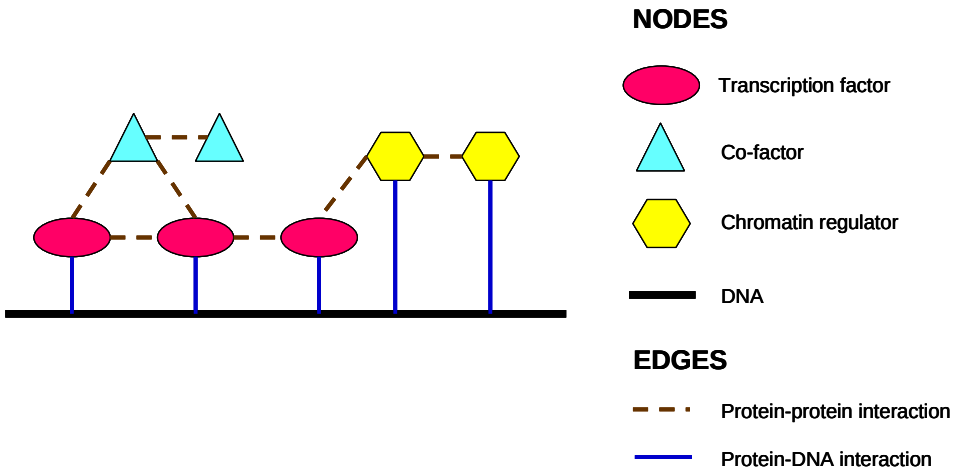


Figure 1.3. Nodes and edges of a transcription network. Schematic diagram shows the nodes and edges of a transcription network. Nodes include transcription factors (pink oval), co-factors (aqua triangle), chromatin regulator (yellow hexagon) and DNA (black line). Edges describe the relationships among the nodes which include protein-protein interactions (brown dotted line) and protein-DNA interaction (blue solid line).

How nodes and edges are related is key to understanding a transcriptional network. Network motifs are small networks with distinct properties, which in combination define the genetic control of the transcription programmes (Lee et al., 2002; Milo et al., 2002; Shen-Orr et al., 2002) (Figure 1.4). An autoregulation motif consists of a transcription factor binding to its own promoter to stimulate

expression. This ensures the stability of its own expression. In contrast, a multi-component loop motif involves more than one transcription factor binding to the promoters of one another to regulate expression. In a feed-forward loop motif, a transcription factor regulates the expression of another transcription factor while both of these transcription factors regulate the expression of a common target gene. A slight modification in the level of a master regulator can result in a significant increase or decrease in the target expression due to the presence of the second regulator which is under the control of the master regulator. A single input motif contains one transcription factor which co-regulates a number of target genes and this often ensures a co-ordinated expression pattern for a certain subset of genes. A multiple input motif involves a set of transcription factors binding to the promoters of the same set of target genes. This allows the expression of the targets to be co-ordinated in response to different signals which stimulates or inhibits expression of the regulator transcription factors. A dense overlapping region utilises a set of transcription factors that overlap to regulate a set of targets where each of these targets is regulated by a different transcription factor combination. A regulator chain motif consists of 3 or more regulator transcription factors in a series where the first transcription factor regulates the second transcription factor, which in turn regulates the third.

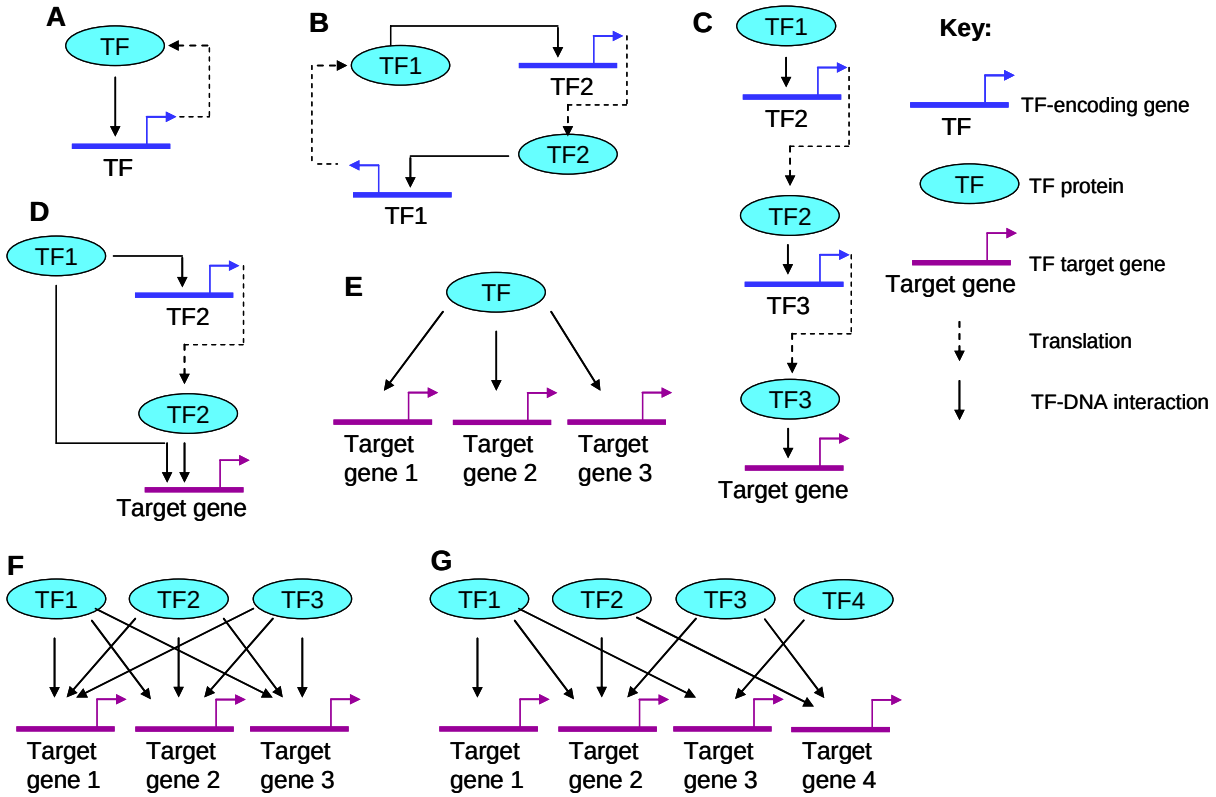


Figure 1.4. Network motifs of transcriptional regulatory networks. Combinations of these motifs regulate expression patterns in a transcriptional network. Black solid arrows indicate TF-DNA interactions and transcription factor regulation while black dotted arrows indicate translation of proteins. A: autoregulation motif. B: multi-component loop. C: regulator chain. D: feed forward loop. E: single input motif. F: multiple input motif. G: dense overlapping region. These motifs are explained in detail in the text.

Many studies have been performed to delineate transcriptional regulatory networks in various biological systems in eukaryotes. For example, in *Saccharomyces cerevisiae*, a comprehensive network of regulator-gene interactions of all known regulators has been determined (Lee et al., 2002). More specifically, yeast has been widely used in the study of transcriptional networks controlling the cell cycle (Lee et al., 2002; Oliva et al., 2005). In mouse, the Sonic hedgehog (Shh)-driven Gli-mediated transcriptional network defining neuronal development and specification has been characterised (Vokes et al., 2007).

1.2.2 Experimental and computational approaches for deciphering regulatory networks

To delineate transcriptional regulatory networks, or aspects thereof, a combination of experimental and computational approaches is required (Figure 1.5). The feasibility of such approaches have been demonstrated most effectively in model organisms such as *Saccharomyces cerevisiae* (Tavazoie et al., 1999) and *Caenorhabditis elegans* (Horner et al., 1998). Methods, such as expression profiling with microarrays (see section 1.3.2), which allow us to visualise the effects of perturbing a particular transcription factor in a biological system, facilitate the identification of direct and secondary target genes co-regulated by a transcription factor, and at the same time provide information about the mode of regulation (i.e. activation or repression). These data enable us to search for the common regulatory elements (e.g. transcription factor binding sites) in the co-regulated genes by computational methods. Further confirmation of the transcription factor binding sites at direct target genes can be achieved by chromatin immunoprepitation (ChIP), while ChIP combined with microarray (ChIP-chip) can be used to map the transcription factor binding sites on a genome-wide scale (Iyer et al., 2001; Ren et al., 2000). Direct and secondary targets of transcription factors can also be distinguished in this way. Chromatin structure affects the binding of transcription factors to regulatory sequences (Hassan et al., 2001; Jenuwein and Allis, 2001). Therefore, the study of chromatin structure and its biochemical modifications is crucial to understanding the complete picture of gene regulation (Lieb et al., 2001). Furthermore, chromosome correlation maps for the chromosomal locations of co-regulated genes can be generated which often show that genes co-localised to specific genomic regions are in open chromatin structures enabling active transcription of the region (Cohen et al., 2000). The combination of all or a subset of these approaches provides insights into the regulatory networks in biological systems (Shannon and Rao, 2002).

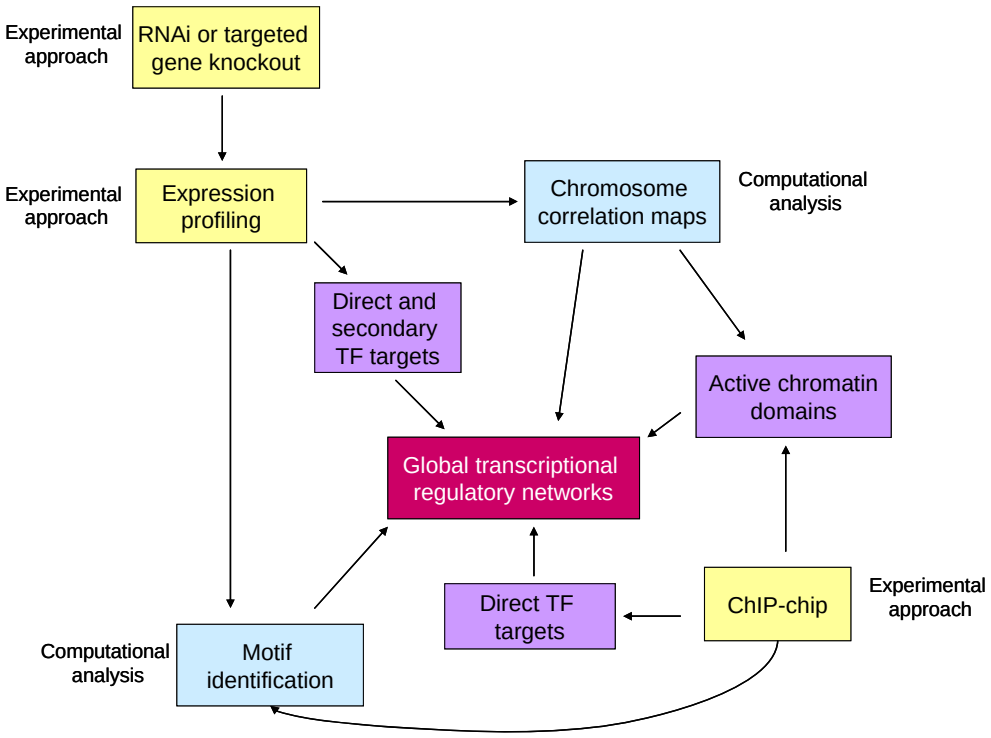


Figure 1.5. Experimental and computational approaches to delineate transcriptional regulatory networks. A combination of methods can be used to identify novel relationships between transcription factors and co-factors with their target genes. Yellow box: experimental approach; blue box: bioinformatics approach; purple box: experimental or informatic outcome. Expression profiling data obtained from gene knockdowns or knockouts enable us to identify direct and indirect targets, generate chromosome correlation maps and search for regulatory motifs. ChIP-chip generates genome-wide data on global TF-DNA binding and histone modifications. These data combined together provide insights into global transcriptional networks.

1.3 Experimental and computational approaches to understand transcriptional regulation

1.3.1 Gene perturbation by RNA interference

Perturbation of the activity of transcription factors has been widely used to study their function and to identify downstream target genes involved in transcriptional programmes. Traditionally, complete knockouts of the gene of interest provide the cleanest experimental paradigm to study. However, generating knockouts are time-consuming and can often result in lethality. With the discovery and advances in RNA interference, transient or stable gene knockdowns can be induced in the cell type of interest and are relatively rapid and inexpensive approaches for the delineation of downstream target genes of transcription factors.

1.3.1.1 Discovery and mechanism of RNA interference

Introduction of double-stranded RNA (dsRNA) was first found to silence genes with complementary sequences in *Caenorhabditis elegans* and has been termed RNA interference or

RNAi (Fire et al., 1991; Fire et al., 1998). Such silencing machinery by dsRNAs was first described as an anti-viral response to protect the organism from RNA viruses and the random integration of transposable elements (Waterhouse et al., 2001). The underlying molecular mechanism of RNAi involves two main steps. dsRNAs are processed into short interfering RNAs (siRNAs), which are about 22 nucleotides in length, by the RNase III enzyme Dicer. These mature siRNAs then associate with various proteins including the Argonaute protein family to form the RNA-induced silencing complex (RISC), where the siRNAs unwind. RISC then uses the unwound strand as the guide which identifies the substrates (Figure 1.6). Subsequent gene silencing occurs at various levels. At the post-transcriptional level, the identification of the target mRNA by the guide siRNA may trigger mRNA degradation by first cleaving the target mRNA. The mRNA cleavage requires siRNA and mRNA base-pairing together with the Argonaute protein which contains an RNaseH-like domain and all the critical active residues for endonucleolytic cleavage (Meister et al., 2004). The resultant cleaved mRNA fragments are directed to the general cellular mRNA degradation pathway which deadenylates the mRNA followed by 3' to 5' or 5' to 3' degradation. Post-transcriptional repression by RNAi can also be achieved by the inhibition of protein translation. In addition to post-transcriptional suppression of gene expression, RNAi is also implicated in silencing at the transcriptional level. siRNAs targeted to the promoter regions of genes can induce transcriptional silencing by DNA methylation in human cells (Kawasaki and Taira, 2004; Morris et al., 2004). However, the mechanism by which siRNAs enter the nucleus for DNA methylation remains unknown. siRNAs can also methylate histone H3 lysine 9 and 27 recruiting chromatin-remodelling complexes such as Mi2/NuRD and Sin3/HDAC resulting in the condensation of chromatin and transcriptional repression (Kawasaki and Taira, 2004; Weinberg et al., 2006).

In plants and *Caenorhabditis elegans*, the RNAi effect can be amplified through the mechanism of transitive RNAi (Sijen et al., 2001). siRNAs targeting the 3' end of a transcript results in the suppression of the mRNA and further production of siRNAs against the same region. siRNAs against sequences upstream of the original targeted region are also generated. Therefore, the RNAi effect is significantly enhanced even with the introduction of minute amount of exogenous dsRNA. Such amplification requires the plant RNA-directed RNA polymerase (RdRP) or the *C. elegans* homologue EGO1 which employ the target mRNA as a template (Schiebel et al., 1998; Smardon et al., 2000). However, this amplification system has not been demonstrated in mammalian systems, where no RdRP homologue has yet been identified.

RNAi also takes place in the endogenous gene silencing machinery using microRNAs (miRNAs). miRNAs are 21 to 23-nucleotide RNA duplexes which are transcribed by miRNA genes and have less than complete complementarities to their targets. Primary microRNAs (pri-miRNAs) are first

processed to form pre-miRNAs by the enzyme Drosha and pre-miRNAs then enter the RNAi pathway (Figure 1.6). The first miRNA, *lin-4*, was discovered in *C. elegans* and was found to control the timing of various stages of larval development by blocking translation of the protein LIN-14 (Lee et al., 1993; Olsen and Ambros, 1999; Wightman et al., 1993). Since then, many more miRNAs have been discovered in invertebrates and mammals and these have been shown to be critical during developmental timing, cell proliferation, differentiation and apoptosis, and signalling pathways. Genomic rearrangements resulting in altered expression of miRNA genes and/or changes in miRNA target sites have also implicated in cancer and other diseases (Kloosterman and Plasterk, 2006).

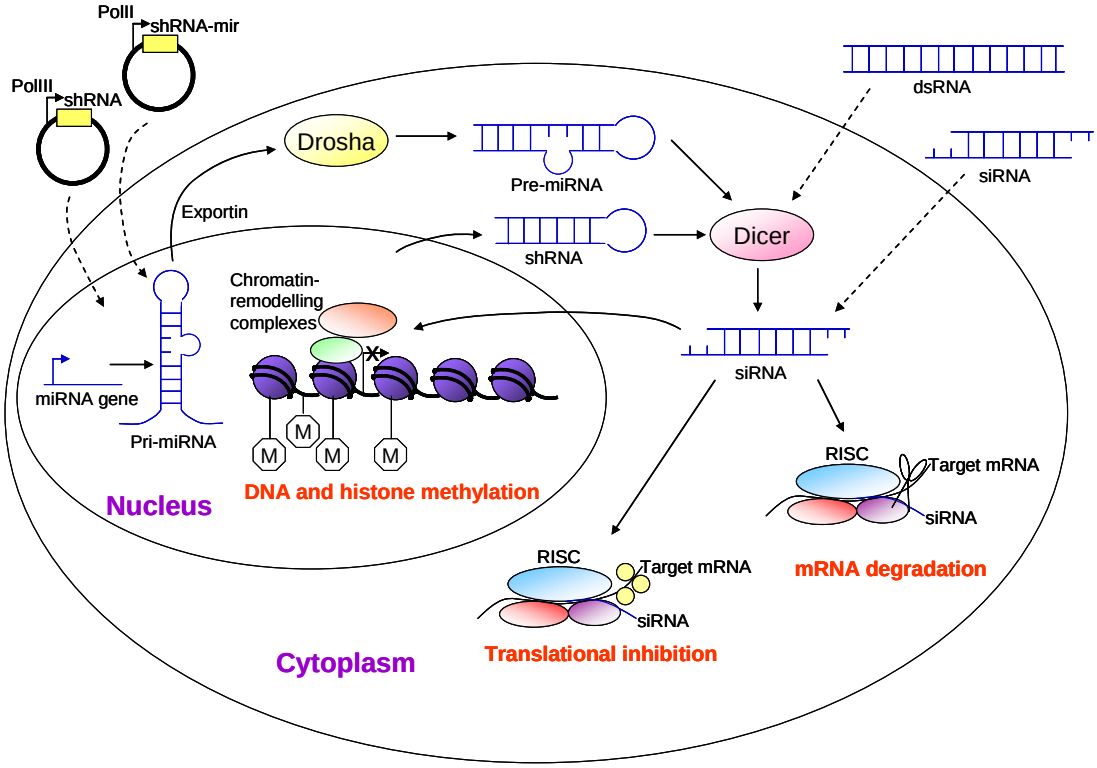


Figure 1.6. The mechanism of RNAi and various ways of triggering RNAi by exogenous sources. In addition to endogenous micro-RNA (miRNA), introduction of exogenous double-stranded RNA (dsRNA), short interfering RNA (siRNA) or plasmids containing short hairpin RNA (shRNA) or short hairpin RNA mir (shRNA-mir) genes all trigger RNA interference in the cell. Transcription of endogenous miRNA genes and exogenous shRNA-mirs inside the nucleus generates pri-miRNAs which are exported to the cytoplasm by exportin and further processed into pre-miRNA by Drosha. Pre-miRNAs, exogenous dsDNAs and shRNAs generated by the exogenous shRNA gene are then processed into siRNA by Dicer. Endogenously-produced siRNAs or exogenous siRNAs incorporate into the RNA-induced silencing complex (RISC) and mediate gene silencing by mRNA degradation, translational inhibition or DNA and histone methylation. Key: dotted arrows: introduction of RNAi triggers outside the cell; ovals: various proteins; dark blue circles inside nucleus: histones; M: methylation.

1.3.1.2 Inducing RNAi in various organisms

A. Invertebrate systems

Inducing RNAi in invertebrate systems such as *C. elegans* and *Drosophila* is relatively straightforward experimentally. Injecting small amounts of long dsRNAs of over 500 bp into the tail of *C. elegans* was first described to induce gene-specific silencing throughout the entire organism and the knockdown was persistent in the progeny (Fire et al., 1998). Similar methods of injection was also described in *Drosophila* (Kennerdell and Carthew, 1998). The spreading effect of RNAi across cell boundaries and inheritance through progeny is mainly due to the ability of *C. elegans* to amplify siRNAs. However, microinjections require expensive equipment and expertise. Simply feeding the worms with *Escherichia coli* expressing the dsRNAs was also found to induce repression in gene expression (Timmons and Fire, 1998). Improvements in the feeding method using a strain of *E. coli* deficient for RNaseIII and engineered to produce high quantities of specific dsRNAs when fed to *C. elegans* resulted in knockdown phenotypes comparable to complete gene knockouts (Timmons et al., 2001). Soaking the worms in solutions containing dsRNA was also demonstrated to be an alternative method of delivery (Tabara et al., 1998). Stable integration of inducible dsRNA-expressing constructs have been developed in *C. elegans* and *Drosophila* embryos and cultured cells where mutant lines can be maintained through multiple generations (Clemens et al., 2000; Kennerdell and Carthew, 2000; Tavernarakis et al., 2000).

B. Mammalian systems

The use of long dsRNA of over 30 bp in mammalian cells to silence genes was found to trigger the innate immune response (Williams, 1997). The enzyme dsRNA-dependent protein kinase (PKR) is activated on binding to long dsRNA, which results in the sequence-independent destruction of all RNAs and generalised repression of protein synthesis. This results in non-specific repression of gene expression within the cell. Therefore, other approaches have been used to induce gene-specific RNAi responses in mammalian cells (Figure 1.6). Each of these is described below and the advantages and disadvantages are summarised in Table 1.2.

(i) siRNAs for RNAi

Introduction of short interfering RNA (siRNA) (shorter than 30 base pairs) into mammalian cells, was found to induce the sequence-specific RNAi pathway (Caplen et al., 2001; Elbashir et al., 2001). These siRNAs are short duplexes of approximately 19 nucleotides in length with 2-nucleotide 3' overhangs on each strand. They bypass the Dicer processing step and enter the RNAi pathway by directly incorporating into the RISC complexes. siRNAs can be synthesised with or without chemical modification to increase their stability and specificity. Alternatively, they can be

generated by *in vitro* transcription of the target cDNA followed by cleavage by recombinant Dicer or bacterial RNase III (Myers et al., 2003; Yang et al., 2002). Dicer cleavage produces siRNA in its natural form for entry into RISC complex. It has been suggested that enzymatically prepared siRNAs can dramatically reduce the off-target effects (Kittler et al., 2007). The silencing effect of siRNA is dependent upon the transfection efficiency and the amount of siRNA used.

(iii) shRNAs for RNAi

Expressing short hairpin RNA (shRNA) from RNA polymerase III promoters in plasmid or viral-based vectors is an efficient way of silencing target genes. shRNAs are produced as single-stranded 50-70 nucleotides molecules which form stem-loop structures. The shRNA mimics the endogenous microRNA (miRNA) pathway to trigger the cleavage of shRNAs generating siRNA for the silencing of specific genes (Brummelkamp et al., 2002; Paddison et al., 2002). The shRNA-encoding DNA fragments can be made by chemically synthesising 50 to 70-nucleotide long oligonucleotides which can be annealed and cloned into a vector. Alternatively, they can be generated by PCR-based methods along with restriction enzyme digestion, which greatly enhance the efficiency of construct generation for RNAi screen (Gou et al., 2003; Sen et al., 2004).

(iii) shRNA-mirs for RNAi

shRNA-mirs are optimised sequences having miRNA-like properties. They are generated by miRNA precursors as the backbone for delivery of hairpin loops flanked by stem sequences found in miRNAs. shRNA-mirs have been demonstrated to successfully induce transient (Zeng et al., 2002) and stable gene knockdowns (Boden et al., 2004b; Dickins et al., 2005). The silencing effects produced by shRNA-mirs are significantly higher than for conventional shRNAs (Boden et al., 2004b; Silva et al., 2005). In some cases, even a single-copy integration can generate potent and stable knockdown (Dickins et al., 2005). This is particularly important for RNAi-based analyses as a reduction in concentration of siRNAs generated *in vivo* can lower the off-target effects (see section 1.3.1.3 B).

RNAi system	Advantages	Disadvantages
siRNA	• Efficient delivery methods available • Pre-validated siRNAs available	Transient silencing due to lack of cellular means to amplify and propagate siRNAs and dilution by actively dividing cells
shRNA	• Stable integration • Expression of shRNAs driven by inducible or constitutive promoter system (Gupta et al., 2004; Matsukura et al., 2003; van de Wetering et al., 2003)	Time and labour consuming in order to carry out the cloning of the constructs and the screening for stable transfectants.
shRNA-mir	• Transient and stable silencing • Stronger silencing effect	Comparatively less well-characterised technique than the other two methods

Table 1.2. Advantages and disadvantages of various RNAi systems in mammalian systems.

(iv) Delivery strategies

Various strategies have been developed to deliver siRNA or shRNA into mammalian cell types and they are summarised in Table 1.3.

Delivery strategy	Remarks	References
Electroporation of siRNA	Transient gene silencing	(MacKeigan et al., 2005)
Lipid-based transfection of siRNA	Transient gene silencing	(Aza-Blanc et al., 2003)
Influenza virosomes with encapsulated siRNA	Transient gene silencing	(de Jonge et al., 2006)
Electroporation of shRNA	Transient gene silencing	(Brummelkamp et al., 2003)
Transfection of shRNA-expressing plasmid vector	Generates stable RNAi by random integration and marker selection	(Brummelkamp et al., 2002)
Delivery of shRNA cassettes using retroviral systems	Used in cell types which are difficult to transfect (e.g. primary cells)	(Barton and Medzhitov, 2002)
Delivery of shRNA cassettes using lentiviral systems	Used in primary cells and non-dividing cells	(Moffat et al., 2006; Ngo et al., 2006; Stewart et al., 2003)
Delivery of shRNA cassettes using adeno-associated virus	Stable integration	(Boden et al., 2004a)
Delivery of shRNA cassettes using adenovirus	Stable integration	(Cao et al., 2005)

Table 1.3. Delivery strategies of RNAi in mammalian systems.

1.3.1.3 Non-specific effects of RNAi

Although PKR activation is not effectively triggered by siRNAs in mammalian systems, other non-specific effects are induced by RNAi triggers. These include the innate immune response (IFN response), off-targeting and saturation of the RNAi pathway as described below.

A. IFN response

dsRNAs which are longer than 30 nucleotides were found to trigger the PKR response in mammalian cells (Williams, 1997). However, it has also been demonstrated that transfection of siRNAs can activate PKR which results in the triggering of the interferon (IFN) pathway and induces a global upregulation of IFN-stimulated genes (ISGs) (Sledz et al., 2003). Some upregulated ISGs are dependent on siRNA concentration while others are not. siRNAs without 2- to 3-nucleotide 3' overhangs were shown to be recognised by the IFN system via the RNA helicase RIG-1 (Marques et al., 2006). These overhangs are the structural characteristics which distinguish synthetic siRNAs from endogenous Dicer-generated ones. siRNAs without 3' overhangs are more likely to be unwound, and this mediates IFN activation. The IFN response is also induced by H1 or U6 promoter-generated shRNAs *in vivo* (Bridge et al., 2003; Pebernard and Iggo, 2004). More detailed analyses of the U6 promoter vectors indicated that ISG induction is a consequence of the presence of an AA dinucleotide motif near the transcription start site of shRNAs.

B. Off-target effects

Off-target effects were first studied by examining the expression profiling of numerous siRNAs directed against the same target genes (Jackson et al., 2003). It was shown that a majority of gene expression patterns were siRNA-specific rather than target-specific. Off-target effects can be elicited by as few as 11 nucleotides of identity between the siRNA and its target. In other studies, different siRNAs against the MEN1 gene were characterised to induce variations in expression levels to different degree in p53 and p21, which are indicators of overall changes in cellular physiology (Scacheri et al., 2004). Off-target effects were further characterised by the ability of various siRNAs to induce changes in cell toxicity in a target-independent manner which generates toxic phenotypic changes (Fedorov et al., 2006). This toxic effect was found to relate to a UGGC motif in the siRNAs. Additional studies have been performed to understand the mechanism of off-target effects. It has been confirmed in various studies that the off-target transcripts have 3' UTR sequence partial complementary to the seed region of the siRNA which is similar to the endogenous miRNA pathway (Birmingham et al., 2006; Jackson et al., 2006).

C. Saturation of the RNAi pathway

Saturation of the RNAi pathway happens when there is an excessive dose of siRNA administered to cells or when shRNAs are highly expressed. Saturation can occur at different levels, depending on the siRNA/shRNA used. Depletion of Dicer and Ago2 was found to up-regulate expression of a large number of genes whose 3' UTRs show an enrichment of putative miRNA target sites (Schmitter et al., 2006). shRNAs or shRNA-mirs may also saturate Drosha or Exportin. Prolonged expression of shRNAs was found to be lethal in mice due to a saturation of Exportin 5 (Grimm et al., 2006).

1.3.1.4 Applications of RNAi

A. Study of gene functions and downstream pathways

RNAi technology has been used to knockdown the expression of specific genes which are of particular interest in certain pathways and diseases. In many cases, it has been used in conjunction with gene expression profiling to identify downstream target genes in signalling pathways (Jazag et al., 2005) and to understand disease mechanisms (Diakos et al., 2007). RNAi, together with expression analyses on microarray and computational or experimental promoter studies, have also been use to dissect transcriptional networks of key transcription factors involved in apoptosis (Elkon et al., 2005) and in embryonic stem cell self-renewal (Jiang et al., 2008).

B. Genetic screens

With the completion of the genome sequence of human and various model organisms, RNAi has been exploited as a tool to screen genes involved in specific pathways or disease. In contrast to the application above, a number of genes in the genome included in an RNAi library are knocked down simultaneously and their effects on specific pathway are studied to identify the genes which are important for the pathway. RNAi-based genetic screens have advantages over conventional knockouts as generating RNAi libraries is relatively cheap and easy. Genetic screens in *C. elegans* and *Drosophila* are summarised in Table 1.4.

Biological pathway studied	Organism	Genome coverage	Reference
Cell division	*C. elegans*	Chromosome III genes	(Gonczy et al., 2000)
DNA-damage responses	*C. elegans*	Whole genome	(van Haaften et al., 2006)
Embryonic developments	*C. elegans*	Whole genome	(Sonnichsen et al., 2005)
miRNA pathway	*C. elegans*	Whole genome	(Parry et al., 2007)
Hedgehog signalling pathway	*D. melanogaster*	Kinases and phosphatase; 43% predicted genes	(Lum et al., 2003)
Cell viability and growth	*D. melanogaster*	Whole genome	(Boutros et al., 2004)
Embryonic development	*D. melanogaster*	Whole genome	(Koizumi et al., 2007)
Chromatin-related transcriptional repression	*D. melanogaster*	Whole genome	(Stielow et al., 2008)

Table 1.4. RNAi genetic screen in *C. elegans* and *Drosophila*. This table summarises the biological pathways studied and the genome coverage of library in *C. elegans* and *Drosophila* RNAi genetic screens.

Genetic screens have also been widely used in mammalian systems (Table 1.5). Due to the size of mammalian genomes and their gene content, initial efforts of RNAi screening focused on libraries representing subsets of genes implicated in various processes and pathway. However, siRNA libraries representing the known human and mouse gene sets are now commercially available from several suppliers such as Ambion, Qiagen and Dharmacon. For stable integration of shRNAs, retroviral or lentiviral libraries targeting either subsets or all human genes have been generated and successfully employed in RNAi screens. Most conventional methods of RNAi screening involve the use of a single well/single gene approach which is relatively time-consuming. A small-scale pooled retroviral vector strategy involving the use of barcoded shRNAs and analyses on microarray was first described by Berns et al (2004) and has since been widely used. This greatly enhances the efficacy of global screening. Second generation plasmid-based shRNA-mir libraries covering all genes in the human and mouse genomes have also been described (Silva et al., 2005). Such libraries enable single-copy expression of the shRNAs which is important for pooled screening applications.

Biological pathway or disease studied	Form of RNAi trigger	Source of siRNA/ delivery of shRNA	Library genome coverage	Analysis strategy/ screen format	Reference
Negative regulation of phosphorylation	siRNA	Chemically synthesised	Phosphoinositide 3-kinase (PI3K) pathway human genes	Single well/ single gene	(Hsieh et al., 2004)
Cell division	siRNA	Endoribonuclease-prepared	> 15000 human genes	Single well/ single gene	(Kittler et al., 2004)
Huamn immunodeficiency virus (HIV) infection required host protein	siRNA	Chemically synthesised	All genes in the human genome (Dharmacon library)	Single well/ single gene	(Brass et al., 2008)
p53 pathway	shRNA	Retroviral delivery	~ 8000 human genes	Single well/ single gene and bar-coded siRNA + microarray screen	(Berns et al., 2004)
RAS activation pathway and tumorigenesis	shRNA	Retroviral delivery	~ 4000 human genes	Soft agar colony growth	(Kolfschoten et al., 2005)
Tumorigenesis	shRNA	Retroviral delivery	~ 7500 human genes	bar-coded siRNA + microarray screen	(Westbrook et al., 2005)
Diffused large B-cell lymphoma	shRNA	Retroviral delivery (Inducible expression)	~ 2500 human genes	bar-coded siRNA + microarray screen	(Ngo et al., 2006)
Cell division and proliferation	shRNA	Lentiviral delivery	~ 12000 human genes	Single well/ single gene	(Moffat et al., 2006)
Cell proliferation in mammary cells	shRNA-mir	Retroviral delivery	All genes in the human genome	bar-coded siRNA + microarray screen	(Silva et al., 2008)
Cell proliferation in cancer cells	shRNA-mir	Retroviral delivery	All genes in the human genome	bar-coded siRNA + microarray screen	(Schlabach et al., 2008)

Table 1.5. RNAi genetic screens in mammalian systems. This table summarises the diseases or biological pathways studied using RNAi screen in mammalian systems. The type and delivery of the RNAi trigger, library coverage and analysis methods are also described.

C. Disease therapy

In addition to using RNAi as an experimental tool, numerous studies have documented applying RNAi technology therapeutically in the treatment of various human diseases. A variety of RNAi triggers and delivery methods have been tested and summerised in Table 1.6. Ultimately, the choice of method of RNAi therapy depends on the disease and organ under treatment. For instance, siRNAs can be degraded by serum nucleases and can only provide a short-term suppression of gene

expression in a specific subset of tissues/organs, thus limiting their therapeutic benefits. Although shRNA delivery is more challenging, it initiates a more sustained therapeutic effect and provides treatment options to a broader range of diseases including viral infections and cancers. In particular, an *ex vivo* treatment protocol has been developed for HIV infection and is now under phase I clinical trial. Here cultured haematopoietic stem cells of HIV patients are incubated with lentiviral vectors carrying the anti-HIV shRNA. These HIV-resistant stem cells are then transplanted into the bone marrow of HIV-affected patients and allowed to proliferate and replace diseased cells.

Whatever method of delivery and RNAi trigger being used, one of the most important issues to be considered and overcome is the safety of RNAi. As mentioned before, RNAi induces a number of innate immune responses and silences non-specific targets which should be taken into account when designing the siRNA or shRNA trigger. Also, due to the high mutation rate in viruses, using a combination of multiple shRNAs against the viral genome, and also against host genes required for infection, are important issues to consider.

Disease	RNAi trigger	Delivery strategy	Route of administration	Reference
Respiratory syncytial virus (RSV) infection	siRNA	Complexing siRNA with nanoparticle	Intranasal	(Zhang et al., 2005a)
Apolipoprotein B (ApoB) related disease	siRNA	Encapsulating siRNA with stable nucleic acid particles (SNALP) Cholesterol-conjugated siRNA	Intravenous injection	(Zimmermann et al., 2006) (Wolfrum et al., 2007)
Human immunodeficiency virus (HIV) infection	siRNA	Complexing siRNA with antibody conjugate	Intravenous injection	(Song et al., 2005)
Haptitis B virus (HBV) infection	shRNA	Adeno-associated virus serotype 8 vector	Intrasplenic injection	(Chen et al., 2007)
Human immunodeficiency virus (HIV) infection	shRNA	Lentiviral vector	*Ex vivo* treatment and transplant	(Li et al., 2006)
Cancer (supression of Interleukin 8, a factor for tumor growth and metastasis)	shRNA	Adenoviral vector	Mouse model xenograft	(Yoo et al., 2008)

Table 1.6. Therapeutic intervention using RNAi. This table summarises the development of RNAi in the treatment of human diseases. The type of RNAi trigger, delivery strategies and route of administration are described.

1.3.2 Gene expression profiling

Measuring the expression of genes in various tissues, different stage during development or during perturbation experiments is essential for understanding complex transcriptional programmes. There are many different ways to profile the expression pattern of genes. These range from traditional low-throughput methods to genome-scale high-throughput methods as described below:

1.3.2.1 Conventional methods

A. Northern blotting

For northern blotting (Alwine et al., 1977), RNA samples are first separated by size via electrophoresis in an agarose or polyacrylamide gel under denaturing conditions. The RNA is then transferred to a membrane, crosslinked and hybridised with a radiolabelled or nonisotopically-labelled probe.

B. Nuclease protection assays

Nuclease protection assay (NPA) involves target-probe hybridisation in solution. A single-stranded labelled probe is incubated with an RNA sample so that DNA-RNA or RNA-RNA hybrids are formed. The mixture is then exposed to ribonucleases that specifically cleave only single-stranded RNA. The probe:target hybrids are precipitated and separated on a denaturing polyacrylamide gel and are either visualised by autoradiography or by secondary detection.

C. Differential display

Differential display (DD) is a technique involving PCR without relying on prior knowledge of gene sequences that can be used to isolate differentially expressed genes (Liang and Pardee, 1992). The mRNA samples are reverse transcribed into cDNA, amplified by PCR and labelled with radioisotopes or fluorescent dyes and separated by denaturing polyacrylamide gels. The cDNA populations from different samples can be visualised and compared, and differentially expressed genes can be identified and sequenced.

D. Quantitative real-time polymerase chain reaction (qRT-PCR)

Quantitative real-time PCR is a highly sensitive kinetics-based quantification technique where PCR products are measured in real time to monitor the concentration of nucleic acids. In qRT-PCR, PCR is performed as normal with a pair of oligonucleotide primers, however fluorescent dyes are used to measure the amount of PCR product. There are two different types of approaches that have been used which will be discussed in Chapter 3. qRT-PCR is commonly used for the validation of microarray data and for quantification where the starting material is limited. It has also been used in combination with other techniques in the study of transcriptional networks during macrophage activation (Nilsson et al., 2006).

Table 1.7 summarised the major advantages and disadvantages of conventional gene expression profiling techniques.

Conventional technique	Advantages	Disadvantages
Northern blotting	• Cheap • Simple procedures	• Not suitable for accurate quantitative analyses • Low-throughput • Low sensitivity • Target transcript sequence required
Nuclease protection assay	• Cheap • Simple procedures • Higher sensitivity • Mutli-probe analyses possible	• Not suitable for accurate quantitative analyses • Low-throughput • Target transcript sequence required
Differential display	• Target transcript sequence not required • Relatively high-throughput	• Not suitable for accurate quantitative analyses • Sequencing may be required
qRT-PCR	• Highly sensitive • Accurate quantitative analyses	• Low-throughput • Target transcript sequence required • Expensive

Table 1.7. Major advantages and disadvantages of conventional methods of gene expression profiling.

1.3.2.2 Sequencing-based methods

A. Serial analysis of gene expression (SAGE)

Serial analysis of gene expression (SAGE) is a sequencing-based technique used to measure gene expression (Velculescu et al., 1995) (http://www.ncbi.nlm.nih.gov/projects/SAGE/). Biotinylated double-stranded cDNAs are cleaved with a restriction enzyme which has a 4-bp recognition site. The 3'ends of the cDNAs are then collected with streptavidin beads. The cDNAs are separated into two pools, ligated with two different linkers (A and B) and each of them are cleaved by a type IIS restriction enzyme which cuts up to 20 bp downstream of the recognition sites. The cleaved fragments, each containing a gene-specific tag, are concatenated (by amplification with primers against the linkers A and B), cloned and sequenced. The number of each gene-specific tag is quantified and the tags are mapped to the annotated genome.

B. Cap analysis of gene expression (CAGE)

Cap analysis of gene expression (CAGE) is a very similar technique to SAGE (Kodzius et al., 2006; Shiraki et al., 2003). Instead of creating signature tag at the 3' end, CAGE clones the 5' ends of cDNA fragments. cDNAs are generated with random primers and isolated by a biotin cap trapper method, where the 5' cap of the mRNA is biotinylated and removed by streptavidin beads. Linkers are attached to the 5' ends of cDNAs to introduce a recognition site for the restriction enzyme. After amplification, the sequencing tags are concatenated for high-throughput sequencing.

C. Massively parallel signature sequencing (MPSS)

Massively parallel signature sequencing (MPSS) combines the technique of non-gel based sequencing and *in vitro* cloning of DNA fragments onto microbeads (Brenner et al., 2000a). Recent developments in sequencing technology such as the Illumina Solexa (Bennett, 2004) and the 454

sequencing platforms (Margulies et al., 2005) have also been developed with a similar principle. The initial steps of *in vitro* cloning are similar to other tagging approaches except that the plasmids used for cloning contains a tag sequence (Brenner et al., 2000b). PCR products having this tag sequence are generated and attached onto the microbeads carrying the anti-tag sequence by base-pairing. Each of these microbeads carries about 10^5 copies of the same cDNA fragments. The high concentration of DNA templates on the microbeads allows high-throughput sequencing to be monitored by detecting fluorescent signals from the beads. MPSS and related sequencing methods have been widely used for studying various biological pathways in different organisms (Table 1.8).

Biological pathway studied	Sequencing technology	Reference
Profile human ES cells markers	MPSS	(Brandenberger et al., 2004)
Profile fetal human neural precursor cells markers	MPSS	(Cai et al., 2006)
Identify cancer related genes	MPSS	(Chen et al., 2005)
Generate an atlas of gene expression in various cells and tisses in human	MPSS	(Jongeneel et al., 2005)
Profile microRNA expression patterns in mammalian embryonic development	MPSS	(Mineno et al., 2006)
Gene profiling and technology evaluation in *Drosophila melanogaster*	454 sequencing	(Torres et al., 2008)
Profile microRNA in human ES cells	Illumina/Solexa	(Morin et al., 2008)

Table 1.8. Applications of massively parallel signature sequencing technology in gene expression profiling.

D. Polony multiplex analysis of gene expression (PMAGE)

Polony multiplex analysis of gene expression (PMAGE) can be used to profile gene expression of rare transcripts and genes with low expression levels (<1 copy per cell) (Kim et al., 2007). Samples are subject to sequencing directly bypassing all the library amplification, concatenation and subcloning steps. These cDNAs samples are amplified with 1-micrometer polony beads carrying adapter primers in emulsion PCRs. Polony beads carrying DNA templates are cross-linked to aminosilylated glass with amino-ester bridges. Thus an *in vitro* library is generated for high-throughput sequencing.

1.3.2.3 Microarray-based methods

Microarray technology was first described in 1995 for quantitative expression analysis in *Arabidopsis* (Schena et al., 1995). Microarrays are libraries of DNA sequences from a genome which are arrayed at high density on a solid support. Since they were first described, the technology has advanced significantly and has been widely used in the expression studies in various organisms. To date, various microarray platforms are available as described below.

Spotted arrays (genomic clones, cDNAs, PCR products or oligonucleotides) were first developed for using array technology. Initially, double-stranded cDNAs were spotted onto glass microscope slides by a robotic device (Schena et al., 1995). The glass slides are usually coated with reactive

molecular groups such as poly-L-lysine or epoxy for DNA fragments to immobilise onto the surface. The major disadvantages of cDNA/PCR product arrays are that it is difficult to control non-specific hybridisation and hybridisation efficiency due to variations in the GC content of cDNAs. In contrast, spotted oligonucleotide arrays are usually 40 to 60-mers and are single-stranded. Thus, problems associated with cross-hybridisation and variations in hybridisation efficiency are theoretically significantly reduced for these arrays. Major drawbacks of spotted arrays lie in the discrepancy among different batches of arrays and the relatively low density of oligonucleotides that are immobilised onto the glass slides.

Oligonucleotides can also be directly synthesised at high density on the surface of the array by photolithography (Affymetrix) (Singh-Gasson et al., 1999), programmable optical mirrors (NimbleGen) (Lipshutz et al., 1999) and ink-jet devices (Agilent) (Hughes et al., 2001). The BeadArray technology (Illumina) has also been developed for synthesis of high density oligonucleotide arrays (Kuhn et al., 2004). This involves the assembly of silica beads carrying hundreds of thousands of copies of a specific oligonucleotide in microwells on fibre optic bundles or planar silica slides.

For comparing expression levels of genes in different RNA populations, either a two-colour or one-colour labeling approach can be used (Figure 1.7). For spotted arrays, target and reference samples are labelled with fluorescent dyes such as Cy3 and Cy5 and hybridised on the same array. The labelled samples will bind to the DNA sequences on the array in a competitive manner and the fluorescence intensities of the two channels are quantitated. For other types of array such as Affymetrix GeneChips and Illumina BeadArray, a one-colour approach is used where different samples labelled with the same fluorescent dye are hybridised onto separate arrays. The fluorescence intensity of a single channel is quantitated and then compared across separate arrays hybridised with either the target or the reference sample.

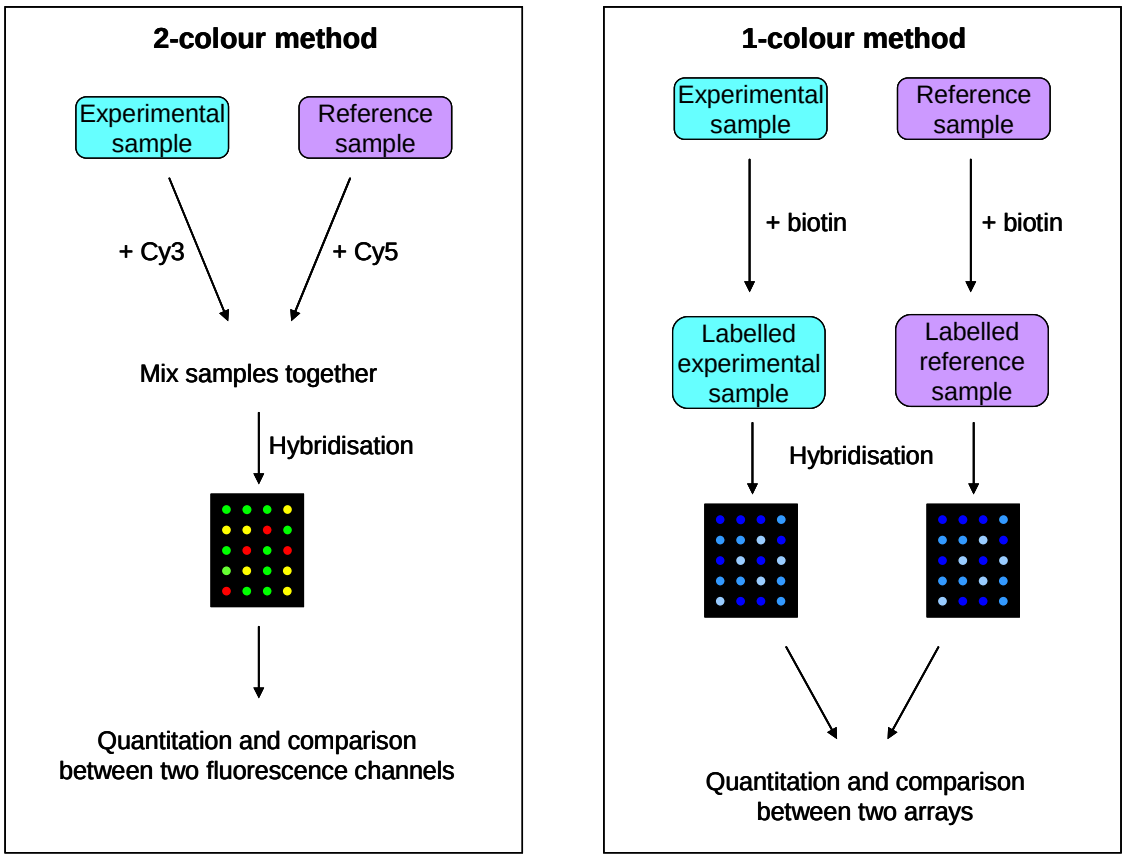


Figure 1.7. Principles of 1-colour and 2-colour microarray hybridisation. Left panel: 2-colour method; right panel: 1-colour method. In the 2-colour approach, experimental and reference samples are labelled, each with one of two different fluorescent dyes. The labelled samples are mixed together and hybridised to the same array. The two fluorescence channels are quantitated and compared. In the 1-colour approach, the experimental and reference samples are labelled with the same fluorescent dye/detection system and the labelled samples are hybridised to two separate arrays. The two arrays are quantitated and compared.

1.3.2.4 Applications of microarrays in gene expression profiling

A. Identification of pathway-specific genes

Typically, global gene expression profiles are monitored throughout a temporal program at different times within a pathway or at different stages of a developmental process. The DNA microarrays usually contain the whole genome or a complete set or subset of open reading frames (ORFs) of the organism. Differentially expressed genes are identified by comparing the expression profile at different time points. For example, microarrays have been used to study metabolic pathways in *S. cerevisiae* and to identify developmental-specific genes of metamorphosis in *Drosophila* (DeRisi et al., 1997; White et al., 1999).

B. Identification of downstream targets using genetic perturbation

Microarray expression profiling is commonly used to identify the effects on patterns of expression which occur when a biological system is perturbed for a gene of interest (for example, knockouts,

over-expression, knockdowns, ectopic expression, or introduction of mutations). Such approaches have been successfully used to identify downstream target genes of some key genes involved in cancer, for example, c-myc, p53 and ras (Coller et al., 2000; Milyavsky et al., 2005). However, this method of profiling cannot distinguish between primary and secondary target genes. Ultimately, determination of direct regulatory programmes controlled by a specific gene product (for example, a transcription factor) must be accompanied by other approaches such as ChIP (see section 1.3.3).

C. Profiling of human diseases and therapeutic responses

Gene expression profiling has also been used to study the molecular basis and identify the gene signatures of human cancer by comparing and classifying patient samples (Ferrando et al., 2002; Ge et al., 2006). It has also been used to study therapeutic effects of drugs and other treatments (Gyorffy et al., 2005; Marton et al., 1998; Shipp et al., 2002).

1.3.3 Characterisation of regulatory elements

Studying only mRNA expression patterns within a biological system cannot allow us to unequivocally identify direct target genes of transcription factors, because they fail to provide evidence about DNA-TF binding events. However, physical interaction between transcription factors and DNA can be determined both experimentally and computationally. This allows us to study where and how the transcription factor regulates the transcription of its target gene, which is important empirical evidence to support our understanding of transcriptional networks.

1.3.3.1 Conventional methods

A. DNase I hypersensitivity assays

As mentioned in section 1.1.2.5, chromatin structures are modified by the combinatorial action of chromatin-remodelling complexes and histone modification rendering the exposure of nucleosome-free DNA. Nucleosome-free DNA regions are often a characteristic of regulatory elements. These nucleosome-free DNA regions are extremely sensitive to the cleavage by DNase I - thus, they are regarded as DNase I hypersensitive sites (HSs). HSs have been shown to be associated with regulatory elements such as promoters, enhancers, silencers etc. (Gross and Garrard, 1988). Traditional and advances of HS assays are summarised in Table 1.9.

DNase I hypersensitivity assays	Desciption	References
Traditional assays	Genomic DNA was cleaved by DNase I followed by DNA purification, restriction enzyme digestion, southern blotting and hybridisation with a labelled probe	(Wu et al., 1979)
DNase I assays coupled with PCR and qPCR	Optimised PCR and quantitative PCR methods were used to increase the resolution, quantitation and sensitivity of the HSs mapping	(Follows et al., 2007; McArthur et al., 2001; Yoo et al., 1996)
Quantitative chromatin profiling (QCP)	Quantitation of tiled amplicons across a locus by quantitative real-time PCR	(Dorschner et al., 2004)
DNase I assays coupled with MPSS	Cloning of hypersensitive sites and analyses by massively parallel signature sequencing (MPSS)	(Crawford et al., 2004; Crawford et al., 2006b)
DNase I assays coupled with tiled microarray	Mapping of HSs by tiled microarray across a particular region of interest and potentially in a global manner in any sequenced genomes	(Crawford et al., 2006a; Follows et al., 2006)
In silico prediction of HSs	*In silico* prediction of HSs in the human genome using a supervised pattern recognition algorithm with high accuracy	(Noble et al., 2005)

Table 1.9. DNase I hypersensitivity assays and advances.

B. DNase I foot-printing

The DNase I foot-printing assay, also called the DNase I protection assay, is an *in vitro* assay used to identify protein-bound DNA elements (Galas and Schmitz, 1978). The procedure involves radioactively-labeling DNA fragments at one end. The DNA fragments are incubated with or without the protein of interest and then subjected to DNase I treatment followed by electrophoresis and autoradiography. DNA bound by proteins or transcription factors is more resistant to cleavage by DNase I than naked DNA and is absent on the autoradiograph (as gaps in the ladder of end-labelled fragments) and can be regarded as footprints for protein-bound regions.

C. Electrophoretic mobility shift assays

Electrophoretic Mobility Shift Assay (EMSA), also called gel shift assay, is another *in vitro* technique for studying protein-DNA interactions (Garner and Revzin, 1981). The gel shift assay is carried out by first incubating a radioactively-end-labelled or fluorescent-labelled (Onizuka et al., 2002) DNA fragment containing the putative protein binding site with or without a protein of interest. The reaction products are then analysed on a non-denaturing polyacrylamide gel followed by autoradiography. The protein-DNA complexes migrate more slowly than naked DNA and are retarded on the gel compared to the control sample.

Other *in vitro* assays include systematic evolution of ligands by exponential enrichment (SELEX) and cyclic amplification and selection of targets (CASTing), both of which screen pools of nucleic acid ligands with the protein of interest (Tuerk and Gold, 1990; Wright et al., 1991).

D. Reporter gene assays

The identification of putative regulatory elements alone fails to provide information on their activity within the cell. Often the functionality of these elements is tested by reporter gene assays (Weber et al., 1984) which can also be adapted for genome-wide screens. The putative regulatory elements of interest (or random genomic fragments for large scale screening), are cloned into a plasmid containing a reporter gene encoding chloramphenicol acetyltransferase (CAT), β-galatosidase, green fluorescent protein (GFP) or luciferase. The plasmid construct is then transfected stably or transiently into cultured cells by electroporation or lipofection and the activity of the reporter is quantified. The reporter construct is made according to the different type of regulatory elements to be tested (Figure 1.8).

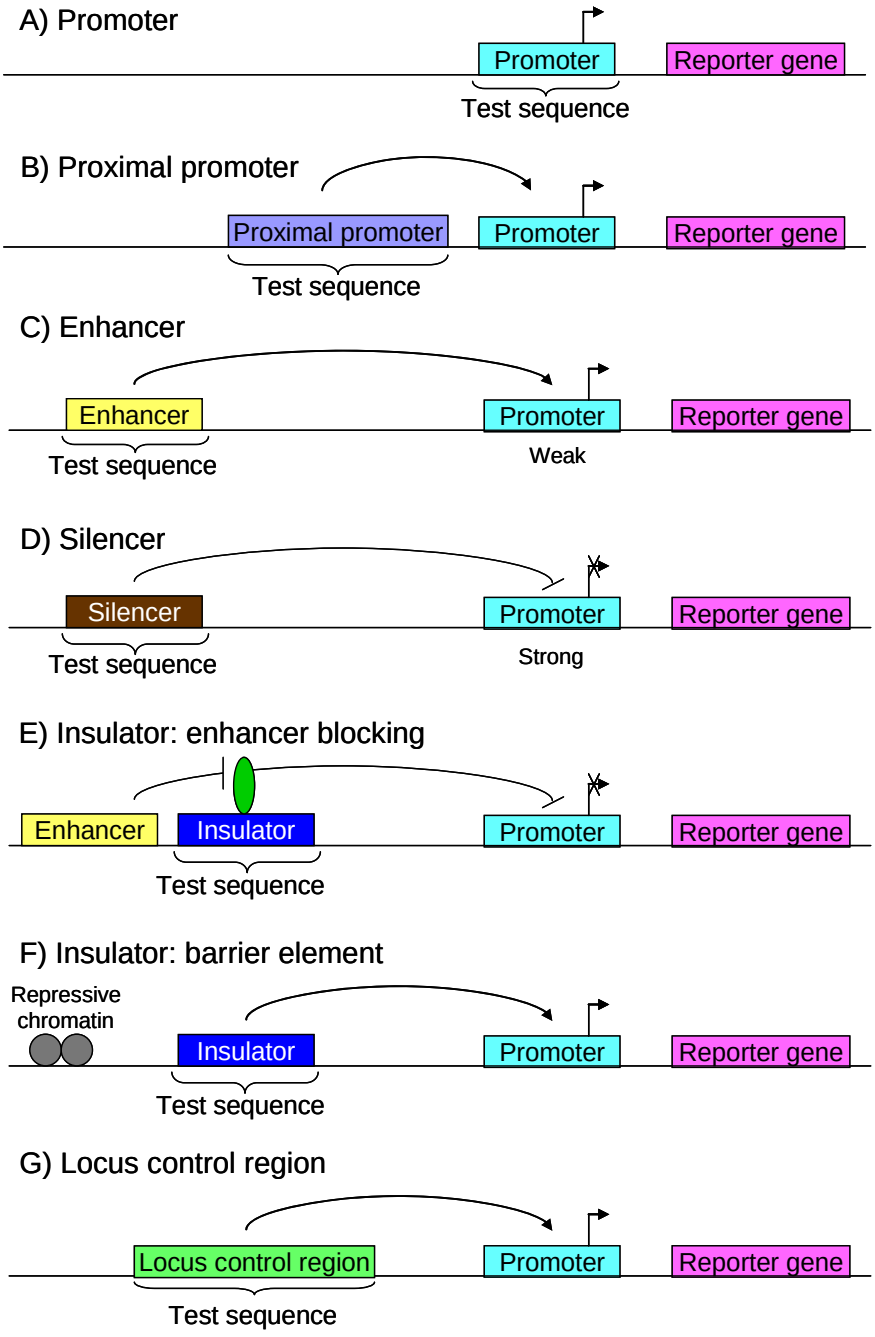


Figure 1.8. Functional reporter gene assays for the identification of regulatory elements. A) A genomic element representing a putative promoter is cloned immediately upstream of a reporter gene lacking an endogenous promoter. B)-D) Sequences representing putative proximal promoters, enhancers and silencers are cloned upstream of a reporter gene directed by an appropriate strength promoter. E) Insulators with an enhancer blocking activity interfere with enhancer-promoter communication and repress gene expression. F) Insulators having a barrier activity avoid the spread of repressive chromatin. G) Locus control regions confer correct gene expression patterns.

1.3.3.2 ChIP-based methods

Chromatin immunoprecipitation (ChIP) is a well developed and powerful technique to study *in vivo* interaction between protein and DNA. This is an approach where protein and DNA can be cross-linked in the native chromatin structure *in vivo* and which overcomes the obstacles presented with the traditional methods which use *in vitro* based assays (Figure 1.9). Cells are grown under the desired experimental condition and fixed with cross-linking agents whilst intact, effectively resulting in covalently interactions between proteins and DNA. The cross-linked chromatins are sonicated to shear the DNA fragments to approximately 200-1000 bp. The protein-DNA complexes are immunoprecipitated with an antibody against the protein of interest. The crosslinks are then reversed and the DNA bound to the protein is purified. The ChIP DNAs can be quantified by Southern blot, PCR or quantitative PCR to identify specifically enriched DNA fragments (Das et al., 2004). An alternative approach named ChIP-on-beads which combines a conventional PCR with tagged primers and captures the products onto microbeads followed by analyses by flow cytometry was developed for larger scale analyses (Szekvolgyi et al., 2006). However, all these methods require prior knowledge of the putative sequence that the protein may bind and are relatively low-throughput. Some of the issues and limitations associated with ChIP are discussed in Chapter 5.

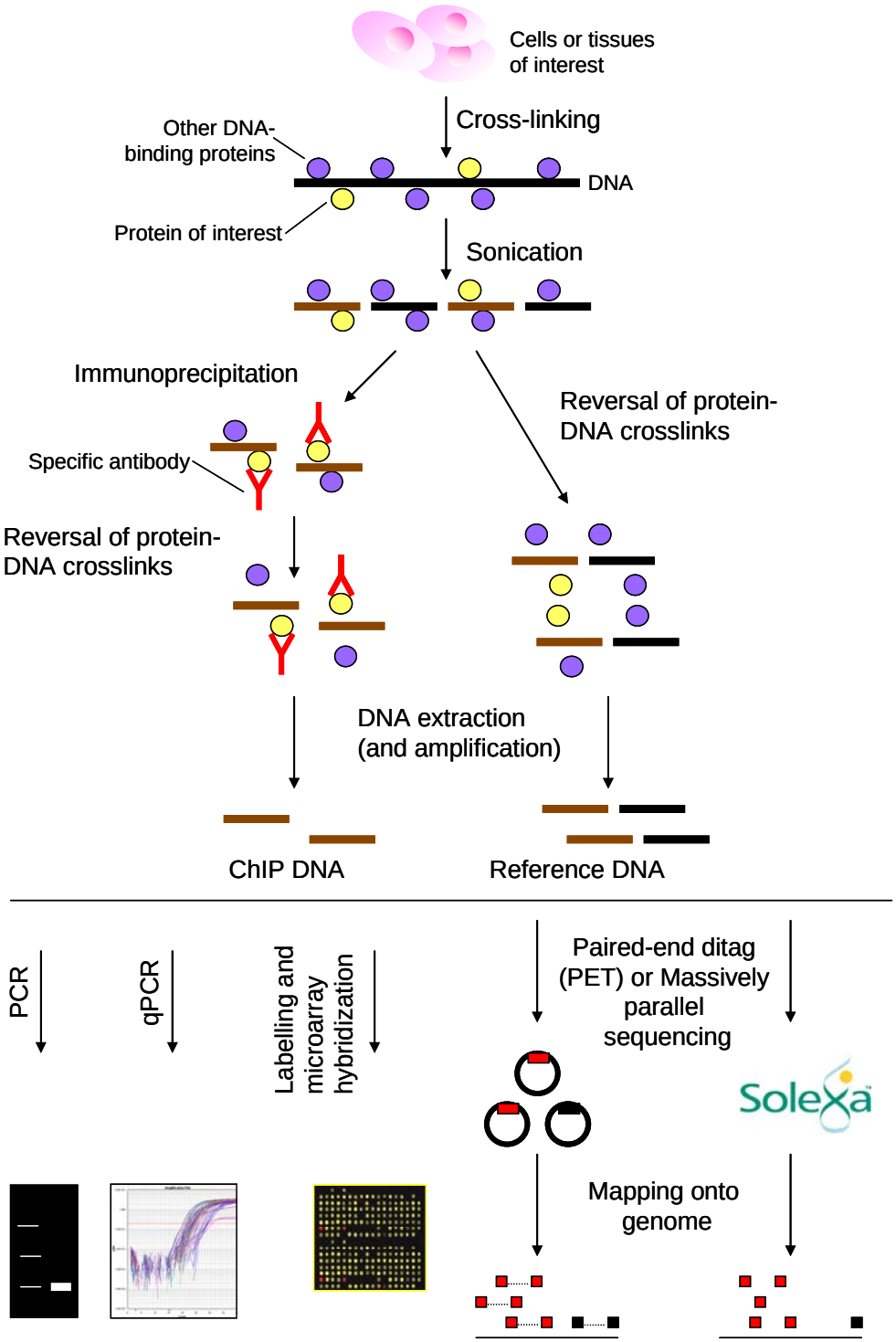


Figure 1.9. A schematic diagram of the chromatin immunoprecipition (ChIP) assay and subsequent analyses. DNA-protein complexes in the cells or tissues of interest are cross-linked, sonicated, and immunoprecipitated followed by reversal of crosslinks and DNA purification (with or without amplification). Purified DNAs are analysed by PCR, qPCR, microarrays, paired-end ditag or massively parallel sequencing.

1.3.3.3 High-throughput ChIP applications

A. ChIP-on-chip

To map protein binding sites on a genome-wide scale, ChIP coupled with microarrays (ChIP-on-chip or ChIP-chip) is an extremely powerful technique which is widely used (Figure 1.9). The DNA

from ChIP is usually amplified or several IPs of material are pooled to provide sufficient DNA for labeling with fluorescent dyes such as Cy3 or Cy5. The labelled DNA is hybridised onto the microarray while DNA that is not immunoprecipitated (or a mock immunoprecipitation) is used as a reference for comparison.

ChIP-on-chip was first successfully demonstrated in yeast, where the complexity of the genome allows one to study all genomic sequences in relatively simple array-based experiments. Table 1.10 summarised some examples of ChIP-on-chip studies performed in yeast.

ChIP-on-chip study in yeast	References
Mapping of binding sites of the transcription factor Rap1	(Lieb et al., 2001)
Mapping of binding sites of the transcription factors SBF and MBF	(Iyer et al., 2001)
Mapping of binding sites of the transcription factors Gal4 and Ste12	(Ren et al., 2000)
Annotation of all transcription factor and DNA binding protein sites across the yeast genome	(Lee et al., 2002)
Study of recruitment of TATA-binding proteins to promoters	(Kim and Iyer, 2004)
Study of recruitment of RNA polymerase II to promoters and open reading frames	(Pokholok et al., 2002)
Study of DNA replication	(Wyrick et al., 2001)
Study of DNA recombination	(Gerton et al., 2000)
Study of chromatin structures and histone modifications	(Bernstein et al., 2002; Kurdistani et al., 2004; Nagy et al., 2002; Pokholok et al., 2005; Robert et al., 2004; Robyr et al., 2002)

Table 1.10. Application of ChIP-on-chip studies in yeast (*S. cerevisiae*).

Because of the complexity of mammalian genomes, ChIP-on-chip studies have traditionally involved the use of arrays which contain features representing only a sub-set of the genome. Human promoter arrays were first used in the mapping of E2F in cell cycle progression and proliferation (Ren et al., 2002). Since then, many similar studies have been performed in human and mouse (Table 1.11). CpG island arrays have also been used in identifying *in vivo* targets of the E2F family, pRb, c-Myc (Mao et al., 2003; Oberley et al., 2003; Weinmann et al., 2002; Wells et al., 2003). CpG island and promoter arrays are particularly useful for studying the global regulation of a particular transcription factor across the whole genome. However, the disadvantage of these arrays is that they only represent a subset of sequences in a genome and, thus, not all transcription factor binding sites can be detected. In otherwords, transcription factors binding to regulatory elements outside promoters and CpG islands such as enhancers and repressors cannot be examined with this biased approach.

To circumvent the limitations of promoter and CpG island arrays in mammals, arrays whose features spanned entire gene loci, chromosomal regions, or whole mammalian genomes have been used in ChIP-on-chip approaches. Tiling path arrays across the human β-globin locus were first

used to map GATA1 binding sites (Horak et al., 2002). Other tiling arrays representing non-repetitive regions in the human genomes were also explored for the study of transcription factor binding and histone modifications (Table 1.11). Using arrays which tiled the ENCyclopedia of DNA Elements (ENCODE) Project (2004a) pilot regions (1% of the human genome) both transcription factors and histone modifications have been characterised (2007; Koch et al., 2007).

ChIP-on-chip study	Organism	Type of array	References
Mapping of E2F in cell cycle progression and proliferation	Human	Promoter array	(Ren et al., 2002)
Mapping of c-Myc and Max in human cancer cells	Human	Promoter array	(Li et al., 2003)
Mapping of HNFs in the human liver and pancreas	Human	Promoter array	(Odom et al., 2004)
Mapping of myogenic regulatory factors (MRFs) and myocytes enhancer factor 2 (MEF2) in muscle development	Human	Promoter array	(Blais et al., 2005)
Mapping of OCT4, SOX2 and NANOG in human embryonic stem pluripotency and self-renewal	Human	Promoter array	(Boyer et al., 2005)
Mapping of the polycomb repressive complexes in the study of embryo developments and pluripotency	Mouse	Promoter array	(Boyer et al., 2006)
Identifying *in vivo* targets of c-Myc	Human	CpG island array	(Mao et al., 2003)
Identifying *in vivo* targets of E2Fs in human cancer cells	Human	CpG island array	(Oberley et al., 2003; Weinmann et al., 2002)
Study of pRb in cell cycles	Human	CpG island array	(Wells et al., 2003)
Mapping of GATA1 binding site across the human β-globin locus	Human	Tiling array	(Horak et al., 2002)
Mapping of the binding sites of CREB (Cyclic AMP-responsive element-binding protein) across the non-repetitive regions of chromosome 22	Human	Tiling array	(Euskirchen et al., 2004)
Mapping of the binding sites of the NF-kappaB family across the non-repetitive regions of chromosome 22	Human	Tiling array	(Martone et al., 2003)
Mapping of binding sites of Sp1, c-Myc and p53 on chromosome 21 and 22	Human	Tiling array	(Cawley et al., 2004)
Study of histone modifications in human chromosome 21 and 22 and comparison with mouse loci	Human and mouse	Tiling array	(Bernstein et al., 2005)
Characterisation of histone modification in 1% of human genome	Human	Tiling array (ENCODE)	(Koch et al., 2007)
Mapping of all promoters in the human genome	Human	Tiling array (All non-repetitive regions)	(Kim et al., 2005b; Kim et al., 2005c)
Mapping of transcriptional initiation sites	Human	Tiling array (All non-repetitive regions)	(Guenther et al., 2007)
Mapping of polycomb repressive complex in human embryonic stem cells and developmental regulators at transcriptional repressive nucleosome regions	Human	Tiling array (Whole genome)	(Lee et al., 2006)

Table 1.11. Application of ChIP-on-chip studies in mammals.

B. Sequencing-based analyses

An alternative approach to analyse the ChIP DNAs is to directly sequence them and map the sequencing reads onto the genome. This is an unbiased approach which is not limited by what is presented on an array. DNA fragments are usually cloned and sequenced (Weinmann et al., 2001). However, the large number of non-specifically immunoprecipitated fragments makes cloning unpractical. The ChIP-display technique concentrates the target DNA sequences and scatters the non-specific ones (Barski and Frenkel, 2004). The ChIP DNAs are digested with a particular restriction enzyme and analysed by gel electrophoresis. DNA elements enriched in the ChIP reaction will show bands of the same size reproducibly on the gel. These concentrated DNA fragment can then be cloned and sequenced. However, ChIP-display is not suitable for mapping histone modifications and transcription factors with a lot of binding sites and cannot produce precise quantification of the enrichments at a particular location. Another technology called STAGE (Sequence Tag Analysis of Genomic Enrichment) was developed which is based on high-throughput sequencing of concatemerised tags derived from target DNA enriched in ChIP (Kim et al., 2005a). Despite its high-throughput nature, this mono-tagging technology leaves ambiguity in the mapping of short sequences onto the genome. A similar technique ChIP-PET (for paired-end ditag) combines the chromatin immunoprecipitation strategy with the paired-end ditag strategy of high-throughput sequencing. The ChIP DNA fragments are cloned to create a library which is further digested and concatemerised to create the PET library having 36-bp signatures of 18 bp of the 5' and 3' ends of the original fragment for sequencing. With the advances in sequencing technology such as the Solexa platform (Bennett, 2004), the time-consuming cloning steps in the above methods can now be circumvented. ChIP-sequencing (ChIP-seq) combines traditional chromatin-immunoprecipitaion with ultra high-throughput Solexa or 454 sequencing platforms for identifying and quantifying enriched DNA elements. Table 1.12 summarised the applications of ChIP coupled with various sequencing techniques in the study of transcriptional regulation.

ChIP coupled with sequencing study	Organism	Sequencing technique	References
Identification of E2F-regulated promoters	Human	ChIP coupled with cloning and traditional sequencing	(Weinmann et al., 2001)
Identification of target genes of RUNX2	Mouse	ChIP-display	(Barski and Frenkel, 2004)
Mapping of target sites of TATA-binding proteins in yeast and E2F4 in human cells	Yeast and human	STAGE	(Kim et al., 2005a)
Mapping of p53 genomic binding sites	Human	ChIP-PET	(Wei et al., 2006)
Mapping of c-Myc genomic binding sites	Human	ChIP-PET	(Zeller et al., 2006)
Delineation of the transcription regulatory networks of Nanog and Oct4 in stem cell pluripotency	Mouse	ChIP-PET	(Loh et al., 2006)
Mapping of *in vivo* binding sites of REST	Human	ChIP-seq (Solexa)	(Johnson et al., 2007)
Generation of high resolution genome-wide map of histone methylation	Human	ChIP-seq (Solexa)	(Barski et al., 2007)
Study of association of histone modification with nucleosomes	Human	ChIP-seq (Solexa)	(Schones et al., 2008)

Table 1.12. Application of ChIP coupled with sequencing techniques.

1.3.3.4 Alternative ChIP approaches

A. DamID assays

An alternative method which circumvents the need for performing ChIP experiments is DamID. This involves the labeling of DNA near the protein binding site (van Steensel and Henikoff, 2000). In this method, the transcription factor of interest is fused with the *Escherichia coli* DNA-adenine methyltransferase (Dam) protein and is expressed in a cell culture system. The Dam protein methylates the adenine base in GATC sites 1.5 to 2 kb around the binding site of the transcription factor-Dam fusion protein. DNAs from this experimental sample and from a control sample, where only the Dam protein is expressed, are extracted, digested with a restriction enzyme (*Dpn* I), labelled and hybridised onto microarrays. This method has been used in *Drosophila* (Orian et al., 2003) and in mammals (Vogel et al., 2006; Vogel et al., 2007).

B. DIP-ChIP

A modification of the ChIP-chip protocol, called DIP-chip, has been developed for the immunoprecipitation of DNA with the protein of interest *in vitro* followed by microarray analyses (Liu et al., 2005). Purified and tagged proteins of interest are mixed with genomic DNA *in vitro* and the protein-bound DNA is isolated by affinity purification, amplified and hybridised onto a genomic array. The advantage of this method is that no specific antibodies are needed as the fusion partner of the protein makes purification much easier.

1.3.4 Computational approaches to study gene regulation

With the completion of genome sequence in many species, computational tools play a significant role in the study of gene regulation and transcription networks, in combination with experimental approaches. Softwares and databases for promoter and transcription factor binding sites (TFBS) predictions and comparative sequence analyses are now widely available.

1.3.4.1 Promoter prediction

Predicting the location of promoters is particularly useful for targeting regions of interest to study with respect to regulatory interactions. However, it is rather challenging considering the core promoter may be distant from the exons and the combination of core elements may differ from promoter to promoter. The most successful programs are based on the analyses of training data sets from known promoter sets as a means of identifying functionally defined sequences conserved across promoters. These programs then scan for these conserved signatures in genomic sequence. These include PromoterInspector (Scherf et al., 2000), FirstEF (Davuluri et al., 2001) and Eponine (Down and Hubbard, 2002). Nevertheless, these programs have limited sensitivity and specificity for genome-scale analyses as they are heavily dependent on the data sets of known promoters (which may have biased representations). Promoters associated with CpG islands are generally well-predicted compared to those which are not (Bajic et al., 2004).

1.3.4.2 Transcription factor binding site prediction

Transcription factor binding sites (TFBSs) are conserved sequences with a certain degree of degeneracy which transcription factors recognise and bind. The binding sites of the most well-characterised transcription factors are compiled in online databases such as TRANSFAC, TRRD and COMPEL (Heinemeyer et al., 1998). Programs such as MATCH (Kel et al., 2003) or online tools such as TESS and TFSEARCH (Akiyama, 1998; Schug, 1997) make use of the TRANSFAC database to identify TFBSs in input genomic sequences. However, one of the major drawbacks of these methods is that there can be a large number of false positive or true negatives owing to the quality of data used initially to populate the databases. Tools such as JASPAR have been developed recently which use more sophisticated statistically-based models of TFBSs (Sandelin et al., 2004).

To overcome the potential problems mentioned above, more intuitive motif discovery approaches identify sets of common sequence motifs in the upstream regions of a set of genes which are likely to be co-regulated. This allows researchers to identify known as well as novel motifs that might be associated with a transcription factor. The algorithms available include AlignACE (Roth et al., 1998), MEME (Bailey and Elkan, 1995), MDScan (Liu et al., 2002) and NestedMICA (Down and Hubbard, 2005). Such method has been used to identify sequence motifs or clusters of motifs in the

promoter regions of co-expressed genes inferred from gene expression data in yeast (Segal et al., 2003).

1.3.4.3 Comparative sequence analyses

Comparative sequence analyses have long been used as a tool to identify evolutionally conserved and functionally important DNA sequences. Traditionally, it has been applied to the coding regions of genomes to predict novel genes, and more recently, for the identification of *cis*-regulatory elements. Many algorithms and softwares have been developed to aid these kinds of analyses. These include, but are not limited to, BLAST (Altschul et al., 1990), FootPrinter (Blanchette and Tompa, 2003), PhastCons (Siepel et al., 2005), LAGAN (Brudno et al., 2003) and VISTA (Visel et al., 2007). One of the early applications of comparative sequence analyses in regulatory element prediction identified a new enhancer in the SCL locus (Gottgens et al., 2000). Large-scale genome comparative analyses have also been perfomed recently to identify enhancers (Pennacchio et al., 2006; Woolfe et al., 2005). In particular, Pennacchio et al. (2006) identified a subset of enhancers which are highly active in neuronal development and functionally validated 45% of them using *in vivo* enhancer trap assays.

However, comparative sequence analyses have limitations. First, not all the conserved regions contain functional regulatory motifs (Balhoff and Wray, 2005). Secondly, transcription factor binding sites may not be conserved among species (Dermitzakis and Clark, 2002). One reason to explain this is that TFBSs have some degree of degeneracy. Therefore, perfect sequence conservation at the binding site may not necessarily be required for function. Recently, a ChIP-on-chip study of four tissue-specific transcription factors in mouse and human hepatocytes revealed that many occupied binding sites for these transcription factors are not conserved between the two species (Odom et al., 2007). It was shown in this study that, in many instances, a transcription factor can bind to a particular TFBS in human, but it binds to a completely different site in the mouse, irrespective of whether sequences are conserved between the two species. This suggests that sequence conservation alone cannot predict transcription factor occupancy.

1.4 Haematopoiesis

Haematopoiesis is an accessible mammalian system to study the processes associated with the regulation of gene expression and the relationships between genes and their protein products in transcriptional networks. The study of human haematopoiesis formed the basis of the biological system used in this thesis.

1.4.1 Embryonic origin and lineages of haematopoiesis

Haematopoiesis is the process of formation of mature blood cells from haematopoietic stem cells (HSCs). Pluripotent HSCs differentiate to form various blood progenitor cells which further give rise to mature and terminally-differentiated blood cells in specific lineages. In mammals, haematopoiesis occurs in two consecutive phases: primitive (or embryonic) haematopoiesis in early embryonic development and definitive haematopoiesis in late embryonic development and adults. Various tissues have been demonstrated to serve as the reservoirs of haematopoietic cells and/or sites of haematopoietic differentiation during different time of the developmental and differentiation process. These include yolk sac, para-aortic-spanchnopleura (PAS), aorta-gonad-mesonephros (AGM), liver, spleen and thymus.

The initial phase of blood development, primitive haematopoiesis, first takes place in the yolk sac around embryonic day 7 (E7) in mice or during the second to third week in human gestation. Here the undifferentiated mesodermal cells form extraembryonic blood islands where endothelial cells, precursors for the formation of blood vessels, differentiate at the edges of the mesoderm while primitive erythrocytes form in the interior regions. Thus, both endothelial and haematopoietic lineages are derived from the same origin. There is evidence supporting the existence of a bi-potential common precursor of endothelial and haematopoietic cells: the haemangioblast (Choi et al., 1998). Primitive haematopoiesis results in the production of mainly large, nucleated erythroblasts, as well as some megakaryocytes and macrophages. It is a robust yet transient process to generate large amount of blood cells for growth and development of the young embryo. Primitive haematopoiesis only occurs at early stages of embryonic development until around day 13 (E13), after which time the yolk sac begins to degenerate (Figure 1.10).

Unlike primitive haematopoiesis which is mainly erythropoietic, definitive haematopoiesis gives rise to all haematopoietic lineages (Figure 1.11). Definitive haematopoiesis occurs both in the extra-embryonic yolk sac and the intraembryonic, mesoderm-derived para-aorta-splanchnopleura (PAS) which later contributes to the aorta-gonad-mesonephros (AGM). Therefore; within the yolk sac, definitive progenitor cells are produced by a population of mesodermal cells having a fetal-adult fate rather than purely a primitive fate. This supports the idea that there is a temporal overlap between primitive and definitive haematopoiesis and that they share a common precursor (Kennedy et al., 1997). These definitive progenitors do not mature in the yolk sac, but instead they migrate to other tissues for maturation.

Definitive haematopoiesis is mainly derived from haematopoietic stem cells (HSCs). HSCs are defined as a cell population which can contribute to the long-term repopulation of the haematopoietic system of irradiated adult mice. HSCs are required for haematopoietic development

during the entire life of an organism. HSCs are characterised by their ability to self-renew and the expression of markers such as CD34 and c-kit. There has been some controversy regarding the origin of HSCs. The yolk sac has long been regarded as the site of HSCs generation as removal of yolk sac was shown to abolish haematopoiesis in the embryo (Moore and Metcalf, 1970). More recent studies have also isolated HSCs in yolk sacs prior to day 9 (E9) and confirmed their long-term multilineage activity (Yoder et al., 1997). It has been proposed that HSCs produced in the yolk sac migrate to the AGM which serves as a reservoir of HSCs. However, there are also findings opposing the yolk sac as the unique origin of HSCs. The AGM was shown to generate and expand the population of HSCs from day 10 (E10) (Medvinsky and Dzierzak, 1996).

Differentiation of HSCs does not occur in the AGM. Instead, HSCs circulate to other intraembryonic tissues such as the fetal liver for terminal differentiation and maturation of haematopoietic cells (Godin et al., 1999). Here enucleated erythrocytes producing adult globins as well as myeloid cells become mature and appear in the circulation around E12. At the same time, the fetal thymus is the site for T-lymphoid development. The fetal spleen becomes the main site of haematopoiesis during late embryogenesis until around the time of birth, when the bone marrow becomes the major site of haematopoiesis throughout the life of the animal (Godin and Cumano, 2002; Kumaravelu et al., 2002).

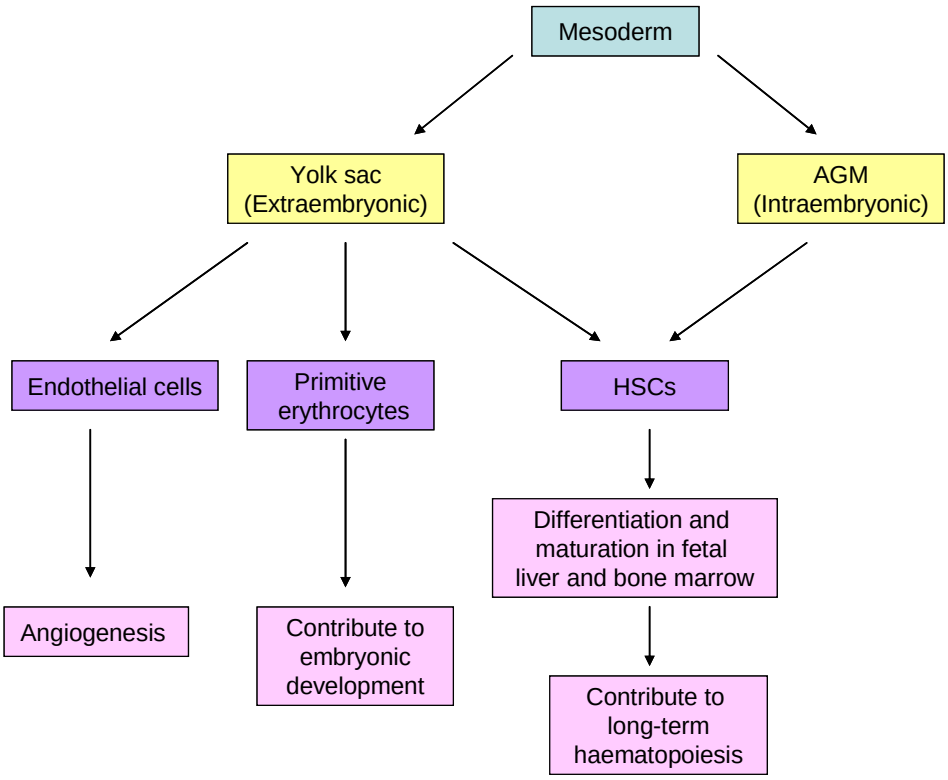


Figure 1.10. A flow diagram of the development of the endothelial lineage and primitive and definitive haematopoeisis from their embryonic origins. The extraembryonic yolk sac from the mesoderm gives rise to endothelial cells, primitive erythrocytes and haematopoietic stem cells (HSCs). Later in embryonic development, the

intraembryonic aorta-gonad-mesonephros (AGM) also gives rise to HSCs. Endothelials cells are implicated in vasculogenesis and angiogenesis. Primitive erythrocytes are critical for supporting embryonic development. HSCs migrate to the fetal liver or bone marrow for differentiation and maturation of various blood lineages and contribute to long-term haematopoiesis.

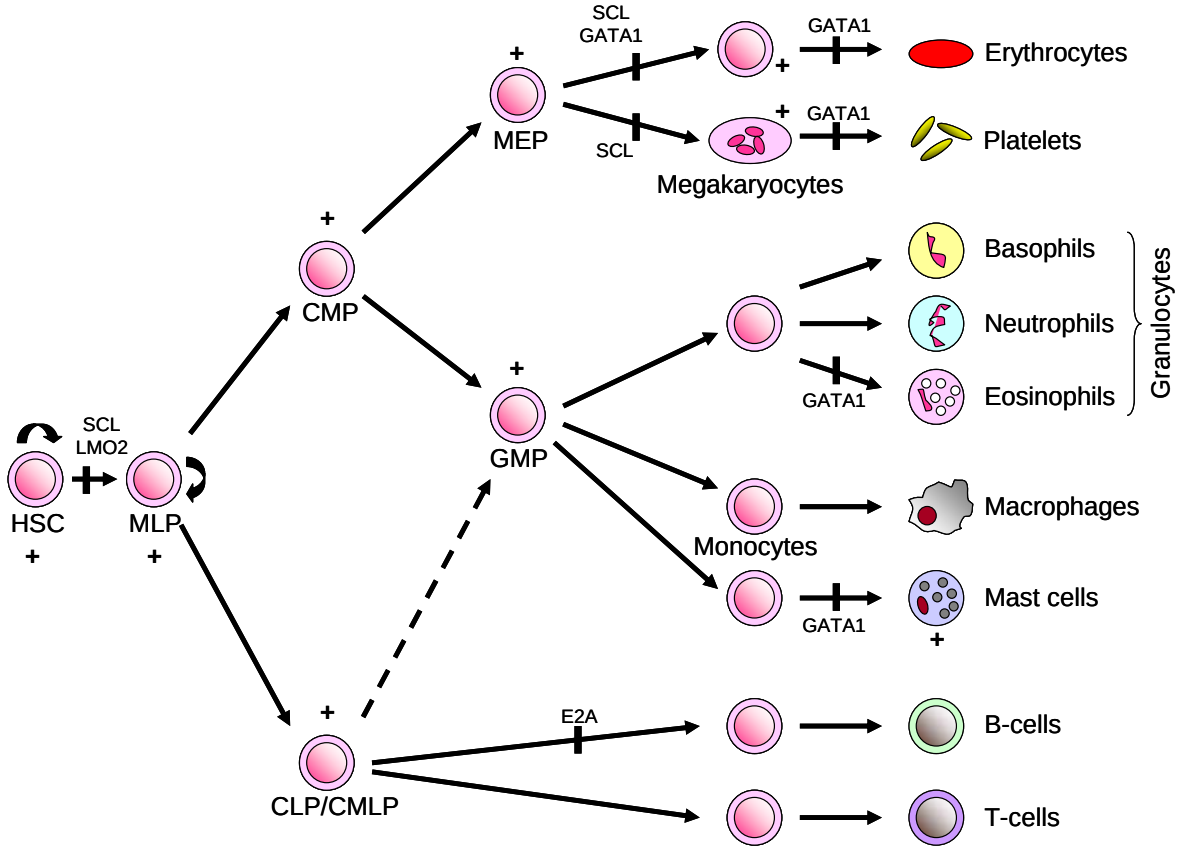


Figure 1.11. A schematic digram of haematopoietic lineage pathways from pluripotent haematopoietic stem cells to mature blood cells. The haematopoietic stem cell (HSC) is the highest in the hierarchy and gives rise to multi-lineage progenitors (MLP) which differentiate to form common myeloid progenitors (CMP) and common lymphoid progenitors (CLP) or common myeloid lymphoid progenitors (CMLP). CLPs give rise to B-cells or T-cells while CMPs give rise to megakaryocyte-erythrocyte progenitors (MEP) or granulocyte-monocyte progenitors (GMP). Alternatively, CMLPs can give rise to GMPs, B-cells and T-cells. GMPs further differentiate to form granulocytes, macrophages and mast cells while MEPs give rise to erythrocytes and platelets. The + signs in the diagram show the major expression pattern common to both SCL and GATA1. Important roles of various transcription factors (discussed in the text) in the haematopoietic lineages are indicated. The bold bars represent developmental blocks when the corresponding protein is removed. Figure modified from Ferreira et al. (2005).

1.4.2 Regulation of haematopoiesis

Haematopoietic commitment and differentiation is regulated by a tightly controlled transcriptional regulatory programme. At different stages of development, different combinations of transcription factors are expressed to further regulate expression of downstream haematopoietic specific genes in the cascade. Thus, a complex transcription network is involved to govern the molecular mechanism leading to differentiation of specific blood lineages. Often transcription factors expressed together

at certain stages form multi-protein complexes which work co-operatively for downstream regulation to take place. The master regulator of haematopoiesis, SCL or TAL1, together with the some of its interacting partners, is discussed below.

1.4.2.1 SCL

A. The SCL gene

SCL, also named TAL1, is a transcription factor which is thought to be a master regulator of haematopoietic development (Begley et al., 1989b). It was first identified in T-cell acute lymphoblastic leukemia (T-ALL) which is resulted from the translocation of human chromosome 1 at p32-33 with the T-cell receptor (TCR) delta locus on chromosome 14q11 (Begley et al., 1989a). The translocation results in the expression of a fusion transcript of TCR-delta and an unknown gene which was thereafter named the stem cell leukaemia (SCL) gene due to its involvement in T-ALL.

The human SCL gene is composed of 8 exons spanning 16 kb on chromosome 1p32-33 (Aplan et al., 1990a). The first five exons are non-coding and there are two promoters (the erythroid-specific promoter 1a and the myeloid-specific promoter 1b) located in the 5' non-coding regions. Alternative splicing events occur at this region generating different mRNA species (Aplan et al., 1990a). An additional promoter, located within exon 4 is cryptically active in T-ALL (Bernard et al., 1992). The murine SCL gene locus is structurally very similar and consists of 7 exons distributed across a 20 kb region of mouse chromosome 4 (Begley et al., 1991).

B. Expression patterns of SCL

Since SCL was identified as a leukaemic translocation fusion protein, it has been shown that SCL expression is critical to various stages of haematopoietic development. Many studies have been performed to examine the expression patterns of SCL. It has been found to be expressed in both the haematopoietic and non-haematopoietic compartments, as discussed below.

- **Haematopoietic expression**

SCL was first found to be expressed in fetal liver, regenerative bone marrow, early myeloid cell lines and leukaemic T-cell lines by northern blot analyses (Begley et al., 1989b). It was later established that SCL is also expressed in human and murine erythroid cell lines, mast cell lines and megakaryocytic cell lines (Green et al., 1992; Green et al., 1991). SCL expression was detected in normal human erythroid, mast, and megakaryocytic cell populations by in situ hybridisation and by RT-PCR (Mouthon et al., 1993). In an *in vitro* differentiation study, expression of SCL was detected and increased during the differentiation of embryonic stem cells to embryoid bodies, the *in vitro* counterpart of haematopoietic progenitors (Elefanty et al., 1997). SCL was found to be expressed in the aorta-associated $CD34^+$ high proliferative potential haematopoietic cells which are

proposed to be HSCs present later in fetal liver and bone marrow (Labastie et al., 1998). In addition to early haematopoietic and erythroid cells, the SCL transcript was also found in pre-B cells (Green et al., 1992). Using a combination of fluorescence-activated cell-sorting (FACS) and RT-PCR for quantitative analysis of expression, SCL expression was confirmed in all haematopoietic cells having an erythroid potential and present, but down-regulated, in common lymphoid progenitors (CLP) and granulocyte/monocyte progenitors (GMP) (Zhang et al., 2005b). No SCL expression was detected in immature and mature cells of the non-myeloid lineage (Figure 1.11).

- **Non-haematopoietic expression**

The overlapping origin of endothelial and haematopoietic lineages suggests a similar expression of key regulators. The SCL transcript was first demonstrated to be expressed in endothelial cells in the spleen (Hwang et al., 1993). The SCL protein was subsequently detected in endothelial progenitors in blood island and in endothelial cells in a variety of tissues including spleen, thymus, placenta and kidney (Kallianpur et al., 1994; Pulford et al., 1995). SCL mRNA has been detected in the endothelial cell clusters of the ventral endothelium of the aorta (Labastie et al., 1998).

SCL is also expressed in the nervous system. It was shown to be expressed in the human neuroepithelial cell lines (Begley et al., 1989b) and subsequently in the murine post-mitotic neurons in the metencephalon and roof of the mesencephalon (Green et al., 1992). Using a knock-in mouse, SCL was shown to be widely expressed in the thalamus, midbrain and hindbrain in the adult and the developing embryonic central nervous system (van Eekelen et al., 2003). SCL expression has also been described in vascular and smooth muscle cells in the aorta and bladder and uterine smooth muscle cells, and in the developing skeleton (Kallianpur et al., 1994; Pulford et al., 1995).

C. Co-ordinated expression pattern of SCL together with GATA1 and LYL1

The expression profiles of SCL and the transcription factors GATA1 and LYL1 are shown to be highly similar in a number of studies. Both SCL and GATA1 were co-ordinately expressed in early haematopoietic and erythroid lineages and their expression undergo biphasic modulations during erythroid and myeloid differentiation in mouse (Green et al., 1992). An early transient decrease followed by an increase of both SCL and GATA1 expression was demonstrated for induced erythroid differentiation. An early transient increase, an initial recovery, followed by a prolonged inhibition was observed during myeloid differentiation (Green et al., 1993). GATA1 and SCL were also found to be co-expressed in erythroid, megakaryocytic and mast cell lineage and down-regulated in terminal erythroid and megakaryocytic maturation (Mouthon et al., 1993). Their expression was shown to be restricted to commited progenitor cells ($CD34^+/CD38^+$) but not the most primitive cells ($CD34^+/CD38^-$). In addition, GATA1 and SCL were found to be expressed in

the extraembryonic mesoderm (precursor of yolk sac), although SCL expression was detected earlier than GATA1 (Silver and Palis, 1997).

The bHLH protein LYL1, which is structurally similar to SCL, also forms a translocation fusion protein during T-ALL (Mellentin et al., 1989) and has overlapping expression pattern with SCL in mouse (Visvader et al., 1991). Similar to SCL, LYL1 is expressed in the erythroid and myeloid lineages and in ascular tissues in mice (Visvader et al., 1991). However, unlike SCL, LYL1 is not expressed in the nervous system (Giroux et al., 2007). Its expression is initiated slightly later than SCL during haematopoietic specification, beginning during haemangioblast differentiation (Chan et al., 2007).

D. Functions of SCL

(i) Haematopoietic development and lineage specification.

SCL is one of the earliest acting regulators of haematopoietic development. Ablation of SCL resulted in embryonic death in mice at E9.5 due to the lack of blood cells (Robb et al., 1995; Shivdasani et al., 1995). Further investigation of $SCL^{-/-}$ embryonic stem (ES) cells in a mouse chimera showed that they are unable to contribute to any haematopoietic lineages, which revealed that SCL is required for both primitive and definitive haematopoiesis (Porcher et al., 1996; Robb et al., 1996) although endothelial cells were still observed in the $SCL^{-/-}$ knockout mice (Robb et al., 1996). Rescue experiments in $SCL^{-/-}$ ES cells revealed that SCL is required for primitive and definitive haematopoiesis at the mesodermal stage (Endoh et al., 2002).

Conditional knockout studies in mice, which circumvent the early lethality observed in $SCL^{-/-}$ mice, demonstrated that SCL is crucial for erythroid and megakaryocytic development in adult mice (Hall et al., 2003). Ablation of SCL completely disrupts erythropoiesis and megakaryopoiesis while the myeloid lineage remains unaffected. Primitive progenitors were also shown to lose their ability to generate erythroid and megakaryocytic cells. However, despite being an important gene for erythropoiesis, SCL is not essential for the generation of mature red blood cells in adults suggesting a possible alternative factor governing this process (Hall et al., 2005).

(ii) Endothelial development.

In addition to being a regulator in haematopoietic development, SCL has been shown to play a crucial role in endothelial development and angiogenesis. In a study of a transgenic knock-in disruption of the SCL locus and an separate study using transgenic rescue of $SCL^{-/-}$ embryos, SCL has been shown to be required for the remodelling of capillary networks to form complex branching vitelline vessels in yolk sacs (Elefanty et al., 1999; Visvader et al., 1998).

SCL has been shown to be required for the generation of blast colonies from blast colony forming cells (BL-CFCs), an *in vitro* equivalent of the haemangioblast (Chung et al., 2002; Robertson et al., 2000). However, contradicting studies showed that $SCL^{-/-}$ cells initiate colony growth but cannot generate endothelial and haematopoietic progeny (D'Souza et al., 2005). The ability to give rise to blast colonies can, however, be rescued by ectopic expression of SCL. This suggests that SCL is essential for commitment of haematopoietic and endothelial lineages from haemangioblast but not for its development.

(iii) HSC self-renewal and repopulating activity.

The role of SCL in the development of HSCs has also been defined. Conditional knockout of SCL in mice demonstrated that SCL is dispensable for the long-term repopulating activity and differentiation into myeloid and lymphoid lineage of HSCs, but is required for the genesis of HSCs (Mikkola et al., 2003). Consistent with this finding, Curtis et al. also suggested that SCL is not required for self-renewal of HSCs but is important for their short-term repopulating capacity (Curtis et al., 2004). A contradicting study on enforced expression in long-term SCID (severe combined immunodeficient) mouse-repopulating cells (LT-SRCs), demonstrated that the expression level of SCL plays a pivotal role in the self-renewal and engraftment of HSCs and this regulation requires the DNA-binding domain of SCL (Reynaud et al., 2005).

(iv) T-cell leukaemia.

Chromosomal rearrangement of SCL is the most common cause of T-ALL and results in the activation of SCL expression in T cells, where it is normally down-regulated. The majority of translocations involves the TCR delta locus which results in the disruption of the promoter and 5' regulatory regions of SCL whilst the full-length coding sequence is unaffected (Begley et al., 1989a; Bernard et al., 1991). Translocation breakpoints were also identified downstream of the SCL coding regions resulting in the formation of an amino truncated protein under the cryptic promoter located in exon 4 (Bernard et al., 1992). An additional rearrangement involved a 5' interstitial deletion in the SCL locus which removes the 5' regulatory elements of SCL and the coding sequence of the SIL gene located immediately upstream of SCL. This results in the expression of SCL under the control of the SIL promoter (Aplan et al., 1990b; Bernard et al., 1991).

There are other known molecular mechanisms, related to SCL binding partners (see section F below), by which chromosomal rearrangement induces tumour formation. LMO1 and LMO2 are also targets of chromosomal rearrangement in T-ALL and are found to be co-expressed with SCL in T-ALL (Wadman et al., 1994). Thus, abberant expression of SCL and LMO2 in T-ALL may induce the expression of genes which are normally silent in T cells including RALDH-2 and TALLA-1

although the relevance of these genes to leukaemogenesis remains obscure (Ono et al., 1997, 1998). A different mechanism involves the sequestering of SCL-interacting partners E2A/TCF3 and HEB by the heterodimerisation between SCL and its partners. This disrupts the homodimerisation or heterodimerisation of E2A/TCF3 and HEB resulting in impaired regulation by these proteins (O'Neil et al., 2004).

E. Transcriptional regulation of SCL

To ensure appropriate expression during haematopoietic differentiation, expression of SCL is tightly regulated as described below (Figure 1.12).

(i) Promoters.

Three promoters have been identified to control transcription of SCL and are conserved in mouse and human: promoter 1a, 1b and an additional promoter in exon 4 (active only in leukaemic T-cells and T-ALL) (Bernard et al., 1992). Promoter 1a is active in erythroid and megakaryocytic lineages and mast cells while promoter 1b is silent in erythroid cells but active in primitive myeloid progenitors and mast cells. GATA1 cooperates with SP1 and SP3 to regulate the promoter 1a of SCL in erythroid cells and mast cells (Bockamp et al., 1998; Lecointe et al., 1994). Promoter 1b is active in primitive myeloid cells but functions in a GATA1-independent manner (Bockamp et al., 1997). Transcription factors PU.1, Elf-1, SP1, and SP3 were found to bind to promoter 1b and transactivate promoter 1b in mast cells (Bockamp et al., 1998).

(ii) Enhancers.

DNase I hypersensitivity assays and reporter assays have been used to identify and characterise putative regulatory elements at the SCL locus in human (Leroy-Viard et al., 1994) and mouse (Fordham et al., 1999; Gottgens et al., 1997). It has been demonstrated that the activity of these regulatory elements have overlapping but distinct features in various haematopoietic cell types. The putative enhancers at the human and mouse locus and their usage in different cell types and tissues are summarised in Figure 1.12.

The stem cell enhancer (+17 /+18 in mouse; +20/+21 in human) was demonstrated to be active in erythroid and mast cells but silent in primitive myeloid cells (Fordham et al., 1999; Gottgens et al., 1997). It was shown to activate both promoters 1a and 1b (Fordham et al., 1999). Further characterisation of this enhancer in transgenic mice demonstrated it targets expression in extra-embryonic mesoderm and both endothelial cells and haematopoietic progenitor cells in the yolk sacs, AGM, fetal liver and bone marrow (Sanchez et al., 1999). Exogenous SCL expression driven by the stem cell enhancer was shown to rescue early haematopoietic development in $SCL^{-/-}$ embryos which further strengthens its involvement in SCL regulation during stem cell development

(Sanchez et al., 2001). Analyses were done to further refine the core region for the enhancer activity and identified a 641 bp region containing the +19 site (Gottgens et al., 2002). This +19 core enhancer contains conserved Myb, Ets and GATA sites and these sites were shown to be bound by GATA2, Fli-1 and Elf-1. The activity of the core +19 enhancer is similar to the +18/+19 enhancer except that it is not sufficient to drive expression in definitive erythroid cells, suggesting addition elements are required for full function (Silberstein et al., 2005). Despite its proven enhancer activity, it was later shown that this stem cell enhancer is dispensable for SCL transcription and haematopoietic cell formation in a mouse knockout study (Gottgens et al., 2004). This suggests that additional regulatory elements are necessary for SCL expression. A -3.8 enhancer was identified subsequently which targets expression in haematopoietic progenitors and endothelium and is bound by Fli-1 and Elf-1 (Gottgens et al., 2004).

A systemic mapping of histone acetylation at the SCL locus identified peaks at the known SCL enhancers and promoters and one additional site 40 kb downstream of exon 1a in mouse (called the +40 region) and at the corresponding conserved +50/+51 region in human (Delabesse et al., 2005). The +40 region was shown to have enhancer activity *in vitro* and target expression in primitive but not definitive erythroid cells *in vivo*. Further analyses of this +40 enhancer indicated that it also targets expression in midbrain but not endothelial cells, and at the same time identified two indispensable GATA/E-box motifs which are bound by SCL and GATA1 in mouse erythroid cells *in vivo* (Ogilvy et al., 2007). The putative +50/+51 enhancer in human was shown to have highly conserved GATA/E-box motifs at +51 and that GATA1, SCL, and LDB1 are bound to this region in a human erythroid cell line (Pawan Dhami, PhD thesis). Transient reporter assays also demonstrated its enhancer activity. Thus, the murine +40 and the human +51 enhancers may function in the auto-regulation of SCL expression in erythroid cells.

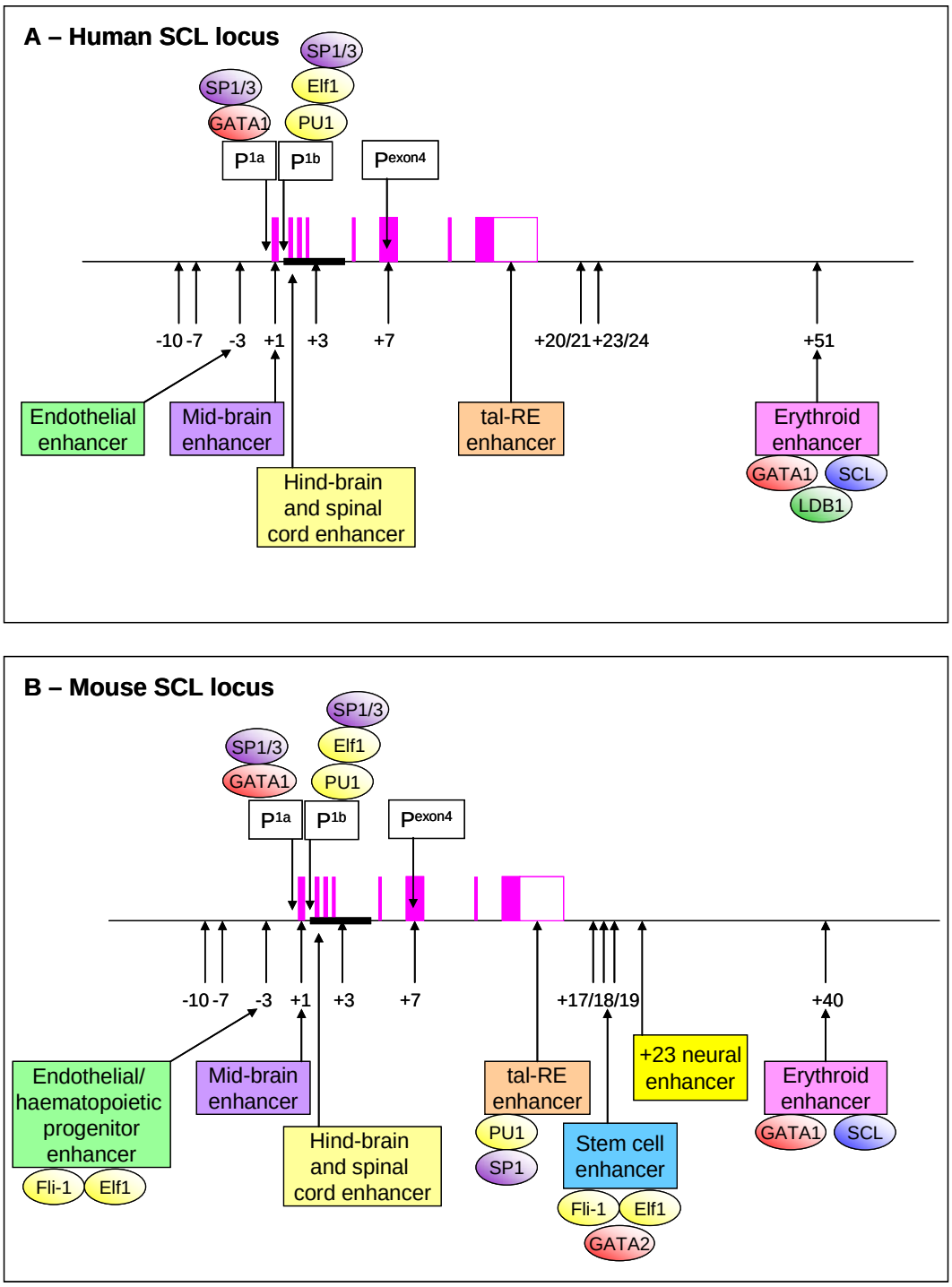


Figure 1.12. Schematic diagrams of regulatory elements at the human and mouse SCL loci. A: human SCL locus; B: mouse SCL locus. The pink boxes show the exons of the SCL gene. +/- numbers refer to the distance in kilobases of each DNase hypersensitive sites (HSs) from promoter 1a. Coloured boxes show the known enhancers while white boxes show the known promoters. Ovals denote proteins which are known to bind to the enhancers/promoters. Detailed description of each component is provided in the text (images shown are not to scale).

F. The SCL protein, interacting partners and downstream targets

The SCL gene encodes a class B basic-helix-loop-helix (bHLH) protein with two isoforms - a 42 kDa full length protein and a 22 kDa amino truncated form (Elwood et al., 1994; Goldfarb et al.,

1992). The bHLH region remains present in both isoforms and is required for nuclear localisation, DNA binding and protein-protein interactions (Hsu et al., 1991; Hsu et al., 1994b).

SCL protein heterodimerises with class A bHLH proteins such as E2A/TCF3, HEB and E2-2 which is a requirement for DNA binding (Hsu et al., 1991; Hsu et al., 1994a). In addition, SCL forms a transactivating protein complex in erythroid cells with other transcription factors including the haematopoietic specific proteins GATA1, LMO2 and ubiquitously expressed proteins E2A/TCF3 (transcript variants E12 and E47) and LDB1 (Wadman et al., 1997) (Figure 1.13). This complex binds to a bipartite DNA motif consisting of an E-box (CANNTG) ~9 bp upstream of a GATA site. The SCL-E2A heterodimer binds to the E-box motif while GATA1 binds to the GATA site. LMO2 and LDB1 do not bind DNA directly - instead, they act as bridging proteins between the SCL-E2A heterodimer and GATA1. More complete descriptions of these binding partners are found in sections 1.4.2.2 – 1.4.2.5.

In addition to these members of the SCL complex, a novel component, ETO2, was found to be recruited to the complex by interacting with E2A (Goardon et al., 2006). ETO2 is a repressor protein and was shown to negatively regulate expression of one target gene GPA (see below). It was demonstrated that changes in the amount of ETO2 protein in this complex governs the expression of erythroid specific genes and is a key determinant in terminal erythroid differentiation.

Since the discovery of SCL and its multiprotein erythroid complex, only a few target genes have thus far been identified in the erythroid lineage (see below). Regulation by SCL in T-cell acute lymphoblastic leukaemia (T-ALL) was studied by comparing the expression of genes in the human genome between SCL-expressing and non-expressing human T-ALL samples using expression arrays and ChIP-on-chip analysis with promoter arrays (Palomero et al., 2006). The results demonstrated that SCL functions as both a repressor and an activator in T-ALL.

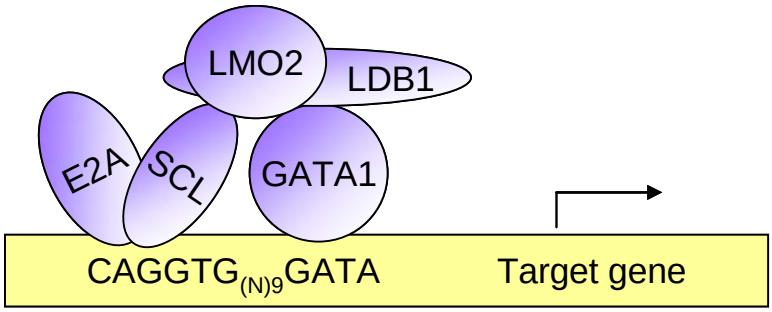


Figure 1.13. The SCL erythroid complex. The ovals or circles indicate the proteins involved in this complex. GAGGTG is the E-box motif; GATA is the GATA site; (N)9 indicates the two motifs are separated by 9 nucleotides. The arrow indicates transcription of target genes. A detailed description of the complex and its target genes are provided in the text.

(i) c-kit.

c-kit was one of the first SCL target genes identified. The c-kit gene encodes a tyrosin receptor kinase which is required for normal haematopoiesis. Expression of c-kit was shown to correlate with SCL expression suggesting a regulatory role of SCL in $CD34^+$ haematopoietic cells (Krosl et al., 1998). Chromatin immunoprecipitation studies revealed that members of the SCL erythroid complex, together with a novel member, specificity protein 1 (Sp1, a zinc-finger protein) occupies the c-kit promoter and the combinatorial interaction of all the members of this complex is essential for the synergistic transactivation of c-kit (Lecuyer et al., 2002). GATA2, another member of the GATA family, was also found in the complex and was shown to convey greater transcriptional activation on the c-kit promoter than GATA1. In a separate study, the pentamer protein complex consisting of SCL, E12, LMO2, LDB1 and the retinoblastoma protein (pRb) was shown to inhibit c-kit expression during erythropoiesis (Vitelli et al., 2000).

(ii) Glycophorin A (GPA).

The erythroid cell-specific glycophorin A gene (GPA) was identified as a target of the SCL erythroid complex (including Sp1) in primary hematopoietic cells (Lahlil et al., 2004). The complex was shown to occupy the GPA promoter *in vivo* and to activate GPA expression with GATA1, rather than GATA2, conveying a greater degree of transcriptional activation.

(iii) α- and β-globin genes.

The SCL erythroid complex was also found to occupy the human β-globin locus control region (LCR) during erythroid differentiation (Song et al., 2007). The long range interaction between the β-globin LCR and the active β-globin promoter requires LDB1 for the formation of the loop structure. The mouse and the human α-globin loci were found to be co-occupied by SCL, E2A, GATA1/2, LMO2 and LDB1 in DNase I hypersensitivity assays and ChIP-on-chip (Anguita et al., 2004; De Gobbi et al., 2007). However, no functional analysis of the α-globin clusters has been performed to investigate the role of the complex in globin regulation.

(iv) Protein 4.2 (P4.2).

The gene for protein 4.2 (P4.2), an important component of the erythrocyte cell membrane skeleton, is also a target gene of the SCL erythroid complex in mouse. SCL, E47, GATA1, LMO2 and LDB1 were demonstrated to activate P4.2 expression via two GATA E-box elements in the P4.2 promoter in erythroid cells (Xu et al., 2003). Maximal transcription requires both GATA and E-box sites and all five members of the complex. The SWI/SNF protein Brg1 was also found to associate with the complex and down-regulate P4.2 expression by recruiting chromatin-remodelling complexes and histone modification enzymes (Xu et al., 2006).

(v) GATA1.

The *cis*-acting regulatory element (HS-1) upstream of the promoter of GATA1 in mouse contains a composite E-box GATA site and was shown to be bound by the SCL eyrthroid complex (Vyas et al., 1999b). Mutations in the GATA1 site, but not the E-box site, significantly abolish the activation activity of the element.

(vi) FLK-1.

The tyrosine receptor kinase FLK-1 is important for the generation of common precursors for both the endothelial and haematopoietic lineages. The FLK intronic enhancer contains two E-box motifs, one indispensable GATA site and two ETS binding sites. These sites have been demonstrated to be bound by SCL, GATA1 and ETS proteins respectively. Mutations on these sites abolished the enhancer activity. Combinatorial action of these transcription factors regulates FLK-1 expression in both haematopoietic and vascular development (Kappel et al., 2000).

(vii) RUNX1/AML1.

The transcription factor RUNX1/AML1 is an important regulator of haematopoiesis and has recently been shown to be regulated by a multiprotein complex containing SCL in mouse. SCL, together with LMO2, LDB1, GATA2 and ETS were found to bind to the putative +23 enhancer of RUNX1 located 23 kb downstream of the transcription start site of RUNX1 *in vivo* in a myeloid progenitor cell line. This +23 enhancer contributes to expression of RUNX1 in early haematopoiesis (Nottingham et al., 2007). Direct binding of SCL, LMO2 and GATA2 was confirmed by chromatin immunoprecipitation in another study and gene expression profiling also revealed that RUNX1, together with RUNX3, are downstream targets of SCL in early haematopoietic development (Landry et al., 2008).

1.4.2.2 GATA1

A. <u>The GATA1 gene and its expression</u>

GATA1, also named NF-E1, NF-1, Ery-1 and GF-1, was first identified as a protein bound to the 3' enhancer of the β-globin gene (Wall et al., 1988). It was later mapped to human chromosome X at Xp21-22 (Zon et al., 1990). The GATA1 gene locus contains six exons where the first exon is non-coding. GATA1 belongs to the GATA family of genes including GATA1 to 6 where GATA1, 2 and 3 are important in haematopoietic development.

GATA1 is widely expressed in various lineages of haematopoietic development. In many respects, its expression patterns mirror those of SCL (see section 1.4.2.1). It is expressed in primitive and definitive erythroid cells (Tsai et al., 1989; Zon et al., 1990), megakaryocytes (Martin et al., 1990), eosinophils (Zon et al., 1993) and mast cells (Martin et al., 1990). It is also expressed in testis

Sertoli cells in mice (Yomogida et al., 1994). At earlier stages of haematopoietic development, GATA1 is expressed in HSCs, common myeloid progenitors (CMP), megakaryocyte/erythrocyte lineage restricted progenitors (MEP) and haemangioblast (Akashi et al., 2000; Kuhl et al., 2005; Yokomizo et al., 2007). During erythroid differentiation from HSCs, GATA1 was found to be expressed at low levels initially while its expression gradually increases as erythroid differentiation progresses.

B. Functions of GATA1

- **Erythropoiesis**

Several lines of evidence have demonstrated a crucial participation of GATA1 in erythroid development. Deletion of GATA1 in mouse ES cells resulted in contribution to all haematopoietic tissues except mature red blood cells in chimeric mice (Pevny et al., 1991). *In vitro* colony assays further suggested that the GATA1 null cells failed to mature beyond proerythroblasts, a cell type found at an early stage of terminal differentiation (Pevny et al., 1995). Similarly, GATA1 null chimeric mice died between E10.5 to E11.5 of anaemia and displayed embryonic erythroid cells arrested at the proerythroblast stage (Fujiwara et al., 1996). This further established the importance of GATA1 in both primitive and definitive erythropoiesis. *In vitro* differentiation of GATA1 null ES cells confirmed that the proerythroblast arrest and death by apoptosis and thus suggesting GATA1 supports the viability of red blood cell precursors by suppressing apoptosis (Weiss et al., 1994; Weiss and Orkin, 1995). Inducible rescue of GATA1 null erythroblasts demonstrated that GATA1 promotes terminal erythroid maturation and G1 cell cycle arrest by suppressing the expression of c-MYC, a proto-oncogene which regulate cell proliferation and differentiation (Rylski et al., 2003). The interaction between GATA1 and its co-factor FOG1 has been shown to be required for terminal erythroid maturation (Rylski et al., 2003).

- **Megakaryopoiesis**

GATA1 also plays a critical role in megakaryocytic development. GATA1-deficient mice were shown to have reduced platelet counts as well as expansion of immature megakaryocytes (Shivdasani et al., 1997). These megakaryocytes have abnormal morphology, are unable to mature and exhibit a marked hyperproliferation *in vivo* and *in vitro* (Vyas et al., 1999a). At the molecular level, GATA1 activates transcription of megakaryocyte specific genes including NF-E2, GP1bα and platelet factor 4.

- **Eosinophils and mast cells development**

It was first shown that GATA1 could convert chicken myeloblasts, mouse common lymphoid progenitors and human myeloid progenitors to eosinophils (Hirasawa et al., 2002; Iwasaki et al.,

2003; Kulessa et al., 1995). Disruption of GATA sites in the GATA1 promoter resulted in selective loss of the eosinophil lineage (Yu et al., 2002a). Mast cell development was also shown to be disrupted in GATA1^{low} mice where the first enhancer and distal promoter of GATA1 are deleted (Migliaccio et al., 2003); the mast cells produced were defective in terminal maturation and had increased apoptosis. At the molecular level, GATA1 has been shown to activate expression of eosinophil specific genes such as MBP (Yamaguchi et al., 1998; Yamaguchi et al., 1999).

- **GATA1 and leukaemia**

One of the more well-studied disorders associated with mutation in GATA1 is the transient myeloproliferative disorder (TMD) which occurs in about 10% of children with Down syndrome. In 20% of the TMD cases, patients develop Down syndrome-related acute megakaryocytic leukaemia (DS-AMKL) later in life. In most cases, the mutations in GATA1 introduce a premature stop codon or a splice site in the N-terminal activation domain which results in the translation of a GATA1s isoform lacking the N-terminal activation domain (Wechsler et al., 2002). GATA1s has diminished transactivation potential in *in vitro* assays and causes a reduction in differentiation of megakaryoctic precursor cells.

A number of missense mutations in the N-finger of GATA1 have also been found in patients with X-linked thrombocytopenia and anaemia (Table 1.13). In most of these cases, the ability of GATA1 to interact with FOG1 or to bind DNA is affected. The severity of the disease depends on the particular type of mutation.

Mutation	FOG1-binding	DNA-binding	Phenotype	References
V205M	Strongly reduced	Not affected	Macrothrombocytopenia; Dyserythropoietic anaemia	(Nichols et al., 2000)
D218G	Reduced	Not affected	Macrothrombocytopenia; Dyserythropoiesis without anaemia	(Freson et al., 2001)
D218Y	Strongly reduced	Not affected	Macrothrombocytopenia; anaemia	(Freson et al., 2002)
G208S	Reduced	Not affected	Macrothrombocytopenia	(Mehaffey et al., 2001)
R216Q	Not affected	Reduced binding to complex and palindromic sites	Macrothrombocytopenia; β-thalassemia	(Balduini et al., 2004; Yu et al., 2002b)

Table 1.13. Mutations of GATA1 in X-linked thrombocytopenia and anaemia.

C. Transcriptional regulation of GATA1

- **Cis-regulatory elements**

Together with its upstream region, the first untranslated exon of GATA1 contains regulatory elements for GATA1 expression which are conserved across vertebrates (Figure 1.14). The erythroid-specific promoter region located upstream of the erythroid first exon (IE) contains a CACCC box and a double GATA site necessary and sufficient to drive expression in erythroid cells

(Zon et al., 1990). The CACCC box is essential for initiation of GATA1 gene expression as mutations or deletions therein completely disrupted promoter activity in zebrafish (Meng et al., 1999). The GATA sites, in contrast, are not essential for lineage-specific expression (Nicolis et al., 1991; Yu et al., 2002a). In mouse, there is an additional promoter upstream of the first exon testis-specific exon (IT) which is used in Sertoli cells (Onodera et al., 1997b).

Upstream regions of the GATA1 IE promoter there is an enhancer element for GATA1 expression in primitive and definitive erythropoiesis, as well as in megakaryocytes and eosinophils (Onodera et al., 1997a). This region is denoted as the enhancer G1HE. The core regions of G1HE which contains a GATA site or a GATA site plus a CACCC box are required for expression in erythroid cells and megakaryocytes respectively (Nishimura et al., 2000; Vyas et al., 1999b). An additional intronic enhancer intron-SP located in the first intron which contains GATA and AP1 repeats is required for efficient expression in definitive erythroid cells (Onodera et al., 1997a). The testis-specific enhancer in mouse G1TAR is required for activation of the IT promoter (Wakabayashi et al., 2003).

- **Trans-acting proteins**

The presence of a number of GATA sites in the regulatory region of GATA1 suggests that GATA factors may bind to these regions. Indeed, GATA1 has been shown to bind to the G1HE, the double GATA site in promoter IE and the intronic enhancer intron-SP *in vivo* by chromatin-immunprecipitation assays (Valverde-Garduno et al., 2004). Overexpression of GATA1 upregulates a transgenic GATA1 reporter gene in zebrafish and the self-association of GATA1 is required for this regulation (Kobayashi et al., 2001; Nishikawa et al., 2003; Shimizu et al., 2007). Furthermore, suppression of GATA2 gene expression down-regulates GATA1 expression in blast cells but has no effect in differentiated cells (Lugus et al., 2007; Tsai and Orkin, 1997). All these findings suggest that GATA1 is under the control of GATA2 at an early stage of development, while GATA1 is involved in autoregulation later in development.

Other transcription factors also play important roles in GATA1 transcriptional regulation. PU.1 antagonises GATA1 expression by hindering the binding of GATA1 to the GATA1 locus and thus inhibits autoregulation (Zhang et al., 2000). It is also suggested that PU.1 inhibits GATA1 expression by creating a repressive chromatin structure (Stopka et al., 2005). The SCL erythroid complex containing SCL, LMO2, LDB1, E2A and GATA1 (Section 1.4.2.1 F) is recruited to the G1HE in erythroid cells *in vivo* (Valverde-Garduno et al., 2004). CP2 has been shown to bind to the upstream region of erythroid specific first exon at two CP2-binding sites adjacent to the double GATA site bound by GATA1. Mutation in these CP2 sites impair promoter activity in erythroid

cells (Bose et al., 2006). This suggests a functional cooperation of the two factors in controlling expression of GATA1.

The Sp1/Krüppel-like factor (KLF) family binds to the CACCC box. Expression of KLF2 and KLF6 has been shown to correlate with expression of GATA1 in mice suggesting a regulatory role played by these factors (Basu et al., 2005; Matsumoto et al., 2006).

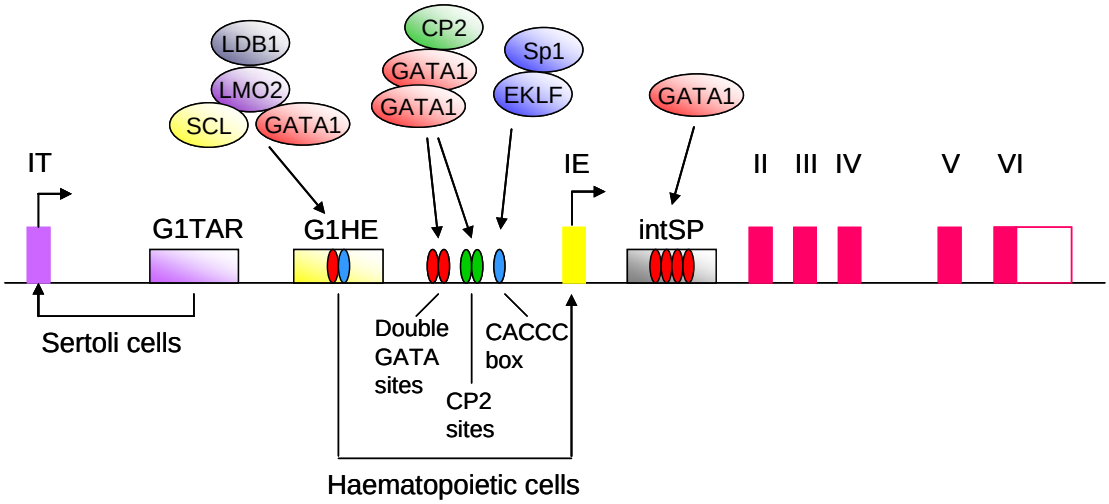


Figure 1.14. A schematic diagram of the regulatory regions of GATA1. The pink boxes show the exons of the GATA1 gene. The GATA1 testis activation region (G1TAR) activates transcription from the testis-specific first exon (IT). Both G1TAR and IT are only found in mouse. The GATA1 haematopoietic enhancer (G1HE) activates transcription from the haematopoietic-specific first exon (IE). The intSP is an erythroid-specific enhancer found in the first intron. Small ovals indicate specific motifs in the regulatory elements: red: GATA sites; blue: CACCC box; green: CP2 sites. Large ovals denote proteins which are known to bind to the enhancers/promoters. Detailed description of each component is provided in the text (image shown is not to scale).

D. The GATA1 protein and interacting partners

GATA1 encodes a protein which belongs to the GATA family of transcription factors (GATAs 1 through 6) and contains three functional domains: the N- and C-terminal zinc finger motifs and the N-terminal activation domain. The C-terminal zinc finger binds to the DNA consensus sequence (A/T)GATA(A/G) whereas the N-terminal zinc finger functions by binding to DNA and recruiting co-factors and contributes to the stability and specificity of DNA-binding. The N-terminal activation domain confers transcriptional activation to target genes (Martin and Orkin, 1990).

GATA1 has been shown to physically interact with a variety of nuclear proteins, as well as to self-dimerise. Such interactions are essential for the function of GATA1 as a transcriptional regulator and are pivotal in haematopoietic development. Table 1.14 summarises the co-factors or transcription factors which interact with GATA1.

Co-factor	Protein functions	Mechanism of action	Reference
GATA1	Zinc-finger protein required for erythroid and megakaryocytic development	Homodimerisation requires N-finger-C-finger contacts and induce transcriptional activation.	(Mackay et al., 1998)
Friend of GATA1 (FOG1)	Zinc-finger protein required for erythroid and megakaryocytic development	Interacts with N-finger of GATA1; co-operates with GATA1 in erythroid and megakaryocytic differentiation; both synergises with or represses GATA1-mediated activation.	(Fox et al., 1999; Tsang et al., 1997)
LMO2	LIM-domain protein implicated in T-ALL	Interacts with GATA1 as a requirement for the formation of the SCL erythroid complex together with LDB1 and E2A; whole complex trans-activates a number of erythroid-specific genes,	(Wadman et al., 1997) Section 1.4.2.1E
p300/CBP	Histone acetyltransferase	Interacts with GATA1 *in vivo* and *in vitro*; stimulates GATA1 activity by acetylating GATA1.	(Boyes et al., 1998)
Erythroid Krüppel-like factor (EKLF)	Erythroid-specific Zinc finger protein of the Krüppel-like factor family	Interacts with C-finger of GATA1; functions as a co-regulator with GATA1.	(Gregory et al., 1996; Merika and Orkin, 1995)
Sp1	Krüppel-like factor family protein required for early embryonic development	Interacts with C-finger of GATA1; activates GATA1 transcriptional activity.	(Gregory et al., 1996; Merika and Orkin, 1995)
PU.1	Ets protein required for lymphoid and granulocytic development	Interacts with GATA1 via the DNA-binding domain; antagonises GATA1 transcriptional activation by preventing its binding to DNA.	(Rekhtman et al., 1999; Zhang et al., 2000)
CP2	Transcription factor which stimulates α-globin expression	Interact with C-finger of GATA; functions as a co-regulator with GATA1 in regulation of erythroid genes	(Bose et al., 2006)
c-myb	Oncogene required for haematoppoiesis	Interacts with GATA1 and inhibits its DNA-binding activity.	(Matsumura et al., 2000)
Fli-1	Ets protein important for megakaryocyte differentiation	Interacts with zinc fingers of GATA1 via its own Ets domain; functions as a co-activator for genes involved in terminal megakaryocytic differentiation.	(Eisbacher et al., 2003)

Table 1.14. Interacting partners of GATA1.

E. Downstream targets of GATA1.

A large set of genes, especially those related to haematopoiesis, have been characterised as target genes of GATA1 and some of the more well-characterised target genes are discussed below. Genes whose regulation is mediated through the SCL erythroid complex containing GATA1 are not discussed here (refer to section 1.4.2.1 F).

(i) α- and β-globin genes.

GATA1 was first identified by its binding to an enhancer at the β-globin locus (Wall et al., 1988). ChIP-on-chip analyses of the β-globin locus demonstrated that GATA1 binds to a region of the HS2 core element and an additional region upstream of γ-globin gene (Horak et al., 2002). Subsequently,

GATA1 was shown to bind to the α-globin locus (Evans et al., 1988). It has been shown that GATA2 drives α-globin expression in multipotent progenitors while GATA1 replaces GATA2 in committed erythroid progenitors where it is bound to the α-globin promoter (Anguita et al., 2004).

(ii) EKLF.

In addition to being an interacting partner of GATA1, EKLF was also identified as a target gene of GATA1. Forced expression of GATA1 in non-erythroid cells induced activation of the EKLF promoter while one of the GATA sites in the promoter of EKLF was found to be indispensable for promoter function (Crossley et al., 1994). GATA1 was shown to bind to two GATA sites in a GATA-E-box-GATA motif in the promoter of EKLF which is essential for EKLF expression (Anderson et al., 1998, 2000). Functional interaction of CP2 and GATA1 may contribute to the regulation at the EKLF promoter (Bose et al., 2006).

(iii) GATA2.

GATA1 has been shown to repress expression of GATA2 in erythroid differentiation (Weiss et al., 1994). GATA1 binds to a region upstream of promoter 1G of GATA2 which is normally bound by GATA2 itself (Grass et al., 2003). GATA2, when bound to its own promoter, recruits CBP leading to histone acetylation and transcriptional activation. Displacement of GATA2 by GATA1 disrupts this autoregulation and thus represses GATA2 expression.

(iv) Epo and EpoR.

Erythropoietin (Epo) is a major growth factor for erythroid cells which binds to the Epo receptor (EpoR), a cell surface marker, resulting in proliferation and differentiation of erythroid progenitors. GATA1 was found to bind and transactivate the EpoR promoter (Zon et al., 1991). Conversely, GATA1 acts as a repressor for Epo expression and binds to a GATA site in the Epo promoter (Imagawa et al., 2002; Imagawa et al., 1997).

(v) NF-E2.

Abrogation of GATA1 expression was shown to significantly reduce expression of NF-E2 in megakaryocytes (Vyas et al., 1999a). GATA1-mediated activation acts in concert with human FOG2 (Holmes et al., 1999). Further analyses of the NF-E2 promoter in mouse demonstrated that GATA1 bind to the proximal promoter 1B located in the first intron (Moroni et al., 2000).

(vi) GFI-1B.

Gfi-1B is an erythroid-specific transcription factor which plays an essential role in erythropoiesis. ChIP assays demonstrated that GATA1 binds to the promoter region of GFI-1B (Huang et al., 2004). Ectopic expression of GATA1 in non-erythroid cells activates the GFI-1B promoter. This

direct activation is also dependent on NF-Y which also binds to GFI-1B promoter. GFI-1B itself suppresses the GATA1-mediated activation by protein-protein interaction (Huang et al., 2005).

(vii) FOG1.

FOG1 is a co-factor and binding partner of GATA 1 (see Table 1.14). Global expression analysis revealed that FOG1 expression is rapidly induced by GATA1 expression and ChIP studies confirmed the binding of GATA1 to GATA motifs in the *cis*-regulatory elements of FOG1 (Welch et al., 2004). This suggests a regulatory hierarchy where GATA1 first induces expression of its co-factor for a co-operative activation of the β-globin gene (Welch et al., 2004).

(viii) c-MYC.

The proto-oncogene c-MYC is a transcription factor which binds to E-box motifs and recruits histone acetyltransferases. GATA1 has been shown to repress c-MYC expression and binds to its promoter in mouse erythroid cells (Rylski et al., 2003).

1.4.2.3 E2A/TCF3

A. <u>The E2A/TCF3 gene and gene products</u>

The E2A gene, also named TCF3, was first identified as two highly similar cDNA clones whose dimerised products bind specifically to the human immunoglobulin kappa chain enhancer (Murre et al., 1989). The gene was mapped on chromosome 19p13.3 and contains 19 exons. E12 and E47 are two splicing variants produced by alternative splicing of exons 17 and 18.

E12 and E47 are the founding members of basic helix-loop-helix (bHLH) family of transcription factors. They belong to class A of the bHLH proteins including HEB and E2-2 which bind to DNA elements with the consensus E-box sequence CANNTG. Both E12 and E47 are virtually identical except that the C-terminal bHLH domains are slightly different (due to the alternative splicing of exons 17 and 18). There are two activation domains, AD1 and AD2, located at the N-terminus and in the central region of the protein, which mediate transcriptional activation by recruiting histone acetyltransferases (Massari et al., 1999; Qiu et al., 1998). E12 and E47 form homodimers or heterodimers with class B bHLH protein such as MyoD where the protein interaction is mediated by the bHLH domain. E47 homodimers, and heterodimers between MyoD and E47 or E12 can bind DNA; whereas E12 homodimers fail to bind DNA due to the presence of an inhibitory domain in the basic region of E12 (Sun and Baltimore, 1991).

Like other class A bHLH proteins, E2A proteins are ubiquitously expressed in a variety of cell types and tissues. However, expression of E2A has been shown to be up-regulated during B-cell lineage commitment (Zhuang et al., 2004).

B. Functions of E2A

- **Regulation of tissue-specific differentiation**

Despite its ubiquitous expression pattern, E2A can still function as a co-regulator in tissue-specific differentiation. This is mediated by the formation of heterodimers between E2A and class B tissue-specific bHLH proteins. One of the more well-studied examples is the regulation of myogenesis. E2A dimerises with MyoD and regulates expression of several downstream muscle-specific regulators which, in turn, control muscle differentiation (Lassar et al., 1991). Another example is the dimerisation of E2A and SCL which leads to the formation of multiprotein complexes and direct transcriptional activation or repression of erythroid-specific genes (Section 1.4.2.1 F).

- **Transcriptional activation of B-cell specific genes**

E2A is a key transcription factor regulating transcription of B-cell specific genes. Early B-cell factor (EBF), an important regulator of B-cell commitment and lineage-specific gene expression, is one of the more well-characterised targets of E2A. Ectopic expression of E12 induced expression of EBF and the promoter of EBF functionally interacts with E47 (Kee and Murre, 1998; Smith et al., 2002). However, E2A itself is not sufficient to drive EBF expression, as PU.1 has been shown to work independently or in a cooperative manner with E2A to direct EBF expression (Medina et al., 2004).

E2A and EBF are involved in the regulation of an overlapping set of B-lineage specific genes including genes crucial for gene rearrangement and BCR expression (Mansson et al., 2004). However, some of these genes may be secondary targets of E2A mediated by EBF activation, although a subset have been shown to have direct association between E2A and their regulatory region in ChIP (Greenbaum and Zhuang, 2002).

- **Regulation of lymphoid development**

Homozygous E2A mutant mice or knockout mice contained no mature B cells while all other haematopoietic lineages were intact (Bain et al., 1994; Zhuang et al., 1994). Detailed examination of the defect in B-cell differentiation revealed that B-cell development was blocked at the stage before IgH DJ rearrangement and before the Pro-B cell formation. More recent studies in E2A knockout haematopoietic progenitor cells show a characteristic pro-B cell signature indicative that these cells are pluripotent (Ikawa et al., 2004); they expressed genes specific to other lineage but not the B-cell lineage, and they could contribute to all blood lineage except B-cells. Taken together, these data indicate that E2A is required for B-lineage restriction and commitment.

E2A, together with other class A bHLH proteins like HEB, are also involved in T-cell development and lineage commitment which requires the formation of heterodimers between E2A and HEB

(Barndt et al., 2000). In contrast to the B- and T- lineages, suppression of E2A function has been implicated in natural killer (NK)-cell development. This is mediated by dimerisation of E2A with the Id protein, thus inhibiting E2A from binding to DNA (Heemskerk et al., 1997).

- **Translocation and leukaemia**

Chromosomal rearrangements involving the E2A gene result in acute lymphoblastic leukemia (ALL). Translocation between chromosome 1 and 19 resulted in the formation of a fusion E2A-PBX1 protein while translocation between chromosome 17 and 19 leads to expression of E2A-HLF (Inaba et al., 1992; Nourse et al., 1990). Such translocations disrupt the normal gene regulatory networks of the proteins involved and the fusion proteins may also cause abnormal transcriptional upregulation of its target genes.

1.4.2.4 LMO2

A. <u>The LMO2 gene and gene product</u>

LMO2 belongs to the LIM-only family of genes which was first discovered by virtue of its translocation product in T-ALL (Boehm et al., 1991). It is located on chromosome 11p13 and contains three exons. It encodes a protein comprising two tandem LIM domains which are zinc-binding finger-like domains structurally similar to the DNA-binding GATA fingers. However, unlike the GATA fingers, LIM domains have not been shown to bind DNA but are restricted to protein-protein interaction. LMO2 is ubiquitously expressed in blood progenitor cells and endothelial cells (Delassus et al., 1999; Yamada et al., 2000).

B. <u>Functions of LMO2</u>

- **Regulation of haematopoietic and endothelial development**

LMO2 plays a critical role in erythropoietic and endothelial development. LMO2 null mice display defects in primitive erythropoiesis and lethality around E10.5 (Warren et al., 1994). Additional studies of homozygous mutant $LMO2^{-/-}$ mouse ES cells showed that all haematopoietic lineages are disrupted in these $LMO2^{-/-}$ cells while ectopic expression of LMO2 rescues this phenotype (Yamada et al., 1998). Thus, LMO2 is necessary for all stages of definitive haematopoiesis and it functions at least at the level of pluripotent stem cell. These studies, when considered together, show that LMO2 has a very similar function as SCL in early haematopoietic development. This suggests that protein-protein interaction between LMO2 and SCL and possibly other transcription factors governs haematopoiesis (Wadman et al., 1997). A further LMO2 null study has also demonstrated that LMO2 is not required for the generation of the primary capillary network (vasculogenesis) but it is

crucial for remodelling of this capillary network into the mature vascular system (angiogenesis) (Yamada et al., 2000).

- **Translocation and leukaemia**

As mentioned above, LMO2 is involved in chromosome translocation between chromosome 11 and 14 or 11 and 7 which causes T-ALL. The translocation breakpoint is upstream of the LMO2 promoter and thus enforced expression of the full-length LMO2 protein is observed in T-cells. Ectopic expression of LMO2 in T-cells from transgenic mice resulted in a marked accumulation of immature T-cells indicating that LMO2 induces cell proliferation and blocks T-cell differentiation (Neale et al., 1995). Enforced expression of both SCL and LMO2 further enhanced this effect leading to the hypothesis that interaction between these two proteins can alter T cell development and potentiate tumorigenesis (Larson et al., 1996).

There are two models describing the mechanism of tumorigenesis by LMO2 translocations. In the first model, an LMO2 complex was described which is similar to its analog in erythroid cells (Grutz et al., 1998). This complex involves two E2A/SCL heterodimers which bind to two E-box motifs separated by one helix turn. LMO2 and LDB2 proteins are bridging protein for this multimer complex. This complex may regulate a specific subset of target genes involved in tumor development. The second model suggests that the abnormal expression of LMO2 may sequester members of protein complexes or dimmers, thus disrupting their normal functions.

1.4.2.5 LDB1

A. The LDB1 gene and gene product

The LIM-domain binding protein 1 (LDB1), also named CLIM2 or NL1, was first discovered due to its ability to interact with LIM domain proteins LIM-homeo-domain (LIM-HD) and LIM-only (LMO) (Agulnick et al., 1996). It was mapped on chromosome 10q24-25 and contains 11 exons. LDB1 is ubiquitously expressed in various tissues. The LDB1 protein contains three domains: a conserved nuclear localisation sequence (NLS), an N-terminal dimerisation domain (DD) and a C-terminal LIM interaction domain (LID). The DD mediates homodimerisation of LDB1, while LID mediates interaction with LIM-HD or LMO proteins. However, no DNA-binding or enzymatic activity has been observed in LDB1. Thus, it is likely that LDB1 functions exclusively through protein-protein interaction.

B. Functions of LDB1

- **Developmental regulator**

Deletion of LDB1 in mice induced severe defects in anterior-posterior patterning, truncation of head structures, posterior axis duplication and a lack of heart embryonic cells (Mukhopadhyay et al., 2003). This suggests that LDB1 plays important roles in diverse developmental process. The exact mechanism underlying the regulation of developmental processes regulated by LDB1 is not fully understood. However, the precise stoichiometry of LDB1 and its interacting partners LIM-HD or LMO proteins may be a critical criterion in determining the downstream pathways regulating various biological processes. One piece of evidence is that overexpression of LDB1 inhibits erythroid differentiation (Visvader et al., 1997).

- **Transcriptional modulator**

Although LDB1 does not bind DNA directly, its interaction with LIM domain proteins contributes significantly to its role as a transcriptional regulator. In additional to the target genes in the SCL erythroid complex containing LDB1 (described in section 1.4.2.1 F), LDB1 has also been found to repress the synergistic activation of the insulin enhancer by LMX1 and E47 (Jurata and Gill, 1997).

1.4.2.6 Transcriptional regulatory networks in haematopoietic development

Haematopoietic differentiation and lineage-specific commitment is a complex process regulated by multiple transcription factors or developmental critical genes. The 5 genes described above (SCL, GATA1, E2A, LMO2 and LDB1) belong to only a small subset of these regulators. These regulators, of which many are transcription factors and co-factors, both physically interact and/or are transcriptionally regulated by one another. This results in the generation of a global regulatory network (Section 1.2).

Swiers et al. 2006 first attempted to build a network for erythropoiesis based on data in mouse (Swiers et al., 2006). The authors identified the network motifs first described in *E. coli* and yeast which play essential roles at different stages of haematopoietic development. For instance, multi-input motifs such as SCL and Hex which are both regulated by the cooperation of GATA2, Fli-1 and Elf-1 are important for co-ordinating gene expression in specific lineage. Feed-forward motifs such as GATA1 → FOG1 → β-globin control the temporal expression of lineage-specific genes and prevent immature activation of certain genes. Autoregulation, such as in the case of GATA1, generates a forward momentum and stabilises expression of GATA1 in specific cell types. The authors concluded that, in summary, network motifs function together to provide a complex regulation of haematopoiesis.

1.5 Aims of this thesis

Although a lot is known about the functional roles of SCL and its binding partners during haematopoietic development, the molecular means by which these functions are regulated are not well understood. For example, only a handful of target genes have been identified in human or mouse for the SCL erythroid complex. The central aim of this thesis was to further characterise the regulatory network governed by this complex in human erythroid cells using a combination of approaches. Specifically, the aims of this work were:

1. To develop perturbation assays for SCL, GATA1, E2A, LMO2 and LDB1 by siRNA-mediated knockdown in the human erythroid cell line K562.
2. To study the downstream effects of each of the knockdowns and identify putative primary or secondary target genes by genome-wide expression analyses using Affymetrix GeneChips.
3. To identify and confirm direct target genes of the 5 transcription factors using ChIP coupled with a human transcription factor promoter array (ChIP-chip).
4. To generate a transcription network of the SCL erythroid complex using an integration of various experimental approaches described in aims 1-3.

(2004a). The ENCODE (ENCyclopedia Of DNA Elements) Project. Science *306*, 636-640.
(2004b). Finishing the euchromatic sequence of the human genome. Nature *431*, 931-945.
(2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature *447*, 799-816.
Agulnick, A.D., Taira, M., Breen, J.J., Tanaka, T., Dawid, I.B., and Westphal, H. (1996). Interactions of the LIM-domain-binding factor Ldb1 with LIM homeodomain proteins. Nature *384*, 270-272.
Akashi, K., Traver, D., Miyamoto, T., and Weissman, I.L. (2000). A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. Nature *404*, 193-197.
Akiyama, Y. (1998). TFSEARCH: Searching Transcription Factor Binding Sites. http://wwwrwcporjp/papia/.
Allfrey, V.G., Faulkner, R., and Mirsky, A.E. (1964). Acetylation And Methylation Of Histones And Their Possible Role In The Regulation Of Rna Synthesis. Proc Natl Acad Sci U S A *51*, 786-794.
Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J Mol Biol *215*, 403-410.
Alwine, J.C., Kemp, D.J., and Stark, G.R. (1977). Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. Proc Natl Acad Sci U S A *74*, 5350-5354.
Anderson, K.P., Crable, S.C., and Lingrel, J.B. (1998). Multiple proteins binding to a GATA-E box-GATA motif regulate the erythroid Kruppel-like factor (EKLF) gene. J Biol Chem *273*, 14347-14354.
Anderson, K.P., Crable, S.C., and Lingrel, J.B. (2000). The GATA-E box-GATA motif in the EKLF promoter is required for in vivo expression. Blood *95*, 1652-1655.
Anguita, E., Hughes, J., Heyworth, C., Blobel, G.A., Wood, W.G., and Higgs, D.R. (2004). Globin gene activation during haemopoiesis is driven by protein complexes nucleated by GATA-1 and GATA-2. Embo J *23*, 2841-2852.
Antes, T.J., Namciu, S.J., Fournier, R.E., and Levy-Wilson, B. (2001). The 5' boundary of the human apolipoprotein B chromatin domain in intestinal cells. Biochemistry *40*, 6731-6742.
Aplan, P.D., Begley, C.G., Bertness, V., Nussmeier, M., Ezquerra, A., Coligan, J., and Kirsch, I.R. (1990a). The SCL gene is formed from a transcriptionally complex locus. Mol Cell Biol *10*, 6426-6435.
Aplan, P.D., Lombardi, D.P., Ginsberg, A.M., Cossman, J., Bertness, V.L., and Kirsch, I.R. (1990b). Disruption of the human SCL locus by "illegitimate" V-(D)-J recombinase activity. Science *250*, 1426-1429.

Aza-Blanc, P., Cooper, C.L., Wagner, K., Batalov, S., Deveraux, Q.L., and Cooke, M.P. (2003). Identification of modulators of TRAIL-induced apoptosis via RNAi-based phenotypic screening. Mol Cell *12*, 627-637.

Baeuerle, P.A., and Baltimore, D. (1988). I kappa B: a specific inhibitor of the NF-kappa B transcription factor. Science *242*, 540-546.

Bailey, T.L., and Elkan, C. (1995). The value of prior knowledge in discovering motifs with MEME. Proceedings / International Conference on Intelligent Systems for Molecular Biology ; ISMB *3*, 21-29.

Bain, G., Maandag, E.C., Izon, D.J., Amsen, D., Kruisbeek, A.M., Weintraub, B.C., Krop, I., Schlissel, M.S., Feeney, A.J., van Roon, M.*, et al.* (1994). E2A proteins are required for proper B cell development and initiation of immunoglobulin gene rearrangements. Cell *79*, 885-892.

Bajic, V.B., Tan, S.L., Suzuki, Y., and Sugano, S. (2004). Promoter prediction analysis on the whole human genome. Nature biotechnology *22*, 1467-1473.

Balduini, C.L., Pecci, A., Loffredo, G., Izzo, P., Noris, P., Grosso, M., Bergamaschi, G., Rosti, V., Magrini, U., Ceresa, I.F.*, et al.* (2004). Effects of the R216Q mutation of GATA-1 on erythropoiesis and megakaryocytopoiesis. Thrombosis and haemostasis *91*, 129-140.

Balhoff, J.P., and Wray, G.A. (2005). Evolutionary analysis of the well characterized endo16 promoter reveals substantial variation within functional sites. Proceedings of the National Academy of Sciences of the United States of America *102*, 8591-8596.

Banerji, J., Olson, L., and Schaffner, W. (1983). A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. Cell *33*, 729-740.

Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. Cell *27*, 299-308.

Barndt, R.J., Dai, M., and Zhuang, Y. (2000). Functions of E2A-HEB heterodimers in T-cell development revealed by a dominant negative mutation of HEB. Mol Cell Biol *20*, 6677-6685.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Barski, A., and Frenkel, B. (2004). ChIP Display: novel method for identification of genomic targets of transcription factors. Nucleic Acids Res *32*, e104.

Barton, G.M., and Medzhitov, R. (2002). Retroviral delivery of small interfering RNA into primary cells. Proceedings of the National Academy of Sciences of the United States of America *99*, 14943-14945.

Basu, P., Morris, P.E., Haar, J.L., Wani, M.A., Lingrel, J.B., Gaensler, K.M., and Lloyd, J.A. (2005). KLF2 is essential for primitive erythropoiesis and regulates the human and murine embryonic beta-like globin genes in vivo. Blood *106*, 2566-2571.

Begley, C.G., Aplan, P.D., Davey, M.P., Nakahara, K., Tchorz, K., Kurtzberg, J., Hershfield, M.S., Haynes, B.F., Cohen, D.I., Waldmann, T.A., *et al.* (1989a). Chromosomal translocation in a human leukemic stem-cell line disrupts the T-cell antigen receptor delta-chain diversity region and results in a previously unreported fusion transcript. Proc Natl Acad Sci U S A *86*, 2031-2035.

Begley, C.G., Aplan, P.D., Denning, S.M., Haynes, B.F., Waldmann, T.A., and Kirsch, I.R. (1989b). The gene SCL is expressed during early hematopoiesis and encodes a differentiation-related DNA-binding motif. Proc Natl Acad Sci U S A *86*, 10128-10132.

Begley, C.G., Visvader, J., Green, A.R., Aplan, P.D., Metcalf, D., Kirsch, I.R., and Gough, N.M. (1991). Molecular cloning and chromosomal localization of the murine homolog of the human helix-loop-helix gene SCL. Proc Natl Acad Sci U S A *88*, 869-873.

Bell, A.C., West, A.G., and Felsenfeld, G. (1999). The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. Cell *98*, 387-396.

Bennett, S. (2004). Solexa Ltd. Pharmacogenomics *5*, 433-438.

Bernard, O., Azogui, O., Lecointe, N., Mugneret, F., Berger, R., Larsen, C.J., and Mathieu-Mahul, D. (1992). A third tal-1 promoter is specifically used in human T cell leukemias. J Exp Med *176*, 919-925.

Bernard, O., Lecointe, N., Jonveaux, P., Souyri, M., Mauchauffe, M., Berger, R., Larsen, C.J., and Mathieu-Mahul, D. (1991). Two site-specific deletions and t(1;14) translocation restricted to human T-cell acute leukemias disrupt the 5' part of the tal-1 gene. Oncogene *6*, 1477-1488.

Berns, K., Hijmans, E.M., Mullenders, J., Brummelkamp, T.R., Velds, A., Heimerikx, M., Kerkhoven, R.M., Madiredjo, M., Nijkamp, W., Weigelt, B., *et al.* (2004). A large-scale RNAi screen in human cells identifies new components of the p53 pathway. Nature *428*, 431-437.

Bernstein, B.E., Humphrey, E.L., Erlich, R.L., Schneider, R., Bouman, P., Liu, J.S., Kouzarides, T., and Schreiber, S.L. (2002). Methylation of histone H3 Lys 4 in coding regions of active genes. Proc Natl Acad Sci U S A *99*, 8695-8700.

Bernstein, B.E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D.K., Huebert, D.J., McMahon, S., Karlsson, E.K., Kulbokas, E.J., 3rd, Gingeras, T.R., *et al.* (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. Cell *120*, 169-181.

Birmingham, A., Anderson, E.M., Reynolds, A., Ilsley-Tyree, D., Leake, D., Fedorov, Y., Baskerville, S., Maksimova, E., Robinson, K., Karpilow, J., *et al.* (2006). 3' UTR seed matches, but not overall identity, are associated with RNAi off-targets. Nature methods *3*, 199-204.

Blais, A., Tsikitis, M., Acosta-Alvear, D., Sharan, R., Kluger, Y., and Dynlacht, B.D. (2005). An initial blueprint for myogenic differentiation. Genes Dev *19*, 553-569.

Blanchette, M., and Tompa, M. (2003). FootPrinter: A program designed for phylogenetic footprinting. Nucleic Acids Res *31*, 3840-3842.

Bockamp, E.O., Fordham, J.L., Gottgens, B., Murrell, A.M., Sanchez, M.J., and Green, A.R. (1998). Transcriptional regulation of the stem cell leukemia gene by PU.1 and Elf-1. J Biol Chem *273*, 29032-29042.

Bockamp, E.O., McLaughlin, F., Gottgens, B., Murrell, A.M., Elefanty, A.G., and Green, A.R. (1997). Distinct mechanisms direct SCL/tal-1 expression in erythroid cells and CD34 positive primitive myeloid cells. J Biol Chem *272*, 8781-8790.

Bode, J., Benham, C., Knopp, A., and Mielke, C. (2000). Transcriptional augmentation: modulation of gene expression by scaffold/matrix-attached regions (S/MAR elements). Critical reviews in eukaryotic gene expression *10*, 73-90.

Bode, J., Goetze, S., Heng, H., Krawetz, S.A., and Benham, C. (2003). From DNA structure to gene expression: mediators of nuclear compartmentalization and dynamics. Chromosome Res *11*, 435-445.

Boden, D., Pusch, O., Lee, F., Tucker, L., and Ramratnam, B. (2004a). Efficient gene transfer of HIV-1-specific short hairpin RNA into human lymphocytic cells using recombinant adeno-associated virus vectors. Mol Ther *9*, 396-402.

Boden, D., Pusch, O., Silbermann, R., Lee, F., Tucker, L., and Ramratnam, B. (2004b). Enhanced gene silencing of HIV-1 specific siRNA using microRNA designed hairpins. Nucleic acids research *32*, 1154-1158.

Boehm, T., Foroni, L., Kaneko, Y., Perutz, M.F., and Rabbitts, T.H. (1991). The rhombotin family of cysteine-rich LIM-domain oncogenes: distinct members are involved in T-cell translocations to human chromosomes 11p15 and 11p13. Proceedings of the National Academy of Sciences of the United States of America *88*, 4367-4371.

Bose, F., Fugazza, C., Casalgrandi, M., Capelli, A., Cunningham, J.M., Zhao, Q., Jane, S.M., Ottolenghi, S., and Ronchi, A. (2006). Functional interaction of CP2 with GATA-1 in the regulation of erythroid promoters. Mol Cell Biol *26*, 3942-3954.

Boutros, M., Kiger, A.A., Armknecht, S., Kerr, K., Hild, M., Koch, B., Haas, S.A., Paro, R., and Perrimon, N. (2004). Genome-wide RNAi analysis of growth and viability in Drosophila cells. Science (New York, NY *303*, 832-835.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., *et al.* (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947-956.

Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K., *et al.* (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature *441*, 349-353.

Boyes, J., Byfield, P., Nakatani, Y., and Ogryzko, V. (1998). Regulation of activity of the transcription factor GATA-1 by acetylation. Nature *396*, 594-598.

Brandenberger, R., Khrebtukova, I., Thies, R.S., Miura, T., Jingli, C., Puri, R., Vasicek, T., Lebkowski, J., and Rao, M. (2004). MPSS profiling of human embryonic stem cells. BMC developmental biology *4*, 10.

Brass, A.L., Dykxhoorn, D.M., Benita, Y., Yan, N., Engelman, A., Xavier, R.J., Lieberman, J., and Elledge, S.J. (2008). Identification of host proteins required for HIV infection through a functional genomic screen. Science (New York, NY *319*, 921-926.

Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D.H., Johnson, D., Luo, S., McCurdy, S., Foy, M., Ewan, M., *et al.* (2000a). Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat Biotechnol *18*, 630-634.

Brenner, S., Williams, S.R., Vermaas, E.H., Storck, T., Moon, K., McCollum, C., Mao, J.I., Luo, S., Kirchner, J.J., Eletr, S., *et al.* (2000b). In vitro cloning of complex mixtures of DNA on microbeads: physical separation of differentially expressed cDNAs. Proc Natl Acad Sci U S A *97*, 1665-1670.

Brent, R., and Ptashne, M. (1985). A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. Cell *43*, 729-736.

Bridge, A.J., Pebernard, S., Ducraux, A., Nicoulaz, A.L., and Iggo, R. (2003). Induction of an interferon response by RNAi vectors in mammalian cells. Nature genetics *34*, 263-264.

Brudno, M., Do, C.B., Cooper, G.M., Kim, M.F., Davydov, E., Green, E.D., Sidow, A., and Batzoglou, S. (2003). LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. Genome Res *13*, 721-731.

Brummelkamp, T.R., Bernards, R., and Agami, R. (2002). A system for stable expression of short interfering RNAs in mammalian cells. Science (New York, NY *296*, 550-553.

Brummelkamp, T.R., Nijman, S.M., Dirac, A.M., and Bernards, R. (2003). Loss of the cylindromatosis tumour suppressor inhibits apoptosis by activating NF-kappaB. Nature *424*, 797-801.

Cai, J., Shin, S., Wright, L., Liu, Y., Zhou, D., Xue, H., Khrebtukova, I., Mattson, M.P., Svendsen, C.N., and Rao, M.S. (2006). Massively parallel signature sequencing profiling of fetal human neural precursor cells. Stem cells and development *15*, 232-244.

Cao, H.B., Wang, A., Martin, B., Koehler, D.R., Zeitlin, P.L., Tanawell, A.K., and Hu, J. (2005). Down-regulation of IL-8 expression in human airway epithelial cells through helper-dependent adenoviral-mediated RNA interference. Cell research *15*, 111-119.

Caplen, N.J., Parrish, S., Imani, F., Fire, A., and Morgan, R.A. (2001). Specific inhibition of gene expression by small double-stranded RNAs in invertebrate and vertebrate systems. Proceedings of the National Academy of Sciences of the United States of America *98*, 9742-9747.
Cawley, S., Bekiranov, S., Ng, H.H., Kapranov, P., Sekinger, E.A., Kampa, D., Piccolboni, A., Sementchenko, V., Cheng, J., Williams, A.J., *et al.* (2004). Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. Cell *116*, 499-509.
Chambeyron, S., and Bickmore, W.A. (2004). Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. Genes Dev *18*, 1119-1130.
Chan, W.Y., Follows, G.A., Lacaud, G., Pimanda, J.E., Landry, J.R., Kinston, S., Knezevic, K., Piltz, S., Donaldson, I.J., Gambardella, L., *et al.* (2007). The paralogous hematopoietic regulators Lyl1 and Scl are coregulated by Ets and GATA factors, but Lyl1 cannot rescue the early Scl-/- phenotype. Blood *109*, 1908-1916.
Chen, C.C., Ko, T.M., Ma, H.I., Wu, H.L., Xiao, X., Li, J., Chang, C.M., Wu, P.Y., Chen, C.H., Han, J.M., *et al.* (2007). Long-term inhibition of hepatitis B virus in transgenic mice by double-stranded adeno-associated virus 8-delivered short hairpin RNA. Gene therapy *14*, 11-19.
Chen, J.D., and Evans, R.M. (1995). A transcriptional co-repressor that interacts with nuclear hormone receptors. Nature *377*, 454-457.
Chen, L., and Widom, J. (2005). Mechanism of transcriptional silencing in yeast. Cell *120*, 37-48.
Chen, Y.T., Scanlan, M.J., Venditti, C.A., Chua, R., Theiler, G., Stevenson, B.J., Iseli, C., Gure, A.O., Vasicek, T., Strausberg, R.L., *et al.* (2005). Identification of cancer/testis-antigen genes by massively parallel signature sequencing. Proc Natl Acad Sci U S A *102*, 7940-7945.
Choi, K., Kennedy, M., Kazarov, A., Papadimitriou, J.C., and Keller, G. (1998). A common precursor for hematopoietic and endothelial cells. Development *125*, 725-732.
Chung, Y.S., Zhang, W.J., Arentson, E., Kingsley, P.D., Palis, J., and Choi, K. (2002). Lineage analysis of the hemangioblast as defined by FLK1 and SCL expression. Development *129*, 5511-5520.
Clemens, J.C., Worby, C.A., Simonson-Leff, N., Muda, M., Maehama, T., Hemmings, B.A., and Dixon, J.E. (2000). Use of double-stranded RNA interference in Drosophila cell lines to dissect signal transduction pathways. Proceedings of the National Academy of Sciences of the United States of America *97*, 6499-6503.
Cohen, B.A., Mitra, R.D., Hughes, J.D., and Church, G.M. (2000). A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. Nat Genet *26*, 183-186.

Coller, H.A., Grandori, C., Tamayo, P., Colbert, T., Lander, E.S., Eisenman, R.N., and Golub, T.R. (2000). Expression analysis with oligonucleotide microarrays reveals that MYC regulates genes involved in growth, cell cycle, signaling, and adhesion. Proc Natl Acad Sci U S A *97*, 3260-3265.
Courey, A.J., and Tjian, R. (1988). Analysis of Sp1 in vivo reveals multiple transcriptional domains, including a novel glutamine-rich activation motif. Cell *55*, 887-898.
Crossley, M., Tsang, A.P., Bieker, J.J., and Orkin, S.H. (1994). Regulation of the erythroid Kruppel-like factor (EKLF) gene promoter by the erythroid transcription factor GATA-1. J Biol Chem *269*, 15440-15444.
Curtis, D.J., Hall, M.A., Van Stekelenburg, L.J., Robb, L., Jane, S.M., and Begley, C.G. (2004). SCL is required for normal function of short-term repopulating hematopoietic stem cells. Blood *103*, 3342-3348.
Cuthbert, G.L., Daujat, S., Snowden, A.W., Erdjument-Bromage, H., Hagiwara, T., Yamada, M., Schneider, R., Gregory, P.D., Tempst, P., Bannister, A.J., *et al.* (2004). Histone deimination antagonizes arginine methylation. Cell *118*, 545-553.
D'Souza, S.L., Elefanty, A.G., and Keller, G. (2005). SCL/Tal-1 is essential for hematopoietic commitment of the hemangioblast but not for its development. Blood *105*, 3862-3870.
Dahmus, M.E. (1995). Phosphorylation of the C-terminal domain of RNA polymerase II. Biochimica et biophysica acta *1261*, 171-182.
Daniel, J.A., Torok, M.S., Sun, Z.W., Schieltz, D., Allis, C.D., Yates, J.R., 3rd, and Grant, P.A. (2004). Deubiquitination of histone H2B by a yeast acetyltransferase complex regulates transcription. J Biol Chem *279*, 1867-1871.
Das, P.M., Ramachandran, K., vanWert, J., and Singal, R. (2004). Chromatin immunoprecipitation assay. BioTechniques *37*, 961-969.
Davuluri, R.V., Grosse, I., and Zhang, M.Q. (2001). Computational identification of promoters and first exons in the human genome. Nat Genet *29*, 412-417.
De Gobbi, M., Anguita, E., Hughes, J., Sloane-Stanley, J.A., Sharpe, J.A., Koch, C.M., Dunham, I., Gibbons, R.J., Wood, W.G., and Higgs, D.R. (2007). Tissue-specific histone modification and transcription factor binding in alpha globin gene expression. Blood *110*, 4503-4510.
de Jonge, J., Holtrop, M., Wilschut, J., and Huckriede, A. (2006). Reconstituted influenza virus envelopes as an efficient carrier system for cellular delivery of small-interfering RNAs. Gene therapy *13*, 400-411.
Delabesse, E., Ogilvy, S., Chapman, M.A., Piltz, S.G., Gottgens, B., and Green, A.R. (2005). Transcriptional regulation of the SCL locus: identification of an enhancer that targets the primitive erythroid lineage in vivo. Mol Cell Biol *25*, 5215-5225.

Delassus, S., Titley, I., and Enver, T. (1999). Functional and molecular analysis of hematopoietic progenitors derived from the aorta-gonad-mesonephros region of the mouse embryo. Blood *94*, 1495-1503.
DeRisi, J.L., Iyer, V.R., and Brown, P.O. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. Science (New York, NY *278*, 680-686.
Dermitzakis, E.T., and Clark, A.G. (2002). Evolution of transcription factor binding sites in Mammalian gene regulatory regions: conservation and turnover. Molecular biology and evolution *19*, 1114-1121.
Diakos, C., Krapf, G., Gerner, C., Inthal, A., Lemberger, C., Ban, J., Dohnal, A.M., and Panzer-Gruemayer, E.R. (2007). RNAi-mediated silencing of TEL/AML1 reveals a heat-shock protein- and survivin-dependent mechanism for survival. Blood *109*, 2607-2610.
Dickins, R.A., Hemann, M.T., Zilfou, J.T., Simpson, D.R., Ibarra, I., Hannon, G.J., and Lowe, S.W. (2005). Probing tumor phenotypes using stable and regulated synthetic microRNA precursors. Nature genetics *37*, 1289-1295.
Down, T.A., and Hubbard, T.J. (2002). Computational detection and location of transcription start sites in mammalian genomic DNA. Genome Res *12*, 458-461.
Down, T.A., and Hubbard, T.J. (2005). NestedMICA: sensitive inference of over-represented motifs in nucleic acid sequence. Nucleic Acids Res *33*, 1445-1453.
Eisbacher, M., Holmes, M.L., Newton, A., Hogg, P.J., Khachigian, L.M., Crossley, M., and Chong, B.H. (2003). Protein-protein interaction between Fli-1 and GATA-1 mediates synergistic expression of megakaryocyte-specific genes through cooperative DNA binding. Mol Cell Biol *23*, 3427-3441.
Elbashir, S.M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., and Tuschl, T. (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature *411*, 494-498.
Elefanty, A.G., Begley, C.G., Hartley, L., Papaevangeliou, B., and Robb, L. (1999). SCL expression in the mouse embryo detected with a targeted lacZ reporter gene demonstrates its localization to hematopoietic, vascular, and neural tissues. Blood *94*, 3754-3763.
Elefanty, A.G., Robb, L., Birner, R., and Begley, C.G. (1997). Hematopoietic-specific genes are not induced during in vitro differentiation of scl-null embryonic stem cells. Blood *90*, 1435-1447.
Elkon, R., Rashi-Elkeles, S., Lerenthal, Y., Linhart, C., Tenne, T., Amariglio, N., Rechavi, G., Shamir, R., and Shiloh, Y. (2005). Dissection of a DNA-damage-induced transcriptional network using a combination of microarrays, RNA interference and computational promoter analysis. Genome biology *6*, R43.
Elwood, N.J., Green, A.R., Melder, A., Begley, C.G., and Nicola, N. (1994). The SCL protein displays cell-specific heterogeneity in size. Leukemia *8*, 106-114.

Endoh, M., Ogawa, M., Orkin, S., and Nishikawa, S. (2002). SCL/tal-1-dependent process determines a competence to select the definitive hematopoietic lineage prior to endothelial differentiation. Embo J *21*, 6700-6708.
Euskirchen, G., Royce, T.E., Bertone, P., Martone, R., Rinn, J.L., Nelson, F.K., Sayward, F., Luscombe, N.M., Miller, P., Gerstein, M., *et al.* (2004). CREB binds to multiple loci on human chromosome 22. Mol Cell Biol *24*, 3804-3814.
Evans, T., Reitman, M., and Felsenfeld, G. (1988). An erythrocyte-specific DNA-binding factor recognizes a regulatory sequence common to all chicken globin genes. Proc Natl Acad Sci U S A *85*, 5976-5980.
Fedorov, Y., Anderson, E.M., Birmingham, A., Reynolds, A., Karpilow, J., Robinson, K., Leake, D., Marshall, W.S., and Khvorova, A. (2006). Off-target effects by siRNA can induce toxic phenotype. RNA (New York, NY *12*, 1188-1196.
Ferrando, A.A., Neuberg, D.S., Staunton, J., Loh, M.L., Huard, C., Raimondi, S.C., Behm, F.G., Pui, C.H., Downing, J.R., Gilliland, D.G., *et al.* (2002). Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. Cancer cell *1*, 75-87.
Ferreira, R., Ohneda, K., Yamamoto, M., and Philipsen, S. (2005). GATA1 function, a paradigm for transcription factors in hematopoiesis. Mol Cell Biol *25*, 1215-1227.
Fire, A., Albertson, D., Harrison, S.W., and Moerman, D.G. (1991). Production of antisense RNA leads to effective and specific inhibition of gene expression in C. elegans muscle. Development *113*, 503-514.
Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature *391*, 806-811.
Fordham, J.L., Gottgens, B., McLaughlin, F., and Green, A.R. (1999). Chromatin structure and transcriptional regulation of the stem cell leukaemia (SCL) gene in mast cells. Leukemia *13*, 750-759.
Fox, A.H., Liew, C., Holmes, M., Kowalski, K., Mackay, J., and Crossley, M. (1999). Transcriptional cofactors of the FOG family interact with GATA proteins by means of multiple zinc fingers. Embo J *18*, 2812-2822.
Freeman, L.A., and Garrard, W.T. (1992). DNA supercoiling in chromatin structure and gene expression. Critical reviews in eukaryotic gene expression *2*, 165-209.
Freson, K., Devriendt, K., Matthijs, G., Van Hoof, A., De Vos, R., Thys, C., Minner, K., Hoylaerts, M.F., Vermylen, J., and Van Geet, C. (2001). Platelet characteristics in patients with X-linked macrothrombocytopenia because of a novel GATA1 mutation. Blood *98*, 85-92.

Freson, K., Matthijs, G., Thys, C., Marien, P., Hoylaerts, M.F., Vermylen, J., and Van Geet, C. (2002). Different substitutions at residue D218 of the X-linked transcription factor GATA1 lead to altered clinical severity of macrothrombocytopenia and anemia and are associated with variable skewed X inactivation. Human molecular genetics *11*, 147-152.

Fujiwara, Y., Browne, C.P., Cunniff, K., Goff, S.C., and Orkin, S.H. (1996). Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. Proc Natl Acad Sci U S A *93*, 12355-12358.

Galas, D.J., and Schmitz, A. (1978). DNAse footprinting: a simple method for the detection of protein-DNA binding specificity. Nucleic Acids Res *5*, 3157-3170.

Garner, M.M., and Revzin, A. (1981). A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system. Nucleic Acids Res *9*, 3047-3060.

Ge, Y., Dombkowski, A.A., LaFiura, K.M., Tatman, D., Yedidi, R.S., Stout, M.L., Buck, S.A., Massey, G., Becton, D.L., Weinstein, H.J.*, et al.* (2006). Differential gene expression, GATA1 target genes, and the chemotherapy sensitivity of Down syndrome megakaryocytic leukemia. Blood *107*, 1570-1581.

Gershenzon, N.I., and Ioshikhes, I.P. (2005). Synergy of human Pol II core promoter elements revealed by statistical sequence analysis. Bioinformatics (Oxford, England) *21*, 1295-1300.

Gerton, J.L., DeRisi, J., Shroff, R., Lichten, M., Brown, P.O., and Petes, T.D. (2000). Inaugural article: global mapping of meiotic recombination hotspots and coldspots in the yeast Saccharomyces cerevisiae. Proc Natl Acad Sci U S A *97*, 11383-11390.

Giroux, S., Kaushik, A.L., Capron, C., Jalil, A., Kelaidi, C., Sablitzky, F., Dumenil, D., Albagli, O., and Godin, I. (2007). lyl-1 and tal-1/scl, two genes encoding closely related bHLH transcription factors, display highly overlapping expression patterns during cardiovascular and hematopoietic ontogeny. Gene Expr Patterns *7*, 215-226.

Goardon, N., Lambert, J.A., Rodriguez, P., Nissaire, P., Herblot, S., Thibault, P., Dumenil, D., Strouboulis, J., Romeo, P.H., and Hoang, T. (2006). ETO2 coordinates cellular proliferation and differentiation during erythropoiesis. Embo J *25*, 357-366.

Godin, I., and Cumano, A. (2002). The hare and the tortoise: an embryonic haematopoietic race. Nat Rev Immunol *2*, 593-604.

Godin, I., Garcia-Porrero, J.A., Dieterlen-Lievre, F., and Cumano, A. (1999). Stem cell emergence and hemopoietic activity are incompatible in mouse intraembryonic sites. The Journal of experimental medicine *190*, 43-52.

Goldfarb, A.N., Goueli, S., Mickelson, D., and Greenberg, J.M. (1992). T-cell acute lymphoblastic leukemia--the associated gene SCL/tal codes for a 42-Kd nuclear phosphoprotein. Blood *80*, 2858-2866.
Gonczy, P., Echeverri, C., Oegema, K., Coulson, A., Jones, S.J., Copley, R.R., Duperon, J., Oegema, J., Brehm, M., Cassin, E.*, et al.* (2000). Functional genomic analysis of cell division in C. elegans using RNAi of genes on chromosome III. Nature *408*, 331-336.
Gonzalez, G.A., and Montminy, M.R. (1989). Cyclic AMP stimulates somatostatin gene transcription by phosphorylation of CREB at serine 133. Cell *59*, 675-680.
Gottgens, B., Barton, L.M., Gilbert, J.G., Bench, A.J., Sanchez, M.J., Bahn, S., Mistry, S., Grafham, D., McMurray, A., Vaudin, M.*, et al.* (2000). Analysis of vertebrate SCL loci identifies conserved enhancers. Nature biotechnology *18*, 181-186.
Gottgens, B., Broccardo, C., Sanchez, M.J., Deveaux, S., Murphy, G., Gothert, J.R., Kotsopoulou, E., Kinston, S., Delaney, L., Piltz, S.*, et al.* (2004). The scl +18/19 stem cell enhancer is not required for hematopoiesis: identification of a 5' bifunctional hematopoietic-endothelial enhancer bound by Fli-1 and Elf-1. Mol Cell Biol *24*, 1870-1883.
Gottgens, B., McLaughlin, F., Bockamp, E.O., Fordham, J.L., Begley, C.G., Kosmopoulos, K., Elefanty, A.G., and Green, A.R. (1997). Transcription of the SCL gene in erythroid and CD34 positive primitive myeloid cells is controlled by a complex network of lineage-restricted chromatin-dependent and chromatin-independent regulatory elements. Oncogene *15*, 2419-2428.
Gottgens, B., Nastos, A., Kinston, S., Piltz, S., Delabesse, E.C., Stanley, M., Sanchez, M.J., Ciau-Uitz, A., Patient, R., and Green, A.R. (2002). Establishing the transcriptional programme for blood: the SCL stem cell enhancer is regulated by a multiprotein complex containing Ets and GATA factors. Embo J *21*, 3039-3050.
Gou, D., Jin, N., and Liu, L. (2003). Gene silencing in mammalian cells by PCR-based short hairpin RNA. FEBS letters *548*, 113-118.
Grant, P.A., Eberharter, A., John, S., Cook, R.G., Turner, B.M., and Workman, J.L. (1999). Expanded lysine acetylation specificity of Gcn5 in native complexes. J Biol Chem *274*, 5895-5900.
Grass, J.A., Boyer, M.E., Pal, S., Wu, J., Weiss, M.J., and Bresnick, E.H. (2003). GATA-1-dependent transcriptional repression of GATA-2 via disruption of positive autoregulation and domain-wide chromatin remodeling. Proc Natl Acad Sci U S A *100*, 8811-8816.
Green, A.R., Lints, T., Visvader, J., Harvey, R., and Begley, C.G. (1992). SCL is coexpressed with GATA-1 in hemopoietic cells but is also expressed in developing brain. Oncogene *7*, 653-660.
Green, A.R., Rockman, S., DeLuca, E., and Begley, C.G. (1993). Induced myeloid differentiation of K562 cells with downregulation of erythroid and megakaryocytic transcription factors: a novel experimental model for hemopoietic lineage restriction. Exp Hematol *21*, 525-531.

Green, A.R., Salvaris, E., and Begley, C.G. (1991). Erythroid expression of the 'helix-loop-helix' gene, SCL. Oncogene *6*, 475-479.
Greenbaum, S., and Zhuang, Y. (2002). Identification of E2A target genes in B lymphocyte development by using a gene tagging-based chromatin immunoprecipitation system. Proceedings of the National Academy of Sciences of the United States of America *99*, 15030-15035.
Gregory, R.C., Taxman, D.J., Seshasayee, D., Kensinger, M.H., Bieker, J.J., and Wojchowski, D.M. (1996). Functional interaction of GATA1 with erythroid Kruppel-like factor and Sp1 at defined erythroid promoters. Blood *87*, 1793-1801.
Grimm, D., Streetz, K.L., Jopling, C.L., Storm, T.A., Pandey, K., Davis, C.R., Marion, P., Salazar, F., and Kay, M.A. (2006). Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways. Nature *441*, 537-541.
Gross, D.S., and Garrard, W.T. (1988). Nuclease hypersensitive sites in chromatin. Annu Rev Biochem *57*, 159-197.
Grosveld, F., van Assendelft, G.B., Greaves, D.R., and Kollias, G. (1987). Position-independent, high-level expression of the human beta-globin gene in transgenic mice. Cell *51*, 975-985.
Grutz, G.G., Bucher, K., Lavenir, I., Larson, T., Larson, R., and Rabbitts, T.H. (1998). The oncogenic T cell LIM-protein Lmo2 forms part of a DNA-binding complex specifically in immature T cells. The EMBO journal *17*, 4594-4605.
Gu, W., and Roeder, R.G. (1997). Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain. Cell *90*, 595-606.
Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. Cell *130*, 77-88.
Gupta, S., Schoer, R.A., Egan, J.E., Hannon, G.J., and Mittal, V. (2004). Inducible, reversible, and stable RNA interference in mammalian cells. Proceedings of the National Academy of Sciences of the United States of America *101*, 1927-1932.
Gyorffy, B., Surowiak, P., Kiesslich, O., Denkert, C., Schafer, R., Dietel, M., and Lage, H. (2005). Gene expression profiling of 30 cancer cell lines predicts resistance towards 11 anticancer drugs at clinically achieved concentrations. Int J Cancer.
Hahn, S. (2004). Structure and mechanism of the RNA polymerase II transcription machinery. Nat Struct Mol Biol *11*, 394-403.
Hall, M.A., Curtis, D.J., Metcalf, D., Elefanty, A.G., Sourris, K., Robb, L., Gothert, J.R., Jane, S.M., and Begley, C.G. (2003). The critical regulator of embryonic hematopoiesis, SCL, is vital in the adult for megakaryopoiesis, erythropoiesis, and lineage choice in CFU-S12. Proc Natl Acad Sci U S A *100*, 992-997.

Hall, M.A., Slater, N.J., Begley, C.G., Salmon, J.M., Van Stekelenburg, L.J., McCormack, M.P., Jane, S.M., and Curtis, D.J. (2005). Functional but abnormal adult erythropoiesis in the absence of the stem cell leukemia gene. Mol Cell Biol *25*, 6355-6362.

Harris, M.B., Mostecki, J., and Rothman, P.B. (2005). Repression of an interleukin-4-responsive promoter requires cooperative BCL-6 function. J Biol Chem *280*, 13114-13121.

Hassa, P.O., Haenni, S.S., Elser, M., and Hottiger, M.O. (2006). Nuclear ADP-ribosylation reactions in mammalian cells: where are we today and where are we going? Microbiol Mol Biol Rev *70*, 789-829.

Hassan, A.H., Neely, K.E., and Workman, J.L. (2001). Histone acetyltransferase complexes stabilize swi/snf binding to promoter nucleosomes. Cell *104*, 817-827.

Heemskerk, M.H., Blom, B., Nolan, G., Stegmann, A.P., Bakker, A.Q., Weijer, K., Res, P.C., and Spits, H. (1997). Inhibition of T cell and promotion of natural killer cell development by the dominant negative helix loop helix factor Id3. The Journal of experimental medicine *186*, 1597-1602.

Heinemeyer, T., Wingender, E., Reuter, I., Hermjakob, H., Kel, A.E., Kel, O.V., Ignatieva, E.V., Ananko, E.A., Podkolodnaya, O.A., Kolpakov, F.A., *et al.* (1998). Databases on transcriptional regulation: TRANSFAC, TRRD and COMPEL. Nucleic Acids Res *26*, 362-367.

Heinzel, T., Lavinsky, R.M., Mullen, T.M., Soderstrom, M., Laherty, C.D., Torchia, J., Yang, W.M., Brard, G., Ngo, S.D., Davie, J.R., *et al.* (1997). A complex containing N-CoR, mSin3 and histone deacetylase mediates transcriptional repression. Nature *387*, 43-48.

Hirasawa, R., Shimizu, R., Takahashi, S., Osawa, M., Takayanagi, S., Kato, Y., Onodera, M., Minegishi, N., Yamamoto, M., Fukao, K., *et al.* (2002). Essential and instructive roles of GATA factors in eosinophil development. J Exp Med *195*, 1379-1386.

Holmes, M., Turner, J., Fox, A., Chisholm, O., Crossley, M., and Chong, B. (1999). hFOG-2, a novel zinc finger protein, binds the co-repressor mCtBP2 and modulates GATA-mediated activation. J Biol Chem *274*, 23491-23498.

Hope, I.A., Mahadevan, S., and Struhl, K. (1988). Structural and functional characterization of the short acidic transcriptional activation region of yeast GCN4 protein. Nature *333*, 635-640.

Hoppe, K.L., and Francone, O.L. (1998). Binding and functional effects of transcription factors Sp1 and Sp3 on the proximal human lecithin:cholesterol acyltransferase promoter. Journal of lipid research *39*, 969-977.

Horak, C.E., Mahajan, M.C., Luscombe, N.M., Gerstein, M., Weissman, S.M., and Snyder, M. (2002). GATA-1 binding sites mapped in the beta-globin locus by using mammalian chIp-chip analysis. Proc Natl Acad Sci U S A *99*, 2924-2929.

Horner, M.A., Quintin, S., Domeier, M.E., Kimble, J., Labouesse, M., and Mango, S.E. (1998). pha-4, an HNF-3 homolog, specifies pharyngeal organ identity in Caenorhabditis elegans. Genes Dev *12*, 1947-1952.

Hsieh, A.C., Bo, R., Manola, J., Vazquez, F., Bare, O., Khvorova, A., Scaringe, S., and Sellers, W.R. (2004). A library of siRNA duplexes targeting the phosphoinositide 3-kinase pathway: determinants of gene silencing for use in cell-based screens. Nucleic acids research *32*, 893-901.

Hsu, H.L., Cheng, J.T., Chen, Q., and Baer, R. (1991). Enhancer-binding activity of the tal-1 oncoprotein in association with the E47/E12 helix-loop-helix proteins. Mol Cell Biol *11*, 3037-3042.

Hsu, H.L., Wadman, I., and Baer, R. (1994a). Formation of in vivo complexes between the TAL1 and E2A polypeptides of leukemic T cells. Proc Natl Acad Sci U S A *91*, 3181-3185.

Hsu, H.L., Wadman, I., Tsan, J.T., and Baer, R. (1994b). Positive and negative transcriptional control by the TAL1 helix-loop-helix protein. Proc Natl Acad Sci U S A *91*, 5947-5951.

Huang, D.Y., Kuo, Y.Y., and Chang, Z.F. (2005). GATA-1 mediates auto-regulation of Gfi-1B transcription in K562 cells. Nucleic Acids Res *33*, 5331-5342.

Huang, D.Y., Kuo, Y.Y., Lai, J.S., Suzuki, Y., Sugano, S., and Chang, Z.F. (2004). GATA-1 and NF-Y cooperate to mediate erythroid-specific transcription of Gfi-1B gene. Nucleic Acids Res *32*, 3935-3946.

Hughes, T.R., Mao, M., Jones, A.R., Burchard, J., Marton, M.J., Shannon, K.W., Lefkowitz, S.M., Ziman, M., Schelter, J.M., Meyer, M.R., *et al.* (2001). Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. Nat Biotechnol *19*, 342-347.

Hwang, L.Y., Siegelman, M., Davis, L., Oppenheimer-Marks, N., and Baer, R. (1993). Expression of the TAL1 proto-oncogene in cultured endothelial cells and blood vessels of the spleen. Oncogene *8*, 3043-3046.

Ikawa, T., Kawamoto, H., Wright, L.Y., and Murre, C. (2004). Long-term cultured E2A-deficient hematopoietic progenitor cells are pluripotent. Immunity *20*, 349-360.

Imagawa, S., Suzuki, N., Ohmine, K., Obara, N., Mukai, H.Y., Ozawa, K., Yamamoto, M., and Nagasawa, T. (2002). GATA suppresses erythropoietin gene expression through GATA site in mouse erythropoietin gene promoter. International journal of hematology *75*, 376-381.

Imagawa, S., Yamamoto, M., and Miura, Y. (1997). Negative regulation of the erythropoietin gene expression by the GATA transcription factors. Blood *89*, 1430-1439.

Inaba, T., Roberts, W.M., Shapiro, L.H., Jolly, K.W., Raimondi, S.C., Smith, S.D., and Look, A.T. (1992). Fusion of the leucine zipper gene HLF to the E2A gene in human acute B-lineage leukemia. Science *257*, 531-534.

Iwasaki, H., Mizuno, S., Wells, R.A., Cantor, A.B., Watanabe, S., and Akashi, K. (2003). GATA-1 converts lymphoid and myelomonocytic progenitors into the megakaryocyte/erythrocyte lineages. Immunity *19*, 451-462.

Iyer, V.R., Horak, C.E., Scafe, C.S., Botstein, D., Snyder, M., and Brown, P.O. (2001). Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. Nature *409*, 533-538.

Jackson, A.L., Bartz, S.R., Schelter, J., Kobayashi, S.V., Burchard, J., Mao, M., Li, B., Cavet, G., and Linsley, P.S. (2003). Expression profiling reveals off-target gene regulation by RNAi. Nat Biotechnol *21*, 635-637.

Jackson, A.L., Burchard, J., Schelter, J., Chau, B.N., Cleary, M., Lim, L., and Linsley, P.S. (2006). Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity. RNA (New York, NY *12*, 1179-1187.

Jazag, A., Ijichi, H., Kanai, F., Imamura, T., Guleng, B., Ohta, M., Imamura, J., Tanaka, Y., Tateishi, K., Ikenoue, T.*, et al.* (2005). Smad4 silencing in pancreatic cancer cell lines using stable RNA interference and gene expression profiles induced by transforming growth factor-beta. Oncogene *24*, 662-671.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science *293*, 1074-1080.

Jiang, J., Chan, Y.S., Loh, Y.H., Cai, J., Tong, G.Q., Lim, C.A., Robson, P., Zhong, S., and Ng, H.H. (2008). A core Klf circuitry regulates self-renewal of embryonic stem cells. Nature cell biology *10*, 353-360.

Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. Science *316*, 1497-1502.

Johnson, E.S. (2004). Protein modification by SUMO. Annu Rev Biochem *73*, 355-382.

Jongeneel, C.V., Delorenzi, M., Iseli, C., Zhou, D., Haudenschild, C.D., Khrebtukova, I., Kuznetsov, D., Stevenson, B.J., Strausberg, R.L., Simpson, A.J.*, et al.* (2005). An atlas of human gene expression from massively parallel signature sequencing (MPSS). Genome Res *15*, 1007-1014.

Jurata, L.W., and Gill, G.N. (1997). Functional analysis of the nuclear LIM domain interactor NLI. Molecular and cellular biology *17*, 5688-5698.

Kallianpur, A.R., Jordan, J.E., and Brandt, S.J. (1994). The SCL/TAL-1 gene is expressed in progenitors of both the hematopoietic and vascular systems during embryogenesis. Blood *83*, 1200-1208.

Kappel, A., Schlaeger, T.M., Flamme, I., Orkin, S.H., Risau, W., and Breier, G. (2000). Role of SCL/Tal-1, GATA, and ets transcription factor binding sites for the regulation of flk-1 expression during murine vascular development. Blood *96*, 3078-3085.

Kawasaki, H., and Taira, K. (2004). Induction of DNA methylation and gene silencing by short interfering RNAs in human cells. Nature *431*, 211-217.
Kee, B.L., and Murre, C. (1998). Induction of early B cell factor (EBF) and multiple B lineage genes by the basic helix-loop-helix transcription factor E12. The Journal of experimental medicine *188*, 699-713.
Kel, A.E., Gossling, E., Reuter, I., Cheremushkin, E., Kel-Margoulis, O.V., and Wingender, E. (2003). MATCH: A tool for searching transcription factor binding sites in DNA sequences. Nucleic Acids Res *31*, 3576-3579.
Kennedy, M., Firpo, M., Choi, K., Wall, C., Robertson, S., Kabrun, N., and Keller, G. (1997). A common precursor for primitive erythropoiesis and definitive haematopoiesis. Nature *386*, 488-493.
Kennerdell, J.R., and Carthew, R.W. (1998). Use of dsRNA-mediated genetic interference to demonstrate that frizzled and frizzled 2 act in the wingless pathway. Cell *95*, 1017-1026.
Kennerdell, J.R., and Carthew, R.W. (2000). Heritable gene silencing in Drosophila using double-stranded RNA. Nat Biotechnol *18*, 896-898.
Kim, J., Bhinge, A.A., Morgan, X.C., and Iyer, V.R. (2005a). Mapping DNA-protein interactions in large genomes by sequence tag analysis of genomic enrichment. Nature methods *2*, 47-53.
Kim, J., and Iyer, V.R. (2004). Global role of TATA box-binding protein recruitment to promoters in mediating gene expression profiles. Mol Cell Biol *24*, 8104-8112.
Kim, J.B., Porreca, G.J., Song, L., Greenway, S.C., Gorham, J.M., Church, G.M., Seidman, C.E., and Seidman, J.G. (2007). Polony multiplex analysis of gene expression (PMAGE) in mouse hypertrophic cardiomyopathy. Science (New York, NY *316*, 1481-1484.
Kim, T.H., Barrera, L.O., Qu, C., Van Calcar, S., Trinklein, N.D., Cooper, S.J., Luna, R.M., Glass, C.K., Rosenfeld, M.G., Myers, R.M.*, et al.* (2005b). Direct isolation and identification of promoters in the human genome. Genome Res *15*, 830-839.
Kim, T.H., Barrera, L.O., Zheng, M., Qu, C., Singer, M.A., Richmond, T.A., Wu, Y., Green, R.D., and Ren, B. (2005c). A high-resolution map of active promoters in the human genome. Nature *436*, 876-880.
Kittler, R., Pelletier, L., Heninger, A.K., Slabicki, M., Theis, M., Miroslaw, L., Poser, I., Lawo, S., Grabner, H., Kozak, K.*, et al.* (2007). Genome-scale RNAi profiling of cell division in human tissue culture cells. Nature cell biology *9*, 1401-1412.
Kittler, R., Putz, G., Pelletier, L., Poser, I., Heninger, A.K., Drechsel, D., Fischer, S., Konstantinova, I., Habermann, B., Grabner, H.*, et al.* (2004). An endoribonuclease-prepared siRNA screen in human cells identifies genes essential for cell division. Nature *432*, 1036-1040.
Kloosterman, W.P., and Plasterk, R.H. (2006). The diverse functions of microRNAs in animal development and disease. Developmental cell *11*, 441-450.

Kobayashi, M., Nishikawa, K., and Yamamoto, M. (2001). Hematopoietic regulatory domain of gata1 gene is positively regulated by GATA1 protein in zebrafish embryos. Development (Cambridge, England) *128*, 2341-2350.
Koch, C.M., Andrews, R.M., Flicek, P., Dillon, S.C., Karaoz, U., Clelland, G.K., Wilcox, S., Beare, D.M., Fowler, J.C., Couttet, P., *et al.* (2007). The landscape of histone modifications across 1% of the human genome in five human cell lines. Genome Res *17*, 691-707.
Kodzius, R., Kojima, M., Nishiyori, H., Nakamura, M., Fukuda, S., Tagami, M., Sasaki, D., Imamura, K., Kai, C., Harbers, M., *et al.* (2006). CAGE: cap analysis of gene expression. Nature methods *3*, 211-222.
Koizumi, K., Higashida, H., Yoo, S., Islam, M.S., Ivanov, A.I., Guo, V., Pozzi, P., Yu, S.H., Rovescalli, A.C., Tang, D., *et al.* (2007). RNA interference screen to identify genes required for Drosophila embryonic nervous system development. Proceedings of the National Academy of Sciences of the United States of America *104*, 5626-5631.
Kolfschoten, I.G., van Leeuwen, B., Berns, K., Mullenders, J., Beijersbergen, R.L., Bernards, R., Voorhoeve, P.M., and Agami, R. (2005). A genetic screen identifies PITX1 as a suppressor of RAS activity and tumorigenicity. Cell *121*, 849-858.
Krosl, G., He, G., Lefrancois, M., Charron, F., Romeo, P.H., Jolicoeur, P., Kirsch, I.R., Nemer, M., and Hoang, T. (1998). Transcription factor SCL is required for c-kit expression and c-Kit function in hemopoietic cells. J Exp Med *188*, 439-450.
Kuhl, C., Atzberger, A., Iborra, F., Nieswandt, B., Porcher, C., and Vyas, P. (2005). GATA1-mediated megakaryocyte differentiation and growth control can be uncoupled and mapped to different domains in GATA1. Mol Cell Biol *25*, 8592-8606.
Kuhn, K., Baker, S.C., Chudin, E., Lieu, M.H., Oeser, S., Bennett, H., Rigault, P., Barker, D., McDaniel, T.K., and Chee, M.S. (2004). A novel, high-performance random array platform for quantitative gene expression profiling. Genome Res *14*, 2347-2356.
Kulessa, H., Frampton, J., and Graf, T. (1995). GATA-1 reprograms avian myelomonocytic cell lines into eosinophils, thromboblasts, and erythroblasts. Genes Dev *9*, 1250-1262.
Kumaravelu, P., Hook, L., Morrison, A.M., Ure, J., Zhao, S., Zuyev, S., Ansell, J., and Medvinsky, A. (2002). Quantitative developmental anatomy of definitive haematopoietic stem cells/long-term repopulating units (HSC/RUs): role of the aorta-gonad-mesonephros (AGM) region and the yolk sac in colonisation of the mouse embryonic liver. Development *129*, 4891-4899.
Kurdistani, S.K., Tavazoie, S., and Grunstein, M. (2004). Mapping global histone acetylation patterns to gene expression. Cell *117*, 721-733.
Labastie, M.C., Cortes, F., Romeo, P.H., Dulac, C., and Peault, B. (1998). Molecular identity of hematopoietic precursor cells emerging in the human embryo. Blood *92*, 3624-3635.

Lahlil, R., Lecuyer, E., Herblot, S., and Hoang, T. (2004). SCL assembles a multifactorial complex that determines glycophorin A expression. Mol Cell Biol *24*, 1439-1452.
Lamond, A.I., and Earnshaw, W.C. (1998). Structure and function in the nucleus. Science *280*, 547-553.
Landry, J.R., Kinston, S., Knezevic, K., de Bruijn, M.F., Wilson, N., Nottingham, W.T., Peitz, M., Edenhofer, F., Pimanda, J.E., Ottersbach, K.*, et al.* (2008). Runx genes are direct targets of Scl/Tal1 in the yolk sac and fetal liver. Blood *111*, 3005-3014.
Larson, R.C., Lavenir, I., Larson, T.A., Baer, R., Warren, A.J., Wadman, I., Nottage, K., and Rabbitts, T.H. (1996). Protein dimerization between Lmo2 (Rbtn2) and Tal1 alters thymocyte development and potentiates T cell tumorigenesis in transgenic mice. The EMBO journal *15*, 1021-1027.
Lassar, A.B., Davis, R.L., Wright, W.E., Kadesch, T., Murre, C., Voronova, A., Baltimore, D., and Weintraub, H. (1991). Functional activity of myogenic HLH proteins requires hetero-oligomerization with E12/E47-like proteins in vivo. Cell *66*, 305-315.
Lecointe, N., Bernard, O., Naert, K., Joulin, V., Larsen, C.J., Romeo, P.H., and Mathieu-Mahul, D. (1994). GATA-and SP1-binding sites are required for the full activity of the tissue-specific promoter of the tal-1 gene. Oncogene *9*, 2623-2632.
Lecuyer, E., Herblot, S., Saint-Denis, M., Martin, R., Begley, C.G., Porcher, C., Orkin, S.H., and Hoang, T. (2002). The SCL complex regulates c-kit expression in hematopoietic cells through functional interaction with Sp1. Blood *100*, 2430-2440.
Lee, D.H., Gershenzon, N., Gupta, M., Ioshikhes, I.P., Reinberg, D., and Lewis, B.A. (2005a). Functional characterization of core promoter elements: the downstream core element is recognized by TAF1. Mol Cell Biol *25*, 9674-9686.
Lee, D.Y., Teyssier, C., Strahl, B.D., and Stallcup, M.R. (2005b). Role of protein methylation in regulation of transcription. Endocrine reviews *26*, 147-170.
Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell *75*, 843-854.
Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K.*, et al.* (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. Cell *125*, 301-313.
Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.T., Thompson, C.M., Simon, I.*, et al.* (2002). Transcriptional regulatory networks in Saccharomyces cerevisiae. Science *298*, 799-804.
Lee, T.I., and Young, R.A. (2000). Transcription of eukaryotic protein-coding genes. Annual review of genetics *34*, 77-137.

Leroy-Viard, K., Vinit, M.A., Lecointe, N., Mathieu-Mahul, D., and Romeo, P.H. (1994). Distinct DNase-I hypersensitive sites are associated with TAL-1 transcription in erythroid and T-cell lines. Blood *84*, 3819-3827.

Letting, D.L., Chen, Y.Y., Rakowski, C., Reedy, S., and Blobel, G.A. (2004). Context-dependent regulation of GATA-1 by friend of GATA-1. Proc Natl Acad Sci U S A *101*, 476-481.

Leuther, K.K., Salmeron, J.M., and Johnston, S.A. (1993). Genetic evidence that an activation domain of GAL4 does not require acidity and may form a beta sheet. Cell *72*, 575-585.

Li, M., Li, H., and Rossi, J.J. (2006). RNAi in combination with a ribozyme and TAR decoy for treatment of HIV infection in hematopoietic cell gene therapy. Annals of the New York Academy of Sciences *1082*, 172-179.

Li, Q., Peterson, K.R., Fang, X., and Stamatoyannopoulos, G. (2002). Locus control regions. Blood *100*, 3077-3086.

Li, Z., Van Calcar, S., Qu, C., Cavenee, W.K., Zhang, M.Q., and Ren, B. (2003). A global transcriptional regulatory role for c-Myc in Burkitt's lymphoma cells. Proc Natl Acad Sci U S A *100*, 8164-8169.

Liang, P., and Pardee, A.B. (1992). Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science (New York, NY *257*, 967-971.

Lieb, J.D., Liu, X., Botstein, D., and Brown, P.O. (2001). Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. Nat Genet *28*, 327-334.

Lim, C.Y., Santoso, B., Boulay, T., Dong, E., Ohler, U., and Kadonaga, J.T. (2004). The MTE, a new core promoter element for transcription by RNA polymerase II. Genes Dev *18*, 1606-1617.

Lipshutz, R.J., Fodor, S.P., Gingeras, T.R., and Lockhart, D.J. (1999). High density synthetic oligonucleotide arrays. Nat Genet *21*, 20-24.

Liu, X., Noll, D.M., Lieb, J.D., and Clarke, N.D. (2005). DIP-chip: rapid and accurate determination of DNA-binding specificity. Genome Res *15*, 421-427.

Liu, X.S., Brutlag, D.L., and Liu, J.S. (2002). An algorithm for finding protein-DNA binding sites with applications to chromatin-immunoprecipitation microarray experiments. Nature biotechnology *20*, 835-839.

Loh, Y.H., Wu, Q., Chew, J.L., Vega, V.B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J., *et al.* (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat Genet *38*, 431-440.

Lorch, Y., Zhang, M., and Kornberg, R.D. (1999). Histone octamer transfer by a chromatin-remodeling complex. Cell *96*, 389-392.

Louie, M.C., Yang, H.Q., Ma, A.H., Xu, W., Zou, J.X., Kung, H.J., and Chen, H.W. (2003). Androgen-induced recruitment of RNA polymerase II to a nuclear receptor-p160 coactivator complex. Proc Natl Acad Sci U S A *100*, 2226-2230.
Lugus, J.J., Chung, Y.S., Mills, J.C., Kim, S.I., Grass, J., Kyba, M., Doherty, J.M., Bresnick, E.H., and Choi, K. (2007). GATA2 functions at multiple steps in hemangioblast development and differentiation. Development (Cambridge, England) *134*, 393-405.
Lum, L., Yao, S., Mozer, B., Rovescalli, A., Von Kessler, D., Nirenberg, M., and Beachy, P.A. (2003). Identification of Hedgehog pathway components by RNAi in Drosophila cultured cells. Science (New York, NY *299*, 2039-2045.
Mackay, J.P., Kowalski, K., Fox, A.H., Czolij, R., King, G.F., and Crossley, M. (1998). Involvement of the N-finger in the self-association of GATA-1. J Biol Chem *273*, 30560-30567.
MacKeigan, J.P., Murphy, L.O., and Blenis, J. (2005). Sensitized RNAi screen of human kinases and phosphatases identifies new regulators of apoptosis and chemoresistance. Nature cell biology *7*, 591-600.
Mahadevan, L.C., Willis, A.C., and Barratt, M.J. (1991). Rapid histone H3 phosphorylation in response to growth factors, phorbol esters, okadaic acid, and protein synthesis inhibitors. Cell *65*, 775-783.
Mansson, R., Tsapogas, P., Akerlund, M., Lagergren, A., Gisler, R., and Sigvardsson, M. (2004). Pearson correlation analysis of microarray data allows for the identification of genetic targets for early B-cell factor. The Journal of biological chemistry *279*, 17905-17913.
Mao, D.Y., Watson, J.D., Yan, P.S., Barsyte-Lovejoy, D., Khosravi, F., Wong, W.W., Farnham, P.J., Huang, T.H., and Penn, L.Z. (2003). Analysis of Myc bound loci identified by CpG island arrays shows that Max is essential for Myc-dependent repression. Curr Biol *13*, 882-886.
Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., *et al.* (2005). Genome sequencing in microfabricated high-density picolitre reactors. Nature *437*, 376-380.
Marques, J.T., Devosse, T., Wang, D., Zamanian-Daryoush, M., Serbinowski, P., Hartmann, R., Fujita, T., Behlke, M.A., and Williams, B.R. (2006). A structural basis for discriminating between self and nonself double-stranded RNAs in mammalian cells. Nat Biotechnol *24*, 559-565.
Martin, D.I., and Orkin, S.H. (1990). Transcriptional activation and DNA binding by the erythroid factor GF-1/NF-E1/Eryf 1. Genes Dev *4*, 1886-1898.
Martin, D.I., Zon, L.I., Mutter, G., and Orkin, S.H. (1990). Expression of an erythroid transcription factor in megakaryocytic and mast cell lineages. Nature *344*, 444-447.

Marton, M.J., DeRisi, J.L., Bennett, H.A., Iyer, V.R., Meyer, M.R., Roberts, C.J., Stoughton, R., Burchard, J., Slade, D., Dai, H., *et al.* (1998). Drug target validation and identification of secondary drug target effects using DNA microarrays. Nat Med *4*, 1293-1301.
Martone, R., Euskirchen, G., Bertone, P., Hartman, S., Royce, T.E., Luscombe, N.M., Rinn, J.L., Nelson, F.K., Miller, P., Gerstein, M., *et al.* (2003). Distribution of NF-kappaB-binding sites across human chromosome 22. Proc Natl Acad Sci U S A *100*, 12247-12252.
Massari, M.E., Grant, P.A., Pray-Grant, M.G., Berger, S.L., Workman, J.L., and Murre, C. (1999). A conserved motif present in a class of helix-loop-helix proteins activates transcription by direct recruitment of the SAGA complex. Molecular cell *4*, 63-73.
Matsukura, S., Jones, P.A., and Takai, D. (2003). Establishment of conditional vectors for hairpin siRNA knockdowns. Nucleic acids research *31*, e77.
Matsumoto, N., Kubo, A., Liu, H., Akita, K., Laub, F., Ramirez, F., Keller, G., and Friedman, S.L. (2006). Developmental regulation of yolk sac hematopoiesis by Kruppel-like factor 6. Blood *107*, 1357-1365.
Matsumura, I., Kawasaki, A., Tanaka, H., Sonoyama, J., Ezoe, S., Minegishi, N., Nakajima, K., Yamamoto, M., and Kanakura, Y. (2000). Biologic significance of GATA-1 activities in Ras-mediated megakaryocytic differentiation of hematopoietic cell lines. Blood *96*, 2440-2450.
Medina, K.L., Pongubala, J.M., Reddy, K.L., Lancki, D.W., Dekoter, R., Kieslinger, M., Grosschedl, R., and Singh, H. (2004). Assembling a gene regulatory network for specification of the B cell fate. Developmental cell *7*, 607-617.
Medvinsky, A., and Dzierzak, E. (1996). Definitive hematopoiesis is autonomously initiated by the AGM region. Cell *86*, 897-906.
Meersseman, G., Pennings, S., and Bradbury, E.M. (1992). Mobile nucleosomes--a general behavior. Embo J *11*, 2951-2959.
Mehaffey, M.G., Newton, A.L., Gandhi, M.J., Crossley, M., and Drachman, J.G. (2001). X-linked thrombocytopenia caused by a novel mutation of GATA-1. Blood *98*, 2681-2688.
Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., and Tuschl, T. (2004). Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. Mol Cell *15*, 185-197.
Mellentin, J.D., Smith, S.D., and Cleary, M.L. (1989). lyl-1, a novel gene altered by chromosomal translocation in T cell leukemia, codes for a protein with a helix-loop-helix DNA binding motif. Cell *58*, 77-83.
Meng, A., Tang, H., Yuan, B., Ong, B.A., Long, Q., and Lin, S. (1999). Positive and negative cis-acting elements are required for hematopoietic expression of zebrafish GATA-1. Blood *93*, 500-508.

Merika, M., and Orkin, S.H. (1995). Functional synergy and physical interactions of the erythroid transcription factor GATA-1 with the Kruppel family proteins Sp1 and EKLF. Mol Cell Biol *15*, 2437-2447.
Mermod, N., O'Neill, E.A., Kelly, T.J., and Tjian, R. (1989). The proline-rich transcriptional activator of CTF/NF-I is distinct from the replication and DNA binding domain. Cell *58*, 741-753.
Metzger, E., Wissmann, M., Yin, N., Muller, J.M., Schneider, R., Peters, A.H., Gunther, T., Buettner, R., and Schule, R. (2005). LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. Nature *437*, 436-439.
Migliaccio, A.R., Rana, R.A., Sanchez, M., Lorenzini, R., Centurione, L., Bianchi, L., Vannucchi, A.M., Migliaccio, G., and Orkin, S.H. (2003). GATA-1 as a regulator of mast cell differentiation revealed by the phenotype of the GATA-1low mouse mutant. J Exp Med *197*, 281-296.
Mikkola, H.K., Klintman, J., Yang, H., Hock, H., Schlaeger, T.M., Fujiwara, Y., and Orkin, S.H. (2003). Haematopoietic stem cells retain long-term repopulating activity and multipotency in the absence of stem-cell leukaemia SCL/tal-1 gene. Nature *421*, 547-551.
Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002). Network motifs: simple building blocks of complex networks. Science *298*, 824-827.
Milyavsky, M., Tabach, Y., Shats, I., Erez, N., Cohen, Y., Tang, X., Kalis, M., Kogan, I., Buganim, Y., Goldfinger, N., *et al.* (2005). Transcriptional programs following genetic alterations in p53, INK4A, and H-Ras genes along defined stages of malignant transformation. Cancer research *65*, 4530-4543.
Mineno, J., Okamoto, S., Ando, T., Sato, M., Chono, H., Izu, H., Takayama, M., Asada, K., Mirochnitchenko, O., Inouye, M., *et al.* (2006). The expression profile of microRNAs in mouse embryos. Nucleic acids research *34*, 1765-1771.
Moffat, J., Grueneberg, D.A., Yang, X., Kim, S.Y., Kloepfer, A.M., Hinkle, G., Piqani, B., Eisenhaure, T.M., Luo, B., Grenier, J.K., *et al.* (2006). A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. Cell *124*, 1283-1298.
Molkentin, J.D. (2000). The zinc finger-containing transcription factors GATA-4, -5, and -6. Ubiquitously expressed regulators of tissue-specific gene expression. J Biol Chem *275*, 38949-38952.
Moore, M.A., and Metcalf, D. (1970). Ontogeny of the haemopoietic system: yolk sac origin of in vivo and in vitro colony forming cells in the developing mouse embryo. British journal of haematology *18*, 279-296.
Morin, R.D., O'Connor, M.D., Griffith, M., Kuchenbauer, F., Delaney, A., Prabhu, A.L., Zhao, Y., McDonald, H., Zeng, T., Hirst, M., *et al.* (2008). Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome Res.

Moroni, E., Mastrangelo, T., Razzini, R., Cairns, L., Moi, P., Ottolenghi, S., and Giglioni, B. (2000). Regulation of mouse p45 NF-E2 transcription by an erythroid-specific GATA-dependent intronic alternative promoter. J Biol Chem *275*, 10567-10576.

Morris, K.V., Chan, S.W., Jacobsen, S.E., and Looney, D.J. (2004). Small interfering RNA-induced transcriptional gene silencing in human cells. Science (New York, NY *305*, 1289-1292.

Mouthon, M.A., Bernard, O., Mitjavila, M.T., Romeo, P.H., Vainchenker, W., and Mathieu-Mahul, D. (1993). Expression of tal-1 and GATA-binding proteins during human hematopoiesis. Blood *81*, 647-655.

Mukhopadhyay, M., Teufel, A., Yamashita, T., Agulnick, A.D., Chen, L., Downs, K.M., Schindler, A., Grinberg, A., Huang, S.P., Dorward, D.*, et al.* (2003). Functional ablation of the mouse Ldb1 gene results in severe patterning defects during gastrulation. Development (Cambridge, England) *130*, 495-505.

Murre, C., McCaw, P.S., and Baltimore, D. (1989). A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. Cell *56*, 777-783.

Myers, J.W., Jones, J.T., Meyer, T., and Ferrell, J.E., Jr. (2003). Recombinant Dicer efficiently converts large dsRNAs into siRNAs suitable for gene silencing. Nat Biotechnol *21*, 324-328.

Myers, L.C., and Kornberg, R.D. (2000). Mediator of transcriptional regulation. Annu Rev Biochem *69*, 729-749.

Nagy, P.L., Griesenbeck, J., Kornberg, R.D., and Cleary, M.L. (2002). A trithorax-group complex purified from Saccharomyces cerevisiae is required for methylation of histone H3. Proc Natl Acad Sci U S A *99*, 90-94.

Neale, G.A., Rehg, J.E., and Goorha, R.M. (1995). Ectopic expression of rhombotin-2 causes selective expansion of CD4-CD8- lymphocytes in the thymus and T-cell tumors in transgenic mice. Blood *86*, 3060-3071.

Nelson, C.J., Santos-Rosa, H., and Kouzarides, T. (2006). Proline isomerization of histone H3 regulates lysine methylation and gene expression. Cell *126*, 905-916.

Ngo, V.N., Davis, R.E., Lamy, L., Yu, X., Zhao, H., Lenz, G., Lam, L.T., Dave, S., Yang, L., Powell, J.*, et al.* (2006). A loss-of-function RNA interference screen for molecular targets in cancer. Nature *441*, 106-110.

Nichols, K.E., Crispino, J.D., Poncz, M., White, J.G., Orkin, S.H., Maris, J.M., and Weiss, M.J. (2000). Familial dyserythropoietic anaemia and thrombocytopenia due to an inherited mutation in GATA1. Nature genetics *24*, 266-270.

Nicolis, S., Bertini, C., Ronchi, A., Crotta, S., Lanfranco, L., Moroni, E., Giglioni, B., and Ottolenghi, S. (1991). An erythroid specific enhancer upstream to the gene encoding the cell-type specific transcription factor GATA-1. Nucleic Acids Res *19*, 5285-5291.

Nilsson, R., Bajic, V.B., Suzuki, H., di Bernardo, D., Bjorkegren, J., Katayama, S., Reid, J.F., Sweet, M.J., Gariboldi, M., Carninci, P., *et al.* (2006). Transcriptional network dynamics in macrophage activation. Genomics *88*, 133-142.

Nishikawa, K., Kobayashi, M., Masumi, A., Lyons, S.E., Weinstein, B.M., Liu, P.P., and Yamamoto, M. (2003). Self-association of Gata1 enhances transcriptional activity in vivo in zebra fish embryos. Mol Cell Biol *23*, 8295-8305.

Nishimura, S., Takahashi, S., Kuroha, T., Suwabe, N., Nagasawa, T., Trainor, C., and Yamamoto, M. (2000). A GATA box in the GATA-1 gene hematopoietic enhancer is a critical element in the network of GATA factors and sites that regulate this gene. Mol Cell Biol *20*, 713-723.

Nottingham, W.T., Jarratt, A., Burgess, M., Speck, C.L., Cheng, J.F., Prabhakar, S., Rubin, E.M., Li, P.S., Sloane-Stanley, J., Kong, A.S.J., *et al.* (2007). Runx1-mediated hematopoietic stem-cell emergence is controlled by a Gata/Ets/SCL-regulated enhancer. Blood *110*, 4188-4197.

Nourse, J., Mellentin, J.D., Galili, N., Wilkinson, J., Stanbridge, E., Smith, S.D., and Cleary, M.L. (1990). Chromosomal translocation t(1;19) results in synthesis of a homeobox fusion mRNA that codes for a potential chimeric transcription factor. Cell *60*, 535-545.

Nunez, E., Kwon, Y.S., Hutt, K.R., Hu, Q., Cardamone, M.D., Ohgi, K.A., Garcia-Bassets, I., Rose, D.W., Glass, C.K., Rosenfeld, M.G., *et al.* (2008). Nuclear receptor-enhanced transcription requires motor- and LSD1-dependent gene networking in interchromatin granules. Cell *134*, 189.

O'Neil, J., Shank, J., Cusson, N., Murre, C., and Kelliher, M. (2004). TAL1/SCL induces leukemia by inhibiting the transcriptional activity of E47/HEB. Cancer Cell *5*, 587-596.

Oberley, M.J., Inman, D.R., and Farnham, P.J. (2003). E2F6 negatively regulates BRCA1 in human cancer cells without methylation of histone H3 on lysine 9. J Biol Chem *278*, 42466-42476.

Odom, D.T., Dowell, R.D., Jacobsen, E.S., Gordon, W., Danford, T.W., MacIsaac, K.D., Rolfe, P.A., Conboy, C.M., Gifford, D.K., and Fraenkel, E. (2007). Tissue-specific transcriptional regulation has diverged significantly between human and mouse. Nat Genet *39*, 730-732.

Odom, D.T., Zizlsperger, N., Gordon, D.B., Bell, G.W., Rinaldi, N.J., Murray, H.L., Volkert, T.L., Schreiber, J., Rolfe, P.A., Gifford, D.K., *et al.* (2004). Control of pancreas and liver gene expression by HNF transcription factors. Science *303*, 1378-1381.

Ogilvy, S., Ferreira, R., Piltz, S.G., Bowen, J.M., Gottgens, B., and Green, A.R. (2007). The SCL +40 enhancer targets the midbrain together with primitive and definitive hematopoiesis and is regulated by SCL and GATA proteins. Mol Cell Biol *27*, 7206-7219.

Oliva, A., Rosebrock, A., Ferrezuelo, F., Pyne, S., Chen, H., Skiena, S., Futcher, B., and Leatherwood, J. (2005). The cell cycle-regulated genes of Schizosaccharomyces pombe. PLoS Biol *3*, e225.

Olsen, P.H., and Ambros, V. (1999). The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation. Developmental biology *216*, 671-680.
Onizuka, T., Endo, S., Hirano, M., Kanai, S., and Akiyama, H. (2002). Design of a fluorescent electrophoretic mobility shift assay improved for the quantitative and multiple analysis of protein-DNA complexes. Bioscience, biotechnology, and biochemistry *66*, 2732-2734.
Ono, Y., Fukuhara, N., and Yoshie, O. (1997). Transcriptional activity of TAL1 in T cell acute lymphoblastic leukemia (T-ALL) requires RBTN1 or -2 and induces TALLA1, a highly specific tumor marker of T-ALL. J Biol Chem *272*, 4576-4581.
Ono, Y., Fukuhara, N., and Yoshie, O. (1998). TAL1 and LIM-only proteins synergistically induce retinaldehyde dehydrogenase 2 expression in T-cell acute lymphoblastic leukemia by acting as cofactors for GATA3. Mol Cell Biol *18*, 6939-6950.
Onodera, K., Takahashi, S., Nishimura, S., Ohta, J., Motohashi, H., Yomogida, K., Hayashi, N., Engel, J.D., and Yamamoto, M. (1997a). GATA-1 transcription is controlled by distinct regulatory mechanisms during primitive and definitive erythropoiesis. Proc Natl Acad Sci U S A *94*, 4487-4492.
Onodera, K., Yomogida, K., Suwabe, N., Takahashi, S., Muraosa, Y., Hayashi, N., Ito, E., Gu, L., Rassoulzadegan, M., Engel, J.D., *et al.* (1997b). Conserved structure, regulatory elements, and transcriptional regulation from the GATA-1 gene testis promoter. Journal of biochemistry *121*, 251-263.
Orian, A., van Steensel, B., Delrow, J., Bussemaker, H.J., Li, L., Sawado, T., Williams, E., Loo, L.W., Cowley, S.M., Yost, C., *et al.* (2003). Genomic binding by the Drosophila Myc, Max, Mad/Mnt transcription factor network. Genes Dev *17*, 1101-1114.
Pabo, C.O., and Sauer, R.T. (1992). Transcription factors: structural families and principles of DNA recognition. Annu Rev Biochem *61*, 1053-1095.
Paddison, P.J., Caudy, A.A., Bernstein, E., Hannon, G.J., and Conklin, D.S. (2002). Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. Genes & development *16*, 948-958.
Pal, S., Cantor, A.B., Johnson, K.D., Moran, T.B., Boyer, M.E., Orkin, S.H., and Bresnick, E.H. (2004). Coregulator-dependent facilitation of chromatin occupancy by GATA-1. Proc Natl Acad Sci U S A *101*, 980-985.
Palomero, T., Odom, D.T., O'Neil, J., Ferrando, A.A., Margolin, A., Neuberg, D.S., Winter, S.S., Larson, R.S., Li, W., Liu, X.S., *et al.* (2006). Transcriptional regulatory networks downstream of TAL1/SCL in T-cell acute lymphoblastic leukemia. Blood *108*, 986-992.

Parry, D.H., Xu, J., and Ruvkun, G. (2007). A whole-genome RNAi Screen for C. elegans miRNA pathway genes. Curr Biol *17*, 2013-2022.
Pebernard, S., and Iggo, R.D. (2004). Determinants of interferon-stimulated gene induction by RNAi vectors. Differentiation; research in biological diversity *72*, 103-111.
Pennacchio, L.A., Ahituv, N., Moses, A.M., Prabhakar, S., Nobrega, M.A., Shoukry, M., Minovitsky, S., Dubchak, I., Holt, A., Lewis, K.D.*, et al.* (2006). In vivo enhancer analysis of human conserved non-coding sequences. Nature *444*, 499-502.
Pevny, L., Lin, C.S., D'Agati, V., Simon, M.C., Orkin, S.H., and Costantini, F. (1995). Development of hematopoietic cells lacking transcription factor GATA-1. Development (Cambridge, England) *121*, 163-172.
Pevny, L., Simon, M.C., Robertson, E., Klein, W.H., Tsai, S.F., D'Agati, V., Orkin, S.H., and Costantini, F. (1991). Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. Nature *349*, 257-260.
Pokholok, D.K., Hannett, N.M., and Young, R.A. (2002). Exchange of RNA polymerase II initiation and elongation factors during gene expression in vivo. Molecular cell *9*, 799-809.
Pokholok, D.K., Harbison, C.T., Levine, S., Cole, M., Hannett, N.M., Lee, T.I., Bell, G.W., Walker, K., Rolfe, P.A., Herbolsheimer, E.*, et al.* (2005). Genome-wide map of nucleosome acetylation and methylation in yeast. Cell *122*, 517-527.
Porcher, C., Swat, W., Rockwell, K., Fujiwara, Y., Alt, F.W., and Orkin, S.H. (1996). The T cell leukemia oncoprotein SCL/tal-1 is essential for development of all hematopoietic lineages. Cell *86*, 47-57.
Privalsky, M.L. (2004). The role of corepressors in transcriptional regulation by nuclear hormone receptors. Annual review of physiology *66*, 315-360.
Ptashne, M., and Gann, A. (1997). Transcriptional activation by recruitment. Nature *386*, 569-577.
Pulford, K., Lecointe, N., Leroy-Viard, K., Jones, M., Mathieu-Mahul, D., and Mason, D.Y. (1995). Expression of TAL-1 proteins in human tissues. Blood *85*, 675-684.
Qiu, Y., Sharma, A., and Stein, R. (1998). p300 mediates transcriptional stimulation by the basic helix-loop-helix activators of the insulin gene. Mol Cell Biol *18*, 2957-2964.
Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B.D., Sun, Z.W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D.*, et al.* (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature *406*, 593-599.
Rekhtman, N., Radparvar, F., Evans, T., and Skoultchi, A.I. (1999). Direct interaction of hematopoietic transcription factors PU.1 and GATA-1: functional antagonism in erythroid cells. Genes Dev *13*, 1398-1411.

Ren, B., Cam, H., Takahashi, Y., Volkert, T., Terragni, J., Young, R.A., and Dynlacht, B.D. (2002). E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints. Genes Dev *16*, 245-256.

Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E.*, et al.* (2000). Genome-wide location and function of DNA binding proteins. Science *290*, 2306-2309.

Reynaud, D., Ravet, E., Titeux, M., Mazurier, F., Renia, L., Dubart-Kupperschmitt, A., Romeo, P.H., and Pflumio, F. (2005). SCL/TAL1 expression level regulates human hematopoietic stem cell self-renewal and engraftment. Blood *106*, 2318-2328.

Rippe, K., von Hippel, P.H., and Langowski, J. (1995). Action at a distance: DNA-looping and initiation of transcription. Trends Biochem Sci *20*, 500-506.

Robb, L., Elwood, N.J., Elefanty, A.G., Kontgen, F., Li, R., Barnett, L.D., and Begley, C.G. (1996). The scl gene product is required for the generation of all hematopoietic lineages in the adult mouse. Embo J *15*, 4123-4129.

Robb, L., Lyons, I., Li, R., Hartley, L., Kontgen, F., Harvey, R.P., Metcalf, D., and Begley, C.G. (1995). Absence of yolk sac hematopoiesis from mice with a targeted disruption of the scl gene. Proc Natl Acad Sci U S A *92*, 7075-7079.

Robert, F., Pokholok, D.K., Hannett, N.M., Rinaldi, N.J., Chandy, M., Rolfe, A., Workman, J.L., Gifford, D.K., and Young, R.A. (2004). Global position and recruitment of HATs and HDACs in the yeast genome. Molecular cell *16*, 199-209.

Robertson, S.M., Kennedy, M., Shannon, J.M., and Keller, G. (2000). A transitional stage in the commitment of mesoderm to hematopoiesis requiring the transcription factor SCL/tal-1. Development *127*, 2447-2459.

Robyr, D., Suka, Y., Xenarios, I., Kurdistani, S.K., Wang, A., Suka, N., and Grunstein, M. (2002). Microarray deacetylation maps determine genome-wide functions for yeast histone deacetylases. Cell *109*, 437-446.

Robzyk, K., Recht, J., and Osley, M.A. (2000). Rad6-dependent ubiquitination of histone H2B in yeast. Science *287*, 501-504.

Roth, F.P., Hughes, J.D., Estep, P.W., and Church, G.M. (1998). Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation. Nature biotechnology *16*, 939-945.

Rylski, M., Welch, J.J., Chen, Y.Y., Letting, D.L., Diehl, J.A., Chodosh, L.A., Blobel, G.A., and Weiss, M.J. (2003). GATA-1-mediated proliferation arrest during erythroid maturation. Mol Cell Biol *23*, 5031-5042.

Saha, A., Wittmeyer, J., and Cairns, B.R. (2006). Chromatin remodelling: the industrial revolution of DNA around histones. Nature reviews *7*, 437-447.
Salghetti, S.E., Caudy, A.A., Chenoweth, J.G., and Tansey, W.P. (2001). Regulation of transcriptional activation domain function by ubiquitin. Science *293*, 1651-1653.
Sanchez, M., Gottgens, B., Sinclair, A.M., Stanley, M., Begley, C.G., Hunter, S., and Green, A.R. (1999). An SCL 3' enhancer targets developing endothelium together with embryonic and adult haematopoietic progenitors. Development *126*, 3891-3904.
Sanchez, M.J., Bockamp, E.O., Miller, J., Gambardella, L., and Green, A.R. (2001). Selective rescue of early haematopoietic progenitors in Scl(-/-) mice by expressing Scl under the control of a stem cell enhancer. Development *128*, 4815-4827.
Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W.W., and Lenhard, B. (2004). JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res *32*, D91-94.
Sassone-Corsi, P., Mizzen, C.A., Cheung, P., Crosio, C., Monaco, L., Jacquot, S., Hanauer, A., and Allis, C.D. (1999). Requirement of Rsk-2 for epidermal growth factor-activated phosphorylation of histone H3. Science *285*, 886-891.
Scacheri, P.C., Rozenblatt-Rosen, O., Caplen, N.J., Wolfsberg, T.G., Umayam, L., Lee, J.C., Hughes, C.M., Shanmugam, K.S., Bhattacharjee, A., Meyerson, M., *et al.* (2004). Short interfering RNAs can induce unexpected and divergent changes in the levels of untargeted proteins in mammalian cells. Proceedings of the National Academy of Sciences of the United States of America *101*, 1892-1897.
Schena, M., Shalon, D., Davis, R.W., and Brown, P.O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science (New York, NY *270*, 467-470.
Scherf, M., Klingenhoff, A., and Werner, T. (2000). Highly specific localization of promoter regions in large genomic sequences by PromoterInspector: a novel context analysis approach. J Mol Biol *297*, 599-606.
Schiebel, W., Pelissier, T., Riedel, L., Thalmeir, S., Schiebel, R., Kempe, D., Lottspeich, F., Sanger, H.L., and Wassenegger, M. (1998). Isolation of an RNA-directed RNA polymerase-specific cDNA clone from tomato. The Plant cell *10*, 2087-2101.
Schlabach, M.R., Luo, J., Solimini, N.L., Hu, G., Xu, Q., Li, M.Z., Zhao, Z., Smogorzewska, A., Sowa, M.E., Ang, X.L., *et al.* (2008). Cancer proliferation gene discovery through functional genomics. Science (New York, NY *319*, 620-624.
Schmitter, D., Filkowski, J., Sewer, A., Pillai, R.S., Oakeley, E.J., Zavolan, M., Svoboda, P., and Filipowicz, W. (2006). Effects of Dicer and Argonaute down-regulation on mRNA levels in human HEK293 cells. Nucleic acids research *34*, 4801-4815.

Schones, D.E., Cui, K., Cuddapah, S., Roh, T.Y., Barski, A., Wang, Z., Wei, G., and Zhao, K. (2008). Dynamic regulation of nucleosome positioning in the human genome. Cell *132*, 887-898.
Schug, J., Overton, G.C. (1997). TESS: Transcriptional Element Search Software on the WWW. Technical Report CBIL-TR-1997-1001-v0.0. . Computational Biology and Informatics Laboratory, School of Medicine, University of Pennsylvania.
Segal, E., Shapira, M., Regev, A., Pe'er, D., Botstein, D., Koller, D., and Friedman, N. (2003). Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. Nat Genet *34*, 166-176.
Sen, G., Wehrman, T.S., Myers, J.W., and Blau, H.M. (2004). Restriction enzyme-generated siRNA (REGS) vectors and libraries. Nature genetics *36*, 183-189.
Sertil, O., Kapoor, R., Cohen, B.D., Abramova, N., and Lowry, C.V. (2003). Synergistic repression of anaerobic genes by Mot3 and Rox1 in Saccharomyces cerevisiae. Nucleic Acids Res *31*, 5831-5837.
Shabalina, S.A., and Spiridonov, N.A. (2004). The mammalian transcriptome and the function of non-coding DNA sequences. Genome Biol *5*, 105.
Shannon, M.F., and Rao, S. (2002). Transcription. Of chips and ChIPs. Science *296*, 666-669.
Shen-Orr, S.S., Milo, R., Mangan, S., and Alon, U. (2002). Network motifs in the transcriptional regulation network of Escherichia coli. Nat Genet *31*, 64-68.
Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., and Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell *119*, 941-953.
Shi, Y.J., Matson, C., Lan, F., Iwase, S., Baba, T., and Shi, Y. (2005). Regulation of LSD1 histone demethylase activity by its associated factors. Mol Cell *19*, 857-864.
Shimizu, R., Trainor, C.D., Nishikawa, K., Kobayashi, M., Ohneda, K., and Yamamoto, M. (2007). GATA-1 self-association controls erythroid development in vivo. J Biol Chem *282*, 15862-15871.
Shipp, M.A., Ross, K.N., Tamayo, P., Weng, A.P., Kutok, J.L., Aguiar, R.C., Gaasenbeek, M., Angelo, M., Reich, M., Pinkus, G.S.*, et al.* (2002). Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. Nat Med *8*, 68-74.
Shiraki, T., Kondo, S., Katayama, S., Waki, K., Kasukawa, T., Kawaji, H., Kodzius, R., Watahiki, A., Nakamura, M., Arakawa, T.*, et al.* (2003). Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. Proc Natl Acad Sci U S A *100*, 15776-15781.
Shivdasani, R.A., Fujiwara, Y., McDevitt, M.A., and Orkin, S.H. (1997). A lineage-selective knockout establishes the critical role of transcription factor GATA-1 in megakaryocyte growth and platelet development. Embo J *16*, 3965-3973.

Shivdasani, R.A., Mayer, E.L., and Orkin, S.H. (1995). Absence of blood formation in mice lacking the T-cell leukaemia oncoprotein tal-1/SCL. Nature *373*, 432-434.
Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S., *et al.* (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res *15*, 1034-1050.
Sijen, T., Fleenor, J., Simmer, F., Thijssen, K.L., Parrish, S., Timmons, L., Plasterk, R.H., and Fire, A. (2001). On the role of RNA amplification in dsRNA-triggered gene silencing. Cell *107*, 465-476.
Silberstein, L., Sanchez, M.J., Socolovsky, M., Liu, Y., Hoffman, G., Kinston, S., Piltz, S., Bowen, M., Gambardella, L., Green, A.R., *et al.* (2005). Transgenic analysis of the stem cell leukemia +19 stem cell enhancer in adult and embryonic hematopoietic and endothelial cells. Stem Cells *23*, 1378-1388.
Silva, J.M., Li, M.Z., Chang, K., Ge, W., Golding, M.C., Rickles, R.J., Siolas, D., Hu, G., Paddison, P.J., Schlabach, M.R., *et al.* (2005). Second-generation shRNA libraries covering the mouse and human genomes. Nature genetics *37*, 1281-1288.
Silva, J.M., Marran, K., Parker, J.S., Silva, J., Golding, M., Schlabach, M.R., Elledge, S.J., Hannon, G.J., and Chang, K. (2008). Profiling essential genes in human mammary cells by multiplex RNAi screening. Science (New York, NY *319*, 617-620.
Silver, L., and Palis, J. (1997). Initiation of murine embryonic erythropoiesis: a spatial analysis. Blood *89*, 1154-1164.
Singh-Gasson, S., Green, R.D., Yue, Y., Nelson, C., Blattner, F., Sussman, M.R., and Cerrina, F. (1999). Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. Nat Biotechnol *17*, 974-978.
Sledz, C.A., Holko, M., de Veer, M.J., Silverman, R.H., and Williams, B.R. (2003). Activation of the interferon system by short-interfering RNAs. Nature cell biology *5*, 834-839.
Smale, S.T., and Kadonaga, J.T. (2003). The RNA polymerase II core promoter. Annu Rev Biochem *72*, 449-479.
Smardon, A., Spoerke, J.M., Stacey, S.C., Klein, M.E., Mackin, N., and Maine, E.M. (2000). EGO-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in C. elegans. Curr Biol *10*, 169-178.
Smith, E.M., Gisler, R., and Sigvardsson, M. (2002). Cloning and characterization of a promoter flanking the early B cell factor (EBF) gene indicates roles for E-proteins and autoregulation in the control of EBF expression. J Immunol *169*, 261-270.
Song, E., Zhu, P., Lee, S.K., Chowdhury, D., Kussman, S., Dykxhoorn, D.M., Feng, Y., Palliser, D., Weiner, D.B., Shankar, P., *et al.* (2005). Antibody mediated in vivo delivery of small interfering RNAs via cell-surface receptors. Nat Biotechnol *23*, 709-717.

Song, S.H., Hou, C., and Dean, A. (2007). A positive role for NLI/Ldb1 in long-range beta-globin locus control region function. Molecular cell *28*, 810-822.
Sonnichsen, B., Koski, L.B., Walsh, A., Marschall, P., Neumann, B., Brehm, M., Alleaume, A.M., Artelt, J., Bettencourt, P., Cassin, E., *et al.* (2005). Full-genome RNAi profiling of early embryogenesis in Caenorhabditis elegans. Nature *434*, 462-469.
Spicuglia, S., Kumar, S., Yeh, J.H., Vachez, E., Chasson, L., Gorbatch, S., Cautres, J., and Ferrier, P. (2002). Promoter activation by enhancer-dependent and -independent loading of activator and coactivator complexes. Mol Cell *10*, 1479-1487.
Spilianakis, C.G., Lalioti, M.D., Town, T., Lee, G.R., and Flavell, R.A. (2005). Interchromosomal associations between alternatively expressed loci. Nature *435*, 637-645.
Stewart, S.A., Dykxhoorn, D.M., Palliser, D., Mizuno, H., Yu, E.Y., An, D.S., Sabatini, D.M., Chen, I.S., Hahn, W.C., Sharp, P.A., *et al.* (2003). Lentivirus-delivered stable gene silencing by RNAi in primary cells. RNA (New York, NY *9*, 493-501.
Stielow, B., Sapetschnig, A., Kruger, I., Kunert, N., Brehm, A., Boutros, M., and Suske, G. (2008). Identification of SUMO-dependent chromatin-associated transcriptional repression components by a genome-wide RNAi screen. Mol Cell *29*, 742-754.
Stopka, T., Amanatullah, D.F., Papetti, M., and Skoultchi, A.I. (2005). PU.1 inhibits the erythroid program by binding to GATA-1 on DNA and creating a repressive chromatin structure. Embo J *24*, 3712-3723.
Studitsky, V.M., Clark, D.J., and Felsenfeld, G. (1994). A histone octamer can step around a transcribing polymerase without leaving the template. Cell *76*, 371-382.
Sun, X.H., and Baltimore, D. (1991). An inhibitory domain of E12 transcription factor prevents DNA binding in E12 homodimers but not in E12 heterodimers. Cell *64*, 459-470.
Suzuki, Y., Tsunoda, T., Sese, J., Taira, H., Mizushima-Sugano, J., Hata, H., Ota, T., Isogai, T., Tanaka, T., Nakamura, Y., *et al.* (2001). Identification and characterization of the potential promoter regions of 1031 kinds of human genes. Genome research *11*, 677-684.
Swiers, G., Patient, R., and Loose, M. (2006). Genetic regulatory networks programming hematopoietic stem cells and erythroid lineage specification. Dev Biol *294*, 525-540.
Szekvolgyi, L., Balint, B.L., Imre, L., Goda, K., Szabo, M., Nagy, L., and Szabo, G. (2006). Chip-on-beads: flow-cytometric evaluation of chromatin immunoprecipitation. Cytometry A *69*, 1086-1091.
Tabara, H., Grishok, A., and Mello, C.C. (1998). RNAi in C. elegans: soaking in the genome sequence. Science (New York, NY *282*, 430-431.
Tavazoie, S., Hughes, J.D., Campbell, M.J., Cho, R.J., and Church, G.M. (1999). Systematic determination of genetic network architecture. Nat Genet *22*, 281-285.

Tavernarakis, N., Wang, S.L., Dorovkov, M., Ryazanov, A., and Driscoll, M. (2000). Heritable and inducible genetic interference by double-stranded RNA encoded by transgenes. Nature genetics *24*, 180-183.

Thomson, S., Clayton, A.L., Hazzalin, C.A., Rose, S., Barratt, M.J., and Mahadevan, L.C. (1999). The nucleosomal response associated with immediate-early gene induction is mediated via alternative MAP kinase cascades: MSK1 as a potential histone H3/HMG-14 kinase. Embo J *18*, 4779-4793.

Thut, C.J., Chen, J.L., Klemm, R., and Tjian, R. (1995). p53 transcriptional activation mediated by coactivators TAFII40 and TAFII60. Science *267*, 100-104.

Timmons, L., Court, D.L., and Fire, A. (2001). Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in Caenorhabditis elegans. Gene *263*, 103-112.

Timmons, L., and Fire, A. (1998). Specific interference by ingested dsRNA. Nature *395*, 854.

Torres, T.T., Metta, M., Ottenwalder, B., and Schlotterer, C. (2008). Gene expression profiling by massively parallel sequencing. Genome Res *18*, 172-177.

Tsai, F.Y., and Orkin, S.H. (1997). Transcription factor GATA-2 is required for proliferation/survival of early hematopoietic cells and mast cell formation, but not for erythroid and myeloid terminal differentiation. Blood *89*, 3636-3643.

Tsai, S.F., Martin, D.I., Zon, L.I., D'Andrea, A.D., Wong, G.G., and Orkin, S.H. (1989). Cloning of cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells. Nature *339*, 446-451.

Tsang, A.P., Visvader, J.E., Turner, C.A., Fujiwara, Y., Yu, C., Weiss, M.J., Crossley, M., and Orkin, S.H. (1997). FOG, a multitype zinc finger protein, acts as a cofactor for transcription factor GATA-1 in erythroid and megakaryocytic differentiation. Cell *90*, 109-119.

Tuerk, C., and Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science *249*, 505-510.

Utley, R.T., Ikeda, K., Grant, P.A., Cote, J., Steger, D.J., Eberharter, A., John, S., and Workman, J.L. (1998). Transcriptional activators direct histone acetyltransferase complexes to nucleosomes. Nature *394*, 498-502.

Valverde-Garduno, V., Guyot, B., Anguita, E., Hamlett, I., Porcher, C., and Vyas, P. (2004). Differences in the chromatin structure and cis-element organization of the human and mouse GATA1 loci: implications for cis-element identification. Blood *104*, 3106-3116.

van de Wetering, M., Oving, I., Muncan, V., Pon Fong, M.T., Brantjes, H., van Leenen, D., Holstege, F.C., Brummelkamp, T.R., Agami, R., and Clevers, H. (2003). Specific inhibition of gene expression using a stably integrated, inducible small-interfering-RNA vector. EMBO reports *4*, 609-615.

van Eekelen, J.A., Bradley, C.K., Gothert, J.R., Robb, L., Elefanty, A.G., Begley, C.G., and Harvey, A.R. (2003). Expression pattern of the stem cell leukaemia gene in the CNS of the embryonic and adult mouse. Neuroscience *122*, 421-436.
van Haaften, G., Romeijn, R., Pothof, J., Koole, W., Mullenders, L.H., Pastink, A., Plasterk, R.H., and Tijsterman, M. (2006). Identification of conserved pathways of DNA-damage response and radiation protection by genome-wide RNAi. Curr Biol *16*, 1344-1350.
van Steensel, B., and Henikoff, S. (2000). Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. Nature biotechnology *18*, 424-428.
Velculescu, V.E., Zhang, L., Vogelstein, B., and Kinzler, K.W. (1995). Serial analysis of gene expression. Science (New York, NY *270*, 484-487.
Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A.*, et al.* (2001). The sequence of the human genome. Science *291*, 1304-1351.
Vilar, J.M., and Saiz, L. (2005). DNA looping in gene regulation: from the assembly of macromolecular complexes to the control of transcriptional noise. Current opinion in genetics & development *15*, 136-144.
Visel, A., Minovitsky, S., Dubchak, I., and Pennacchio, L.A. (2007). VISTA Enhancer Browser--a database of tissue-specific human enhancers. Nucleic Acids Res *35*, D88-92.
Visvader, J., Begley, C.G., and Adams, J.M. (1991). Differential expression of the LYL, SCL and E2A helix-loop-helix genes within the hemopoietic system. Oncogene *6*, 187-194.
Visvader, J.E., Fujiwara, Y., and Orkin, S.H. (1998). Unsuspected role for the T-cell leukemia protein SCL/tal-1 in vascular development. Genes Dev *12*, 473-479.
Visvader, J.E., Mao, X., Fujiwara, Y., Hahm, K., and Orkin, S.H. (1997). The LIM-domain binding protein Ldb1 and its partner LMO2 act as negative regulators of erythroid differentiation. Proceedings of the National Academy of Sciences of the United States of America *94*, 13707-13712.
Vitelli, L., Condorelli, G., Lulli, V., Hoang, T., Luchetti, L., Croce, C.M., and Peschle, C. (2000). A pentamer transcriptional complex including tal-1 and retinoblastoma protein downmodulates c-kit expression in normal erythroblasts. Mol Cell Biol *20*, 5330-5342.
Vogel, M.J., Guelen, L., de Wit, E., Peric-Hupkes, D., Loden, M., Talhout, W., Feenstra, M., Abbas, B., Classen, A.K., and van Steensel, B. (2006). Human heterochromatin proteins form large domains containing KRAB-ZNF genes. Genome Res *16*, 1493-1504.
Vogel, M.J., Peric-Hupkes, D., and van Steensel, B. (2007). Detection of in vivo protein-DNA interactions using DamID in mammalian cells. Nature protocols *2*, 1467-1478.

Vokes, S.A., Ji, H., McCuine, S., Tenzen, T., Giles, S., Zhong, S., Longabaugh, W.J., Davidson, E.H., Wong, W.H., and McMahon, A.P. (2007). Genomic characterization of Gli-activator targets in sonic hedgehog-mediated neural patterning. Development *134*, 1977-1989.

Vyas, P., Ault, K., Jackson, C.W., Orkin, S.H., and Shivdasani, R.A. (1999a). Consequences of GATA-1 deficiency in megakaryocytes and platelets. Blood *93*, 2867-2875.

Vyas, P., McDevitt, M.A., Cantor, A.B., Katz, S.G., Fujiwara, Y., and Orkin, S.H. (1999b). Different sequence requirements for expression in erythroid and megakaryocytic cells within a regulatory element upstream of the GATA-1 gene. Development (Cambridge, England) *126*, 2799-2811.

Wadman, I., Li, J., Bash, R.O., Forster, A., Osada, H., Rabbitts, T.H., and Baer, R. (1994). Specific in vivo association between the bHLH and LIM proteins implicated in human T cell leukemia. Embo J *13*, 4831-4839.

Wadman, I.A., Osada, H., Grutz, G.G., Agulnick, A.D., Westphal, H., Forster, A., and Rabbitts, T.H. (1997). The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. Embo J *16*, 3145-3157.

Wakabayashi, J., Yomogida, K., Nakajima, O., Yoh, K., Takahashi, S., Engel, J.D., Ohneda, K., and Yamamoto, M. (2003). GATA-1 testis activation region is essential for Sertoli cell-specific expression of GATA-1 gene in transgenic mouse. Genes Cells *8*, 619-630.

Wall, L., deBoer, E., and Grosveld, F. (1988). The human beta-globin gene 3' enhancer contains multiple binding sites for an erythroid-specific protein. Genes Dev *2*, 1089-1100.

Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R.S., and Zhang, Y. (2004a). Role of histone H2A ubiquitination in Polycomb silencing. Nature *431*, 873-878.

Wang, Y., Wysocka, J., Sayegh, J., Lee, Y.H., Perlin, J.R., Leonelli, L., Sonbuchner, L.S., McDonald, C.H., Cook, R.G., Dou, Y.*, et al.* (2004b). Human PAD4 regulates histone arginine methylation levels via demethylimination. Science *306*, 279-283.

Warren, A.J., Colledge, W.H., Carlton, M.B., Evans, M.J., Smith, A.J., and Rabbitts, T.H. (1994). The oncogenic cysteine-rich LIM domain protein rbtn2 is essential for erythroid development. Cell *78*, 45-57.

Waterhouse, P.M., Wang, M.B., and Lough, T. (2001). Gene silencing as an adaptive defence against viruses. Nature *411*, 834-842.

Weber, F., de Villiers, J., and Schaffner, W. (1984). An SV40 "enhancer trap" incorporates exogenous enhancers or generates enhancers from its own sequences. Cell *36*, 983-992.

Wechsler, J., Greene, M., McDevitt, M.A., Anastasi, J., Karp, J.E., Le Beau, M.M., and Crispino, J.D. (2002). Acquired mutations in GATA1 in the megakaryoblastic leukemia of Down syndrome. Nature genetics *32*, 148-152.

Wei, C.L., Wu, Q., Vega, V.B., Chiu, K.P., Ng, P., Zhang, T., Shahab, A., Yong, H.C., Fu, Y., Weng, Z., *et al.* (2006). A global map of p53 transcription-factor binding sites in the human genome. Cell *124*, 207-219.

Weinberg, M.S., Villeneuve, L.M., Ehsani, A., Amarzguioui, M., Aagaard, L., Chen, Z.X., Riggs, A.D., Rossi, J.J., and Morris, K.V. (2006). The antisense strand of small interfering RNAs directs histone methylation and transcriptional gene silencing in human cells. RNA (New York, NY *12*, 256-262.

Weinmann, A.S., Bartley, S.M., Zhang, T., Zhang, M.Q., and Farnham, P.J. (2001). Use of chromatin immunoprecipitation to clone novel E2F target promoters. Mol Cell Biol *21*, 6820-6832.

Weinmann, A.S., Yan, P.S., Oberley, M.J., Huang, T.H., and Farnham, P.J. (2002). Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis. Genes Dev *16*, 235-244.

Weiss, M.J., Keller, G., and Orkin, S.H. (1994). Novel insights into erythroid development revealed through in vitro differentiation of GATA-1 embryonic stem cells. Genes Dev *8*, 1184-1197.

Weiss, M.J., and Orkin, S.H. (1995). Transcription factor GATA-1 permits survival and maturation of erythroid precursors by preventing apoptosis. Proc Natl Acad Sci U S A *92*, 9623-9627.

Welch, J.J., Watts, J.A., Vakoc, C.R., Yao, Y., Wang, H., Hardison, R.C., Blobel, G.A., Chodosh, L.A., and Weiss, M.J. (2004). Global regulation of erythroid gene expression by transcription factor GATA-1. Blood *104*, 3136-3147.

Wells, J., Yan, P.S., Cechvala, M., Huang, T., and Farnham, P.J. (2003). Identification of novel pRb binding sites using CpG microarrays suggests that E2F recruits pRb to specific genomic sites during S phase. Oncogene *22*, 1445-1460.

West, A.G., Gaszner, M., and Felsenfeld, G. (2002). Insulators: many functions, many mechanisms. Genes Dev *16*, 271-288.

Westbrook, T.F., Martin, E.S., Schlabach, M.R., Leng, Y., Liang, A.C., Feng, B., Zhao, J.J., Roberts, T.M., Mandel, G., Hannon, G.J., *et al.* (2005). A genetic screen for candidate tumor suppressors identifies REST. Cell *121*, 837-848.

White, K.P., Rifkin, S.A., Hurban, P., and Hogness, D.S. (1999). Microarray analysis of Drosophila development during metamorphosis. Science (New York, NY *286*, 2179-2184.

Wightman, B., Ha, I., and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. Cell *75*, 855-862.

Williams, B.R. (1997). Role of the double-stranded RNA-activated protein kinase (PKR) in cell regulation. Biochem Soc Trans *25*, 509-513.
Williams, R.R., Azuara, V., Perry, P., Sauer, S., Dvorkina, M., Jorgensen, H., Roix, J., McQueen, P., Misteli, T., Merkenschlager, M.*, et al.* (2006). Neural induction promotes large-scale chromatin reorganisation of the Mash1 locus. J Cell Sci *119*, 132-140.
Wolfrum, C., Shi, S., Jayaprakash, K.N., Jayaraman, M., Wang, G., Pandey, R.K., Rajeev, K.G., Nakayama, T., Charrise, K., Ndungo, E.M.*, et al.* (2007). Mechanisms and optimization of in vivo delivery of lipophilic siRNAs. Nat Biotechnol *25*, 1149-1157.
Wolstein, O., Silkov, A., Revach, M., and Dikstein, R. (2000). Specific interaction of TAFII105 with OCA-B is involved in activation of octamer-dependent transcription. J Biol Chem *275*, 16459-16465.
Wong, J.M., and Bateman, E. (1994). TBP-DNA interactions in the minor groove discriminate between A:T and T:A base pairs. Nucleic Acids Res *22*, 1890-1896.
Wood, A., Krogan, N.J., Dover, J., Schneider, J., Heidt, J., Boateng, M.A., Dean, K., Golshani, A., Zhang, Y., Greenblatt, J.F.*, et al.* (2003). Bre1, an E3 ubiquitin ligase required for recruitment and substrate selection of Rad6 at a promoter. Mol Cell *11*, 267-274.
Woolfe, A., Goodson, M., Goode, D.K., Snell, P., McEwen, G.K., Vavouri, T., Smith, S.F., North, P., Callaway, H., Kelly, K.*, et al.* (2005). Highly conserved non-coding sequences are associated with vertebrate development. PLoS biology *3*, e7.
Wright, W.E., Binder, M., and Funk, W. (1991). Cyclic amplification and selection of targets (CASTing) for the myogenin consensus binding site. Mol Cell Biol *11*, 4104-4110.
Wyrick, J.J., Aparicio, J.G., Chen, T., Barnett, J.D., Jennings, E.G., Young, R.A., Bell, S.P., and Aparicio, O.M. (2001). Genome-wide distribution of ORC and MCM proteins in S. cerevisiae: high-resolution mapping of replication origins. Science *294*, 2357-2360.
Xu, Z., Huang, S., Chang, L.S., Agulnick, A.D., and Brandt, S.J. (2003). Identification of a TAL1 target gene reveals a positive role for the LIM domain-binding protein Ldb1 in erythroid gene expression and differentiation. Mol Cell Biol *23*, 7585-7599.
Xu, Z., Meng, X., Cai, Y., Koury, M.J., and Brandt, S.J. (2006). Recruitment of the SWI/SNF protein Brg1 by a multiprotein complex effects transcriptional repression in murine erythroid progenitors. The Biochemical journal *399*, 297-304.
Yamada, Y., Pannell, R., Forster, A., and Rabbitts, T.H. (2000). The oncogenic LIM-only transcription factor Lmo2 regulates angiogenesis but not vasculogenesis in mice. Proceedings of the National Academy of Sciences of the United States of America *97*, 320-324.

Yamada, Y., Warren, A.J., Dobson, C., Forster, A., Pannell, R., and Rabbitts, T.H. (1998). The T cell leukemia LIM protein Lmo2 is necessary for adult mouse hematopoiesis. Proceedings of the National Academy of Sciences of the United States of America *95*, 3890-3895.
Yamaguchi, Y., Ackerman, S.J., Minegishi, N., Takiguchi, M., Yamamoto, M., and Suda, T. (1998). Mechanisms of transcription in eosinophils: GATA-1, but not GATA-2, transactivates the promoter of the eosinophil granule major basic protein gene. Blood *91*, 3447-3458.
Yamaguchi, Y., Nishio, H., Kishi, K., Ackerman, S.J., and Suda, T. (1999). C/EBPbeta and GATA-1 synergistically regulate activity of the eosinophil granule major basic protein promoter: implication for C/EBPbeta activity in eosinophil gene expression. Blood *94*, 1429-1439.
Yang, D., Buchholz, F., Huang, Z., Goga, A., Chen, C.Y., Brodsky, F.M., and Bishop, J.M. (2002). Short RNA duplexes produced by hydrolysis with Escherichia coli RNase III mediate effective RNA interference in mammalian cells. Proceedings of the National Academy of Sciences of the United States of America *99*, 9942-9947.
Yoder, M.C., Hiatt, K., and Mukherjee, P. (1997). In vivo repopulating hematopoietic stem cells are present in the murine yolk sac at day 9.0 postcoitus. Proc Natl Acad Sci U S A *94*, 6776-6780.
Yokomizo, T., Takahashi, S., Mochizuki, N., Kuroha, T., Ema, M., Wakamatsu, A., Shimizu, R., Ohneda, O., Osato, M., Okada, H., *et al.* (2007). Characterization of GATA-1(+) hemangioblastic cells in the mouse embryo. Embo J *26*, 184-196.
Yomogida, K., Ohtani, H., Harigae, H., Ito, E., Nishimune, Y., Engel, J.D., and Yamamoto, M. (1994). Developmental stage- and spermatogenic cycle-specific expression of transcription factor GATA-1 in mouse Sertoli cells. Development (Cambridge, England) *120*, 1759-1766.
Yoo, J.Y., Kim, J.H., Kim, J., Huang, J.H., Zhang, S.N., Kang, Y.A., Kim, H., and Yun, C.O. (2008). Short hairpin RNA-expressing oncolytic adenovirus-mediated inhibition of IL-8: effects on antiangiogenesis and tumor growth inhibition. Gene therapy *15*, 635-651.
Yu, C., Cantor, A.B., Yang, H., Browne, C., Wells, R.A., Fujiwara, Y., and Orkin, S.H. (2002a). Targeted deletion of a high-affinity GATA-binding site in the GATA-1 promoter leads to selective loss of the eosinophil lineage in vivo. J Exp Med *195*, 1387-1395.
Yu, C., Niakan, K.K., Matsushita, M., Stamatoyannopoulos, G., Orkin, S.H., and Raskind, W.H. (2002b). X-linked thrombocytopenia with thalassemia from a mutation in the amino finger of GATA-1 affecting DNA binding rather than FOG-1 interaction. Blood *100*, 2040-2045.
Zeller, K.I., Zhao, X., Lee, C.W., Chiu, K.P., Yao, F., Yustein, J.T., Ooi, H.S., Orlov, Y.L., Shahab, A., Yong, H.C., *et al.* (2006). Global mapping of c-Myc binding sites and target gene networks in human B cells. Proc Natl Acad Sci U S A *103*, 17834-17839.

Zeng, Y., Wagner, E.J., and Cullen, B.R. (2002). Both natural and designed micro RNAs can inhibit the expression of cognate mRNAs when expressed in human cells. Mol Cell *9*, 1327-1333.
Zenzie-Gregory, B., Khachi, A., Garraway, I.P., and Smale, S.T. (1993). Mechanism of initiator-mediated transcription: evidence for a functional interaction between the TATA-binding protein and DNA in the absence of a specific recognition sequence. Mol Cell Biol *13*, 3841-3849.
Zhang, P., Zhang, X., Iwama, A., Yu, C., Smith, K.A., Mueller, B.U., Narravula, S., Torbett, B.E., Orkin, S.H., and Tenen, D.G. (2000). PU.1 inhibits GATA-1 function and erythroid differentiation by blocking GATA-1 DNA binding. Blood *96*, 2641-2648.
Zhang, W., Yang, H., Kong, X., Mohapatra, S., San Juan-Vergara, H., Hellermann, G., Behera, S., Singam, R., Lockey, R.F., and Mohapatra, S.S. (2005a). Inhibition of respiratory syncytial virus infection with intranasal siRNA nanoparticles targeting the viral NS1 gene. Nature medicine *11*, 56-62.
Zhang, Y., Payne, K.J., Zhu, Y., Price, M.A., Parrish, Y.K., Zielinska, E., Barsky, L.W., and Crooks, G.M. (2005b). SCL expression at critical points in human hematopoietic lineage commitment. Stem Cells *23*, 852-860.
Zhao, H., and Dean, A. (2004). An insulator blocks spreading of histone acetylation and interferes with RNA polymerase II transfer between an enhancer and gene. Nucleic Acids Res *32*, 4903-4919.
Zhao, Z., Tavoosidana, G., Sjolinder, M., Gondor, A., Mariano, P., Wang, S., Kanduri, C., Lezcano, M., Sandhu, K.S., Singh, U., *et al.* (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. Nature genetics *38*, 1341-1347.
Zhu, B., Zheng, Y., Pham, A.D., Mandal, S.S., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. (2005). Monoubiquitination of human histone H2B: the factors involved and their roles in HOX gene regulation. Mol Cell *20*, 601-611.
Zhu, P., Zhou, W., Wang, J., Puc, J., Ohgi, K.A., Erdjument-Bromage, H., Tempst, P., Glass, C.K., and Rosenfeld, M.G. (2007). A histone H2A deubiquitinase complex coordinating histone acetylation and H1 dissociation in transcriptional regulation. Mol Cell *27*, 609-621.
Zhuang, Y., Jackson, A., Pan, L., Shen, K., and Dai, M. (2004). Regulation of E2A gene expression in B-lymphocyte development. Molecular immunology *40*, 1165-1177.
Zhuang, Y., Soriano, P., and Weintraub, H. (1994). The helix-loop-helix gene E2A is required for B cell formation. Cell *79*, 875-884.
Zimmermann, T.S., Lee, A.C., Akinc, A., Bramlage, B., Bumcrot, D., Fedoruk, M.N., Harborth, J., Heyes, J.A., Jeffs, L.B., John, M., *et al.* (2006). RNAi-mediated gene silencing in non-human primates. Nature *441*, 111-114.

Zon, L.I., Tsai, S.F., Burgess, S., Matsudaira, P., Bruns, G.A., and Orkin, S.H. (1990). The major human erythroid DNA-binding protein (GF-1): primary sequence and localization of the gene to the X chromosome. Proc Natl Acad Sci U S A *87*, 668-672.

Zon, L.I., Yamaguchi, Y., Yee, K., Albee, E.A., Kimura, A., Bennett, J.C., Orkin, S.H., and Ackerman, S.J. (1993). Expression of mRNA for the GATA-binding proteins in human eosinophils and basophils: potential role in gene transcription. Blood *81*, 3234-3241.

Zon, L.I., Youssoufian, H., Mather, C., Lodish, H.F., and Orkin, S.H. (1991). Activation of the erythropoietin receptor promoter by transcription factor GATA-1. Proc Natl Acad Sci U S A *88*, 10638-10641.

Chapter 2

Materials and Methods

2.1 Composition of solutions

Cell lysis buffer (for nuclear protein extraction)

- 10 mM Tris-HCL pH 8
- 10 mM NaCl
- 0.2% Igepal (Sigma)
- 1% protease inhibitor cocktail (Sigma)

Extraction buffer (for nuclear protein extraction)

- 20 mM HEPES pH 7.9
- 0.2 mM EDTA
- 25% Glycerol
- 1.5 mM $MgCl_2$
- 0.42 M NaCl
- 0.001 M DTT (Invitrogen)
- 1% protease inhibitor cocktail (Sigma)

10 × dNTP mix (for DNA labelling of ChIP assays)

- 1 mM dCTP
- 2 mM each of dGTP, dTTP and dATP

Tecan hybridisation buffer

- 50% formamide (Fluka)
- 5% dextran sulphate
- 0.1% Tween 20 (BDH)
- 2 × SSC
- 10 mM Tris pH 7.4

PBS/0.05% Tween 20 (hybridisation wash solution 1)

PBS/0.05% Tween 20 for washing the arrays was prepared by dissolving the following salts in one litre of HPLC water.

- 7.33 g NaCl
- 2.36 g Na_2HPO_4
- 1.52 g $NaH_2PO_4 \cdot H2O$
- 500 µl Tween 20 (Sigma)

Cell lysis buffer (CLB)

- 10 mM Tris-HCl pH 8.0
- 10 mM NaCl
- 0.2% Igepal (Sigma)
- 10 mM sodium butyrate (Sigma)
- 50 µg/ml PMSF (Sigma)
- 1 µg/ml leupeptin

Nuclei lysis buffer (NLB)

- 50 mM Tris-HCl pH 8.1
- 10 mM EDTA
- 1% SDS
- 10 mM sodium butyrate (Sigma)
- 50 µg/ml PMSF (Sigma)
- 1 µg/ml leupeptin

IP dilution buffer (IPDB)

- 20 mM Tris-HCl pH 8.1
- 150 mM NaCl
- 2 mM EDTA
- 1% Triton X-100 (Sigma)
- 0.01% SDS

- 10 mM sodium butyrate (Sigma)
- 50 µg/ml PMSF (Sigma)
- 1 µg/ml leupeptin

IP wash buffer 1 (IPWB1)

- 20 mM Tris-HCl pH 8.1
- 50 mM NaCl
- 2 mM EDTA
- 1% Triton X-100 (Sigma)
- 0.1% SDS

IP wash buffer 2 (IPWB2)

- 10 mM Tris-HCl pH 8.1
- 250 mM LiCl
- 1 mM EDTA
- 1% Igepal (Sigma)
- 1% deoxycholic acid

IP elution buffer (IPEB)

- 100 mM $NaHCO_3$
- 1% SDS

TE (pH 8.0)

- 10 mM Tris base (pH 8.0)
- 1 mM EDTA

1 × PBS (for ChIP assays)

1 X PBS used for washing the cells in ChIP assay was prepared by dissolving the following salts in 1 litre of HPLC water and the pH was adjusted to 7.4 with HCl.

- 8 g NaCl
- 0.2 g KCl
- 1.44 g Na_2PO_4

- 0.24 g KH_2PO_4

20 × SSC

20 × SSC used in the washing steps in the TECAN hybridisation was prepared by dissolving the following salts in 1 litre of HPLC water.

- 175.3 g NaCl
- 88.2 g tri-sodium citrate

10 × Tris-buffered saline (TBS)

10 × TBS used in western blotting was prepared by dissolving the following salts in 1 litre of deionised water and the pH was adjusted to 7.6 with HCl.

- 24.4 g Tris-HCl
- 80 g NaCl

1 × Tris-buffered saline Tween 20 (TBST)

1 × TBST used in western blotting was prepared by diluting 100 ml 10 × TBS in 1 litre of deionised water and adding 1 ml of Tween 20.

12 × MES stock buffer

12 × MES stock buffer used in Affymetrix GeneChip hybridisations was prepared by dissolving the following salts in 1 litre of molecular biology grade water and filtered through a 0.2 μm filter.

- 64.61 g of MES hydrate
- 193.3 g of MES sodium salt

2× hybridisation buffer (for Affymetrix GeneChips)

- 2 × MES Stock Buffer
- 1 M NaCl
- 20 mM EDTA
- 0.01% Tween-20

Wash buffer A: non-stringent wash buffer (for Affymetrix GeneChips)

- 6 × SSPE
- 0.01% Tween-20

Wash buffer B: stringent wash buffer (for Affymetrix GeneChips)

- 2 × MES stock buffer
- 0.1 M NaCl
- 0.01% Tween-20

2X stain buffer (for Affymetrix GeneChips)

- 2 × MES stock buffer
- 1 M NaCl
- 0.05% Tween-20

2.2 Reagents

Antibodies

- Complete lists of antibodies, with company names and catalogue numbers, used in western blotting and chromatin immunoprecipitation assays are included in Appendix 3A and 3B respectively.
- anti-glycophorin A PE (BD Biosciences)
- biotinylated anti-streptavidin (Vector Laboratories)

siRNAs

- All siRNAs used in the RNAi assays were designed and synthesised by Eurogentec or Ambion. A complete list of siRNAs, manufacturers and sequences is included in Chapter 3.

Primer pairs

- All primer pairs used in the quantitative real time PCR assays were synthesised by Sigma. Complete lists of the sequences of the primer pairs are included in Appendix 1.

Enzymes

- proteinase K (Invitrogen)
- RNase A (ICN Biochemicals)
- RNase-free DNase I (Ambion)
- Klenow fragment (Invitrogen)
- SuperScript^{TM} II RNase H^{-} reverse transcriptase (Invitrogen)

Flourophores

- Cy3 dCTP (GE Healthcare)
- Cy5 dCTP (GE Healthcare)
- anti-glycophorin A PE (BD Biosciences)
- FITC conjugated GATA1 siRNA (Eurogentec)

2.3 Cell lines

Human erythroleukaemic cell line K562 (Lozzio and Lozzio 1975) was a gift from Professor Anthony Green, Department of Haematology, University of Cambridge, UK.

Human erythroleukaemic cell line HEL 92.1.7 (Martin and Papayannpoulou, 1982) was obtained from American Type Culture Collection (ATCC).

2.4 Tissue culture

2.4.1 Culturing and propagation of cell lines

Cell lines K562 and HEL 92.1.7 were cultured in RPMI 1640 (Sigma) supplemented with 10% v/v fetal calf serum (FCS) (Invitrogen) and 100 μg/ml penicillin/streptomycin (Sigma) (named thereafter 'supplemented media'). The cells were maintained at 37^{o}C and 5% CO_2 at a cell density of 0.5×10^6.

Sub-culturing was performed as follows:

1. Cells were counted every two days and spun down at 1200 rpm for 5 minutes.
2. Cells were resuspended in fresh warm supplemented media and maintained at a cell density of 0.5×10^6 in 75 cm^3 tissue culture flasks with vented caps (Corning) with a maximum of 50 ml of cells.
3. For addition sub-culturings, cells were counted and resuspended as above and maintained in 150 cm^3 flask with a maximum of 100 ml of media/cell suspension.

2.4.2 Cryopreservation of cell lines

For cryopreservation of cell lines, cells were counted and spun down at 1200 rpm for 5 minutes and resuspended in freezing media [FCS in 10% v/v dimethylsulphoxide (DMSO, Sigma)] at a density of 4×10^6 /ml. The cells were transferred to polypropylene cryotubes (Nunc) and cooled overnight at -70 ^{o}C at a rate of 1-2 ^{o}C per minute. The cryotubes were transferred to gas phase liquid nitrogen tank at -180 ^{o}C for long term storage.

To defrost frozen cell stock, cells were thawed rapidly in a 37 ^{o}C water bath, washed with 10 ml fresh media once and resuspended in 10 ml fresh supplemented media and maintained in a 25 cm^3 tissue culture flask with vented cap.

2.5 Transfection of siRNA

1. siRNAs for GATA1, SCL, E2A, LDB1, LMO2 and firefly luciferase were designed and synthesised by Ambion or Eurogentec (Chapter 3).
2. Media for K562 cells was changed one day before transfection and cells were maintained at a cell density of 0.5×10^6.
3. 5×10^6 cells were transfected with 2 µl of 100 µM of siRNAs in 100 µl of RPMI 1640 (10% v/v FCS and 100 µg/ml penicillin/streptomycin) using the Nucleofector^{TM} II system (Amaxa Biosystems) with programme T16.
4. Transfected cells were resuspended in 10 ml RPMI 1640 (10% v/v FCS and 100 µg/ml penicillin/streptomycin) in 25 cm^3 flask with vented cap and incubated at 37^{o}C and 5% CO_2. The final concentration of the siRNA was 20 nM.
5. 1×10^7 cells were transfected in two separate transfections (5 million cells in each) for each time point for each siRNA.
6. Cells were harvested at 12 hr, 24 hr, 36 hr and 48 hr after transfection for RNA and protein extraction as described in section 2.6 and 2.7.

2.6 RNA extraction

K562 cells were harvested (after transfection or without transfection) and total cellular RNA was extracted and purified as follows:

2.6.1 Total RNA extraction by TRIZOL

1. 1.5×10^6 of the transfected cells were harvested at 12 hr, 24 hr, 36 hr and 48 hr after siRNA transfection or 3×10^6 untransfected cells were harvested by centrifugation at 1200 rpm for 5 minutes.
2. Cell pellets were resuspended with 1 ml TRIZOL reagent (Invitrogen).
3. 0.2 ml of chloroform (Sigma) was added per 1 ml of TRIZOL reagent used.
4. The samples were mixed by shaking vigorously for 15 seconds and incubated at room temperature for 2-3 minutes.
5. After incubation, the samples were centrifuged at 12000 rcf for 15 minutes at 4°C.

6. The aqueous phases (upper layer) were transferred to new 1.5 ml microfuge tubes and 0.5 ml of isopropanol (Sigma) was added. The samples were mixed by inverting the tube a few times.
7. The samples were incubated at room temperature for 10 minutes and centrifuged at 12000 rcf for 10 minutes at 4°C. The RNA should now be visible as a pellet at the bottom of the tube.
8. The supernatants were removed and the pellets were washed once with 1 ml 75% ethanol (Sigma).
9. The samples were mixed by inverting the tube a few times and centrifuged at 12000 rcf for 5 minutes at 4°C.
10. The supernatants were removed and the pellets were air-dried.
11. The pellets were resuspended in 42 µl RNase-free water (Ambion) by pipetting until the pellets were completely dissolved.

2.6.2 DNase treatment for RNA samples

1. The RNA samples were subsequently treated with 6 units of RNase-free DNase I (Ambion) in a 50 µl reaction.
2. The samples were incubated at 37°C for 30 minutes.
3. The reaction mix was then heat inactivated at 100°C for 5 min and cooled on ice.

2.6.3 Phenol-chloroform extraction and ethanol precipitation

1. The RNA samples prepared as in section 2.6.2 were diluted to 100 µl with RNase-free water.
2. Equal volume of phenol-chloroform-isopropanol (Ambion) (100 µl) was added to the RNA samples.
3. The reaction mix was mixed thoroughly by vortexing and centrifuged at 13500 rpm for 5 minutes.
4. The aqueous phase (upper layer) of the centrifuged samples was transferred to a new 1.5 ml microfuge tube and 10 µl of RNase-free 3M sodium acetate pH5.2 (Ambion) and 250 µl 100% ethanol were added to each tube.
5. The RNAs were precipitated at -20°C for 1 hour.
6. RNAs were precipitated by centrifugation at 13500 rpm for 20 minutes at 4°C.
7. The supernatants were removed and the pellets were washed once with 500 µl 75% ethanol (Sigma).

8. The samples were mixed by inverting the tube a few times and centrifuged at 13500 rpm for 5 minutes at 4°C.
9. The supernatants were removed and the pellets were air-dried.
10. The pellets were resuspended in 30 - 50 µl RNase-free water (Ambion) by pipetting until the pellets were completely dissolved.
11. The total RNA extracted was quantified using Nanodrop ND-1000 spectrophotomic system (Labtech).
12. The RNA quality was assessed by electrophoresis of 500 ng of the sample on a 1% agarose/1 × TBE minigel (made up with RNase-free water).
13. 3 × volumes of 100% ethanol (Sigma) was added to the aqueous sample and the samples were then stored at -70°C.

2.7 Reverse transcription

First strand cDNAs were synthesised using SuperScript^{TM} II RNase H^{-} reverse transcriptase (Invitrogen) as follows.

1. 1 µg of RNA and 1 µl of 50 ng/µl random primers (Invitrogen) were mixed to a volume of 12 µl with RNase-free water.
2. The samples were incubated at 100°C for 1 minute and cooled on ice for 2 minutes.
3. 4 µl of 5X first strand buffer (Invitrogen), 2 µl of 0.1M DTT (Invitrogen) and 1 µl of 10 mM dNTPs (Invitrogen) were added to the samples.
4. 1 µl of SuperScript^{TM} II RNase H^{-} reverse transcriptase (Invitrogen) was added to the samples
5. The reaction mix was mixed gently by flicking and incubated at 42°C for 1 hour.
6. The samples were heat-inactivated at 100°C for 5 minutes and cooled on ice.
7. The resulting cDNAs were diluted to 10 ng/µl by TE pH 8.

2.8 Quantitative real-time PCR

2.8.1 Primer design

1. Primer pairs for all the real-time PCR assays were designed by using the Primer Express software version 2.0 (Applied Biosystems) or Primer 3 (http://frodo.wi.mit.edu/, Rozen and Skaletsky 2000) following the guidelines below:
    - primer length: 18-30 bases

- GC content: 40 – 60 %
- Tm: 64 °C
- dTm: 2°C
- amplicon length: 80 – 150 bp
- intron-spanning where possible

2. Primer pair sequences were checked for specificity by BLAST comparison with the entire human genome and *in silico* PCR (UCSC Genome Browser).
3. Primer specificity was confirmed by dissociation curve analyses in the real-time PCR. Only one peak was observed in all cases.

2.8.2 Quantitative real-time PCR

Quantitative real-time (qRT)–PCR was used to assess (i) the knockdown efficiency in the siRNA assays, (ii) to investigate expression of putative target genes in the siRNA assays and (iii) to confirm enrichment levels in the ChIP assays.

1. The SYBR green PCRs were set up in 96-well optical reaction plate (Applied Biosystems) in a 25 µl reaction in duplicate by mixing the following reagents on ice:

• water	variable
• 2 µM forward and reverse primer mix	5 µl
• 2 × SYBR green PCR mastermix (Applied Biosystems)	12.5 µl
• 10 ng/µl cDNA samples or 10 × diluted ChIP DNA samples	4 µl or 5 µl
Total volume:	25 µl

2. PCR was performed on a 7700 sequence detection system (Applied Biosystems) using the following conditions: 2 min at 50 ^{o}C; 10 min at 95 ^{o}C; 40 cycles of 95 ^{o}C for 15 s and 60 ^{o}C for 1 min.

2.8.3 Data analyses

Ct values were extracted using Sequence Detector 1.7a (Applied Biosystems) with the same threshold and data analyses were performed as follows:

<u>For expression assays (cDNA quantification):</u>

ΔCt = Ct (house-keeping gene) - Ct (gene of interest)

$\Delta\Delta$Ct = ΔCt (luciferase control) - ΔCt (siRNA knockdown)

Fold repression = $(1+\text{PCR yield})^{\Delta\,\Delta CT}$

% of mRNA remained after knockdown = 100/ Fold repression

For ChIP-PCR assays:

$\Delta Ct = Ct\ (input) - Ct\ (ChIP\ sample)$

$Fold\ enrichment = (1+PCR\ yield)^{\Delta CT}$

2.9 Protein extraction

2.9.1 Protein extraction

1. 8.5×10^6 of the transfected cells were harvested at 12, 24, 36 and 48 hour time points after siRNA transfection by centrifugation at 1200 rpm for 5 minutes.
2. The cell pellets were washed once with 10 ml ice-cold PBS followed by centrifugation at 1200 rpm for 5 minutes.
3. Cells were resuspended with 250 µl cell lysis buffer and incubated for 5 minutes on ice.
4. Nuclei were obtained by centrifugation at 11000 rcf for 1 minute at 4 ^oC.
5. Nuclei were washed once with 250 µl cell lysis buffer followed by centrifugation 11000 rcf for 1 minute at 4 ^oC.
6. Nuclei were resuspended in 70 µl of extraction buffer and snap frozen with dry ice in 100% ethanol.
7. Protein samples were stored at -80 ^oC.

2.9.2 Protein quantification

Nuclear proteins were quantified using the Bio-Rad Protein Assay Kit II.

1. 0.8 ml of filtered deionised water was added to 1.5 ml spectrophotometer cuvettes (Bio-Rad).
2. 0, 2, 5, 10, 15 or 20 µl of BSA (1 mg/ml) (Invitrogen) was added to the cuvettes in order to generate a standard curve.
3. 2 µl of nuclear protein extract was added to the cuvettes.
4. 200 µl of Dye Reagent Concentrate (Bio-Rad) was added to each cuvette.
5. The reaction mixtures were mixed well by pipetting.
6. The samples were incubated for 20 minutes at room temperature.
7. The absorbance of each sample at 595 nm was measured using the CECIL spectrophotometer CE2020 (CECIL Instruments).

8. The concentration of the nuclear proteins was calculated in Excel using the BSA standard curve in Excel.

2.10 Western blotting

2.10.1 Sample preparation

Nuclear protein samples were prepared under reducing or non-reducing conditions as follows:

For reducing samples:

- 10, 20 or 30 µg of nuclear proteins variable
- 4 × LDS loading buffer (Invitrogen) 7.5 µl
- sample reducing agent (Invitrogen) 3 µl
- deionised water variable

Total volume: 30 µl

For non-reducing samples:

- 10, 20 or 30 µg of nuclear proteins variable
- 4 × LDS loading Buffer (Invitrogen) 7.5 µl
- deionised water variable

Total volume: 30 µl

The samples were mixed by vortexing and denatured at 100 ^{o}C for 2 minutes.

2.10.2 SDS-PAGE

1. NuPAGE 4-20% Novex Bis-Tris gels (Invitrogen) were washed with distilled water and combs were removed and the wells were washed with 1 × MOPS running buffer (Invitrogen) with a syringe.
2. The gels were assembled in XCell SureLock^{TM} Mini-Cell (Invitrogen).
3. The inner and outer chambers of the Mini-Cell tank were filled with 200 µl and 500 µl of 1 × MOPS Running Buffer respectively (N.B. 500 µl of anti-oxidant (Invitrogen) was added to the inner chambers.)
4. The denatured proteins and 5µl of See Blue Plus Standard (Invitrogen) were loaded into the wells.
5. The proteins were electrophoresed for 1 to 1.5 hours at constant voltage of 200 V and a starting current of 125 mA/ gel at 4 ^{o}C.

2.10.3 Blotting

1. 1 × NuPAGE Transfer Buffer (500ml) was prepared as follows:
   - 20 × NuPAGE Transfer Buffer (Invitrogen) 25 ml
   - methanol 50 ml

- deionised water 425 ml

Total volume: 500 ml

2. The transfer buffer and 1 litre deionised water were kept in the fridge for at least 30 min before use.
3. The gels were dissembled after electrophoresis.
4. Blotting pads, filter papers and gels were equilibrated in cold transfer buffer for 10 seconds.
5. PVDF membranes (Millipore) were equilibrated in 100% methanol for 10 seconds and transferred to the cold transfer buffer.
6. The blot modules were assembled as follows:

For 1 gel	For 2 gels
• top (+) • 2 blotting pads • 2 filter papers • transfer membrane • gel • 2 filter papers • 2 blotting pads • bottom (-)	• top (+) • 2 blotting pads • 2 filter paper • transfer membrane • second gel • 2 filter paper • 1 blotting pad • 2 filter paper • transfer membrane • first gel • 2 filter paper • 2 blotting pads • bottom (-)

7. Any air bubbles in blotting pads and between the gel and membrane were removed.
8. The blot module was clipped together firmly and placed into a transfer tank.
9. The blot module was filled with transfer buffer until the gel/membrane sandwich was covered in transfer buffer.
10. The outer chamber was filled with cold deionised water to the top.
11. The blotting was performed at a constant voltage of 30 V for one gel and 35 V for 2 gels and a starting current of 170 mA/ gel at 4 ^{o}C for 2 hours.

2.10.4 Immunoblotting and detection

1. The membrane was blocked using blocking buffer (5% non-fat dry milk in tris-buffered-saline-tween 20 (TBST) at room temperature for 1 hour.
2. The membrane was then incubated with primary antibodies at the appropriate dilutions (Appendix 3A) in 10 ml blocking buffer at 4 ^{o}C overnight.

3. The membrane was then washed four times with TBST for 1 hour
4. The membrane was incubated with secondary antibodies at the appropriate dilutions (Appendix 3A) in 10 ml blocking buffer at room temperature for 1 hour.
5. The membrane was then washed again four times with TBST for 1 hour.
6. The membrane was incubated with ECL plus (GE Healthcare) for 5 minutes.
7. Signals were developed on ECL Hyperfilm (GE Healthcare).
8. The membrane was stained with 10 ml of water-diluted Dye Reagent Concentrate (Bio-Rad) at a dilution of 1:5 to visualise loading control.

2.11 Flow cytometry analysis for fluorescent oligo transfection

1. Media for K562 cells were changed one day before transfection and cells were maintained at a cell density of 0.5×10^6.
2. 5×10^6 cells were transfected in 100 µl of RPMI 1640 (10% v/v FCS and 100 µg/ml penicillin/streptomycin) using the Nucleofector™ II system (Amaxa Biosystems) with programme T16 with the following conditions:

   <u>0 hour time point:</u>

   - 100 µM FITC-labelled GATA1a siRNA 2 µl
   - 100 µM unlabelled GATA1a siRNA 2 µl

   <u>24 hour time point:</u>

   - 100 µM FITC-labelled GATA1a siRNA 2 µl
   - 100 µM unlabelled GATA1a siRNA 2 µl
3. Transfected cells were resuspended in 10 ml RPMI 1640 (10% v/v FCS and 100 µg/ml penicillin/streptomycin) in 25 cm^3 flask with vented cap and incubated at 37°C and 5% CO_2. The final concentration of the siRNA was 20 nM.
4. Cells were counted and 1×10^6 cells were harvested by centrifugation at 1200 rpm for 5 minutes.
5. Cells were washed with 10 ml PBS followed by centrifugation at 1200 rpm for 5 minutes.
6. Cells were resuspended in 1 ml PBS and filtered with 30um mesh filter (CellTrics®, Partec GmbH).
7. The resuspended cells were kept on ice and covered with aluminium foil.
8. Cells were flow-sorted using the Cytomics FC 500 flow cytometer (Beckman Coulter; Fullerton, CA) equipped with an air-cooled 20mW 488nm Argon laser.

9. Data were analysed with the WinMDI 2.8 software (http://facs.scripps.edu/software.html) and percentage of transfected cells were calculated.

2.12 Cell morphology studies

1. Media for K562 cells and HEL 92.1.7 cells was changed one day before transfection and cells were maintained at a cell density of 0.5×10^6.
2. 5×10^6 cells were either not transfected, sham transfected (placing the cuvette in the electroporator without pressing the button) or transfected with siRNAs or water in 100 µl of RPMI 1640 (10% v/v FCS and 100 µg/ml penicillin/streptomycin) using the Nucleofector™ II system (Amaxa Biosystems) with programme T16 as described in section 2.5.
3. The following transfections were set up and cells were harvested at corresponding time points:

Cell line	Transfection	Time points for harvest			
		0hrs	1hr	24hrs	48hrs
K562	nil	√			
K562	water		√	√	√
K562	LUC siRNA		√	√	√
K562	GATA1 siRNA		√	√	√
K562	E2A siRNA		√	√	√

4. Cells were counted and 1×10^6 cells were harvested by centrifugation at 1200 rpm for 5 minutes.
5. Cells were washed with 10 ml PBS followed by centrifugation at 1200 rpm for 5 minutes and cells were resuspended in 10 ml PBS and placed on ice.
6. 300 µl of the resuspended cells were placed into Cytofunnel disposable sample chambers (Shannon) on top of a microscope slide.
7. The samples were spun at 200 rpm for 5 minutes in Cytospin 3 (Shannon) to precipitate the cells on the slide.
8. The slides were allowed to air dry.
9. The slides were stained in Stain Quick-Staining Kit (Lamb).
10. The slides were washed with tap water to remove excessive stain and air dried at room temperature.
11. The dried slides were mounted in Depex Polystyrene (DPX) (BDH) and round cover slips (Shannon) and observed under a light microscope.

12. For a blinded test, the slides were blindly labeled and 100 cells were randomly scored with the following cell morphologies: cells with small blebs, cells with large blebs, cells with 2 nuclei, cells with > 2 nuclei and others.

2.13 Growth arrest studies

The growth patterns of K562 untransfected or transfected cells under different conditions were studied.

1. Media for K562 cells was changed one day before transfection and cells were maintained at a cell density of 0.5×10^6.
2. 5×10^6 cells were transfected in 100 µl of RPMI 1640 (10% v/v FCS and 100 µg/ml penicillin/streptomycin) using the Nucleofector™ II system (Amaxa Biosystems) with programme T16 as described in section 2.5.
3. The following transfections were set up and cells were harvested at corresponding time points:

Transfection	**Time points for harvest**			
	0 hr	**24 hr**	**48 hr**	**72 hr**
nil	√	√	√	√
Water	√	√	√	√
LUC siRNA	√	√	√	√
GATA1 siRNA	√	√	√	√

4. An aliquot of cells was taken at corresponding time points and stained with equal volume of Trypan Blue (Sigma) for 5 minutes at room temperature.
5. No. of viable cells were scored in two independent aliquots.

2.14 Glycophorin A expression detection in K562 and HEL 92.1.7

1. K562 and HEL 92.1.7 cells were cultured to confluence as described in section 2.4.1.
2. Cells were counted and 1×10^6 cells were harvested by centrifugation at 1200 rpm for 5 minutes.
3. Cells were washed twice with 10 ml PBS followed by centrifugation at 1200 rpm for 5 minutes.
4. Cells were resuspended in 100 µl of PBS and placed on ice.
5. Cells were stained by adding 20 µl of anti-glycophorin A PE (BD Biosciences) to the resuspended cells and incubating for 1 hour at 4 ^{o}C in the dark.
6. Stained cells were washed once with 10 ml PBS.

7. Cells were resuspended in 500 µl of PBS and analyzed by Cytomics FC 500 flow cytometer (Beckman Coulter; Fullerton, CA) equipped with an air-cooled 20mW 488nm Argon laser. Fluorescence emitted from PE was collected using 575BP (FL2) bandpass filter.
8. Alternatively, stained cells were resuspended in 500 µl of PBS with 3% formaldehyde for short term storage in the dark until ready for flow-sorting.
9. Percentage of cells expressing glycophorin A was calculated using WinMDI 2.8 software (http://facs.scripps.edu/software.html).

2.15 Differentiation study of K562

1. K562 cells were cultured to confluence as described in section 2.4.1.
2. Cells were counted and 1×10^6 cells were collected by centrifugation at 1200 rpm for 5 minutes and fixed in 500 µl of PBS with 3% formaldehyde.
3. Cells were counted and 10×10^6 cells were collected by centrifugation at 1200 rpm for 5 minutes.
4. Cells were resuspended with 20 ml of fresh supplemented media.
5. 40 µl of 25 mM hemin was added to the resuspended cells.
6. 1×10^6 cells were collected by centrifugation at 1200 rpm for 5 minutes and fixed in 500 µl of PBS with 3% formaldehyde at 24, 48, 72 and 96 hours.
7. Glycophorin A expression was quantified in all samples as described in section 2.14.

2.16 Affymetrix GeneChip expression analysis

Total RNAs extracted from siRNA-transfected K562 cells were subject to hybridisation to Affymetrix GeneChip Expression array. The RNAs were processed with the One-Cycle Target Labelling and Control Reagents provided by Affymetrix as follows:

2.16.1 Eukaryotic target preparation

Preparation of poly-A RNA spike controls

The poly-A RNA dilutions for 5 µg of total RNA were prepared with the Eukaryotic Poly-A RNA Control Kit as follows:

1. 2 µl of the Poly-A control stock was added to 38 µl of Poly-A control dilution buffer for the first dilution (1:20).

2. The samples were mixed thoroughly and spun down to collect the liquid at the bottom of the tube.
3. 2 µl of the first dilution was added to 98 µl of Poly-A control dilution buffer to prepare the second dilution (1:50).
4. The samples were mixed thoroughly and spun down to collect the liquid at the bottom of the tube.
5. 2 µl of the second dilution was added to 18 µl of Poly-A control dilution buffer to prepare the third dilution (1:10).
6. The samples were mixed thoroughly and spun down to collect the liquid at the bottom of the tube.
7. 2 µl of the third dilution was added to 5 µg of sample total RNA.

Double-stranded cDNA synthesis

RNA samples were subject to one-cycle cDNA synthesis with the One-Cycle cDNA synthesis kit.

1. RNA sample, diluted poly-A RNA controls, and T7-Oligo(dT) primer were mixed as follows:
   - RNA sample (5 µg) variable
   - diluted poly-A RNA controls 2 µl
   - T7 Oligo(dT) primer 50 µM 2 µl
   - RNase-free water variable

   Total volume: 12 µl
2. The tube was gently flicked a few times to mix, and then centrifuged briefly (~5 seconds) to collect the reaction at the bottom of the tube.
3. The reaction was incubated for 10 minutes at 70°C.
4. The sample was cooled at 4°C for at least 2 minutes and centrifuged briefly (~5 seconds) to collect the sample at the bottom of the tube.
5. In a separate tube, the first-strand master mix was assembled as follows:
   - 5 × 1st strand reaction mix 4 µl
   - DTT 0.1 M 2 µl
   - dNTP 10 mM 1 µl

   Total volume: 7 µl
6. The tube was gently flicked a few times to mix, and then centrifuged briefly (~5 seconds) to collect the reaction at the bottom of the tube.
7. 7 µl of first-strand master mix was transferred to each RNA/T7-Oligo(dT) primer mix for a final volume of 19 µl.

8. The tube was gently flicked a few times to mix, and then centrifuged briefly (~5 seconds) to collect the reaction at the bottom of the tube.
9. The reaction mix was incubated for 2 minutes at 42°C.
10. 1 µl of SuperScript II was added to each RNA sample for a final volume of 20 µl.
11. The tube was gently flicked a few times to mix, and then centrifuged briefly (~5 seconds) to collect the reaction at the bottom of the tube.
12. The samples were incubated for 1 hour at 42°C; then cooled for at least 2 minutes at 4°C.
13. In a separate tube, second-strand master mix was assembled as follows:

- RNase-free water 91 µl
- 5 × 2nd strand reaction mix 30 µl
- dNTP 10 mM 3 µl
- *E. coli* DNA ligase 1 µl
- *E. coli* DNA Polymerase 1 µl
- RNase H 1 µl

Total volume: 130 µl

14. The tube was gently flicked a few times to mix, and then centrifuged briefly (~5 seconds) to collect the reaction at the bottom of the tube.
15. 130 µl of second-strand master mix was added to each first-strand synthesis sample for a total volume of 150 µl.
16. The tube was gently flicked a few times to mix, and then centrifuged briefly (~5 seconds) to collect the reaction at the bottom of the tube.
17. The reaction mix was incubated for 2 hours at 16°C.
18. 2 µl of T4 DNA Polymerase was added to each sample and incubated for 5 minutes at 16°C.
19. After incubation with T4 DNA Polymerase, 10 µl of 0.5 M EDTA was added to the samples.

Cleanup of double-stranded cDNA

The cDNA samples were cleaned up using the Sample Cleanup Module as follows:

1. 600 µl of cDNA binding buffer was added to the double-stranded cDNA synthesis preparation and the samples were mixed by vortexing for 3 seconds.
2. 500 µl of the sample was applied to the cDNA Cleanup Spin Column sitting in a 2 ml collection tube, and the column was centrifuged for 1 minute at 10000 rpm.
3. The flow-through was discarded.
4. The spin column was reloaded with the remaining mixture and centrifuged as above.

5. The flow-through and collection tube were discarded.
6. The spin column was transferred into a new 2 ml collection tube.
7. 750 µl of the cDNA wash buffer was added onto the spin column.
8. The column was centrifuged for 1 minute at 10000 rpm. The flow-through was discarded.
9. The cap of the spin column was opened and centrifuged for 5 minutes at 13200 rpm. The flow-through and collection tube was discarded.
10. The spin column was transferred to a 1.5 ml collection tube, and 14 µl of cDNA elution buffer was added directly onto the spin column membrane. The resulting spin column was incubated for 1 minute at room temperature and centrifuge for 1 minute at 13200 rpm to elute. The recovered volume of cDNA was 12 µl.

Synthesis of biotin-labeled cRNA

Biotin-labeled cRNA was synthesised with the GeneChip IVT Labeling Kit as follows:

1. The reaction mixture was assembled at room temperature as follows:
   - template cDNA 12 µl
   - RNase-free Water 8 µl
   - 10 × IVT labeling buffer 4 µl
   - IVT labeling NTP mix 12 µl
   - IVT labeling enzyme mix 4 µl

   Total volume: 40 µl
2. The reagents were carefully mixed and the mixture was collected at the bottom of the tube by brief (~5 seconds) centrifugation.
3. The reaction was incubated at 37°C for 16 hours in an oven.

Cleanup and quantification of biotin-labeled cRNA

The cRNA samples were cleaned up using the Sample Cleanup Module as follows:

1. 60 µl of RNase-free water was added to the IVT reaction and the resulting sample was mixed by vortexing for 3 seconds.
2. 350 µl IVT cRNA binding buffer was added to the sample and the reaction mix was mixed by vortexing for 3 seconds.
3. 250 µl 100% ethanol was added to the lysate, and the reaction mix was mixed well by pipetting.
4. 700 µl of the sample was added to the IVT cRNA Cleanup Spin Column sitting in a 2 ml collection tube.

5. The column was centrifuged for 15 seconds at 10000 rpm. The flow-through and collection tube were discarded.
6. The spin column was transferred into a new 2 ml collection tube.
7. 500 µl of IVT cRNA wash buffer was added onto the spin column.
8. The column was centrifuged for 15 seconds at 10000 rpm. The flow-through was discarded.
9. 500 µl of 80% ethanol was added onto the spin column and the column was centrifuged for 15 seconds at 10000 rpm. The flow-through was discarded.
10. The cap of the spin column was opened and the column was centrifuged for 5 minutes at 13200 rpm. The flow-through and collection tube were discarded.
11. The spin column was transferred into a new 1.5 ml collection tube, and 11 µl of RNase-free water was added directly onto the spin column membrane. The column was centrifuged for 1 minute at 13200 rpm to elute.
12. 10 µl of RNase-free water was added directly onto the spin column membrane. The column was centrifuged for 1 minute at 13200 rpm to elute to the same tube.
13. An aliquot of the cRNA was diluted by 1:100 fold and quantified using Nanodrop ND-1000 spectrophotomic system (Labtech).
14. An adjusted cRNA yield was calculated to reflect carryover of unlabeled total RNA with the formula below:

    Adjusted cRNA yield = RNAm - (total RNAi) (y)

    RNAm = amount of cRNA measured after IVT (µg)

    Total RNAi = starting amount of total RNA (µg)

    y = fraction of cDNA reaction used in IVT

cRNA fragmentation

1. cRNAs were fragmented to 35 to 200 base fragments. The reaction mixture was assembled as follows:

    - cRNA (adjusted concentration) 20 µg
    - 5× fragmentation buffer 8 µl
    - RNase-free water variable

    Total volume: 40 µl

2. The reaction mix was incubated at 94°C for 35 minutes. The samples were placed on ice following the incubation.

2.16.2 Target hybridisation

1. A hybridisation cocktail was set up for each array by mixing the following reagents:

• fragmented cRNA (15 µg)	30 µl
• control oligonucleotide B2 (3 nM)	5 µl
• 20 × eukaryotic hybridisation controls (bioB, bioC, bioD, cre)	15 µl
• Herring sperm DNA (10 mg/ml) (Sigma)	3 µl
• Acetylated BSA (50 mg/ml) (Invitrogen)	3 µl
• 2 × hybridisation buffer	150 µl
• DMSO (Sigma)	30 µl
• RNase-free water	64 µl
Total volume:	300 µl

   N.B. The 20 × eukaryotic hybridisation controls and Herring Sperm DNA were heated to 65°C for 5 minutes to resuspend the cRNA or DNA before dispensing into aliquots.

2. The Affymetrix GeneChip® Human Genome U133 plus 2.0 probe array was equilibrated to room temperature immediately before use.

3. The hybridisation cocktail was heated to 99°C for 5 minutes in a heat block.

4. The array was wetted by filling it through one of the septa with 200 µl of 1 × hybridisation buffer using a micropipettor and appropriate tips.

5. The probe array filled with 1 × hybridisation buffer was heated at 45°C for 10 minutes with rotation of 60 rpm in the hybridisation oven.

6. The hybridisation cocktail which has been heated at 99°C was transferred to a 45°C for 5 minutes.

7. The hybridisation cocktail was spun at maximum speed in a microcentrifuge for 5 minutes to remove any insoluble material from the hybridisation mixture.

8. The buffer solution was removed from the probe array cartridge and the cartridge was filled with 200 µl of the clarified hybridisation cocktail.

9. The probe array was placed in the rotisserie box in the 45°C oven with rotation at 60 rpm for hybridisation for 16 hours.

2.16.3 Fluidics station setup

1. To wash, stain, and scan a probe array in the Fluidics station 450, an experiment was first registered in the GeneChip Operating Software (GCOS) with the following information:
   - Experiment Name
   - Probe Array Type
   - Array barcode
   - Sample Name

- Sample Type
- Project

2. The fluidics station was first primed. In the GCOS, Run → Fluidics → Protocol → Prime_450 were selcted.

3. The intake buffer reservoir A was changed to non-stringent wash buffer and intake buffer reservoir B was changed to stringent wash buffer.

4. All modules were selected to begin priming.

2.16.4 Probe array washing, staining and scanning

1. After 16 hours of hybridisation, the hybridisation cocktail was removed from the probe array and the probe array was filled completely with the 200 µl of non-stringent wash buffer (wash buffer A).

2. The staining reagents were prepared as follows:

   Streptavidin phycoerythrin (SAPE) stain solution (prepared immediately before use)

• 2 × MES stain buffer	600 µl
• 50 mg/ml acetylated BSA (Invitrogen)	48 µl
• 1 mg/ml Streptavidin Phycoerythrin (SAPE) (Molecular Probes)	12 µl
• RNase-free water	540 µl
Total volume:	1200 µl

   The solutions were mixed and aliquoted into two light-protected tubes of 600 µl each and used for stains 1 and 3.

   Antibody solution

• 2 × MES stain buffer	600 µl
• 50 mg/ml acetylated BSA (Invitrogen)	24 µl
• 10 mg/ml normal goat IgG (Sigma)	6 µl
• 0.5 mg/ml biotinylated anti-streptavidin (Vector Laboratories)	3.6 µl
• RNase-free water	266.4 µl
Total volume:	600 µl

3. The probe arrays were then washed with wash buffers A and B and stained in the Fluidic Station with protocol EukGE_WS2v5_450.

4. The resulting arrays were scanned in GeneChip® Scanner 3000 with AutoLoader.

2.16.5 Data analysis

1. The scanned arrays were analysed with two separate softwares: the Bioconductor package and the GeneSpring GX 7.3.1.

2. The quality of the probe arrays was accessed by the BioC Affy package of the Bioconductor. The overall perfect match signal intensity, 3' to 5' ration of housekeeping control genes,

uniformity of hybridisation and array-array correlation of signal intensity were calculated for each of the array. Arrays which differ significantly from others were discarded from further analyses. Criteria used for quality control were discussed in Chapter 4.

3. Signal intensities of all array elements were calculated using Robust Multichip Average (RMA) method in GeneSpring GX 7.3.1.
4. An independent experiment was created for each of the TFs under study in GeneSpring. For example, for GATA1, an experiment with all the biological replicates of luciferase siRNA transfected K562 and GATA1 siRNA transfected K562 was created.
5. Three levels of normalisation steps were performed for all arrays in the experiments:
   - Data Transformation: Values below 0.01 were set to 0.01.
   - Per Chip: Each measurement was divided by the 50^{th} percentile of all measurements in that sample.
   - Per Gene: Each gene was divided by the median of its measurements in all samples.
6. The normalised intensity values were extracted from GeneSpring as an Excel spreadsheet. TF/Luc ratios of normalised intensity were calculated and normalised to the median.
7. Two methods were used for the statistical analyses: the average method and the Venn method. In the average method, an average of the intensity values was taken in the three biological replicates. Ratios of these average values in the TF against the Luciferase negative control were calculated. Standard deviations of the median normalised ratios were calculated. Genes that are 2 standard deviations above or below the mean were selected to be the repressed or activated gene lists. In the Venn method, each biological replicate was treated independently, TF/Control ratios were calculated and activated or repressed genes were selected as in the average method. The activated or repressed genes were compared in each biological replicate and the overlapping genes were chosen as putative targets.
8. Gene Ontology classifications were done using GO Term Finder (http://go.princeton.edu/cgi-bin/GOTermFinder) (Boyle et al., 2004).

2.17 Chromatin immunoprecipitation on microarrays (ChIP-on-chip)

Chromatin immunoprecipitations were performed with wild type or transfected K562 and HEL cells as follows:

Wild type cells:

1×10^8 cells were cultured as mentioned above and media were changed one day before chromatin extraction. A 1×10^6 aliquot of cells was taken for flow-sorting to determine the proportion of dividing cells in the population (Cytomation MoFlo High Performance Cell Sorter, Dako Cytomation). Aliquot of cells was washed with 5 ml of PBS and fixed with 5 ml of 70% ethanol.

Transfected cells:

1. K562 cells were cultured as mentioned in section 2.4.
2. 60×10^6 of cells were transfected with siRNAs (5 million cells in each transfection) according to section 2.5.
3. Two transfections (10 million cells) were pooled and resuspended in 10 ml supplemented media and incubated in 25 cm^3 vented flasks for 24 hours. Six flasks were incubated in total.
4. A 1×10^6 aliquot of cells was taken for flow-sorting (Cytomation MoFlo High Performance Cell Sorter, Dako Cytomation).
5. Aliquot of cells was washed with 5 ml of PBS and fixed with 5 ml of 70% ethanol.
6. 2×10^6 cells were taken for RNA extraction while 7×10^6 cells were taken for protein extraction.
7. The remaining transfected cells (50×10^6) were harvested for chromatin extraction.

2.17.1 Chromatin preparation

1. Wild type or transfected cells were collected by centrifuging at 1200 rpm for 5 minutes, washed once with serum-free RPMI and resuspended in 50 ml serum-free RPMI in a glass flask.
2. DNA-protein interactions were cross-linked by adding 1.35 ml of 37% formaldehyde solution drop-by-drop (BDH AnalaR) (final concentration 1.0%) to the cells. Cross-linking was performed at room temperature with gentle agitation for 10 minutes and stopped by adding 3.15 ml of 2 M glycine (final concentration of 0.125 M) and incubated for 5 minutes at room temperature with gentle shaking.
3. The cells were transferred to a 50 ml Falcon tube on ice and centrifuged at 1200 rpm for 6 minutes at 4ºC. The pellet was washed with 10 ml ice-cold phosphate buffered saline (PBS) and centrifuged again at 2000 rpm for 5 minutes at 4ºC.

4. The cell pellet was resuspended in 3 ml of cell lysis buffer (CLB) and incubated for 10 minutes on ice. The nuclei were collected by centrifuging at 2500 rpm for 5 minutes at 4ºC.
5. Nuclei were lysed by resuspending in 1.2 ml of nuclei lysis buffer (NLB) and incubating on ice for 10 minutes. 0.72 ml of IP dilution buffer (IPDB) was then added to the nuclei and the content was transferred to a 5 ml Falcon tube.
6. The sample was sonicated using the Sanyo/MES Soniprep sonicator, with the settings as follows:

   Amplitude: 14 microns

   Number of bursts: 8

   Length of bursts: 30 seconds

   The sample was cooled for 1 minute in an ice/ethanol bath between each pulse. The DNA was sheared to approximately 300-1000 bp fragments.
7. The sheared chromatin was centrifuged at 13000 rpm for 5 minutes at 4ºC. The supernatant was transferred to a 15 ml Falcon and 4.1 ml of IPDB was added to each tube to bring the ratio of NLB:IPDB to 1:4. The chromatin was then snap frozen with liquid nitrogen and stored at -70 ºC.

2.17.2 Immunoprecipitation

1. Chromatin was precleared by adding 100 µl of normal rabbit IgG (Upstate Biotechnology) and incubated for 2 hours at 4ºC on a rotating wheel.
2. 200 µl of the homogeneous protein G-agarose suspension (Roche) was added and the chromatin was incubated for 5 hours at 4ºC on a rotating wheel.
3. The beads were centrifuged at 3000 rpm for 2 minutes at 4ºC. The supernatant was used to set up the following conditions in 2 ml tubes:
   - Control IgG: 0.675 ml chromatin + 0.675 ml $IPDB^{mod}$ (NLB+IPDB at a ratio of 1:4) + 10 µg IgG
   - Antibody under study: 0.675 ml chromatin + 0.675 ml $IPDB^{mod}$ + 10 µg of antibodies against the transcription factors under study
4. The samples were incubated overnight at 4ºC with rotation.
5. 270 µl of the chromatin was used to set up an input control and stored at -20ºC

6. The samples were centrifuged at 13000 rpm for 5 minutes at 4ºC. The supernatants were transferred to new 2 ml microcentrifuge tubes and 50 µl of the homogenous protein G-agarose suspension (25 µl bed volume) was added to each sample. The samples were incubated for 6 hours at 4ºC with rotation.
7. The protein G-agarose beads (Roche) were centrifuged at 13000 rpm for 20 seconds at 4ºC. The pellets were washed twice with 750 µl of IP wash buffer 1 (IPWB1), once with 750 µl of IP wash buffer 2 (IPWB2) and twice with 750 µl of TE pH8.0. For each wash, the samples were mixed briefly by vortexing and centrifuged at 7500 rpm for 2 minutes at 4ºC.
8. The DNA-protein-antibody complexes were eluted from the beads by adding 225 µl of IP elution buffer (IPEB) at room temperature followed by centrifugation at 7500 rpm for 2 minutes. This step was repeated and both elutions were combined in the same tube.

2.17.3 Reversal of crosslinking and DNA extraction

1. 0.2 µl of RNase A (ICN Biochemicals, 10 mg/ml) and 27 µl of 5 M NaCl (final concentration of 0.3 M) were added to each sample. 0.1 µl of RNase A (10 mg/ml stock) and 16.2 µl of 5 M NaCl were added to the input sample. The samples were incubated at 65ºC for 6 hours.
2. 9 µl of proteinase K (Invitrogen, 10 mg/ml) was added to each sample and the samples were incubated at 45ºC for overnight.
3. 2 µl of yeast tRNA (Invitrogen, 5 mg/ml) was added to each sample immediately before adding 500 µl of phenol/chloroform. The samples were mixed well by vortexing and centrifuged at 13200 rpm for 5 minutes at room temperature. The aqueous layer was transferred to new 2 ml microcentrifuge tubes. This step was repeated once with 500 µl of chloroform.
4. 5 µg of glycogen (Roche, 5 mg/ml), 1 µl of yeast tRNA (5 mg/ml stock) and 50 µl of 3 M sodium acetate pH 5.2 were added to each sample. The samples were mixed well and 1.25 ml of 100% ethanol was added. The DNAs were precipitated at -70ºC for 1 hour.
5. The samples were centrifuged at 13200 rpm for 20 minutes at 4ºC. The pellets were washed with 500 µl of ice-cold 70% ethanol.
6. The pellets were allowed to air-dry for 10 minutes and resuspended in 100 µl of HPLC water for the input and 50 µl of HPLC water for the other samples.

2.17.4 Labelling of ChIP DNA and input DNA with cyanine dyes

1. 20 µl of ChIP DNA was mixed with 60 µl of 2.5 × Random Primers Solution (BioPrime Labelling Kit, Invitrogen) and 50.5 µl of HPLC water.

2. 2 µl of input DNA for wild type cells (or 4 µl of input DNA for transfected cells) was mixed with 60 µl of 2.5 × Random Primers Solution and 50.5 µl of HPLC water.

3. ChIP and input DNAs were denatured in a heat block for 10 min at 100°C, and immediately cooled on ice.

4. The following reagents were added to the samples on ice and the contents were mixed thoroughly:

   - dNTP mix 15 µl
   - Cy3 (for ChIP samples) or Cy5 (for input samples) labelled dCTP (1 mM) (GE Healthcare) 1.5 µl
   - Klenow Fragment (BioPrime Labeling Kit, Invitrogen) 3 µl

5. The samples were incubated at 37°C overnight in darkness and the labelling reactions were stopped by adding 15 µl of stop buffer (BioPrime Labeling Kit, Invitrogen).

6. Unlabelled nucleotides were removed from DNA labelling reactions with G-50 microspin columns (GE Healthcare). Three columns (each column has only 50-60 µl maximum capacity) were required for each sample.

7. The resins in the G50 columns were resuspended by gentle vortexing. The cap was loosened by one-quarter turn and the bottom closure was snapped off.

8. The columns were placed in a 1.5 ml screw-cap microcentrifuge tube for support and centrifuged at 4000 rpm for 1 minute. The resins were washed once with 50 µl HPLC water and centrifuged at 4000 rpm for 1 minute.

9. The columns were placed in a new 1.5 ml tube and 50 µl of the labelling reactions was applied to the centre of the angled surface of the compacted resin bed of each of the columns. The columns were spun at 4000 rpm for 2 minutes. The flowthrough samples were retained and combined.

2.17.5 Hybridisation of the human transcription factor promoter array

1. Hybridisation DNA mixtures were prepared for precipitation as follows:

   - ChIP Cy3 labelled DNA ~180 µl
   - Input Cy5 labelled DNA ~180 µl
   - Human Cot1 DNA (Invitrogen) 135 µl
   - 3 M NaAc pH 5.2 55 µl
   - 100% EtOH (cold) 1200 µl

2. All the tubes were mixed gently, covered with aluminium foil and precipitated at -70°C for 60 minutes.

3. The precipitated DNAs were centrifuged for 15 minutes at 13000 rpm at 4 ^{o}C. The pellets were washed with 500 µl 80% EtOH, and centrifuged at 13000 rpm for 5 minutes. Supernatants were removed and tubes were re-spun at 13000 rpm for 1 minute. The pellets were air-dried
4. The DNA pellets were resuspended in 130 ųl of hybridisation buffer (2 × SSC, 50% deionised formamide, 10 mM Tris-HCl pH 7.4, 5% dextran sulphate, 0.1% Tween 20) and 3 ųl of yeast tRNA was added to each sample
5. The hybridisation DNAs were denatured for 10 minutes at 100°C and then immediately quenched on ice.
6. The samples were pulse spun. The hybridisation DNAs were incubated at 37°C for 60 minutes in the dark.
7. The TECAN automatic hybridisation/wash station was prepared, by placing the human transcription factor promoter array slides into the appropriate clean chambers, priming the wash buffer pumps and loading the appropriate hybridisation/ washing program protocol.
8. 110-120 ųl of hybridisation buffer was injected into a TECAN slide chamber containing the promoter array avoiding air bubbles.
9. The slides were allowed to prehybridise for 60 minutes at 37°C on a medium agitation setting, after which time the slides are automatically washed and dried in preparation for injection of the labelled hybridisation mixture.
10. The labelled hybridisation DNAs were pulse spun and 110-120 ųl was injected into a TECAN slide chamber containing the promoter array avoiding air bubbles.
11. The slides were allowed to hybridise for 45 hours at 37°C.

2.17.6 Slide washing, scanning and data analyses

1. The slides were washed in the TECAN station as follows, followed by drying with nitrogen gas:

- Ten washes in PBS with 0.05% Tween 20 at 37°C, each last for 1 minute with additional 30 seconds soak time for each wash
- Five washes in 0.1X SSC at 52°C, each last for 1 minute with additional 2 minute soak time for each wash
- 10 washes in PBS with 0.05% Tween 20 at 23°C, each last for 1 minute with additional 30 seconds soak time for each wash
- HPLC grade water at 23°C for 1 minute and 30 seconds

2. The Cy3 and Cy5 images were scanned with ScanArray 4000 XL scanner (Perkin Elmer) at 5 μm resolutions using a laser power of 100% and a photo multiplier tube (PMT) value of between 80%-85%.

3. Fluorescent intensities of each spot on the array were quantitated using the ScanArray Express software (Perkin Elmer) using the adaptive circle quantitation method and the total normalisation method. The spots representing the array elements were located automatically by the software and the mean signal intensity values against background were calculated for each channel. The mean ratios of the Cy5/Cy3 channels were reported in the resulting Excel datasheet.

4. Statistical analyses of the ChIP-on-chip data were performed in Microsoft Excel. Quality control for the hybridisation of the arrays was carried out by investigating the average signal intensity of the array and the signal/ noise ratios. Arrays with significantly lower signal intensity and signal/ noise ratios are discarded from further analyses.

5. All the 'unfound' spots on the array were not included in the statistical analyses.

6. Mean ratios, standard deviations (SDs) and coefficients of variation (CVs) were calculated for the two replicate spots representing each array element. The mean ratios were normalised against the median values of all the mean ratios.

7. As the positive control of ChIP, enrichments of elements on the SCL tiling array (included on the TF promoter array) were visualised by plotting the mean ratios of all array elements along the y-axis and the respective genomic positions along the x-axis. High quality ChIP should have significant fold enrichments in the +51 enhancer region of the SCL locus. Any ChIPs with low enrichments in this region were discarded from further analyses.

8. For the statistical analyses of the promoter elements, two methods were used: the average method (method B) and the Venn method (method A). In the average method, an average of the mean ratios of promoters was taken in the three biological replicates. These average values in the TF ChIPs were normalised with their corresponding negative control IgG ChIPs. Promoter elements which were 2 standard deviations above the mean were selected to be the enriched promoters. In the Venn method, each biological replicate was treated independently, IgG normalised and enriched promoters were selected as in the average method. The enriched promoters were compared in each biological replicate and the promoters enriched from all three biological replicates (overlapping in the Venn diagram) were chosen as putative targets.

2.18 Sequence analysis of promoters

2.18.1 Motif discovery for putative targets

Promoters of putative target genes co-regulated by SCL, GATA1 and E2A in Affymetrix expression analyses were selected for NestedMICA analyses (Down and Hubbard, 2005) for common regulatory motifs.

2.18.2 Conserved transcription factor binding sites identification

1. A 4 kb window (3 kb upstream and 1 kb downstream) around the transcription start sites of enriched promoters identified in ChIP-chip analyses was taken from Ensembl (http://www.ensembl.org/index.html).
2. TF binding sites were identified using TFSEARCH (http://www.cbrc.jp/research/db/TFSEARCH.html) and TESS (http://www.cbil.upenn.edu/cgi-bin/tess/tess).
3. Local comparative genomic sequence alignments of the 3 kb window in across various species were obtained from UCSC Genome Browser (http://genome.ucsc.edu/).
4. Conserved TF binding sites were identified and ChIP-qPCR confirmation of TF binding was performed around these sites.

2.19 Transcriptional network generation

Network diagrams combining ChIP-on-chip and expression data were generated in BioTapestry software (http://www.biotapestry.org/) (Longabaugh et al., 2005).

Chapter 3
Developing working siRNAs for members of the SCL erythroid complex in K562 cells

3.1 Introduction

3.1.1 The SCL erythroid complex and its downstream regulation

The SCL erythroid complex was first described in 1997 and contains at least 5 members (SCL, GATA1, E2A, LDB1 and LMO2) (Wadman et al., 1997). This complex recognises the consensus E-box and GATA motifs separated by approximately 9 nucleotides. SCL and E2A dimerise with each other and bind to the E-box motif while GATA1 binds to the GATA site. LDB1 and LMO2 act as bridges for the SCL/E2A heterodimer and GATA1 protein. Whereas the downstream regulation by some of the members of this complex has been widely studied, the regulation by SCL and its interacting partners in the SCL erythroid complex is poorly defined. Only three genes (glycophorin A, c-kit and α-globin) have been shown to be directly regulated by this complex in human erythroid cells (discussed in Chapter 1, Section 1.4.2.1 F). Therefore; the focus of this thesis is to further delineate and characterise additional transcriptional targets of the SCL erythroid complex in the erythroid lineage.

In previous studies, overexpression or targeted gene knockdown or knockout have been used to study the downstream targets of members of the SCL erythroid complex. Ectopic expression of SCL in transgenic mice has been shown to correlate with increase in c-kit expression (Lecuyer et al., 2002). GPA expression was shown to be dependent on the levels following ectopic expression of SCL in the human TF-1 cell line and in primary cells (Lahlil et al., 2004). siRNAs targeting SCL was used to study the requirement of SCL in endothelial differentiation and angiogenesis (Lazrak et al., 2004). The same RNAi sequence targeting SCL was used in a shRNA-expressing construct to study the regulation of some putative direct target genes of SCL in a T-ALL cell line Jurkat (Palomero et al., 2006). Global gene expression patterns were compared before and after the induction of GATA1 expression in the GATA1-null erythroblast cell line G1E-ER4 (Rylski et al., 2003; Welch et al., 2004) and between wild type and GATA1-deficient murine megakaryocytes (Muntean and Crispino, 2005). Ectopic expression of GATA1 in a non-erythroid cell line, U937, was shown to stimulate expression of its direct target GFI-1B (Huang et al., 2004). Conversely, siRNA knockdown of GATA1 was used to study its requirement for the auto-regulation of GFI-1B in K562 cells (Huang et al., 2005).

3.1.2 The RNA interference system

RNA interference (RNAi) is the endogenous pathway of suppression of gene expression at the post-transcriptional level. The use of RNA interference to silence transcription of genes is a powerful way to identify putative downstream target genes of transcription factors. Typically, the gene or transcription factor of interest is silenced by RNAi and the downstream effect is studied. Follow-up assays to study such downstream effects include conventional expression assays such as northern blotting and quantitative PCR and high-throughput genome-wide studies such as expression microarrays and sequencing. The type of assays chosen depends on what kind of information the researcher wants to generate.

3.1.2.1 Comparison between RNAi and traditional knockouts

Many different RNAi approaches have been used to silence genes in mammalian systems (discussed in Chapter 1, Section 1.3.1.2). Regardless of the approach of the RNAi trigger, RNAi has a number of advantages over traditional knockouts. Firstly, traditional knockouts usually require a number of rather complicated cloning steps for mammalian cells. In addition, stable integration and selection are also needed for the generation of knockouts. Secondly, complete elimination of a transcription factor may lead to lethality which makes subsequent analyses difficult. However, the major disadvantage of using the RNAi system is that 100% knockdown is hard to achieve and therefore, some putative downstream targets may not be detected.

3.1.2.2 Components of a good RNAi system

A number of factors should be taken into account when generating a good working RNAi system in mammalian cells. Some of the most important factors are discussed below.

(i) The RNAi system

Different types of RNAi triggers have been documented to generate efficient knockdowns in mammalian cells (Chapter 1, Section 1.3.1.2). These include the use of siRNA, shRNA and shRNA-mir. The choice of the RNAi trigger depends on the type of studies the researcher wants to perform and achieve. Typically, if a transient assay is sufficient, siRNA would be the best choice due to its ease of use and availability of validated siRNA sequences. If stable gene silencing is required for the study, shRNA and shRNA-mir would be required to generate stable transfectant.

During the design of siRNA or shRNA sequences, many considerations should be taken into account. These will be discussed in further details in Section 3.1.2.3. However, the target regions of the siRNA or shRNA should also be chosen with care. Some genes have a number of splice variants. Thus, when designing the regions to be targeted, a region common to all splice variants

(usually the 3' ends of the mRNA) would be more desirable so that all the splice variants are silenced in the assay. Conversely, if a particular splice variant is of interest, a region unique to the relevant variant should be used as the basis for designing the siRNA or shRNA.

(ii) The cell system and delivery strategy

To deliver siRNA or shRNA constructs into mammalian cells, various transfection methods can be employed dependent on the cell types (Chapter 1, Section 1.3.1.2). An optimal delivery strategy should be used for a specific cell type to achieve the highest possible knockdown efficiency. Regardless of the delivery strategy being used, the transfection efficiency should be monitored. This is because the transfection efficiency of siRNA or shRNA constructs directly affects the silencing level of the targeted gene. This can be done by transfecting a fluorescently-labelled siRNA or a shRNA construct expressing a fluorescent protein such as GFP and monitoring the fluorescence intensity. In addition, the effect of transfection on the cell should also be studied to identify any possible non-specific effects induced by the transfection method or the RNAi system. Studying the cell morphology or growth patterns of the cells are ways to determine such non-specific effects.

To trace the delivery of siRNAs into cells and test the transfection efficiency of the electroporation system, chemically synthesised siRNAs can be modified to have a fluorescent dye attached at the 3' or 5' end. Thus, when this labelled siRNA is transfected, cells containing the labelled siRNA will emit fluorescence which can be detected when the fluorophores bound to the siRNA molecules are excited by the appropriate laser. The choice of the fluorescent dye is crucial. Considering a siRNA duplex is a very small molecule, attaching a fluorescent dye with a high molecular weight may affect the uptake by the cells during electroporation and thus, such transfections may not truly mimic normal siRNA delivery. Because of the relatively low molecular weight of fluorescein (FITC), studies have demonstrated that attaching this fluorophore to the 3' end of siRNA duplexes will not reduce its efficacy in transfection (Holen et al., 2002). FITC labelled siRNAs have previously been used to monitor siRNA delivery by lipofectamine using fluorescence microscopy. It was demonstrated that delivery was the highest at 2-4 hours after transfection but diminished at later time points (Holen et al., 2002).

(iii) Minimizing non-specific effects

A number of non-specific effects have been described for RNAi (Chapter 1, Section 1.3.1.3). These non-specific effects should be minimised and monitored during the design of the RNAi experiment.

IFN response

To limit the IFN response, several considerations should be made during the design of RNAi experiments. Firstly, expression levels of IFN-stimulated genes (ISGs) should be monitored in any

RNAi experiments. IFIT1 is one of the most sensitive markers for ISG activation (Marques et al., 2006). Secondly, siRNAs should be of high purity and their concentration should be titrated to the lowest effective dose. Thirdly, earlier time points after induction of siRNA knockdown are preferable for studying their effects in perturbation experiments. This is because the hydrolysis of the 3' overhangs which leads to accumulation of blunt-ended siRNAs can trigger the IFN response, and these usually occur at later, rather than earlier, time points. Fourthly, certain sequence motifs such as UCUCU and GUCCUUCAA should be avoided in designing siRNAs as they have been shown to induce the IFN response in immune cells (Judge et al., 2005). Finally, for shRNAs expressed by a vector under the control of the H1 or U6 promoter, AA dinucleotide motifs near the transcription start sites should be avoided.

Off-target effects

Chemical modification of siRNAs where a 2'-O-methyl ribosyl substitution at position 2 of the guide strand has been demonstrated to reduce off-target effects (Jackson et al., 2006). This modification could also eliminate siRNA-induced undesirable and toxic phenotypes (Fedorov et al., 2006). In addition to siRNA design, other experimental parameters such as rescue and redundancy experiments should also be considered (Echeverri et al., 2006). Rescue experiments are performed by expressing a functional, though mutant, version of the target gene which is resistant to the siRNA trigger. If the phenotype can be rescued, one can be convinced that the phenotype is caused by a specific siRNA induction. Redundancy experiments involve the use to two or more siRNAs with different sequences raised against the same target gene. This can significantly reduce the probability that the resulted phenotype or gene expression changes are caused by off-target effects.

Saturation of RNAi pathway

Whereas saturation effects induced by siRNAs can be effectively suppressed by optimizing siRNA concentrations (Semizarov et al., 2003), expression of shRNAs in mammalian systems is more difficult to monitor as random integration of the expression construct can result in varied levels of shRNA expression. One solution to this problem would be to screen transfected cell lines for the copy number of integration events.

(iv) Functional validation

As different siRNA or shRNA generate different levels of gene knockdown, it is essential to evaluate the silencing capability before moving forward with large-scale experiments. Many different ways can be used to detect the expression changes in the gene being targeted at the mRNA and protein levels. These include northern blot, qRT-PCR, western blot, immuno-fluorescent (IF) studies, and fluorescence-activated cell sorting (FACS). These methods allow us to detect the expression of the endogenous gene and their advantages and disadvantages are summarised in Table

3.1. One of the biggest issues regarding the detection of changes in protein level is the availability of antibodies against the proteins of interest. Another option is to clone the targeted gene with an epitope tag (such as FLAG and MYC) so that the expression of level of the fusion protein can be detected using antibodies against the tag. This method circumvents the problem associated with requirement of good antibodies against the protein being targeted but requires the expression of an exogenous fusion protein.

Method	Detection level	Advantage	Disadvantage	Throughput
Northern blot	Endogenous mRNA	Easy	Non-quantitative RNA isolation	Low
qRT-PCR	Endogenous mRNA	Sensitive Quantitative	RNA isolation Primer design	High
Western blot, Immunofluorecence (IF), FACS etc	Endogenous protein	Easy	Antibody availability	Low
Western blot, Immunofluorecence (IF), FACS etc. on epitope tag	Exogenous fusion protein	Same antibody for detection	Cloning	High

Table 3.1. Commonly used assays for functional validation of mammalian RNAi.

qRT-PCR is particularly useful in the functional validation of knockdown as it is relatively fast and easy. Two major methods have been developed for qRT-PCR assays: the Taqman™ system and the SYBR Green system. Fluorescent reporter probes methods such as the Taqman™ system (AppliedBiosystems Inc.) is the most expensive but accurate way to quantify the PCR products. It involves the use of a probe which is specific to the amplified sequence and has a reporter fluorescent dye at the 5'end and a quencher dye at the 3'end. The probe, when unbound to any DNA, is non-fluorescent due to fluorescence quenching by the quencher dye when it is close to the fluorescent dye. However, when it binds to the target DNA during the PCR, the DNA polymerase (which has 5' to 3' exonuclease activity) cleaves the probe separating the fluorescent and quenching dyes leading to the emission of fluorescence.

Another type of probe is the DNA-binding dye such as SYBR Green (Morrison et al., 1998) which binds to double stranded DNA during the PCR amplification of the target DNA and emits fluorescence upon binding. The fluorescence can be detected with a real-time thermocycler. The intensity of fluorescence directly correlates with the exponential increase in the PCR products and can be determined by the threshold cycle (Ct). Thus, by comparing the experimental and reference samples, changes in expression level of genes can be detected. DNA-binding dyes are comparatively cheaper than the Taqman™ system but they are less accurate as they bind to any double-stranded DNAs including primer dimers.

(v) Knockdown effectiveness

The ">70% knockdown" threshold is considered to be a benchmark by which commercial providers of siRNAs assess whether a siRNA has been validated as a good working assay against a gene of interest. This threshold is also supported in the wider scientific community. In a RNAi screen of the PI3K pathway, a 70% knockdown efficiency was used as a cut-off for screening of effective siRNA assays (Hsieh et al., 2004). Subsequent proof-of-principle cell-based genetic screen showed that only siRNAs with >70% knockdown scored in a functional assay (Hsieh et al., 2004). However, even if a siRNA has a 70% knockdown efficiency at a specific time point after transfection, it may be possible to increase the efficiency by performing time-course experiments and identifying additional time points where the knockdown is at its highest level. However, one must also be aware, as mentioned in Chapter 1, that accumulation of blunt-ended siRNAs through time (as a result of the hydrolysis of the 3' overhang) will enhance the non-specific IFN response (Chapter 1, section 1.3.1.3 A). Therefore, it is also a common view that using an earlier, rather than a later, time point in perturbation studies reduces the likelihood of identifying non-specific effects at the level of gene expression. It has been shown, in at least one study, that a maximum knockdown was observed at the 24 hour time point for a human coagulation tissue factor (Holen et al., 2002), although this is likely to vary from gene to gene and between different siRNA assays for a given gene. Indeed, the time required for the maximum RNAi efficiency was shown to be proportional to the half-life of the target protein (Choi et al., 2005). Thus, all of these factors need to be considered when deciding on the appropriate experimental conditions to analyse the biological effects of siRNA knockdowns.

3.1.2.3 Advantages of using siRNA in the current study

In this thesis, transient knockdown of members of SCL eythroid complex by siRNAs was employed for a number of reasons, as follows:

(i) Rapid and easy of use: Transfection of small nucleic acid molecules, such as siRNAs, into mammalian cells is a relatively simple process. Delivery methods using lipofection and electroporation are well-developed and reagents and optimised protocols are available commercially.

(ii) Commercial siRNAs: Commercially- or custom- designed siRNAs are widely available (from companies such as Ambion, Dharmacon and Invitrogen) for most human and mouse genes. In some cases, functionally validated siRNAs can also be purchased which further facilitates the use of working siRNA assays for the gene of interest.

(iii) Optimisation of siRNA design: siRNA design is arguably the most critical step developing an effective knockdown assay for the gene of interest. One of the major advantages of using a siRNA

platform is that the experimenter can optimise the design of siRNAs. *In silico* design of siRNAs utilises different algorithms for sequence design and chemical modifications of the siRNA duplexes. Such optimisation can enhance their knockdown efficiency, reduce non-specific effects, improve their stability in culture systems and lower their cellular toxicity.

Many researchers have studied ways to optimise sequence design for generating effective siRNAs. Elbashir et al (2002) elaborated several guidelines for chemical synthesis of siRNAs (Elbashir et al., 2002) and these are summarised in Table 3.2. However, in addition to these rules, one should also take into account the secondary structure prediction of the mRNA and sequence comparison of the siRNA with the entire genome to reduce off-target effects. Some commercial companies, such as Ambion, also provide public resources for designing siRNA sequences (http://www.ambion.com/techlib/misc/siRNA_finder.html).

Length	19 nucleotides
GC content	30 to 70%
mRNA regions to target	Between 100 nucleotides from start codon and stop codon Avoiding 5' and 3' UTR
Overhangs	Two 3' 2-deoxythymidine residues

Table 3.2. Criteria for designing siRNAs (Elbashir et al., 2002).

Furthermore, chemical modifications have also been demonstrated to increase siRNA efficiency and stability, used either alone or in combination (Table 3.3).

Property	Modification
Increased thermal stability	2'-fluoro; 2'-O-methyl
Increased stability to digestion by nucleases	2'-fluoro pyrimidines; most chemically modified bases at the 3' and 5' termini
Reduced off-target effects	2'-O-methyl ribosyl substitution at position 2 of the guide strand

Table 3.3. siRNA properties that can be improved by the introduction of chemical modification (Corey, 2007).

(iv) Transient knockdown: Transient siRNA knockdown occurs from between 1 to 4-5 days after delivery of the siRNA to the cell where the mRNA level fully recovers after 4-5 days (Holen et al., 2002). This time interval is normally sufficient to experimentally observe transcriptional changes in downstream target genes. Such rapid changes in expression are often unobservable in stable knockdown cell lines, since a lot of selection and induction steps are required in order to obtain cell populations that display the knockdown phenotype.

3.1.3 The cell culture system under study

The human cell culture system used in this study is the erythroleukemic K562 cell line originally isolated from a chronic myeloid leukemia (CML) patient in blast crisis (Lozzio and Lozzio, 1977). It carries the BCR-ABL translocation and is thought to represent the common myeloid progenitor (CMP) stage of myeloid development which can give rise to both the megakaryocytic and erythroid lineages. It can be induced to erythroid differentiation by hemin and to megakaryocytic

differentiation by phorbol 12-myristate 13-acetate (PMA) (Huo et al., 2006). K562 cells have an approximate doubling time of 24 hours and it is well-characterised and studied for various aspects of haematopoietic function. Thus, its relevant biological characteristics, ease by which it can be differentiated and transfected makes K562 an excellent starting point from which to identify targets of the SCL erythroid TF complex.

K562 has been widely used in the study of erythroid development and transcriptional regulation of erythroid-specific genes. It was used in the study of GATA1 regulation of its target genes GFI1B and EKLF (Bose et al., 2006; Huang et al., 2005). This cell line was also used in the mapping of GATA1 binding sites along the β-globin locus (Horak et al., 2002) as well as in the study of histone modification and transcription factor binding in the α-globin locus (De Gobbi et al., 2007). These studies together demonstrated that K562 is a reliable system for the investigation of transcriptional regulation in erythroid cells.

3.2 Aims of this chapter

The aims of work presented in this chapter were:

1. To develop working siRNAs for five members of the SCL erythroid complex (SCL, GATA1, E2A, LMO2 and LDB1) in K562 cells.
2. To determine the effectiveness of the delivery strategy of the siRNAs into K562 cells.
3. To study the effect of these working siRNAs phenotypically as a function of the morphology and growth of K562 cells.
4. To characterise the efficacy of the siRNA knockdown in time-course experiments at the mRNA and protein level.

3.3 Overall strategy

Generating efficient knockdowns for each TF in the SCL erythroid complex were required for studying the transcriptional downstream target genes regulated by members of this complex (see Chapters 4, 5, and 6). To this end, a strategy was developed which allowed knockdown efficiencies to be monitored and studied in a variety of ways. The overall strategy is summarised in Figure 3.1. Firstly, commercially-designed custom-made siRNAs against each of five TFs (GATA1, E2A, LMO2, LDB1 and SCL) were transfected into cells by electroporation and knockdown efficiencies of the relevant mRNAs were measured by quantitative PCR (see section 3.4.1). At the same time, antibodies against the five TFs were tested and characterised in western blotting. Secondly, the efficiency of the delivery strategy was monitored by transfecting fluorescently-labelled siRNAs and

analysing the proportion of cells which showed fluorescence by FACS. Thirdly, physiological and morphological effects of siRNA-induced knockdown and transfections on cultured cells were also monitored to further characterise any gross phenotypic changes induced by the knockdowns. Finally, after identifying siRNAs which resulted in knockdowns by at least 70% of physiological mRNA levels, time-course experiments were performed over a period of 48 hours to study the changes in mRNA and protein expression during siRNA knockdown. This would allow timepoints to be identified at which the maximum knockdown efficiencies at the mRNA and protein levels were observed.

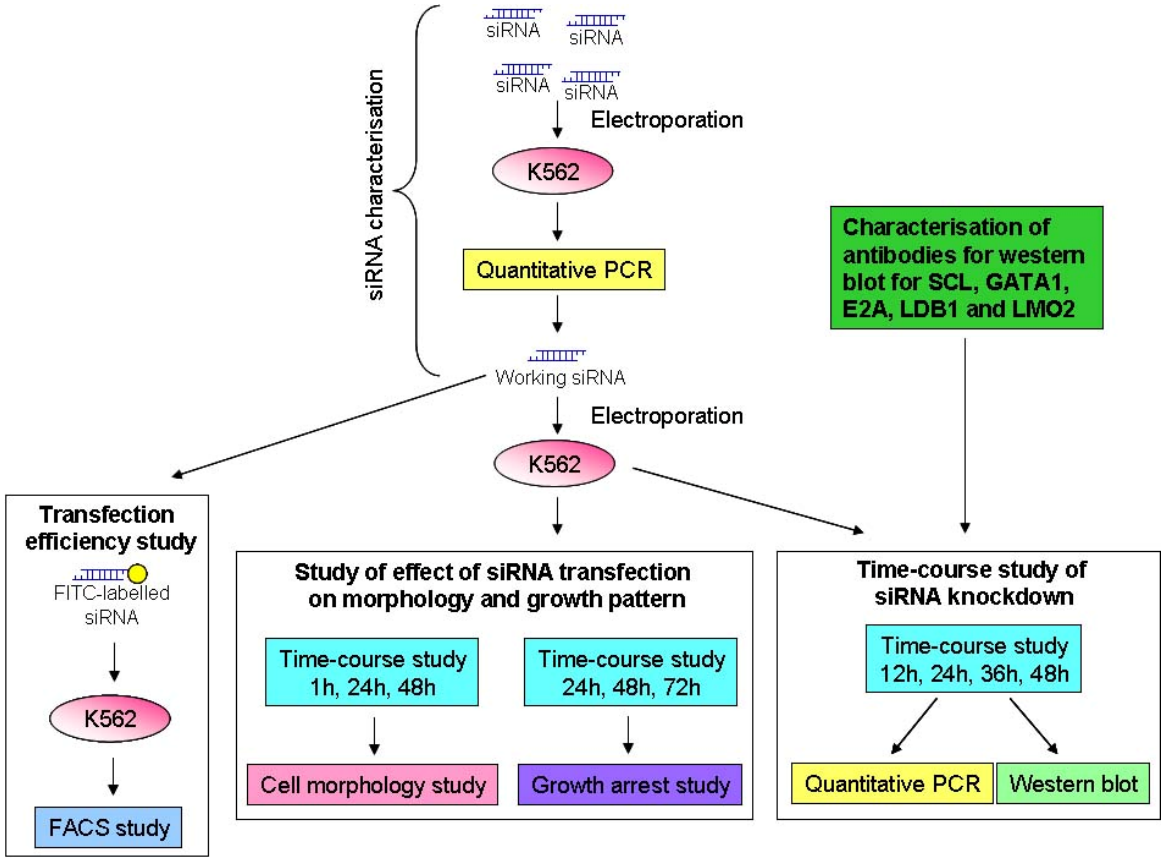


Figure 3.1. Overall strategy of siRNA knockdown analysis of TFs in the SCL erythroid complex. siRNAs and antibodies against each of five TFs were first characterised. The efficiency of the delivery strategy as well as the physiological and morphological effects of siRNA-induced knockdown were monitored. Time-course experiments were performed over a period of 48 hours to study the changes in mRNA and protein expression during siRNA knockdown.

3.4 Results

3.4.1 Developing working siRNA assays against TFs in the SCL erythroid complex (SEC)

A number of siRNAs were tested against each member of the SCL erythroid complex (SCL, GATA1, E2A, LMO2 and LDB1) to obtain siRNAs with high knockdown efficiencies. siRNAs were commercially designed and chemically synthesised by Eurogentec or Ambion without modifications. For all the five TFs under study, siRNAs were designed against a region common to

all transcript variants if applicable. Both of the commercial suppliers provided a guarantee service where replacements for custom siRNAs against any of the TFs tested were given when the siRNAs did not generate efficient knockdowns. For example, three siRNAs were supplied for each order and replacements were given so that at least two working siRNAs were obtained (please see below for criteria of working siRNAs). Transfections of siRNAs were performed using the Amaxa Nucleofector II. Amaxa has developed an optimised electroporation protocol for K562 cells which generates a transfection efficiency of more than 90% when monitored 24 hours after transfection (note: this efficiency is based on transfection of plasmids) (www.amaxa.com). To avoid any non-specific or stress responses (Semizarov et al., 2003), the siRNAs were transfected at final concentrations of 20 nM in the transfection media (although higher levels were shown in the Vetrie laboratory to result in similar knockdown efficiencies but with higher levels of non-specific effects; Philippe Couttet, unpublished observations) .

Knockdown at the mRNA level of each TF was monitored by quantitative real time PCR using SYBR assays at 24 hours after transfection. For each TF, two siRNAs targeting different regions of the gene with a knockdown efficiency of approximately 70% at 24 hours after transfection were chosen for further analyses. Using two siRNAs per TF lowers the likelihood of identifying off-or non-specific- targets in subsequent analyses (since the same off-targets are not normally found with two different siRNAs to the same gene; see also Chapter 1, Section 1.3.1.3 B). In addition, a siRNA against a gene which is not present in the human genome (firefly luciferase) was also tested and used as a negative control for all of the siRNA experiments performed for this thesis. Only one negative control siRNA was used as it does not target any regions in the human genome and should not elicit the off-target effect. Expression levels of the TF knockdown conditions were normalised against the expression levels found in the luciferase siRNA condition; this ensured that the changes in mRNA levels were due to the siRNAs targeting the gene of interest and not due to any generalised effects from electroporation or siRNA transfection. For a working siRNA to be chosen in the screening, less than approximately 30% of the mRNA level of the gene being targeted should remain 24 hours after transfection of the corresponding siRNAs at a concentration of 20 nM in the transfection media. Table 3.4 summarised the sequences, target exons and mRNA knockdown efficiencies of all the siRNAs tested. 16 siRNAs were tested in total and 11 of them passed the screening. Out of the 7 siRNA designed by Eurogentec (excluding the firefly pGL3 luciferase siRNA), 5 of them passed the siRNA screening. In contrast, out of the 6 siRNAs designed by Ambion, 3 of them passed. All the siRNAs which passed the screening (except those for LDB1) targeted the last exon. Among all the five TFs studied, the siRNAs for GATA1 gave the best knockdown efficiency with only 4 - 10% of the mRNA remaining after siRNA transfection. Figure 3.2 summarised the knockdown efficiency of all the siRNAs tested in the screening in one replicate.

Numbering for Figure 3.2	TF	Sources	Sense sequence (5' to 3')	Antisense sequence (5' to 3')	Exon	% of mRNA remained at 24 hour after transfection	Screening results	Designations
	Firefly Luciferase	Eurogentec	CUUACGCUGAGUACUUCGAtt	UCGAAGUACUCAGCGUAAGtt		---	---	LUC
1	SCL	D. Mathieu	GAAGCUCAGCAAGAAUGAGtt	CUCAUUCUUGCUGAGCUUCtt	4	29	Passed	SCLa
2	SCL	D. Mathieu	GGGAAUCACAUCUUUUAAGtt	CUUAAAAGAUGUGAUUCCCtt	4	31	Passed	SCLb
3	GATA1	Eurogentec	GGAUGGUAUUCAGACUCGAtt	UCGAGUCUGAAUACCAUCCtt	6	4	Passed	GATA1a
4	GATA1	Ambion	UGCGGAAGGAUGGUAUUCAtt	UGAAUACCAUCCUUCCGCAt	6	10	Passed	GATA1b
5	GATA1	Ambion	CAGGCCACUACCUACGCAAtt	UUGCAUAGGUAGUGGCCUGtc	6	71	Failed	---
6	E2A	Ambion	GGAAAAGGUGUCAGGUGUGtt	CACACCUGACACCUUUUCCtc	18	20	Passed	E2Aa
7	E2A	Eurogentec	CCUGGCUUAUUCUUCUAAAtt	UUUAGAAGAAUAAGCCAGGtt	18	16	Passed	E2Ab
8	E2A	Ambion	GCUCAAUGCCUGGUAUCUGtt	CAGAUACCAGGCAUUGAGCtg	18	100	Failed	---
9	E2A	Ambion	GCAGCCUGUUUGAAACGGCtt	GCCGUUUCAAACAGGCUGCtt	18	100	Failed	---
10	E2A	Ambion	GGUCUCCUUUUCUGGUCUUtt	AAGACCAGAAAAGGAGACCtg	18	30	Passed	---
11	E2A	Eurogentec	GUUCGGAGGUUCAGGUCUUtt	AAGACCUGAACCUCCGAACtt	2	39	Failed	---
12	LMO2	Eurogentec	CAAGCGGAUUCGUGCCUAUtt	AUAGGCACGAAUCCGCUUGtt	6	23	Passed	LMO2a
13	LMO2	D. Mathieu	GCAUCCAAGUGGCAUAAUUtt	AAUUAUGCCACUUGGAUGCtt	6	30	Passed	LMO2b
14	LDB1	Eurogentec	GGAUGGACCAAAGAGAUAUtt	AUAUCUCUUUGGUCCAUCCtt	5	16	Passed	LDB1a
15	LDB1	Eurogentec	CCUCCGACUCUGUGUGAUAtt	UAUCACACAGAGUCGGAGGtt	8-9	17	Passed	LDB1b
16	LDB1	Eurogentec	GGCAUUCCACAGCAACUUUtt	AAAGUUGCUGUGGAAUGCCtt	6	36	Failed	---

Table 3.4. Characterisation of siRNAs for knockdown of the SCL erythroid complex. The siRNA sequences, target exons and % of mRNA remaining at 24 hour after transfection are shown in the table. A siRNA passed the screening should be able to silence the target mRNA at a final concentration of 20 nM in the transfection media so that less than approximately 30% of the target mRNA remained 24 hour after transfection. The designation shows the symbols designed for each working siRNA which were used in subsequent analyses throughout this thesis.

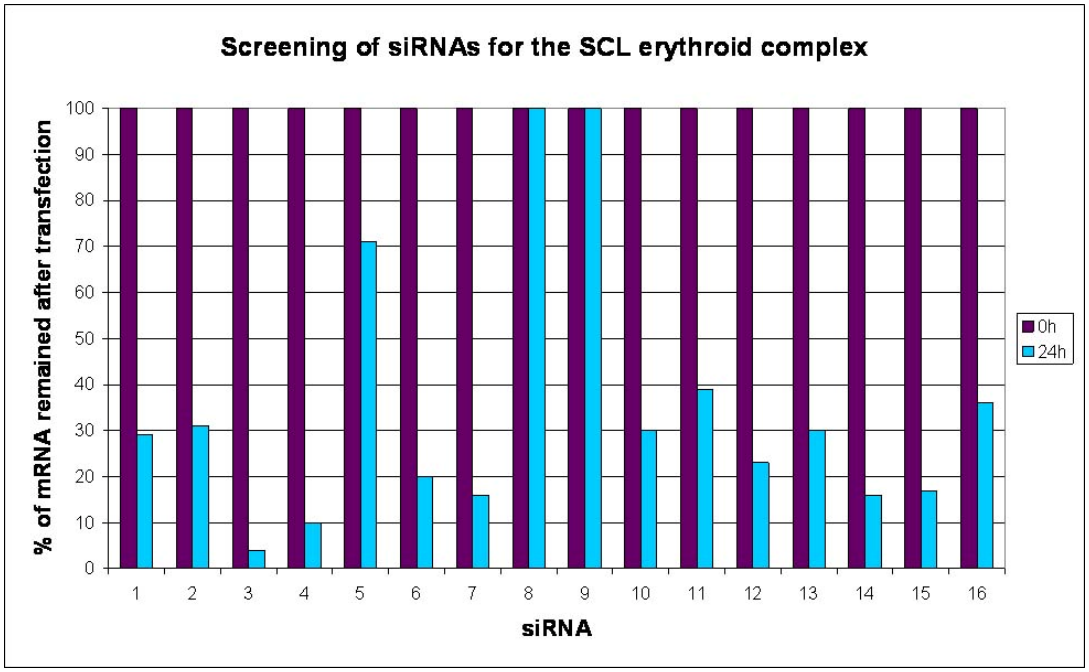


Figure 3.2. Screening of siRNAs for five members of the SCL erythroid complex. Y-axis: % of mRNA of the targeted gene remaining after transfection; x-axis: siRNA numbering as shown in Table 3.4. Figure shows results of one replicate in the initial screening

3.4.2 Characterisation of antibodies for western blotting analyses

The knockdown of specific mRNAs using siRNAs does not preclude that the effect is the same at the protein level. Thus, for the purposes of the work for this thesis, the time point after transfection when the maximum effects of a knockdown were observed were based on when the relevant proteins were reduced to their lowest levels in K562 cells. Therefore, it was necessary to develop western blot assays to quantify the protein levels for each of the TFs in the SCL erythroid complex before and after siRNA transfection in K562 cells. A number of commercially-available antibodies against each of the TFs were tested in western blotting assays with nuclear protein extracts of K562 cells. Antibody concentrations used in the initial western analyses were the highest recommended concentration by the manufacturers. However, it was also necessary to test some of the antibodies in appropriate dilution series to minimise background effects in the detection of the relevant proteins. All the antibodies tested and the how they performed in western assays were summarised in Appendix 3A.

For SCL, four antibodies were tested. The polyclonal antibody (Active Motif) detected a band on for the SCL protein which was approximately 45 kDa on western blots (Figure 3.3 B). The monoclonal anti-TAL1 3BTL73 antibody detected the appropriate band size (data not shown), whereas the unpurified sera generated a high background (Figure 3.3 C). The TAL1 Abcam antibody showed bands which were slightly higher in molecular weight than the expected SCL one

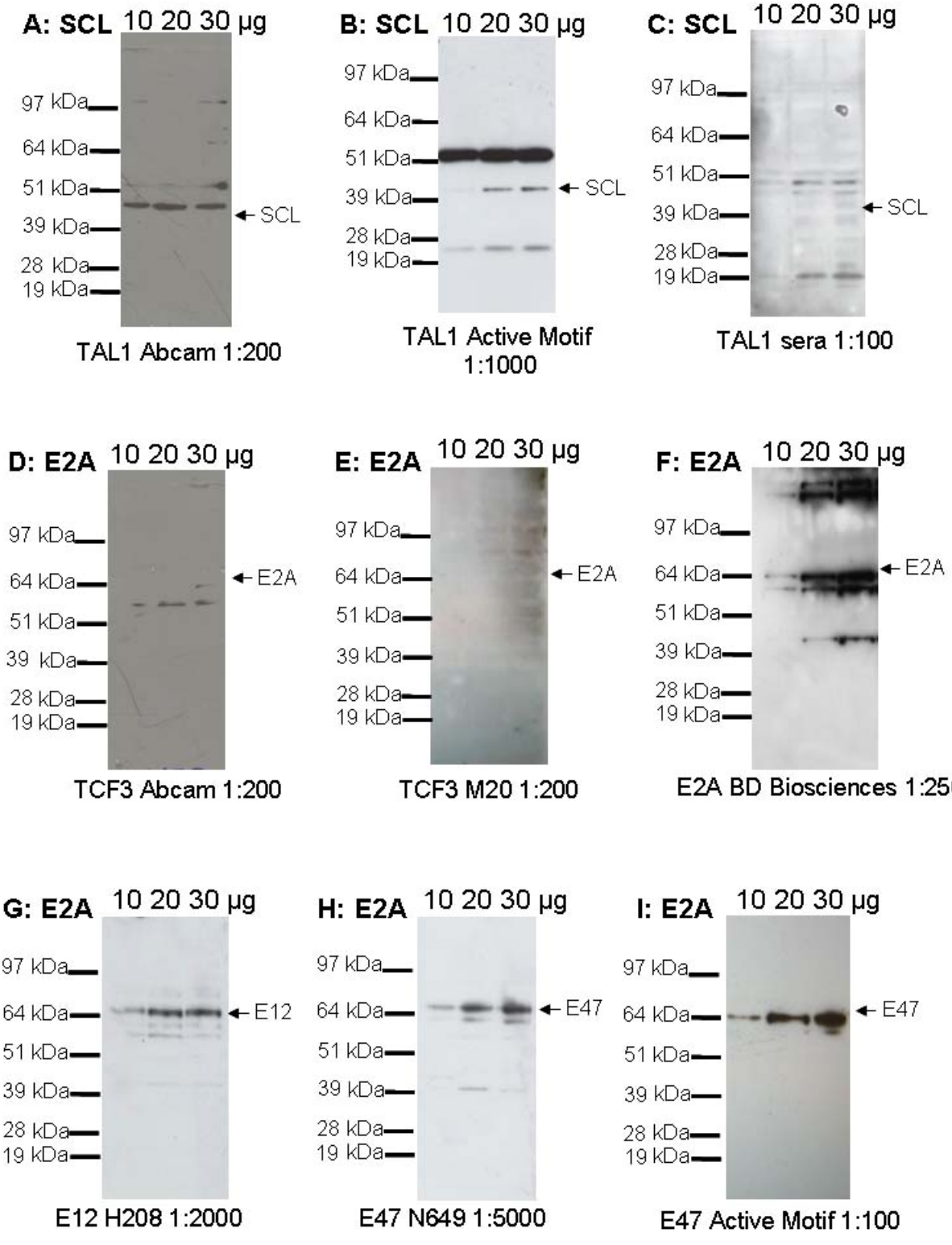
at 45 kDa (Figure 3.3 A). However, only the TAL1 Active Motif antibody showed a reduction in band intensity after knockdown by SCL siRNA (Section 3.4.5.1). Therefore, the Active Motif SCL antibody was used in subsequent analyses.

For E2A, seven antibodies were tested, including three which were raised against E2A (TCF3) and four raised against the splice variants E12 and E47. The TCF3 Abcam and TCF3 M20 Santa Cruz antibodies for E2A did not identify the predicted E2A protein band on western blots which was approximately 67 kDa in K562 cells (Figure 3.3 D and E). The E2A BD Biosciences antibody detected a band at approximately 64 kDa (Figure 3.3 F) and this band did not diminish in western analysis of E2A siRNA knockdown (Figure 3.4 A). Similarly, protein bands of approximately 64 kDa were observed for the E47 antibodies (Active Motif and Merck) (Figure 3.3 I and J respectively) but no reduction in band intensity was seen in siRNA knockdown (Figure 3.4 B and C respectively). Protein bands of the predicted size (67 kDa) for E12 and E47 protein were detected using the E12 H208 and E47 N649 antibodies (Figure 3.3 G and H respectively). The identities of these bands were confirmed in the knockdown experiment as they showed diminished intensities in E2A siRNA transfected cells (Section 3.4.5.3).

For GATA1, a polyclonal antibody from Santa Cruz gave a band of the predicted size of 45 kDa (Figure 3.3 K). This antibody generated very low background and did not cross-react with a closely related member of the GATA family, GATA2 (molecular mass of 51 kDa), which is also expressed in K562. The identity of the band observed in western analysis with this antibody was confirmed in the knockdown study (Section 3.4.5.2).

Similarly, an antibody for LDB1 from Santa Cruz was characterised which detected the LDB1 isoforms of the correct predicted size (43 kDa) (Figure 3.3 L). The identity of the band observed in this antibody was confirmed in the knockdown study where both isoforms were knocked down (Section 3.4.5.4.).

Three commercially available antibodies were tested for detection of the LMO2 protein which is approximately 18 kDa (Figure 3.3 M to O). However, none of them detected bands of the predicted size. Furthermore, high background and non-specific bands were detected even under both reducing and non-reducing electrophoresis conditions. One possible reason for the inability to detect this protein by western analysis is that the expression level of LMO2 in K562 cells may be below the limits of detection. However, it is equally likely to be an issue with the performance of the antibodies.


A: SCL 10 20 30 µg
97 kDa
64 kDa
51 kDa
39 kDa
28 kDa
19 kDa
SCL
TAL1 Abcam 1:200
B: SCL 10 20 30 µg
97 kDa
64 kDa
51 kDa
39 kDa
28 kDa
19 kDa
SCL
TAL1 Active Motif 1:1000
C: SCL 10 20 30 µg
97 kDa
64 kDa
51 kDa
39 kDa
28 kDa
19 kDa
SCL
TAL1 sera 1:100
D: E2A 10 20 30 µg
97 kDa
64 kDa
51 kDa
39 kDa
28 kDa
19 kDa
E2A
TCF3 Abcam 1:200
E: E2A 10 20 30 µg
97 kDa
64 kDa
51 kDa
39 kDa
28 kDa
19 kDa
E2A
TCF3 M20 1:200
F: E2A 10 20 30 µg
97 kDa
64 kDa
51 kDa
39 kDa
28 kDa
19 kDa
E2A
E2A BD Biosciences 1:250
G: E2A 10 20 30 µg
97 kDa
64 kDa
51 kDa
39 kDa
28 kDa
19 kDa
E12
E12 H208 1:2000
H: E2A 10 20 30 µg
97 kDa
64 kDa
51 kDa
39 kDa
28 kDa
19 kDa
E47
E47 N649 1:5000
I: E2A 10 20 30 µg
97 kDa
64 kDa
51 kDa
39 kDa
28 kDa
19 kDa
E47
E47 Active Motif 1:100

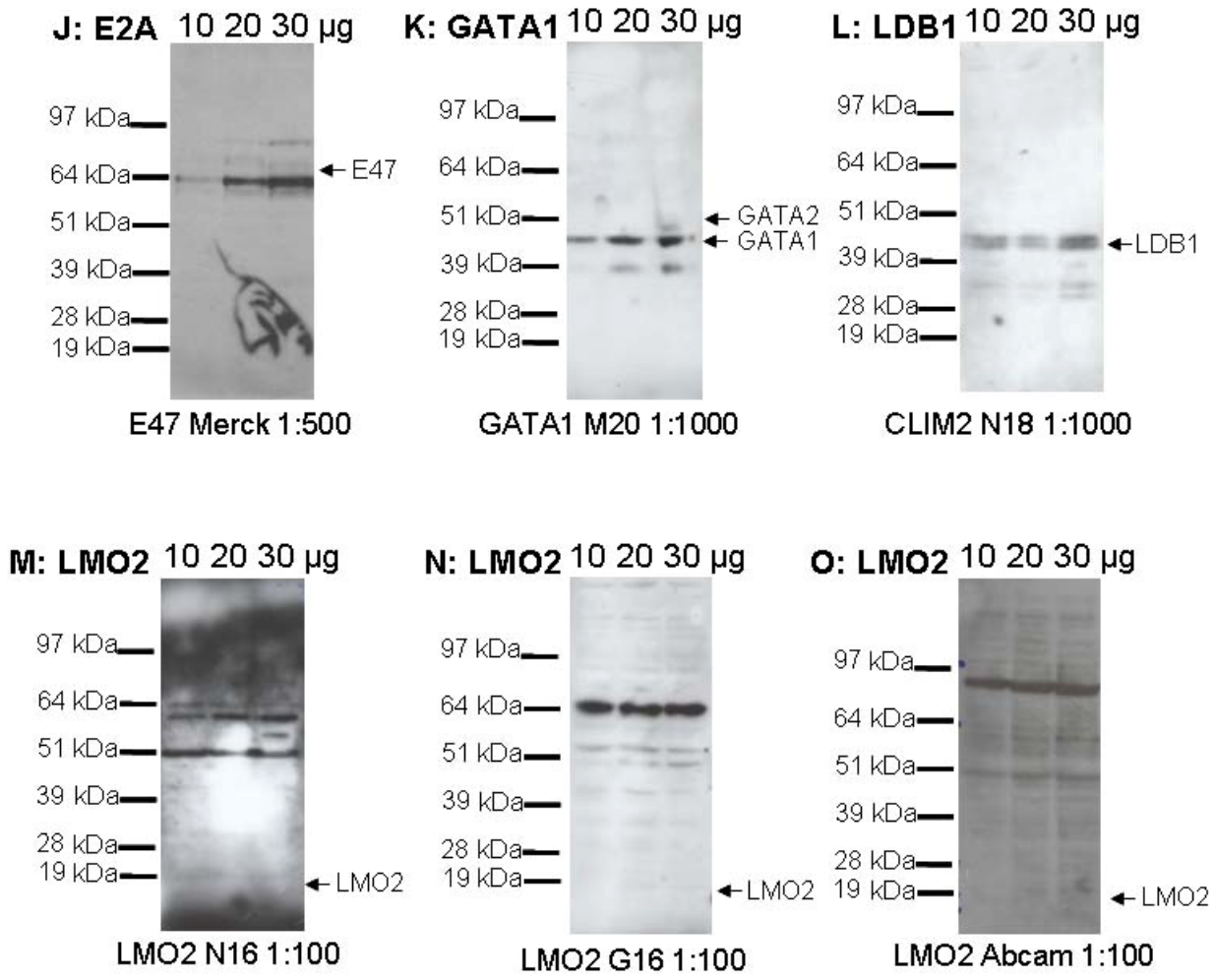


Figure 3.3. Western blotting analyses for the characterisation of antibodies against members of the SCL erythroid complex (SEC). 10, 20 and 30 µg of K562 nuclear protein extracts were used for western blot analyses to characterise the antibodies. SDS-PAGE was performed under denaturing and non-reducing conditions except for the results shown in panel B where reducing condition were used. In the analyses of LMO2 antibodies, both reducing and non-reducing conditions were tested. Each panel shows the x-ray films developed by chemiluminescence. The commercial names and dilutions of antibodies used are stated at the bottom of each panel. The dilutions used for western analyses were the recommended dilution from the company and they were titrated for optimisation in those working ones (panels B, G, H, K and L). The arrows on the right of each blot indicate the predicted protein size of the transcription factor under study. Size markers are shown on the left of each panel. A: TAL1 Abcam antibody; B: TAL1 Active Motif antibody; C: TAL1 2BTL73 anti-sera; D: TCF3 Abcam antibody; E: TCF3 M20 Santa Cruz antibody; F: E2A BD Biosciences antibody; G: E12 H208 Santa Cruz antibody; H: E47 N649 Santa Cruz antibody; I: E47 Active Motif antibody; J: E47 Merck antibody; K: GATA1 M20 Santa Cruz antibody; L: CLIM2 (LDB1) N18 Santa Cruz antibody; M: LMO2 N16 Santa Cruz antibody; N: LMO2 G16 Santa Cruz antibody; O: LMO2 Abcam antibody.

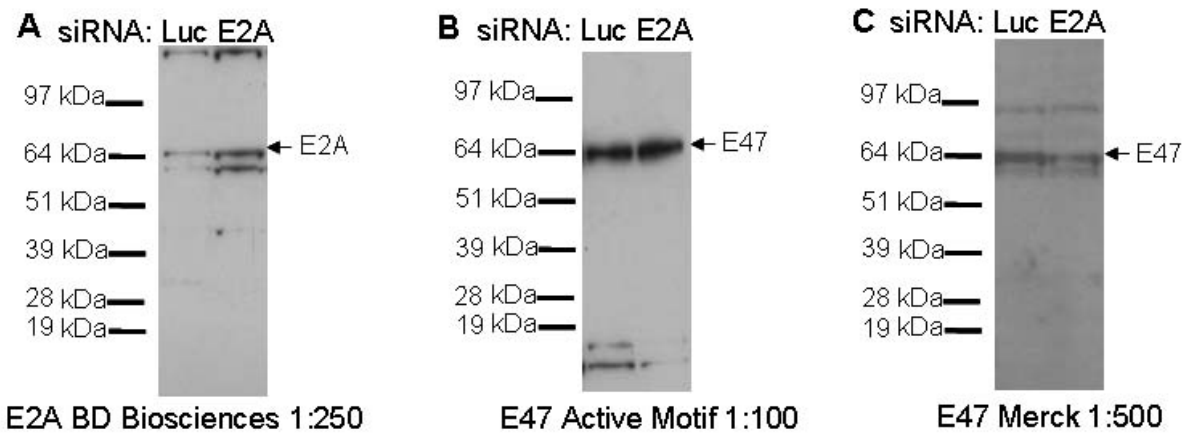
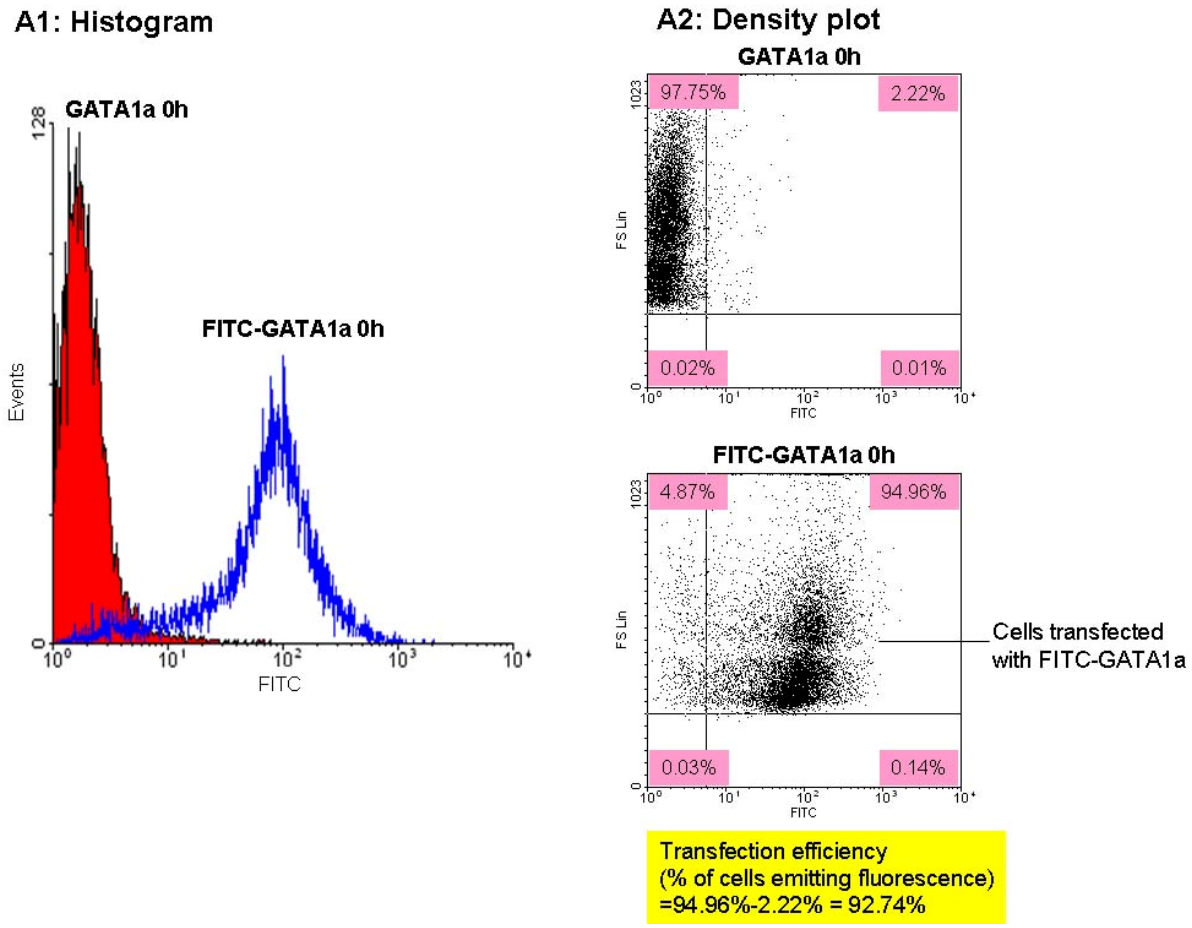


Figure 3.4. Western blot analyses for E2A knockdown. Nuclear proteins were extracted from firefly luciferase siRNA and E2Aa siRNA transfected K562 cells at 24 hour. 30 µg of nuclear protein extract were used for western blot analyses to characterise the E2A or E47 antibodies. SDS-PAGE was performed under denaturing and non-reducing conditions. Each panel shows the x-ray films developed by chemiluminescence. The commercial names and dilutions of antibodies used are stated at the bottom of each panel. The dilutions used for western analyses were the recommended dilution from the company. The arrows on the right of each blot indicated the predicted protein size of E2A or E47. Size markers are shown on the left of each panel.

3.4.3 Determination of the transfection efficiency of siRNAs

After working siRNAs were selected for each TF according to the criteria described above, the efficiency of delivery of the siRNAs into K562 cells by electroporation was studied. It was important to perform these experiments, since the K562 electroporation procedure used here was developed and validated by Amaxa using plasmids, not siRNAs. To this end, the efficiency of transfection was monitored using a 3' fluorescein (FITC)-labelled GATA1a siRNA (one of the GATA1 siRNAs which was used in further studies in this thesis). This experiment would be used as a model to provide evidence that the transfection efficiency was not a limiting factor in obtaining good knockdowns using the siRNAs for each TF. Fluorescein was used as a tag due its relatively low molecular weight compared to other fluorophores. Thus, the effect of the fluorescein tag on the transfection efficiency of the labelled siRNA could be minimised. FITC-labelled and unlabelled GATA1a siRNAs were transfected into K562 cells and aliquots of cells from both conditions were taken 24 hours after transfection. The cells were subjected to FACS and the proportions of cells carrying fluorescence were determined. The percentage of cells emitting fluorescence in the transfection of FITC-GATA1a siRNA was compared to the background of cells transfected with unlabelled GATA1a siRNA. There was a significant shift in the detection of fluorescence in the FITC-GATA1a siRNA transfected cells compared with GATA1a siRNA transfected cells at both 0 hour and 24 hour time points (Figure 3.5 A1 and B1). By setting the fluorescence intensity of the GATA1a siRNA transfected cells as the baseline in the density plot, percentages of cells which

carry FITC fluorescence were calculated for 0 hour and 24 hour time points (Figure 3.5 A2 and B2). The percentage of FITC-GATA1a cells emitting fluorescence was approximately 93% at the 0 hour time point and 84% at the 24 hour time point. It was also noted that the overall FITC fluorescence of the FITC-GATA1a transfected cells was higher at 0 hour than at 24 hour. The transfection efficiency of the FITC labelled siRNA (93%) is consistent with the knockdown efficiency of the GATA1a siRNA (approximately 4% of GATA1 mRNA remained after transfection) (Table 3.4). Taken together, the Amaxa Nucleofector II system provided a high transfection efficiency in K562 cells which was consistent with the knockdown efficiency of the corresponding siRNA. Therefore, transfection efficiency *per se* was not a limiting factor in obtaining knockdowns for the relevant TFs reported in this thesis.



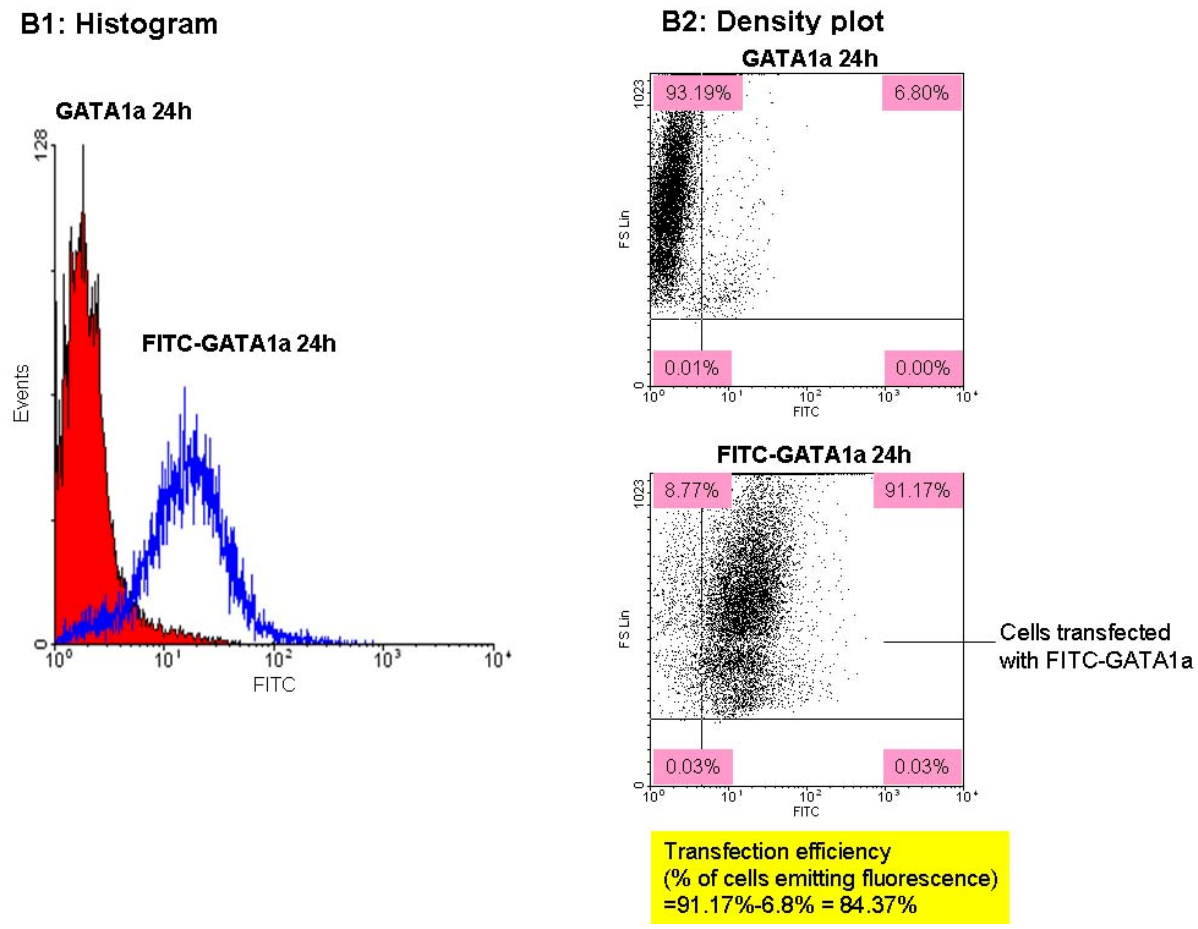


Figure 3.5. Flow cytometric analysis of the FITC labeled siRNA transfection into K562 cells. A) 0 hour and B) 24 hour. Panels A1 and B1 show histogram plots of GATA1a (red curve) and FITC-GATA1a (blue curve) transfections with the y-axis showing the number of cells (events) and x-axis showing FITC intensity. Panels A2 and B2 show density plots of GATA1a (top) and FITC-GATA1a (bottom) transfections with the y-axis showing the forward scatter (an indicator of cell size) and x-axis showing FITC intensity. Numbers shown in pink boxes show the percentage of cells in each of the quadrants. Top left quadrants contain cells emitting background fluorescence while top right quadrants contain cells emitting FITC fluorescence. Transfection efficiencies were calculated as described in the yellow boxes: transfection efficiency = % of cells in the top right quadrant in the FITC minus the control.

3.4.4 Changes in K562 cell growth and morphology induced by siRNA transfection

To determine whether electroporation of siRNAs had any physiological effects on K562 cells which were not specific to the knockdown of the TFs, changes in growth rates and morphology of K562 cells were assessed before and after transfection. For growth rates studies, the total number of K562 cells were determined at four time points (0, 24, 48 and 72 hours) after transfection with the same numbers of cells under each of the following conditions: a) no transfection; b) transfection with water; c) transfection with the luciferase siRNA and d) transfection with the GATA1a siRNA. In the case of the 0 hour time point, cells were harvested and counted immediately after transfection (or at the same time equivalent for cells that were not transfected). Transfection with water was included to investigate any effects solely due to electroporation itself rather than due to siRNA

transfection. In turn, transfection with luciferase siRNA was used to study any effects induced by the RNAi pathway which were not specific to the TF of interest (in this case, GATA1). As in section 3.4.3, GATA1a siRNA was used as the model for these studies. The number of viable cells at each of the 24, 48 and 72 hour time points (determined by trypan-blue staining) was determined relative to the 0 hour time point and plotted against time (Figure 3.6). In this way, cells which died as a result of electroporation were not included in the calculations. There were observable decreases in the growth rates of K562 cells at 24, 48 and 72 hour transfection conditions compared with the equivalent conditions for untransfected cells. The slope of the curve indicated how rapidly the cells were dividing. The effect on growth rate was the greatest between 0-24 hours as the slope was the smallest in the transfected cells compared to the wild type cells. After the 24 hour time point, the effect on growth rate diminished as the slopes were more similar between the wild type and transfected cells. Moreover, the growth patterns for transfections with water, luciferase siRNA and GATA1a siRNA were very similar at all time points. This indicates that the growth arrest observed in K562 cells was an effect of electroporation but not due to effects of siRNA transfection or induction of the RNAi pathway.

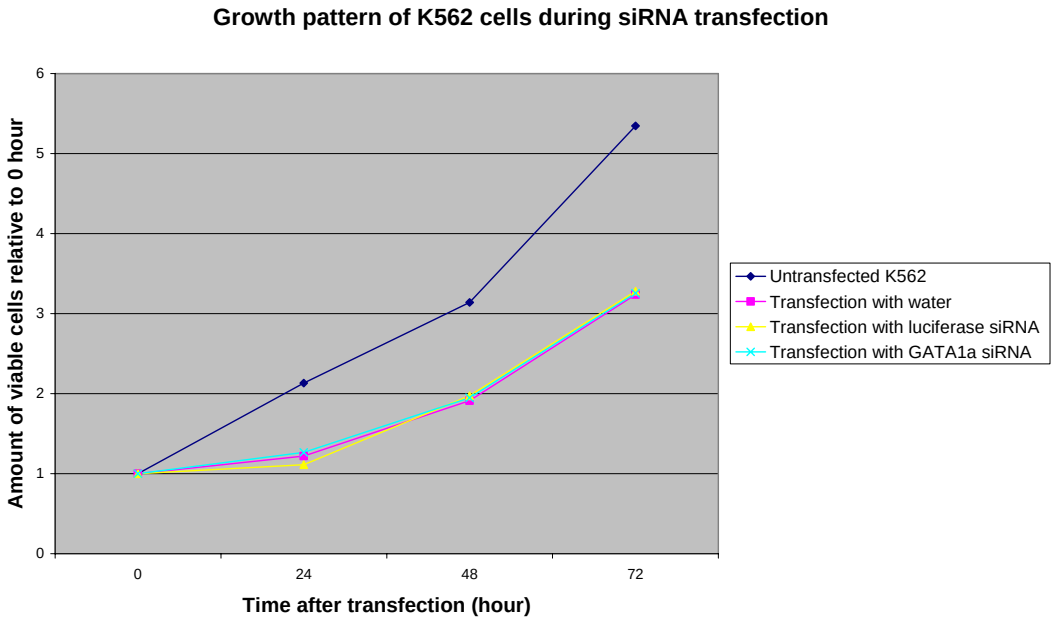
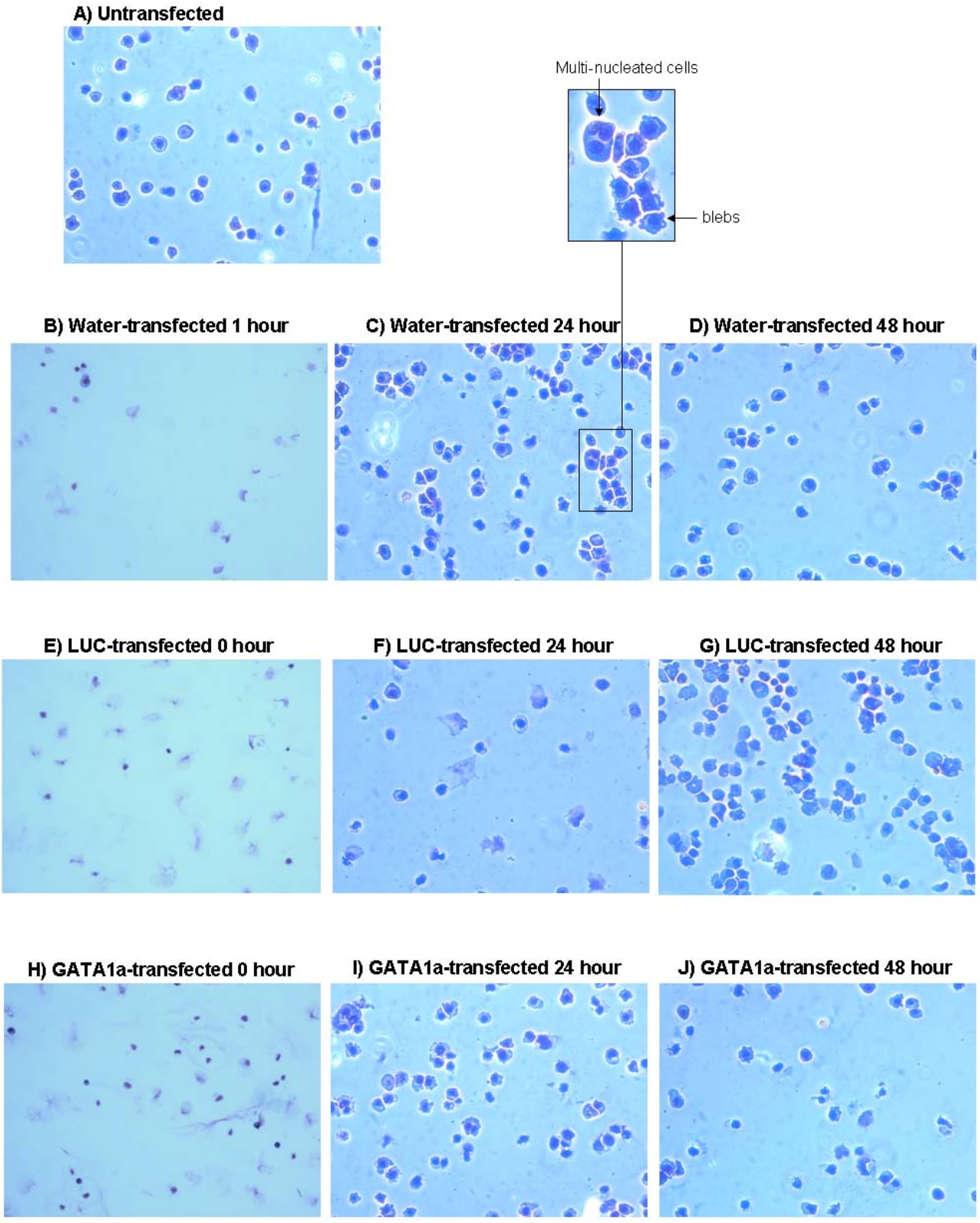


Figure 3.6. Growth pattern of K562 cells during siRNA transfections. The growth patterns of K562 cells under different transfection conditions were studied. Dark blue curve: untransfected K562 cells; pink curve: K562 cells transfected with water; yellow curve: K562 cells transfected with the luciferase siRNA; light blue curve: K562 cells transfected with the GATA1a siRNA. The graph shows the number of viable cells at different time points after treatment relative to the 0 hour time point. Results shown were average of two independent biological replicates.

Similarly, changes to cell morphologies of K562 cells under various transfection conditions were also studied. In addition to the various controls (no transfection, transfection with water, and

transfection with the luciferase siRNA), the effects of siRNAs against either GATA1 (GATA1a) and E2A (E2Ab) were monitored at three time points (1 hour, 24 hour and 48 hour) after transfection. As growth characteristics of cells were difficult to assess objectively, the experiment was performed blind and the E2Ab siRNA was also included as an additional control to monitor for any effects induced by siRNAs against specific TFs. Cells were harvested, spun down on a glass slide using a cytospin, stained with Stain Quick-Staining Kit (Lamb) and characterised under a light microscope. Untransfected cells were collected at the 0 hour time points, while transfected cells were allowed to recover from transfection and only collected at the 1 hour, 24 hour and 48 hour time points (the 1 hour time was used as cells at the 0 hour time point were considered to be fragile and often burst during preparation with the cytospin). Cells with different morphologies including cells with small blebs, large blebs, 2 nuclei and more than 2 nuclei were quantitated in a blind test. Wherever possible, 100 cells of different morphologies were counted by two independent scientists. The percentages of cells having different morphologies were calculated (Figure 3.8).

The untransfected cells were round, their cell membranes were smooth without any projections and were primarily mono-nucleated (Figure 3.7 A). The transfected cells, on the other hand, under all four transfection regimes, were smaller and had noticeable blebs (projections) of a variety of sizes on the cell membranes at both the 24 hour and 48 hour time points. For the transfected cells, more multi-nucleated cells were also observed (Figure 3.7 B-I). No differences were observed between the various transfection conditions. Thus, changes in the morphologies of K562 cells during transfection were mainly due to electroporation, and not as a result of siRNA transfection or induction of the RNAi pathway. This was further supported by the results of the blind test (Figure 3.8). Only less than 5% of wild type cells were shown to have blebs or be multi-nucleated whereas up to more than half of the cells were shown to possess these morphologies in the transfected cells, particularly at the 48 hour time point. It should be noted from Figure 3.8 that the blind counts at the 1 hour time point was not representative as the majority of the cells were dead after cytospin and only a small number of cells could be counted. These results, taken together with those described for the growth arrest studies, suggest that, at the level of gross morphology and growth, off-target or side-effects of the siRNA studies performed for this thesis could largely be attributed to the effects of electroporation.


A) Untransfected
Multi-nucleated cells
blebs
B) Water-transfected 1 hour
C) Water-transfected 24 hour
D) Water-transfected 48 hour
E) LUC-transfected 0 hour
F) LUC-transfected 24 hour
G) LUC-transfected 48 hour
H) GATA1a-transfected 0 hour
I) GATA1a-transfected 24 hour
J) GATA1a-transfected 48 hour

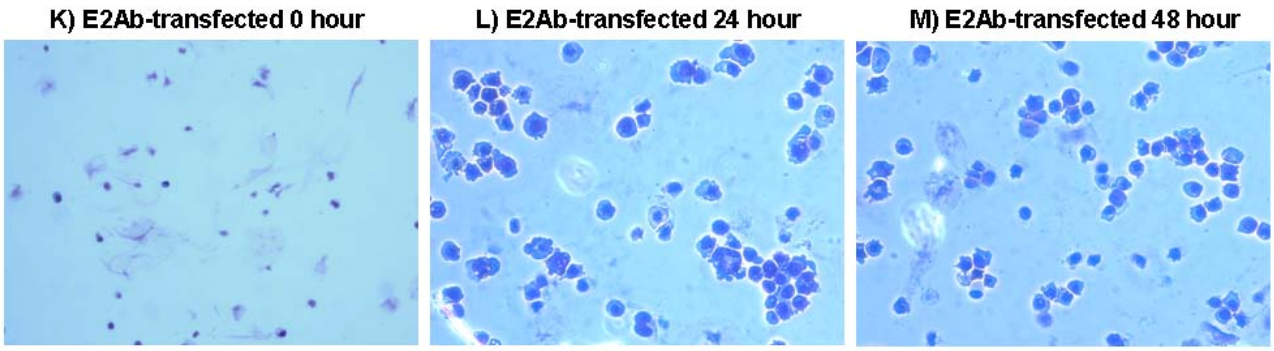


Figure 3.7. Cell morphological studies of K562 cells during siRNA transfection. K562 cells were untransfected or transfected with different conditions and collected at different time points. A: untransfected K562 collected at the beginning of experiment (0 hour); B, C and D: K562 cells transfected with water only and collected at 0 hour, 24 hour and 48 hour time points respectively; E, F and G: K562 cells transfected with luciferase siRNA and collected at 0 hour, 24 hour and 48 hour time points respectively; H, I and J: K562 cells transfected with GATA1a siRNA and collected at 0h, 24 hour and 48 hour time points respectively; K, L and M: K562 cells transfected with E2Ab siRNA and collected at 0 hour, 24 hour and 48 hour time points respectively. Multinucleated cells and cells with blebs are shown in the zoomed-in window for figure C.

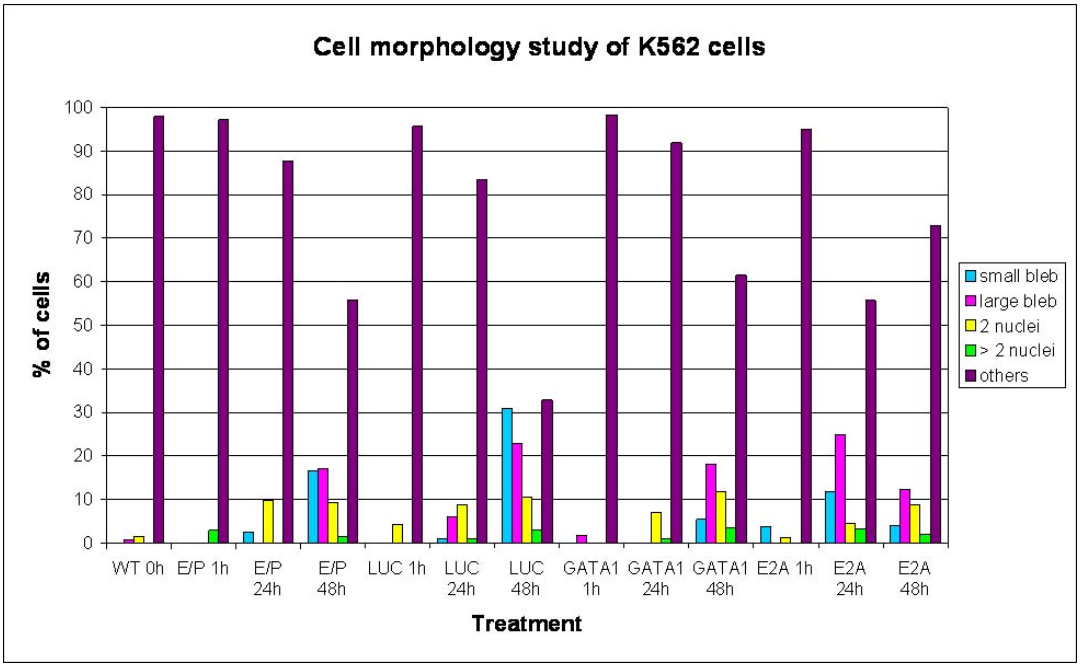


Figure 3.8. Quantitative analysis of cell morphology studies of K562 cells. Approximately 100 cells were counted randomly for different morphologies for each transfection at each time point. Y-axis: % of cells with different morphologies; x-axis: transfection of K562 cells. WT: wild type; E/P: electroporated with water; LUC: transfection with firefly pGL3 luciferase siRNA; GATA1: transfection with GATA1a siRNA; E2A: transfection with E2Ab siRNA. Key on the right shows the colour bars representing different morphologies.

3.4.5 Time-course study of siRNA knockdown

Studying the knockdown at one particular time point fails to provide a full picture of how the siRNAs are reacting within the cells and how the mRNAs or proteins are being silenced across

time. Therefore, time-course studies were required to investigate the changes in gene expression after siRNA transfections and to identify the time point where the best knockdown effects were induced. The following criteria were used in choosing the optimal time point in the time-course study:

1. The average remaining protein level of the targeted protein by the two independent siRNAs should be less than 30% compared to the firefly luciferase siRNA control. For each independent siRNA, the remaining targeted protein level should be 40% or less.
2. If protein level cannot be assessed due to the lack of a suitable antibody for western analysis, the average remaining mRNA level of the targeted gene by two independent siRNA should be less than 30%. For each independent siRNA, the remaining targeted mRNA level should be 40% or less.
3. The earliest time point where criteria 1 and 2 can be achieved should be chosen to reduce off-target effects.
4. The same time point should be used for both siRNAs to reduce discrepancies due to variations in the growth patterns of cells and induction of the RNAi pathway.
5. The earliest time point that can be chosen is 24 hours. The 12 hour time point should not be chosen as sufficient time is needed to allow the cells to be recovered after electroporation. As shown in the growth pattern study in section 3.4.4 (Figure 3.6), the growth rate of electroporated cells was lower in the first 24 hours and the cells slowly recovered after 24 hours. Also, the amount and quality of RNA extracted at the 12 hour time point were lower than other time points making subsequent analyses difficult.

To this end, time course experiments for GATA1, E2A, SCL, LMO2 and LDB1 were performed and the knockdowns were analysed at the mRNA and protein levels. mRNA and nuclear protein samples were collected at 12, 24, 36 and 48 hours after siRNA transfection and quantified by quantitative real time PCR and western blotting. For the mRNA quantification by real-time PCR, expression levels of the TF knockdown were normalised against the expression levels found in the luciferase siRNA condition for reasons described in section 3.4.1. The internal house-keeping controls, β-actin, GAPDH, β-tubulin and RPL19, were also included in the normalisation to minimise effects of variations of RNA concentration and quality across samples. For protein quantifications by western blotting, the nuclear protein extracts from transfected cells were first quantified and then equal amounts of protein were loaded into each well prior to electrophoresis and western analysis. The relevant protein bands detected from the western blots were quantified by densitometry and knockdown levels were determined relative to the luciferase control. Equal

loading of each lane of the westerns was verified by staining the membrane as described in Chapter 2 (Section 2.10.4). The results of each TF knockdown time course analysis are described in the following sections. Three biological replicates were performed for each siRNA time-course and the qRT-PCR and western blot data from one representative replicate time-course were reported in the following sections.

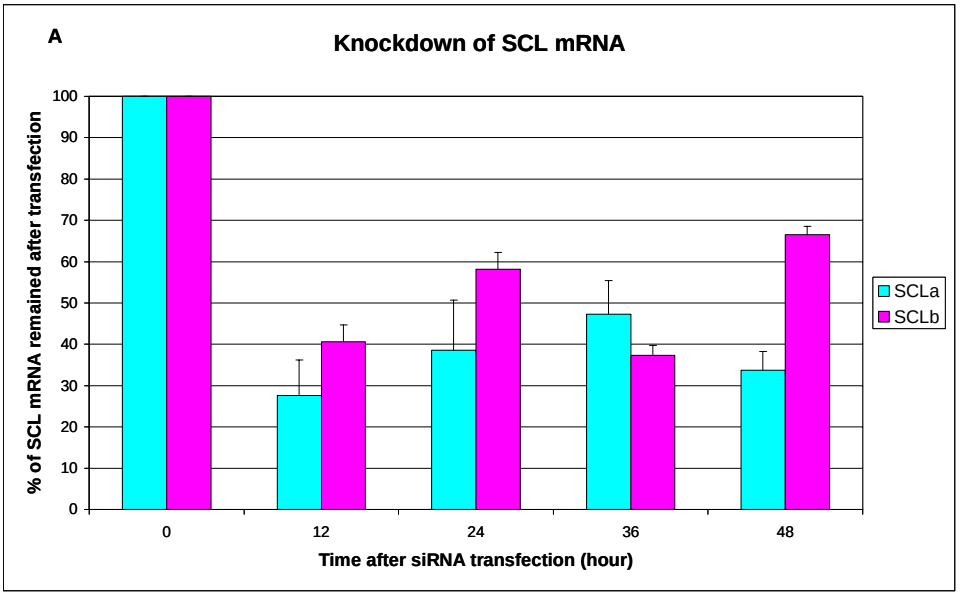
3.4.5.1 Knockdown of SCL

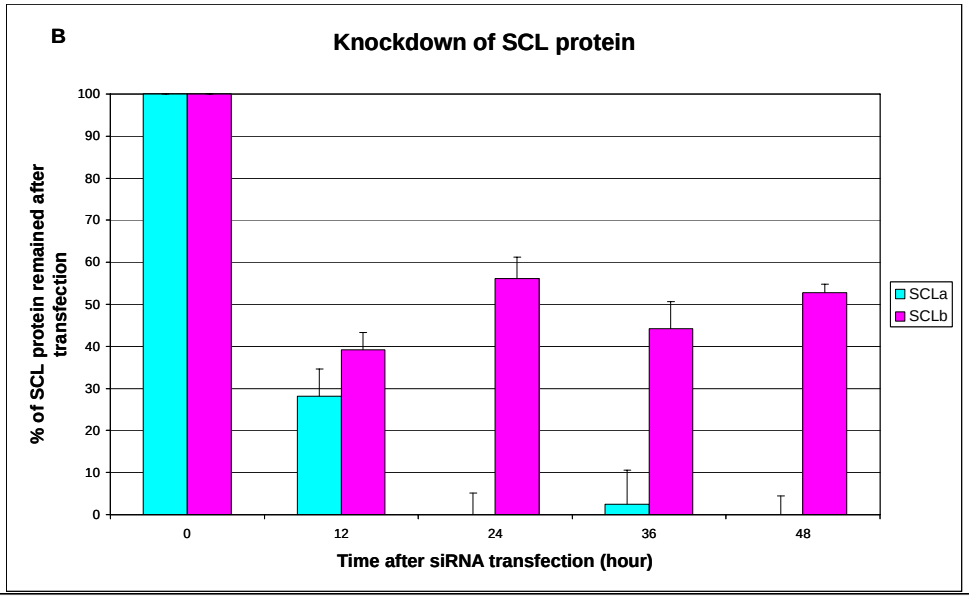
The knockdown of SCL with two independent siRNAs (SCLa and SCLb) was studied through a 48 hour time-course experiment as shown in Figure 3.9. SCL mRNA levels in the SCLa siRNA transfected cells decreased substantially to approximately 40% at 12 hour and fluctuated in the later time points at between 30-60% of its original physiological level. The knockdown at the protein level for siRNA SCLa was consistent with the mRNA level after 12 hours, but then showed a dramatic reduction to less than 10% its original level to nearly 100% knockdown at the 24 hour, 36 hour and 48 hour time points. These results were in marked contrast to the mRNA levels detected during these later time points. The SCLb siRNA, however, did not show reduction in the mRNA levels that had been anticipated based on the results of the initial screening described in section 3.4.1. The SCLb siRNA demonstrated a relatively weak knockdown after 24 hours with 60% of the mRNA remaining – this siRNA had previously given a knockdown to 31% of the physiological level of the SCL mRNA in the initial screens. However, SCLb reported better knockdowns at the 12 and 36 hour time points, where it achieved knockdowns with approximately 42% and 32% of the SCL mRNA remaining, respectively. SCL protein levels across the SCLb time-course showed the maximum knockdown after 12 hours (40% of protein remaining) which was consistent with its mRNA levels. However, whilst the maximum knockdown at the mRNA level was achieved at the 36 hour time point, this was not reflected at the protein level (maximum knockdown 43% protein remaining).

An optimal time point satisfying the four criteria listed above for both siRNAs against SCL could not be identified. The protein level of SCL was knocked down to less than 40% remaining at the 24 hour time point for SCLa but not for SCLb. The required remaining protein level (40%) for SCL was not reached for SCLb at all the time points tested (24 hour, 36 hour and 48 hour). Although less than 40% of the protein remained at 12 hour for SCLb, this time point was not considered according to the criteria set above. The 24 hour time point was chosen as the optimal time point for further study for the SCLa siRNA.

The time-course experiment for the SCLa siRNA suggested that the protein has a much shorter half-life than the mRNA because there was a lag in the mRNA levels being knocked down. However, from the data obtained for SCLb, the half-lives of mRNA and protein seem to be similar. Therefore,

it is difficult to draw any conclusion on the half-lives of mRNA and protein of SCL from these results.





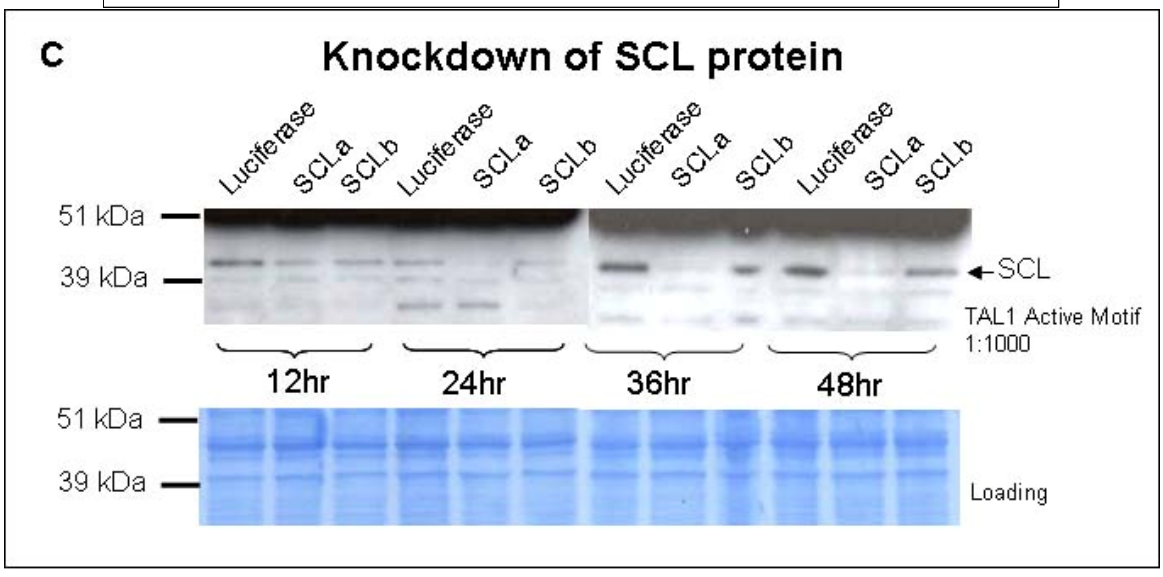


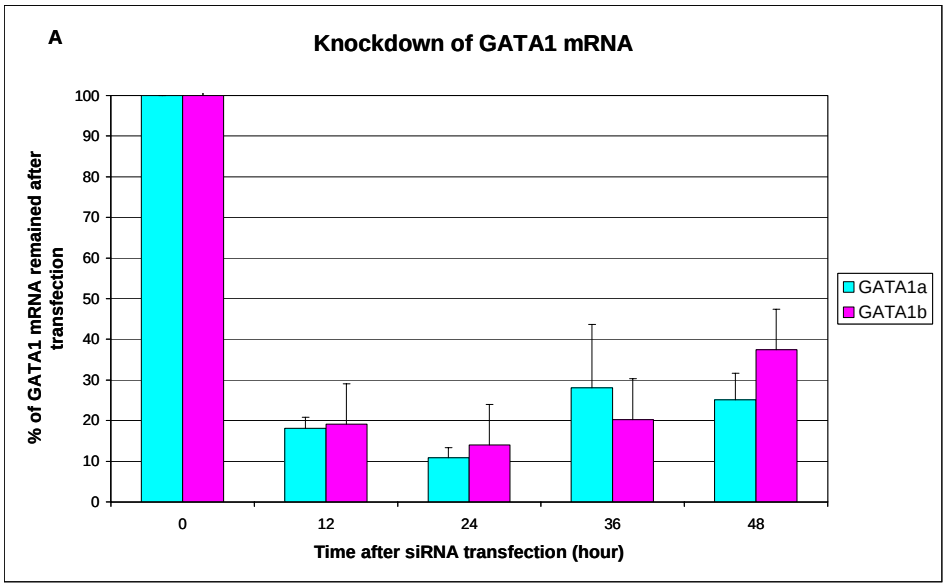
Figure 3.9. siRNA knockdown time-course study of SCL. Two siRNAs directed against SCL were used: SCLa and SCLb. A: Knockdown of SCL at the mRNA level was quantified by quantitative PCR as described in the text. Bar chart shows the mRNA level of SCL remaining (y-axis) after siRNA transfection relative to luciferase siRNA transfection across the time points (x-axis) Blue bars indicate mRNA levels in the SCLa condition while the pink bars indicate mRNA levels in the SCLb condition. The error bars show the standard error of the mean between the three independent biological replicates. B: Knockdown of SCL at the protein level by densitometry of bands determined by immuno-detection of the relevant protein band on western blots. Bar chart shows the protein level of SCL remaining after siRNA transfection relative to luciferase siRNA transfection. Blue bars indicate protein levels in the SCLa condition while pink bars indicate the protein levels remaining in the SCLb condition. The error bars show the standard error of the mean between the three independent biological replicates. C: Western blot analyses of SCL protein knockdowns. Upper panel shows the bands detected by immuno-detection of the western blot with the TAL1 Active Motif antibody. The lower panel shows the blot stained with Bradford reagent as a protein loading control. The arrow shows the predicted size of the SCL protein.

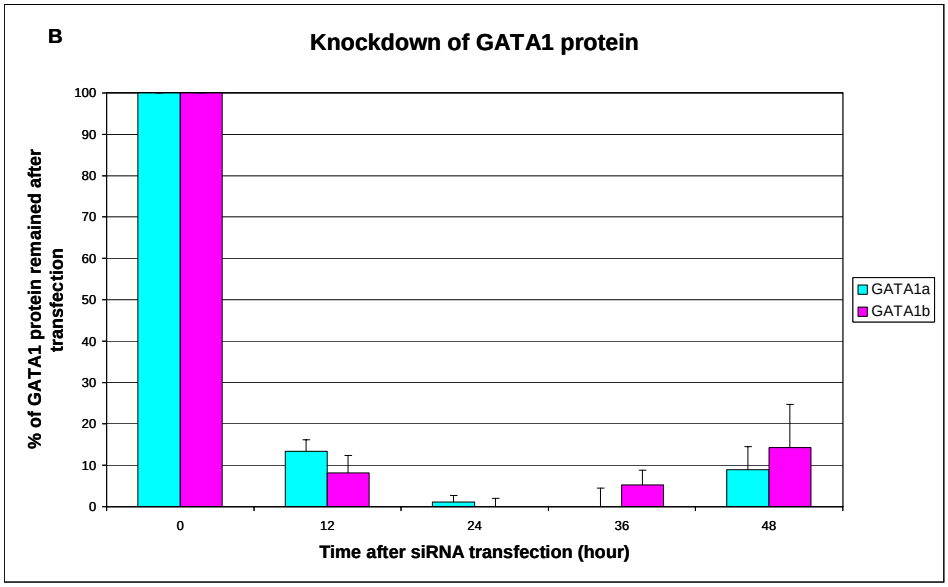
3.4.5.2 Knockdown of GATA1

The time-course analysis of the knockdown of GATA1 with two independent siRNAs (GATA1a and GATA1b) is shown in Figure 3.10. In general, both siRNAs generated substantial knockdowns at both mRNA and protein level. However, the GATA1a siRNA induced a marginally better knockdown effect at both mRNA and protein level at the majority of time points studied. GATA1 mRNA level in cells transfected with either of the siRNAs decreased dramatically to less than 20% of the original mRNA level at the 12 hour time point, further reducing to approximately 10% at 24 hour, then gradually increased in the later time points. The knockdown at the protein level had a similar trend with almost 0% of protein remaining at the 24 hour time point, and a gradual increase at the 36 hour and 48 hour time points.

Both siRNAs GATA1a and GATA1b gave the maximum knockdown at the mRNA and protein level at the 24 hour time point. This time point was the earliest time point where both siRNAs were able to knock down the GATA1 protein to less than 30% remaining on average. Although such knockdown was also observed at the 12 hour time point, this time point was not considered as the optimal time point for the reasons mentioned above. Therefore, the 24 hour time point was chosen for subsequent analysis.

The time-course experiment suggests that the half-lives of mRNA and protein of GATA1 were similar as a lag between the mRNA and protein being knocked down was not observed in both siRNAs. Also, the rapid reduction of both the mRNA and protein level at the 12 hour time point suggested that the half-lives of GATA1 mRNA and protein are relatively short.


A
Knockdown of GATA1 mRNA
% of GATA1 mRNA remained after transfection
100
90
80
70
60
50
40
30
20
10
0
0
12
24
36
48
Time after siRNA transfection (hour)
GATA1a
GATA1b


B
Knockdown of GATA1 protein
% of GATA1 protein remained after transfection
100
90
80
70
60
50
40
30
20
10
0
0
12
24
36
48
Time after siRNA transfection (hour)
GATA1a
GATA1b

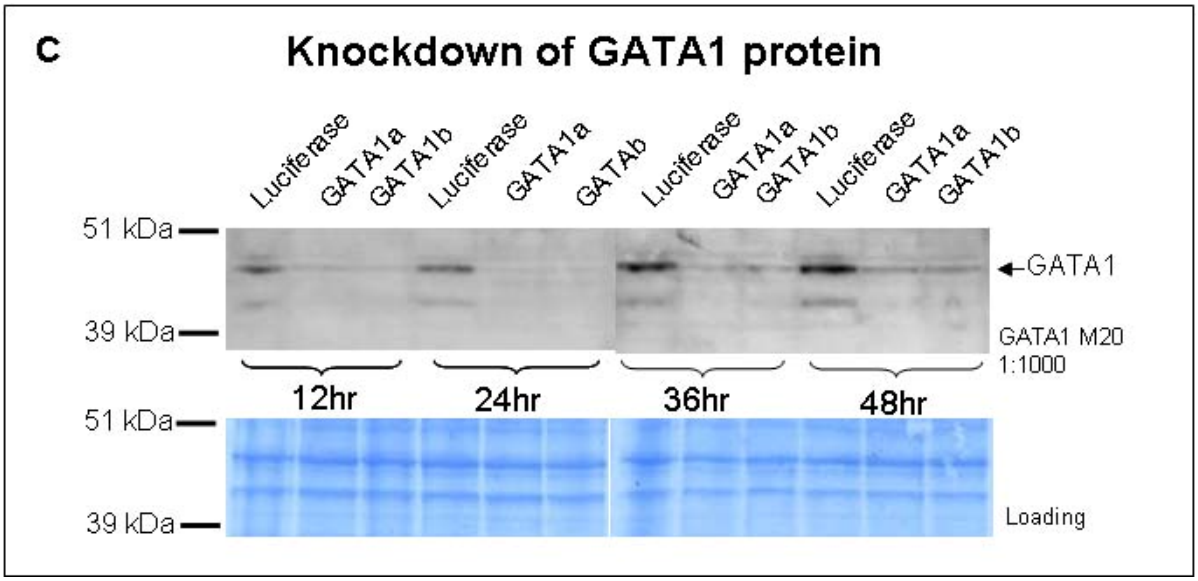

C
Knockdown of GATA1 protein
Luciferase
GATA1a
GATA1b
Luciferase
GATA1a
GATAb
Luciferase
GATA1a
GATA1b
Luciferase
GATA1a
GATA1b
51 kDa
39 kDa
GATA1
GATA1 M20
1:1000
12hr
24hr
36hr
48hr
51 kDa
39 kDa
Loading

Figure 3.10. siRNA knockdown time-course study of GATA1. Two siRNAs directed against GATA1 were used: GATA1a and GATA1b. A: Knockdown of GATA1 at the mRNA level was quantified by quantitative PCR as described in the text. Bar chart shows the mRNA level of GATA1 remaining (y-axis) after siRNA transfection relative to luciferase siRNA transfection across the time points (x-axis) Blue bars indicate mRNA levels in the GATA1a condition while the pink bars indicate mRNA levels in the GATA1b condition. The error bars show the standard error of the mean between the three independent biological replicates. B: Knockdown of GATA1 at the protein level by densitometry of bands determined by immuno-detection of the relevant protein band on western blots. Bar chart shows the protein level of GATA1 remaining after siRNA transfection relative to luciferase siRNA transfection. Blue bars indicate protein levels in the GATA1a condition while pink bars indicate the protein levels remaining in the GATA1b condition. The error bars show the standard error of the mean between the three independent biological replicates. C: Western blot analyses of GATA1 protein knockdowns. Upper panel shows the bands detected by immuno-detection of the western blot with the GATA1 M20 Santa Cruz antibody. The lower panel shows the blot stained with Bradford reagent as a protein loading control. The arrow shows the predicted size of the GATA1 protein.

3.4.5.3 Knockdown of E2A

The knockdown of E2A was performed using two independent siRNAs (E2Aa and E2Ab) targeting regions of the coding sequence found in both the E12 and E47 transcript variants. The results of the time-course analysis are shown in Figure 3.11. In general, both siRNAs generated similar and substantial knockdowns at both the mRNA and protein level for both transcript variants, but the effect on E12 was different from that on E47. E12 mRNA levels in cells transfected with either of the siRNAs decreased to their lowest levels of approximately 30% the original mRNA level at the 12 hour time point and remained at similar levels at all subsequent time points. In contrast, E47 mRNA levels achieved their lowest levels (< 30% of original mRNA levels) at the 36 hour time point and then began to increase again by 48 hours.

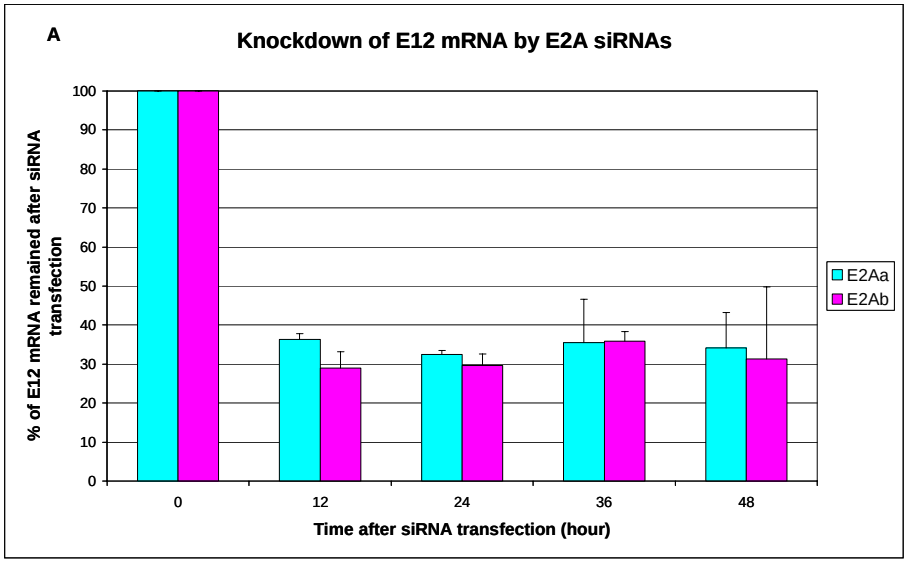
The knockdown at the protein level at the E12 and E47 isoforms could not be studied on the western blot due to issues associated with possible cross-reactivities of the E12 and E47 antibodies. There is no evidence from the antibody supplier showing that the two antibodies do not cross-react with both isoforms. The following interpretation is based on the assumption that the antibodies did not cross-react. The E2Ab siRNA produced a marked drop in E12 levels at the 12 hour time point, while the E2Aa siRNA had no effect after 12 hours. The largest knockdown effect of the E12 protein was achieved at the 24 hour time point for both siRNAs with only 15% of the protein remaining; subsequent time points showed a general increase in E12 protein levels. The E47 protein levels decreased gradually during the time course for both siRNAs and were at their lowest level (<30% of the original level) at 36 hours and had begun to increase substantially by 48 hours.

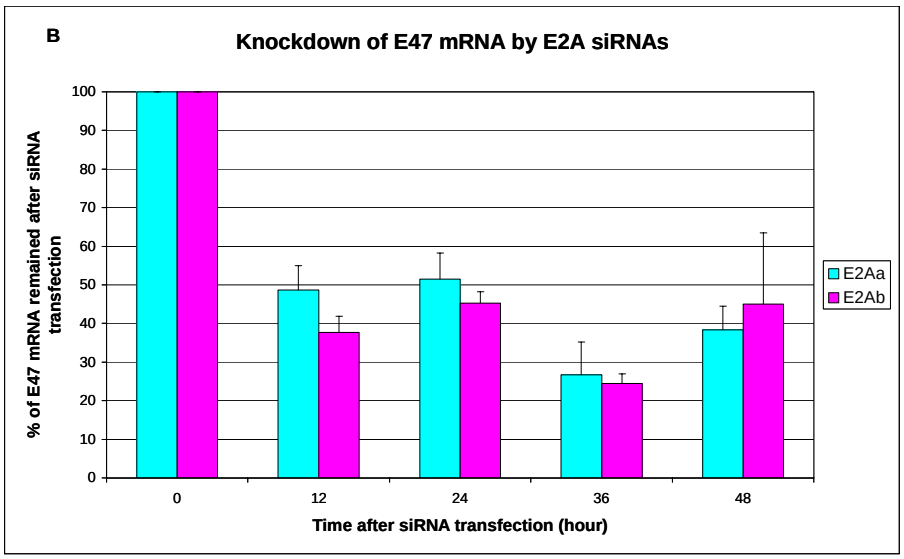
Assuming that the E12 and E47 antibodies did cross-react, western blots probed with either antibody would show the knockdown of E2A (the sum total of E12 and E47 isoforms). Based on this assumption, the greatest knockdown of E2A was achieved at the 24 hour time point for both

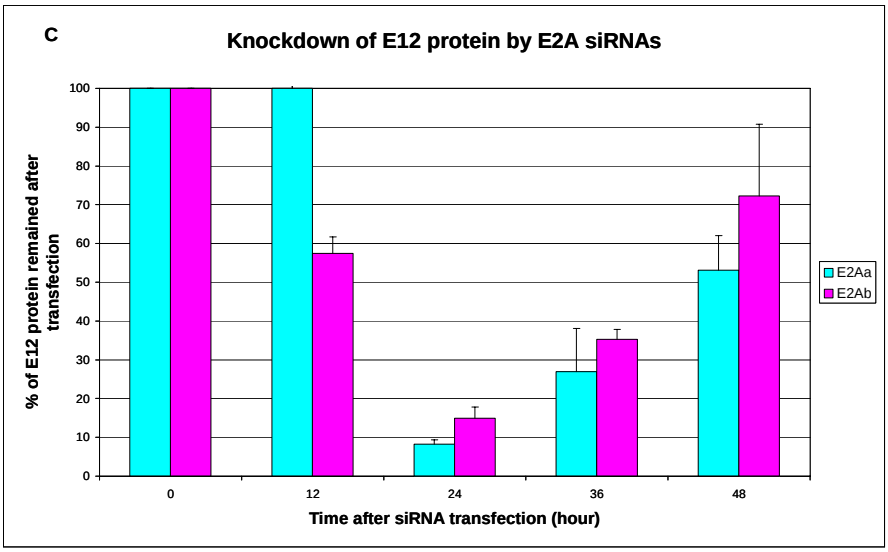
siRNAs with only 15% of the protein remaining using the E12 antibody. In contrast, the greatest knockdown of E2A was achieved at the 36 hour time point for both siRNAs with less than 30% of the protein remaining using the E47 antibody.

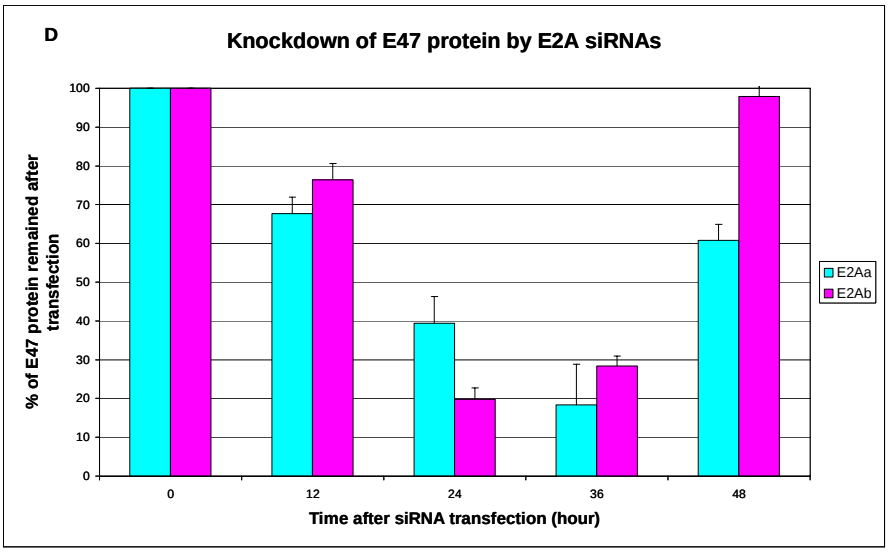
Both siRNAs E2Aa and E2Ab gave a knockdown of less than 30% protein remaining (on average of both siRNA) at the protein level at the 24 hour time point for the detection by both E12 and E47 antibodies. This time point was the earliest time point where both siRNAs were able to knock down the E2A protein to less than 30% remaining on average. Therefore, the 24 hour time point was chosen for subsequent analysis for both siRNAs.

The time-course experiment suggests that the half-life of the E2A mRNA was shorter than that of the E2A protein as a time lag between the mRNA and protein being knocked down was observed in both siRNAs. The mRNAs of E12 and E47 transcripts were reduced to less than 50% remaining at the 12 hour time point whereas the protein level of E2A remained at about 70% of its physiological level (on average). This suggests that the half-life of E2A mRNA appeared to be shorter than that of the protein.




B
Knockdown of E47 mRNA by E2A siRNAs
% of E47 mRNA remained after siRNA transfection
100
90
80
70
60
50
40
30
20
10
0
0
12
24
36
48
Time after siRNA transfection (hour)
E2Aa
E2Ab


C
Knockdown of E12 protein by E2A siRNAs
% of E12 protein remained after transfection
100
90
80
70
60
50
40
30
20
10
0
0
12
24
36
48
Time after siRNA transfection (hour)
E2Aa
E2Ab


D
Knockdown of E47 protein by E2A siRNAs
% of E47 protein remained after transfection
100
90
80
70
60
50
40
30
20
10
0
0
12
24
36
48
Time after siRNA transfection (hour)
E2Aa
E2Ab

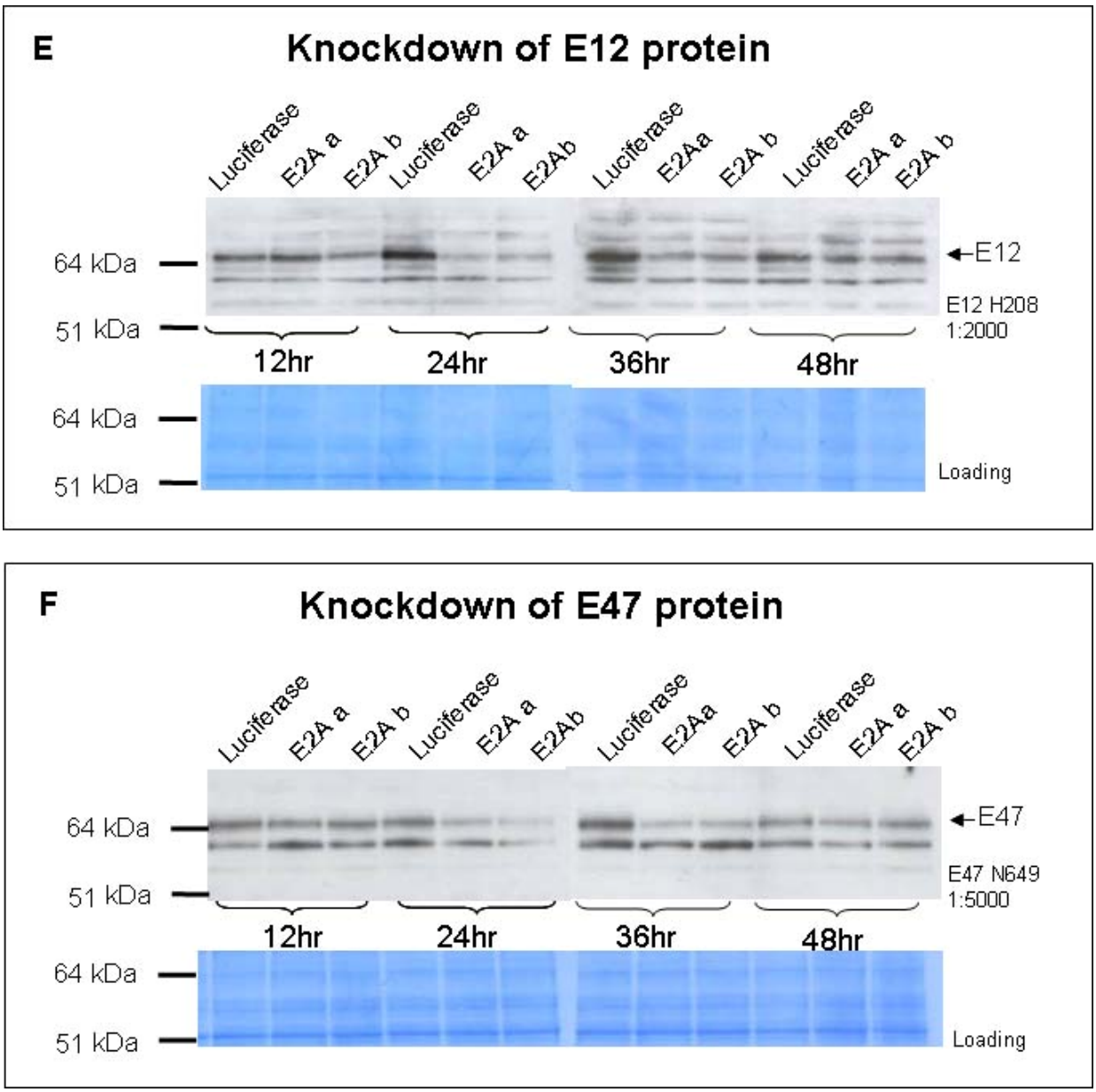


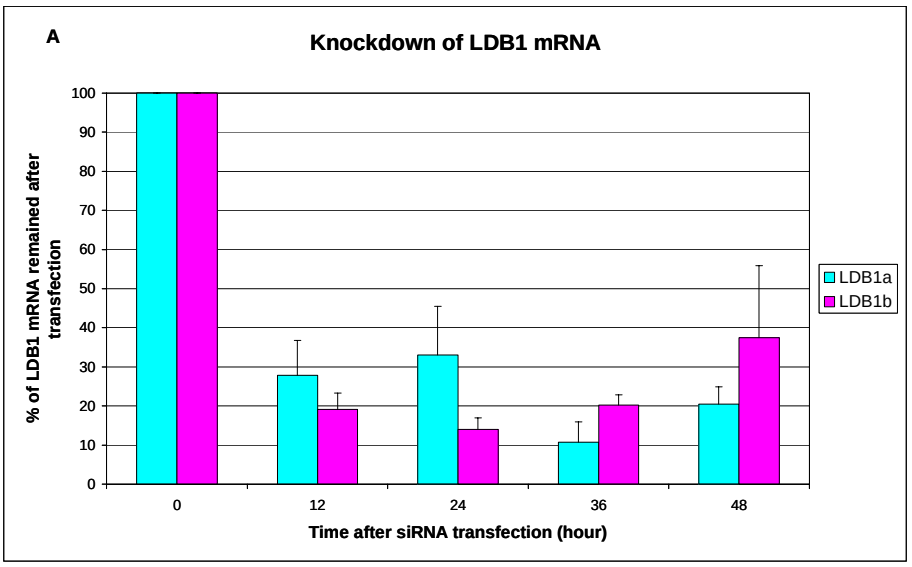
Figure 3.11. siRNA knockdown time-course study of E2A. Two siRNAs directed against E2A were used: E2Aa and E2Ab. A and B: Knockdown of E12 and E47 respectively at the mRNA level was quantified by quantitative PCR as described in the text. Bar chart shows the mRNA level of E12 or E47 remaining (y-axis) after siRNA transfection relative to luciferase siRNA transfection across the time points (x-axis) Blue bars indicate mRNA levels in the E2Aa condition while the pink bars indicate mRNA levels in the E2Ab condition. The error bars show the standard error of the mean between the three independent biological replicates. C and D: Knockdown of E12 and E47 respectively at the protein level by densitometry of bands determined by immuno-detection of the relevant protein band on western blots. Bar chart shows the protein level of E12 or E47 remaining after siRNA transfection relative to luciferase siRNA transfection. Blue bars indicate protein levels in the E2Aa condition while pink bars indicate the protein levels remaining in the E2Ab condition. The error bars show the standard error of the mean between the three independent biological replicates. E and F: Western blot analyses of E12 and E47 protein knockdowns respectively. Upper panel shows the bands detected by immuno-detection of the western blot with the E12 H208 and E47 N649 Santa Cruz antibody. The lower panel shows the blot stained with Bradford reagent as a protein loading control. The arrow shows the predicted size of the E2A protein.

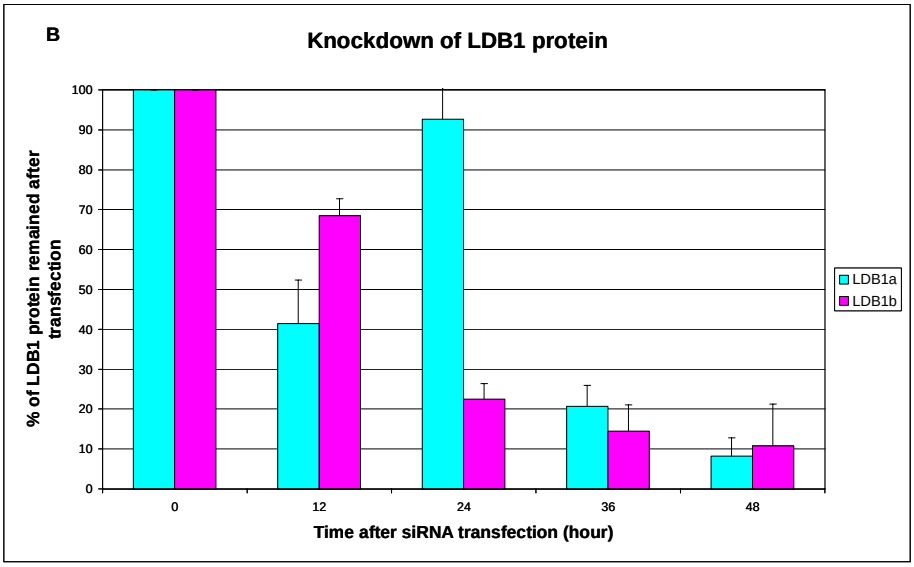
3.4.5.4 Knockdown of LDB1

The knockdown of LDB1 by two independent siRNAs (LDB1a and LDB1b) was studied through a 48 hour time-course experiment (Figure 3.12). Both siRNAs generated substantial knockdowns at both the mRNA and protein level. Overall, LDB1 mRNA levels in cells transfected with either of the siRNAs decreased dramatically to less than 20% of their original levels at the 12 hour time point. However, the mRNA levels in cells transfected with the LDB1a siRNA were at its lowest (only 20% of the original level of the mRNA remaining) at 36 hour and increased up to 30% by the 48 hour time point. The knockdown by the LDB1b siRNA was slightly different - approximately 35% of the original mRNA level was detected at the 24 hour time point and was further reduced to less than 10% at the 36 hour time point and increased up to 30% by the 48 hour time point. In contrast, the knockdown of both LDB1 isoforms was not significant at 12 hours in either siRNA condition (whereas only 20% of the original mRNA level was seen at the same time point). The protein knockdown was gradual for both siRNAs and reached a maximum knockdown to 10% of its original level at the 48 hour time point for both isoforms.

Both siRNAs LDB1a and LDB1b induced a substantial knockdown at the protein level with only 20% of the original remaining at the 36 hour time point. This time point was the earliest time point where both siRNAs were able to knock down the LDB1 protein to less than 30% remaining on average. Although a greater knockdown was achieved at the 48 hour time point, this time point was not considered as the optimal time point as an earlier time point is more desirable to reduce off-target effects. Therefore, the 36 hour time point was chosen for subsequent analysis.

Both LDB1 siRNAs were shown to induce knockdown to a similar level, however the maximum knockdown at the mRNA and protein levels was achieved at different time points. This suggests that the half-lives of mRNA and protein of LDB1 were different as a time lag between the mRNA and protein being knocked down was observed in both siRNAs. The mRNA of LDB1 was reduced to less than 20% remaining at the 12 hour time point whereas the protein level of LDB1 stayed at about 55% remaining on average for both siRNAs at the same time point. Also, the maximum knockdown at the mRNA level was achieved at the 36 hour time point whereas it was observed at the 48 hour time point at the protein level. This suggests that the half-life of the LDB1 mRNA was shorter than that of the protein.


A
Knockdown of LDB1 mRNA
% of LDB1 mRNA remained after transfection
100
90
80
70
60
50
40
30
20
10
0
0
12
24
36
48
Time after siRNA transfection (hour)
LDB1a
LDB1b


B
Knockdown of LDB1 protein
% of LDB1 protein remained after transfection
100
90
80
70
60
50
40
30
20
10
0
0
12
24
36
48
Time after siRNA transfection (hour)
LDB1a
LDB1b

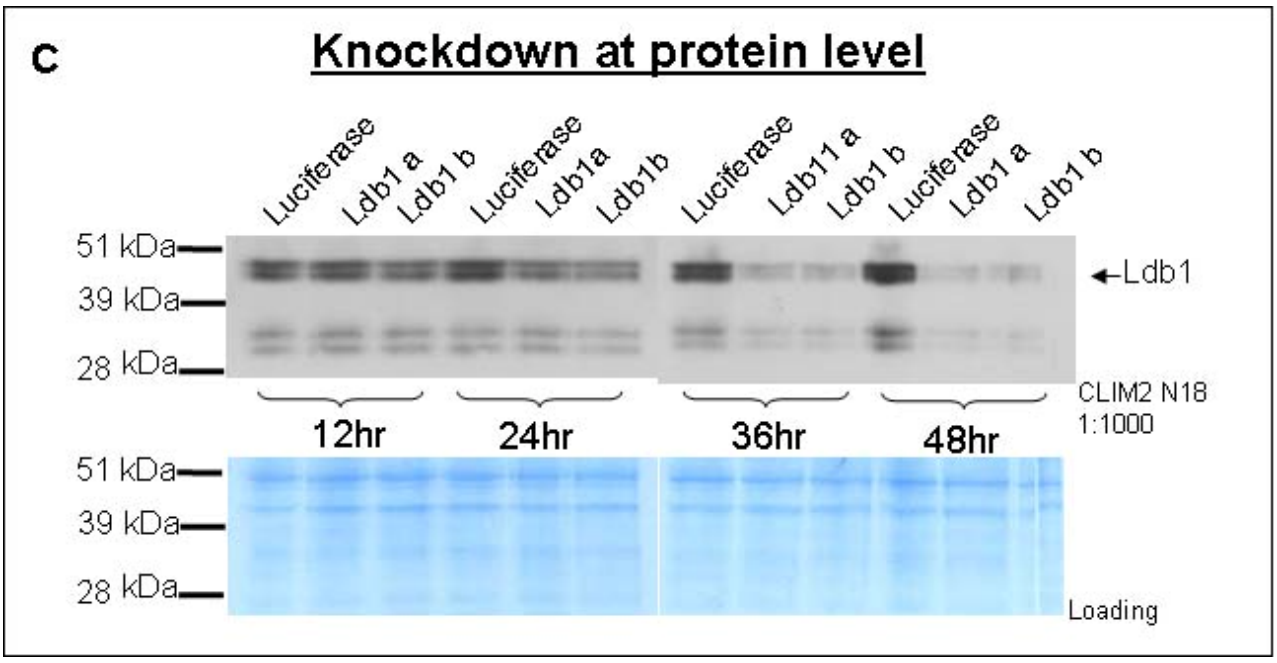

c
Knockdown at protein level
Luciferase
Ldb1 a
Ldb1 b
Luciferase
Ldb1a
Ldb1b
Luciferase
Ldb11 a
Ldb1 b
Luciferase
Ldb1 a
Ldb1 b
51 kDa
39 kDa
28 kDa
←Ldb1
CLIM2 N18
1:1000
12hr
24hr
36hr
48hr
51 kDa
39 kDa
28 kDa
Loading

Figure 3.12. siRNA knockdown time-course study of LDB1. Two siRNAs directed against LDB1 were used: LDB1a and LDB1b. A: Knockdown of LDB1 at the mRNA level was quantified by quantitative PCR as described in the text. Bar chart shows the mRNA level of LDB1 remaining (y-axis) after siRNA transfection relative to luciferase siRNA transfection across the time points (x-axis) Blue bars indicate mRNA levels in the LDB1a condition while the pink bars indicate mRNA levels in the LDB1b condition. The error bars show the standard error of the mean between the three independent biological replicates. B: Knockdown of LDB1 at the protein level by densitometry of bands determined by immuno-detection of the relevant protein band on western blots. Bar chart shows the protein level of LDB1 remaining after siRNA transfection relative to luciferase siRNA transfection. Blue bars indicate protein levels in the LDB1a condition while pink bars indicate the protein levels remaining in the LDB1b condition. The error bars show the standard error of the mean between the three independent biological replicates. C: Western blot analyses of LDB1 protein knockdowns. Upper panel shows the bands detected by immuno-detection of the western blot with the LDB1 CLIM2 N18 Santa Cruz antibody. The lower panel shows the blot stained with Bradford reagent as a protein loading control. The arrow shows the predicted size of the LDB1 protein.

3.4.5.5 Knockdown of LMO2

The knockdown of LMO2 by two independent siRNAs (LMO2a and LMO2b) was studied through a 48 hour time-course experiment (Figure 3.13). As no working antibodies were available (see section 3.4.4), only mRNA levels of LMO2 were monitored in the time-course experiments. Across the time-course, the two siRNAs behaved similarly although LMO2b siRNA generated a greater knockdown effect at all time points. The mRNA levels in cells transfected with LMO2b was similar (between 15-20% of the original mRNA level remained) at all the time points. The mRNA levels in cells transfected with LMO2a was slightly different with a similar mRNA level observed at the 12 hour, 24 hour and 36 hour time points (between 20 to 28%) and subsequently increasing to approximately 70% at the 48 hour time point.

Since the knockdown at the protein level could not be studied due to the lack of a working antibody, the choice of the optimal time point was based solely on the mRNA level. Both siRNAs LMO2a and LMO2b induced a significant knockdown at the mRNA level with only 20% of the original mRNA remaining on average at the 24 hour time point. This time point was the earliest time point where both siRNAs were able to knock down the LMO2 protein to less than 30% remaining on average. Although such knockdown was also observed at the 12 hour time point, this time point was not considered as the optimal time point for the reasons mentioned above. Therefore; the 24 hour time point was chosen for subsequent analysis.

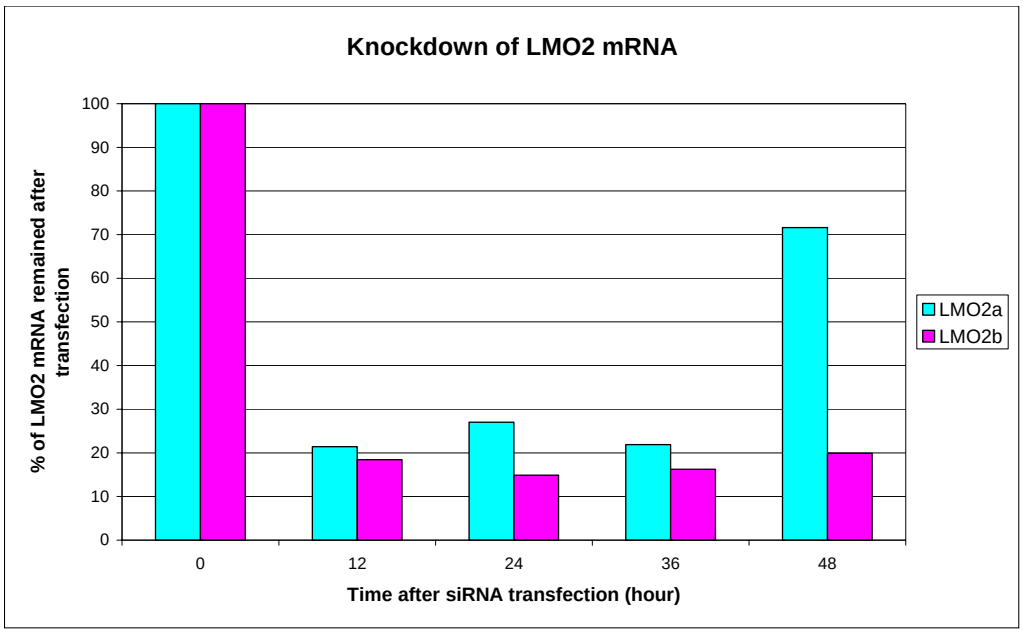


Figure 3.13. siRNA knockdown time-course study of LMO2. Two siRNAs directed against LMO2 were used: LMO2a and LMO2b. Knockdown of LMO2 at the mRNA level was quantified by quantitative PCR as described in the text. Bar chart shows the mRNA level of LMO2 remaining (y-axis) after siRNA transfection relative to luciferase siRNA transfection across the time points (x-axis) Blue bars indicate mRNA levels in the LMO2a condition while the pink bars indicate mRNA levels in the LMO2 condition. The error bars show the standard error of the mean between the three independent biological replicates.

3.5 Discussion

The work presented in this chapter describes the screening of working siRNA assays for five members of the SCL erythroid complex and the characterisation of these siRNAs in various studies. The siRNAs were shown to be delivered into K562 cells with high efficiency (>90% of cells contained the siRNA) by electroporation. Further characterisation of some of the siRNAs delivered to K562 cells demonstrated that effects on cell growth and morphology were as a result of electroporation and not due to the siRNA induction of the RNAi machinery or specific siRNA effects. Time-course knockdown experiments were performed to identify appropriate time points where the knockdown was greatest at the mRNA and proteins levels – these time points would serve as useful guides for further experiments aimed at using siRNA knockdowns to identify downstream targets of the SCL erythroid complex (SEC).

3.5.1 siRNA delivery

One of the most crucial factors affecting the efficacy of knockdown by siRNA is the efficiency of delivering siRNA into the cells. If the delivery efficiency is low, even a good siRNA cannot induce a high knockdown effect. Therefore; it is important that the delivery of the siRNAs is monitored to ensure that the strategy is optimised for subsequent experimental approaches which will use

knockdowns to address biological questions. Both electroporation and lipofection have been widely used in RNAi experiments, and electroporation by the Amaxa Nucleofector II system was used in this project due mainly to the high transfection efficiency which has been optimised by the supplier for K562 cells.

For similar reasons as discussed in the introduction to this chapter, fluorescein FITC-labelled siRNAs were used here to study the efficiency of delivery by the Amaxa Nucleofector II system. In contrast to fluorescence microscopy used in previously published studies, FACS analysis using an appropriate laser which can allow detection of FITC was used here. FACS was used in preference to microscopy because FITC emits a green fluorescence which is often difficult to distinguish from background fluorescence using microscopy. Furthermore, K562 cells, being a cell line which grows in suspension, requires the use of a cytospin to prepare microscope slides. Treatment of cells in this way can result in a high proportion of fragmented cells, making microscopic examination more difficult. Moreover, FACS analysis allows for the measurement of many more cells emitting fluorescence, when compared to microscopy, thus resulting in a more precise estimate of transfection efficiency.

Using FACS analysis, the transfection efficiency by a FITC-labelled siRNA, which showed a very high knockdown effect (GATA1a), was greater than 90% immediately after transfection. This efficiency diminished slightly after 24 hours - approx. 84% of cells carried the siRNA and the overall fluorescence levels emitted by the cells had also decreased. This is consistent with the transfection efficiency reported by Amaxa for K562 cells (www.amaxa.com) and reflects the knockdown efficiency of GATA1a, which is also over 90%. The reduction in % of cells emitting fluorescence after 24 hours was likely due to the presence of increased cell numbers in culture (due to on-going cell division), an ever-increasing proportion of which would not carry the labelled siRNA. The reduction in the fluorescence intensity emitted by the cells after 24 hours was due to the high photobleaching rate of FITC.

An alternative approach to study the transfection efficiencies of siRNAs would be to use quantum dots (QDs) which are highly photostable but relatively small fluorescent nanocrystals which are both brighter than conventional fluorescent dyes and easier to detect among *in vivo* background (Chan and Nie, 1998; Gao et al., 2004). QDs has been used to track RNAi by co-transfecting siRNA and QDs together into cells using lipofectamine (Chen et al., 2005). No chemical labelling with the QD is required for the siRNA during its synthesis, as QD/siRNA complexes are formed during transfection. This can be more cost efficient as well as preventing any undesirable effects of the chemical modification of siRNA on the delivery and induction of the RNAi pathways. To date, QD siRNAs have not been tested using delivery into cells by electroporation.

3.5.2 Effect of siRNA transfection on phenotypic changes

As mentioned in Chapter 1, induction of the RNAi pathway can induce non-specific effects including off-target effects, immune responses and saturation of the pathway itself (Chapter 1, section 1.3.1.3). Changes in growth patterns and cell morphologies upon transfection with siRNAs through time were also investigated to determine whether such changes were the result of specific knockdown effects of the protein of interest, or due to non-specific effects.

The cell morphology studies showed that electroporation of water, the control siRNA for luciferase and specific siRNA against GATA1 or E2A all induced changes in K562 cell morphology after 24 hours, resulting in the formation of small or big projections on the cell membranes and increased levels of multinucleated cells (the latter suggesting an increased rate of cell division may be occurring). These features were not present in cells which had not been electroporated. Furthermore, these features persisted at later time points (e.g. 48 hours after transfection). Nevertheless, it is clear from this data that all cells subjected to electroporation behaved in a similar way, suggesting that these morphological changes were not due to specific changes induced by the knockdown of the GATA1 or E2A mRNAs and proteins, or due to effects of introduction of siRNAs into K562 cells.

Similarly, the growth patterns of all cells which had been electroporated (with water, GATA1 and luciferase siRNAs) were different from the patterns observed with cells that had not been electroporated. The growth rate of the electroporated cells were significantly lower during the first 24 hours compared to their non-electroporated counterparts. Therefore, electroporation *per se* and not effects due to transfection with siRNA was the likely cause of such changes in growth rates. These growth rate effects appeared to diminish after the 24 hour time point, suggesting that the effect was transient, as one would expect given that the effect of electroporation would likely diminish as cells have had more time to recover in culture.

All of this data, taken together, suggests that transfection of siRNAs, regardless of whether they are against specific genes of interest or to luciferase controls, has no effect on cell morphology or growth pattern of K562 cells. Electroporation, however, does generate a stress response on the cells leading to visible changes in morphology and growth pattern. These results underlie the importance of comparing the effects of siRNA knockdowns with relevant controls for the cellular responses elicited by electroporation. Furthermore, by using a luciferase siRNA control, the effects of both electroporation and any generalised siRNA effects can also be taken into consideration when assessing the effects of siRNA knockdowns of specific genes of interest. The relevance of such controls are highlighted further in Chapters 4, 5 and 6, as they apply to our understanding of how

such effects can also be mediated at the molecular level and can have a bearing on our interpretation of data in both expression and ChIP-on-chip studies.

3.5.3 siRNA-induced knockdown of the SCL erythroid complex in time-course study

(i) Defining and quantitating knockdown levels for siRNA assays

During the screening of siRNAs for members of the SCL erythroid complex described in this Chapter, a knockdown efficiency of 70% (30% of original mRNA level) was considered to be the benchmark level (see introduction of this Chapter), above which siRNAs were considered to be working effectively enough to warrant their use in further aspects of this project. Two siRNAs were selected for each TF under study to reduce the chance of identifying off-targets in the subsequent expression analysis. In addition, time-course experiments monitoring the knockdown of each TF with the validated siRNAs were also performed. Determining the time points at which maximum knockdowns were achieved were crucial for the subsequent expression profiling analyses described in Chapter 4.

The knockdown levels for each siRNA assay were determined at both the mRNA and protein level for each TF using quantitative real time PCR (qPCR) and western blotting respectively. qPCR is a very sensitive and quantitative assay, although the SYBR green assays used in this study is less quantitative, but more economical, than Taqman assays. To circumvent any quantitation issues, accurate normalisation with internal housekeeping controls is required (Lupberger et al., 2002; Vandesompele et al., 2002). In this Chapter, β-actin, β-tubulin, GAPDH and RPL19 were used. These genes are normally highly and constitutively expressed in most tissues. Furthermore, by using more than one control, one can account for sample to sample variations in mRNA levels in some, but not all of the genes, which may affect accurate normalisation. Whilst other studies have used ribosomal RNAs (rRNAs) as normalisation controls, some researchers have demonstrated that rRNAs are not appropriate controls as there is an imbalance between rRNA and mRNA fractions and rRNA cannot truly reflect the mRNA levels (Vandesompele et al., 2002).

Although working antibodies were characterised for the majority of the TF under study, no antibodies were found to work in western blotting for LMO2. LMO2 is a very small protein (~18 kDa) which might be susceptible to degradation or denaturation during electrophoresis. In such cases, where it is not possible to use western blotting, immunofluorescence-based assays using microscopy or flow cytometry assays can be used to track the expression of the protein of interest. However, these assays are much more time-consuming and optimisation of assays is required – and this was beyond the scope of the work presented in this thesis. Also, accurate quantification cannot be easily performed, particularly for immunofluorescence-based assays.

The K562 cell line used in this study has the ability to spontaneously differentiate down the erythroid or megakaryocytic lineage in the absence of differentiating agents (Lozzio et al., 1981). Although the cell culture conditions were tightly monitored to avoid differentiation, it cannot be completely avoided. Thus, the siRNA knockdown efficiency across various replicates may differ due to self differentiation of K562 cells.

From the results obtained for all the time-course studies for the 5 TFs, it was demonstrated that the knockdown at the mRNA and protein levels do not necessarily agree with each other. For GATA1, the knockdown at both mRNA and protein levels were similar while the knockdown at the mRNA level was shown to be more rapid than that of the protein level for LDB1 and E2A. This suggests that mRNA and proteins half-lives may be different. Furthermore, the time points at which a maximum knockdown was obtained were not the same for all the siRNA assays. For instance, the maximum knockdown for LDB1 was obtained at a later time point from the other TFs, demonstrating that protein and mRNA half-lives also vary from gene to gene.

(i) The SCL knockdown

Two siRNAs were tested in the initial screening and both passed the criteria of selecting working siRNAs - i.e. reduction of mRNA level to 30% of its original level at the 24 hour time point. Both of the chosen siRNAs targeted exon 4 of the SCL gene. In the initial screening, SCLb only marginally passed the cut-off of 70% knockdown efficiency. Unfortunately, further characterisation in the time-course study showed that SCLb could not induce a sufficient knockdown at the protein level to 40% of its original level at 24 hour, 36 hour and 48 hour time points except at the 12 hour time point. One possible reason to explain this discrepancy is that the K562 cells may have undergone a degree of spontaneous differentiation that changed the expression level of SCL (which is known to vary during myeloid differentiation). This may have affected the degree to which SCL could be silenced by the SCLb siRNA. Whilst the 12 hour time point did show the appropriate level of knockdown for SCLb, it could not be chosen as the optimal time point due to the various reasons mentioned in section 3.4.5. Thus, no time points were shown to be suitable for further characterisation for this siRNA. One possible solution would have been to test more siRNAs against SCL and select another one which satisfied the required selection criteria. However, due to the time constraint for this project, this additional screening could not be performed. Thus, SCLb siRNA was not used in the gene expression profiling experiments described in Chapter 4.

An SCL knockdown to less than 10% of its original protein level was observed at 24 hours after SCLa siRNA transfection. For this reason, knockdown samples at this time point were chosen for

subsequent analysis by gene expression profiling studies described in Chapter 4. The knockdown at the mRNA level generated a reduction to approximately 30% of its original level at this time point which is higher than that at the protein level. This might be because the mRNA of SCL has considerably lower turn-over rate and less susceptible to degradation inside the cells than the SCL protein. This further illustrates the requirement to observe both mRNA and protein levels during knockdown, as the levels of one do not necessarily marry up with the levels of the other.

(ii) The GATA1 knockdown

The selection of optimal time point for GATA1 in the gene expression profiling experiment was comparatively simple. Both GATA1a and GATA1b siRNAs generated the maximum knockdown for both mRNA and protein levels at the 24 hour time points. The knockdown efficiency of these two siRNAs was high - with close to 0% of the protein remaining and 10% of the mRNA remaining at the 24 hour time point. Furthermore, the mRNA and protein knockdown levels correlated with each other for GATA1. The mRNA and protein levels were reduced to similar levels at all of the time points (within 10% of each other) and the maximum knockdown was observed at the same time point for mRNA and protein. This may be due to the fact that the mRNA and protein of GATA1 have similar stabilities.

(iii) The E2A knockdown

Studying the knockdown of E2A was particularly challenging for a number of reasons. Firstly, the E2A gene produces two mRNA transcript variants, E12 and E47, both of which encode functional proteins of similar molecular masses and which only differ in the amino acid content encoded by a single exon. Because of this, it was particularly difficult to choose siRNAs which could knockdown one variant but not the other. Secondly, monitoring knockdowns of the two protein isoforms was further complicated by the fact that polyclonal antibodies for these variants were most likely to cross-react.

For these reasons, two siRNAs against a region common to both variants were selected in the screening of working siRNA assays. This ensured that both variants were targeted. However, there was no guarantee that one may be silenced more than the other even with these siRNAs. In fact, from the time-course experiments, the knockdown of E12 and E47 appeared different across time. At the mRNA level, a maximum knockdown was achieved at the 24 hour time point for E12 and at the 36 hour time point for E47. The differences in the protein level of these two variants were shown to be hard to monitor due to the possible cross-reactivity of the antibodies. For subsequent gene expression profiling experiments, the 24 hour time point was chosen because a significant

knockdown at the protein level (with less than 30% protein remaining on average of both siRNAs) was induced as shown by the detection with both E12 and E47 antibodies. Although evidence was provided that E2A (or its isoform E47) showed a maximum knockdown for the protein at 36 hours, it could be argued that this time point would also satisfy the selection criteria for use in expression profiling. However, using the earlier 24 hour time point is always more desirable in such studies to avoid issues related to the induction of the innate immune response.

(iv) The LDB1 knockdown

The knockdown of LDB1 was another example where the mRNA and protein levels did not correlate. The maximum knockdowns at the mRNA and protein levels were achieved at different time points (36 hour time point for mRNA and 48 hour time point for protein). Furthermore, the reduction of mRNA levels was shown to be more rapid than at the protein level. The mRNA level dramatically reduced to 20% of its original level after 12 hours whereas the protein level only reduced to 50% of its original level initially and gradually reduced to less than 10% remaining at later time points. For subsequent expression profiling, the 36 hour time point was used as both siRNAs were able to knock down the LDB1 protein to less than 30% remaining on average. The 48 hour time point was not chosen even though the maximum protein knockdown was observed. This is again because an earlier time point with significant knockdown is more desirable to avoid non-specific effects.

(v) The LMO2 knockdown

Without a working antibody for LMO2 in western analysis, it was more difficult to identify a time point after siRNA transfection for subsequent expression profiling. From the time-course experiment described in this Chapter, it was assumed that a significant knockdown at the protein level was achieved at the optimal time point for mRNA knockdown. Therefore, the 24 hour time point after siRNA transfection was chosen. However, given that evidence was provided in this Chapter that mRNA and protein levels do not always correlate during knockdown, and that this effect can be gene-specific, there is no way of knowing whether the LMO2 protein was knocked down to appreciable levels at this time point. Thus, expression profiling data from the LMO2 knockdown experiments, described in Chapter 4, must be considered with this in mind. Further validation of LMO2 antibodies which perform well in western analysis would be required to resolve this issue.

3.6 Conclusions

The work presented in this Chapter demonstrated that RNAi is a relatively straightforward technique to knockdown gene expression of specific genes, provided that one is prepared to perform

the appropriate controls and develop assays to monitor both mRNA and protein levels across time courses. That said, the development of siRNA assays for members of the SCL erythroid complex has thus provided a means to elucidate the transcriptional targets of this complex in subsequent Chapters of this thesis.

Bose, F., Fugazza, C., Casalgrandi, M., Capelli, A., Cunningham, J.M., Zhao, Q., Jane, S.M., Ottolenghi, S., and Ronchi, A. (2006). Functional interaction of CP2 with GATA-1 in the regulation of erythroid promoters. Mol Cell Biol *26*, 3942-3954.
Chan, W.C., and Nie, S. (1998). Quantum dot bioconjugates for ultrasensitive nonisotopic detection. Science (New York, NY *281*, 2016-2018.
Chen, A.A., Derfus, A.M., Khetani, S.R., and Bhatia, S.N. (2005). Quantum dots to monitor RNAi delivery and improve gene silencing. Nucleic acids research *33*, e190.
Choi, I., Cho, B.R., Kim, D., Miyagawa, S., Kubo, T., Kim, J.Y., Park, C.G., Hwang, W.S., Lee, J.S., and Ahn, C. (2005). Choice of the adequate detection time for the accurate evaluation of the efficiency of siRNA-induced gene silencing. Journal of biotechnology *120*, 251-261.
Corey, D.R. (2007). Chemical modification: the key to clinical application of RNA interference? The Journal of clinical investigation *117*, 3615-3622.
De Gobbi, M., Anguita, E., Hughes, J., Sloane-Stanley, J.A., Sharpe, J.A., Koch, C.M., Dunham, I., Gibbons, R.J., Wood, W.G., and Higgs, D.R. (2007). Tissue-specific histone modification and transcription factor binding in alpha globin gene expression. Blood *110*, 4503-4510.
Echeverri, C.J., Beachy, P.A., Baum, B., Boutros, M., Buchholz, F., Chanda, S.K., Downward, J., Ellenberg, J., Fraser, A.G., Hacohen, N., *et al.* (2006). Minimizing the risk of reporting false positives in large-scale RNAi screens. Nature methods *3*, 777-779.
Elbashir, S.M., Harborth, J., Weber, K., and Tuschl, T. (2002). Analysis of gene function in somatic mammalian cells using small interfering RNAs. Methods (San Diego, Calif *26*, 199-213.
Fedorov, Y., Anderson, E.M., Birmingham, A., Reynolds, A., Karpilow, J., Robinson, K., Leake, D., Marshall, W.S., and Khvorova, A. (2006). Off-target effects by siRNA can induce toxic phenotype. RNA (New York, NY *12*, 1188-1196.
Gao, X., Cui, Y., Levenson, R.M., Chung, L.W., and Nie, S. (2004). In vivo cancer targeting and imaging with semiconductor quantum dots. Nat Biotechnol *22*, 969-976.
Holen, T., Amarzguioui, M., Wiiger, M.T., Babaie, E., and Prydz, H. (2002). Positional effects of short interfering RNAs targeting the human coagulation trigger Tissue Factor. Nucleic acids research *30*, 1757-1766.
Horak, C.E., Mahajan, M.C., Luscombe, N.M., Gerstein, M., Weissman, S.M., and Snyder, M. (2002). GATA-1 binding sites mapped in the beta-globin locus by using mammalian chIp-chip analysis. Proc Natl Acad Sci U S A *99*, 2924-2929.
Hsieh, A.C., Bo, R., Manola, J., Vazquez, F., Bare, O., Khvorova, A., Scaringe, S., and Sellers, W.R. (2004). A library of siRNA duplexes targeting the phosphoinositide 3-kinase pathway: determinants of gene silencing for use in cell-based screens. Nucleic acids research *32*, 893-901.
Huang, D.Y., Kuo, Y.Y., and Chang, Z.F. (2005). GATA-1 mediates auto-regulation of Gfi-1B transcription in K562 cells. Nucleic Acids Res *33*, 5331-5342.
Huang, D.Y., Kuo, Y.Y., Lai, J.S., Suzuki, Y., Sugano, S., and Chang, Z.F. (2004). GATA-1 and NF-Y cooperate to mediate erythroid-specific transcription of Gfi-1B gene. Nucleic Acids Res *32*, 3935-3946.
Huo, X.F., Yu, J., Peng, H., Du, Z.W., Liu, X.L., Ma, Y.N., Zhang, X., Zhang, Y., Zhao, H.L., and Zhang, J.W. (2006). Differential expression changes in K562 cells during the hemin-induced erythroid differentiation and the phorbol myristate acetate (PMA)-induced megakaryocytic differentiation. Molecular and cellular biochemistry *292*, 155-167.
Jackson, A.L., Burchard, J., Leake, D., Reynolds, A., Schelter, J., Guo, J., Johnson, J.M., Lim, L., Karpilow, J., Nichols, K., *et al.* (2006). Position-specific chemical modification of siRNAs reduces "off-target" transcript silencing. RNA (New York, NY *12*, 1197-1205.

Judge, A.D., Sood, V., Shaw, J.R., Fang, D., McClintock, K., and MacLachlan, I. (2005). Sequence-dependent stimulation of the mammalian innate immune response by synthetic siRNA. Nat Biotechnol *23*, 457-462.
Lahlil, R., Lecuyer, E., Herblot, S., and Hoang, T. (2004). SCL assembles a multifactorial complex that determines glycophorin A expression. Mol Cell Biol *24*, 1439-1452.
Lazrak, M., Deleuze, V., Noel, D., Haouzi, D., Chalhoub, E., Dohet, C., Robbins, I., and Mathieu, D. (2004). The bHLH TAL-1/SCL regulates endothelial cell migration and morphogenesis. J Cell Sci *117*, 1161-1171.
Lecuyer, E., Herblot, S., Saint-Denis, M., Martin, R., Begley, C.G., Porcher, C., Orkin, S.H., and Hoang, T. (2002). The SCL complex regulates c-kit expression in hematopoietic cells through functional interaction with Sp1. Blood *100*, 2430-2440.
Lozzio, B.B., and Lozzio, C.B. (1977). Properties of the K562 cell line derived from a patient with chronic myeloid leukemia. Int J Cancer *19*, 136.
Lozzio, B.B., Lozzio, C.B., Bamberger, E.G., and Feliu, A.S. (1981). A multipotential leukemia cell line (K-562) of human origin. Proceedings of the Society for Experimental Biology and Medicine Society for Experimental Biology and Medicine (New York, NY *166*, 546-550.
Lupberger, J., Kreuzer, K.A., Baskaynak, G., Peters, U.R., le Coutre, P., and Schmidt, C.A. (2002). Quantitative analysis of beta-actin, beta-2-microglobulin and porphobilinogen deaminase mRNA and their comparison as control transcripts for RT-PCR. Molecular and cellular probes *16*, 25-30.
Marques, J.T., Devosse, T., Wang, D., Zamanian-Daryoush, M., Serbinowski, P., Hartmann, R., Fujita, T., Behlke, M.A., and Williams, B.R. (2006). A structural basis for discriminating between self and nonself double-stranded RNAs in mammalian cells. Nat Biotechnol *24*, 559-565.
Morrison, T.B., Weis, J.J., and Wittwer, C.T. (1998). Quantification of low-copy transcripts by continuous SYBR Green I monitoring during amplification. BioTechniques *24*, 954-958, 960, 962.
Muntean, A.G., and Crispino, J.D. (2005). Differential requirements for the activation domain and FOG-interaction surface of GATA-1 in megakaryocyte gene expression and development. Blood *106*, 1223-1231.
Palomero, T., Odom, D.T., O'Neil, J., Ferrando, A.A., Margolin, A., Neuberg, D.S., Winter, S.S., Larson, R.S., Li, W., Liu, X.S., *et al.* (2006). Transcriptional regulatory networks downstream of TAL1/SCL in T-cell acute lymphoblastic leukemia. Blood *108*, 986-992.
Rylski, M., Welch, J.J., Chen, Y.Y., Letting, D.L., Diehl, J.A., Chodosh, L.A., Blobel, G.A., and Weiss, M.J. (2003). GATA-1-mediated proliferation arrest during erythroid maturation. Mol Cell Biol *23*, 5031-5042.
Semizarov, D., Frost, L., Sarthy, A., Kroeger, P., Halbert, D.N., and Fesik, S.W. (2003). Specificity of short interfering RNA determined through gene expression signatures. Proc Natl Acad Sci U S A *100*, 6347-6352.
Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., and Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome biology *3*, RESEARCH0034.
Wadman, I.A., Osada, H., Grutz, G.G., Agulnick, A.D., Westphal, H., Forster, A., and Rabbitts, T.H. (1997). The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. Embo J *16*, 3145-3157.
Welch, J.J., Watts, J.A., Vakoc, C.R., Yao, Y., Wang, H., Hardison, R.C., Blobel, G.A., Chodosh, L.A., and Weiss, M.J. (2004). Global regulation of erythroid gene expression by transcription factor GATA-1. Blood *104*, 3136-3147.

Chapter 4

Expression profiling analyses of siRNA knockdowns of the SCL erythroid complex

4.1 Introduction

As discussed in Chapter 1, in order to identify downstream targets of transcription factors, one of the key analyses is to identify gene expression changes which occur when you perturb the function of a transcription factor of interest in a biological system. The siRNA knockdown studies described in Chapter 3 provide a means for perturbing the function of transcription factors of interest. With the characterisation of siRNAs for each transcription factor in the SCL erythroid complex in time-course experiments, the optimal time points for subsequent perturbation studies were determined. Thus, further analyses to identify downstream target genes using microarray gene expression analyses are described in this chapter.

4.1.1 Information generated using expression profiling of perturbation of transcription factors

A. Direct and indirect targets

Studying where transcription factor binds in the genome only allows us to determine the direct target genes they regulate - these are referred to as the primary targets of a particular transcription factor. However, in complex transcriptional pathways or networks, regulation can be achieved at many levels. For example, one transcription factor may regulate another transcription factor, and in turn, this transcription factor may regulate a third, and so on. Studying the direct binding by a transcription factor only reveals the first level of interactions between the transcription factor and its targets. Whole genome gene expression profiling of a transcription factor perturbation, on the other hand, enables us to identify both direct target genes regulated by the transcription factor and as well as other downstream genes regulated at subsequent levels (so-called indirect targets) (Figure 4.1). This is because perturbations at any one point in the network can affect the entire cascade of transcriptional events occurring further down the network of interactions.

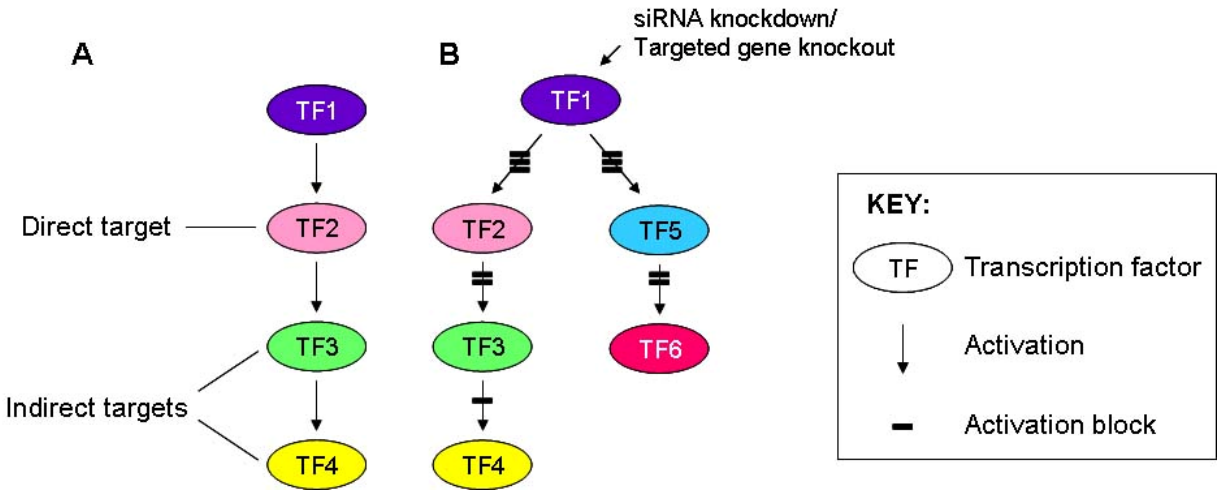


Figure 4.1. Direct and indirect target genes regulated by transcription factors. Panel A: illustration of direct and indirect target genes regulated by transcription factor 1 (TF1). TF2 is the direct target gene of TF1 while TF3 and TF4 are indirect target genes regulated by TF1. Panel B: changes in regulation of direct and indirect target genes when TF1 is silenced by siRNA knockdown or targeted gene knockout. The dash illustrates an activation block and the number of dashes describes the degree of activation block after TF1 is silenced. Typically, the activation block of the direct target genes is the highest (as demonstrated by 3 dashes) whilst that of the indirect target genes is lower (as demonstrated by 1-2 dashes).

B. Mode of regulation

Transcription factor-binding studies such as ChIP-on-chip allow us to study where the transcription factor binds but it does not directly provide information on how this binding event is impinging on the expression of its target gene. However, expression profiling allows one to determine whether a target gene is being activated or repressed by the transcription factor binding event, or whether the binding of the transcription factor has no immediate effect on gene expression. In the case of the latter, the binding of a transcription factor to the regulatory regions of genes may not induce or suppress the expression of a target gene – quite often, the binding of a transcription factor results in a "poised" state of the target for activation or repression later in a developmental programme, when other transcription factors or chromatin-remodelling factors required for regulation are expressed (Chapter 1, Section 1.1.2.5).

4.1.2 Expression profiling studies of the SCL erythroid complex in literature

Several studies have addressed the regulation of SCL and GATA1 target genes in high-throughput assays using expression microarrays. Palomero et al. (2006) delineated downstream targets of SCL in T-cell acute lymphoblastic leukaemia (T-ALL) where SCL is over-expressed due to translocation (Chapter 1, section 1.4.2.1 F) (Palomero et al., 2006). Genome-wide expression profiles of SCL-expressing and non-expressing human T-ALL samples were compared using Affymetrix U133 arrays to identify putative target genes induced by SCL. Lin and Aplan (2007) studied the changes

in expression in the mouse genome in thymic tumors from precursor T-cell lymphoblastic lymphoma/leukaemia (pre-T LBL) derived from transgenic mouse overexpressing SCL, LMO1 and NHD13 (Lin and Aplan, 2007). In a very recent study by Landry et al. (2008), a Nimblegen mouse 60-mer oligonucleotide expression microarray platform was used to study the changes in expression after the reintroduction of SCL into $SCL^{-/-}$ mouse yolk sac. This study identified RUNX1, a transcription factor required for definitive haematopoiesis (Landry et al., 2008), as a target of SCL. Welch et al. (2004) studied the expression changes in a sub-set of mouse genes using an Affymetrix GeneChip array before and after the induction of GATA1 expression in the GATA1-null erythroblast cell line G1E-ER4 (Welch et al., 2004). A number of genes were identified which were either up-regulated or down-regulated and both rapid and delayed responses were demonstrated. Affymetrix mouse expression arrays were also used to profile the expression patterns of wild type and GATA1-deficient murine megakaryocytes (Muntean and Crispino, 2005).

While the studies mentioned above described the role of SCL and/or GATA1 in leukaemia, early haematopoiesis or myeloid cells, none of them addressed the role of these transcription factors in regulating genes during erythroid development. In fact, few well characterised downstream targets genes of SCL and GATA1 in erythroid cells have been described in the published literature (Chapter 1, section 1.4.2.1 and 1.4.2.2). Furthermore, downstream targets of E2A, LMO2 and LDB1 in erythroid cells have thus far not been reported. Therefore, genome-wide scale analyses of the five transcription factors in the SCL erythroid complex studied here are necessary in order to have a more complete understanding of their target gene repertoire and roles in gene expression during erythroid development.

4.1.3 The Affymetrix GeneChip expression array

Many methods can be used to study the expression of genes as summarised in Chapter 1, section 1.3.2. Depending on the scale and accuracy required for a particular experimental system, these methods have different strength and weaknesses. For the study of downstream regulation by a particular transcription factor during perturbation, analyses by quantitative PCR or other low-throughput methods can be time-consuming and they often require *a priori* knowledge of the genes of interest. Thus, some important target genes may be excluded in the analyses. Therefore, for identifying targets of transcription factors, genome-wide analyses are desirable because they provide unbiased views of gene expression programmes. To this end, genome-wide profiling by microarrays is a rapid method to study all possible gene expression outputs (depending on the genome coverage of the microarray) - although there can still be biases in the genes represented on such platforms. At the time the project described in this thesis was initiated, whole genome expression microarrays were widely used to analyse gene expression outputs obtained from gene

perturbation studies (i.e., the work pre-dates the development of massively parallel sequencing-based methods).

In the work described in this Chapter, GeneChip expression arrays produced by Affymetrix were used. The GeneChip probe arrays generated by Affymetrix use a combination of photolithography and combinatorial chemistry in a series of cycles to construct arrays of oligonucleotides (Singh-Gasson et al., 1999). A glass substrate is coated with linkers containing photolabile protecting groups. This glass substrate is then covered with a mask which exposes selected portions of the probe array to ultraviolet light. Upon illumination, the photolabile protecting groups are removed at the exposed regions enabling selective nucleotide addition to the surface. The nucleotides added at each step also contains light-sensitive protecting group. Different masks are applied and the cycle of illumination and chemical coupling is repeated until the probes reach their full length (25 nucleotides). In the end, a specific set of oligonucleotide probes synthesised at particular known locations on the array are generated.

The GeneChip arrays contain a large number of highly specific probe sets representing each gene (Figure 4.2). Such specificity is very important when measuring the expression of two very similar genes. Within each probe set, a gene is represented by millions of copies of eleven probe pairs (oligos) of 25 bp which are found throughout the mRNA sequence of the gene. The use of multiple probes generates high sensitivity and reproducibility while reducing background noise. A probe pair contains two probes. Probes that are perfectly complementary to the target sequence, called Perfect Matches (PM), are intended to measure mainly specific hybridisation. A second set of probes identical to PM except for a single nucleotide in the centre of the probe sequence (the 13^{th} nucleotide), called Mismatches (MM), are intended to quantify non-specific hybridisation. A PM and its corresponding MM constitutes a probe pair. Such PM and MM probes are essential elements for eliminating effect of non-specific binding.

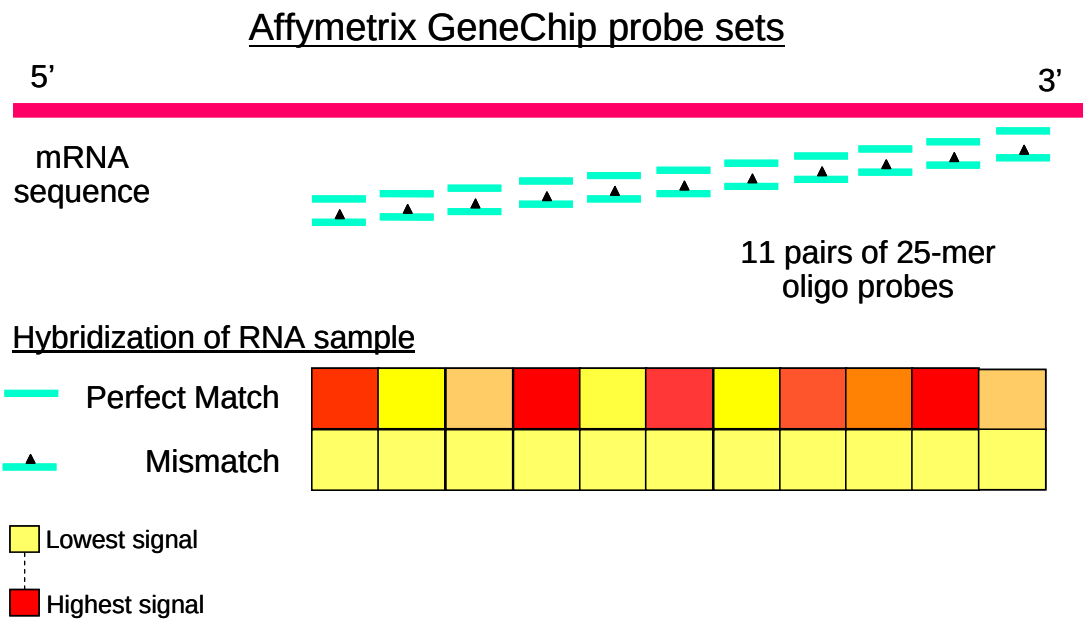


Figure 4.2. The Affymetrix expression GeneChip probe sets. 11 pairs of 25-mer oligo probes were designed for an mRNA sequence. Each probe pair includes the perfect match probe and the mismatch probe where the middle nucleotide is replaced by a different one. During hybridisation, if the RNA samples contain fragments matching the probe sets, they will generate a signal with the perfect match probes, while no or very low signals will be detected for the mismatch probes.

The Affymetrix GeneChip expression array system is a one-colour microarray system. In a one-colour array, control and experimental samples are hybridised onto different arrays, detected with the same fluorescent dye, and comparisons are made across different hybridisations. The Affymetrix GeneChip has standard and optimised protocols for sample manipulation and hybridisation (Figure 4.3). To perform hybridisation, total RNA or mRNA extracted from the cell or tissues of interest is first reverse-transcribed using a T7-oligo(dT) promoter primer to generate double-stranded cDNA. The cDNA then undergoes an *in vitro* transcription (IVT) reaction in the presence of T7 RNA polymerase and biotinylated ribonucleotides to generate biotin-labelled complementary RNAs (cRNAs). The biotinylated cRNAs are fragmented (to optimise target-probe hybridisation kinetics) and hybridised onto the probe array. The hybridised probe array is stained with a streptavidin phycoerythrin (PE) conjugate and scanned. The PE conjugate is excited by laser and emits fluorescence for detection.

The GeneChip Human Genome U133 Plus 2.0 array provides a comprehensive coverage of protein coding genes the human genome. This chip includes 54 000 probe sets (11 in each set) representing over 47 000 human transcripts and variants, all of which are analysed in a single hybridisation. The sequences from which the probe sets were derived were selected from the GeneBank, dbEST and RefSeq databases and the probe sets themselves have been annotated onto the human genome sequence.

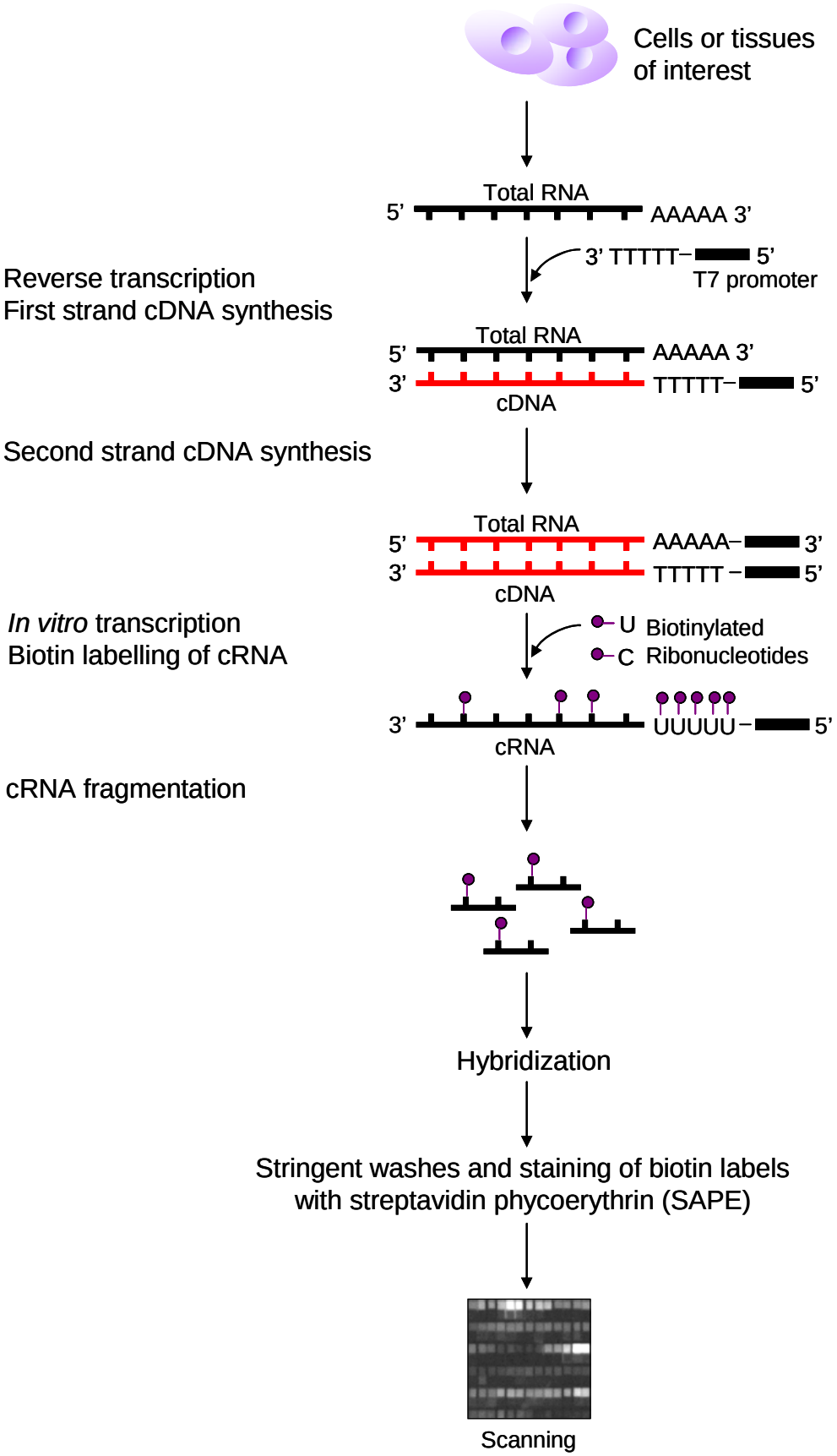


Figure 4.3. Target labelling and hybridisation of Affymetrix GeneChip arrays. Total cellular mRNA samples from the cells or tissues of interest are first reverse transcribed to generate double stranded cDNA with a T7 promoter. Complementary RNAs (cRNA) are generated by *in vitro* transcription with biotin-labelled ribonucleotides. The cRNAs are fragmented and hybridised on the array (please see text for details).

The use of commercial microarrays has many advantages over in-house custom-made microarrays. Firstly, for large-scale genome-wide analysis, generating in-house arrays is very time-consuming and requires a well developed informatics and array manufacture pipeline (which is not always available in academic laboratories). Thus, commercial arrays provide a widely available "off-the-shelf'' alternative. Secondly, commercial arrays are usually tested, validated and quality-controlled by both academic and commercial sources. Thirdly, target preparation and hybridisation protocols are well-established and usually require no further optimisation.

4.1.4 Microarray data analyses

Microarray experiments, regardless of whether they are one-colour or two-colour experiments, involve the measurement of the expression levels of a large number of genes in only a few replicate samples, given that microarrays are expensive and sometimes the biological samples are limiting. Developing appropriate statistical techniques to determine which changes are relevant is thereby very important. Typically, microarray analyses involve five main parts which are discussed below: quantitation, normalisation, inferential statistics, descriptive statistics and data mining.

A. Data processing methods (Quantitation)

Quantitation is the process of measuring the fluorescence intensity of spots or probes on the array while correcting it against the background intensity - which is another source of measurable fluorescence on the image.

Three different ways of processing and measuring probe set intensities on Affymetrix arrays have been developed, namely Affymetrix Microarray Suite v.5 (MAS5) (Affymetrix), robust multichip average (RMA) (Irizarry et al., 2003b) and GC-RMA (Wu and Irizarry, 2005). MAS5 was developed by Affymetrix where the weighted average of the 2% of probes having the lowest intensities was selected as background. It utilises the mismatch probe signals to adjust the perfect match intensity. For RMA analysis, each array is assumed to have a common mean background and the mismatch probes are ignored. GC-RMA is a modified version of RMA which models probe intensity as a function of GC-content. Comparison between the MAS5 and RMA softwares indicated that RMA has better precision to detect low expressing genes and has higher specificity and sensitivity for detecting differential expression (Irizarry et al., 2003a). In addition, GC-RMA was shown to over-correct the G+C content within probe sets whereas RMA introduce less bias than both MAS5 and GC-RMA (Siddiqui et al., 2006).

B. Normalisation

Normalisation is the process of removing systematic bias in the data across different samples while preserving the variation in gene expression that occurs because of biologically relevant changes in

transcription. Normalisation is also essential to allow the comparison of gene expression across multiple microarray experiments.

A basic assumption of the normalisation process is that the average gene does not change in an experiment. In the global normalisation procedure, two main steps are involved: scaling and centering. In scaling, the intensity for all the gene expression measurements in one channel for two-colour arrays or one array for single-colour array are multiplied by a constant factor so that the mean measurement equals to one. In centering, the intensity of the measurements is centered to ensure that the mean and the standard deviations of all the distributions are equal. Other normalisation procedures include normalising the measurements to some house-keeping genes e.g. GAPDH and β-actin but this is based on the assumption that the expressions of these genes do not change across samples.

C. Determining Relevant Expression Differences (Inferential statistics)

Determination of the genes which are differentially expressed between two RNA samples is one of the most important yet difficult issues associated with high-throughput microarray analyses. A variety of procedures can be applied to extract the most biologically relevant and significant expression differences. A few examples of ways of determining these significant differences are described below:

- Fold change

The ratios of signal intensity of a gene between the experimental condition and the control conditions are calculated. A ratio is chosen as the threshold or cut-off (usually two fold) to determine genes having a significant change in expression. In otherwords, all genes having a ratio which exceeds the threshold are considered to be *bona fide* gene expression differences between the two samples. However, this method has low specificity and low sensitivity since the fold change chosen is entirely arbitrary and is prone to generate both false positives and false negatives in the analyses.

- Standard deviation

This method assumes the ratios between control and experimental values form a continuous normal distribution. Genes are selected according to their distance from the mean values of the control-to-experimental ratios. Usually the distances are taken to be ±2 or ±3 standard deviations. Two standard deviations from the mean represent a 95.45% confidence level whereas three standard deviations from the mean represent a 99.73% confidence level. In other words, for genes lying more than two standard deviations away from the mean, the probability that the genes selected are

differentially expressed is 95.45%. For those genes lying more than three standard deviations away from the mean, the probability that the genes are differentially expressed is 99.73%.

- Univariate statistics

Univariate statistical test such as a *t*-test can be used to assign a probability (P value) to a gene which is being differentially regulated above a given threshold, when the log ratios of the control-to-experiment values follow a normal distribution. A *t*-test is used to determine the difference between the means of two populations. The *t*-test compares the size of the difference between means with the standard error of that difference. From a *t*-test, a *t* statistic is converted to a probability value P. But suppose you are measuring the expression levels of 5,000 genes, instead of applying the standard cut-off for statistical significance of $p<0.05$, it is appropriate to correct the P value estimate by dividing the number of gene expression measurements you are making, i.e. set P to the far more stringent value of $p<(0.05/5,000)$ or $p< 1x10^{-5}$. This is called a Bonferroni correction. However, such correction is sometimes too stringent and no differentially expressed genes may be reported.

This method is particular useful when replicates are present for the microarray analysis. This is a better method than the methods listed above as the variations across replicates can be assessed so that statistically-significant genes across replicates can be chosen. However, this method assumes that the changes in expression level of genes are highly correlated across replicates. This may not be true depending on the manipulation of the samples for hybridisation. Sometimes, large variations in gene expression levels of real differentially-expressed genes may be observed across replicates and they will be missed out when this method is used for analysis.

However, regardless of which analysis being used, false positives may still be identified. The percentage of false positives identified by chance is described as the false discovery rate (FDR). The false discovery rate can range from 10 to 80% depending on the statistical analyses (Tusher et al., 2001). One way to minimise the FDR is to increase the sample size.

D. Descriptive statistics

The patterns or signature of gene expression should be identified in all the gene expression values obtained in an experiment. This type of question is addressed using descriptive statistics or exploratory analysis. Clustering trees can show the relationships between samples (such as normal versus diseased cells), between genes, or both. Hierarchical clustering such as that used in the program Cluster/TreeView (http://rana.lbl.gov/EisenSoftware.htm) (Eisen et al., 1998), is probably the most popular way for making trees with microarray data. This method groups genes and/or samples with similar expression patterns into family trees. Gene expression values are colour coded

from bright red (most up-regulated) to bright green (most down-regulated). This allows one to visualize large amounts of data.

Principal components analysis (PCA) is a different exploratory technique used to find patterns in gene expression data from microarray experiments. The central idea behind PCA is to transform a number of variables into a smaller number of uncorrelated variables called principal components. In a typical microarray experiment, the point of PCA is to detect and remove redundancies in the data (such as genes whose expression values do not change) in order to reduce the noise in the data set and to identify outliers (or clusters of outliers) that might be of interest to study.

E. Data mining

Once differentially-expressed genes are identified in the microarray analysis, these data must be interpreted in terms of gene functions and functional relationship between genes using existing biological knowledge. The Gene Ontology Consortium (http://www.geneontology.org/) addressed the need for consistent description of gene products with different databases. The GO project describes functions of gene products in three different categories: cellular components, biological processes and molecular functions. Such effort has made interpretation of differentially-expressed genes a manageable task.

4.1.5 Confirmation and validation of data

Technical and biological variability generated experimentally, and due to data processing and statistical methods affect the results obtained for microarray experiments. Increasing the number of replicates will decrease the false discovery rate (FDR) and thus the chance of getting false positive genes. However, sometimes it is difficult to increase the number of replicates considering that microarray experiments are expensive to perform and samples may often be limiting (e.g. patient samples). Therefore, the results obtain from microarray studies should be verified by other approaches.

- Comparison with existing literature.

The microarray data can be compared with information available in literature and databases. If there is agreement between the microarray analysis and data from other sources, this provides a general confidence level that the data accurately reflects the biological processes involved. Taking the analysis in this Chapter as an example, published target genes have been identified for SCL and GATA1 (Chapter 1, section 1.4.2.1 and 1.4.2.2). If these published target genes are also identified in the expression profiling analysis, this provides evidence that the data is likely to be meaningful.

- Other gene expression assays

Experimental approaches should be used to further confirm the results obtained from microarray analyses. qPCR and other assays mentioned in Chapter 1 (Chapter 1, section 1.3.2) are useful and sensitive assays to confirm the changes in gene expression obtained from microarray analyses.

4.2 Aims of this chapter

The overall aim of work presented in this Chapter was to identify putative target genes of the SCL erythroid as follows:

1. To study the changes in global gene expression patterns identified by siRNA knockdown of each of five members of the SCL erythroid complex. This would be accomplished by using the Affymetrix expression GeneChips.
2. To validate the gene expression differences obtained in the microarray analyses by q-PCR.
3. To identify differentially-expressed genes which are common to the 5 knockdown states. Such co-regulated genes would be considered to be putative targets of the SCL erythroid complex.
4. To search for common motifs in the regulatory regions of these co-regulated genes as a means of identifying the locations where these transcription factors bind in order to regulate them.

4.3 Overall strategy

In Chapter 3, siRNAs with knockdown efficiencies that satisfied specific criteria were selected for each transcription factor in the SCL erythroid complex. In this Chapter, the changes in expression of other genes in the genome which were a consequence of these knockdowns were studied by Affymetrix expression arrays. Three biological replicates for each transcription factor knockdown and the luciferase negative controls were performed. The qualities of the hybridised arrays were monitored and only the arrays passed the quality control were used in statistical analyses to identify differentially expressed genes by comparing the data from luciferase siRNA transfected cells and the cells transfected with specific siRNAs against transcription factors in the SCL erythroid complex. Confirmation of differentially expressed genes was addressed by performing quantitative PCR from the knockdown condition mRNA samples. Comparisons of the differentially expressed gene sets for different transcription factor knockdowns were used to identify co-regulated genes which were considered to be putative targets of the SCL erythroid complex. Computational analyses of common DNA binding motifs were also performed using the NestedMICA programme (Down and Hubbard, 2005) as a means of determining the binding location of transcription factors in the complex. The overall strategy of this expression study was summarised in Figure 4.4.

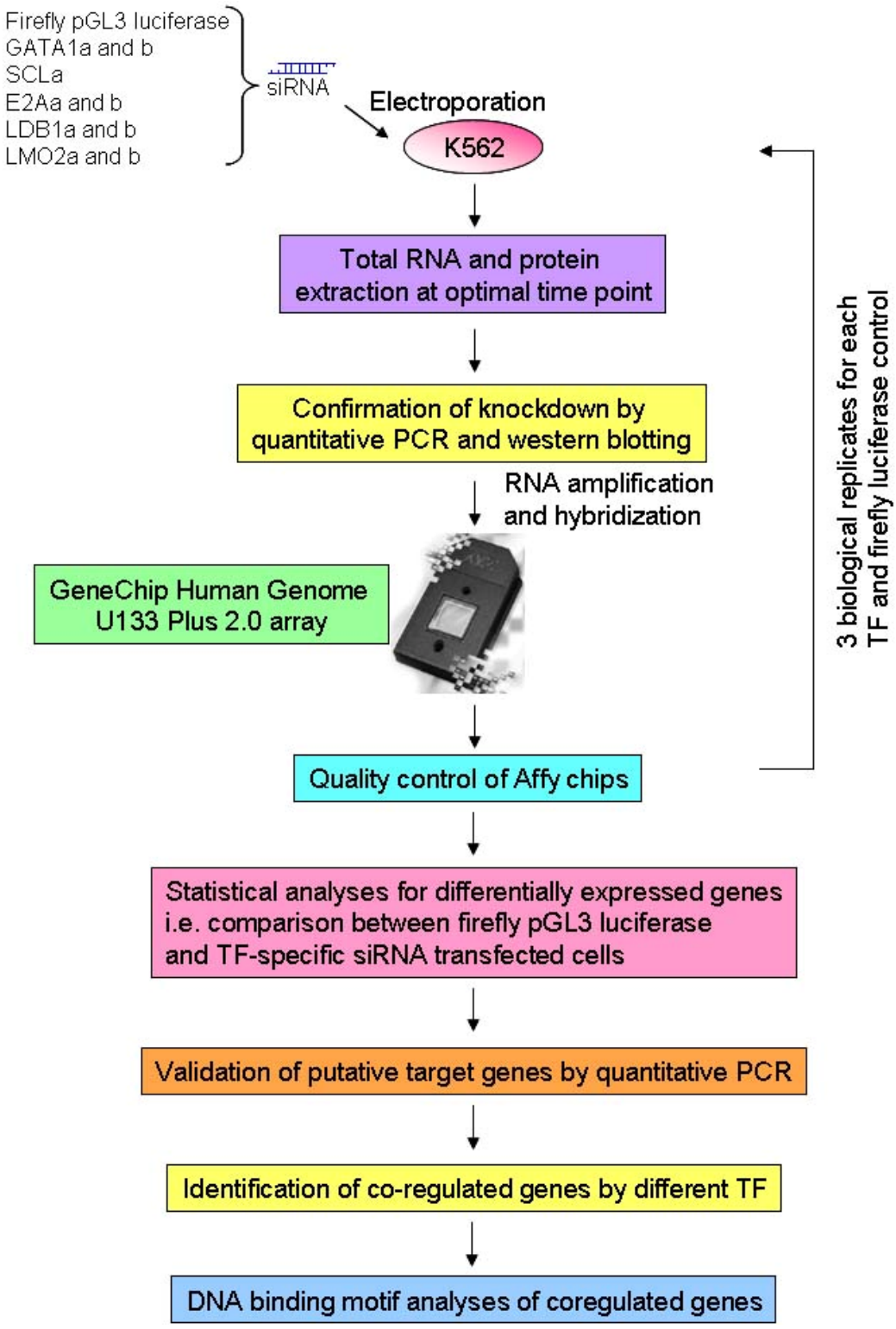


Figure 4.4. Overall strategy of expression profiling study of the silencing of the SCL erythroid complex. The effect of silencing of 5 members of the SCL erythroid complex on the expression of genes in the human genome was studied using Affymetrix expression GeneChip arrays. High quality hybridised GeneChips for 3 biological replicates for each transcription factor were used for statistical analyses to identify differentially expressed genes. Validation of differentially expressed genes was performed using SYBR green qPCR. Identification of co-regulated genes for each transcription factor was done by comparing the differentially-expressed genes of the 5 transcription factors. DNA motif analysis was also performed using NestedMICA.

4.4 Results

4.4.1 Preparation and quality control of samples

In order to minimise variation in the data obtained from three independent bioreplicates of each transcription factor knockdown, a number of parameter were controlled for the preparation of samples to be used on the Affymetrix GeneChips as follows:

A. Culturing of cells

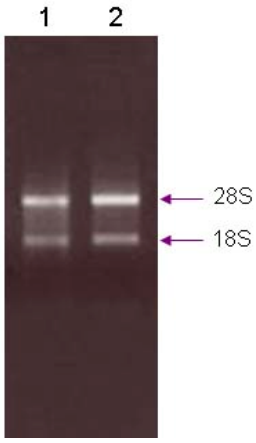
K562 cells were cultured and maintained at a concentration of 0.5 to 1 million/ml according to the ATCC specification. To ensure that transfections performed for individual bioreplicates behaved in a consistent manner, K562 cells were cultured for no more than a week before siRNA transfections were performed. K562 cells were split and fresh media were added one day before transfections.

B. RNA quality

RNA can be easily degraded by RNases and this can affect the quality of the RNA samples in subsequent manipulations and analyses. To control the quality of total cellular RNA samples used in the Affymetrix experiments, electrophoresis of the total RNA samples was performed to check if there were any signs of degradations. In the total cellular RNA, mRNA only comprises 1-3% of the total amount whereas ribosomal RNA (rRNA) makes up over 80% of the sample. After electrophoresis, only the rRNAs (28S, 18S, 5.8S and 5S) are visualised on the gel and can be used as a reference to monitor the overall RNA quality. For intact RNA samples, the ratio of 28S and 18S should be approximately 2:1 - this is traditionally used as the benchmark to monitor RNA degradation. In the Affymetrix experiments, all the RNA samples were checked for degradation. Figure 4.5 shows examples of the RNA samples extracted from E2Aa siRNA transfected and untransfected K562 cells, and the quantification of the 28S and 18S rRNA subunits. The bands for 28S and 18S rRNA subunits were quantified and yielded a 28S/18S ratio of approximately 2.

In addition to the 28S/18S ratio, the contamination of RNA samples by organic solvents and protein was also monitored. The ratio of absorbance at 260 nm and 280 nm indicates the purity of the RNA sample and should fall into the range between 1.8 and 2.1. Ratios of more than 2.1 indicate RNA degradation while ratios below 1.8 indicate protein contamination. The 260/280 absorbance readings of all the RNA samples were measured and fell between the range of 1.8 to 2.1. The readings in Figure 4.5 showed the 260/280 ratios of the two RNA samples described above - both samples showed ratios between 1.8 and 2.1.

A



B

Lane	28S intensity	18S intensity	28S/18S Ratio	260/280 Ratio
1	16064	8206	1.96	1.99
2	20170	9396	2.15	2.01

Figure 4.5. Agarose gel electrophoresis and 260/280 nm absorbance check for total RNA samples. Total cellular RNAs were visualised by ethidium bromide staining of a 1% denaturing TBE agarose gel. Panel A: Gel picture shows subunits of rRNA. Lane 1, total RNA sample extracted from E2A siRNA transfection; lane 2, total RNA sample extracted from untransfected K562 cells. Purple arrows on the right indicate the positions of the 28S and 18S rRNA subunits. The 5S rRNA subunit could not be detected on the gel. Panel B: Table shows the quantification of 28S and 18S bands on the gel by Labworks and the 260/280 absorbance ratios. The ratios of 28S to 18S were close to 2 while the 260/280 ratios were between 1.8 and 2.1 in both instances.

C. Amplification rate

Sample preparation for hybridisation to the Affymetrix GeneChip expression array required only a small amount of starting total RNA. This was because the RNA was reverse-transcribed to generate double-stranded cDNA containing a T7 promoter. *In vitro* transcription (IVT) of the cDNA was carried out under the control of the T7 promoter and large amounts of complementary RNA (cRNA) were generated. This amplification process allowed the synthesis of sufficient amounts of cRNA for hybridisation onto the array when the initial RNA sample was limiting (Figure 4.3). During the amplification process, the RNA could have become degraded which may have resulted in low amplification rates. Therefore, it was crucial to assess the amplification process before hybridisation. With a starting total RNA quantity of 5µg, an adjusted complimentary RNA (cRNA) amount of over 60 µg was expected according to the manufacturer's protocol (adjusted cRNA amount was the amount of cRNA measured after IVT minus the starting amount of total RNA). If the yield of cRNA was substantially lower, there may have been RNA degradation during the amplification or the amplification may have been inefficient due to RNA purity or degradation of the starting material. In all the biological replicates of Affymetrix array hybridisation performed for this study, the amount of amplified cRNA was over 60 µg (Table 4.1).

siRNA	Selected optimal time point	Amount of adjusted amplified cRNA (µg)		
		Biorep 1	Biorep 2	Biorep 3
Luciferase	24 hr	---	---	121
Luciferase	36 hr	191	165	138
GATA1a	24 hr	---	140	127
GATA1b	24 hr	60	98	87
SCLa	24 hr	127	101	194
E2Aa	24 hr	62	68	130
E2Ab	24 hr	102	110	74
E12	24 hr	---	104	83
E47	24 hr	90	111	115
LDB1a	36 hr	128	119	109
LDB1b	36 hr	159	122	159
LMO2a	24 hr	---	80	97
LMO2b	24 hr	---	66	103

Table 4.1. Amount of adjusted amplified cRNA in all biological replicates. Adjusted cRNA amount is the amount of cRNA measured after IVT minus the starting amount of total RNA. Note: the quantities of cRNAs were not available for samples indicated by a ---. cRNAs for these samples were generated by others in the lab (Amanda Hall, Sanger Institute). However, these samples gave a cRNA yield greater than the 60 µg threshold.

D. Amplified RNA quality

In addition to checking the amplification rate, the quality of the amplified RNA was also monitored. This was because the resultant cRNAs may have been degraded during the amplification procedure. cRNA quality was assessed by agarose gel electrophoresis and by using a Bioanalyzer. The Bioanalyzer was not used in this study as it was not available in the lab when this project was carried out. Using the former method, the typical size distribution of the unfragmented cRNA below 1 kb was observed as shown in Figure 4.6. All the amplified cRNA samples used in this study were visualised by agarose gel electrophoresis before hybridisation to the Affymetrix GeneChips. Any samples which did not show the expected size distribution were discarded and the process was repeated until high quality cRNA was obtained.

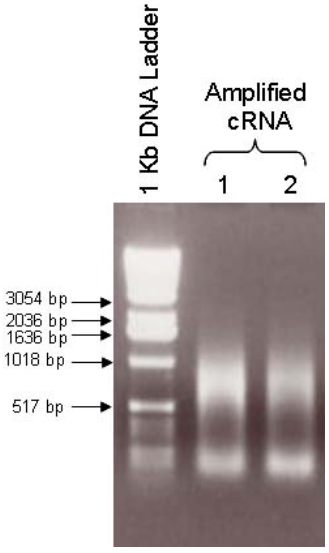


Figure 4.6. Agarose gel electrophoresis of amplified cRNAs. The purified and unfragmented cRNAs obtained from IVT were visualised by ethidium bromide staining of a 1% denaturing TBE agarose gel. The left lane shows the 1 kb DNA ladder and the corresponding sizes of the bands are labelled. Lane 1, luciferase control siRNA cRNA sample; lane 2, GATA1a siRNA knockdown cRNA sample.

4.4.2 Quality control of Affymetrix GeneChips

Once the cRNA samples were checked for quality control, they were hybridised onto the Affymetrix GeneChip expression arrays. However, the hybridisations themselves were subjected to strict quality control criteria even before any of the arrays were analysed with respect to the biological study being performed. Any GeneChips which do not pass the criteria for quality controls were discarded. The following criteria were used to assess the quality of Affymetrix Gene Chip hybridisations.

A. Probe array image inspection

One of the first criteria to be checked was the scanned GeneChip image. This was done to determine the overall quality of the hybridisation. The presence of observable image artifacts such as scratches, uneven signal intensity across array etc. was inspected by eye. Each probe cell was also visualised by zooming in. None of the GeneChips hybridised for this study showed obvious and visible artifacts. An example of a high quality GeneChip hybridisation from this study is shown in Figure 4.7.

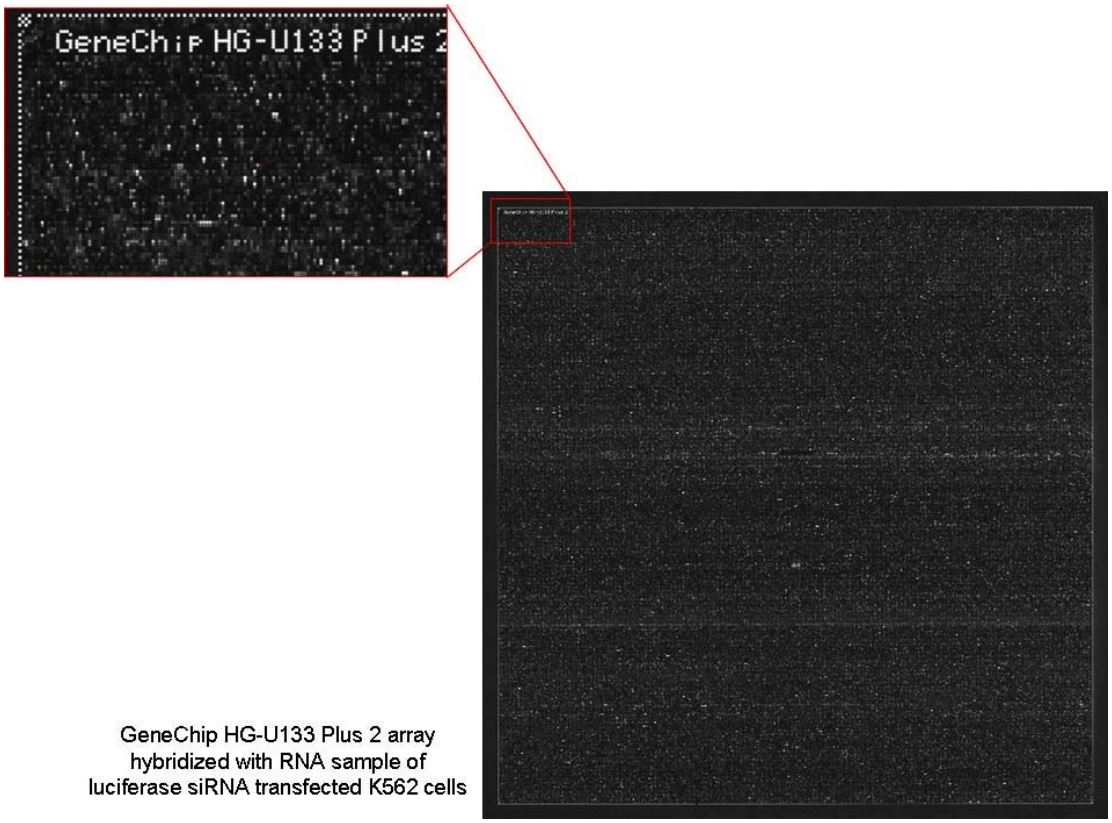


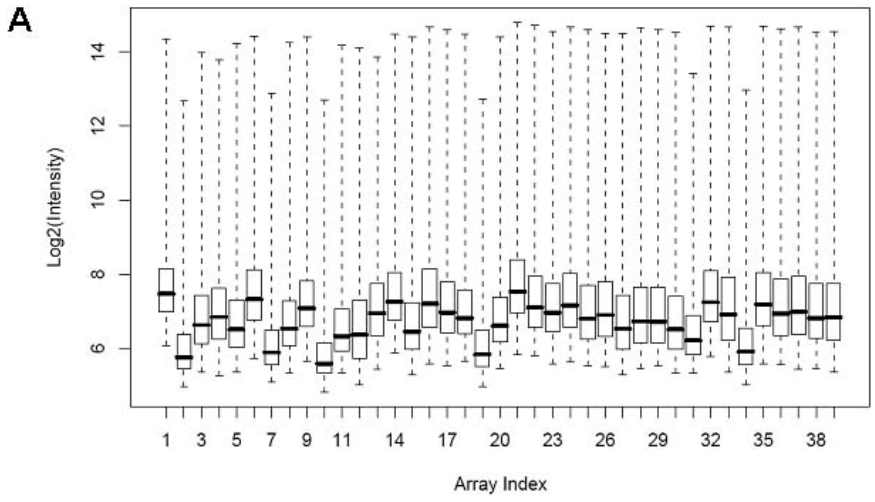
Figure 4.7. Image of hybridised Affymetrix GeneChip. GeneChip HG-U133 Plus 2 GeneChip arrays hybridised with cRNA derived from K562 cells transfected with luciferase siRNA at the 36 hour time point is shown. Image on the right shows the entire scanned GeneChip. Image on the left in the red box shows the zoomed image of the top left hand corner of the GeneChip. This is an example of a high quality GeneChip image with no visible artifacts.

B. Intensity correlation

Further quality control of the Affymetrix GeneChips was carried out by analysing the signal intensity and control gene profiles. This was done using the AffyQC Report package of Bioconductor. The signal intensity of all arrays included in the data analyses was assessed. The AffyQC Report package generated the log2 intensity of all the perfect match probes in various GeneChips and the density plots of these intensity values. Regardless of the samples being hybridised, the overall signal intensity of all the GeneChips should be similar since the majority of probe signals (i.e., gene expression levels) are not changing amongst the samples. All the GeneChips hybridised for this project showed similar patterns in the density of intensity values with the 50% of probes having a log2 intensity value between 5 to 8 (Figure 4.8). This indicated that the GeneChips all showed similar hybridisation characteristics and passed the signal intensity quality control criteria.

siRNA	GeneChip Index		
	Biorep 1	**Biorep 2**	**Biorep 3**
Luciferase (24 hr)	1	2	3
Luciferase (36 hr)	37	38	39
GATA1a	4	5	6
GATA1b	7	8	9
SCLa	10	11	12
E2Aa	19	20	21
E2Ab	22	23	24
E12	13	14	15
E47	16	17	18
LDB1a	25	26	27
LDB1b	28	29	30
LMO2a	31	32	33
LMO2b	34	35	36

Table 4.2. GeneChip (array) index used in the affyQCReport.



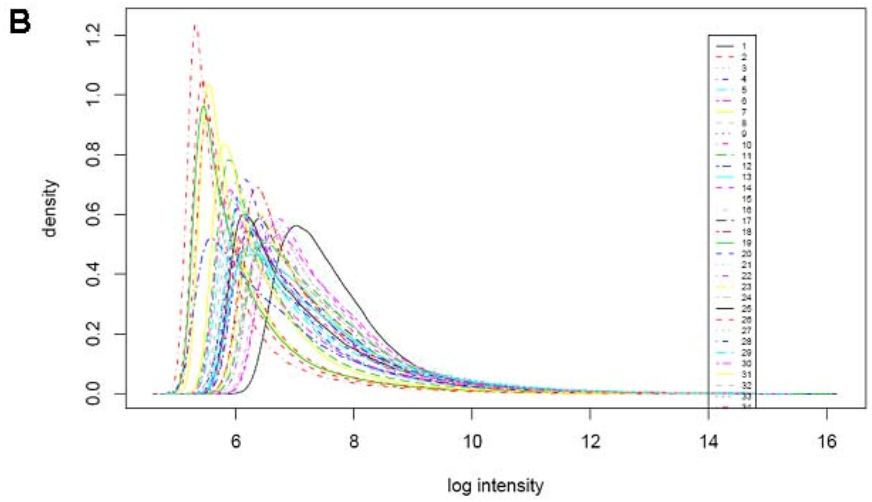


Figure 4.8. Perfect match intensity of all GeneChips studied in this project. Perfect match probe log2 intensities were calculated in the AffyQC Report package of Bioconductor. Panel A: box plot of log intensity of all perfect match probes across all GeneChips used in this project. The boxes contain the median centred 50% of the datapoints for each GeneChip hybridisation. The x-axis shows the GeneChip index (Table 4.2) while the y-axis shows the log2 intensity of probes. Panel B: density plot of log intensity of all perfect match probes across all GeneChips used in this project. The x-axis shows the log intensity of probes while the y-axis shows the kernel density of probes having a particular log intensity. The numbering of GeneChips and the corresponding samples are shown in Table 4.2.

C. Housekeeping gene profiles

The intensity signals of housekeeping genes on the GeneChips were also used as a means of assessing hybridisation quality. The signal intensity of the 3' probe sets for the house-keeping genes β-actin and GAPDH were compared to the signal intensity of the corresponding 5' probe sets. For good quality hybridisation samples, the 3' to 5' ratio should be less than 3 as degradation usually occurs from the 5' end of mRNA, resulting in an accumulation of 3' fragments. A high 3' to 5' ratio may also indicate inefficient transcription of double-stranded cDNA (ds cDNA) or biotinylated cRNA as the antisense cRNA was transcribed from the sense strand of the ds cDNA via the T7 promoter at the 3' end of the sense strand. The 3' to 5' ratios for β-actin and GAPDH for all the GeneChip hybridisations for this project were below 3 as shown in Figure 4.9.

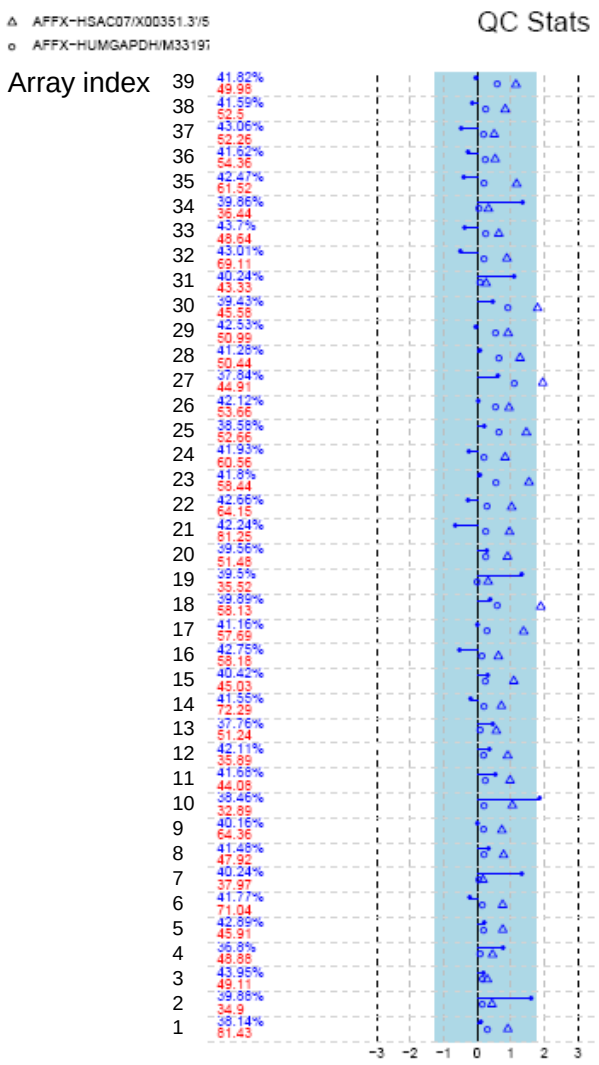


Figure 4.9. Internal house-keeping control gene profile and present call profile. The black numbers on the left are indicative of the GeneChip index (see Table 4.2). The blue numbers and red numbers next to the GeneChip index show the percentage of present call probes and the average background intensity respectively. The dotted vertical lines delineate the scale of -3 to 3 for the 3' to 5' ratios of the house-keeping genes. The triangles show the 3' to 5' ratios for β-actin while the circles showed the 3' to 5' ratios for GAPDH. When the circles and triangles are coloured in blue, they were within the acceptable quality control ratios, otherwise they are coloured in red.

D. Average background intensity and percentage of present genes

The background intensity of the hybridisation signals on the GeneChips has a great impact of quantification of probe intensity and therefore it was also monitored as one of the quality control criteria. According to the documentation in the Affymetrix manual, the typical average background values should range from 20 to 100. The GeneChip hybridisations obtained for this project all had average background intensities falling within this range (see Figure 4.9; red numbers).

The number of probe sets called 'present' relative to the total number of probe sets on the GeneChip is described as the percentage of present genes. This percentage is an indication of sample quality and is dependent on the cell type and biological or environmental stimuli. Low percentage values imply poor sample quality whilst replicates are expected to have similar percentage values. All the arrays hybridised have percentage of present calls of approximately 40% indicating that the hybridisations and the sample qualities were similar for the GeneChips analysed for this project (see Figure 4.9; blue numbers).

E. Border elements intensity correlation

During the hybridisation of samples onto the Affymetrix array, control "spikes" were included. The control oligo B2 was spiked into the hybridisation mix and it hybridised to features along the outer edges and corners of the GeneChip (so-called "border elements"). These hybridisation controls were independent of RNA sample quality and amplification and were used as positive controls for even hybridisation characteristics across the GeneChip and were also used by the software for automatic grid alignment over the image during quantitation of signals. To assess for even hybridisation of the GeneChips, the intensities for all border elements were collected. Elements with an intensity of 1.2 times above the mean were regarded as "signal" controls (positive controls). Elements with a signal less that 0.8 of the mean were regarded as "background" controls (negative controls). The intensities of positive and negative border elements for each GeneChip should be similar. Large variations in the positive control elements indicate non-uniform hybridisation or gridding problems. Variations in the negative controls indicate background fluctuations. At least 50% of positive border elements of the GeneChip hybridisations had intensity values below 20000 with an average of 11774 of all the positive border elements in all arrays. The average intensity of all negative border elements in all of the GeneChips was 141 with 50% of them close to 0 (Figure 4.10).

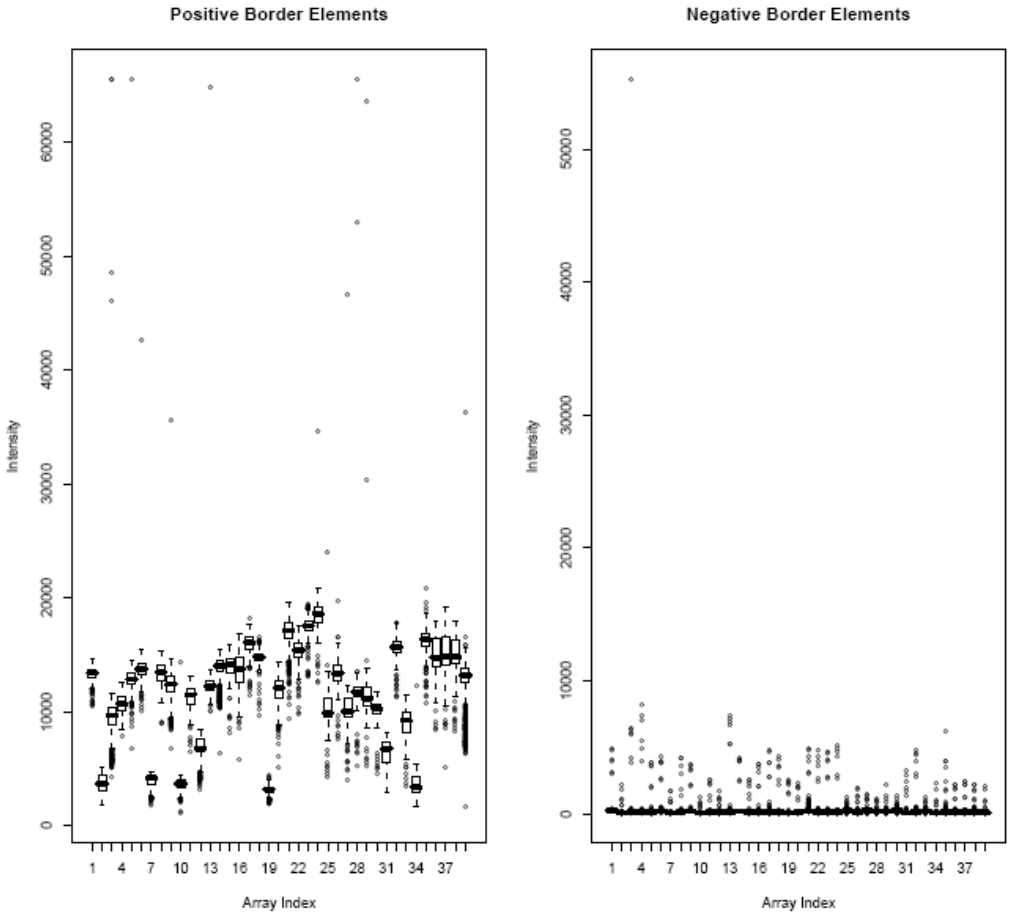


Figure 4.10. Boxplot of intensity of positive and negative border elements. Left panel shows the intensity of the positive border elements and right panel shows the intensity of the negative border elements. The boxes contain the median centred 50% of the datapoints for each GeneChip hybridisation. In both panels, the y-axis is the signal intensity and the x-axis is the GeneChip (array) index.

F. Hybridisation and Poly-A controls

In addition to the housekeeping control genes discussed above, the quality of the entire amplification and labelling process, and the sensitivity of the GeneChips was assessed using exogenous positive control poly-A mRNA "spikes". These "spikes" were poly-A mRNAs derived from *in vitro* synthesised, polyadenylated transcripts for several *B. subtilis* genes (*lys*, *phe*, *thr* and *dap*). Probe sets for these genes were represented on the GeneChips. These control mRNAs were added to the starting RNAs at different concentrations and were amplified and labelled together with the RNA samples. Assessing the signal intensity generated for these controls helped monitor the amplification and labelling process independent of the RNA sample. The median of signal intensities of these spike controls in the 39 GeneChips in this experiment is shown in Figure 4.11. The "spike" controls showed increasing and linear signal intensities with increasing concentrations in the starting RNA samples. The lowest concentration of these controls allowed messages which were represented at 1 copy in 50 000 mRNAs to be detected on the GeneChip. All the GeneChips hybridised for this Chapter showed similar patterns for the poly-A spike controls.

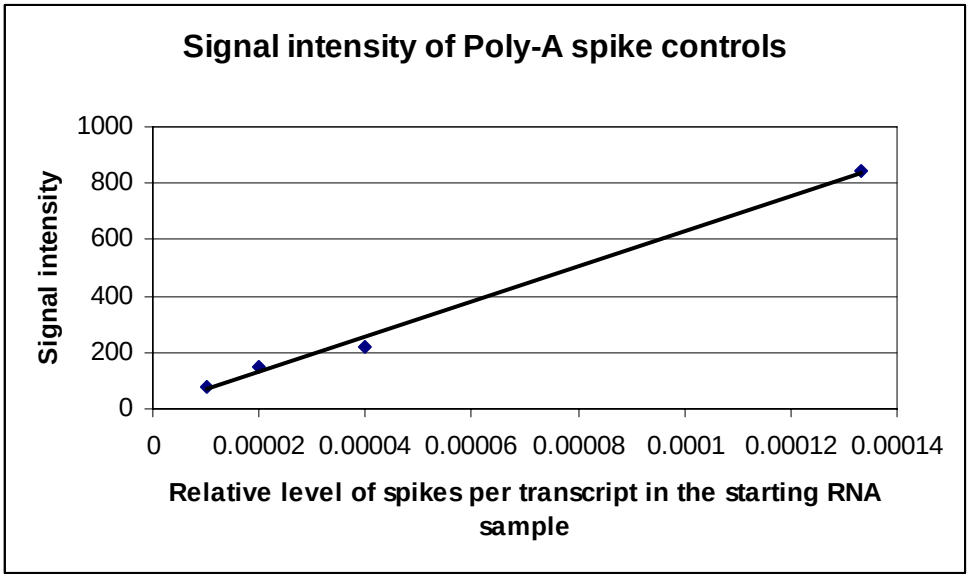


Figure 4.11. Signal intensity of poly-A RNA "spike" controls of the Affymetrix GeneChip hybridised with cRNA derived from K562 cells transfected with the various siRNAs. The x-axis is the relative level of each of the *B. subtilis* "spike" control transcripts per transcript in the starting RNA sample. Blue dots indicate each of the four *B. subtilis* spike control transcripts with increasing mRNA concentration from left to right: *lys*, *phe*, *thr* and *dap*. The y-axis is the median values of signal intensity of these transcripts on the 39 GeneChips hybridised in this experiment.

Additional controls were also included in the hybridisation to the GeneChips. These controls were used to evaluate the hybridisation efficiency independent of the RNA preparation and amplification procedure. These mRNA transcript controls were derived from genes in the biotin synthesis pathway of *E. coli* (*Cre*, *BioB*, *BioC* and *BioD*). Probe sets for these genes were represented on the GeneChips. Like the poly-A RNA controls, the hybridisation controls were added at different concentrations (1.5 pM, 5 pM, 25 pM and 100 pM for *BioB*, *BioC*, *BioD* and *Cre* respectively*)*. However, unlike the poly-A RNA controls, these mRNAs were labelled separately from the starting RNA samples and were added directly into the hybridisation mixture. The signal intensity of these genes should increase according to their relative concentrations if the hybridisation was performed according to manufacturer's standard. The median of signal intensities of these hybridisation controls in the 39 GeneChips in this experiment is shown in Figure 4.12. The hybridisation controls showed increasing and linear signal intensities with increasing concentrations in the hybridisation mixture. All the GeneChips hybridised for this Chapter showed similar patterns for the hybridisation controls.

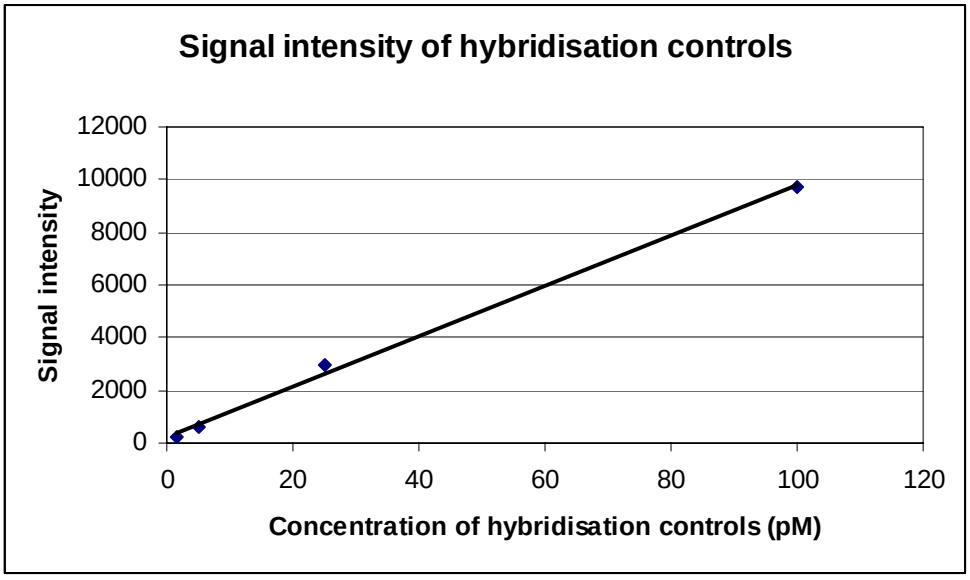


Figure 4.12. Signal intensity of the hybridisation controls of the Affymetrix GeneChip hybridised cRNA derived from K562 cells transfected with the various siRNAs. The x-axis is the concentration of the hybridisation controls of *E. coli* genes. Blue dots indicate the spike control genes with increasing concentration from left to right: *BioB*, *BioC*, *BioD* and *Cre*. The y-axis is the signal intensity of these genes in the 39 GeneChips hybridised in this experiment.

4.4.3 Data analysis of Affymetrix GeneChips

Once the hybridisations onto Affymetrix GeneChips passed the quality control criteria, the data derived from the biological study was analysed by statistical methods in order to determine differentially expressed genes between the relevant luciferase control (time points 24 or 36 hrs) and its corresponding transcription factor-specific siRNA knockdown conditions. Many methods of

quantification of probe sets have been developed e.g. MAS5, RMA and GC-RMA (Section 4.1.4). In the analyses described in this Chapter, the RMA method was used as it was shown to be more sensitive and more specific while introducing less bias to G+C content of probes (Irizarry et al., 2003a; Siddiqui et al., 2006). To handle such large data sets involving large numbers of probe sets and transcript information from across entire human genome, the GeneSpring GX 7.3.1 data analysis software was used.

4.4.3.1 Normalisation and statistical analyses of Affymetrix GeneChip data

Figure 4.14 outlined the strategy used to determine the genes that were differentially expressed between the control and experimental conditions. Signal intensities of the probe sets in all the 39 scanned GeneChips were imported into the GeneSpring analysis suite and quantitated by RMA. Experiments were created in GeneSpring to include all three biological replicates of the luciferase controls, the three biological replicates of the siRNAa transfections and the three biological replicates of the siRNAb transfections (except for SCL, where only the siRNAa assay was used). The signal intensities of all probes/genes were normalised in the following ways:

1. values of lower than 0.01 were set at 0.01,
2. to the median probe intensity of all measurements per hybridisation,
3. to the median of all gene intensities in all the samples in the experiment.

These normalised intensity values of all genes were then exported to Microsoft Excel and statistical analyses of differentially expressed genes were performed. The analyses of the two siRNAs for each transcription factor were done separately. The mean signal intensities of each gene were derived for the 3 biological replicates for luciferase, siRNAa and siRNAb respectively. Comparisons were made between luciferase and siRNAa as well as between luciferase and siRNAb in order to derive ratios of differences in gene expression. For statistical purposes, it was assumed that the ratios of gene expression for any one experiment occur as a normal distribution centered around the mean (Figure 4.13). Genes which were differentially expressed between the luciferase and experimental siRNAs by more than 2 standard deviations away from the mean ratio of the entire dataset were chosen for further analyses. Two standard deviations were used as a cut-off as it represented a 95.45% confidence level – in other words, the genes identified were statistically significant in terms of differential expression.

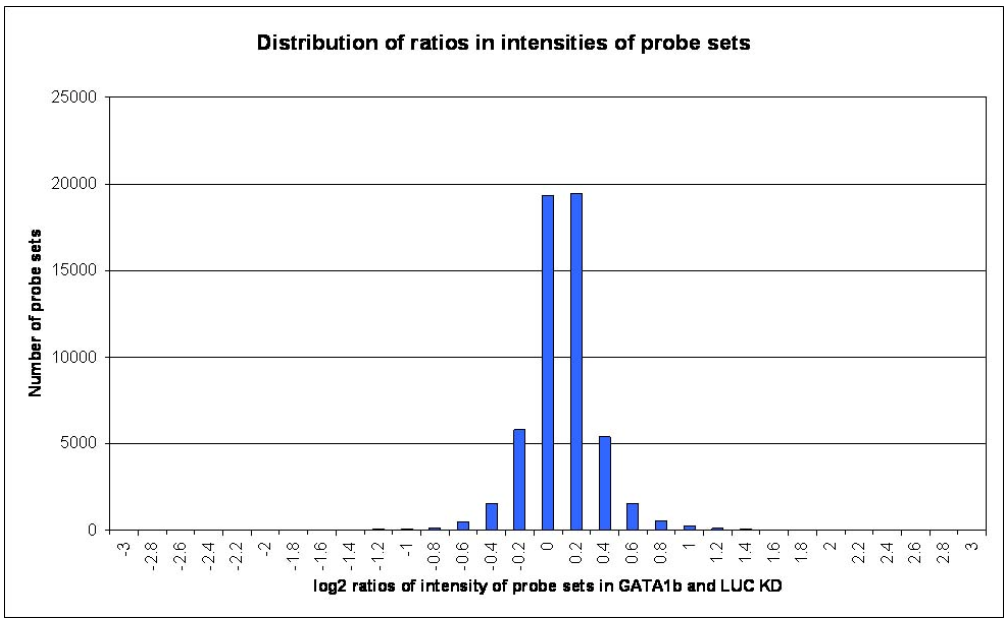


Figure 4.13. Distribution of log2 ratios of intensity of probe sets in GATA1b knockdown against luciferase knockdown. Y-axis: number of probe sets; x-axis: log2 ratios of the intensity of probe sets in the GATA1b knockdown against luciferase knockdown. The probe sets are centred around the mean in a normal distribution.

For each transcription factor knockdown experiment, four gene lists were obtained: genes down-regulated in siRNAa, genes down-regulated in siRNAb, genes up-regulated in siRNAa and genes up-regulated in siRNAb. Gene lists for the two siRNAs (a and b) for each transcription factor were treated independently up to this point because different siRNAs for the same gene can generate different off-targeting effects (Chapter 1, section 1.3.1.3 B). To filter away these off-target genes from further analyses, the down-regulated gene lists of the two siRNAs for each transcription factor were compared while the up-regulated gene lists of the two siRNAs for each transcription factor were also compared. These comparisons were performed using Venn diagrams and would allow for the identification of genes which were differentially expressed by both siRNAs. The gene lists identified by each siRNA are shown in the Venn diagrams of Figure 4.15. The genes found in the overlaps of the Venn circles (either up- or down-regulated) were considered as putative target genes of each transcription factor. Three points should be noted when interpreting the data from the Venn diagrams:

1. The numbers shown in the Venn diagrams are number of probe sets rather than numbers of genes. On the Affymetrix GeneChip, a gene can be represented by more than one probe set. Thus the actual numbers of genes found to be up- or down-regulated by each transcription factor are less than the numbers shown in Figure 4.15.

2. Genes/probe sets which were down-regulated by the knockdown of a transcription factor were considered to be putative target genes which were activated by the transcription factor.

3. Gene/probe sets which were up-regulated by the knockdown of a transcription factor were considered to be putative target genes which were repressed by the transcription factor.

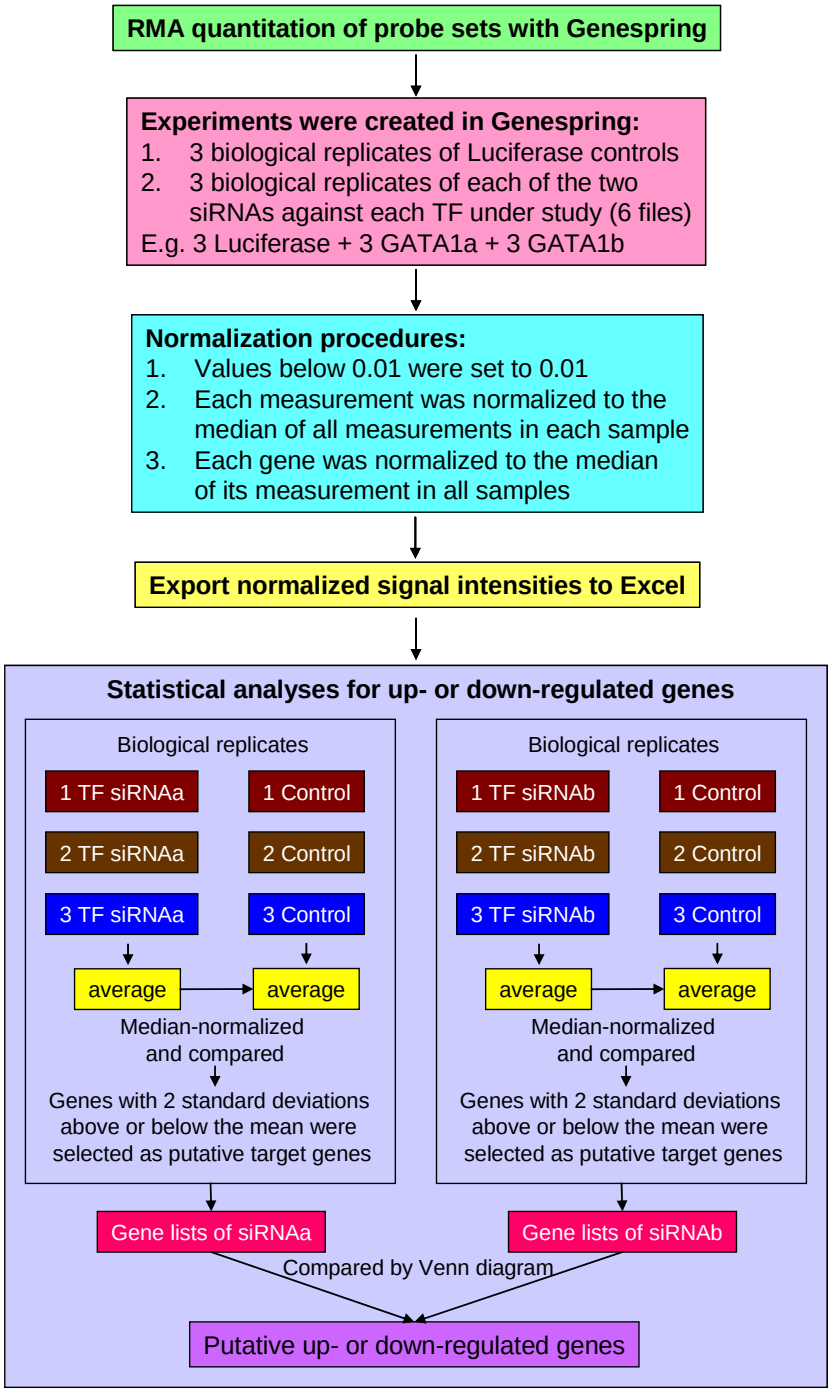
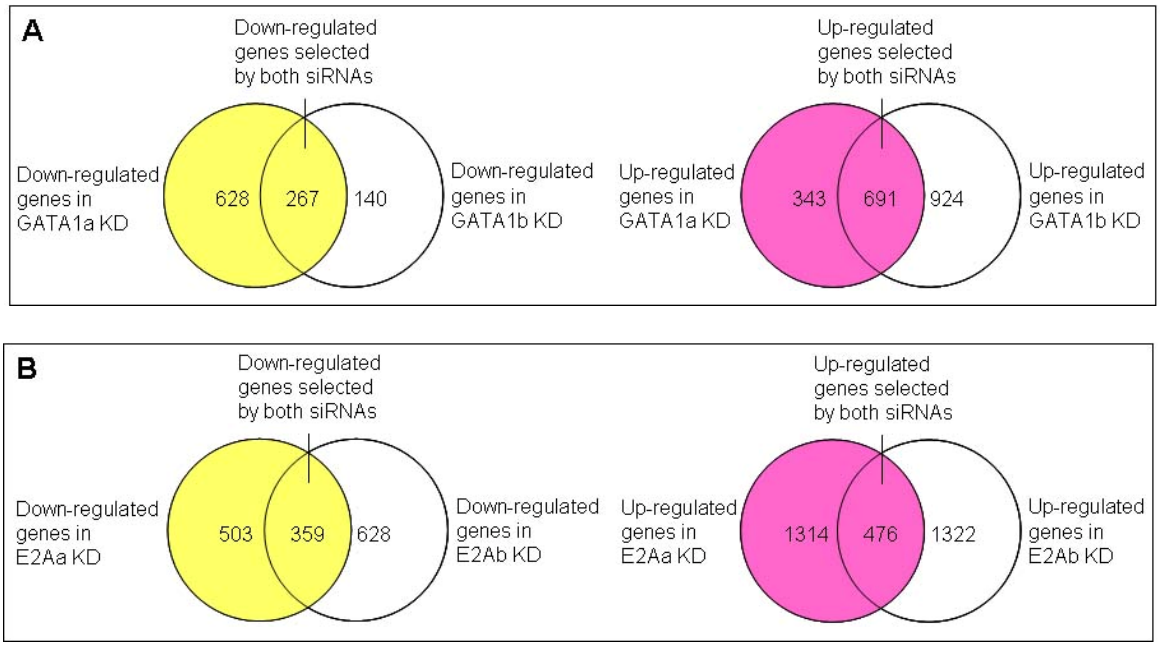


Figure 4.14. Flow diagram of statistical analyses of differentially-expressed genes in Affymetrix GeneChip. Signal intensities of the probe sets were quantitated by RMA. Experiments were created in GeneSpring to include all three biological replicates of the luciferase controls, the three biological replicates of the siRNAa transfections and the three biological replicates of the siRNAb transfections. The signal intensities of all probes/genes were normalised at three levels. The statistical analyses of the two siRNAs for each transcription factor were done separately. The mean signal intensities of each gene were derived for the 3 biological replicates for luciferase, siRNAa and siRNAb respectively. Comparisons were made between luciferase and siRNAa as well as between luciferase and siRNAb in order to derive ratios of differences in gene expression. Genes which were differentially expressed between the luciferase and

experimental siRNAs by more than 2 standard deviations away from the mean ratio of the entire dataset were chosen for further analyses.

4.4.3.2 Differentially-expressed genes and comparison of two siRNAs

Based on the Affymetrix studies described above, the general functional roles of GATA1, SCL, E2A, LDB1 and LMO2 were determined with respect to how they affected gene expression patterns across the entire human genome in K562 cells. These roles were based on the numbers of up-regulated and down-regulated genes (probe sets) which were identified to be common to both the siRNA a and b knockdowns for each transcription factor (Figure 4.15). For the GATA1 knockdowns, 267 probe sets were found to be down-regulated (activated by the transcription factor) while 691 probe sets were found to be up-regulated (repressed by the transcription factor) for both siRNAs. This suggests that GATA1 is primarily a repressor in K562 cells. 486 and 359 probe sets were shown to be up- and down-regulated by E2A respectively. This suggests that E2A acts as both a repressor and an activator in K562 cells. Similarly, for the LDB1 knockdowns 716 probe sets were up-regulated, while 822 probe sets were down-regulated, suggesting that LDB1 acts as both an activator and a repressor in K562 cells. LMO2 was seen to act mainly as a repressor - 1063 probe sets were up-regulated by LMO2 while 54 probe sets were down-regulated in the knockdown experiments. As only one siRNA was found to be effective in silencing the expression of SCL in K562 cells, it was difficult to determine its general role in regulating gene expression in K562 cells. 897 probe sets were found to be down-regulated by SCL while 1811 probe sets were found to be up-regulated. This suggests that SCL is more likely to be a repressor. However, as only one siRNA knockdown was used in the Affymetrix expression study, some of these genes might be off-target genes. Therefore, it was not possible to draw any firm conclusions of how SCL normally affects gene expression in this cell line.



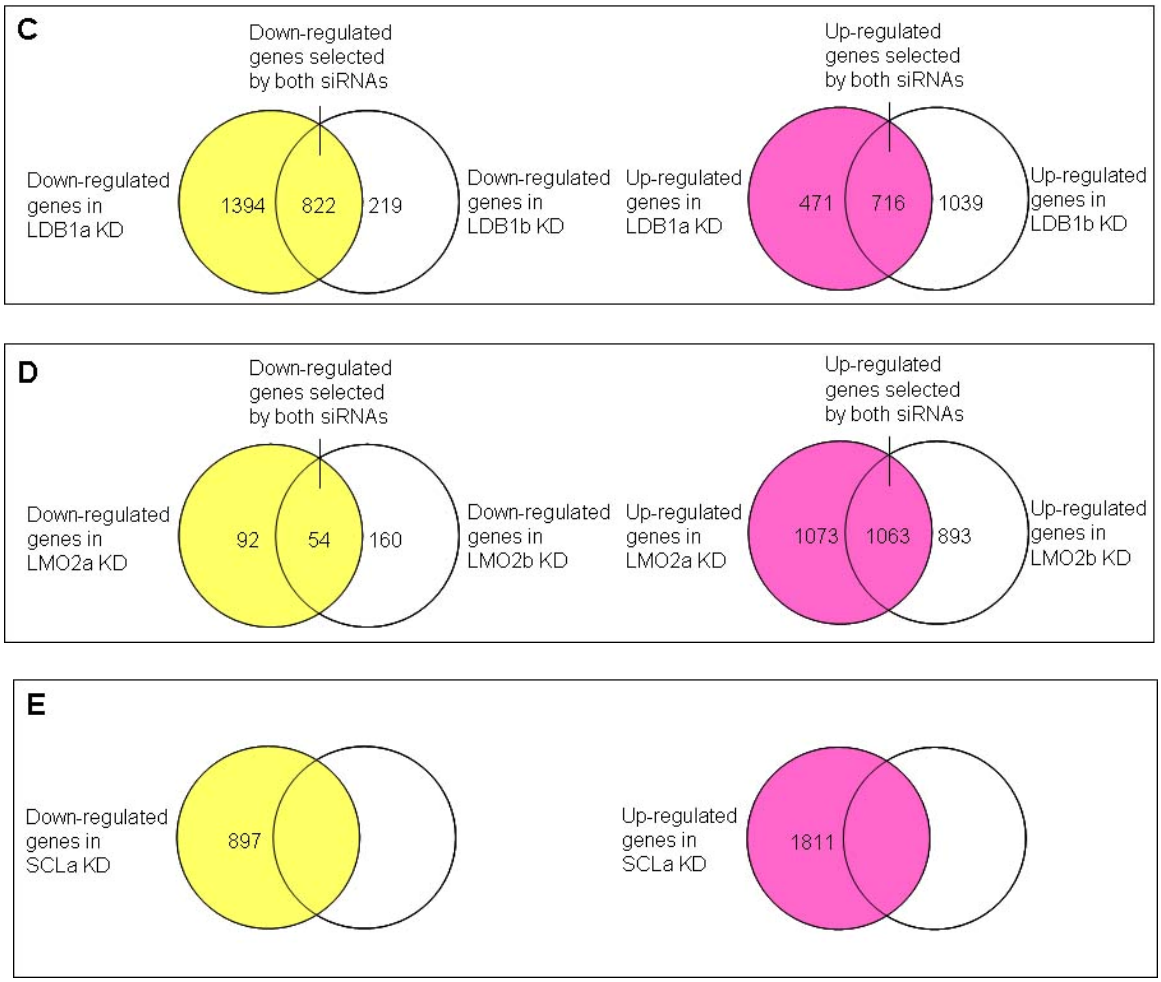


Figure 4.15. Venn diagram comparison of genes (probe sets) identified by siRNA knockdown experiments in K562 cells for each of five transcription factors found in the SCL erythroid complex. Down-regulated genes are shown in the Venn diagrams on the left (yellow) while up-regulated genes are shown in the Venn diagrams on the right (pink). Numbers shown in the Venn circles are numbers of probe sets for genes in the human genome identified in the relevant siRNA knockdown studies. The numbers shown in the overlap of the Venn circles denote those probe sets found in both the siRNA a and b knockdown conditions. Panel A: Venn diagram of GATA1 knockdowns; panel B: Venn diagram of E2A knockdown; panel C: Venn diagram of LDB1 knockdowns; panel D: Venn diagram of LMO2 knockdowns; panel E: Venn diagram of SCL knockdown study.

4.4.3.3 Validation of selected differentially-expressed genes by quantitative PCR

False positive expression differences are common in microarray analyses (Tusher et al., 2001). To determine whether the data obtained from the Affymetrix GeneChip analyses represented *bona fide* expression differences between the control luciferase and transcription factor knockdown conditions, a subset of up-regulated and down-regulated genes were further studied by quantitative PCR. Such validation allows us to determine and evaluate the Affymetrix GeneChip technology as a means of studying differential expression. For this purpose, the differentially-expressed genes of GATA1, SCL and E2A were investigated. Up-regulated and down-regulated genes which were transcription factors were chosen in the validation as they are the key components of a transcription

network and are thus excellent genes to study in the context of understanding transcriptional cascades downstream of the SCL erythroid complex in future work (see Chapter 7).

Table 4.3 lists the transcription factors that were up- or down-regulated in the knockdown studies that were chosen for validation. The majority of the genes selected were implicated in transcriptional regulation in various developmental processes including haematopoietic development. Within these genes, some of them are published targets of the corresponding transcription factors (MYC, EKLF, NFE2 and GFI1B). They were included in the validation to evaluate the sensitivity of the qPCR assays.

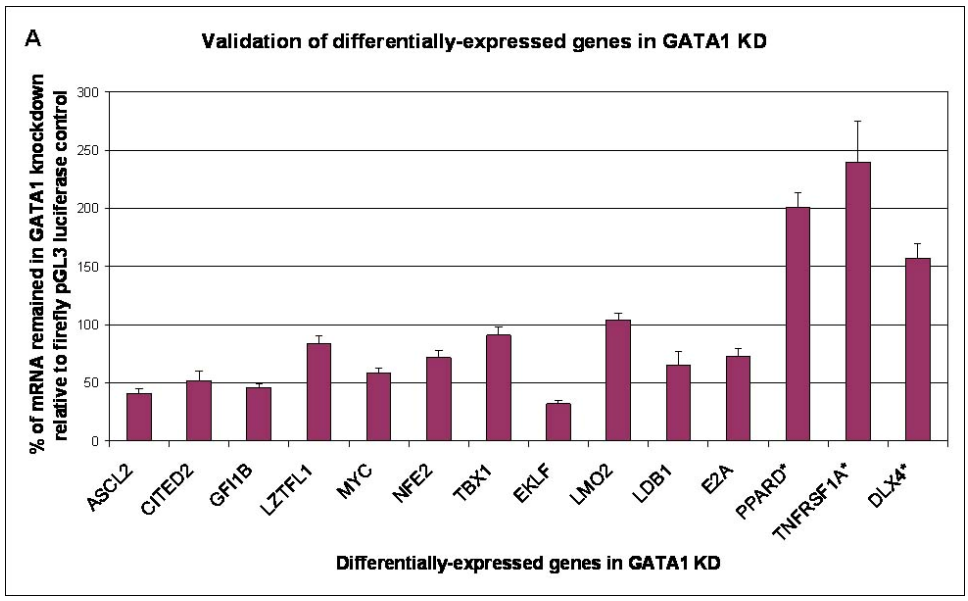
The RNAs used for the quantitative PCR validation were those used in the Affymetrix GeneChip analyses. They represented samples obtained from two independent biological replicates of the siRNA knockdown studies of GATA1, SCL and E2A. In each case, the mRNA levels of the putative target genes were compared between the knockdown sample and the control firefly luciferase sample. The mRNA levels remaining after knockdown relative to the luciferase control are shown in the bar charts in Figure 4.16. As described previously for the knockdown time-course experiments in Chapter 3 (Section 3.4.5), four house-keeping genes, β-actin, GAPDH, β-tubulin and RPL19, were used as controls for normalisation to minimise variations of RNA quality and concentrations.

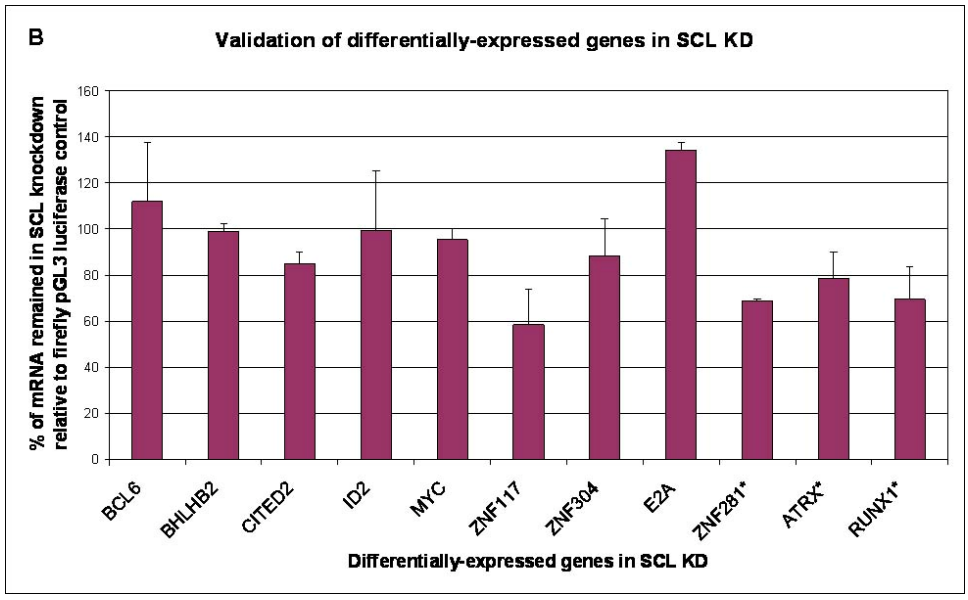
The cut-off in fold change used to determine whether a differentially-expressed gene showed a *bona fide* expression difference between the transcription factor knockdown and the luciferase control was different for each transcription factor. This cut-off was determined following the fold change observed for two standard deviations from the mean ratios in the Affymetrix GeneChip experiment. The fold change cut-offs used for GATA1, SCL and E2A were 1.39, 1.41 and 1.39 respectively. Ten out of fourteen of the differentially-expressed genes in GATA1 knockdown were found to have a fold change above the cut-off except for E2A, LMO2, LZTFL1 and TBX1 (Figure 4.16 A). For SCL and E2A, the validation rates were substantially lower (Figure 4.16 B and C). Only 1 gene (ZNF117) out of 11 genes and 2 genes (GTF2I and HHEX) out of 10 genes were validated for SCL and E2A respectively (Figure 4.16 D). Overall, a validation rate of 37% (13/35 assays) was achieved for the target genes studied. This data would suggest that the Affymetrix GeneChips identified a relatively high proportion of false positives in the knockdown studies described here.

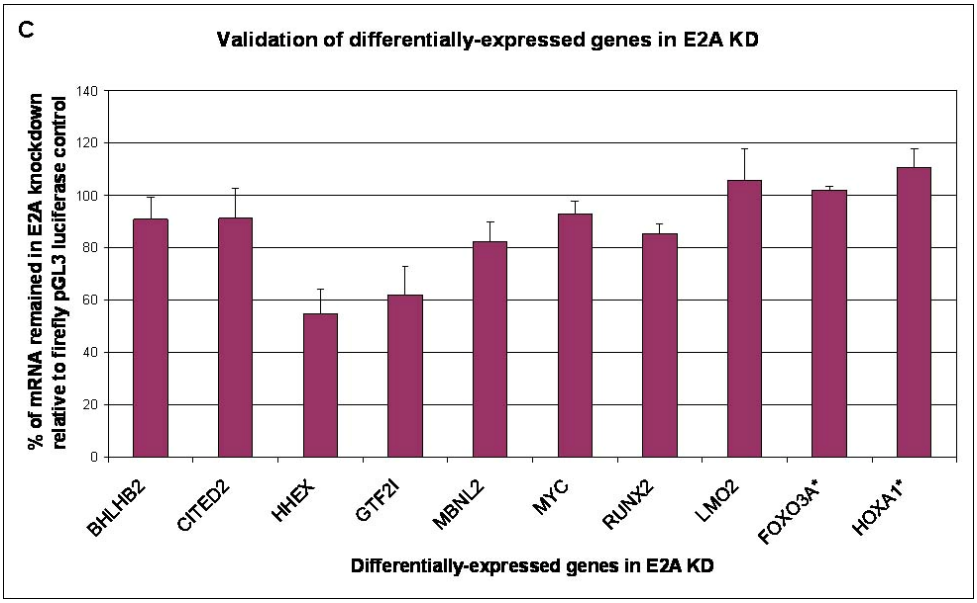
Name of putative target gene	Transcription factor regulating target	Mode of regulation	Protein subunit/ family	Functions of putative target gene
ASCL2	GATA1	Activation	bHLH family	lineage-specific transcription factors essential for development of the trophectoderm
CITED2	GATA1, SCL, E2A	Activation	C-terminal domain binds to CBP/p300 CH1 domain	Transactivates transcription factor AP2, an important regulator of neural and cardiac development
GFI1B	GATA1	Activation	Zinc finger protein	Represses transcription by recruiting corepressors and histone modifiers such as histone deacetylases (HDACs). Plays important roles in erythropoiesis.
LZTFL1	GATA1	Activation	Leucine zipper family	
MYC	GATA1, SCL, E2A	Activation	MYC family bHLH/Leucine Zipper domain	Oncogene of leukemia. Activates transcription of growth-promoting genes and represses growth-arrest genes by dimerizing MAX. Induces epigenetic reprogramming of human cells to pluripotency.
NFE2	GATA1	Activation	Leucine zipper family	Activates β-globin gene expression. Required for megakaryocytes maturation and platelet production.
TBX1	GATA1	Activation	T-box DNA binding domain family	Required for the development of epithelial cells and auditory organs
EKLF	GATA1	Activation	Krüppel-like factor family Zinc finger protein	Expressed in erythroid lineage. Activates β-globin gene expression.
LMO2	GATA1, E2A	Activation	LIM domain protein	Regulates erythroipoietic and endothelial development. Member of the SCL erythroid complex.
LDB1	GATA1	Activation	LIM-domain interacting protein	Regulates developmental processes. Member of the SCL erythroid complex.
E2A	GATA1, SCL	Activation	bHLH family	Activates transcription of B-cell specific genes. Regulates B-cell lineage development. Member of the SCL erythroid complex.
PPARD	GATA1	Repression	Peroxisome proliferator-activated receptor (PPAR) superfamily	Represses transcription of adipogenesis.
TNFRSF1A	GATA1	Repression	Tumor necrosis factor receptor superfamily	Required for inflammatory response
DLX4	GATA1	Repression	Homeobox family	Represses β-globin gene expression by binding to two silencer elements.
BCL6	SCL	Activation	Zinc finger protein	Acts as a sequence-specific transcriptional repressor by recruiting histone deacetylases.

				Chromosomal translocation results in B-cell lymphomas.
BHLHB2	SCL, E2A	Activation	bHLH family	Regulates chondrocyte differentiation via cAMP pathway.
ID2	SCL	Activation	ID family HLH protein domain	Inhibits function of bHLH transcription factors by heterodimerisation in a dominant neg manner. Regulate cell proliferation and differentiation.
ZNF117	SCL	Activation	Zinc finger protein	Unknown
ZNF304	SCL	Activation	Zinc finger protein	Unknown
ZNF281	SCL	Repression	Zinc finger protein	Unknown
ATRX	SCL	Repression	Zinc finger protein (PHD finger)	Mutations in the XH2 gene cause the alpha-thalassemia/mental retardation syndrome. Represses α-globin expression. Interacts with EZH2, a chromatin regulator.
RUNX1	SCL	Repression	RUNX family Runt domain protein	Chromosome translocations of RUNX1 are associated with leukaemia. Fusion partner of ETO in acute leukaemia. Required for definitive haematopoiesis and bone cell development.
HHEX	E2A	Activation	Homeobox family	Functions as a transcriptional repressor in liver cells and may be involved in the differe and/or maintenance of the differentiated state in hepatocytes. Implicated in haematopoietic and endothelial development. Regulatory region contains the SCL stem cell enhancer.
GTF2I	E2A	Activation	Zipper-like motif Helix-loop/span-helix motif	General transcription factor. Subunit of a chromatin-modifying complex.
MBNL2	E2A	Activation	Zinc finger protein	Implicated in myotonic dystrophy, a neuromuscular disorder.
RUNX2	E2A	Activation	RUNX family Runt domain protein	Master regulator of bone development. Transcriptional regulator of bone lineage specific genes.
FOXO3A	E2A	Repression	Forkhead domain	Chromosomal translocation involved in acute lymphoblastic leukaemia. Triggers apoptosis by inducing the expression of genes that are critical for cell death. Regulates erythroid development. Implicated in haematopoietic cell renewal.
HOXA1	E2A	Repression	Homeobox family	Implicated in neural, inner ear and cardiovascular development.

Table 4.3. Differentially-expressed genes of GATA1, SCL and E2A selected for validation by quantitative PCR. The transcription factor regulating the putative target genes, mode of regulation, protein family and functions of putative target genes are listed in each column.


A
Validation of differentially-expressed genes in GATA1 KD
% of mRNA remained in GATA1 knockdown relative to firefly pGL3 luciferase control
300
250
200
150
100
50
0
ASCL2
CITED2
GFI1B
LZTFL1
MYC
NFE2
TBX1
EKLF
LMO2
LDB1
E2A
PPARD*
TNFRSF1A*
DLX4*
Differentially-expressed genes in GATA1 KD


B
Validation of differentially-expressed genes in SCL KD
% of mRNA remained in SCL knockdown relative to firefly pGL3 luciferase control
160
140
120
100
80
60
40
20
0
BCL6
BHLHB2
CITED2
ID2
MYC
ZNF117
ZNF304
E2A
ZNF281*
ATRX*
RUNX1*
Differentially-expressed genes in SCL KD


C
Validation of differentially-expressed genes in E2A KD
% of mRNA remained in E2A knockdown relative to firefly pGL3 luciferase control
140
120
100
80
60
40
20
0
BHLHB2
CITED2
HHEX
GTF2I
MBNL2
MYC
RUNX2
LMO2
FOXO3A*
HOXA1*
Differentially-expressed genes in E2A KD

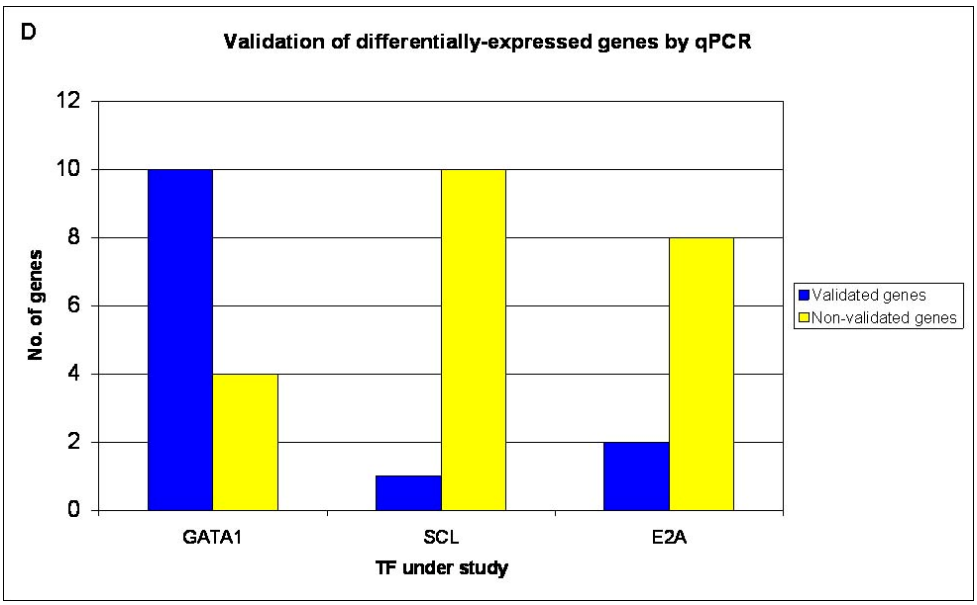


Figure 4.16. Validation of differentially-expressed genes by quantitative PCR. Expression of differentially-expressed transcription factor of GATA1, SCL and E2A were studied by quantitative PCR. mRNA of differentially-expressed genes was compared between the specific knockdown and the luciferase control in two independent biological replicates. In panels A, B and C, the bar charts shows the mRNA level of differentially-expressed genes in the knockdown compared to the control in the qPCR analyses. The error bars show the standard error between the two independent biological replicates. The genes marked with an asterisk are genes identified to be up-regulated in the original Affymetrix experiments. Panel A, validation of differentially-expressed genes in GATA1 knockdown; panel B, validation of differentially-expressed genes in SCL knockdown; panel C, validation of differentially-expressed genes in E2A knockdown. Panel D showed the number of validated and non-validated differentially-expressed genes for each transcription factor under study with the selected fold change cut-off for each knockdown.

The changes in mRNA expression of the chosen differentially-expressed genes were compared between the results obtained in the quantitative PCR and Affymetrix GeneChip (Table 4.4). In general, the changes in mRNA expression were shown to be larger in the results obtained in the Affymetrix GeneChip. The median coefficient of variation of the validated genes (labelled in yellow boxes in Table 4.4) was 9.74% while that of the non-validated genes was 43.37%. This indicates the change in expression of the non-validated genes deviated more than 4 times more from the Affymetrix GeneChip data the validated gene set.

A) GATA1 KD

Differentially expressed gene	% of mRNA remained		CV
	Affy Gene Chip	qPCR	
ASCL2	42.27	40.74	2.60
CITED2	37.03	51.79	23.49
GFI1B	36.94	46.20	15.76
LZTFL1	58.89	83.69	24.60
MYC	42.47	58.41	22.34
NFE2	46.10	71.90	30.93
TBX1	44.80	90.89	48.04
EKLF	13.79	32.08	56.38
LMO2	53.86	103.81	44.81
LDB1	59.38	65.58	7.02
E2A	54.13	73.17	21.16
PPARD*	174.85	200.72	9.74
TNFRSF1A*	310.87	239.91	18.22
DLX4*	149.59	156.94	3.39

B) SCL KD

Differentially expressed gene	% of mRNA remained		CV
	Affy Gene Chip	qPCR	
BCL6	36.88	111.93	71.33
BHLHB2	47.38	98.77	49.73
CITED2	46.59	85.06	41.33
ID2	30.05	99.20	75.66
MYC	63.15	95.43	28.79
ZNF117	54.26	58.38	5.17
ZNF304	57.58	88.42	29.87
E2A	65.21	134.08	48.88
ZNF281*	140.96	68.72	48.72
ATRX*	199.55	78.57	61.51
RUNX1*	142.87	69.39	48.95

C) E2A KD

Differentially expressed gene	% of mRNA remained		CV
	Affy Gene Chip	qPCR	
BHLHB2	46.03	90.83	46.29
CITED2	49.46	91.15	41.93
HHEX	58.86	54.58	5.34
GTF2I	54.57	62.02	9.03
MBNL2	53.32	82.19	30.13
MYC	52.94	92.89	38.74
RUNX2	57.81	85.38	27.23
LMO2	47.14	105.80	54.25
FOXO3A*	159.09	102.03	30.90
HOXA1*	158.44	110.79	25.03

Table 4.4. Comparison of changes in mRNA expression of differentially-expressed genes between Affymetrix GeneChip and quantitative PCR. The % of mRNA remained after siRNA knockdown analysed in Affymetrix GeneChip and qPCR and the coefficient of variation (CV) between the results obtained in the two assays are shown in the tables. Table A: differentially-expressed genes in GATA1 knockdown; Table B: differentially-expressed genes in SCL knockdown; Table C: differentially-expressed genes in E2A knockdown. The validated genes (selected with a cut-off of described above) are highlighted in yellow. Up-regulated genes are marked with an asterisk.

4.4.3.4 Further study and classification of differentially-expressed genes

To provide evidence for the reliability of the Affymetrix GeneChip datasets, the up- or down-regulated probes identified from the analysis above were further studied in terms of their functional classifications, comparison with published target genes and auto-regulation of the SCL erythroid complex.

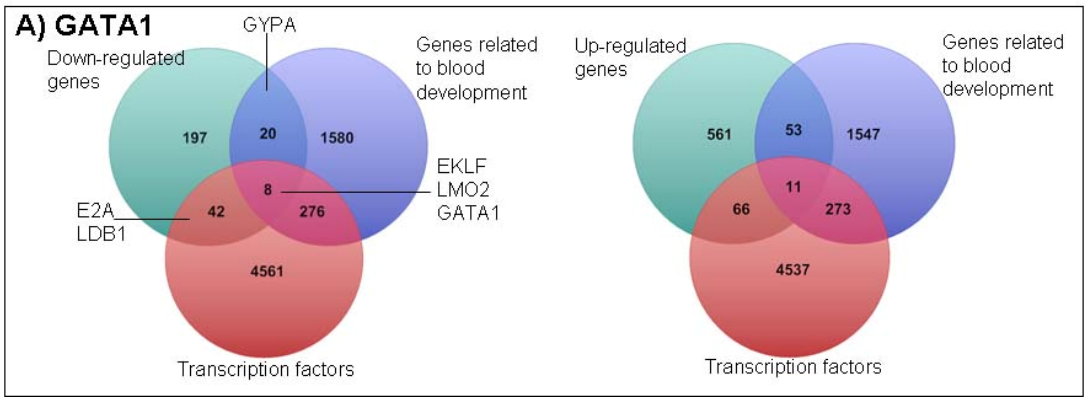
(i) Transcription factors and genes involved in haematopoiesis

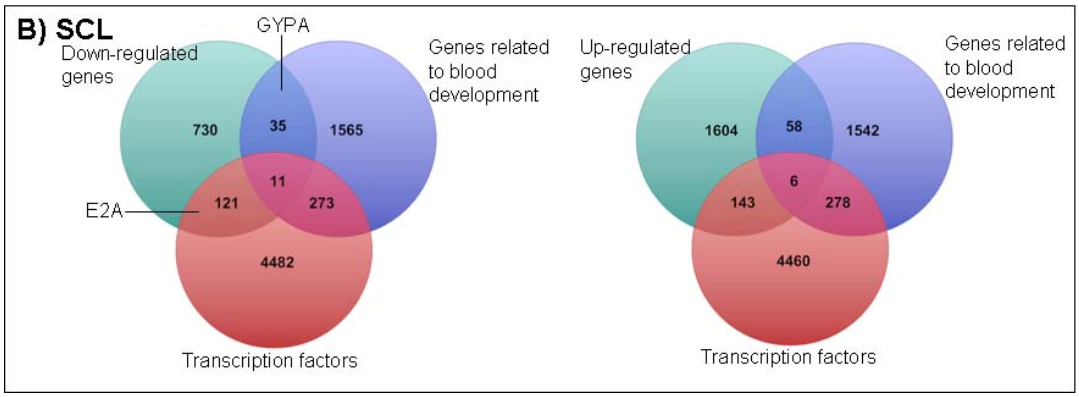
The putative target gene lists derived from the transcription factor knockdown experiments were examined to (i) identify those genes that have been previously shown to be involved in haematopoiesis, and (ii) those genes that were transcription factors. This would allow us to determine (i) whether targets of the SCL erythroid complex identified in K562 cells were representative of haematopoiesis, and (ii) allow a direct comparison of transcription factor targets found in ChIP-on-chip studies using a transcription factor promoter array (see Chapter 5). Venn diagrams were used to study the number of transcription factors and haematopoietic-specific genes for each activated or repressed gene list (Figure 4.17). Gene lists for transcription factors and haematopoietic-specific genes were defined by Philippe Couttet and David Vetrie (Sanger Institute) using lists of all known human transcription factors downloaded from ENSEMBL (including transcription factors and chromatin modifiers/remodelers) and genes known to be expressed and have specific roles during haematopoiesis (including genes important in both haematopoietic and endothelial lineages since both share a common early precursor, the haemangioblast). In total, 1884 and 4887 probe sets found on the Affymetrix GeneChips were found to represent haematopoietic-specific genes and genes encoding transcription factors respectively. These figures were used to derive the percentages of target genes in each class as shown in Table 4.5. P-values associated with each class were also derived using the chi-squared test, which tests a null hypothesis at the frequency distribution of certain events observed in a sample is consistent with a particular theoretical distribution.

Based on the proportion of probe sets on the Affymetrix arrays which were haematopoietic-specific (1884 out of 54614), the likelihood of detecting differentially-expressed haematopoietic-specific genes by chance was approximately 3.4%. However, 9.5% (P-value <0.0001) of the GATA1 target

genes were haematopoietic-specific, suggesting that the GATA1 knockdown experiment was able to enrich for the identification of haematopoietic-specific genes. Similarly, E2A perturbed in the knockdown experiments also showed enrichment for haematopoietic-specific genes (6.8%, P-value 0.0190). This provided confidence that the K562 biological system and the experimental approach were not identifying random events unrelated to blood development. However, no significant enrichments were seen for SCL, LDB1 and LMO2 (3.8% to 4.3%) (see Discussion of this Chapter).

Based on the proportion of probe sets on the Affymetrix arrays which were specific for genes encoding transcription factors (4887 out of 54614), the likelihood of detecting differentially expressed genes encoding transcription factors by chance was approximately 8.9%. However, 14.9% (P-value 0.0360) of target genes identified by the knockdown experiment of E2A were transcription factors (P-value<0.05 is considered to be significant). Between 9-13.3% of the target genes for the other three transcription factors (SCL, GATA1, LMO2 and LDB1) were transcription factors. However, only the down-regulated gene lists of SCL (14.7%, P-value 0.0360) and LMO2 (22.2%, P-value<0.0001) were enriched with transcription factors. This suggests that at least one transcription factor specifically enriched for other transcriptional regulators, suggesting that the SCL erythroid complex may have an important role in regulating transcriptional cascades in K562 cells (see also the discussion for this Chapter).





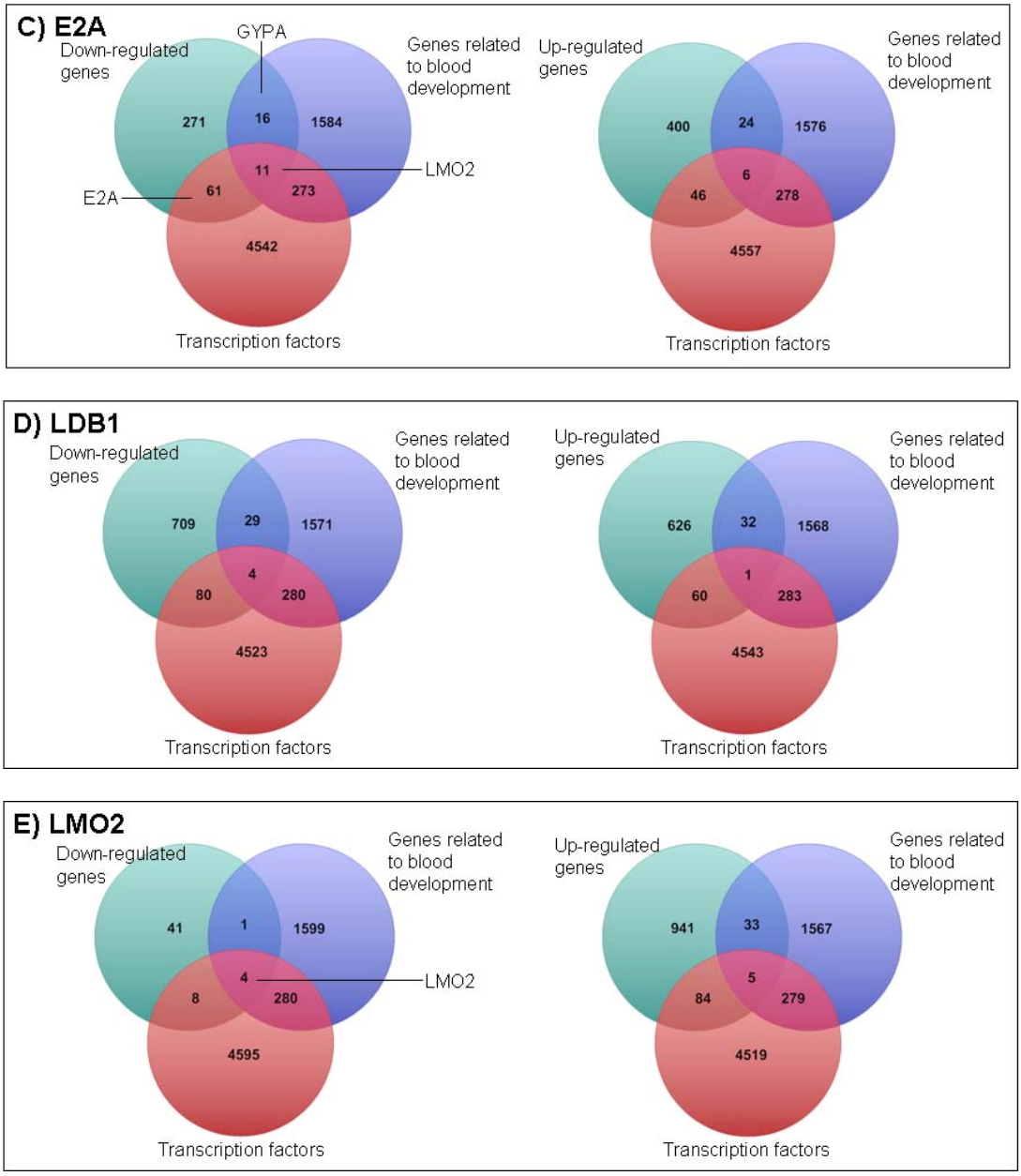


Figure 4.17. Venn diagrams comparison of up- or down-regulated gene lists with haematopoietic-specific gene list and transcription factors. Down-regulated genes are shown in the Venn diagrams on the left while up-regulated genes are shown in the Venn diagrams on the right. Numbers shown in the Venn diagrams are numbers of probe sets representing different or same genes in the human genome. In each Venn diagram, the top left green circle represents the up- or down-regulated genes picked up in the siRNA knockdown study, the top right blue circle represents the haematopoietic genes and the lower red circle represents transcription factors. Some interesting target genes are labelled in the Venn diagram. Panel A: Venn diagrams of GATA1; panel B: Venn diagrams of SCL; panel C: Venn diagrams of E2A; panel D: Venn diagrams of LDB1; panel E: Venn diagram of LMO2.

TF	Category (HG: haematopoietic genes; TF: transcription factor)	Down-regulated genes		Up-regulated genes		All differentially-expressed genes	
		Percentage	P-value	Percentage	P-value	Percentage	P-value
GATA1	HG	10.5%	<0.0001	9.3%	0.0004	9.5%	<0.0001
	TF	18.7%	0.0005	11.1%	0.4846	13.3%	0.1622
SCL	HG	5.1%	0.2410	3.5%	0.5577	4.1%	0.5577
	TF	14.7%	0.0360	8.2%	0.7268	10.4%	0.7268
E2A	HG	7.5%	0.0034	6.3%	0.0786	6.8%	0.0190
	TF	20.1%	0.0001	10.9%	0.4846	14.9%	0.0360
LDB1	HG	4%	0.5577	4.6%	0.2410	4.3%	0.5577
	TF	10.2%	0.7268	8.5%	1.0000	9.4%	1.0000
LMO2	HG	9.3%	0.0004	3.6%	0.5577	3.8%	0.5577
	TF	22.2%	<0.0001	8.4%	0.7268	9%	1.0000

Table 4.5. Percentages and P-values of haematopoietic genes and transcription factors in the differentially-expressed gene lists for each member of the SCL erythroid complex.

(ii) Gene Ontology classification

The activated or repressed genes for each transcription factor were also classified according to the terms found in the Gene Ontology (GO) database. The GO project describes functions of gene products in three different categories: cellular components, biological processes and molecular functions. The differentially-expressed genes of each transcription factor knockdown were studied to identify statistically significant GO terms using GO Term Finder (http://go.princeton.edu/cgi-bin/GOTermFinder) (Boyle et al., 2004). GO terms which are over-represented in the differentially-expressed genes lists compared to the whole human genome with a P-value of <0.01 were identified. The GO terms in the three categories, the associated P-values, the percentage in the differentially-expressed gene lists and in the human genome are included in Appendix 2.

The GO terms in the biological process and molecular function categories which appeared in more than one differentially-expressed gene list are shown Table 4.6. Three GO biological process terms (chromatin modification, regulation of transcription from RNA polymerase II promoter and transcription from RNA polymerase II promoter) are related to the regulation of gene expression. Four molecular function terms (transcription activator activity, transcription regulator activity, transcription cofactor activity and transcription factor binding) are also related to transcription. This illustrates that the five members of the SCL erythroid complex regulate a number of downstream target genes which are related to the regulation of transcription. Nine of the GO biological process terms are related to programmed cell death or apoptosis. This indicates that members of the SCL erythroid complex may also regulate a number of genes related to the apoptotic pathway.

	GATA1		SCL		E2A		LDB1		LMO2	
Biological process	Down-regulated	Up-regulated	Down-regulated	Up-regulated	Down-regulated	Up-regulated	Down-regulated	Up-regulated	Down-regulated	Up-regulated
regulation of cell proliferation										
regulation of developmental process										
regulation of macromolecule metabolic process										
regulation of programmed cell death										
regulation of signal transduction										
anatomical structure development										
apoptosis										
biological regulation										
biopolymer metabolic process										
cell cycle										
cell cycle process										
cell death										
cell motion										
cell proliferation										
cellular component organization and biogenesis										
cellular developmental process										
cellular metabolic process										
chromatin modification										
death										
developmental process										
endomembrane system										
establishment of protein localization										
gene expression										
intracellular signaling cascade										
macromolecule metabolic process										
localization of cell										
macromolecule localization										
mRNA processing										
multicellular organismal development										
multicellular organismal process										
negative regulation of apoptosis										
negative regulation of biological process										
negative regulation of cell proliferation										
negative regulation of cellular process										
negative regulation of developmental process										
negative regulation of programmed cell death										
nucleobase, nucleoside, nucleotide and nucleic acid metabolic process										
organ development										
positive regulation of biological process										
positive regulation of cellular process										
positive regulation of developmental process										
primary metabolic process										
programmed cell death										
protein kinase cascade										
protein localization										
protein transport										
regulation of apoptosis										
regulation of cell cycle										
regulation of cell proliferation										
regulation of cellular metabolic process										
regulation of developmental process										
regulation of macromolecule metabolic process										
regulation of metabolic process										
regulation of programmed cell death										
regulation of transcription from RNA polymerase II promoter										
RNA metabolic process										
RNA splicing										
system development										
transcription from RNA polymerase II promoter										
Molecular function										
enzyme binding										
kinase binding										
protein binding										
RNA binding										
transcription activator activity										
transcription cofactor activity										
transcription factor binding										
transcription regulator activity										

Table 4.6. Gene Ontology classification of differentially-expressed genes for each of the five members of the SCL erythroid complex. The GO terms associated with biological processes (top) and molecular functions (bottom) and significantly enriched in more than one of the differentially-expressed gene lists are shown. The blue boxes indicate the GO terms which are statistically significant in the up- or down-regulated gene lists in the knockdown study of five members of the SCL erythroid complex (P value < 0.1).

(iii) Identification of published target genes

The differentially-expressed genes were compared with the published target genes of the transcription factors (see Chapter 1, section 1.4.2). GATA1, SCL and E2A were all found to regulate one of the three known target genes of the SCL erythroid complex - GYPA (c-kit and α-globin being the other two). GATA1 was shown to regulate 6 out of the 11 published target genes and these included GYPA, EKLF, NFE2, EPOR, MYC and GFI-1B. This indicates that the siRNA-induced knockdown in combination with expression profiling with the GeneChip identified at least some published targets for these transcription factors.

(iv) Auto-regulation of the SCL erythroid complex

Based on the Affymetrix data, the transcription factors of the SCL erythroid complex were also found to regulate other members of the complex itself. For example, GATA1 activated expression of E2A, LDB1 and LMO2. SCL activated expression of E2A while E2A activated expression of LMO2. This indicates that members of the SCL erythroid complex are involved in auto-regulatory loops to regulate the transcription of other proteins involved in the complex.

4.4.3.5 Co-regulation of transcription factors in the SCL erythroid complex

Whilst each of the transcription factors studied here may function alone or in combination with other transcription factors in regulating gene expression, the aim of this project was to identify targets of the SCL erythroid complex. Therefore comparing the differentially-expressed gene lists for each transcription factor and determining which genes were found in more than one list would provide insights into which genes are targets of the SCL erythroid complex. Gene lists were analysed in several ways, by varying the number of members of the SEC in the comparisons, and by including data at the expression outcome (activated or repressed) of the targets (since it was likely that *bona fide* targets of the SCL erythroid complex would be affected in the same way during knockdown of any one of the five transcription factors).

(i) Identification and classification of co-regulated genes

Initially, the putative target genes of GATA1, SCL and E2A were compared as these three transcription factors are bound to DNA directly in the SCL erythroid complex. 102 probe sets representing 92 genes were found to be co-regulated by GATA1, SCL and E2A (Figure 4.18 A).

These 102 probe sets were further classified and studied (see below). To further assess the roles played by the bridging proteins LDB1 and LMO2 in the SCL erythroid complex, the putative target gene lists of GATA1, SCL and E2A were also compared against those of LDB1 and LMO2. Unlike the co-regulation among GATA1, SCL and E2A, only a very small portion of genes (up to 7 probes) were found to be co-regulated by the 3 transcription factors: LDB1, LMO2 and either GATA1, SCL or E2A (Figure 4.18 B-D). However, no genes were found to be co-regulated by all five members of the complex.

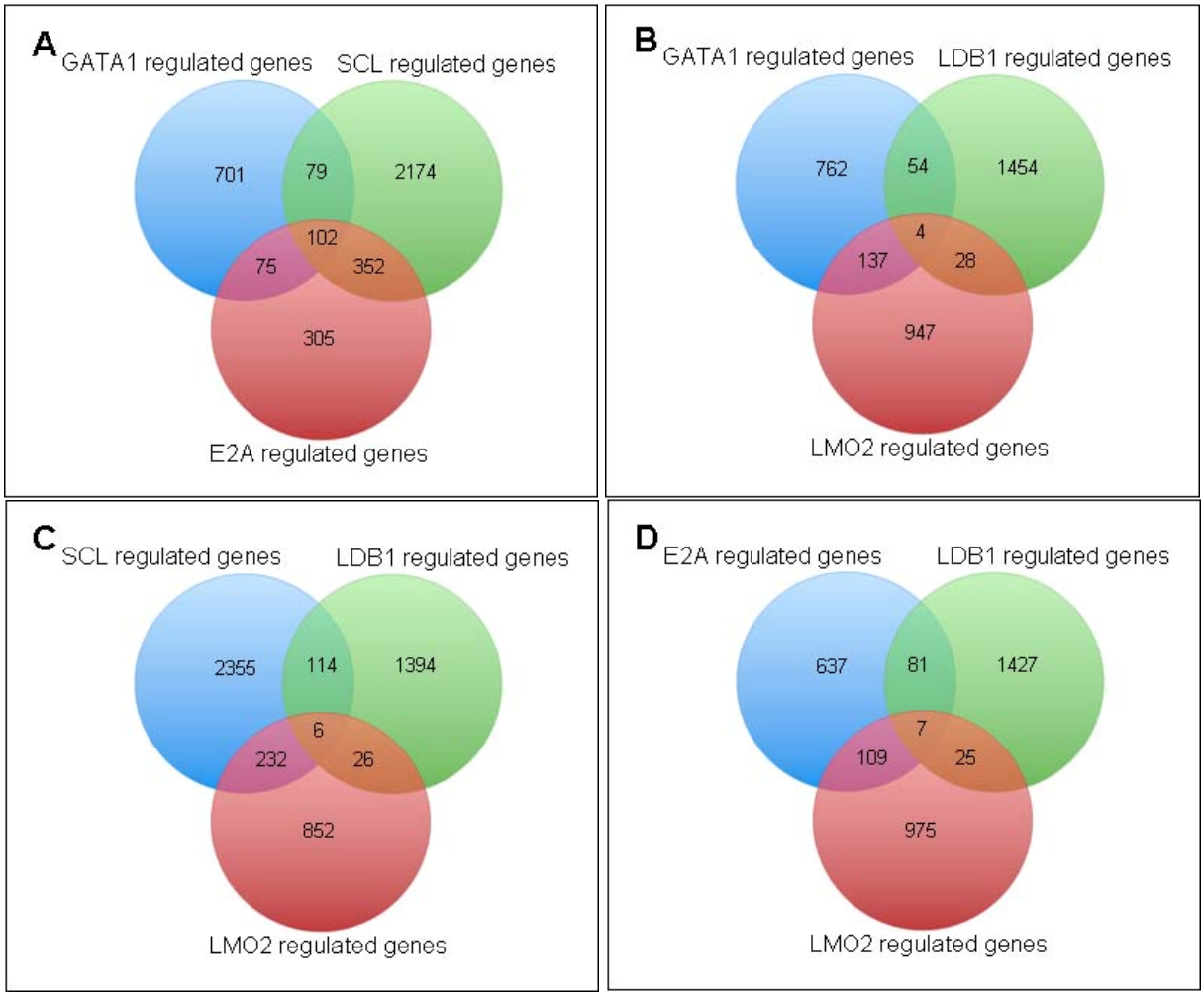


Figure 4.18. Co-regulation of target genes by members of the SCL erythroid complex. Numbers shown in the pie charts are numbers of probe sets representing genes in the human genome. Panel A: co-regulation of GATA1, SCL and E2A; panel B: co-regulation of GATA1, LDB1 and LMO2; panel C: co-regulation of SCL, LDB1 and LMO2; panel D: co-regulation of E2A, LDB1 and LMO2.

Within the group of 92 genes found to be co-regulated by GATA1, SCL and E2A, 19 were transcription factors. Therefore, not surprisingly, these 92 genes were enriched in GO terms related to transcription from RNA polymerase II promoter (Table 4.7). In addition, these 92 genes were also enriched in the protein binding GO term which indicates that these genes may be involved in protein-protein interaction required for the regulation of transcription. These GO classifications again reinforce the idea that the SCL erythroid complex may play a critical role in transcriptional

regulation by regulating other transcription factors and associated factors which are involved in the regulation of transcription and signal transduction activities.

Biological process	P-value	% in gene list	% in genome
regulation of transcription from RNA polymerase II promoter	0.00347	8.51	1.28
RNA metabolic process	0.00467	26.60	12.33
transcription from RNA polymerase II promoter	0.00602	9.57	1.82
Molecular function			
Protein binding	3.6E-06	48.94	26.66
RNA binding	6.3E-05	13.83	2.87
Cellular component			
intracellular part	8.6E-05	58.51	39.25
intracellular organelle	8.8E-05	52.13	32.57
Organelle	8.9E-05	52.13	32.58
intracellular membrane-bounded organelle	0.00013	45.74	26.73
membrane-bounded organelle	0.00013	45.74	26.74
Nucleus	0.00369	32.98	18.30
Intracellular	0.00773	58.51	44.61

Table 4.7. Gene Ontology classification of GATA1, SCL and E2A co-regulated genes. GO terms associated with biological process, molecular function and cellular component significantly enriched in the GATA1, SCL and E2A co-regulated genes are shown. The P-values associated with each GO term, percentage of genes belonging to the GO term in the gene list and in the human genome are also shown.

(ii) Identification of known co-regulated target genes

Within these 92 genes co-regulated by GATA1, SCL and E2A, glycophorin A (GYPA) - which is a published known target of the SCL erythroid complex - was identified. This confirmed that the Affymetrix expression data could detect at least one co-regulated target out of the 3 published co-regulated genes (GYPA, c-kit and α-globin) found in the SCL erythroid complex.

(iii) Auto-regulation of the SCL erythroid complex

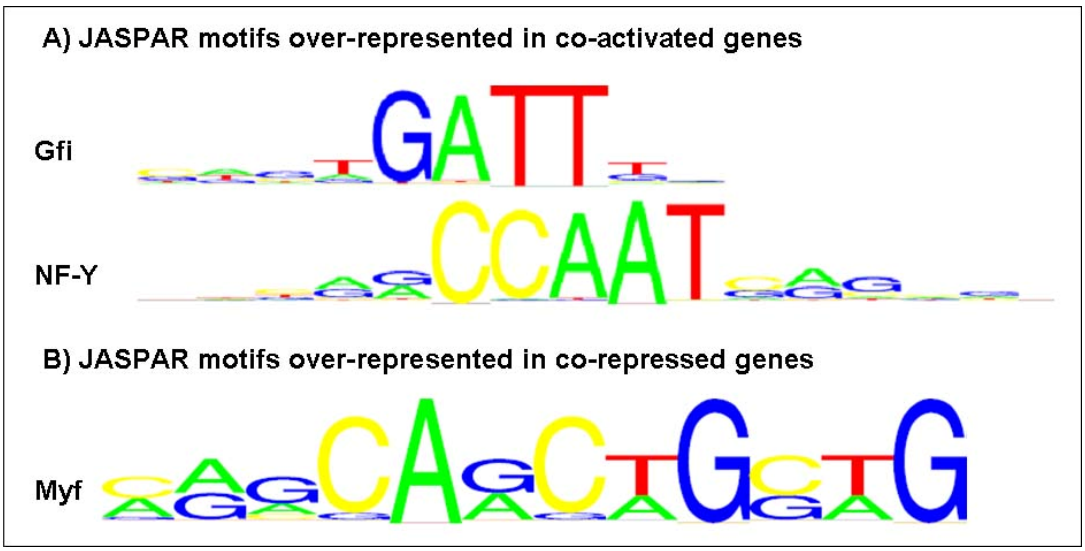
The auto-regulation of the SCL erythroid complex (first mentioned in section 4.4.3.4 part iv) also appeared to extend to co-regulation by more than one member of the complex. Two members of the SEC were identified as putative target genes of GATA1, SCL and E2A. E2A was shown to be activated by both GATA1 and SCL while LMO2 was shown to be activated by GATA1 and E2A.

4.4.4 Motif discovery of co-regulated putative target genes

It is known that the SCL erythroid complex binds to a composite E-box/GATA motif (Wadman et al., 1997) which directly binds to the SCL/E2A heterodimer and GATA1. As a means of confirming whether the putative target genes identified by the knockdown experiments were *bona fide*, motif analysis was performed to determine whether this motif, or any similar ones, was found in common

for the putative target genes of the transcription factor knockdown experiments. The 92 genes co-regulated by SCL, GATA1 and E2A (described in section 4.4.3.5) were studied by two methods to identify transcription factor binding motifs in a one kilobase region covering their known promoter regions identified using FirstEF (Davuluri et al., 2001). Promoter regions were chosen for this analysis, although regulation involving enhancers may play a crucial role as well – however the location of any enhancers was not known. In the first method, the vertebrate motif database JASPAR CORE was used to identify known transcription factor binding motifs that were over-represented within this 1 kb region around the transcription start sites of the genes. The JASPAR CORE database is an open-access database containing curated, non-redundant transcription factor binding site profiles for multicellular eukaryotes which were derived from experimentally verified DNA sequences bound by transcription factors (Sandelin et al., 2004). Two transcription factor binding motifs, Gfi and NF-Y, were found to be over-represented in the co-activated gene list while one motif Myf was over-represented in the co-repressed gene list (Figure 4.19 A and B). The Gfi motif is recognised by the zinc finger protein Gfi family containing the C2H2 motif (Zweidler-Mckay et al., 1996). The NF-Y motif is recognised by the nuclear transcription factor Y family and has a characteristic CCAAT motif (Becker et al., 1991). The Myf motif is a bHLH motif recognised by the myogenic factor family (Wasserman and Fickett, 1998). The composite E-box/GATA motif was not identified by this analysis.

In the second method, the NestedMICA programme was used to perform unbiased motif discovery (Down and Hubbard, 2005) (Chapter 1, section 1.3.4.2). This method allowed us to identify possible novel DNA motifs in the promoter regions of the 92 genes co-regulated by GATA1, SCL and E2A. In this case, three DNA motifs were identified reproducibly in the promoters of the activated genes; no recurrent motifs were found in the repressed genes (Figure 4.19 C). Once again, the composite E-box/GATA motif was not identified by this analysis.



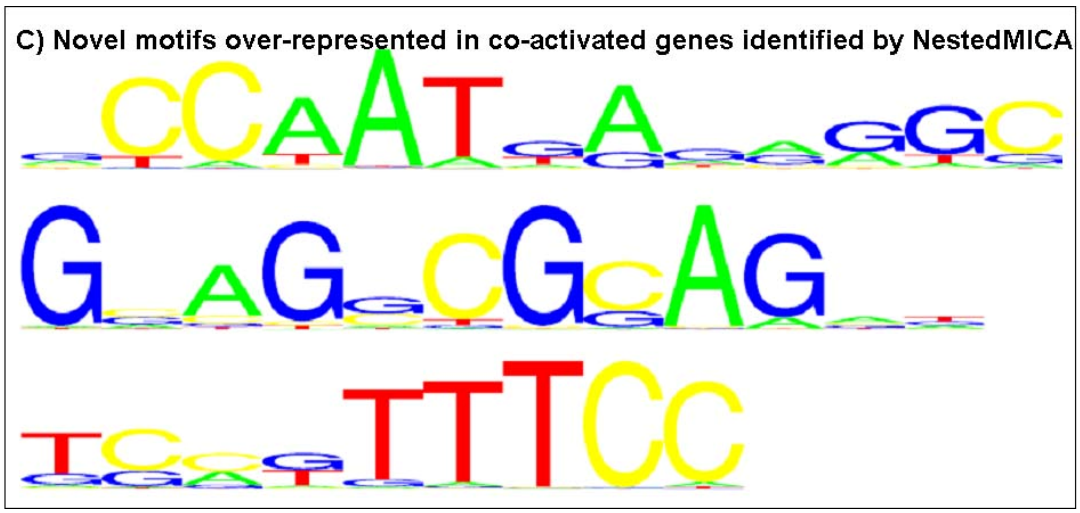


Figure 4.19. Motif discovery of co-regulated genes by GATA1, SCL and E2A. Panel A: annotated motifs identified for the co-activated genes in the JASPAR database; panel B: annotated motifs identified for the co-repressed genes in the JASPAR database; panel C: novel motifs identified for the co-activated genes using NestedMICA. DNA logos are presented for each motif and the height of each letter indicates relative occurrence of nucleotides in the identified binding sites.

4.5 Discussion

The work presented in this Chapter demonstrated the use of a commercial expression array platform, the Affymetrix GeneChip, to study the effects on gene expression in the K562 cell line when members of the SCL erythroid complex were knocked down using siRNAs. This array analysis permitted the identification of putative target genes regulated by each transcription factor in the complex, as these targets would be very likely to change in their levels of expression during the knockdown conditions. The data presented in this Chapter will now be discussed as follows:

4.5.1 Affymetrix GeneChips as a platform of expression profiling

- **Low validation rate by qPCR**

The results obtained from the expression profiling of downstream effects of siRNA knockdown using Affymetrix GeneChip demonstrated that it is not a particular good method of identifying target genes of the transcription factors under study. This conclusion was based on the low validation rate of differentially-expressed genes by qPCR. This suggests that the differentially-expressed genes identified on the Affymetrix GeneChip included a large number of false positive targets. This may be because the TF knockdowns were not sufficient to elicit profound and reproducible changes in target gene expression profiles (because a proportion of the knocked down TF was still present). Thus, quantitative measurements on the Affymetrix platform may not detect such subtleties in expression changes or may have detected changes which were not reproducible

between bioreplicates. This was supported by the fact that GATA1, which had the highest knockdown efficiency, showed the highest rate of qPCR validation.

- **Identification of published targets**

Despite the caveats mentioned above, known targets of members of the complex were identified by Affymetrix analysis. Taking GATA1 as an example, the majority of its published target genes was detected on the Affymetrix expression array. These included EPOR, GYPA, GFI1B, NFE2, MYC and EKLF. However, other published targets, α- and β-globin genes, Epo, FOG-1 and GATA2, were not detected. One of the key downstream targets of the SCL erythroid complex, glycophorin A (GYPA), was shown to be co-activated by GATA1, SCL and E2A in the expression study. However, GYPA was not detected as an activated gene by LDB1 and LMO2, although it is known to be a target gene of the whole SCL erythroid complex (Lahlil et al., 2004). Furthermore, the two other published targets of the SCL erythroid complex - c-kit and α-globin - were not detected by any members of the complex.

A number of reasons may explain why some of the published and novel targets of the complex may not have been identified in the GeneChip analysis:

(i) The knockdown of the transcription factors under study was not 100%. The remaining level of the transcription factors may be sufficient to drive the expression of their target genes. Thus, the change in expression of these target genes during the knockdown may not be significant or reproducible for detection on the expression array.

(ii) For target genes which are regulated by the whole SCL erythroid complex, some members of the complex may be dispensable for the regulation. This may be particularly relevant to LDB1 and LMO2, which do not bind DNA directly, but are bridging proteins. The roles of these proteins may be to stabilise the complex and not participate in direct regulation *per se*.

(iii) It is also possible that other transcription factors, apart from the five members of the complex in question, can compensate for the knockdown effects, thus allowing regulation of target genes even in the absence of a member of the complex.

(iv) The three DNA binding partners (GATA1, SCL and E2A), or combinations thereof, are able to interact within other regulatory complexes which do not include LMO2 or LDB1. This would add an additional layer of complexity onto the analysis and make gene list comparisons more complex.

(v) It may be difficult to determine co-regulation by multiple members of the SCL eyrthroid complex because they could also be acting on target genes independent of the SCL erythroid complex. Thus major effects may be elicited by some knockdowns, but not others.

(vi) Only one time point was studied on the expression array for each knockdown assay. The effects on gene expression may be transient, occur earlier than that was monitored or may take substantially more time after the silencing of the transcription factor. Therefore; not every target gene can be detected at the time point selected.

(vii) The stringency of fold change use in the statistical analyses may also be an issue. The expression changes of some target genes may be very subtle and not satisfy the criteria for selection as differentially-expressed genes.

4.5.2 The SCL erythroid complex regulates transcription factors

The differentially-expressed genes identified in the gene expression profiling for each transcription factor were over-represented for transcription factors. Between 9% and 14.9% of target genes identified by the five transcription factor knockdown experiments were transcription factors (section 4.4.3.4). This suggests that the SCL erythroid complex may play a crucial role regulating haematopoietic transcriptional networks in K562 cells. This makes sense, given the role of SCL as a master regulator of haematopoiesis.

4.5.3 Identification of haematopoietic-related genes regulated by members of the SCL erythroid complex

Enrichments of haematopoietic-related genes were observed in the differentially-expressed gene lists for members of the SCL erythroid complex during knockdown (section 4.4.3.4). The percentage of haematopoietic-specific genes of the 5 transcription factors ranged from 3.8% to 9.5%. This data confirms that knockdown of members of the SCL erythroid complex does induce changes to genes which have known roles in haematopoietic development. The percentage of haematopoietic-specific genes in the GATA1 study was the highest among the 5 transcription factors (9.5%). This is because it is an important regulator of erythroid development and the knockdown efficiency with siRNA was the highest for the five TFs studied. The percentages for SCL and LDB1 were the lower (4.1% and 4.3% respectively). For SCL, only one siRNA was used in the expression profiling and thus many of the differentially-expressed genes identified may be off-targets, thus resulting in a lower haematopoietic-specific effect. LDB1 is a ubiquitously-expressed gene. Therefore, its target gene list may reflect other cellular events than those associated purely with haematopoiesis. Yet, E2A is a ubiquitously-expressed gene and should also have identified a high degree of non-haematopoietic-related target genes. However, E2A identified 1.5 times as many haematopoietic targets as LDB1 (6.8%). This may be due to the fact that it is a known interacting partner of SCL, and such dimerisation is a requirement for DNA-binding (Hsu et al., 1994). LMO2 is expressed in haematopoietic progenitors and is required for erythropoiesis and

theoretically, a large percentage of its putative target genes should also have been haematopoietic-related. However, LMO2 showed the lowest levels of enrichment for haematopoietic-related genes (3.8%). One possible reason is that the siRNA-induced knockdown could not be monitored at the protein level for LMO2 (due to the lack of an antibody which worked well in western analysis). Therefore, there was no way of knowing whether the time point chosen to identify relevant targets was appropriate.

4.5.4 Auto-regulation of the SCL erythroid complex

Previous studies have demonstrated that GATA1 is a regulator of SCL expression. Indeed, the SCL +51 enhancer has also been shown to be bound by at least three members of the SCL erythroid complex SCL, GATA1 and LDB1 (Chapter 1, section 1.4.2.1). Thus, SCL may indeed be regulated by the whole SEC complex or by no fewer than three of its members. The expression data obtained from this Chapter further characterised the auto-regulatory role of this complex. E2A was found to be activated by both GATA1 and SCL while LMO2 was activated by E2A and GATA1. LDB1 was also an activated target gene by GATA1. This data suggests that there are tightly controlled regulatory complexities which may govern the activity of the SEC.

This also highlights a further challenge to analysing the targets of the SEC - the knockdown of one member of the complex may change the expression of another member of the complex. This idea will be explored further in Chapter 6 of this thesis.

4.5.5 Motif discovery at target genes

The motif discovery analyses for the 92 genes co-regulated by GATA1, SCL and E2A offered little insights into their regulation by the transcription factors of the SCL erythroid complex. This was because the expected E-box/GATA composite motif was not identified in the promoters of this set of target genes. However, there are plausible explanations for this. Firstly, a 1 kb region around the TSS was used in these motif discovery analyses. Gene regulation and binding by transcription factors may occur outside this 1 kb window at regulatory elements such as proximal or distal enhancers, or silencers or even at distal promoters. Secondly, the siRNA-induced knockdown in combination with Affymetrix expression analysis identifies both direct and indirect target genes regulated by the transcription factor. Thus, the gene list which was used in the motif discovery contains both types of target genes - and it is highly unlikely that indirect targets would require an E-box GATA consensus motif. This would make consensus motifs difficult to derive from such a mixed set of targets. Despite this, a few additional motifs were identified in these targets. These motifs could possibly represent (i) new TFBS for other transcription factors which are required by direct targets which also bind the SEC, or (ii) are sites which bind factors required by both direct

and indirect targets of the SEC to mediate transcriptional control at various levels of a transcriptional cascade. The delineation of direct targets of the SCL erythroid complex, from indirect ones, will be the basis of the next chapter in this thesis.

4.6 Conclusions

The work presented in this Chapter identified genes involved in haematopoiesis and transcription factors as the downstream targets of members of the SCL erythroid complex. However, the validation data demonstrated that the Affymetrix GeneChip platform generated a high false positive rate. Therefore; results from this Chapter should be carefully interpreted when further analyses are being performed and compared in the following Chapters.

Becker, D.M., Fikes, J.D., and Guarente, L. (1991). A cDNA encoding a human CCAAT-binding protein cloned by functional complementation in yeast. Proceedings of the National Academy of Sciences of the United States of America *88*, 1968-1972.
Boyle, E.I., Weng, S., Gollub, J., Jin, H., Botstein, D., Cherry, J.M., and Sherlock, G. (2004). GO::TermFinder--open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. Bioinformatics (Oxford, England) *20*, 3710-3715.
Davuluri, R.V., Grosse, I., and Zhang, M.Q. (2001). Computational identification of promoters and first exons in the human genome. Nat Genet *29*, 412-417.
Down, T.A., and Hubbard, T.J. (2005). NestedMICA: sensitive inference of over-represented motifs in nucleic acid sequence. Nucleic Acids Res *33*, 1445-1453.
Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. Proceedings of the National Academy of Sciences of the United States of America *95*, 14863-14868.
Hsu, H.L., Huang, L., Tsan, J.T., Funk, W., Wright, W.E., Hu, J.S., Kingston, R.E., and Baer, R. (1994). Preferred sequences for DNA recognition by the TAL1 helix-loop-helix proteins. Mol Cell Biol *14*, 1256-1265.
Irizarry, R.A., Bolstad, B.M., Collin, F., Cope, L.M., Hobbs, B., and Speed, T.P. (2003a). Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res *31*, e15.
Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U., and Speed, T.P. (2003b). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics *4*, 249-264.
Lahlil, R., Lecuyer, E., Herblot, S., and Hoang, T. (2004). SCL assembles a multifactorial complex that determines glycophorin A expression. Mol Cell Biol *24*, 1439-1452.
Landry, J.R., Kinston, S., Knezevic, K., de Bruijn, M.F., Wilson, N., Nottingham, W.T., Peitz, M., Edenhofer, F., Pimanda, J.E., Ottersbach, K.*, et al.* (2008). Runx genes are direct targets of Scl/Tal1 in the yolk sac and fetal liver. Blood *111*, 3005-3014.
Lin, Y.W., and Aplan, P.D. (2007). Gene expression profiling of precursor T-cell lymphoblastic leukemia/lymphoma identifies oncogenic pathways that are potential therapeutic targets. Leukemia *21*, 1276-1284.
Muntean, A.G., and Crispino, J.D. (2005). Differential requirements for the activation domain and FOG-interaction surface of GATA-1 in megakaryocyte gene expression and development. Blood *106*, 1223-1231.
Palomero, T., Odom, D.T., O'Neil, J., Ferrando, A.A., Margolin, A., Neuberg, D.S., Winter, S.S., Larson, R.S., Li, W., Liu, X.S.*, et al.* (2006). Transcriptional regulatory networks downstream of TAL1/SCL in T-cell acute lymphoblastic leukemia. Blood *108*, 986-992.
Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W.W., and Lenhard, B. (2004). JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res *32*, D91-94.
Siddiqui, A.S., Delaney, A.D., Schnerch, A., Griffith, O.L., Jones, S.J., and Marra, M.A. (2006). Sequence biases in large scale gene expression profiling data. Nucleic Acids Res *34*, e83.
Singh-Gasson, S., Green, R.D., Yue, Y., Nelson, C., Blattner, F., Sussman, M.R., and Cerrina, F. (1999). Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. Nat Biotechnol *17*, 974-978.
Tusher, V.G., Tibshirani, R., and Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. Proceedings of the National Academy of Sciences of the United States of America *98*, 5116-5121.
Wadman, I.A., Osada, H., Grutz, G.G., Agulnick, A.D., Westphal, H., Forster, A., and Rabbitts, T.H. (1997). The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. Embo J *16*, 3145-3157.

Wasserman, W.W., and Fickett, J.W. (1998). Identification of regulatory regions which confer muscle-specific gene expression. J Mol Biol *278*, 167-181.
Welch, J.J., Watts, J.A., Vakoc, C.R., Yao, Y., Wang, H., Hardison, R.C., Blobel, G.A., Chodosh, L.A., and Weiss, M.J. (2004). Global regulation of erythroid gene expression by transcription factor GATA-1. Blood *104*, 3136-3147.
Wu, Z., and Irizarry, R.A. (2005). Stochastic models inspired by hybridization theory for short oligonucleotide arrays. J Comput Biol *12*, 882-893.
Zweidler-Mckay, P.A., Grimes, H.L., Flubacher, M.M., and Tsichlis, P.N. (1996). Gfi-1 encodes a nuclear zinc finger protein that binds DNA and functions as a transcriptional repressor. Molecular and cellular biology *16*, 4024-4034.

Chapter 5

ChIP-on-chip analyses of the SCL erythroid complex

5.1 Introduction

The expression analyses of siRNA knockdown described in the previous Chapters allowed us to identify putative target genes regulated by each of five transcription factors found in the SCL erythroid complex. However, the limitations of these types of studies mean that they do not provide direct information regarding the binding of transcription factors to the regulatory regions of target genes. Complementary methods are required to further investigate such protein-DNA interactions at *cis*-regulatory elements of target genes, thus allowing such genes to be considered as *bona fide* direct target genes of the transcription factors. Many methods, both traditional and high-throughput, have been developed and characterised for the study of protein-DNA binding and for the identification of regulatory DNA elements (Chapter 1, section 1.3.3). Traditional low-throughput methods are time-consuming, and in many cases, they are based on DNA-protein binding *in vitro*. The development of ChIP-on-chip as an *in vivo* technique in the last decade has significantly enhanced the scale and spectrum of specificity for identifying transcription factor or other protein-bound DNA elements. At the time this project was first initiated, massively parallel sequencing had not been fully developed - microarrays were still playing the leading role in high-throughput genome-wide ChIP studies. Therefore, ChIP-on-chip analysis was used to identify direct targets for the five transcription factors of the SCL eythroid complex as described in this Chapter.

5.1.1 ChIP-on-chip: principles and issues

In ChIP-on-chip, cells or tissues are extracted and the DNA-protein complexes are cross-linked with formaldehyde. The cross-linked complexes are sonicated to shear the DNA into fragments – the amount of sonication determines the extent of shearing and typically the DNA is sheared to between 200 bp and 1 kb. The DNA-protein complexes are then immunoprecipitated with antibodies specific to a protein bound to the DNA. The immunoprecipitated and (non-immunoprecipitated “input” control sample) DNA-protein complexes are then de-crosslinked, and the ChIP and input DNA are extracted. Because of the amount of ChIP DNA recovered, it is quite often amplified by PCR prior to use in microarray analyses. The ChIP DNA and input DNA are then fluorescently labelled with two different dyes, such as Cy5 and Cy3, and hybridised onto genomic arrays of interest (Figure 5.1).

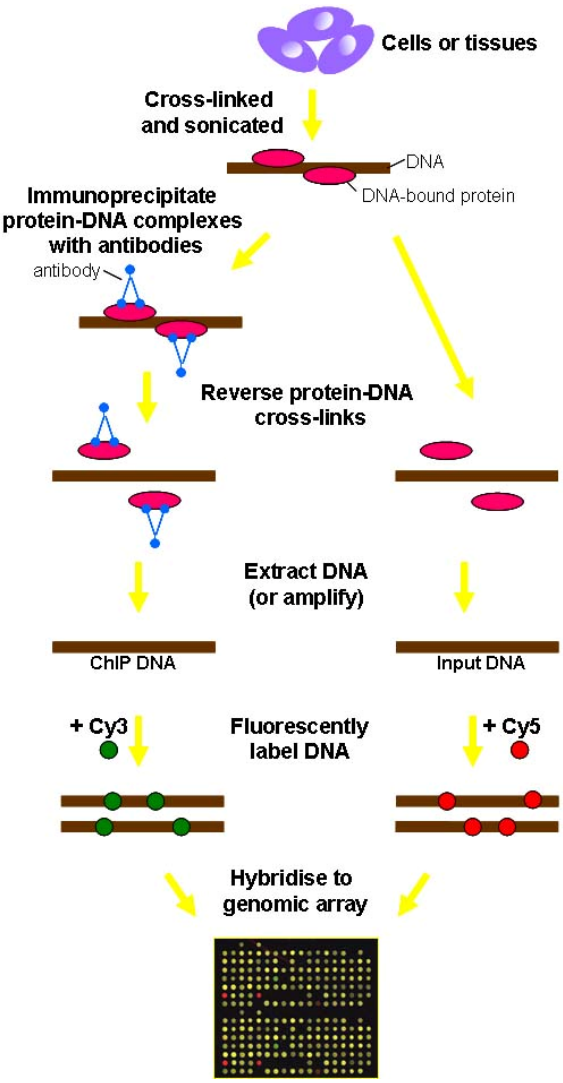


Figure 5.1. The principle of ChIP-on-chip. Flow diagram shows the steps involved in the method. Briefly, DNA-protein complexes are crosslinked, sonicated and immunoprecipitated with specific antibodies. Crosslinks of DNA-protein complexes are reversed and DNA extracted, labelled with fluorescent dyes and hybridised onto genomic arrays.

A number of issues should be considered carefully when performing ChIP-on-chip analyses. These factors ensure that the data obtained is of high quality. Some of them are discussed in details below.

- **Cross-linking**

Cross-linking between protein and DNA is the key factor affecting subsequent steps in a ChIP-on-chip experiment. Formaldehyde is commonly used for cross-linking between protein and DNA, as well as among proteins (Orlando et al., 1997). DNA elements bound by multiprotein complexes, where many of the protein components do not directly bind DNA, can also be studied. The type of protein-DNA interaction being cross-linked depends on the concentration of formaldehyde and the

length of time of cross-linking. As a result, different cell types or different protein-DNA interactions may require optimisation. For large DNA-binding protein complexes, long-range cross-linkers such as dimethyl adipimidate (DMA) can be used in combination with formaldehyde (Zeng et al., 2006). In cases where the protein-DNA interaction is relatively strong such as histone proteins, native ChIP, where no cross-linking is required, can be performed (O'Neill and Turner, 1996; O'Neill et al., 2006).

- **Antibodies**

The quality of antibodies used in ChIP-on-chip experiment is the most critical parameter determining the experimental outcome. Some commercial antibodies are validated and marketed for ChIP applications. However, for most antibodies, validation is performed by the experimenter, and often several antibodies are tested for each ChIP assay. To select antibodies that work well in ChIP, ChIP-qPCR of DNA regions where the protein is known to bind (if known) is useful to perform. Ultimately the best antibodies are those which can pick up specific protein *in vivo* and do not cross-react with other proteins or proteins of the same family. The easiest way to check the specificity of an antibody is by western blotting. To further ensure that the antibody only bind to the protein under study, siRNA knockdown can be used to silence the protein in the cell type and the knockdown of the relevant protein can be quantified by western blotting. Furthermore, the epitope recognised by the antibodies should be carefully selected. DNA-binding motifs or protein-interacting motifs of transcription factors or histone proteins are usually involved directly in DNA or protein binding and are masked during cross-linking. In these cases, it would be difficult for the antibodies to recognise these masked epitopes.

- **Cell numbers**

Traditionally, a large number of cells (usually 10^7 cells) is required for each ChIP assay. This is the main limiting factor for ChIP experiments performed in primary cells or cells/tissue types where the cell number is limiting (such as stem cells), especially in mammalian systems. Many protocols have been developed to circumvent this issue. Carrier ChIP (CChIP) was developed to perform ChIP in combination with qPCR with as few as 100 cells with the addition of *Drosophila* cells as the carrier agent in native non-crosslinked condition. This was successfully applied in mouse for the study of histone modifications (O'Neill et al., 2006). However, one of the drawbacks of this method is that the carrier agents may interfere with the profile of the native protein-DNA interaction. Also, CChIP cannot be used in formaldehyde cross-linked materials due to the low recovery rate. Other methods have also been developed to solve the cell number issue in cross-linked material. MiniChIP was developed for the study of histone modifications in mouse haematopoietic stem cells and progenitor cells with 50,000 cells by qPCR (Attema et al., 2007). The Q^2ChIP protocol has been demonstrated

to detect histone modifications in as few as 100 cells by qPCR (Dahl and Collas, 2007). MicroChIP has been recently developed with 10,000 cells for the study of RNA Pol II and histone H3 modifications in combination with genome-scale microarrays (Acevedo et al., 2007).

- **ChIP DNA yield**

The amount of DNA recovered after the ChIP experiment is usually ten to a few hundred nanograms (based on experience in the Vetrie laboratory). For highly sensitive applications such as qPCR and ChIP-seq, only nanograms of ChIP DNA are required for analyses. However, for hybridisation onto genome-scale microarray, micrograms of DNA are usually needed. Therefore, an amplification of ChIP DNA is often required in most cases to generate enough starting material for hybridisation. Various amplification protocols have been developed and used in ChIP-on-chip studies. These include the ligation-PCR method where a double-stranded linker is ligated to the end of the DNA fragment for PCR amplification (Ren et al., 2000), the random-priming method where random primers are annealed to the DNA for PCR amplification (Iyer et al., 2001), and the T7-based linear amplification where poly dTs are added to the ends of DNA fragments and polyA dT primers are used for PCR (Bernstein et al., 2005). However, all these methods of PCR amplification may introduce biases for certain sequences or fragment lengths which will affect subsequent analyses on microarrays. Unamplified ChIP DNA has also been successfully used in microarray analyses on the ENCODE tiling arrays to study histone modifications (Koch et al., 2007).

- **Array platform and data analysis**

Depending on the type of analysis that is required, different array platforms can be employed for the downstream analysis of ChIP DNA. These include tiling arrays, promoter arrays, CpG island arrays and whole-genome arrays (Chapter 1, section 1.3.3.3 A). A few parameters should be considered when choosing the appropriate array platforms. These include:

(i) genome coverage of array,

(ii) resolution of array elements,

(iii) density and duplicates of array elements, and

(iv) reproducibility of genomic enrichments.

The analysis of ChIP-on-chip datasets obtained from the microarray is critical for identifying significant protein-bound DNA elements. Similar to expression microarray analysis, normalisation is required as the initial step of data analysis for ChIP-on-chip to account for signal-dependent issues, variation between replicates and scanning conditions. In addition, normalisation with arrays

hybridised with samples generated using IgGs as the antibodies should also be considered to eliminate any non-specific binding by the corresponding IgGs (Pawan Dhami, PhD thesis).

5.1.2 Human transcription factor promoter array platform

The array platform used in the ChIP-on-chip studies described in this Chapter was an in-house transcription factor promoter array. This array contains duplicates of array elements of two main components: the SCL tiling path (Pawan Dhami, PhD thesis) and the promoters of the majority of human transcription factors. These will be discussed in more detail in the following sections. The array was generated using a single-stranded technology developed at the Sanger Institute (Dhami et al., 2005). In this system, single-stranded DNA fragments derived from double-stranded PCR products are immobilised on the surface of the array. During the PCR amplification, primers with a 5'-aminolink modification were used to amplify the sequence from genomic DNA resulting in the generation of double-stranded PCR products containing the modification on one strand only. The double-stranded PCR products are spotted onto the array surface and covalent interactions between the aminolink modification and the array surface occurs. The unmodified strand is then removed by chemical or physical denaturation leaving only the modified single-strand attached to the array surface. This single-stranded array system has a high sensitivity as the resultant single-stranded DNAs cannot reanneal making them effective targets for hybridisation with the labelled samples (Figure 5.2). It has been shown that this array system generates a higher signal:noise ratios than double-stranded PCR product arrays (Dhami et al., 2005).

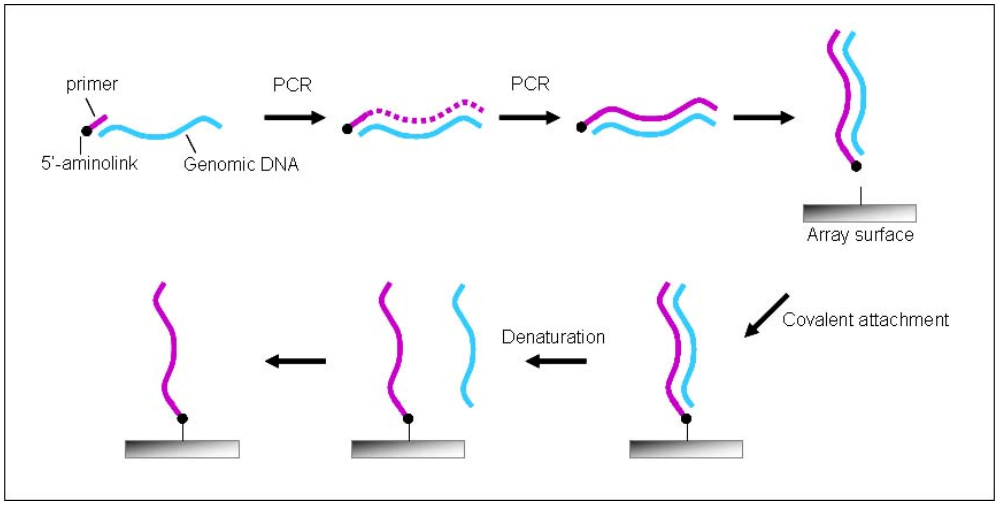


Figure 5.2. The single-stranded array platform. Schematic diagram showed the generation of arrays with the single-stranded array platform. Double-stranded PCR products are generated with a 5'-aminolink primer. 5'aminolink modified strands (purple strands) are attached to the array surface by covalent interaction while the unmodified strands (blue strands) are denatured. Please see text for detailed description. Figure was modified from Dhami et al. 2005 with permission.

- **The SCL tiling path**

GATA1, SCL and LDB1 were shown to bind to the +51 enhancer of SCL (Pawan Dhami, PhD thesis) which is equivalent to the previously described +40 enhancer in mouse (Delabesse et al., 2005) (Chapter 1, section 1.4.2.1 E). As a positive control for selection of ChIP-working antibodies and quality control of the ChIP-on-chip experiment, an SCL tiling path was included on the transcription factor promoter array (see next section). The SCL tiling path was generated by Dr. Pawan Dhami (Pawan Dhami, PhD thesis) which spans approximately 256 kb across the human SCL locus at a resolution of 400 bp. It includes two genes upstream of SCL (KCY and SIL) and three genes downstream (CYP4Z1, CYP4A22 and MAP17) (Figure 5.3 A). Using antibodies against GATA1, SCL and LDB1, significant enrichments were observed by ChIP-on-chip in a novel regulatory region designated as the +51 region (Figure 5.3 B) (Pawan Dhami, PhD thesis). The DNA sequence of this +51 region has hallmarks of the recognition sequence of the SCL erythroid complex originally identified by Wadman et al. (1997) where the E-box and GATA motifs were separated by 9 nucleotides. Therefore, the other members of the SCL erythroid complex may bind to the +51 enhancer.

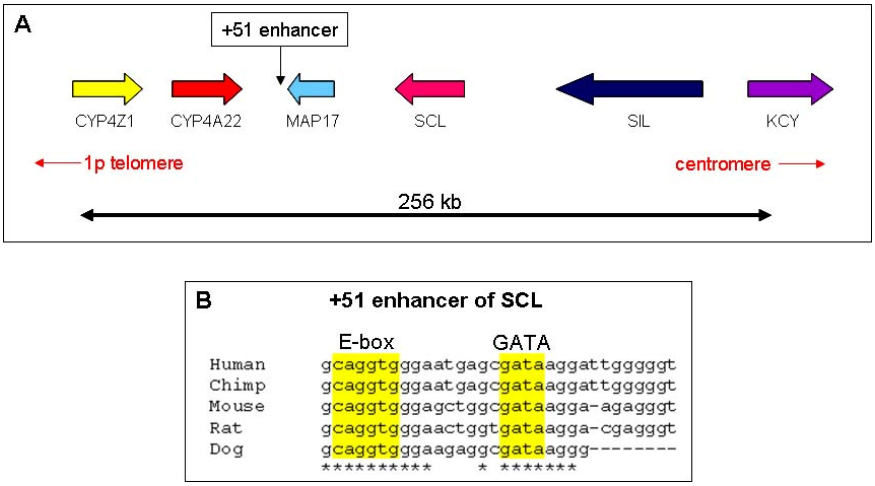


Figure 5.3. The SCL tiling path and the +51 enhancer of SCL. Panel A: schematic diagram showing the genomic region of the SCL locus included on the SCL tiling path array. The black two-way arrow shows the 256 kb region included in the array. The thick coloured arrows represent the genes. The red arrows show the orientation of the locus. The small black arrow shows the position of the +51 enhancer of SCL. Panel B: multiple sequence alignment of the +51 enhancer of SCL. Nucleotides shaded in yellow show the conserved E-box and GATA motifs. Asterisks at the bottom showed the conserved nucleotides across 5 species.

- **The transcription factor promoter array**

As it was not possible to study the entire human genome by ChIP-on-chip when this project was initiated, a sub-set of genomic sequences were studied by ChIP-on-chip. Given that transcription

factors are the key regulators of transcriptional cascades, the focus of the ChIP-on-chip studies for this project was based on the use of an in-house transcription factor promoter array. This array contains approximately 1600 promoters of human transcription factors as well as promoters of a selected handful of haematopoietic genes known to be targets of members of SCL erythroid complex (for example, EPOR, which is a target of GATA1). Gene list for transcription factors was defined by Philippe Couttet and David Vetrie (Sanger Institute) using lists of all known human transcription factors downloaded from ENSEMBL (including transcription factors and chromatin modifiers/remodelers). The haematopoietic gene EPOR was included on the array as a positive control for ChIP as GATA1 was shown to bind to the EPOR promoter (Zon et al., 1991). To generate this array, the locations of promoters were first determined using the *in silico* promoter prediction algorithm FirstEF. FirstEF is a software which identifies CpG islands, promoter regions and first exon splice-donor sites in the genome with high accuracy and low false-positive rate (Davuluri et al., 2001). Using FirstEF, predicted promoters of both human and mouse transcription factor genes were shown to be closely clustered within 1 kb around known transcription start sites (Figure 5.4). Therefore; a 1 kb region around the TSS was selected for each transcription factor gene. 1 kb regions for each transcription factor promoter were PCR amplified and included on the array.

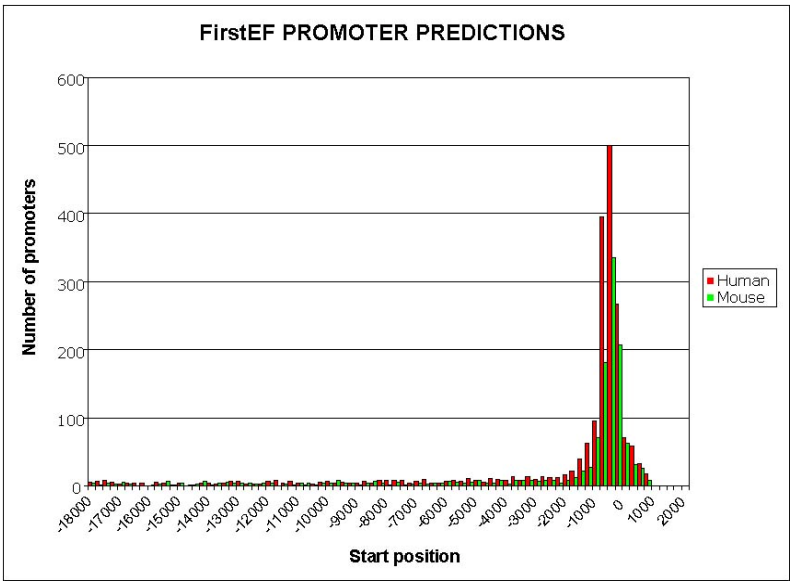


Figure 5.4. FirstEF prediction of human and mouse transcription factor promoters. The histogram shows the number of promoters for transcription factors predicted at various positions relative to the transcription start site. X-axis: start position in bps; y-axis: number of promoters; red bars: human promoters; green bars: mouse promoters. Start position 0 indicates the transcription start site while positive values indicate sequences downstream of the TSS and negative values indicate sequences upstream of the TSS. FirstEF analysis was performed by Dr. Robert Andrews (Wellcome Trust Sanger Institute).

5.1.3 ChIP studies of transcription factors in the SCL erythroid complex in the literature

A number of studies have been performed by previous researchers to identify gene targets bound by members of the SCL erythroid complex. ChIPs in combination with PCR, qPCR or microarray were used to identify and characterise the direct binding of the transcription factors to the promoters or enhancers of their target genes (Table 5.1). In particular, ChIP in combination with a human promoter array was used to identify 71 promoters showing significant binding of SCL in the human T-ALL Jurkat cells (Palomero et al., 2006). ChIP-qPCR of E2A demonstrated the association of E2A with approximately 60% of the SCL target genes in this study. A tiling array across 130 kb of the mouse α-globin locus was used to map GATA1 binding regions at various stages of haematopoietic development in mouse and to study the recruitment of interacting partners using ChIP (Anguita et al., 2004). ChIP-on-chip analysis was used to map GATA1 binding sites in the human β-globin locus in K562 cells identified both known and novel binding regions (Horak et al., 2002). However, a thorough study of the five members of the SCL erythroid complex using ChIP-on-chip in erythroid cells is lacking.

Transcription factor studied	Target gene	Technique used	Organism	Cell type	References
SCL	GYPA promoter	ChIP-PCR	Human	Haematopoietic cell line (TF1)	(Lahlil et al., 2004)
GATA1, SCL and LDB1	P4.2 promoter	ChIP-PCR	Mouse	Erythroid cell line (MEL)	(Xu et al., 2003)
SCL and E2A	c-kit promoter	ChIP-PCR	Human	Haematopoietic cell line (TF1)	(Lecuyer et al., 2002)
SCL, GATA1 and LMO2	β-globin locus control region	ChIP-qPCR	Human	Erythroid progenitor cell line (K562)	(Song et al., 2007)
SCL	71 human genes	ChIP + promoter array	Human	T-ALL cell line (Jurkat)	(Palomero et al., 2006)
GATA1	GFI1B promoter	ChIP-PCR	Human	Erythroid progenitor cell line (K562)	(Huang et al., 2004)
GATA1	HS2 region of the β-globin locus	ChIP-PCR	Mouse	Erythroid cell line (MEL)	(Johnson et al., 2002)
GATA1	MYC promoter	ChIP-PCR	Mouse	GATA1-null erythroblast cell line (G1E-ER4)	(Rylski et al., 2003)
GATA1	FOG-1 enhancer	ChIP-qPCR	Mouse	GATA1-null erythroblast cell line (G1E-ER4)	(Welch et al., 2004)
GATA1	α-globin locus	ChIP + tiling array	Mouse	Erythroid cell line (MEL)	(Anguita et al., 2004)
GATA1	β-globin locus	ChIP + tiling array	Human	Erythroid progenitor cell line (K562)	(Horak et al., 2002)

Table 5.1. ChIP studies of various members of the SCL erythroid complex. The target gene, technique, organism and cell type used in the ChIP studies are listed in the table.

5.2 Aims of this chapter

The aims of work presented in this Chapter were:

1. To test and validate antibodies targeting five members of the SCL erythroid complex for ChIP-on-chip applications.
2. To identify putative promoters bound by each member of the SCL erythroid complex in K562 cells by ChIP in combination with the transcription factor promoter array.
3. To investigate the transcription factor binding sites (TFBS) in the putative promoters and perform comparative genomic sequence analyses of these TFBSs.
4. To validate the putative target genes by ChIP-qPCR in K562 and HEL erythroid cell lines.

5.3 Overall strategy

The overall aim of the work described in this Chapter was to confirm and identify direct target genes regulated by each of five members (GATA1, SCL, E2A, LMO2 and LDB1) of the SCL erythroid complex using ChIP-on-chip. Working ChIP assays for each transcription factor were validated in ChIP-on-chip in K562 cells using the SCL tiling path/transcription factor promoter array. As mentioned in section 5.1.2, GATA1, SCL and LDB1 were shown to bind to the +51 enhancer (Pawan Dhami, PhD thesis) and antibodies against these transcription factors were previously characterised for ChIP assays by Dr. Pawan Dhami. Since the +51 enhancer contains the consensus E-box and GATA1 motifs separated by 9 nucleotides as first described for the SCL erythroid complex by Wadman et al (1997), the other members of the SCL erythroid complex (E2A and LMO2) may also bind to this enhancer. Based on enrichments obtained for the +51 enhancer and promoters at the SCL locus, the best performing antibodies were chosen as the working ChIP assays for these two transcription factors. Three biological replicates of ChIP-on-chip for each of the five transcription factors and their corresponding IgG controls were performed using the SCL tiling path/transcription factor promoter array. Although it has been shown that dye-specific bias is a source of error in 2-colour array experiments and cannot be removed during normalisation (Dobbin et al, 2005), dye-swap experiments were not performed as this will double the cost required for the experiments. The quality of each ChIP-on-chip assay was assessed at various steps during the experiments (Section 5.4.1). Enrichments of each promoter in the ChIP-on-chip study of the transcription factors were normalised with their enrichments in the corresponding IgG controls. Statistical analyses of enriched promoters were carried out for the ChIP-on-chip experiments which passed quality control. Cross-comparison between the putative targets identified by each of the five members of the SCL erythroid complex was performed to identify targets bound by all five

members of the complex. To identify the DNA sequence motifs which were likely to bind the five transcription factors, sequences in the enriched promoters were analysed by TESS and TFSearch together with comparative genomic sequence analyses. Confirmation of promoter binding events was addressed by ChIP in combination with quantitative PCRs. The overall strategy for this ChIP-on-chip study is summarised in Figure 5.5.

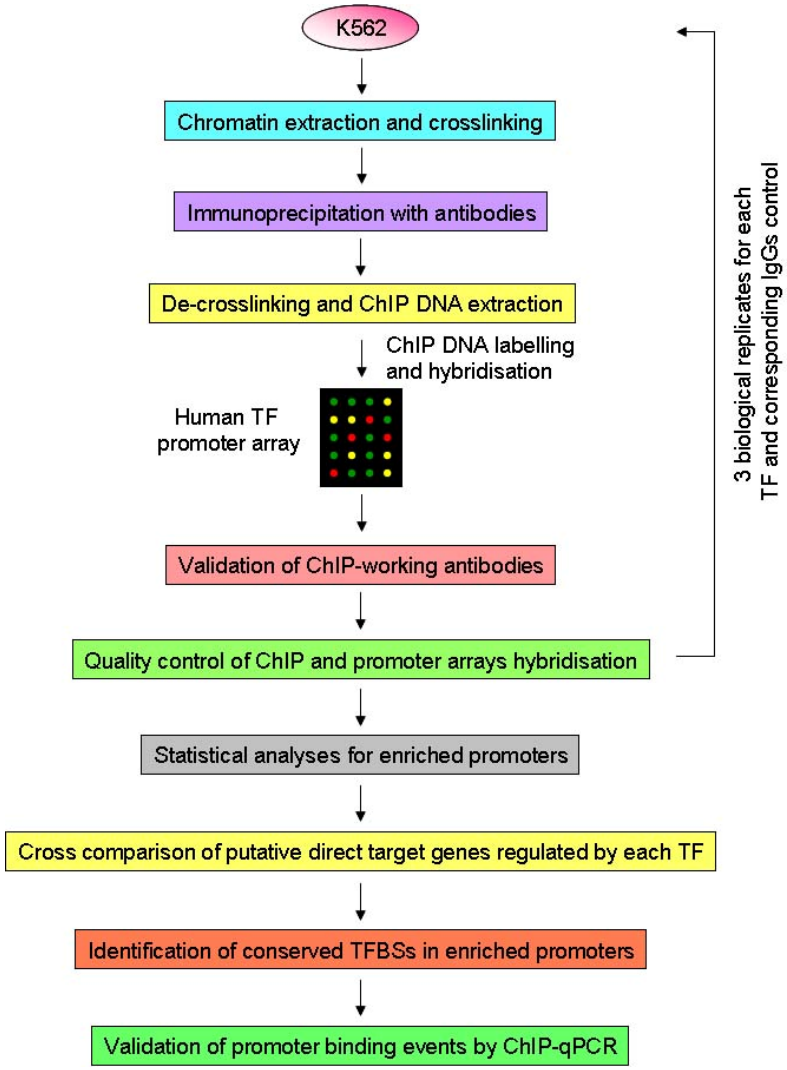


Figure 5.5. Overall strategy for the ChIP-on-chip analyses of the SCL erythroid complex. ChIP assays in K562 were performed as described in Chapter 2. Chromatin from K562 cells was extracted and sonicated while DNA bound by the transcription factor under study was immunoprecipitated by specific antibodies followed by de-crosslinking and extraction. Working ChIP assays for each transcription factor were validated based on enrichments obtained for the +51 enhancer at the SCL locus. Quality control of various steps of the ChIP-on-chip assays was performed. Three biological replicates of ChIP-on-chip for each of the five transcription factors and their corresponding IgG controls were performed. Normalisation of enrichments was done against the IgG controls. Statistical analyses of enriched promoters were carried out for the ChIP-on-chip experiments. Cross-comparison between the enriched promoters identified by

each of the five members of the SCL erythroid complex was performed. Sequences in the enriched promoters were analysed by TESS and TFSearch together with comparative genomic sequence analyses to identify conserved transcription factor binding sites. Confirmation of promoter binding events was addressed by ChIP in combination with quantitative PCRs.

5.4 Results

5.4.1 Quality control of various steps of chromatin-immunoprecipitation

To ensure that experiments done at different times for the various biological replicates were consistent, a variety of steps were analysed throughout the ChIP-on-chip procedure as described below.

5.4.1.1 Culturing of cells

Cell lines (K562 and HEL) were cultured for no more than a week at concentrations of 0.5 million cells per millilitre before chromatin extractions were performed. Fresh media were added one day before extraction. To further reduce the variability across replicates, the same passage of cells was defrosted for biological replicates performed at different times. From the cultured cells, aliquots of cells were taken for flow analysis prior to chromatin extraction. The proportion of actively dividing cells was monitored by flow analysis to determine the DNA content of the cells (Figure 5.6) as a measure of the number of actively dividing cells (actively dividing cells in S or G2/M phases of the cell cycle have higher DNA contents due to DNA replication). Only cell populations with similar growth characteristics were used for subsequent analyses. For example, for all of the experiments performed for K562, approximately 60-70% of cells were actively dividing in all three bioreplicates.

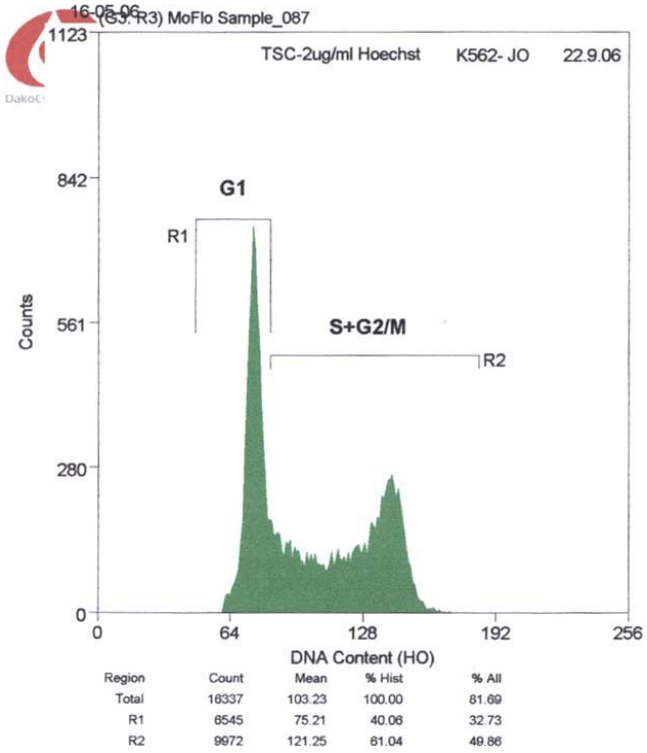


Figure 5.6. Flow analysis of growth pattern of cell lines. K562 or HEL cells used in the ChIP-on-chip experiment described in this Chapter were subjected to flow-analysis by staining with the DNA binding dye Hoechst 33342 to determine the DNA content of each cell. The percentages of cells in the G1 (labelled as R1) and S and G2 or M (labelled as R2) phases of the cell cycle were determined by measuring the fluorescence intensity (shown at the bottom of the image). Actively dividing cells in S or G2/M phases have higher DNA contents due to DNA replication and this could be used as a measure of the proportion of cells in the population which were actively dividing (this study was performed by Bee Ling Ng, Wellcome Trust Sanger Institute).

5.4.1.2 Preparation of cross-linked chromatin

The initial step of ChIP-on-chip is the cross-linking of protein and DNA in the chromatin. The cross-linking condition used here was 1% formaldehyde for 10 minutes (this was based on titration experiments performed by Dr. Pawan Dhami in the laboratory). The resultant protein-DNA complexes were sonicated to shear the DNA into fragments with a size distribution in the range of of 600-3000 bp (average size around 1000 bp). To ensure that the cross-linking and sonication consistently resulted in DNA fragments of the correct size distribution; a small aliquot of the cross-linked and sonicated material was analysed by agarose gel electrophoresis (Figure 5.7). A smear was observed with an average size distribution of approximately 1000-1500 bp. Purified DNA from this crude chromatin extract was subsequently shown to give a size distribution with an average DNA fragment size of approximately 1000 bp (See Figure 5.8).

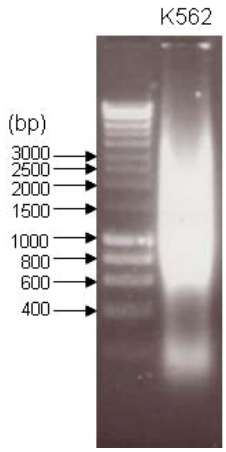


Figure 5.7. Agarose gel electrophoresis of cross-linked and sonicated chromatin. Chromatin extracted from K562 after cross-linking and sonication was analysed by electrophoresis of a 1% agarose gel made with 1 X TBE and visualised by ethidium bromide staining. A 1 kb ladder was loaded in the left lane and 5 µl of the K562 chromatin cross-linked in 1% formaldehyde for 10 minutes is shown in the lane to the right of the size markers.

5.4.1.3 Extraction of ChIP DNA

Similarly, agarose gel electrophoresis was used to examine the size distributions and recoveries of input and ChIP DNAs (Figure 5.8). Input DNA is the material extracted after de-crosslinking of the chromatin which did not undergo any immunoprecipitation. ChIP DNA, in contrast, is the DNA extracted after immunoprecipitation with specific antibodies. On agarose gels, input DNA normally showed a visible DNA smear of similar size distribution to the crude chromatin (Figure 5.7). ChIP DNAs, in contrast, were difficult to visualise on agarose gels because of the amount of material recovered from ChIP assays. Thus, relative amounts of DNA recovered in ChIP samples were monitored by comparing the intensity of the yeast tRNA which was co-precipitated in the ChIP samples.

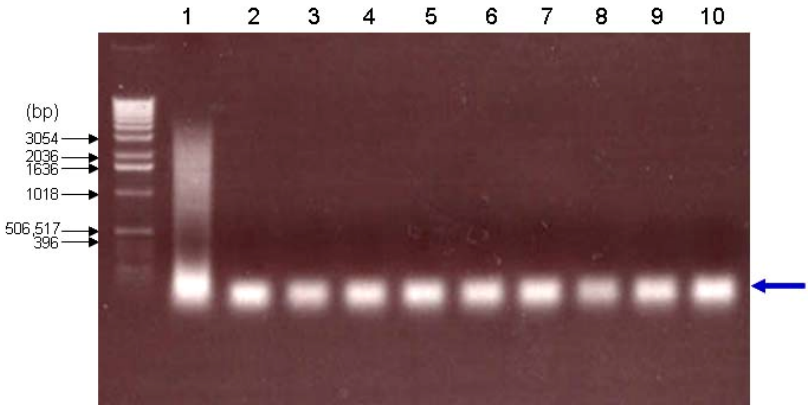


Figure 5.8. Agarose gel electrophoresis of input and ChIP DNA. 5 µl of input and ChIP DNAs using antibodies for transcription factor and IgGs were extracted and electrophoresed on a 1% agarose gel in 1 X TBE and visualised by ethidium bromide staining. A 1 kb ladder was loaded in the lane on the left of the image. Lane 1: input DNA; lane 2: ChIP DNA of LMO2 G16 antibody; lane 3: ChIP DNA of LMO2 N16 antibody; lane 4: ChIP DNA of LMO2 Abcam antibody; lane 5: ChIP DNA of LDB1 N18 antibody; lane 6: ChIP DNA of SCL serum; lane 7: ChIP DNA of E12 H208 antibody; lane 8: ChIP DNA of goat IgG; lane 9: ChIP DNA of mouse IgG; lane 10: ChIP DNA of rabbit IgG. The

input DNA in lane 1 shows a smear of the appropriate size distribution while only yeast tRNA was observed in the ChIP DNA samples at the bottom of the gel (shown by the blue arrow).

5.4.1.4 Labelling of input and ChIP DNA

Input and ChIP DNAs were labelled with cyanine dyes (Cy3 and Cy5) for array hybridisations. The DNA labelling process was performed by random priming with Klenow fragments lacking the 3' to 5' and 5' to 3' exonuclease activity (Lieu et al., 2005). Due to the intrinsic strand displacement activity of Klenow, the labelled fragments were exponentially amplified (Walker, 1993) (Figure 5.9).

The labelled input and ChIP DNAs were analysed by agarose electrophoresis to evaluate the labelling and amplification efficiency (Figure 5.10). In both input DNA and ChIP DNA, smears were observed across a broad size range, with the majority of the labelled fragments being less than 200 bp in size. Compared with the original unlabelled ChIPs DNA (Figure 5.8) where $1/10^{th}$ the original material was loaded onto agarose gels, more obvious smears were observed after labelling (1/30 of the labelled material) indicating large quantities of DNA were amplified during labelling.

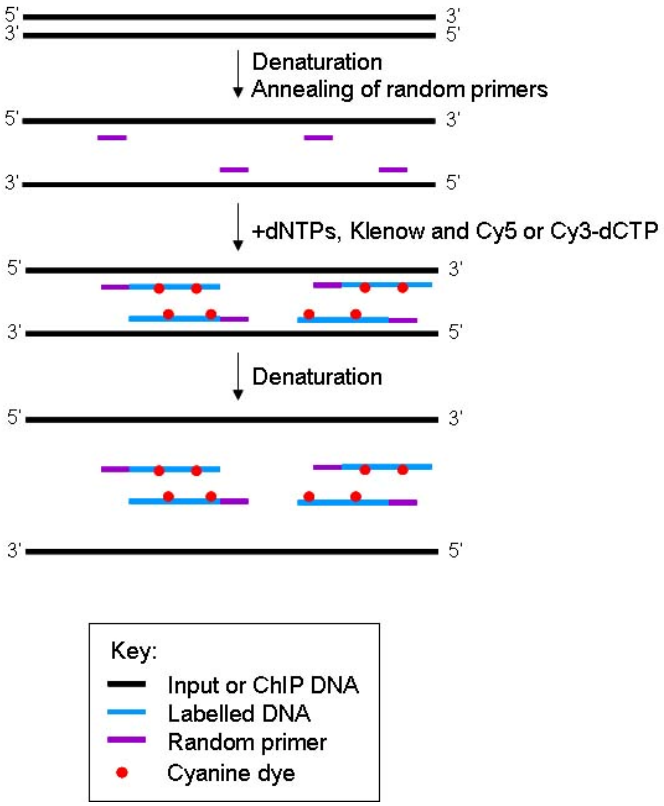


Figure 5.9. Random priming and cyanine labelling of input and ChIP DNA. Input and ChIP DNAs were labelled with a random priming method involving the use of the Klenow fragment with a strand displacement activity. The DNA being labelled was first denatured and primers were annealed. The resulted DNA were amplified with Klenow enzyme and labelled with cyanine dyes.

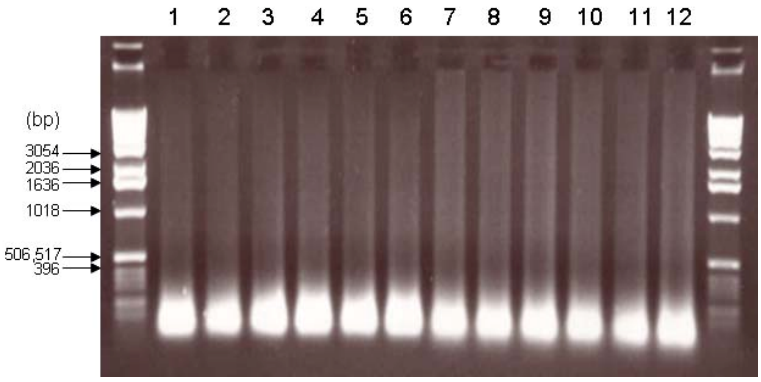


Figure 5.10. Agarose gel electrophoresis of labelled input and ChIP DNA. 5 µl of input and ChIP DNA samples from K562 were labelled by the random priming method and electrophoresed on a 1% agarose gel made with 1 X TBE and visualised by ethidum bromide staining. 1 kb ladder is loaded in the first and last lane. Lanes 1-6: Input samples; lane 7: labelled ChIP DNA of rabbit IgG; lane 8: labelled ChIP DNA of goat IgG; lane 9: labelled ChIP DNA of LMO2 N16 antibody; lane 10: labelled ChIP DNA of LDB1 N18 antibody; lane 11: labelled ChIP DNA of E47 N649 antibody; lane 12: labelled ChIP DNA of E12 H208 antibody.

5.4.1.5 Hybridisation and analyses of the transcription factor promoter array

After hybridisation and scanning, the resultant array images were quality-controlled. Initially they were assessed by eye to identify any visible problems with the array hybridisation which may affect the quantification of spots (high background and various hybridisation artifacts). Array which showed such problems were discarded and the hybridisations were repeated. Given that the array elements were spotted in duplicate, the coefficients of variation (CVs) for the duplicated elements were calculated for each spot to determine reproducibility of datapoints within a single hybridisation. Typically, the median CV (median of all CVs obtained from duplicate array elements) for a given hybridisation was approximately 5% and arrays which deviated substantially from this median value were not included in further analyses (a CV of 6% was used as a cut-off). An example of good quality hybridisation is shown in Figure 5.11.

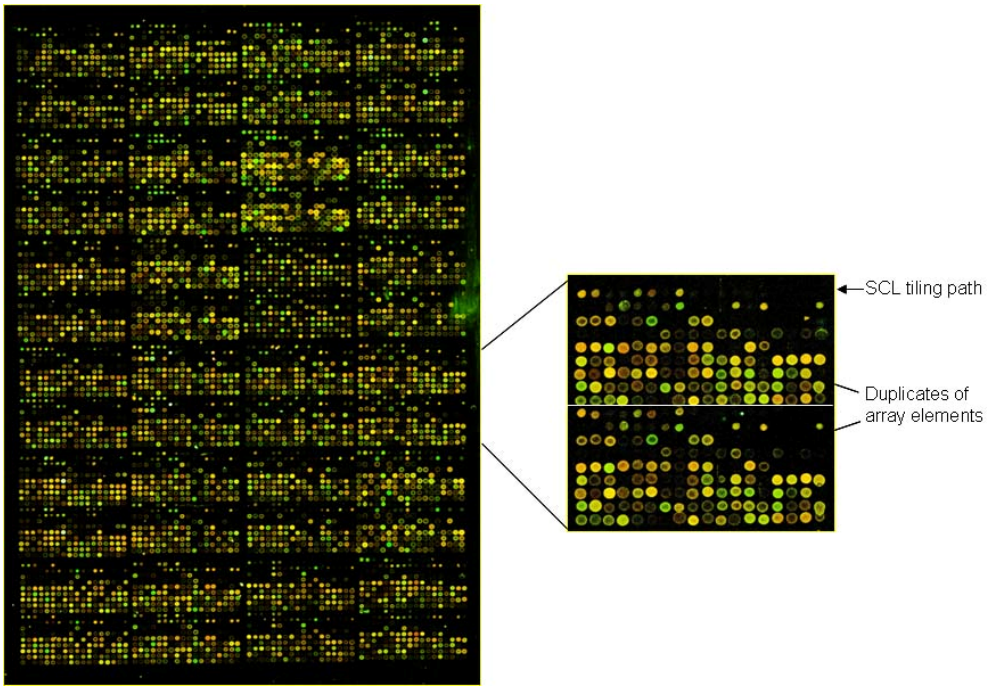


Figure 5.11. A composite image of the human transcription factor promoter array. The promoter array was hybridised with (i) ChIP DNA derived from ChIP with the GATA1 M20 antibody in K562 cells and (ii) the input DNA of K562 cells. The array contains 24 sub-arrays and 4132 spots where each spot represents an array element for either a human transcription factor promoter or a tile of the SCL tiling path. The zoomed-in image on the right illustrates one of the sub-arrays containing the spots for the SCL tiling path and the duplicates of each array element. Green spots represent array elements enriched in the ChIP sample. Red spots represent array elements under-represented in the ChIP sample. Yellow spots represented array elements equally represented in the ChIP and input samples. White spots showed array elements with saturated pixel values in the image for the ChIP sample.

5.4.2 Evaluation of working antibodies by positive control elements of the array

Three criteria were used for the selection of high quality antibodies for use in ChIP-on-chip assays for the five transcription factors of the SCL erythroid complex:

(i) they must show significant enrichments at the +51 enhancer of the SCL locus. The promoter array contained the SCL tiling path (section 5.1.2) which acted as the positive control region for testing the antibodies against the 5 members of the SCL erythroid complex. Each member was expected to bind the +51 enhancer of SCL which contains the consensus E-box/GATA motif and had been shown to bind GATA1, SCL and LDB1 (Pawan Dhami, PhD thesis). This +51 is the equivalent to the +40 enhancer of SCL in mouse (Ogilvy et al., 2007).

(ii) the background in the negative regions must be low. As previously demonstrated by the ChIP-on-chip data of GATA1, SCL and LDB1 on the SCL tiling array, many regions on the locus show

enrichments at or near a value of 1 (baseline). These regions are regarded as the negative regions for assessment of non-specific binding.

(iii) they must be specific as detected by western blotting as described in Chapter 3.

Appendix 3B summarised the characteristics of, and the results obtained for, a variety of antibodies tested in the ChIP-on-chip experiment. Six of these antibodies were used for further ChIP-on-chip experiments and the results of these across the SCL tile path are described below.

(i) GATA1: 15- to 30-fold enrichments for the +51 SCL enhancer were observed for the GATA1 M20 ChIP assay (Figure 5.12 A), replicating results obtained previously with this assay (Pawan Dhami, PhD thesis). Significant enrichments of up to 5-fold were also observed for SCL promoter 1a and +3 and -9/-10 enhancers (Pawan Dhami, PhD thesis). This antibody was also shown to be highly specific on western blotting (Chapter 3, section 3.4.2).

(ii) SCL: Both antibodies tested for SCL showed substantial enrichments at the +51 enhancer where 8- to 12-fold enrichments were observed for the TAL1 Active Motif antibody (Figure 5.12 B) and 20-fold enrichments were shown for the SCL serum. Although the SCL serum showed higher enrichment in the +51 enhancer, the quantity of this antibody was limiting (as it was a gift from a collaborator) and it did not yield the appropriate bands for the SCL protein in western analysis (Chapter 3, section 3.4.2). In contrast, the TAL1 Active Motif antibody was shown to be specific for SCL in western analysis and was therefore used in subsequent ChIP-on-chip analyses.

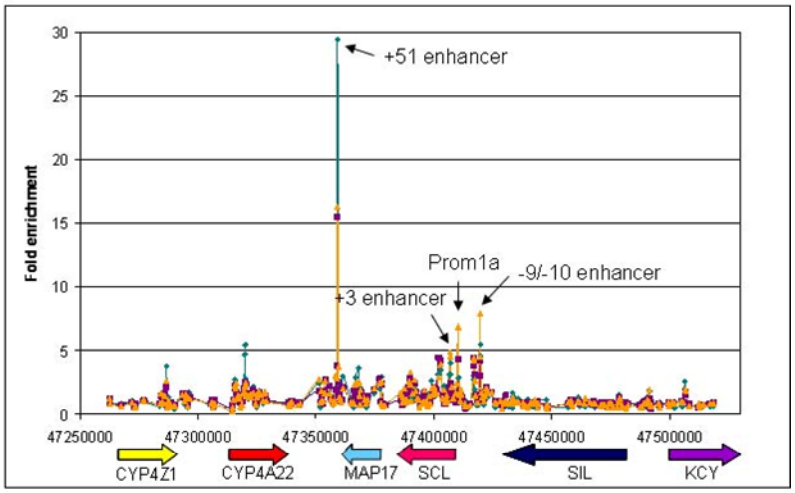
(iii) E2A: Two antibodies were tested for E2A which recognised both E12 and E47 isoforms. The TCF3 antibody from Abcam showed no substantial enrichments in any of the SCL enhancers or promoters while the E2A antibody from BD Biosciences showed an approximately 12-fold enrichment in the +51 region. However, the E2A antibody from BD Bioscience could not identify specific bands for E2A in western analysis (Chapter 3, section 3.4.2). Specific antibodies for the E12 and E47 isoforms were also characterised. The E12 H208 and E47 N649 antibodies both showed up to 60-fold enrichments in the +51 enhancer and enrichments of approximately 8-fold in the +3 and -9/-10 enhancers and the promoter 1a (Figure 5.12 C and D). These two antibodies were also shown to be specific for E12 and E47 in western analysis (Chapter 3, section 3.4.2) and were used for further ChIP-on-chip analyses. However, no information is known about the cross-reactivity of these two antibodies with the other isoform.

(iv) LDB1: Up to 45-fold enrichments for the +51 enhancer were observed for the one antibody tested for LDB1 (Figure 5.12 E). In addition, substantial enrichments of more than 10-fold were also observed for promoter 1a and -9/-10 enhancers and 5-fold enrichments were shown for the +3

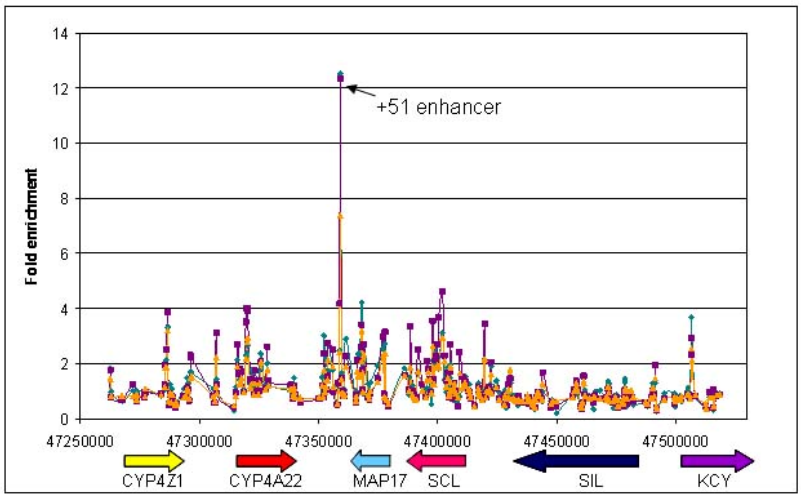
enhancer. This antibody was also shown to be highly specific on western blotting (Chapter 3, section 3.4.2). Therefore, this antibody was used for further ChIP-on-chip analyses.

(v) LMO2: Both antibodies tested for LMO2 did not show high enrichments across the SCL locus and generated a lot non-specific noise (Figure 5.12 F). However, the N16 antibody was slightly better than G16 in terms of the enrichments at the +51 enhancer i.e. up to 10-fold for LMO2 N16 versus 7-fold for LMO2 G16. LMO2 N16 was therefore used in subsequent ChIP-on-chip analyses despite there being no western data to support its specificity.

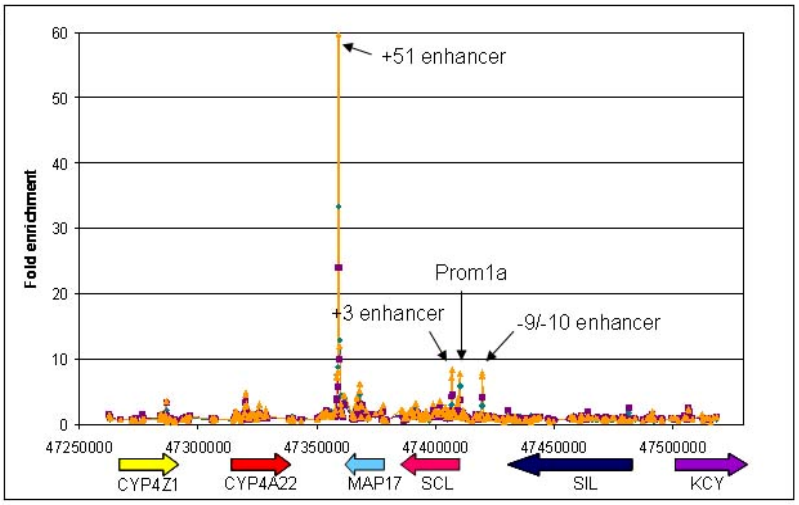
A) GATA1 M20



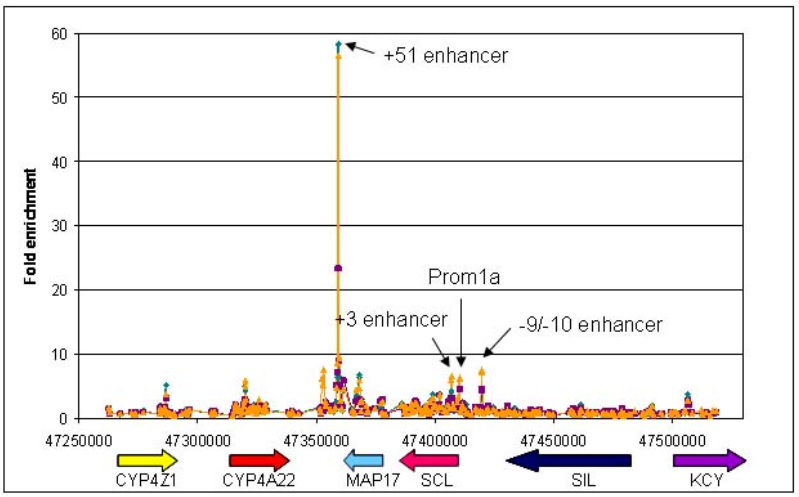
B) TAL1 Active Motif



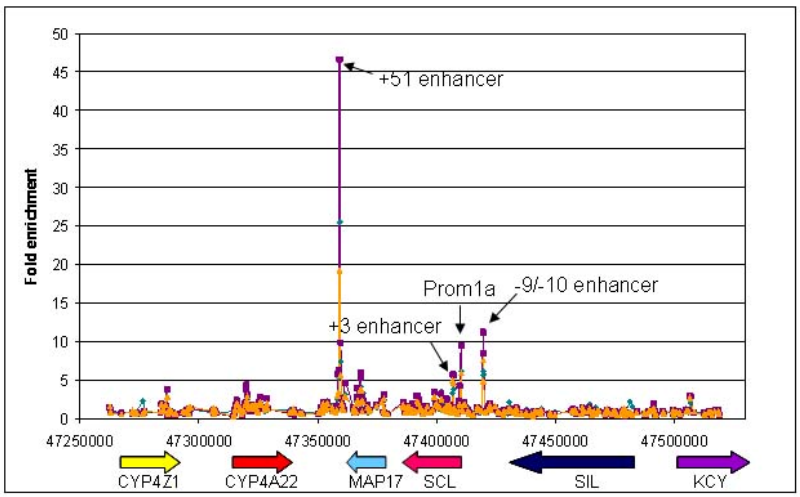
C) E12 H208



D) E47 N649



E) LDB1 N18



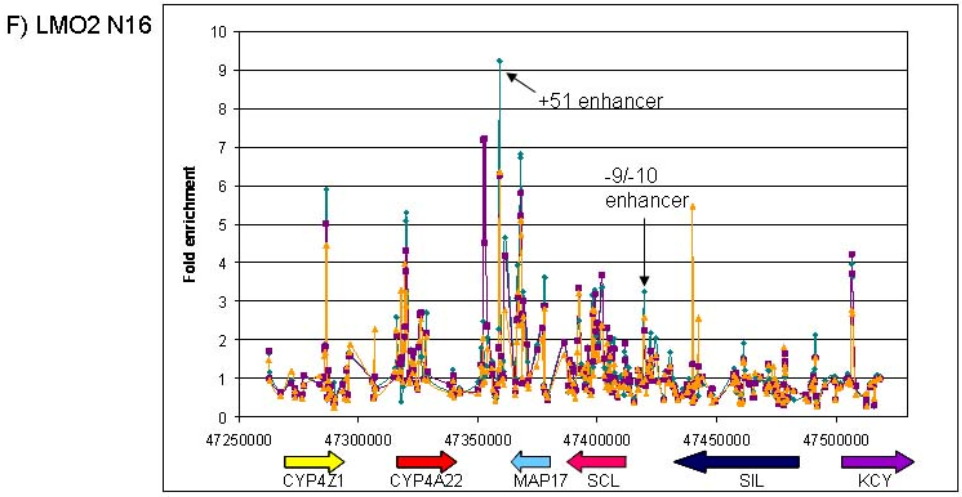


Figure 5.12. ChIP-on-chip profiles of selected working antibodies for the SCL erythroid complex across the SCL locus in K562 cells. The ChIP-on-chip profiles across the SCL locus of antibodies selected for subsequent analyses were shown. Panel A: SCL locus profile of GATA1 M20 antibody; panel B: SCL locus profile of TAL1 Active Motif antibody; panel C: SCL locus profile of E12 H208 antibody; panel D: SCL locus profile of E47 N649 antibody; panel E: SCL locus profile of LDB1 N18 antibody; panel F: SCL locus profile of LMO2 N16 antibody. The x-axis represented the genomic coordinates across the SCL tiling path and the y-axis represented the fold enrichments. The thick coloured arrows showed the position of the genes included on the SCL tiling path. Light blue curve: biological replicate 1; violet curve: biological replicate 2 and orange curve: biological replicate 3. SCL enhancers or promoters which showed enrichments were labelled by black arrows on the graph.

5.4.3 Data analyses of enriched promoters

Having validated the performance of antibodies for each of the five transcription factors in ChIP-on-chip, three bioreplicates for each of the chosen assays were performed on the SCL tiling path/transcription factor promoter array. Two technical replicates were also performed for each biological replicate. Similarly, the host IgG control ChIP-on-chip experiments were performed for each transcription factor assay across three bioreplicates and two technical replicates. The quality of each ChIP and hybridisation was monitored as described in section 5.4.1 and the array hybridisations that passed the quality control criteria were subject to statistical analyses for the selection of enriched promoters which are likely to be bound by the transcription factors under study.

5.4.3.1 Overall strategy of statistical analyses

Figure 5.13 outlines the procedures used for statistical analyses of enriched promoters in the ChIP-on-chip experiment. Signal intensities of the array elements for all the scanned array hybridisations were first quantitated in Scanarray Express. Ratios of Cy3 (ChIP sample) against Cy5 (Input sample) were also generated in Scanarray Express. The ratios for the duplicated array elements in a

given hybridisation were averaged. Ratios for the two technical replicates were averaged to provide a mean ratio for each bioreplicate. The ratio data was transformed by normalisation, at various stages, in three ways:

(i) signal intensities for both channels in each hybridisation were scaled by total intensity in Scanarray Express,

(ii) each ratio measurement for every array element in a given hybridisation was normalised to the median ratio of all measurements.

(iii) the ratios for all array elements in each experiment (either in each bioreplicate or as the mean of bioreplicates) were normalised against the ratios obtained for the host IgG controls. This normalisation procedure would help account for non-specific enrichments from the host IgGs and effectively remove them from the datasets.

Two methods were used to carry out the statistical analyses of the enriched promoters (Figure 5.13). In method A, each biological replicate was treated separately with respect to the generation of mean enrichments and normalisation between the transcription factor ChIP-on-chip assays and the host IgG ChIP-on-chip assays. Enriched promoter array elements which were two standard deviations above the mean were chosen as the putative target promoters. Two standard deviations were used as a cut-off as it represented a 95.45% confidence level – in other words, the promoters identified were statistically significant in terms of enrichment levels away from background. The promoter lists from each of the three biological replicates were compared in a Venn diagram and promoters found to be significantly enriched in all three bioreplicates were chosen as the putative target promoters of the transcription factor under study. In method B, the average ratio of each promoter was obtained from the 3 biological replicates for the transcription factor ChIP-on-chip assay and normalised with the corresponding average ratio of each promoter from the 3 bioreplicates for the host IgG ChIP-on-chip experiments. Promoters which were enriched 2 standard deviations above the mean were chosen as the putative target genes.

Comparatively speaking, method A was a more stringent approach for selecting promoters which are likely to be bound by the transcription factor. Only promoters which were statistically significant in all three biological replicates were chosen as putative target genes. This requires that the transcription factor-promoter binding is strong and significant to show enrichment in each ChIP-on-chip experiment. Method B, however, is less stringent but it was possible to detect binding events which showed a degree of variability in enrichment across the three bioreplicates.

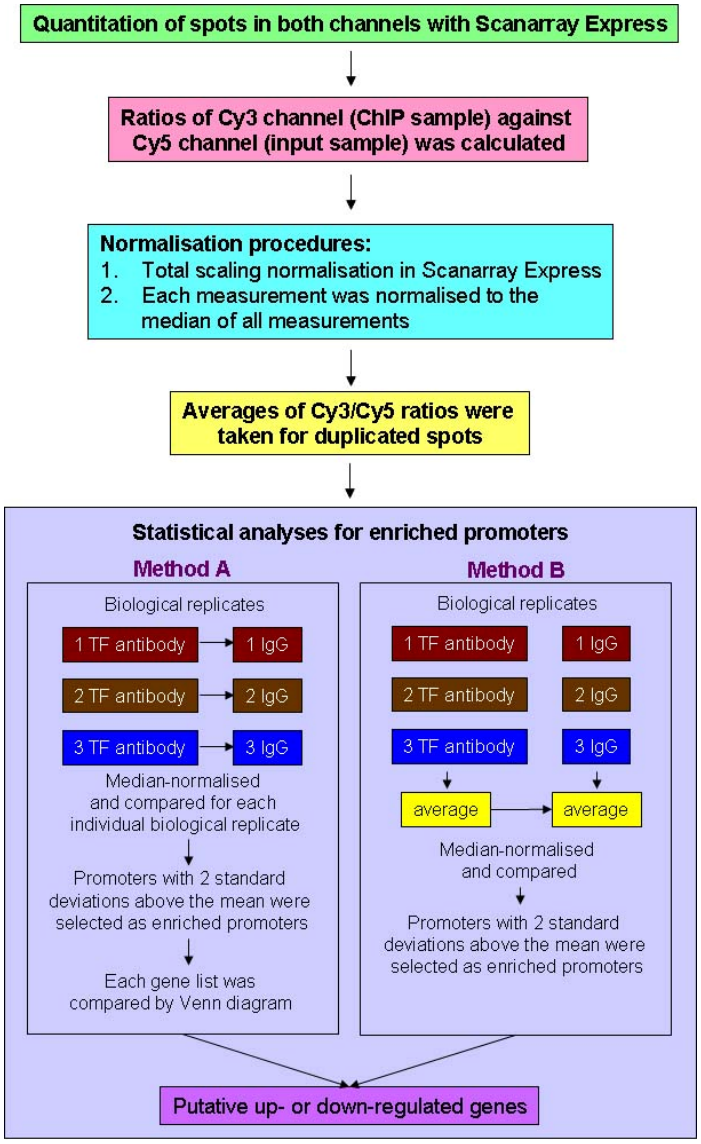


Figure 5.13. Flow diagram of statistical analyses of enriched promoters in ChIP-on-chip. Signal intensities of the array elements were first quantitated in Scanarray Express. Ratios of Cy3 (ChIP sample) against Cy5 (Input sample) were also generated and ratios for the duplicated elements were averaged. Ratios for the two technical replicates were also averaged to provide a mean ratio for each bioreplicate. The ratio data was transformed by normalisation at various levels as described in the text. Statistically significant enriched promoters were identified for each of methods A or B.

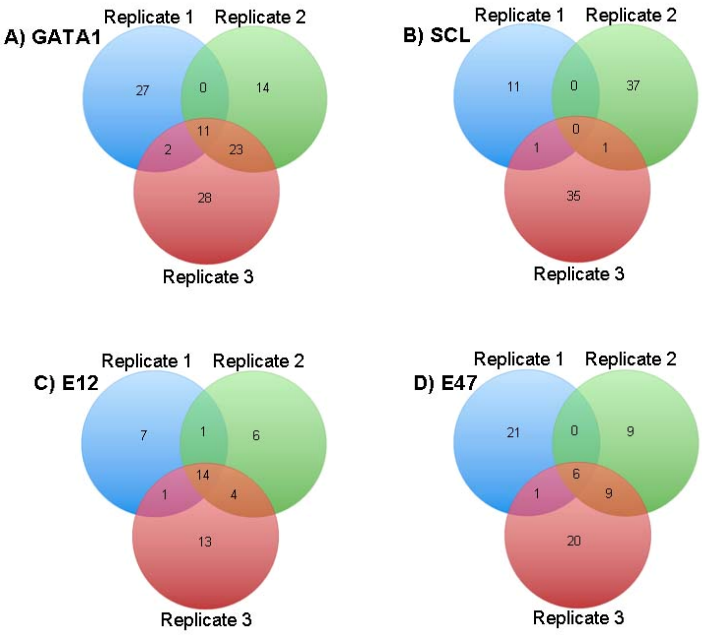
5.4.3.2 Data analyses for the selection of putative target genes

Using the two strategies outlined above, a number of promoters were selected as putative regulatory target genes for each of the transcription factors under study.

Figure 5.14 shows the results for each transcription factor ChIP-on-chip analysis using method A. In ChIP-on-chip analysis for GATA1, E12, E47 and LDB1, between 6 and 14 promoters were

identified in all the three biological replicates. In particular, the known direct target genes (EPOR and EKLF), were found to be enriched in the ChIP-on-chip study of GATA1. Overall, the percentages of promoters being significantly enriched in all bioreplicates for each of these four transcription factor ChIP-on-chip assays was approximately 1% of the total number of promoters on the array. However, no promoters were found to be consistently enriched in all three of the biological replicates for SCL and LMO2. Both of these ChIP-on-chip assays were consistently the worst performing (in terms of enrichments) of all of the assays used.

A larger set of significantly enriched promoters were identified using method B. The number of promoters identified by the transcription factor ChIP-on-chip assays in method B ranged from 15 to 41. Unlike the results obtained for method A, a number of promoters were found to be significantly enriched for SCL and LMO2 using this method. Promoter targets identified by methods A and B were also compared in Venn diagrams. All the promoters identified by method A for a given transcription factor were also identified by method B for the same transcription factor (Figure 5.15). In total, using the two different statistical methods of analyses described here, over 100 promoters of putative target genes were found to be enriched in ChIP-on-chip analysis.



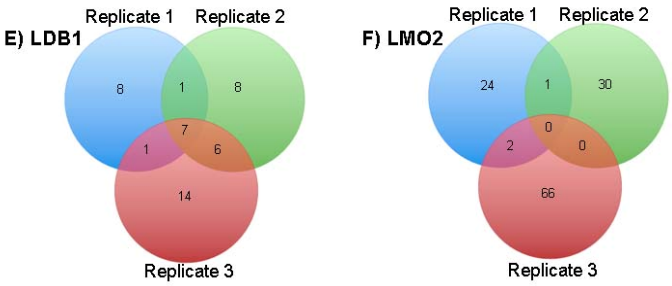


Figure 5.14. Venn diagram comparison of putative target promoters identified in all three bioreplicates for each transcription factor ChIP-on-chip assay using statistical method A. Numbers shown in the Venn diagrams were numbers of promoters identified in each biological replicate of the ChIP-on-chip studies. Panel A: Venn diagram of GATA1 ChIP-on-chip study; panel B: Venn diagram of SCL ChIP-on-chip study; panel C: Venn diagram of E12 ChIP-on-chip study; panel D: Venn diagram of E47 ChIP-on-chip study; panel E: Venn diagram of LDB1 ChIP-on-chip study; panel F: Venn diagram of LMO2 ChIP-on-chip study.

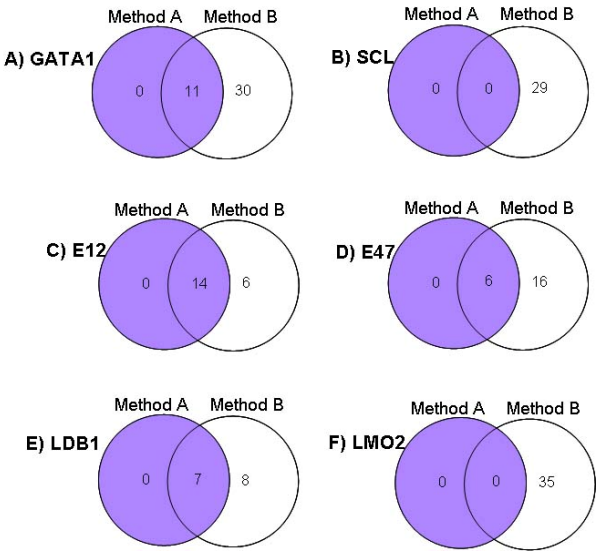


Figure 5.15. Venn diagram comparison of putative target promoters identified in ChIP-on-chip studies for each of the five transcription factors of SCL erythroid complex using statistical methods A and B. Numbers shown in the Venn diagrams were numbers of promoters identified in each biological replicate of the ChIP-on-chip studies. Panel A: Venn diagram of GATA1 ChIP-on-chip study; panel B: Venn diagram of SCL ChIP-on-chip study; panel C: Venn diagram of E12 ChIP-on-chip study; panel D: Venn diagram of E47 ChIP-on-chip study; panel E: Venn diagram of LDB1 ChIP-on-chip study; panel F: Venn diagram of LMO2 ChIP-on-chip study.

5.4.3.3 Classification and literature review of putative target genes

The transcription factor promoters which were identified as being significantly enriched in the ChIP-on-chip analysis described above were considered to represent regulatory interactions of putative direct target genes of members of the SCL erythroid complex. The putative target genes for

GATA1, E12, E47 or LDB1 selected by method A are summarised in Table 5.2. Some of the targets were enriched by more than one transcription factor using the method A criteria (for example, EPOR was identified with GATA1, E12, E47 and LDB1 and the SCL +51 enhancer was identified with all six transcription factor assays). The promoters identified in method B for the 6 transcription factor ChIP-on-chip assays were also cross-compared with the enriched promoters obtained by method A. Indeed, some of these target genes identified by method A for one transcription factor were also identified by method B for another transcription factor. This is also shown in Table 5.2. Taken all together, this data provided further evidence that at least some of the target gene promoters may be bound by the whole erythroid complex or variations thereof.

The vast majority of the targets identified were transcription factors, with the exception of EPOR (the known target of GATA1 in the erythroid lineage). To further understand the nature of the putative targets of members of the SCL erythroid complex, information was obtained from public databases including iHOP (http://www.ihop-net.org/UniPub/iHOP/), OMIM (http://www.ncbi.nlm.nih.gov/sites/entrez?db=OMIM) and Gene Expression Atlas (http://expression.gnf.org/cgi-bin/index.cgi) and from performing literature searches. This information is summarised in Table 5.2 and Figure 5.16. The ChIP-on-chip studies were able to identify additional putative targets with known function in the haematopoietic compartment. Based on the method A analysis, ten of the target genes were known to be expressed in the lymphoid lineage while five others (including SCL) were found to be expressed in early blood progenitors found in the bone marrow. Six of the genes (including SCL) were shown to be involved in haematopoietic development. Furthermore, eight of the target gene encoded proteins involved in chromatin remodelling/chromatin modifications. The putative promoters identified for the 6 transcription factor ChIP-on-chip assays using statistical methods A and method B were classified by function as shown in Figure 5.16.

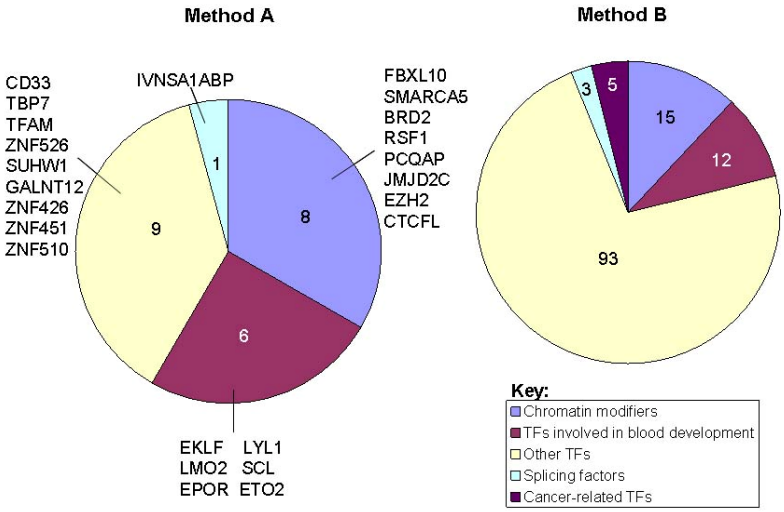


Figure 5.16. Classification of putative target genes of members of the SCL erythroid complex based on ChIP-on-chip studies. Pie charts show the classification of transcription factors identified by one or more transcription factors in the SCL erythroid complex using statistical method A (left panel) or method B (right panel). Numbers indicated in the pie charts show number of target genes in each category. The gene symbols shown in the method A pie chart are further summarised in Table 5.2. Each functional category is depicted by the colour code shown in the key.

Name of putative target gene	TF regulating target (method A)	TF regulating target (method B)	Expression pattern	Functions of putative target gene
FBXL10*	GATA1		Low expression in thymus and $CD4^+$ T-cells	JmjC domain-containing histone demethylation protein. Involved in chromatin modification and recruitment to chromatin. Cooperate with MBD1 to regulate transcription at methylated CpG sequences.
EKLF*	GATA1, E12, E47, LDB1	SCL	Bone marrow	Regulator of erythropoiesis. Transcriptional activator of β-globin expression.
ZNF526	GATA1			
TFAM	GATA1		Low expression in $CD4^+$ T-cells	Regulator of mitochondrial DNA replication.
TBP7	GATA1			Subunit of 26S protease required for ubiquitination.
SMARCA5*	GATA1		Low expression in $CD4^+$ T-cells	Associates with RSF1 and is required for chromatin assembly. Component of a chromatin-remodelling complex.
IVNSA1ABP	GATA1	LDB1	Low expression in $CD4^+$ T-cells	Involved in mRNA nuclear export and pre-mRNA splicing.
SUHW1	GATA1		Low expression in all cell types	
LMO2*	GATA1	E12, LDB1	Bone marrow	Regulator of erythroipoietic and endothelial development. Member of the SCL erythroid complex.
GALNT12	GATA1		$CD4^+$ T-cells and lung	Plays an important role in the initial step of mucin-type oligosaccharide biosynthesis in digestive organs.
EPOR*	GATA1, E12, E47, LDB1		Bone marrow	Required for differentiation and maturation of erythrocytes and programmed cell death
BRD2*	E12	GATA1, E47, LDB1	Low expression in thymus	Associated with E2F and involved in H4 acetylation
CTCFL*	E12	GATA1	Low expression in all cell types	Paralogue of the insulator CTCF which shares the same DNA-binding domain as CTCF and expressed in a mutual exclusive manner as CTCF. Its expression is activated in a wide-range of cancers. Possibly involved in epigenetic reprogramming of CTCF-binding sites.
ZNF 426	E12		$CD4^+$ T-cells	
RSF1*	E12		Low expression in $CD4^+$ T-cells	Associates with SMARCA5 and is required for chromatin assembly.
LYL1*	E12, E47, LDB1		Bone marrow	Dimerises with E2A. Chromosomal translocation leads to T-ALL

				Regulator of erythroid differentiation Highly similar in expression and function with SCL.
ZNF451	E12, E47, LDB1	GATA1	Low expression in $CD4^+$ T-cells	
PCQAP*	E12	E47	Low expression in $CD4^+$ T-cells	Mediates chromatin-directed transcriptional activation through protein complex formation.
JMJD2C*	E12	LDB1	$CD4^+$ T-cells	Contains histone demethylase activity. PHD finger domain protein. Overexpression leads to progression of cancer.
ETO2*	E12, E47, LDB1	SCL	Thymus	Breast-tumor suppressor gene. Repressor of early erythroid gene expression. Fusion partner of RUNX1 in leukemia-related translocation. Member of SCL erythroid complex.
ZNF510	E47, LDB1	GATA1	Low expression in all cell types	
EZH2*	LDB1	E47, GATA1, SCL	Thymus	Histone lysine methyltransferase. Associated with transcriptional repression. Methylate histone H1 and H3.
CD33	LDB1			Antigen expressed in myeloid lineage
SCL*	GATA1, E12, E47, SCL, LDB1, LMO2		High expression in HSCs and erythroid progenitors	Regulator of haematopoeitic development. Member of the SCL erythroid complex.

Table 5.2. Putative target promoters of members of the SCL erythroid complex. This table shows the expression pattern and function of the putative target genes identified for one or more of the 5 members of the SCL erythroid complex using method A (second column) or method B (third column). The genes marked with an asterisk were chosen for further characterisation (section 5.4.4).

5.4.4 Characterisation of a subset of putative target genes

5.4.4.1 Criteria for selection of subset of genes for further studies

In order to make additional characterisation of putative targets possible in the context of this project, the following criteria were used to select a subset of genes for further analyses.

- **Significantly enriched in all biological replicates**

Since method A was the more stringent approach for selecting statistically significant putative target genes, this gene list was used for choosing a subset of genes for further analyses. However, this gene list did not include any putative targets for SCL and LMO2. Therefore, putative target genes of SCL and LMO2 selected by method B were also included for further analysis.

- **Putative targets of more than one member of the SCL erythroid complex**

Since the main objective of this project was to identify direct transcriptional targets of the entire SCL erythroid complex in haematopoietic development, target genes which were identified by more than one of the transcription factor ChIP-on-chip assays were prioritised for further study.

- **Haematopoietic function**

Given that the SCL erythroid complex has been shown to regulate genes in the erythroid lineage, genes with known involvement in erythropoiesis or expression in the erythroid lineage were prioritised for further analysis.

- **Chromatin function**

Surprisingly, a number of target genes were identified whose functions were related to chromatin structure and function. As there is currently tremendous scientific interest in these proteins, the functions of which have widespread effects on the regulation of all genes in transcriptional programmes, these target genes were also selected for follow-up studies.

In summary, fourteen target genes were chosen which satisfied the first or second criteria and at least one of the functional criteria. These targets are highlighted with an asterisk in Table 5.2. Additional studies were then performed to validate the promoter binding events and further characterise the putative target genes. These included:

1. **Transcription factor motif identification in promoter regions:** Given that each promoter array element on the transcription factor array was approximately 1 kb in size, the potential binding site of the transcription factors to DNA sequence motifs in the promoter region were likely to be found within this one kilobase segment (however, the sites of binding could also be close to, but not within, this one kilobase of sequence). To identify the possible binding site of

the transcription factors in the promoter region, the DNA sequences of the promoters of each of the fourteen targets were screened for consensus transcription factor binding sites and the conservation of these sites were then compared across species. The presence of the relevant transcription factor binding motifs in regions of sequence conservation would provide additional evidence that the transcription factors had *bone fide* binding sites within these promoters.

2. **ChIP-qPCR validation of transcription factor binding events:** transcription factor-promoter interactions were then validated by ChIP-qPCR using the putative transcription factor binding motifs as locations around which the qPCR assays were designed. This validation was performed in K562 cells, where the interaction was initially identified and also in a second erythroid cell line, HEL. The validation of the transcription factor binding events in a second cultured cell line of a similar developmental state as K562 would support the biological relevance of these binding sites in regulating these target genes *in vivo*.

3. **Effect of knockdown of transcription factors on target gene expression:** The expression changes of these target genes were also investigated in time-course experiments of siRNA knockdowns of members of the SCL erythroid complex (to be discussed in Chapter 6). This would provide evidence that perturbations of the SCL erythroid complex affect the expression of the target genes.

All these studies together would provide further evidence that these genes are direct targets of the transcription factors which are found in the SCL erythroid complex. Figure 5.17 summarised the studies performed in characterising the putative direct target genes of the SCL erythroid complex.

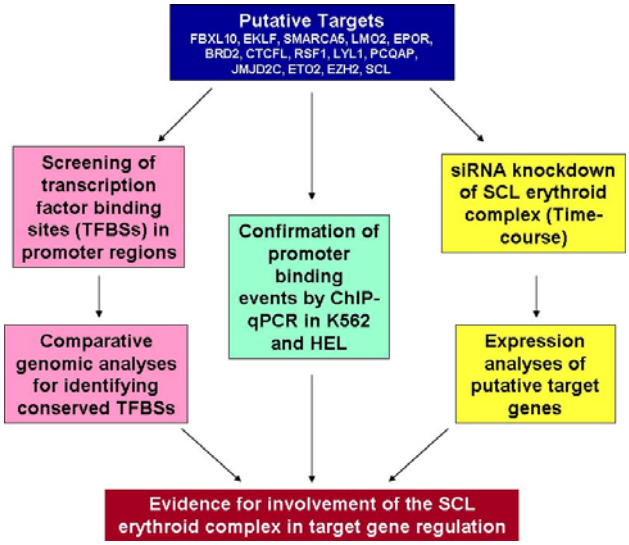


Figure 5.17. Follow-up characterisation of selected putative targets from the ChIP-on-chip experiments. The 14 putative target genes selected in the ChIP-on-chip study were further characterised in three analyses: (i) screening of conserved transcription factor binding sites; (ii) confirmation of promoter binding by the transcription factors by ChIP-qPCR in K562 and HEL; (iii) expression analyses of the 15 putative target genes in siRNA knockdown of the TFs. The information generated in these analyses will provide evidence for the involvement of the SCL erythroid complex in target gene regulation.

5.4.4.2 Transcription factor binding sites (TFBS) studies and comparative genomic analyses of enriched promoters

For screening of transcription factor binding sites in the promoter regions of the target genes, a 4 kb window (1 kb downstream of the TSSs and 3 kb upstream of the TSSs) was used to identify all possible transcription factor binding sites (TFBSs) using TESS and TFSEARCH. Transcription factor binding sites are conserved sequences with a certain degree of degeneracy which TFs recognise and bind. TESS and TFSEARCH are two web-based motif search algorithms which use the TRANSFAC TFBS database to identify TFBSs in genomic sequences (Chapter 1, section 1.3.4.2). Whilst the transcription factor binding events were likely to be present in the region encompassed by the approximately 1 kb contained within the promoter array elements, the windows for TFBS search were expanded to 4 kb to ensure that all possible TFBSs mapping near the TSS were identified. This would ensure that any motifs (and possible locations for transcription factor binding events) which were located close to, but not within, the sequences represented on the array elements, could be identified. In particular, E-box motifs of the E2A/SCL type and GATA motifs were identified within these promoter sequences. Given that the SCL erythroid complex binds to a composite E-box/GATA site separated by 9-12 bases in regulatory elements of its target genes, the location of clusters of E-box and GATA consensus sequences was of particular interest.

Following the mapping of relevant TFBS, the conservation of these binding sites across species was investigated. This allowed us to identify evolutionarily and functionally important DNA-binding motifs (Chapter 1, section 1.3.4.3). Multiple sequence alignments of the 4 kb of sequences around the TSSs were downloaded from the UCSC genome browser (http://genome.ucsc.edu/). These were derived from a variety of species including human, mouse, rat, rabbit, dog and chicken. Any DNA binding sites for E2A/SCL and GATA1 were carefully scrutinised for sequence conservation.

Relevant TFBSs and multi-species sequence alignments at the promoters of the fifteen target genes are shown in Appendix 4. Detailed descriptions of the possible TFBSs are given as follows:

A. BRD2: E-box and GATA motifs separated by 12 bases were identified. They are highly, though not completely, conserved across species.

B. CTCFL: Only one conserved GATA site was identified.

C. EKLF: Two possible TFBSs were identified. In the first one, three conserved E-box motifs separated by 6-12 bases were identified. In the second one, two conserved GATA sites separated by 47 bases were identified. However, no E-box and GATA motifs in close proximity were found in the promoter region studied.

D. EPOR: Three E-box motifs with high conservation across species separated by 6-13 bases were identified.

E. ETO2: E-box and GATA motifs separated by 9 bases were identified. Both motifs are fully conserved across species.

F. EZH2: Three possible TFBSs were identified. In the first two, E-box and GATA motifs separated by 10 to 23 bases were identified. They are highly, though not completely, conserved across species. In the third one, an E-box motif was identified with high conservation across species.

G. FBXL10: E-box and GATA motifs separated by 55 bases were identified. They are highly, though not completely, conserved across species.

H. JMJD2C: Three possible TFBSs were identified. In the first two, E-box and GATA motifs separated by 19 to 56 bases were identified. They are highly, though not completely, conserved across species. In the third one, an E-box motif was identified with full conservation across species.

I. LMO2: Two possible TFBSs were identified. In the first one, two conserved E-box motifs separated by 29 bases were identified. In the second one, one conserved GATA site was identified. However, no E-box and GATA motifs in close proximity were found in the promoter region studied.

J. LYL1: Two possible TFBSs were identified. In the first one, two fully-conserved GATA sites separated by 25 bases were identified. In the second one, one conserved E-box was identified. However, no E-box and GATA motifs in close proximity were found in the promoter region studied.

K. SCL: The +51 enhancer of SCL was selected for qPCR validation. This +51 enhancer contains the consensus E-box/GATA motifs separated by 9 bases. This was included in the validation as a positive control and reference for the qPCR assays.

L. SMARCA5: Three possible TFBSs were identified. In the first one, E-box and GATA motifs separated by 9 bases were identified. They are highly, though not completely, conserved across

species. In the third one, two GATA sites separated by 5 to 61 bases were identified. However, the conservation of these GATA sites was not high.

M. PCQAP: Four possible TFBSs were identified. In the first two, a single E-box motif was identified. The conservation for the first one was not high whereas there was no alignment with other species for the second one. In the other two, E-box and GATA motifs separated by 32 to 61 bases were identified. The conservation for the first one was not high whereas there was no alignment with other species for the second one.

N. RSF1: Only one conserved E-box motif was identified in the promoter region. No E-box and GATA motifs in close proximity were found in the promoter region studied.

5.4.4.3 ChIP-qPCR validation of promoter binding events

Based on the locations of conserved E-box and GATA motifs, qPCR assays were designed and validation of the transcription factor-promoter binding events was performed using ChIP-qPCR. In ChIP-qPCR, the input and ChIP DNAs were subjected to SYBR Green real-time quantitative PCR analyses. TFBS regions amplified in both input and ChIP DNA were quantified and compared. To normalise the fold enrichments above background, ChIP-qPCR was also performed for eleven negative control regions in the SCL locus which do not give enrichments above background for members the SCL erythroid complex (Appendix 1E). The average enrichment for these eleven regions was determined for every ChIP-qPCR assay and this value was used to scale the ChIP-qPCR enrichments of the promoter binding events so that the enrichment for negative control regions was a baseline of 1.

To identify statistical significant enrichments for the transcription factor binding sites tested for the selected putative target, cut-offs for significant enrichment were chosen. These cut-offs were different for different ChIP assays as the efficiency of antibodies differed. The enrichments of the eleven negative regions on the SCL locus were used as the baseline for determining significant fold enrichments. The standard deviations and average of these eleven regions in the two biological replicates for each ChIP assay of transcription factor were calculated. A fold enrichment cut-off was identified as the two standard deviations above the mean of enrichment i.e. a 99.45% confidence level is reached.

Three levels of validations were performed:

(i) Confirmation of ChIP-on-chip data was performed so that the identified TFBSs were tested for enrichment in ChIP-qPCR with ChIP DNAs from the assay which showed enrichments in ChIP-on-chip in K562 cells (analysed by method A).

(ii) Promoter binding events were tested for all transcription factors in the SCL erythroid complex in K562 cells to detect binding events which were missed using ChIP-on-chip.

(iii) Confirmation of binding events in a second, but somewhat similar, cell line was done to test the biological relevance of *in vivo* promoter-binding events in K562 cells.

A. Confirmation of ChIP-on-chip data

Since more than one region was found to be conserved with E-box and GATA motifs in many of the target promoters, all of them were first tested for enrichments in ChIP-qPCR. Initially, these regions were tested for enrichment in ChIP-qPCR with ChIP DNAs from the assay which showed enrichments in ChIP-on-chip in K562 cells (analysed by method A) (Table 5.3). In 10 out of the 14 chosen promoters, at least one qPCR region per promoter was shown to have a significant enrichment above the cut-off for at least one member of the SCL erythroid complex. However, enrichments for all tested regions of FBXL10, PCQAP, EZH2 and RSF1 were less than the cut-offs and unfortunately no other regions in the promoters showed conserved E-box and GATA motifs. These four genes were excluded in subsequent ChIP-qPCR analyses.

Region tested	GATA1 ChIP (>4.4)	E12 ChIP (>3.1)	E47 ChIP (>3.8)	LDB1 ChIP (>2.7)
BRD2		8.28		
CTCFL		6.76		
EKLF (1)	0	3	1.1	0.9
EKLF (2)	8.02	7.94	2.28	5.27
EPOR	20.57	4.24	4.14	3.14
ETO2		31.35	10.83	29.84
FBXL10	2.79			
JMJD2C (1)		31.35		
JMJD2C (2)				
JMJD2C (3)		3.55		
LMO2 (1)	4.53			
LMO2 (2)	1.2			
LYL1 (1)		37.99	14.91	62.64
LYL1 (2)		2.15	0.63	1.1
SCL	13.88	59.98	17.04	44.07
SMARCA5 (1)	0.71			
SMARCA5 (2)	1.53			
SMARCA5 (3)	19.77			
PCQAP (1)		0.78		
PCQAP (2)		0.64		
PCQAP (3)		0.97		
PCQAP (4)		1.55		
EZH2 (1)				1.3
EZH2 (2)				1.1
EZH2 (3)				1
RSF1		1.6		

Table 5.3. Fold enrichments putative target promoters tested in ChIP-qPCR. The fold enrichments of all the regions selected for the 14 putative target promoters in the confirmation by ChIP-qPCR are shown. The cut-offs of fold

enrichments used for each ChIP assay are shown in the first row in brackets. The regions which show a fold enrichment above the cut-offs are highlighted in yellow. These regions were also highlighted with an asterisk in Appendix 4 (where more than one region tested) and were chosen for subsequent ChIP-qPCR analyses.

B. Study of promoter binding for 5 members of the SCL erythroid complex in K562

From the results of the ChIP-on-chip studies, not all the selected putative target gene promoters were found to be bound by all 5 of the transcription factors in the SCL erythroid complex. However, all selected putative target promoters were tested by ChIP-qPCR for all the 5 transcription factors in K562 cells. This would allow for binding events which were missed using ChIP-on-chip to be detected by the more sensitive PCR-based assay. Two independent biological replicates were performed for each experiment.

For the ChIP-qPCR in K562, nine out of ten of putative target promoters showed significant enrichments above background passing the cut-offs in ChIP for at least one of the five transcription factors (Figure 5.18 and Table 5.4). In some cases where enrichments were only observed in ChIP-on-chip experiments for one transcription factor, they were shown to be enriched for some of the other transcription factors by ChIP-qPCR. For example, LYL1 was originally identified in the E12, E47 and LDB1 ChIP-on-chip but not in GATA1 and SCL - but it was shown to be enriched in the GATA1 and SCL ChIP-qPCR.

None of the putative target promoters were shown to be enriched in the ChIP of all five members of the SCL erythroid complex. However, four promoters or enhancers for SCL (CTCFL, LYL1, SCL and SMARCA5) were found to be enriched in the ChIP of four members of the complex including both E2A isoforms, GATA1, LDB1 and SCL. Promoters of EPOR and ETO2 were enriched in the ChIP of three members including GATA1, both E2A isoforms and LDB1 while promoters of BRD2 and EKLF were also enriched in the ChIP of three members including GATA1, the E12 isoform of E2A and LDB1. Promoter of LMO2 was only enriched in the GATA1 ChIP assay. Although the JMJD2C promoter showed significant enrichments above cut-off in the E12 ChIP assay in the initial screening (Table 5.3), no significant enrichments were shown in the ChIP of all five members of the SCL erythroid complex in the current study (Table 5.4).

Among the five members of the SCL erythroid, GATA1 bound to the largest number of promoters or enhancers (for SCL) (9 of them) while E12 and LDB1 bound to 8 of them. LMO2 was shown to bind to none of the promoters or enhancers tested (possibly due to the poor quality of the ChIP assay for LMO2).

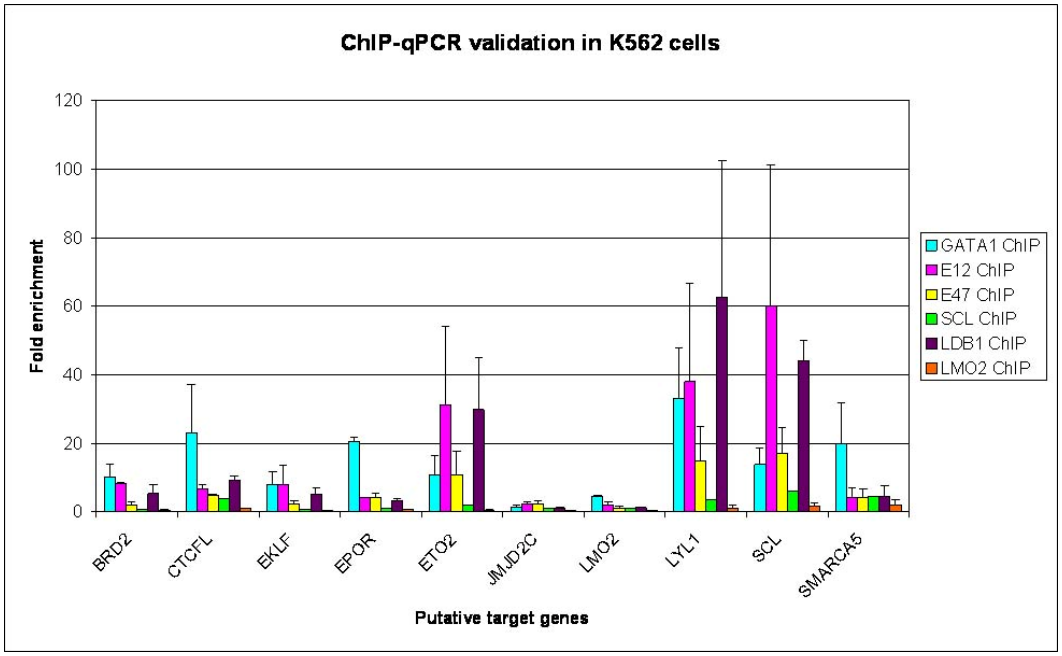


Figure 5.18. ChIP-qPCR analyses of selected putative target genes in K562. Histogram shows the fold enrichments of selected regions for putative target genes in ChIP-qPCR. Y-axis: fold enrichments above background. X-axis: putative target genes. The ChIP experiments represented by the colour bars are shown in the key on the right. Error bars show standard errors of two biological replicates.

Putative target	Fold enrichment					
	GATA1 (>4.4)	E12 (>3.1)	E47 (>3.8)	SCL (>3.1)	LDB1 (>2.7)	LMO2 (>3.1)
BRD2	10.13	8.28	1.97	0.69	5.37	0.44
CTCFL	23.03	6.76	4.99	3.97	9.15	1.02
EKLF	8.02	7.94	2.28	0.74	5.27	0.46
EPOR	20.57	4.24	4.14	1.21	3.14	0.66
ETO2	10.9	31.35	10.83	1.85	29.84	0.37
JMJD2C	1.35	2.33	2.42	1.2	1.17	0.5
LMO2	4.53	1.98	1.2	1.14	1.24	0.43
LYL1	32.98	37.99	14.91	3.58	62.64	1.1
SCL	13.88	59.98	17.04	6.22	44.07	1.67
SMARCA5	19.77	4.3	4.32	4.46	4.52	2.1
Total Validated	9	8	6	4	8	0

validated above threshold

Table 5.4. Fold enrichments of selected putative target promoters in ChIP-qPCR in K562 cells. The fold enrichments of the regions selected for the 10 putative target promoters in ChIP studies of five members of the SCL erythroid complex by ChIP-qPCR are shown. The cut-offs of fold enrichments used for each ChIP assay are shown in

the first row in brackets. The promoters which show a fold enrichment above the cut-offs are highlighted in green boxes. The total number of validated target genes for each member is shown in the bottom of the table.

A comparison of the interactions between promoters or enhancers (in the case of the +51 region of SCL) and transcription factors observed in ChIP-on-chip (analysed with methods A and B) and ChIP-qPCR was performed (Tables 5.5). Twenty-five binding events (56.8%) were observed in both assays (shown in green boxes in Table 5.5). Nine binding events (20.5%) were only observed in ChIP-on-chip (shown in blue boxes in Table 5.5), and six of these were identified by the less stringent method B analysis (which may be less reliable at identifying real binding events). Ten binding events (22.7%) were only observed in ChIP-qPCR (shown in pink boxes in Table 5.5). Overall, this analysis shows that the majority of ChIP-on-chip interactions were confirmed and that both approaches are complimentary at detecting interactions missed by the other method.

	Comparison between ChIP-on-chip and ChIP-qPCR in K562					
Putative target genes	GATA1	E12	E47	SCL	LDB1	LMO2
BRD2			*			
CTCFL						
EKLF				*		
EPOR						
ETO2				*		
JMJD2C					*	
LMO2		*			*	
LYL1						
SCL						
SMARCA5						

Validated in ChIP-on-chip only
Validated in both ChIP-on-chip and ChIP-qPCR
Validated in ChIP-qPCR only
* Interaction picked up by method B

Table 5.5. Comparison between ChIP-on-chip and ChIP-qPCR in K562 cells. The green boxes indicate promoter/enhancer binding events which were observed in both ChIP-on-chip and ChIP-qPCR. The pink and blue boxes indicate promoter/enhancer binding events which were observed in only in ChIP-qPCR or in ChIP-on-chip respectively

C. Study of promoter binding for 5 members of the SCL erythroid complex in HEL

Given that K562 is a cell line originally derived from a patient with chronic myeloid leukaemia (CML), the information derived from the ChIP-on-chip experiment performed in this study may not reflect the *bona fide* binding events found in normal erythroid cells. Confirming the promoter binding events in a second, but somewhat similar, cell line would provide further confidence of the true *in vivo* promoter-binding events (although cell culture may affect these binding events in both cell lines). Therefore, to further characterise the transcription factor binding at specific E-box or

GATA motifs in the selected promoter/enhancer regions, ChIP material from another cell line (HEL) was used. K562 and HEL are both erythroid progenitor cell lines which can be spontaneously differentiated into erythroid cells. However, developmentally, HEL cells represent a more mature erythroid cell population than K562. This was confirmed by flow-analysis of the erythrocytic surface marker glycophorin A which showed that a larger proportion of HEL cells expressed GPA than was found in K562 (Figure 5.19). Furthermore, HEL cells do not contain the BCL-ABL translocation (which is known to affect gene expression) (Martin and Papayannopoulou, 1982), suggesting that gene expression patterns in this cell line may reflect normal erythroid development more so than K562.

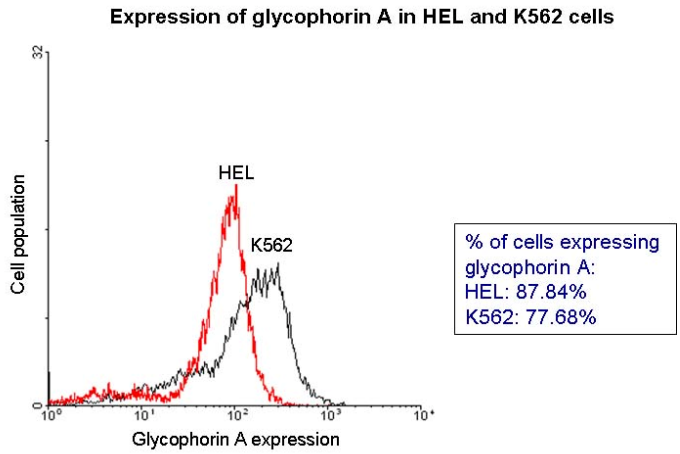


Figure 5.19. Flow analysis of glycophorin A expression in HEL and K562 cell lines. X-axis: Glycophorin A expression; y-axis: number of cells in population. The red curve shows the pattern for HEL cells while the black curve shows the pattern of K562 cells. % of cells in each population expressing GPA was calculated by WINMDI software and is shown in the box on the right.

For the ChIP-qPCR in HEL, eight out of ten of putative target promoters showed significant enrichments above baseline cut-offs in ChIP for at least one of the five transcription factors (Figure 5.20 and Table 5.6). Only the SCL +51 enhancer was shown to be enriched in the ChIP of all five members of the SCL erythroid complex. Two promoters (LYL1 and BRD2) were found to be enriched in the ChIP of four members of the complex including both E2A isoforms, GATA1, LDB1 and SCL. Promoter of ETO2 was enriched in the ChIP of three members including GATA1, both E2A isoforms and LDB1. Promoter of EKLF was enriched in the GATA1 and E12 ChIP assays. Promoters of LMO2 and EKLF were only enriched in the GATA1 ChIP assay. Two promoters were not enriched in the ChIP assays for all members of the SCL erythroid complex (CTCFL and JMJD2C). Again, the validation rates were in agreement with the quality of the ChIP assay – with validation for SCL and LMO2 showing the lowest levels.

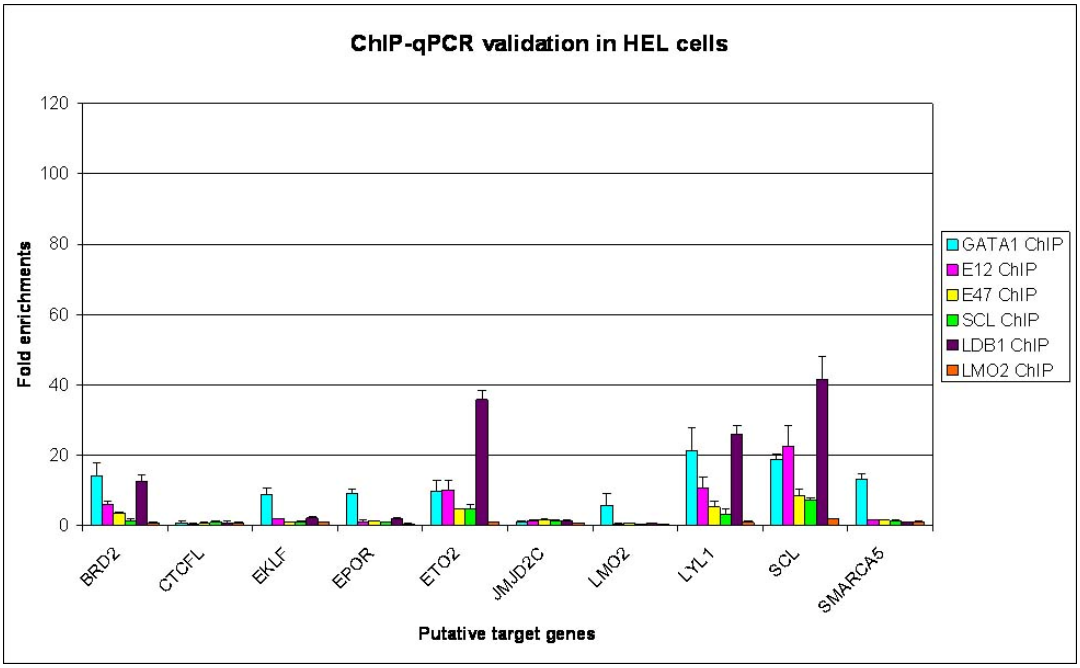


Figure 5.20. ChIP-qPCR analyses of selected putative target genes in HEL. Histogram shows the fold enrichments of selected regions for putative target genes in ChIP-qPCR. Y-axis: fold enrichments above background. X-axis: putative target genes. The ChIP experiments represented by the colour bars were shown in the key on the right. Error bars showed standard errors of two biological replicates.

Putative target	Fold enrichment					
	GATA1 (>3.6)	E12 (>1.5)	E47 (>2.1)	SCL (>1.4)	LDB1 (>2.2)	LMO2 (>1.3)
BRD2	14.02	5.93	3.47	1.41	12.67	0.75
CTCFL	0.74	0.41	0.55	0.95	0.77	0.58
EKLF	8.91	2.02	1.04	1.17	2.12	1
EPOR	9.05	1.05	1.26	1.02	2.1	0.47
ETO2	9.75	10.06	4.63	4.88	35.83	0.91
JMJD2C	1.15	1.35	1.61	1.3	1.33	0.69
LMO2	5.7	0.52	0.53	0.37	0.55	0.43
LYL1	21.28	10.63	5.24	3.06	25.94	0.86
SCL	18.81	22.5	8.37	7.33	41.67	1.81
SMARCA5	13.3	1.54	1.5	1.35	0.88	0.9
Total Validated	8	6	4	3	4	1

validated above threshold

Table 5.6. Fold enrichments selected putative target promoters in ChIP-qPCR in HEL cells. The fold enrichments of the regions selected for the 10 putative target promoters in ChIP studies of five members of the SCL erythroid complex by ChIP-qPCR are shown. The significance cut-offs of fold enrichments used for each ChIP assay are shown in the first row in brackets. The promoters which show a fold enrichment above the cut-offs are highlighted in green boxes. The total number of validated target genes for each member is shown in the bottom of the table.

D. Comparison between the validated targets of K562 and HEL

A comparison of the binding of transcription factors in the SCL erythroid complex to the promoters of their putative target genes was made between the data obtained in K562 and HEL cells. This was done to study the biological relevance of promoter binding in K562 cells. Table 5.7 compared the interactions found in each cell line. Twenty-two binding events (60%) were shown to be the same in both K562 and HEL cells (shown in green boxes in Table 5.7). Twelve binding events (32%) were only observed in K562 (shown in pink boxes in Table 5.7) while three binding events (8%) were only observed in HEL (shown in blue boxes in Table 5.7). Particularly, the CTCFL promoter was only enriched in all the ChIP assays in K562 cells but none in HEL cells. As a large proportion of promoter binding events were found in both cell lines, there is a high level of confidence that the data obtained in K562 is biologically relevant.

Deleted: the

Putative target genes	Validation in the ChIP assays in K562 and HEL cells					
	GATA1	E12	E47	SCL	LDB1	LMO2
BRD2						
CTCFL						
EKLF						
EPOR						
ETO2						
JMJD2C						
LMO2						
LYL1						
SCL						
SMARCA5						

HEL
K562 & HEL
K562

Table 5.7. Comparison between the ChIP-qPCR assays in K562 and HEL cells. The green boxes indicate promoter/enhancer binding events which were observed in both K562 and HEL. The pink and blue boxes indicate promoter/enhancer binding events were observed in only in K562 and HEL respectively.

5.4.4.4 Comparison of ChIP-on-chip, ChIP-qPCR and motif analyses

The results obtained in ChIP-on-chip, ChIP-qPCR and the *in silico* motif analyses were used to deduce whether a particular promoter was regulated by one or more members of the SCL erythroid complex or the whole complex. A summary of the combined data for the 24 putative target genes (described first in Table 5.2) are summarised in Table 5.8. The criteria used to make these deductions were as follows:

1. Target of any one of the five transcription factors:

- There must be evidence of significant enrichments in at least one of the ChIP analyses (ChIP-on-chip, ChIP-qPCR in K562 and ChIP-qPCR in HEL) for a gene to be considered as a direct target of any one transcription factor.

2. Target of the whole SCL erythroid complex (all five members):

- Significant enrichments must be observed in at least three ChIP assays (GATA1, E12 or E47, and LDB1) in either ChIP-on-chip, ChIP-qPCR in K562 or ChIP-qPCR in HEL. Also, both GATA and E-box motifs with the spacing of 9-12 bp must be identified in the promoters. Due to the poor quality of the SCL and LMO2 ChIP assays, a target was not required to demonstrate enrichments for these two TFs.

OR

- Significant enrichments must be observed in at least four ChIP assays (GATA1, E12 or E47, LDB1, and either SCL or LMO2) in either ChIP-on-chip, ChIP-qPCR in K562 or ChIP-qPCR in HEL. No motif data was required (this would allow *trans* interactions between enhancers (containing a motif) and promoters (not containing a motif) to be included as targets.

Based on this analysis, all 24 genes were considered as targets of at least one transcription factor, while 8 genes were considered to be direct targets of the whole SCL erythroid complex. These eight genes were BRD2, CTCFL, EKLF, ETO2, LYL1, SCL, SMARCA5 and EZH2. Four genes (EPOR, LMO2, ZNF451 and ZNF510) were found to be direct targets of GATA1, E2A (E12 or E47 or both) and LDB1. These genes may be direct targets of a novel complex containing GATA1, E2A and LDB1 but they may also be possible targets of the whole SCL erythroid complex due to the poor quality of the SCL and LMO2 antibodies. Only conserved E-box motifs were found in the EPOR and LMO2 promoters while no motif analyses was performed for ZNF451 and ZNF510.

Putative target gene	ChIP-on-chip (methods A & B)						ChIP-qPCR (K562)						ChIP-qPCR (HEL)						Motif analysis				Interpretation
	GATA1	E12	E47	SCL	LDB1	LMO2	GATA1	E12	E47	SCL	LDB1	LMO2	GATA1	E12	E47	SCL	LDB1	LMO2	E-box motif	GATA motif	Motif sequence conservation	Spacing between E-box and GATA	
BRD2																					High	12 bases	Target of SEC
CTCFL																					High		Target of SEC
EKLF																					High		Target of SEC
EPOR																					High		Target of GATA1, E2A and LDB1
ETO2																					Complete	9 bases	Target of SEC
JMJD2C																					High	19 bases	Target of E12 and LDB1
LMO2																					High		Target of GATA1, E12 and LDB1
LYL1																					High		Target of SEC
SCL																					Complete	9 bases	Target of SEC
SMARCA5																					High		Target of SEC
FBXL10																					High	55 bases	Target of GATA1
PCQAP																					Low	32 or 61 bases	Target of E2A
EZH2																					High	10 or 23 bases	Target of SEC
RSF1																					High		Target of E12
TFAM																							Target of GATA1
TBP7																							Target of GATA1
IVNSA1ABP																							Target of GATA1 and LDB1
SUHW1																							Target of GATA1
GALNT12																							Target of GATA1
CD33																							Target of LDB1
ZNF 426																							Target of E12
ZNF 526																							Target of GATA1
ZNF510																							Target of GATA1, E47 and LDB1
ZNF451																							Target of GATA1, E2A and LDB1

Validated targets
Non-validated targets
Not tested
Motif identified
Motif not identified

Table 5.8. Comparision of ChIP-on-chip, ChIP-qPCR and motif analyses. Putative targets identified in ChIP-on-chip, ChIP-qPCR in K562 and HEL are shown as green boxes for validated targets and red boxes for non-validated targets. E-box or GATA motifs identified in the motif analysis and contained within the sequence of the ChIP-qPCR assay are shown as blue boxes. Black boxes indicate no data available. The interpretation in the last column shows whether the putative target is confirmed as a direct target of one or more transcription factor or of the whole SCL erythroid complex (SEC) using the criteria detailed in section 5.4.4.4.

5.5 Discussion

The results of this Chapter describe the use of the ChIP-on-chip method to study the binding events of transcription factors of the SCL erythroid complex to promoter regions of target genes. Five transcription factors (GATA1, SCL, LMO2, LDB1, and two isoforms of E2A – E47 and E12) in the SCL erythroid complex were studied by ChIP-on-chip in K562 cells using an in-house transcription factor promoter array. A number of transcription factors related to haematopoietic development and chromatin remodelling were identified as putative targets of some or all members of the SCL erythroid complex in these ChIP-on-chip studies. These targets were confirmed in subsequent ChIP-qPCR, and by *in silico* transcription factor binding site and comparative sequence analysis.

5.5.1 Validation of promoter-binding events

Three levels of validation were performed in section 5.4.4.3 including (i) confirmation of ChIP-on-chip data, (ii) testing of promoter binding events for all transcription factors in the SCL erythroid complex in targets identified in K562 cells, and (iii) confirmation of biological relevance of K562 cells by identifying TF-binding events of K562 targets in HEL cells. This validation was all performed using ChIP-qPCR. The findings of each of these validation studies are discussed below.

(i) Confirmation of ChIP-on-chip data

ChIP-qPCR was first performed to confirm binding events obtained in ChIP-on-chip. Promoter binding events of members 10 out of 14 putative target genes (71%) were validated in ChIP-qPCR assay in K562 cells. This is approximately twice the validation rate obtained for the Affymetrix GeneChip platform for studies carried out in Chapter 4. Those promoters not validated may be false-positive targets identified on the promoter array. However, given the stringent statistical criteria by which these four were selected for validation (method A criteria), this is unlikely. Alternatively, the ChIP-qPCR assays may have been designed around TFBSs where the transcription factors do not bind. Thus, the ChIP-qPCR assays used here may result in false negatives. Therefore, these four non-validated genes may still be targets of the corresponding transcription factors. Other conserved TFBS inside or outside the promoter regions should be tested by ChIP-qPCR to confirm this.

(ii) Promoter binding events for five in the SCL erythroid complex in K562 cells

The first level of validation provided a general confidence that the promoter was bound by at least one member of the SCL erythroid complex. In the second level of confirmation, ChIP-qPCR was performed to detect binding events for all five members of the complex on each target promoter in K562 cells which may have been missed by ChIP-on-chip. Of 44 TF-DNA interactions detected by either ChIP-PCR or ChIP-on-chip, 56.8% were found by both methods, and 22.7% and 20.5% were found only in ChIP-qPCR or in ChIP-on-chip assays respectively. This would suggest that false negatives as well as false positives could be present in either assay. In fact, a large proportion of the "ChIP-on-chip only" interactions (6 out of 10) were only picked up using the less stringent method B of analyses. This may suggest that method B generates a high level of false positive target promoters.

(iii) Validation of K562 promoter-binding events in HEL cells

K562 is a cancerous cell line containing a BCR-ABL translocation which may induce changes in the expression pattern of genes. In fact, one piece of evidence showing that K562 may be abnormal is that thirteen out of the 24 putative target genes selected from the ChIP-on-chip study are normally expressed in the T-cell lineage (Table 5.2), despite K562 being an erythroid cell line. This indicates that there may be abnormal regulation of genes in K562. The HEL cell line lies at a similar stage of haematopoietic development as K562 cells and is BCR-ABL negative. Therefore it is a good cell line to validate K562 targets by ChIP-qPCR confirmation. However, since HEL is also a cell line, many issues associated with gene regulation in cell culture are not resolved.

A large proportion of promoter binding events (60%) for the 5 TFs of the SCL erythroid complex were observed in both K562 and HEL cells. This indicates that the majority of the data obtained in K562 is likely to be biologically relevant. Furthermore, these common binding events may mean that similar transcriptional programmes are found in both cell lines. Yet, 32% of the binding events were only observed in K562 but not HEL. These transcription/promoter interactions may be induced by the BCR-ABL translocation. However, they may also represent interactions which are found earlier in erythroid differentiation, given that K562 cells may be slightly more immature cells in the erythroid lineage than HEL cells (as determined by GPA expression). Following from that, the 8% of the binding events which were only observed in HEL cells may only be found later in erythroid development. Furthermore, the possibility that variations of the composition of the SCL erythroid complex, or different modes of regulation of these targets are present in K562 and HEL cells. Thus, while all of the targets found in K562 may also be targets in HEL, they may be regulated in different ways or by different TFs.

5.5.2 Validation of known target genes

Only three published target genes have previously been identified for the SCL erythroid complex, namely GYPA, c-kit and α-globin. However, since they are not transcription factors and were not included on the promoter array, they could not be validated in the ChIP-on-chip study in this Chapter. However, other direct target genes of members of the SCL erythroid complex were validated in the ChIP-on-chip assays. GATA1 has previously been shown to bind to the promoters of EPOR and EKLF (Anderson et al., 1998; Zon et al., 1991) and these bindings were confirmed in the data shown in this Chapter. The SCL +51 enhancer was previously reported to be bound by GATA1, SCL and LDB1 (Pawan Dhami, PhD thesis) and these binding events were also confirmed here.

A ChIP-on-chip study was performed for a TF complex containing SCL, HEB and E2A in the leukaemic Jurkat T-cell line (Palomero et al., 2006). The putative target genes obtained in this Chapter for SCL and E2A were compared with those obtained in the Jurkat study. No target genes were found in common in both studies. One possible reason to explain this is that the two studies were performed in cell lines derived from entirely different haematopoietic lineages (erythroid versus lymphoid) and the regulatory pattern may be very different, especially since SCL is expressed in Jurkat because of its involvement in T-acute lymphocytic leukaemia. Also, the transcription factors in these two cell lines may form different multiprotein complexes and thus may bind and regulate different target genes.

5.5.3 Novel targets of the SCL erythroid complex

Based on the criteria used in this study to define TF binding events at promoters, eight genes were likely to be direct targets of the whole SCL erythroid complex. These included BRD2, CTCFL, EKLF, ETO2, LYL1, SCL, SMARCA5 and EZH2 (section 5.4.4.4). However, apart from SCL where there is evidence that LMO2 binds to the +51 enhancer, there is no experimental data showing that the other 7 promoters are bound by LMO2. However, TFBS motif analysis confirmed the presence of conserved E-box and GATA motifs in the SCL, ETO2, BRD2, SMARCA5 and EZH2 promoters, providing additional confidence that they are direct targets of the whole SCL erythroid complex. However, there is still a possibility that LMO2 is not present in the SCL eythroid complex binding to the promoters of these seven genes. Thus, these eight genes may not be regulated by the whole SCL erythroid complex. The possibility that other LMO family members are part of the complex cannot be excluded. This line of reasoning could also explain why it was not possible to confirm binding events for all five members of the complex for these eight target genes. GATA2 has been shown to play the same role as GATA1 in the SCL erythroid complex in binding to the c-kit promoter (Lecuyer et al., 2002). In addition, other transcription factors may also form

novel complexes containing other transcription factors at the promoters of these targets. In fact, transcription factors such as SP1 and ETO2 have been identified as part of the SCL erythroid complex in certain contexts (Goardon et al., 2006; Lecuyer et al., 2002). Binding events for these TFs were not performed in the present study.

It is also important to note only 14 of the 24 genes listed in Table 5.2 were studied by ChIP-qPCR in this Chapter. The remaining 10 may also be direct target of the whole SCL erythroid complex. Thus, this gene list serves as an additional source of targets for analysis in the future.

5.5.4 The sequences of the putative binding sites of the SCL erythroid complex

For the eight targets described in section 5.5.3, conserved E-box and GATA motifs with spacing ranging from 9 to 12 bp were found in the promoter or enhancers of five of them - BRD2, ETO2, SCL, SMARCA5 and EZH2 (Figure 5.21). However, ChIP-qPCR assays showed enrichment around these composite sites for only three of them – SCL, ETO2 and BRD2. According to Wadman et al. (1997), the SCL erythroid complex binds to an E-box motif with consensus sequence of CAGGTG, followed 9 bp downstream by a GATA site. However, this canonical sequence with exactly the same sequence and spacing was only observed in the SCL +51 enhancer (Figure 5.21). Collectively, for the three sites which showed enrichment in ChIP-qPCR assays (SCL, BRD2 and ETO2) variations in (i) sequence, (ii) spacing and (iii) orientation of the sites were observed. This suggests that there is flexibility in terms of the requirements for TF binding to allow the components of the complex to reside on the same face of the DNA molecule.

SCL	n **caggtg**	nnnnnnnn	c**gata**a
BRD2	n **catctg**	nnnnnnnnnnn	t**tatc**t
ETO2	n **catctg**	nnnnnnnn	t**gata**a
SMARCA5[1]	n **cagctg**	nnnnnnnn	t**tatc**c
EZH2[2]	n **catctg**	nnnnnnnnn	g**tatc**c

Figure 5.21. Alignment of composite E-box/GATA motifs found in promoter sequences of five targets of whole SCL eythroid complex. The sequences of the E-box (green) and GATA (red) are highlighted for each target. n = any nucleotide; [1]= no significant ChIP-qPCR enrichment around this site; [2] = no significant ChIP-qPCR enrichment around this site.

In four of the eight target genes (LYL1, SMARCA5, CTCFL, ELKF), GATA sites were found in the regions assayed by ChIP-qPCR. However, the ChIP-qPCR data also suggests that either E2A or SCL binds to these regions in the absence of an obvious E-box motif. A possible reason to explain ChIP enrichment of TFs in these regions is that there is a looping of DNA sequences which brings GATA and E-box motifs on different regulatory elements into close proximity, allowing for the whole complex to bind (also see section 5.5.7). This mechanism could also be invoked to explain targets which have only an E-box motif in their promoters, although none of these eight targets fall

into that category. It is also possible that the SCL erythroid complex may bind to DNA sequences in addition to the consensus sequence suggested by Wadman et al (1997). To further characterise any of these TFBSs, additional assays would need to be performed in order to provide empirical evidence that these are indeed the actual sites of TF binding of members of the SCL erythroid complex. Gel shift assays are *in vitro* analyses that can be employed to confirm the binding of transcription factors to these DNA sequences. Moreover, mutation analysis can be used to investigate the requirement of these binding sites for driving expression in reporter assays.

5.5.5 Biological roles of novel targets of the SCL erythroid complex

The identification of novel targets of the SCL erythroid complex sheds new light on the role that this complex has in controlling transcriptional programmes in erythroid development. Of the eight genes thought to be novel direct targets of the whole SCL erythroid complex, four of them have known roles in haematopoietic development (EKLF, ETO2, LYL1 and SCL). SCL is a member of the complex itself and has been shown to be indispensable for haematopoietic development (see Chapter 1). ETO2 has previously been shown to be an interacting partner of the SCL where such interaction is related to down-regulation of early erythroid gene expression (Schuh et al., 2005). It was later demonstrated to be a novel member of the SCL erythroid complex (Goardon et al., 2006). Knockdown experiments of ETO2 demonstrated its involvement in governing erythroid and megakaryocytic differentiation (Goardon et al, 2006; Hamlett et al, 2008). EKLF is a transcription factor required for terminal erythropoiesis which regulates the expression of β-globin gene (Nuez et al., 1995). In EKLF knockout mice, definitive fetal liver erythropoiesis is disrupted, leading to lethality by embryonic day 15 (Nuez et al., 1995). LYL1 has overlapping expression patterns with SCL in mouse and is expressed in the erythroid and myeloid lineages and in ascular tissues (Visvader et al., 1991) (Chapter 1, section 1.4.2.1 C). LYL1 knockout mice were shown to be viable and have normal blood counts except for a reduced number of B-cells while Lyl1$^{-/-}$ haematopoietic stem cells showed severe defects in repopulation activities (Capron et al., 2006). Therefore, the SCL erythroid complex may play important roles in controlling specific aspects of erythroid development.

Deleted: ing

Deleted: leading to

Formatted: Superscript

A somewhat more surprising set of targets suggests that the SCL erythroid complex plays more generalised roles in controlling wide programmes of gene expression. Four direct target genes of the whole SCL erythroid complex are involved in regulating chromatin (BRD2, CTCFL, SMARCA5 and EZH2). Such regulators are known to have roles in regulating expression of a wide range of genes in many cell types. BRD2 dimerises with E2F and binds to acetylated histone H4 tails (Nakamura et al., 2007). It has also been shown to bind to the entire length of transcribed genes allowing RNA polymerase II to transcribe through nucleosomes (LeRoy et al., 2008). CTCFL

(BORIS) is a paralogue of the insulator CTCF which shares the same DNA-binding domain as CTCF and is expressed in a mutual exclusive manner to CTCF (Loukinov et al., 2002). These insulator proteins are involved in regulating chromatin domains, and three-dimensional chromatin looping structures, thus ensuring appropriate expression of genes. SMARCA5 associates with RSF1 and is required for chromatin assembly (Loyola et al., 2003). It is also a component of some chromatin-remodelling complexes (Bochar et al., 2000; Poot et al., 2000) (it should also be noted that RSF1 was also considered a target for the complex based on the genes listed in Table 5.2). Expression of SMARCA5 was also shown to be dysregulated in acute myeloid leukaemia (AML) and knockout studies also indicated that SMARCA5 is required for proliferation of haematopoietic progenitors (Stopka et al., 2000; Stopka and Skoultchi, 2003). EZH2 is a histone lysine methyltransferase which methylates histone proteins (Cao et al., 2002). Thus, through regulating chromatin factors, the SCL erythroid complex exerts transcriptional control over a large number of genes through epigenetic reprogramming or chromatin structure. This further emphasises its role as a key regulator of blood development.

5.5.6 Autoregulation of members of the SCL erythroid complex

In Chapter 4, evidence was provided from Affymetrix GeneChip analysis that members of the SCL erythroid complex were involved in regulation of the genes for other members of the complex. The data described in this Chapter further provides evidence of this regulation and that the whole SCL erythroid complex directly regulates expression of the genes of its own members. GATA1 was shown to bind to the promoter of LMO2 (section 5.4.4.3) – this confirms the findings of the Affymetrix GeneChip analysis in the GATA1 siRNA knockdown (although this was not confirmed by the qPCR validation of Affyemtrix expression changes). Furthermore, based on ChIP-on-chip and ChIP-qPCR, SCL and ETO2 were shown to be direct targets of the whole SCL erythroid complex [ETO2 can be a member of this SCL erythroid complex (Goardon et al., 2006)]. Regulation by individual members of the complex and regulation by the complex as a whole provides two levels of regulation - ensuring that the expression level of various members of the complex are tightly regulated in erythroid development. This further highlights the complex regulatory network that controls expression of the SEC.

5.5.7 Limitations of the ChIP-on-chip studies

The ChIP-on-chip assays in this Chapter have been demonstrated to identify DNA elements bound by proteins of interest. Over the last 11 years, since the discovery of the SCL erythroid complex in 1997 by Wadman et al, only three direct target genes (GYPA, c-kit, and α-globin) had been identified. In the study described in this Chapter, 8 additional direct target genes of the SCL erythroid complex were identified. However, there are likely to be many more targets of this

complex which have not been identified here. The limitations of using a transcription factor promoter array in ChIP-on-chip studies are discussed below.

- **Off-promoter binding**

The ChIP-on-chip study in this Chapter focused on an in-house array containing 1 kb array elements of promoters of transcription factors. The promoter sequences were identified in the genome using the FirstEF algorithm. However, if promoters were not accurately identified by FirstEF, the actual promoters would not be represented on the array. This may mean that promoter binding events for some target genes are missed in ChIP-on-chip. Furthermore, transcription factors may bind to other regulatory elements such as enhancers, silencers or distal promoters to regulate transcription. These binding events cannot be detected on the promoter array used in this study. Therefore, the current study only allowed the identification of a subset of genes regulated by the transcription factors in the SCL erythroid complex. One possible solution to this limitation is to increase the coverage of the genome represented on the array. Indeed promoter arrays having coverage of 10 kb around promoter regions are commercially available. Ultimately, the best solution would be to use whole genome tiling arrays which would remove any representation bias and ensure all possible binding events to be detected.

- **Resolution of the array**

In ChIP-on-chip studies of TF binding, the resolution of array elements plays a crucial role for localising the binding sites of transcription factors. The promoter array used in this Chapter has a resolution of 1 kb. Thus, the ChIP-on-chip analyses could only detect binding to the 1 kb fragment but could not identify the precise location of TF binding sites. Higher resolution arrays (using oligonucleotides as array elements) which have a greater coverage around promoters (>5 -10 kb around promoters) would resolve this issue. However, given this limitation, one can use TFBS and comparative sequence analysis to help refine the search for the site of TF binding, and then use ChIP-qPCR to identify and confirm specific interactions at precise locations. This was used extensively in this Chapter.

- **Efficiency of ChIP assays**

The antibodies used in the studies of this Chapter were evaluated by both western blotting and by binding to the +51 enhancer element on the SCL tiling array. Although they were shown to perform well in ChIP-on-chip and pick up high enrichments at the +51 region, some of them performed less well than the others. Particularly, the antibodies for SCL and LMO2 showed the lowest enrichments at the +51 region among all the antibodies and the specificity of LMO2 antibody could not be evaluated on western blotting. The inconsistency of the results obtained for SCL and LMO2 on the

promoter array was evident in the datasets generated by both statistical methods A and B. Furthermore, enrichments were also lower in the ChIP-qPCR for SCL and LMO2 assays. Evaluation of additional antibodies for SCL and LMO2 in ChIP would help resolve these issues and provide more reliable information on the binding profiles of these two TFs.

To circumvent issues with specific antibodies, another possible solution would be to express a tagged protein of the transcription factor under study in the cell line of interest. Some researchers have tried to express an epitope-tagged protein for ChIP studies (Greenbaum and Zhuang, 2002; Lee et al., 2002). Others have tried to co-express the target protein fused to a short biotin acceptor domain together with the biotinylating enzyme BirA from *Escherichia coli* (Viens et al., 2004). The resulting protein-DNA complexes could then be purified by streptavidin affinity. However, whether expression of a tagged protein can completely reflect the native binding patterns of the TF is always an issue for these types of studies.

- **Indirect protein-DNA interaction**

As shown from the data obtained in the ChIP-qPCR studies, enrichments were observed in regions where no TFBSs were found. DNA elements not directly bound by the transcription factor under study could be identified in the ChIP study via indirect protein-DNA interaction. During cross-linking, proteins are cross-linked with any DNA sequence in close proximity. It is possible that transcription factors bound to a primary DNA sequence, which interacts with a secondary DNA sequence by chromatin looping, are cross-linked all together with both DNA elements. As a result, enrichments could be observed in both DNA sequences even though there is no direct binding of the transcription factor with the secondary DNA sequence. Such chromatin looping events have been previously described in the literature. Long range interactions between distal *cis*-elements and the promoter of the α- and β-globin genes were reported to co-ordinate the expression of the genes (Song et al., 2007; Vernimmen et al., 2007). A study of the topoisomerase IIα gene confirmed that the recognition sites of Sp1 and Sp3 transcription factors in the distal and proximal promoters interact with each other via DNA looping (Williams et al., 2007). In fact, from the data obtained in the ChIP-qPCR assays in this Chapter, binding events were confirmed for both E2A and SCL on promoters where only GATA motifs were found (section 5.5.4). To test for such long-range interactions, chromosome conformation capture (3C) could be used (Dekker et al., 2002).

5.6 Conclusions

The work described in this Chapter demonstrated the use of ChIP-on-chip as a robust technique to identify promoters bound by the SCL erythroid complex in erythroid cells. Both published and novel direct target genes were identified. The data obtained in this Chapter provides useful

information for the generation of a transcription network governing aspects of erythroid development which will be described in Chapter 6 of this thesis.

Acevedo, L.G., Iniguez, A.L., Holster, H.L., Zhang, X., Green, R., and Farnham, P.J. (2007). Genome-scale ChIP-chip analysis using 10,000 human cells. BioTechniques *43*, 791-797.
Anderson, K.P., Crable, S.C., and Lingrel, J.B. (1998). Multiple proteins binding to a GATA-E box-GATA motif regulate the erythroid Kruppel-like factor (EKLF) gene. J Biol Chem *273*, 14347-14354.
Anguita, E., Hughes, J., Heyworth, C., Blobel, G.A., Wood, W.G., and Higgs, D.R. (2004). Globin gene activation during haemopoiesis is driven by protein complexes nucleated by GATA-1 and GATA-2. Embo J *23*, 2841-2852.
Attema, J.L., Papathanasiou, P., Forsberg, E.C., Xu, J., Smale, S.T., and Weissman, I.L. (2007). Epigenetic characterization of hematopoietic stem cell differentiation using miniChIP and bisulfite sequencing analysis. Proc Natl Acad Sci U S A *104*, 12371-12376.
Bernstein, B.E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D.K., Huebert, D.J., McMahon, S., Karlsson, E.K., Kulbokas, E.J., 3rd, Gingeras, T.R., *et al.* (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. Cell *120*, 169-181.
Bochar, D.A., Savard, J., Wang, W., Lafleur, D.W., Moore, P., Cote, J., and Shiekhattar, R. (2000). A family of chromatin remodeling factors related to Williams syndrome transcription factor. Proceedings of the National Academy of Sciences of the United States of America *97*, 1038-1043.
Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science (New York, NY *298*, 1039-1043.
Dahl, J.A., and Collas, P. (2007). Q2ChIP, a quick and quantitative chromatin immunoprecipitation assay, unravels epigenetic dynamics of developmentally regulated genes in human carcinoma cells. Stem cells (Dayton, Ohio) *25*, 1037-1046.
Davuluri, R.V., Grosse, I., and Zhang, M.Q. (2001). Computational identification of promoters and first exons in the human genome. Nat Genet *29*, 412-417.
Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. Science *295*, 1306-1311.
Delabesse, E., Ogilvy, S., Chapman, M.A., Piltz, S.G., Gottgens, B., and Green, A.R. (2005). Transcriptional regulation of the SCL locus: identification of an enhancer that targets the primitive erythroid lineage in vivo. Mol Cell Biol *25*, 5215-5225.
Dhami, P., Coffey, A.J., Abbs, S., Vermeesch, J.R., Dumanski, J.P., Woodward, K.J., Andrews, R.M., Langford, C., and Vetrie, D. (2005). Exon array CGH: detection of copy-number changes at the resolution of individual exons in the human genome. Am J Hum Genet *76*, 750-762.
Goardon, N., Lambert, J.A., Rodriguez, P., Nissaire, P., Herblot, S., Thibault, P., Dumenil, D., Strouboulis, J., Romeo, P.H., and Hoang, T. (2006). ETO2 coordinates cellular proliferation and differentiation during erythropoiesis. Embo J *25*, 357-366.
Greenbaum, S., and Zhuang, Y. (2002). Identification of E2A target genes in B lymphocyte development by using a gene tagging-based chromatin immunoprecipitation system. Proc Natl Acad Sci U S A *99*, 15030-15035.
Horak, C.E., Mahajan, M.C., Luscombe, N.M., Gerstein, M., Weissman, S.M., and Snyder, M. (2002). GATA-1 binding sites mapped in the beta-globin locus by using mammalian chIp-chip analysis. Proc Natl Acad Sci U S A *99*, 2924-2929.
Huang, D.Y., Kuo, Y.Y., Lai, J.S., Suzuki, Y., Sugano, S., and Chang, Z.F. (2004). GATA-1 and NF-Y cooperate to mediate erythroid-specific transcription of Gfi-1B gene. Nucleic Acids Res *32*, 3935-3946.
Iyer, V.R., Horak, C.E., Scafe, C.S., Botstein, D., Snyder, M., and Brown, P.O. (2001). Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. Nature *409*, 533-538.
Johnson, K.D., Grass, J.A., Boyer, M.E., Kiekhaefer, C.M., Blobel, G.A., Weiss, M.J., and Bresnick, E.H. (2002). Cooperative activities of hematopoietic regulators recruit RNA polymerase II to a tissue-specific chromatin domain. Proc Natl Acad Sci U S A *99*, 11760-11765.

Koch, C.M., Andrews, R.M., Flicek, P., Dillon, S.C., Karaoz, U., Clelland, G.K., Wilcox, S., Beare, D.M., Fowler, J.C., Couttet, P., *et al.* (2007). The landscape of histone modifications across 1% of the human genome in five human cell lines. Genome Res *17*, 691-707.
Lahlil, R., Lecuyer, E., Herblot, S., and Hoang, T. (2004). SCL assembles a multifactorial complex that determines glycophorin A expression. Mol Cell Biol *24*, 1439-1452.
Lecuyer, E., Herblot, S., Saint-Denis, M., Martin, R., Begley, C.G., Porcher, C., Orkin, S.H., and Hoang, T. (2002). The SCL complex regulates c-kit expression in hematopoietic cells through functional interaction with Sp1. Blood *100*, 2430-2440.
Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.T., Thompson, C.M., Simon, I., *et al.* (2002). Transcriptional regulatory networks in Saccharomyces cerevisiae. Science *298*, 799-804.
LeRoy, G., Rickards, B., and Flint, S.J. (2008). The double bromodomain proteins Brd2 and Brd3 couple histone acetylation to transcription. Molecular cell *30*, 51-60.
Lieu, P.T., Jozsi, P., Gilles, P., and Peterson, T. (2005). Development of a DNA-labeling system for array-based comparative genomic hybridization. J Biomol Tech *16*, 104-111.
Loukinov, D.I., Pugacheva, E., Vatolin, S., Pack, S.D., Moon, H., Chernukhin, I., Mannan, P., Larsson, E., Kanduri, C., Vostrov, A.A., *et al.* (2002). BORIS, a novel male germ-line-specific protein associated with epigenetic reprogramming events, shares the same 11-zinc-finger domain with CTCF, the insulator protein involved in reading imprinting marks in the soma. Proceedings of the National Academy of Sciences of the United States of America *99*, 6806-6811.
Loyola, A., Huang, J.Y., LeRoy, G., Hu, S., Wang, Y.H., Donnelly, R.J., Lane, W.S., Lee, S.C., and Reinberg, D. (2003). Functional analysis of the subunits of the chromatin assembly factor RSF. Molecular and cellular biology *23*, 6759-6768.
Martin, P., and Papayannopoulou, T. (1982). HEL cells: a new human erythroleukemia cell line with spontaneous and induced globin expression. Science *216*, 1233-1235.
Nakamura, Y., Umehara, T., Nakano, K., Jang, M.K., Shirouzu, M., Morita, S., Uda-Tochio, H., Hamana, H., Terada, T., Adachi, N., *et al.* (2007). Crystal structure of the human BRD2 bromodomain: insights into dimerization and recognition of acetylated histone H4. The Journal of biological chemistry *282*, 4193-4201.
Nuez, B., Michalovich, D., Bygrave, A., Ploemacher, R., and Grosveld, F. (1995). Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene. Nature *375*, 316-318.
O'Neill, L.P., and Turner, B.M. (1996). Immunoprecipitation of chromatin. Methods in enzymology *274*, 189-197.
O'Neill, L.P., VerMilyea, M.D., and Turner, B.M. (2006). Epigenetic characterization of the early embryo with a chromatin immunoprecipitation protocol applicable to small cell populations. Nat Genet *38*, 835-841.
Ogilvy, S., Ferreira, R., Piltz, S.G., Bowen, J.M., Gottgens, B., and Green, A.R. (2007). The SCL +40 enhancer targets the midbrain together with primitive and definitive hematopoiesis and is regulated by SCL and GATA proteins. Mol Cell Biol *27*, 7206-7219.
Orlando, V., Strutt, H., and Paro, R. (1997). Analysis of chromatin structure by in vivo formaldehyde cross-linking. Methods *11*, 205-214.
Palomero, T., Odom, D.T., O'Neil, J., Ferrando, A.A., Margolin, A., Neuberg, D.S., Winter, S.S., Larson, R.S., Li, W., Liu, X.S., *et al.* (2006). Transcriptional regulatory networks downstream of TAL1/SCL in T-cell acute lymphoblastic leukemia. Blood *108*, 986-992.
Poot, R.A., Dellaire, G., Hulsmann, B.B., Grimaldi, M.A., Corona, D.F., Becker, P.B., Bickmore, W.A., and Varga-Weisz, P.D. (2000). HuCHRAC, a human ISWI chromatin remodelling complex contains hACF1 and two novel histone-fold proteins. The EMBO journal *19*, 3377-3387.
Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., *et al.* (2000). Genome-wide location and function of DNA binding proteins. Science *290*, 2306-2309.

Rylski, M., Welch, J.J., Chen, Y.Y., Letting, D.L., Diehl, J.A., Chodosh, L.A., Blobel, G.A., and Weiss, M.J. (2003). GATA-1-mediated proliferation arrest during erythroid maturation. Mol Cell Biol *23*, 5031-5042.
Schuh, A.H., Tipping, A.J., Clark, A.J., Hamlett, I., Guyot, B., Iborra, F.J., Rodriguez, P., Strouboulis, J., Enver, T., Vyas, P., *et al.* (2005). ETO-2 associates with SCL in erythroid cells and megakaryocytes and provides repressor functions in erythropoiesis. Mol Cell Biol *25*, 10235-10250.
Song, S.H., Hou, C., and Dean, A. (2007). A Positive Role for NLI/Ldb1 in Long-Range beta-Globin Locus Control Region Function. Molecular cell *28*, 810-822.
Sun, L., Huang, L., Nguyen, P., Bisht, K.S., Bar-Sela, G., Ho, A.S., Bradbury, C.M., Yu, W., Cui, H., Lee, S., *et al.* (2008). DNA methyltransferase 1 and 3B activate BAG-1 expression via recruitment of CTCFL/BORIS and modulation of promoter histone methylation. Cancer research *68*, 2726-2735.
Valverde-Garduno, V., Guyot, B., Anguita, E., Hamlett, I., Porcher, C., and Vyas, P. (2004). Differences in the chromatin structure and cis-element organization of the human and mouse GATA1 loci: implications for cis-element identification. Blood *104*, 3106-3116.
Vernimmen, D., De Gobbi, M., Sloane-Stanley, J.A., Wood, W.G., and Higgs, D.R. (2007). Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. The EMBO journal *26*, 2041-2051.
Viens, A., Mechold, U., Lehrmann, H., Harel-Bellan, A., and Ogryzko, V. (2004). Use of protein biotinylation in vivo for chromatin immunoprecipitation. Analytical biochemistry *325*, 68-76.
Visvader, J., Begley, C.G., and Adams, J.M. (1991). Differential expression of the LYL, SCL and E2A helix-loop-helix genes within the hemopoietic system. Oncogene *6*, 187-194.
Walker, G.T. (1993). Empirical aspects of strand displacement amplification. PCR methods and applications *3*, 1-6.
Welch, J.J., Watts, J.A., Vakoc, C.R., Yao, Y., Wang, H., Hardison, R.C., Blobel, G.A., Chodosh, L.A., and Weiss, M.J. (2004). Global regulation of erythroid gene expression by transcription factor GATA-1. Blood *104*, 3136-3147.
Williams, A.O., Isaacs, R.J., and Stowell, K.M. (2007). Down-regulation of human topoisomerase IIalpha expression correlates with relative amounts of specificity factors Sp1 and Sp3 bound at proximal and distal promoter regions. BMC molecular biology *8*, 36.
Xu, Z., Huang, S., Chang, L.S., Agulnick, A.D., and Brandt, S.J. (2003). Identification of a TAL1 target gene reveals a positive role for the LIM domain-binding protein Ldb1 in erythroid gene expression and differentiation. Mol Cell Biol *23*, 7585-7599.
Zeng, P.Y., Vakoc, C.R., Chen, Z.C., Blobel, G.A., and Berger, S.L. (2006). In vivo dual cross-linking for identification of indirect DNA-associated proteins by chromatin immunoprecipitation. BioTechniques *41*, 694, 696, 698.
Zon, L.I., Youssoufian, H., Mather, C., Lodish, H.F., and Orkin, S.H. (1991). Activation of the erythropoietin receptor promoter by transcription factor GATA-1. Proc Natl Acad Sci U S A *88*, 10638-10641.

Chapter 6

Further characterisation of putative target genes of members of the SCL erythroid complex

6.1 Introduction

The previous Chapters described two experimental approaches, expression profiling of siRNA-induced knockdown and ChIP-on-chip for identifying transcriptional targets of the SCL erythroid complex. Both methods provided different types of information and are complimentary to each other. The gene expression profiling studies provided data on the changes in expression of downstream target genes while the ChIP-on-chip analysis determine the transcription factor binding to *cis*-regulatory elements. Thus, both studies facilitate a better understanding of gene expression controlled by the SCL erythroid complex.

6.1.1 Expression profiling of siRNA-induced knockdown of the SCL erythroid complex

The siRNA-induced knockdown of members of the SCL erythroid complex in combination with expression analyses on a genome-wide scale provided a platform for the study of regulation of target genes. Four main caveats from this study should be re-iterated here:

1. Both direct and indirect (downstream) target genes were identified when each of the five TFs were knocked down using siRNA. Which of these are direct or indirect targets cannot be distinguished purely by expression analysis alone (unless a rigorous time-course Affymetrix GeneChip analysis was performed).
2. The efficiency of knockdown using siRNA has a great impact on the results obtained in the subsequent expression profiling. Ultimately the outcomes of changes of expression are likely to rely on how the removal of a substantial amount of the TF in the nucleus influences its binding to the *cis*-regulatory elements.
3. The overall mode of regulation of target genes can be inferred from Affymetrix expression analysis. Based on changes in expression, it is possible to determine which genes are activated or repressed by the TF. However, complex regulation in transcriptional networks (feedback loops, auto-regulation etc.) may mean the overall effect on expression does not relate directly to the effect that an individual TF is having.

4. Expression profiling alone does not provide information on the regulatory elements and all the protein components required for changes in gene expression to be induced.

6.1.2 ChIP-on-chip study of the SCL erythroid complex

The ChIP-on-chip analyses described in Chapter 5 using antibodies against each TF allowed us to study the interaction between transcription factors and their DNA interacting partners *in vivo*. Unlike expression analysis with Affymetrix GeneChips, it provides direct evidence on the DNA-protein interactions between a TF and its target genes. Three main caveats from this study should be re-iterated here:

1. ChIP-on-chip assays provide no information on whether TF binding events actually activate or repress the transcription of genes and how they induce such regulation.
2. As the ChIP assays can identify any DNA sequences in close proximity to the protein under study, it is not possible to tell whether the TF interacts with the DNA directly, through other proteins or protein complexes in co-operative interactions, or via DNA looping.
3. The array used in Chapter 5 was a transcription factor promoter array. Therefore, the data obtained from these ChIP-on-chip studies can only identify a small portion of genes regulated through their promoters but not mediated through other regulatory elements (unless those elements are in close contact with the promoters).

6.1.3 Auto-regulation of the SCL erythroid complex

From the results obtained in Chapter 4 and 5, three different levels of auto-regulatory pattern of members of the SCL erythroid complex were observed. At the first level, a transcription factor was found to bind to its own promoter and/or enhancer. This was observed for SCL in the ChIP-on-chip study where it bound to its own enhancer and promoter. GATA1 has also been shown to bind to the G1HE, the double GATA site in promoter IE and the intronic enhancer intron-SP *in vivo* (Valverde-Garduno et al., 2004). At the second level of auto-regulation, the whole SCL erythroid complex directly regulates expression of its members. SCL and ETO2, both of which are members of the SCL erythroid complex, were shown to be bound by the whole SCL erythroid complex. At the third level, one member regulated another. From the gene expression profiling study in Chapter 4, E2A was found to be activated by both GATA1 and SCL, LMO2 was activated by E2A and GATA1, and LDB1 was activated by GATA1 (note that such expression changes may be direct or indirect). Furthermore, based on ChIP-on-chip and ChIP-qPCR, GATA1 binds to the promoter of LMO2. These three levels of auto-regulation ensure multiple levels of control over the expression of each member of the complex, and thus they tightly control the expression of their own genes.

Understanding the complexities of this auto-regulation is essential for understanding the transcriptional cascades controlled by the SCL erythroid complex.

6.1.4 Regulation of the LYL1 gene

The ChIP-on-chip analysis identified LYL1 as a direct target gene of the whole SCL erythroid complex. LYL1 is a bHLH protein and has overlapping expression pattern in the erythroid and myeloid lineages and in ascular tissues with SCL in mouse (Visvader et al., 1991). Its expression is initiated slightly later than SCL during haematopoietic specification, beginning during haemangioblast differentiation (Chan et al., 2007). Therefore, LYL1 is thought to be a functional paralogue of SCL (Chapter 1, section 1.4.2.1).

The promoters of LYL1 and SCL have similar structure with two GATA sites located in close proximity. However, no E-box/GATA composite motifs were found in their promoters. SCL, however, has a canonical E-box/GATA composite motif in its +51 enhancer sequence – suggesting that the +51 region may mediate the binding of the SCL erythroid complex. Regulation of the gene may therefore be achieved through the interaction of +51 with its cognate promoter – thus facilitated the detection of the erythroid complex on the promoter using ChIP-on-chip. Considering the coordinated expression pattern of SCL and LYL1, LYL1 is therefore an interesting candidate for studying off-promoter binding of the SCL erythroid complex.

6.1.5 Generation of transcription networks

Ultimately, a more complete understanding of the role of the SCL erythroid complex and its target genes would come from an integration of the datasets presented in this thesis. This would result in a transcriptional regulatory network with the SEC at its core. To aid in the integration of these datasets, a number of computational programmes have been developed for building and visualising gene regulation and expression patterns. These include CellDesigner, Cytoscape and BioTapestry (Longabaugh et al., 2005; Oda et al., 2004; Shannon et al., 2003). All of these softwares simplify the representation of transcription regulatory networks using graphical interfaces and are open source packages freely available to the scientific community. CellDesigner (www.CellDesigner.org) was first described in the modelling of a comprehensive molecular interaction pattern in macrophage based on data found in the published literature (Oda et al., 2004). This software provides a wide range of interaction symbols for the drawing of regulatory networks. It also supports representation of multiple compartments such as the cell membrane and the nucleus. However, in this software, genes are treated as other ‘biomolecules’ (a CellDesigner terminology) and cannot be distinguished from each other. Cytoscape (http://www.cytoscape.org/) is another software for simulating protein-DNA, protein-protein and genetic networks (Shannon et al., 2003).

It allows the end-user to integrate interaction networks with expression data. Users can also customise the properties of interaction symbols. Like CellDesigner, however, DNA sequences are not presented explicitly in Cytoscape. This is a drawback, given that a computational representation that specifically describes transcription factor interactions with the *cis*-regulatory DNA elements is required for modelling transcription networks (Figure 6.1). BioTapestry (www.biotapestry.org) was developed to support this kind of representation of gene regulation (Longabaugh et al., 2005). BioTapestry also allows users to include time-course expression data in an interactive interface in addition to other features that BioTapestry, CellDesigner and Cytoscape all provide.

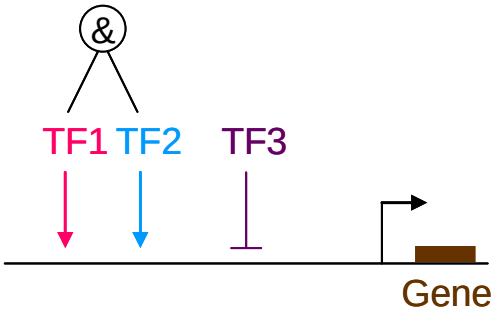


Figure 6.1. A computational representation of interactions between transcription factors and a gene. The black horizontal line represents DNA. The black arrow indicates the transcription start site and the DNA to the left of the arrow represents upstream sequences while the brown box, to the right, represents a gene. TFs 1, 2 and 3 are transcription factors binding to upstream regulatory elements. Coloured arrows indicate activating activities while purple blunt arrow indicated repressing activity. The & symbol shows that TF1 and 2 are both required for activation.

Previous studies have demonstrated the integration of various experimental data for the generation of regulatory networks in haematopoietic development (Chapter 1, section 1.4.2.6) (Swiers et al., 2006). However, the studies performed were based on mouse and a comprehensive study of the SCL erythroid complex was not described. The data obtained in this thesis allowed us to have a more in-depth understanding of the transcription regulatory network of this complex in human haematopoiesis.

6.2 Aims of this chapter

The results obtained in Chapter 4 and Chapter 5 generated different datasets describing different means by which to identify target genes of the SCL erythroid complex. These datasets were compared, evaluated and discussed. Additional experiments were also performed to explain the results obtained from the previous Chapters and further characterise putative target genes of the SCL erythroid complex.

The aims of the work presented in this Chapter were:

1. To compare the putative target genes identified in the expression analyses using Affymetrix GeneChips with the targets identified with ChIP-on-chip and ChIP-qPCR analysis.
2. To study the expression of putative target genes identified in ChIP-on-chip during perturbations of members of the SCL erythroid complex in siRNA-mediated knockdown time-course studies.
3. To further examine aspects of auto-regulation of the SCL erythroid complex observed from the datasets from Chapters 4 and 5. This will be accomplished by looking at the expression level of members of the SCL eythroid complex during perturbations of each members of the SCL erythroid complex in siRNA-mediated knockdown time-course studies.
4. To observe the kinetics of TF binding under the conditions used for Affymetrix analysis. This will be done by characterising the effect of siRNA-mediated GATA1 knockdown on the binding efficiency of GATA1 to promoters.
5. To study the off-promoter binding and regulation of SCL and LYL1 by various TFs in the SCL erythroid complex.
6. To generate a transcriptional regulatory network during erythroid development based on the data obtained in Chapters 4 and 5 and in this Chapter.

6.3 Results

6.3.1 Further characterisation of putative target genes identified in Affymetrix and ChIP-on-chip studies

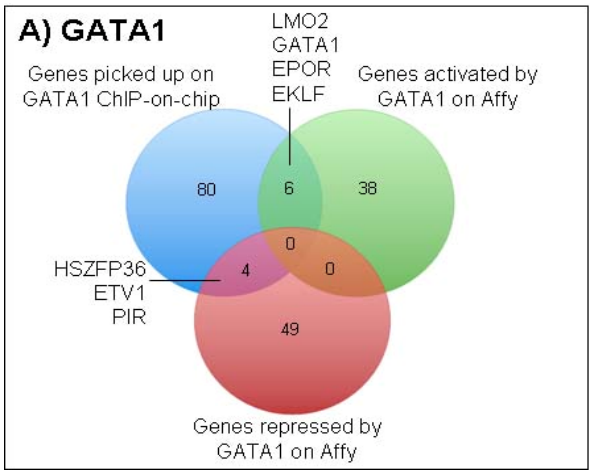
6.3.1.1 Comparison between putative target genes of Affymetrix expression and ChIP-on-chip studies

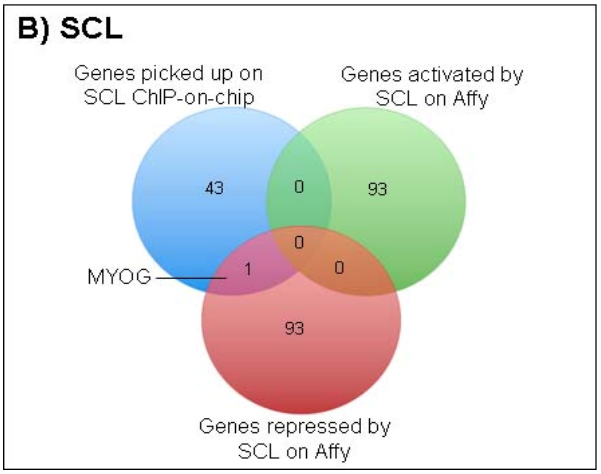
The putative direct target genes of the SCL erythroid complex which were identified in the ChIP-on-chip studies and those validated by ChIP-qPCR in Chapter 5 were compared with the activated or repressed gene lists obtained in the Affymetrix expression studies in Chapter 4. Only the targets for a given TF which were represented on both the promoter array and on the Affymetrix array were used in the comparisons. To increase the level of overlap between the lists, the promoter targets from the ChIP-on-chip study analysed by method B (which included all targets identified by method A plus additional ones) were used in the comparisons. Targets identified by both ChIP-on-chip or ChIP-qPCR, and which showed changes in Affymetrix expression analysis of siRNA-mediated TF knockdowns, were likely to be *bona fide* direct targets for the relevant TF.

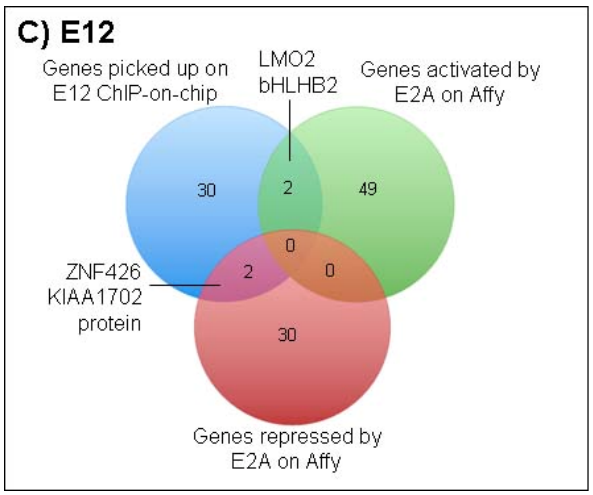
Only a small number of genes were found to be identified in both the Affymetrix expression studies and the ChIP-on-chip studies. In the case of GATA1, 10 target genes were found in both analyses,

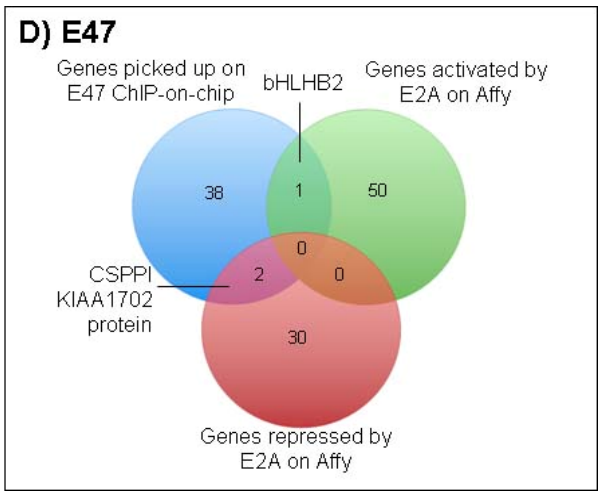
from a total of 90 possible ChIP-on-chip targets and 97 possible Affymetrix targets (Figure 6.2 A). Activation of EKLF by GATA1 was also confirmed in an expression time-course study of siRNA-mediated GATA1 knockdown in K562 cells (Section 6.3.1.2). Only one target gene of SCL (MYOG) was confirmed in both Affymetrix and ChIP-on-chip analyses (Figure 6.2 B). In the case of E2A, ChIP-on-chip targets for either variant E12 or E47 were compared separately against the Affymetrix expression analyses of knockdown by the E2A siRNAs (Figure 6.2 C and D). In total, five targets were identified by both ChIP-on-chip and Affymetrix analysis, out of a total of 34 (E12) or 41 (E47) ChIP-on-chip targets and 81 Affymetrix targets. Two of these targets, bHLHB2 and KIAA1702, were found to be common to both E12 and E47 as ChIP-on-chip targets. No genes were identified as overlapping genes in the ChIP study and the Affymetrix analysis for LDB1 and LMO2 (Figure 6.2 E and F), further suggesting that both of these proteins may be dispensable from the SCL erythroid complex without having consequences at the level of expression (section 4.5.1). The genes co-regulated by SCL, GATA1 and E2A and present on the promoter array were also compared with the putative targets identified in the ChIP-on-chip by SCL, GATA1 or E2A (Figure 6.2 G). No genes were found to be overlapping between these two categories.

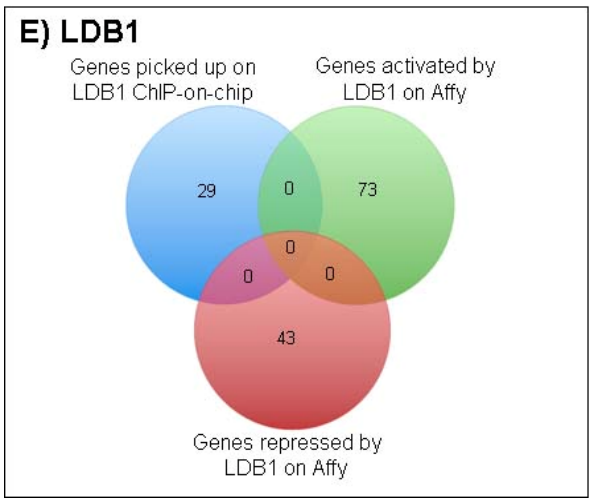
A second way of comparing the data was made by looking at genes identified in any of the Affymetrix GeneChip knockdown experiments with target promoters identified in any of the ChIP-on-chip assays (Figure 6.2 H). This would allow for genes which may have been identified with a particular ChIP-on-chip assay, but not with its corresponding Affymetrix experiment and vise versa, to be identified. Using method B of ChIP-on-chip analyses, 37 target genes were found in both analyses, from a total of 196 possible ChIP-on-chip targets and 331 possible Affymetrix targets. This number of overlapping genes is significantly higher than that of total number of overlapping genes when the five transcription factors were investigated independently (18 target genes). With either analysis, this would suggest that only approximately 5-11% of *bona fide* target genes of members of the SCL erythroid complex, actually change in measureable levels of expression during knockdown. Furthermore, this would argue that the less stringent method B ChIP-on-chip analysis may allow for more direct targets to be identified.

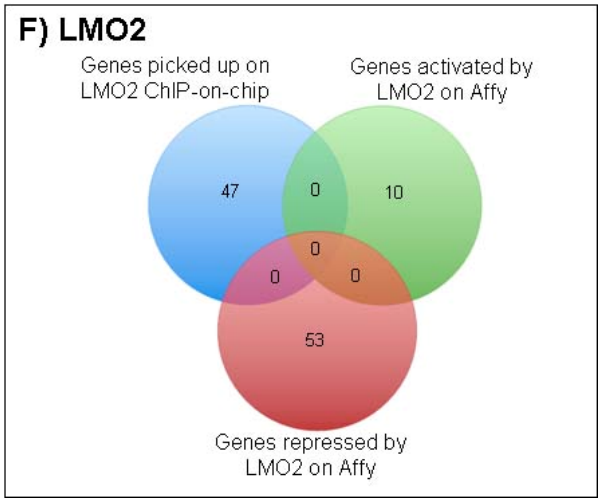

A) GATA1
LMO2
GATA1
EPOR
EKLF
Genes picked up on
GATA1 ChIP-on-chip
Genes activated by
GATA1 on Affy
80
6
38
0
HSZFP36
ETV1
PIR
4
0
49
Genes repressed by
GATA1 on Affy


B) SCL
Genes picked up on
SCL ChIP-on-chip
Genes activated by
SCL on Affy
43
0
93
0
MYOG
1
0
93
Genes repressed by
SCL on Affy


C) E12
Genes picked up on
E12 ChIP-on-chip
LMO2
bHLHB2
Genes activated by
E2A on Affy
30
2
49
0
ZNF426
KIAA1702
protein
2
0
30
Genes repressed by
E2A on Affy


D) E47
Genes picked up on
E47 ChIP-on-chip
bHLHB2
Genes activated by
E2A on Affy
38
1
50
0
CSPPI
KIAA1702
protein
2
0
30
Genes repressed by
E2A on Affy


E) LDB1
Genes picked up on
LDB1 ChIP-on-chip
Genes activated by
LDB1 on Affy
29
0
73
0
0
0
43
Genes repressed by
LDB1 on Affy


F) LMO2
Genes picked up on
LMO2 ChIP-on-chip
Genes activated by
LMO2 on Affy
47
0
10
0
0
0
53
Genes repressed by
LMO2 on Affy

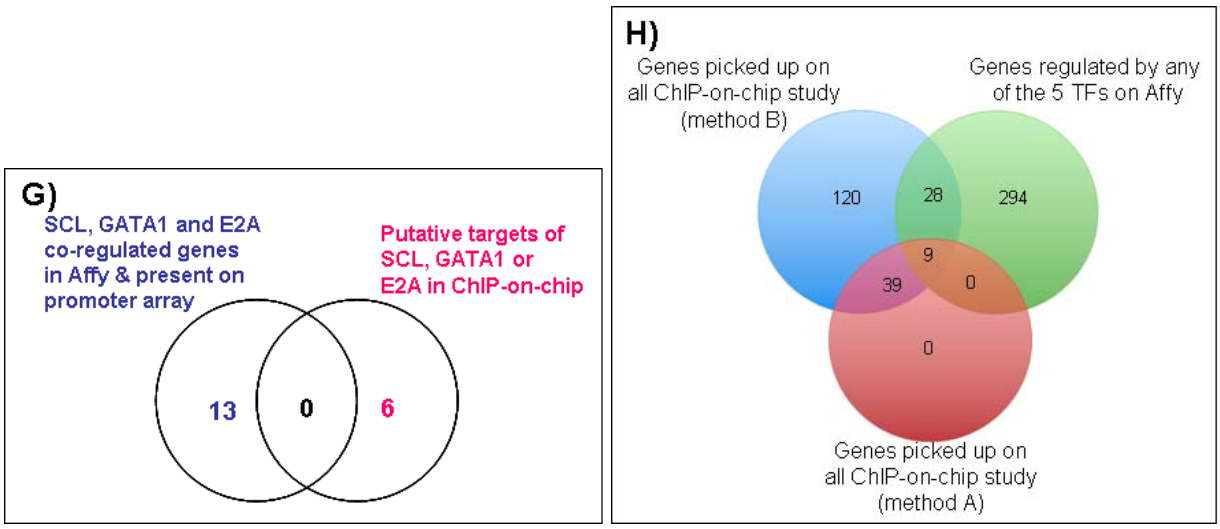


Figure 6.2. Venn diagram comparison of putative target genes identified in Affymetrix expression studies of siRNA-mediated knockdowns and ChIP-on-chip studies of the SCL erythroid complex. Numbers shown in the Venn diagrams are numbers of probes (either Affymetrix probe sets or promoter array elements). Some of the overlapping genes are labelled in the Venn diagram. Panel A: Venn diagram for GATA1; panel B: Venn diagram for SCL; panel C: Venn diagram of E12 ChIP targets compared with E2A siRNA knockdown; panel D: Venn diagram of E47 ChIP targets compared with E2A siRNA knockdown; panel E: Venn diagram of LDB1; panel F: Venn diagram of LMO2; panel G: Venn diagram of SCL, GATA1 and E2A co-regulated genes identified by Affymetrix analysis and the putative target gene promoters identified by either of SCL, GATA1 or E2A in ChIP-on-chip; panel H: Venn diagram of genes regulated by any one of the TF identified by Affymetrix analysis and the putative target gene promoters identified by any one of the five TFs in ChIP-on-chip (analysis method A or B).

6.3.1.2 Expression time-course study of the putative target genes during siRNA knockdown of the SCL erythroid complex

From the analysis described above, there was a low correlation between the Affymetrix data and the ChIP-on-chip/ChIP-qPCR data. There are many possible reasons for this, all of which are discussed in section 6.4.1. One of these reasons, which will be addressed in this section, is that the effect of the knockdown on the expression of the ChIP targets may not be manifested at the time point which was analysed in the Affymetrix analysis (i.e. 24 hours). Thus, it is necessary to examine the effects of siRNA-mediated TF knockdown across a time-course to determine when changes in expression occur for the target genes. To this end, time course studies of up to 48 hours were performed for the knockdown of GATA1, E2A, SCL, LDB1 and LMO2. A range of target genes identified by ChIP-on-chip/ChIP-qPCR were analysed in this expression study - these included targets identified by one TF only, by several, or by all five members of the SCL eythroid complex. The changes in expression of the putative target genes identified by ChIP-on-chip/ChIP-PCR analysis were investigated by quantitative PCR at 12, 24, 36 and 48 hours after knockdown. Except in the case of SCL, where only one siRNA was used, two siRNAs were used for all the other TFs (in a manner similar to that used for the Affymetrix analysis). Two independent biological replicates of each

siRNA were performed for each time-course experiment and qPCR validation of putative target genes (i.e., 4 datapoints). In the qPCR, the normalisation was performed against the luciferase knockdown. The % of mRNA remaining for the putative target genes after siRNA transfection was calculated based on the mean of the four datapoints derived for each target gene (Figure 6.3). The expression level of the TF being knocked down was also included in Figure 6.3 as a positive control for each experiment. Virtually all 14 putative target genes identified in the ChIP-on-chip studies showed changes in expression, to some degree, in the knockdown of all five members of the SCL erythroid complex.

The expression level of four housekeeping genes (β-ACTIN, RPL16, GAPDH and β-TUBULIN) were also monitored in each experiment. The expression of these four housekeeping genes was not expected to change substantially in the time course and therefore they were used as a baseline to determine a significant expression change for the target genes. At each time point, the standard deviation and mean of fold change in the four housekeeping genes in the TF knockdown compared to the luciferase knockdown were calculated. An expression fold change with two standard deviations above or below the mean expression of these four genes at each time point was chosen as the cut-off to determine statistically significant changes in expression in the putative target genes (Table 6.1).

siRNA knockdown		Cut-off in % of mRNA remained after siRNA knockdown			
		12h	24h	36h	48h
GATA1	Down-regulation	54.91	50.35	64.90	65.08
	Up-regulation	182.11	198.61	154.08	153.66
E2A	Down-regulation	67.61	45.63	59.03	79.11
	Up-regulation	147.91	219.16	169.40	126.40
SCL	Down-regulation	42.44	57.81	57.65	58.28
	Up-regulation	235.62	172.99	173.45	171.59
LDB1	Down-regulation	62.98	45.83	63.84	75.57
	Up-regulation	158.79	218.21	156.64	132.34
LMO2	Down-regulation	78.52	60.36	53.01	46.51
	Up-regulation	127.36	165.68	188.65	215.01

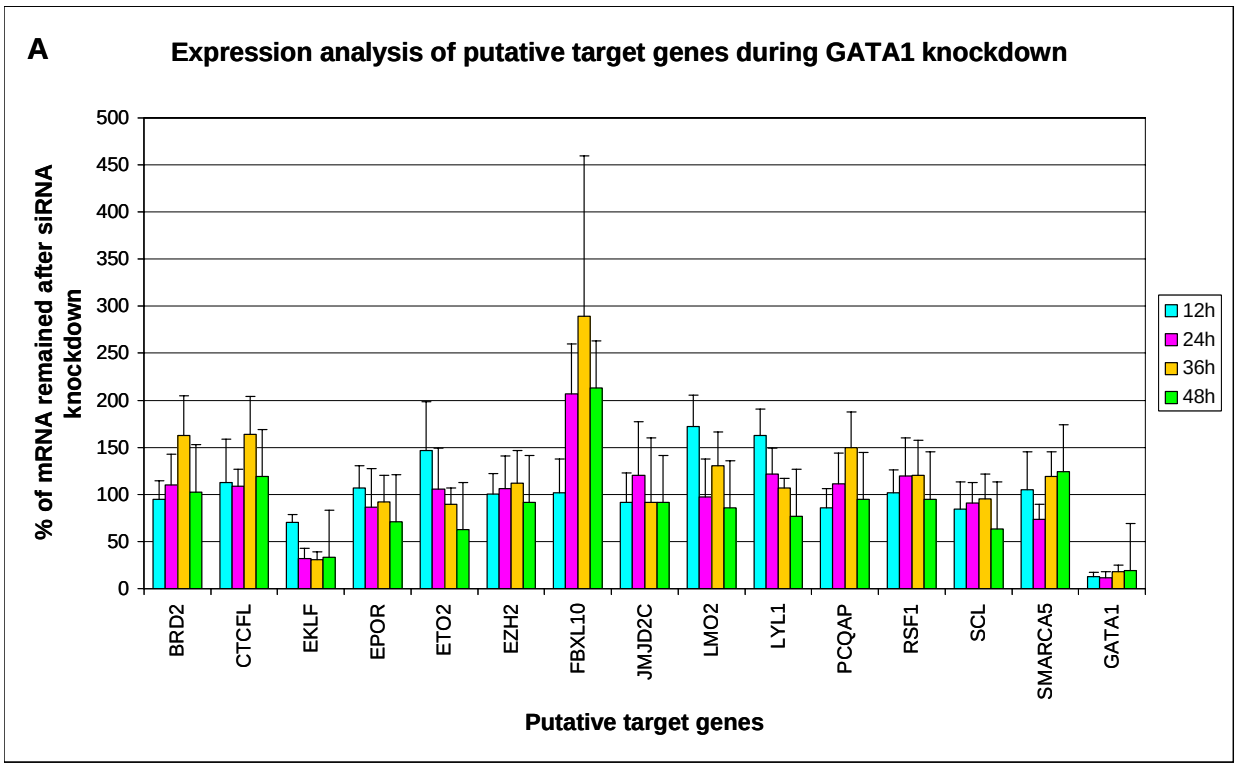
Table 6.1. Cut-off of fold increase or decrease for each siRNA knockdown study. The table shows the percentage of mRNA remained after siRNA knockdown determined by the cut-off calculated for each time point. Both the percentages in mRNA remained for up- or down-regulations were shown.

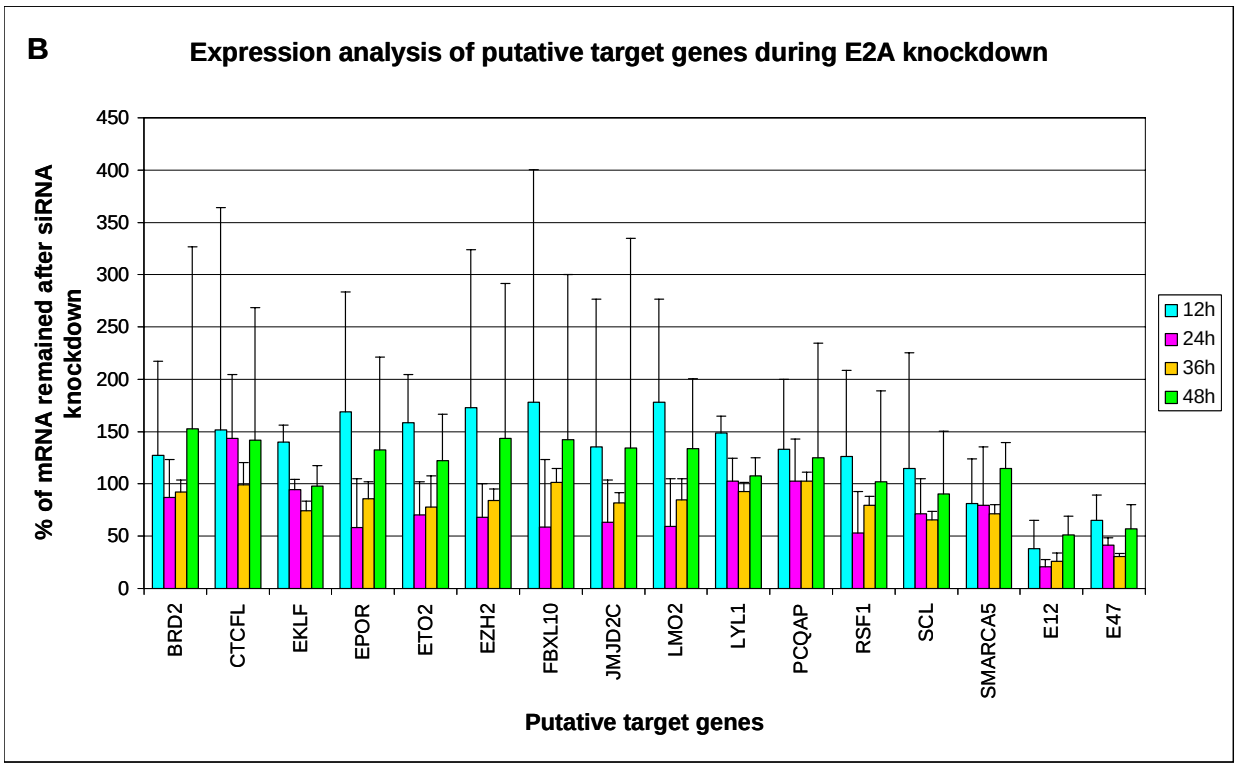
The expression changes of each of the putative target genes in the time-course knockdown study are shown in Table 6.2 and Figure 6.3 and summarised below. At one or more time points in any of the five TF knockdowns, the following interpretations were made:

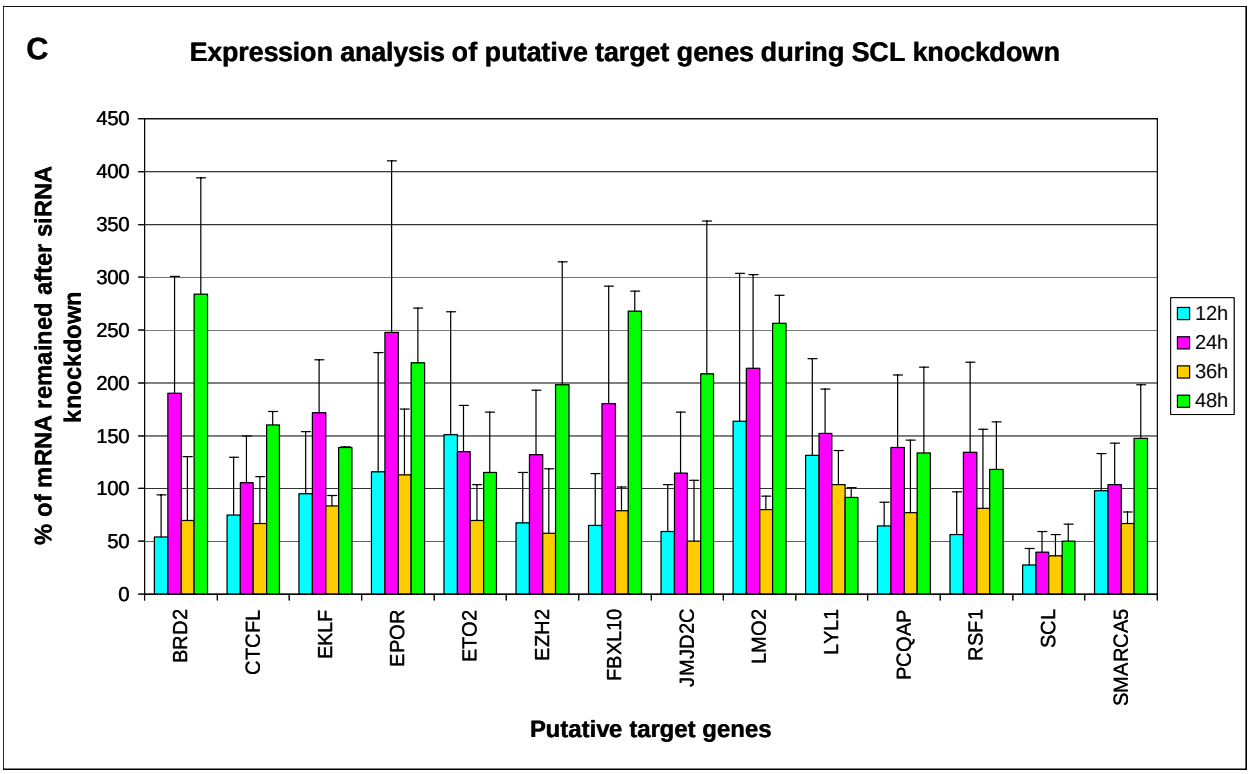
- BRD2 is activated by GATA1, E2A and SCL while repressed by LDB1 and LMO2.

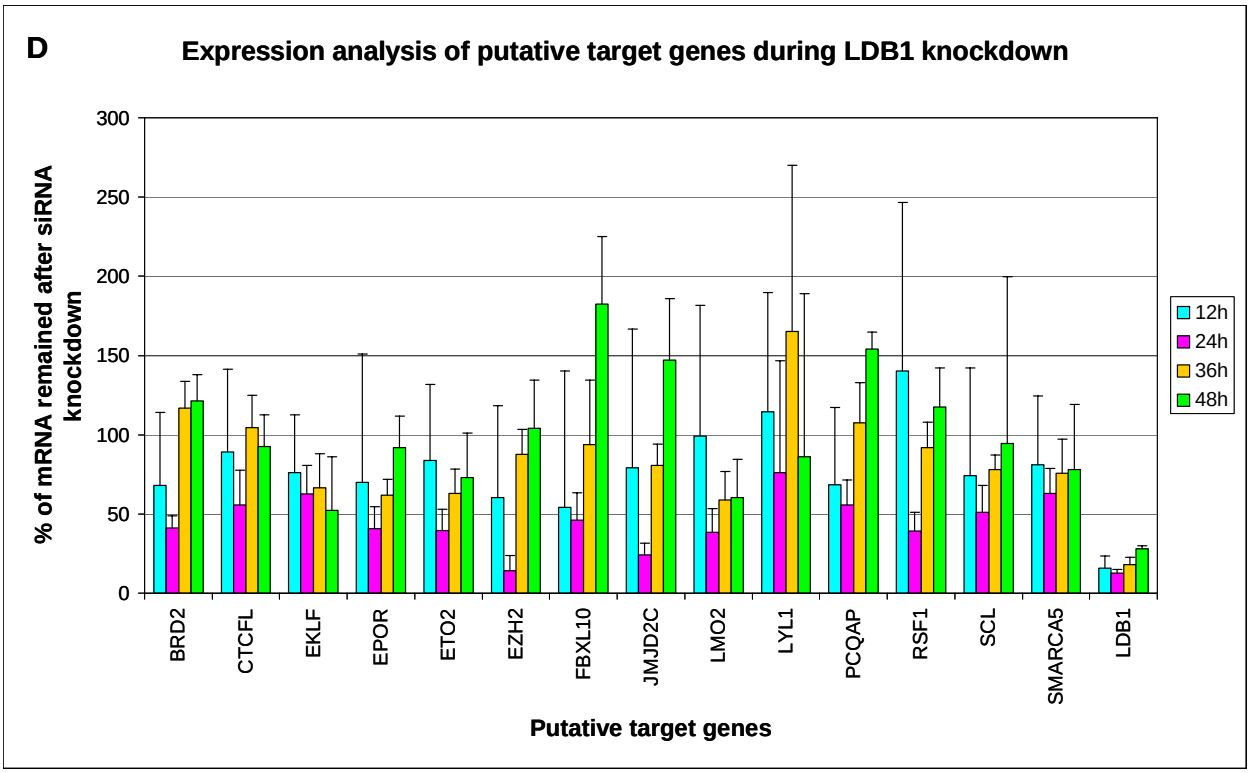
- CTCFL is activated by GATA1 and repressed by E2A.
- EKLF is activated by GATA1 and LDB1.
- EPOR is activated by E2A and SCL and repressed by LDB1.
- ETO2 is activated by GATA1 and LDB1 and repressed by E2A.
- EZH2 is activated by E2A and repressed by LDB1 and LMO2. It was also down-regulated at 36 hour and up-regulated at 48 hour for SCL knockdown. Such fluctuation in expression makes it hard to determine the mode of regulation by SCL.
- FBXL10 is activated by GATA1, E2A and SCL while repressed by LMO2. It was also down-regulated at 12 hour and up-regulated at 48 hour for LDB1 knockdown. Such fluctuation in expression makes it hard to determine the mode of regulation by LDB1.
- JMJD2C is activated by E2A while repressed by LMO2. It was also down-regulated at 36 hour and up-regulated at 48 hour for SCL knockdown and down-regulated at 24 hour and up-regulated at 48 hour LDB1 knockdown. Such fluctuation in expression makes it hard to determine the mode of regulation by SCL and LDB1.
- LMO2 is activated by E2A and SCL while repressed by and LDB1.
- LYL1 is repressed by E2A, LDB1 and LMO2.
- RSF1 is activated by LDB1.
- SCL is activated by GATA1.
- SMARCA5 is activated by LMO2.

Whilst the Affymetrix experiment only detected changes in three genes at the 24 hour time point, the time course qPCR analysis described here detected changes in all 14 target genes for at least one time point in at least one knockdown experiment (63 expression changes in total). Three genes (EZH2, FBXL10 and JMJD2C) showed complex patterns of up- and down-regulation. Furthermore, out of a total of 63 significant expression changes amongst all 14 targets, only 17 of them occur at the time point selected for the Affymetrix experiment (27%), while the rest occur at earlier or later time points. This indicates that the majority of measurable expression changes induced by siRNA knockdown occurred at different time points than the 24 hour time point studied in the Affymetrix GeneChip analyses.


A
Expression analysis of putative target genes during GATA1 knockdown
% of mRNA remained after siRNA knockdown
500
450
400
350
300
250
200
150
100
50
0
12h
24h
36h
48h
BRD2
CTCFL
EKLF
EPOR
ETO2
EZH2
FBXL10
JMJD2C
LMO2
LYL1
PCQAP
RSF1
SCL
SMARCA5
GATA1
Putative target genes


B
Expression analysis of putative target genes during E2A knockdown
% of mRNA remained after siRNA knockdown
450
400
350
300
250
200
150
100
50
0
12h
24h
36h
48h
BRD2
CTCFL
EKLF
EPOR
ETO2
EZH2
FBXL10
JMJD2C
LMO2
LYL1
PCQAP
RSF1
SCL
SMARCA5
E12
E47
Putative target genes


C
Expression analysis of putative target genes during SCL knockdown
% of mRNA remained after siRNA knockdown
450
400
350
300
250
200
150
100
50
0
BRD2
CTCFL
EKLF
EPOR
ETO2
EZH2
FBXL10
JMJD2C
LMO2
LYL1
PCQAP
RSF1
SCL
SMARCA5
Putative target genes
12h
24h
36h
48h


D
Expression analysis of putative target genes during LDB1 knockdown
% of mRNA remained after siRNA knockdown
300
250
200
150
100
50
0
BRD2
CTCFL
EKLF
EPOR
ETO2
EZH2
FBXL10
JMJD2C
LMO2
LYL1
PCQAP
RSF1
SCL
SMARCA5
LDB1
Putative target genes
12h
24h
36h
48h

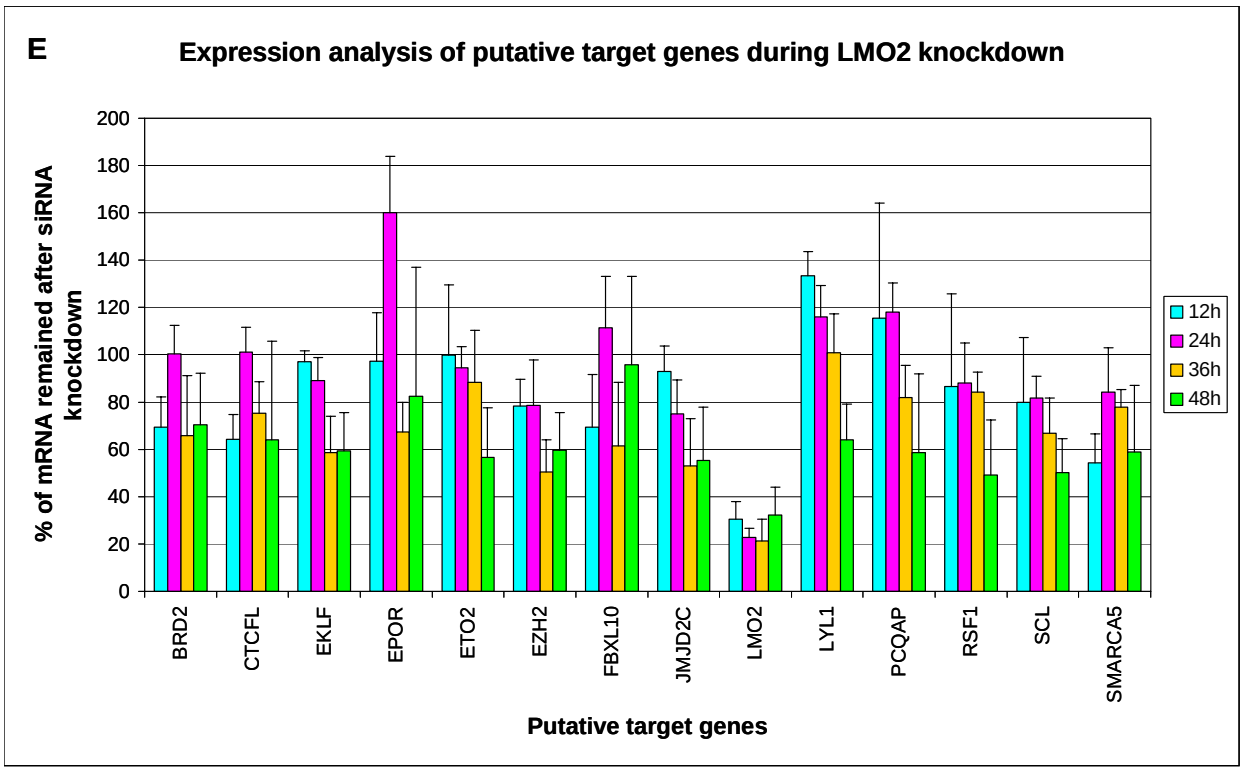


Figure 6.3. Time-course expression analyses of putative target genes during the siRNA knockdown of the SCL erythroid complex. Panel A: GATA1 knockdown; panel B: E2A knockdown; panel C: SCL knockdown; panel D: LDB1 knockdown; panel E: LMO2 knockdown. Y-axis: % of mRNA remaining after siRNA transfection normalised to luciferase siRNA transfection. These percentages are the mean values of 4 datapoints (2 biological replicates for two siRNAs). X-axis: putative target genes. Colour bars indicating the time points are shown on the right of the histograms.

Putative target gene	% of mRNA remained after siRNA knockdown																			
	GATA1 KD				E2A KD				SCL KD				LDB1 KD				LMO2 KD			
	12h	24h	36h	48h	12h	24h	36h	48h	12h	24h	36h	48h	12h	24h	36h	48h	12h	24h	36h	48h
BRD2	95	110	163	103	127	87	92	153	54	190	70	284	68	41	117	122	69	100	66	70
CTCFL	113	109	164	119	152	143	99	142	75	105	67	160	89	56	104	92	64	101	75	64
EKLF	71	32	31	33	140	95	74	98	95	172	83	139	76	63	67	52	97	89	59	60
EPOR	107	87	92	71	169	58	86	132	116	248	113	219	70	41	62	92	97	160	67	83
ETO2	146	106	90	63	158	70	78	122	151	135	70	115	84	40	63	73	100	94	88	57
EZH2	100	107	112	91	173	68	84	144	68	132	58	198	60	14	88	104	78	79	50	60
FBXL10	102	207	289	213	178	59	101	142	65	180	79	268	54	46	94	182	70	111	61	96
JMJD2C	92	120	92	91	135	63	82	134	59	115	50	208	79	24	81	147	93	75	53	55
LMO2	172	97	131	86	178	59	85	133	163	214	80	256	99	39	59	60	31	23	21	32
LYL1	162	122	107	77	149	103	93	108	131	152	104	92	114	76	165	86	133	116	101	64
PCQAP	86	112	150	95	133	103	103	125	64	139	77	134	69	56	107	154	115	118	82	59
RSF1	102	120	120	95	126	53	79	102	56	134	81	118	140	39	92	117	87	88	84	49
SCL	85	91	95	63	115	71	66	90	28	40	36	50	74	51	78	94	80	82	67	50
SMARCA5	105	74	119	124	81	80	71	115	98	104	67	148	81	63	76	78	54	84	78	59

No significant change
Down-regulated below cut-off
Up-regulated above cut-off
TF being knocked down by siRNA

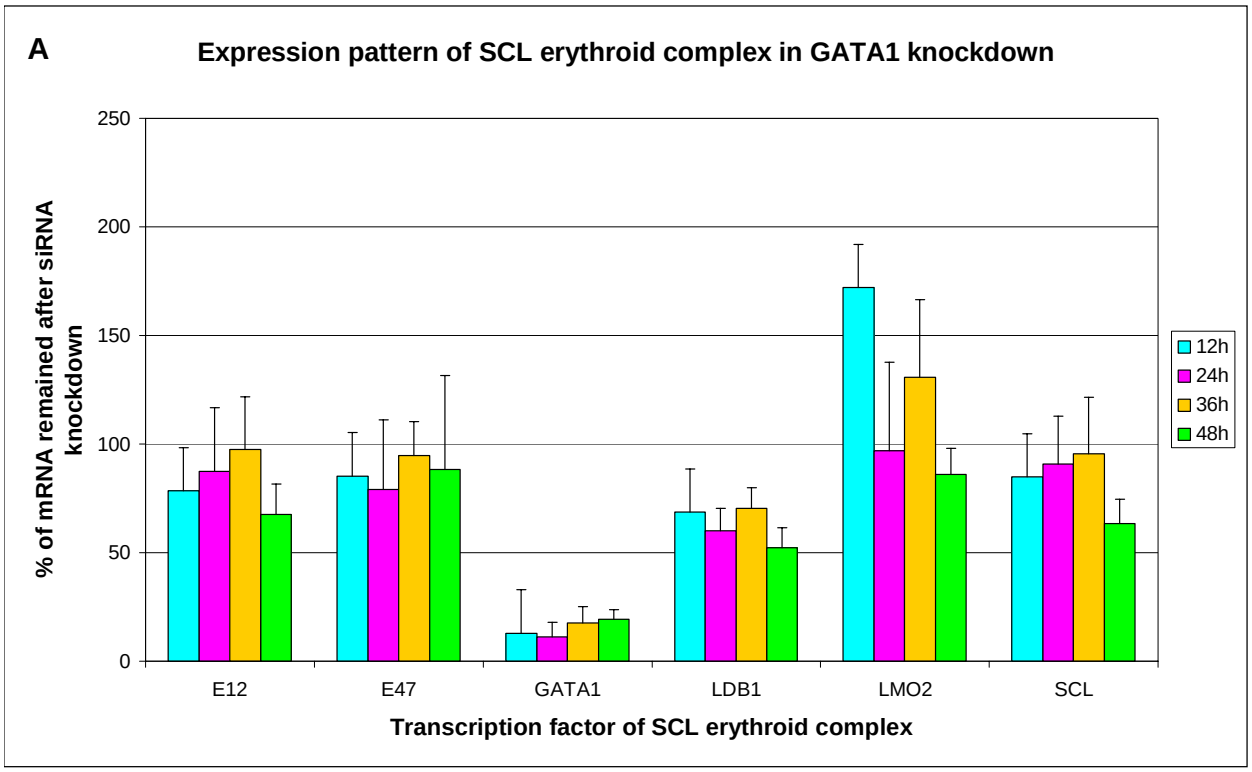
Table 6.2. Putative target genes showing statistically significant change in expression. The percentage of mRNA remained after siRNA knockdown (KD) at 12, 24, 36 and 48 hour time points are shown for each putative target gene. The green boxes indicate down-regulated genes having a change in expression below the cut-off determined in Table 6.2. The red boxes indicate up-regulated genes having a change in expression above the cut-off. The grey boxes the change in expression of the transcription factor being knocked down by siRNA.

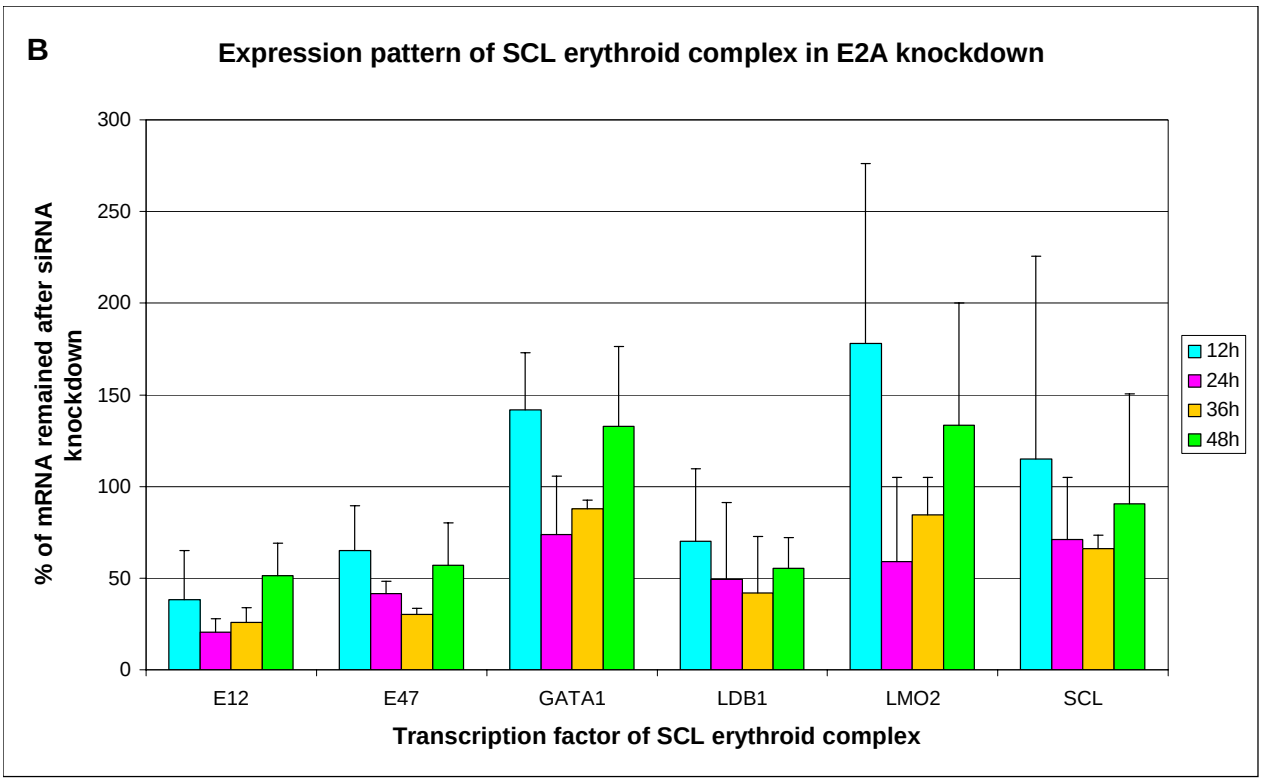
6.3.1.3 Further evidence for auto-regulation of the SCL erythroid complex at the level of gene expression

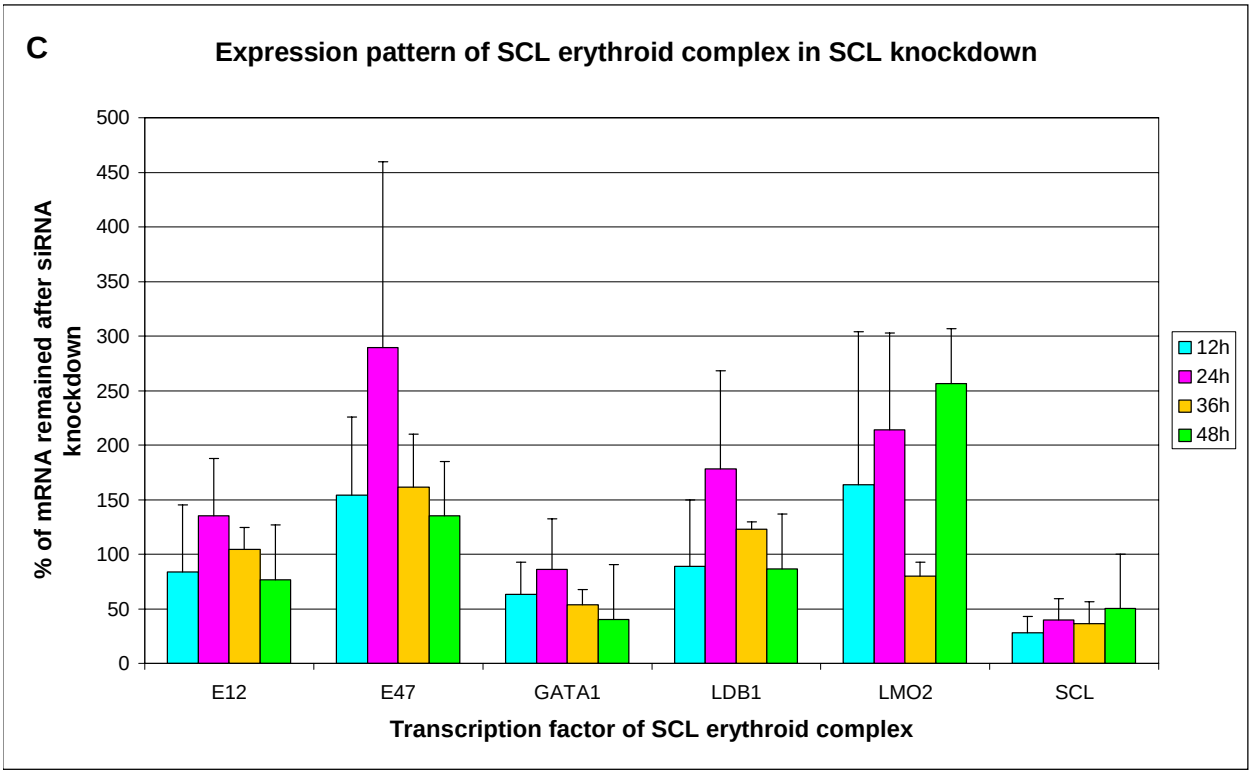
There was evidence presented in Chapters 4 and 5 which suggested that members of the SCL eyrthroid complex were self-regulated by the whole complex, or members therein. To understand this auto-regulation in more detail, the effects of siRNA-mediated knockdown of members of the complex was assessed on the expression levels for each member of the complex. The expression patterns for each member was studied by qPCR in 48 hour time-course experiments as described above for the target gene analysis (Figure 6.4). The expression of other TFs in the complex changed when one of the TFs was being silenced. The same fold change cut-off described in section 6.3.1.2 (Table 6.1) was used to determine statistically significant change in expression of the transcription factors during knockdown (Table 6.3). The data can be summarised as follows:

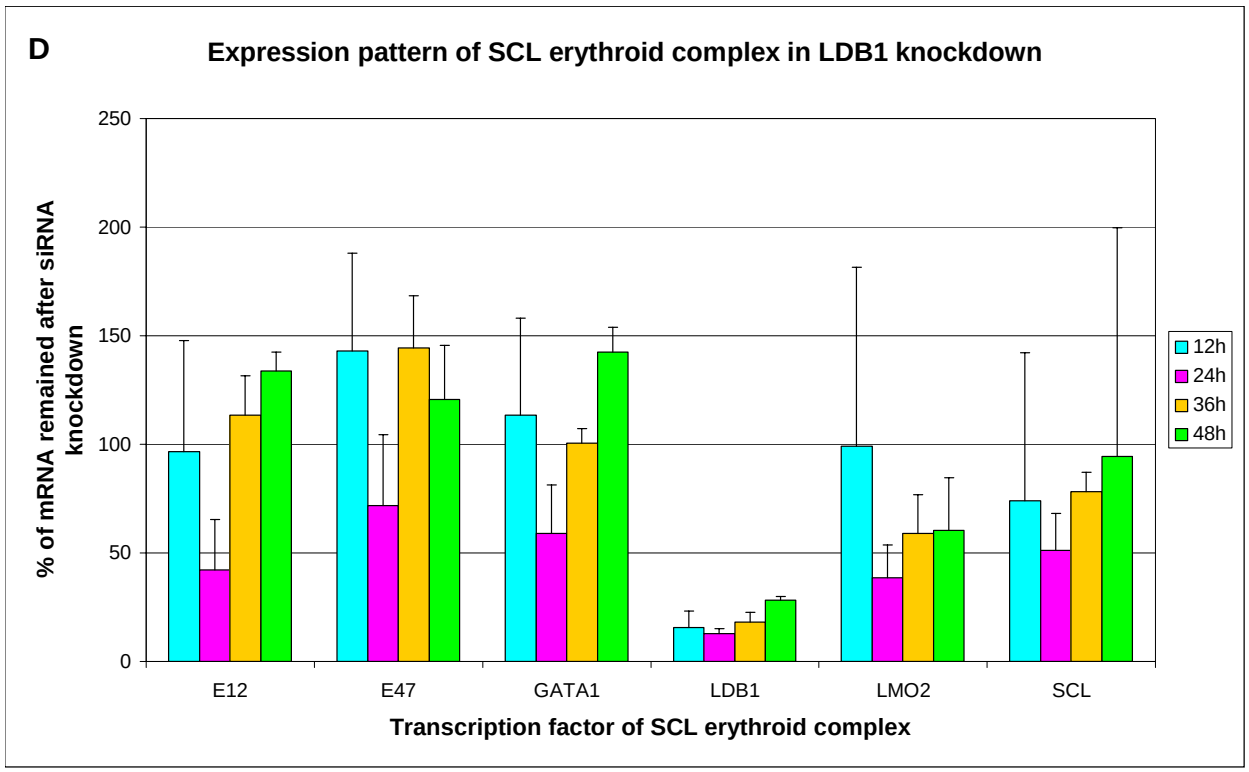
- In the knockdown of GATA1, LDB1 and SCL showed significant down-regulation at the 48 hour time point. Curiously, SCL, a direct target of the whole SCL erythroid complex (Chapter 5) only showed a significant expression change with the knockdown of GATA1 (and none of the other TFs).
- In the knockdown of E2A, significant up-regulation was observed for GATA1 at the 48 hour time point and for LMO2 at the 12 and 48 hour time points. Down-regulation was observed in LDB1 at the 36 and 48 hour time point.
- In the knockdown of SCL, significant up-regulation was also observed for E47 and LDB1 at the 24 hour time point and for LMO2 at the 24 and 48 hour time point. Down-regulation was observed in GATA1 at the 36 and 48 hour time point.
- In the knockdown of LDB1, significant up-regulation was also observed for GATA1 at the 48 hour time point. Expression of E12 was shown to be significantly down-regulated at 24 hour and up-regulated at 48 hour. Down-regulation was observed in LMO2 at the 24, 36 and 48 hour time point.
- In the knockdown of LMO2, significant down-regulation was also observed for E47 at the 12 hour time point.

The results shown here demonstrate that the knockdown of each TF in the complex, affects the expression of other members of the complex in a variety of ways involving both up and down regulation and combinations of both through time. This data further delineates the various modes of auto-regulation which are involved in modulating levels of each member of the TF complex. Such effects would compound the issue of identifying target genes for each member of the complex using knockdown analysis, since expression changes associated with the changing levels of other TFs in the complex, would also be reflected in the final Affymetrix GeneChip analyses.




B
Expression pattern of SCL erythroid complex in E2A knockdown
% of mRNA remained after siRNA knockdown
300
250
200
150
100
50
0
E12
E47
GATA1
LDB1
LMO2
SCL
Transcription factor of SCL erythroid complex
12h
24h
36h
48h


C
Expression pattern of SCL erythroid complex in SCL knockdown
% of mRNA remained after siRNA knockdown
500
450
400
350
300
250
200
150
100
50
0
E12
E47
GATA1
LDB1
LMO2
SCL
Transcription factor of SCL erythroid complex
12h
24h
36h
48h



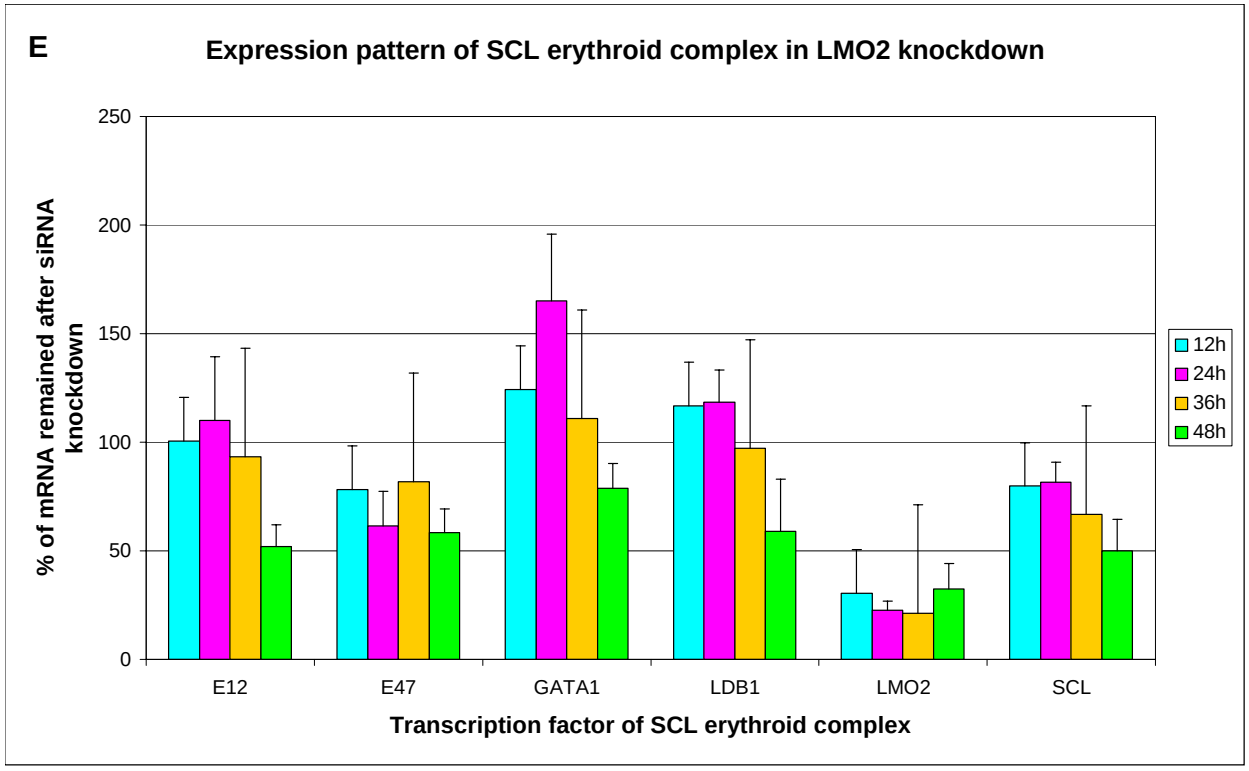


Figure 6.4. Time-course expression analyses for each of the five members of the SCL erythroid complex during siRNA knockdown of members of the complex. Y-axis: % of mRNA remaining after siRNA transfection normalised to luciferase knockdown. These percentages were the mean values of 4 datapoints from two biological replicates of 2 siRNAs used per TF. X-axis: transcription factor of SCL erythroid complex. Colour bars indicating the time points are shown on the right of the histograms.

Transcription factor of SCL erythroid	% of mRNA remained after siRNA knockdown																			
	GATA1 KD				E2A KD				SCL KD				LDB1 KD				LMO2 KD			
	12h	24h	36h	48h	12h	24h	36h	48h	12h	24h	36h	48h	12h	24h	36h	48h	12h	24h	36h	48h
E12	78	88	98	68	38	21	26	51	84	135	105	77	97	42	113	134	101	110	93	52
E47	85	79	95	88	65	42	30	57	154	289	162	135	143	72	144	121	78	61	82	58
GATA1	13	11	18	19	142	74	88	133	63	86	54	40	113	59	100	142	124	165	111	79
LDB1	69	60	70	52	70	49	42	55	89	178	123	87	16	13	18	28	117	118	97	59
LMO2	172	97	131	86	178	59	85	133	163	214	80	256	99	39	59	60	31	23	21	32
SCL	85	91	95	63	115	71	66	90	28	40	36	50	74	51	78	94	80	82	67	50

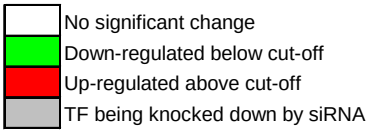


Table 6.3. Members of the SCL erythroid complex showing statistically significant change in expression. The percentage of mRNA remained after siRNA knockdown (KD) at 12, 24, 36 and 48 hour time points are shown for each of the five transcription factor. The green boxes indicate down-regulated genes having a change in expression below the cut-off determined in Table 6.1. The red boxes indicate up-regulated genes having a change in expression above the cut-off. The grey boxes the change in expression of the transcription factor being knocked down by siRNA.

6.3.1.4 ChIP-on-chip study of GATA1 knockdown

Data described in section 6.3.1.1 demonstrated that the TF binding events observed in ChIP-on-chip studies did not correlate, for the most part, with changes in expression of target genes after siRNA-induced knockdown. This may be due to differences in the experimental set-ups between the Affymetrix and ChIP-on-chip studies in combination with biological reasons. It is important to understand why these differences occurred in order to provide confidence that data derived from both experimental approaches was biologically meaningful. Experiments were designed to address these issues as described below.

The ways in which the ChIP-on-chip studies and Affymetrix experiments were conducted in this project were inherently different. For ChIP-on-chip, wild type K562 cells were used to identify targets. For Affymetrix analysis, knockdown samples were used. During knockdown, the kinetics of TF clearance from binding sites on target gene promoters during TF knockdown may not be the same for all targets – some may be removed from targets more rapidly than others. Thus, the effect of TF clearance on transcription may be different for these targets. Furthermore, TF clearance kinetics may not be directly inferred from the knockdown of the protein levels of the TFs themselves – thus, the time at which the majority of the clearance has occurred may not correspond to when the maximal knockdown of the protein was observed. Furthermore, the experimental manipulation of the knockdown cells (i.e., transfection) may affect TF binding events, further complicating the issues.

To provide some clues to why these discrepancies occurred between Affymetrix and ChIP-on-chip data, the effects of siRNA-mediated TF knockdown on the binding of TFs to promoters on the array

were monitored. This was performed for only one of the TFs - GATA1. A ChIP-on-chip study using the TF promoter array was performed for GATA1 during siRNA-mediated GATA1 knockdown at the 24 hour time point, consistent with the experimental set-up used for the Affymetrix analysis. The luciferase, GATA1a and GATA1b siRNAs were transfected into K562 cells by electroporation. After 24 hours, protein, total RNA and chromatin were extracted from the K562 cells. The knockdown of GATA1 was confirmed by quantitative PCR and western blotting for three independent biological replicates (Appendix 5). ChIP-on-chip was performed as previously described for the wild type K562 analysis.

Since the SCL locus was used as a positive control for the ChIP-on-chip studies described in Chapter 5 (and is also a key target of the entire SCL erythroid complex), the binding of GATA1 to various regulatory regions of the SCL locus during GATA1 knockdown was studied initially (Figure 6.5). The profiles of GATA1 binding after luciferase knockdown, GATA1a knockdown and GATA1b knockdown were shown to be very similar when compared to wild type K562 cells with little evidence for substantial loss of the GATA1 protein from all of the regulatory regions of SCL. In the study of GATA1 binding in wild type cells, four regions were shown to be significantly enriched: the -9/-10 enhancer, the +3 enhancer, promoter 1a and the +51 erythroid enhancer (labelled in Figure 6.5). The GATA1 enrichments of these four regions were compared in the luciferase control knockdown against the GATA1a and GATA1b knockdown (Table 6.4). The enrichment at the +51 enhancer did not change substantially after the knockdown of GATA1 whereas the enrichments for -9/-10 enhancer and promoter 1a were reduced by 15% and 24% respectively. The change for the +3 enhancer was the greatest with the enrichment decreasing by approximately 41% after siRNA-induced knockdown of GATA1.

Surprisingly, differences in fold enrichments were also observed between the wild type K562 cells and the luciferase siRNA transfected cells (Table 6.4). Fold enrichments at the +3 and +51 SCL enhancers increased by approximately 25% in the luciferase siRNA transfected cells and decreased by 11% and 14% at the -9/-10 enhancer and promoter 1a respectively.

This analysis of the SCL locus provided some initial evidence that GATA1 clearance from its binding sites does not necessarily reflect the degree of protein knockdown for GATA1 (more than 90% at the 24 hour timepoint), and that electroporation may also affect binding of GATA1 – at least at the SCL locus.

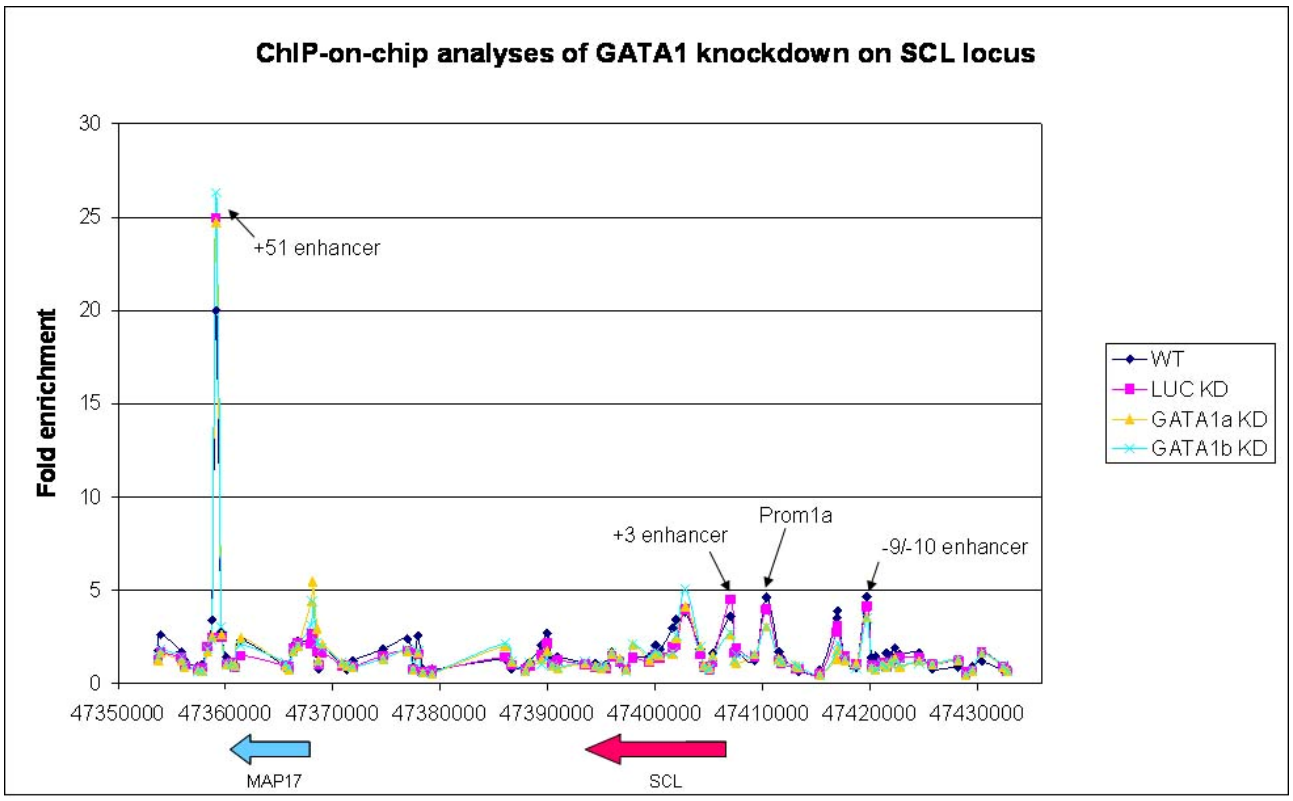


Figure 6.5. GATA1 ChIP-on-chip profile during GATA1 siRNA-mediated knockdown across the SCL locus in K562 cells. The ChIP-on-chip profiles across the SCL locus for wild type cells, luciferase, GATA1a and GATA1b siRNA transfected cells are shown. The x-axis shows the genomic coordinates across the SCL tiling path and the y-axis shows the fold enrichments. The thick-coloured arrows at the bottom of the figure show the position of the genes included on the SCL tiling path and their direction of transcription. Blue curve: profile for wild type cells; pink curve: profile for the luciferase knockdown (KD); yellow curve: profile for the GATA1a knockdown and aqua curve: profile for the GATA1b knockdown. Enhancers or promoters which showed significant enrichments are labelled by black arrows on the graph. The fold enrichments for each region were the mean of three independent biological replicates.

Regulatory elements on SCL locus	Fold enrichment				% change in enrichments in LUC KD against WT	% change in enrichments in GATA1 KD against LUC KD		
	Wild type	LUC KD	GATA1a KD	GATA1b KD	LUC vs WT	GATA1a KD	GATA1b KD	Average
+51 enhancer	20.00	24.98	24.72	26.33	24.9	-1.04	5.40	2.18
+3 enhancer	3.60	4.49	2.61	2.72	24.72	-41.87	-39.42	-40.65
Promoter 1a	4.61	3.97	3.03	3.03	-13.88	-23.68	-23.68	-23.68
-9/-10 enhancer	4.63	4.11	3.50	3.48	-11.23	-14.84	-15.33	-15.09

Table 6.4. Comparison of fold enrichments of enhancers and promoters of SCL in luciferase, GATA1a and GATA1b siRNA transfected K562 cells and wild type cells. This table shows the fold enrichment of the promoters and enhancers of SCL in luciferase, GATA1a and GATA1b siRNA transfected K562 cells and wild type cells and the percentage change for each regulatory element.

The ChIP-on-chip enrichments for the promoters of the putative target genes selected from the ChIP-on-chip studies in Chapter 5 were also investigated in the GATA1 knockdown condition (Figure 6.6 and Table 6.5). For thirteen of the target promoters, there was a reduction in the fold

enrichment in the GATA1 knockdown compared to the luciferase control knockdown with a percentage reduction of 8% to 44% for 13 of the targets - even though the overall protein level of GATA1 was reduced by 85-90%. In contrast, the fold enrichments increased for LYL1 by 13% (Figure 6.7). Differences in fold enrichment were again observed between the ChIP-on-chip performed in wild type K562 cells and the luciferase siRNA transfected cells. 9 out of the 14 promoters studied show decreases in GATA1 binding of up to 39% after luciferase siRNA transfection while the other 5 promoters showed increases of up to 68%. This further confirmed that electroporation with siRNAs may also affect binding of GATA1 to promoters.

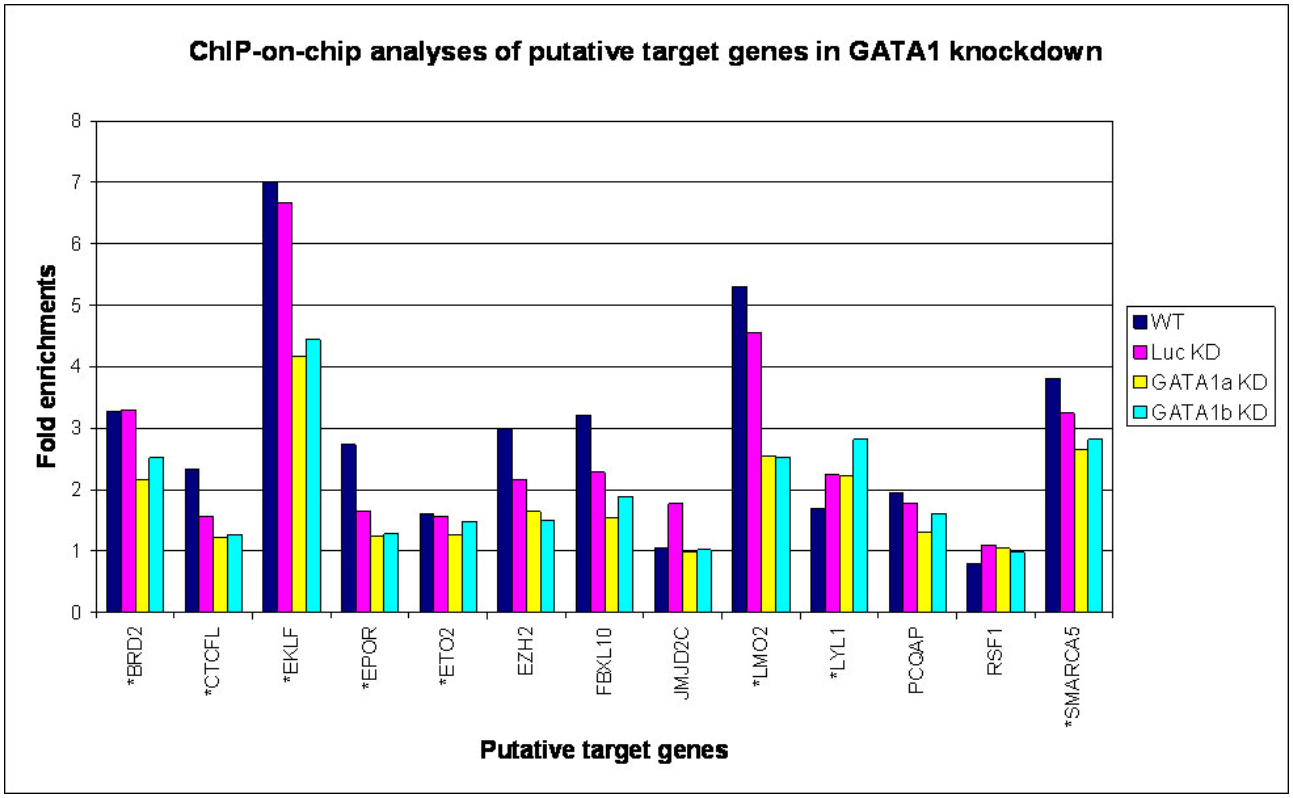


Figure 6.6. ChIP-on-chip analyses of GATA1 binding at target gene promoter during knockdown of GATA1. Histogram showed the fold enrichments for putative target genes in ChIP-on-chip studies in wild type K562 cells, the luciferase siRNA knockdown, the GATA1a siRNA knockdown and the GATA1b siRNA knockdown. Y-axis: fold enrichments. X-axis: putative target gene promoters. The ChIP-on-chip assays represented by the colour bars are shown in the key on the right. The fold enrichments for each target promoter were the averages of three independent biological replicates. The asterisk indicated genes with significant enrichments in the ChIP-on-chip (analysed by both methods A and B) and ChIP-qPCR studies of GATA1 in wild type cells in Chapter 5.

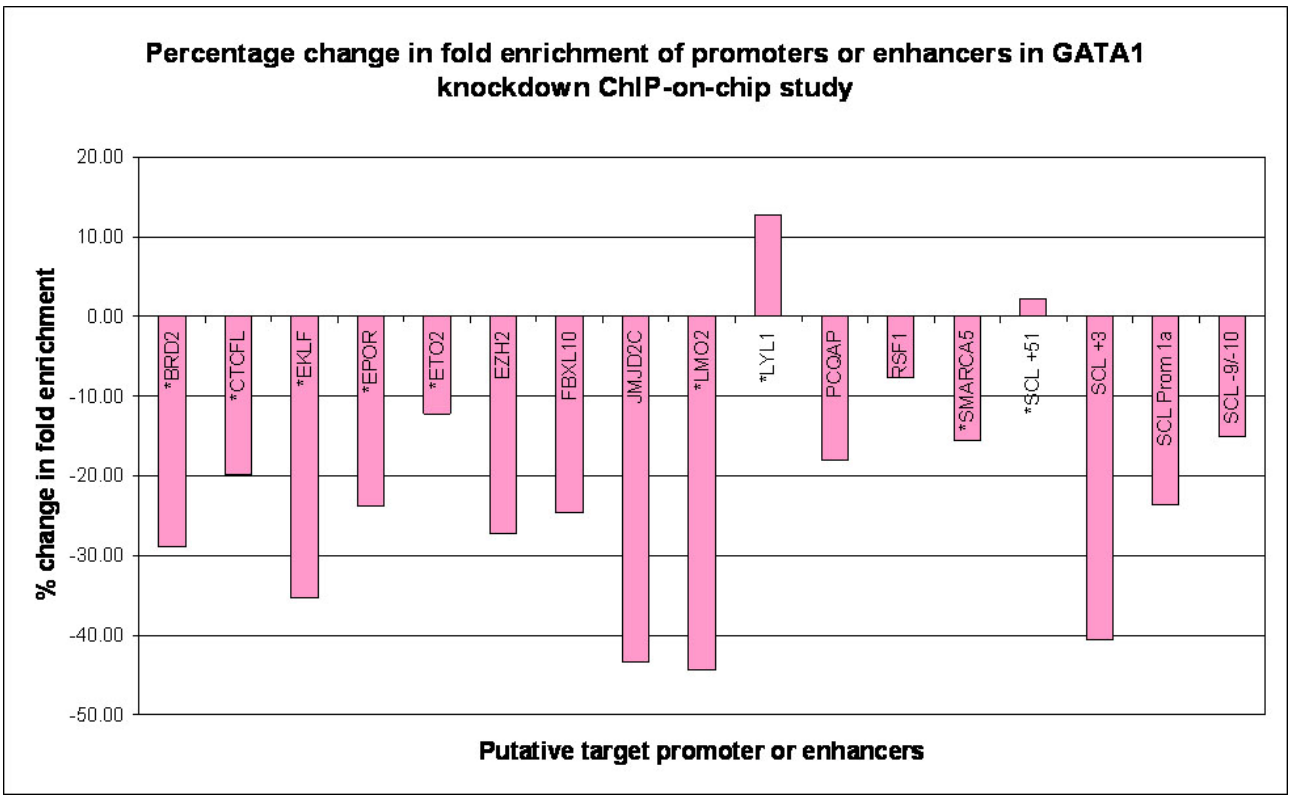


Figure 6.7. Percentage change in fold enrichment of promoters and enhancers in GATA1 knockdown ChIP-on-chip study. Histogram showed the percentage change in fold enrichments for putative target genes and the SCL promoter and enhancers in ChIP-on-chip studies in GATA1 siRNA knockdown compared to luciferase knockdown. Y-axis: percentage change in fold enrichments. X-axis: putative target gene promoters or SCL promoter and enhancers (last four bars). The asterisk indicated promoters or enhancers with significant enrichments in the ChIP-on-chip (analysed by both methods A and B) and ChIP-qPCR studies of GATA1 in wild type cells in Chapter 5.

Putative target gene	Fold enrichments				% change in enrichments in LUC KD against WT	% change in enrichments in GATA1 KD against LUC KD			% of mRNA remained after GATA1 KD at 24 hour
	WT	Luc KD	GATA1a KD	GATA1b KD	LUC vs WT	GATA1a KD	GATA1b KD	Average	Average
*BRD2	3.28	3.29	2.16	2.52	0.32	-34.47	-23.38	-28.93	110.31
*CTCFL	2.33	1.56	1.23	1.26	-33.30	-20.93	-18.81	-19.87	108.90
*EKLF	7.01	6.66	4.17	4.44	-5.02	-37.37	-33.42	-35.40	32.24
*EPOR	2.73	1.66	1.24	1.28	-39.36	-25.02	-22.65	-23.84	86.70
*ETO2	1.62	1.56	1.27	1.47	-3.41	-18.63	-5.76	-12.20	105.79
EZH2	2.99	2.16	1.64	1.50	-27.77	-23.80	-30.54	-27.17	106.56
FBXL10	3.21	2.28	1.55	1.89	-28.92	-31.95	-17.24	-24.59	206.71
JMJD2C	1.05	1.77	0.98	1.02	68.25	-44.47	-42.41	-43.44	120.38
*LMO2	5.31	4.55	2.54	2.53	-14.16	-44.18	-44.56	-44.37	97.02
*LYL1	1.69	2.24	2.22	2.82	32.30	-0.69	26.11	12.71	121.72
PCQAP	1.94	1.78	1.31	1.60	-8.36	-26.22	-10.02	-18.12	111.60
RSF1	0.81	1.10	1.05	0.98	36.13	-4.59	-10.68	-7.63	119.54
*SMARCA5	3.81	3.24	2.65	2.82	-14.95	-18.30	-12.98	-15.64	73.62

Table 6.5. Comparison of fold enrichments of putative target promoters in luciferase, GATA1a and GATA1b siRNA transfected K562 cells and wild type cells. This table shows the fold enrichment of the putative target promoters in luciferase, GATA1a and GATA1b siRNA transfected K562 cells and in wild type cells. The percentage

change in enrichment between the luciferase and GATA1 siRNAs is shown in the fifth column. The percentage of mRNA remained after GATA1 KD at the 24 hour time point from the time-course study in section 6.3.1.2 is shown in the last column. The asterisk indicated genes with significant enrichments in the ChIP-on-chip (analysed by both methods A and B) and ChIP-qPCR studies of GATA1 in wild type cells in Chapter 5.

The percentage clearance of the GATA1 protein from the promoters after GATA1 knockdown was also compared to the expression changes for these genes which were obtained from the qPCR time course studies at the 24 hour time point (section 6.3.1.2) (Table 6.5). For genes which showed the greatest percentage of GATA1 clearance (i.e. LMO2 and JMJD2C), the change in expression was not significant. However, for EKLF, the percentage clearance was 34% and there was a 70% reduction in mRNA level. This suggests that the effect, at the level of expression, of clearance of GATA1 from the promoters of target genes may be different for each target gene. For this reason, the knockdown of GATA1 would have different effects on each of its target genes – and only some may demonstrate measurable expression changes in the knockdown condition. These data again help resolve issues which relate to the inability to detect expression changes for many of the targets of the SCL erythroid complex at the 24 hour time point.

In summary, the results presented here suggest that differential rates of GATA1 clearance from target promoters in knockdown experiments (resulting from effects of GATA1 knockdown and electroporation of siRNAs) may confound attempts to identify measurable expression changes in both qPCR and Affymetrix GeneChip analysis.

6.3.1.5 Off-promoter regulation of SCL and LYL1

The ChIP-on-chip studies of Chapter 5 identified LYL1 as a target gene of four of the five members of the SCL erythroid complex – and is therefore likely to be a good candidate to be regulated by the entire complex. Expression analysis by qPCR confirmed that knockdown of each member of the complex affected the expression of LYL1, further supporting a role for the whole complex in its regulation. However, its promoter does not have an E-box/GATA composite motif which would support the role of the whole SCL erythroid complex involved in its regulation. The absence of an E-box/GATA site in its promoter is analogous to the situation found for SCL – which is thought to be a functional and structural paralogue of LYL1 at both the protein and DNA level (Chapter 1, section 1.4.2.1). SCL, however, has a canonical E-box/GATA composite motif in its +51 enhancer sequence – suggesting that the +51 region may mediate the binding of the SCL erythroid complex. Regulation of the gene may therefore be achieved through the interaction of +51 with its cognate promoter – thus facilitated the detection of the erythroid complex on the promoter using ChIP-on-chip. If such a similar situation were also true for LYL1 (and the paralogy between the two genes extended to regulation), one would expect that a downstream enhancer of LYL1 would contain an

E-box/GATA composite site. This site would therefore also bind the SCL eythroid complex and this would be detected by ChIP.

To elucidate whether this type of regulation occurs outside the promoter region of LYL1, TFBS search by TESS and TFSearch together with comparative genomic analyses was used to determine the level of paralogy between the SCL and LYL1 promoters, and to aid in the identification of a region downstream of the LYL1 promoter which showed structural hallmarks of the SCL +51 enhancer. Figure 6.8 A and B shows that the +51 enhancer contains the consensus E-box/GATA motif while promoter 1a contains two GATA sites residing close to each other. The LYL1 promoter also contains two GATA sites (Figure 6.8 C) and, at approximately 33 kb downstream of the transcription start site of LYL1, a highly conserved E-box and GATA motifs separated by 8 bases was identified (Figure 6.8 D and 6.9). ChIP-qPCR was performed to complement assays performed previously in Chapter 5 to demonstrate the binding of members of the SCL erythroid complex at the SCL and LYL1 promoters and also at the downstream regions of SCL and LYL1 (Figure 6.10). Substantial enrichments of up to 60 fold were seen in both the LYL1 promoter and the +33 region in both K562 and HEL in all five ChIP assays except for LMO2 (where the antibody used did not facilitate good enrichments for any of the experiments performed in this thesis). The enrichments at the SCL promoter were somewhat lower, but, as was previously shown by ChIP-on-chip studies, the +51 enhancer also showed large enrichments of up to 60-fold in both K562 and HEL cells. This data suggests that a putative novel regulatory element for LYL1 was identified (see also Discussion). Moreover, the regulation of both SCL and LYL1 may be similar and is likely to be mediated through interactions between distal elements and their corresponding promoters. This data also demonstrates that the SCL erythroid complex is likely to mediate interactions through other regulatory regions apart from promoters, many of which may not have been detected by TF binding on the promoter array.

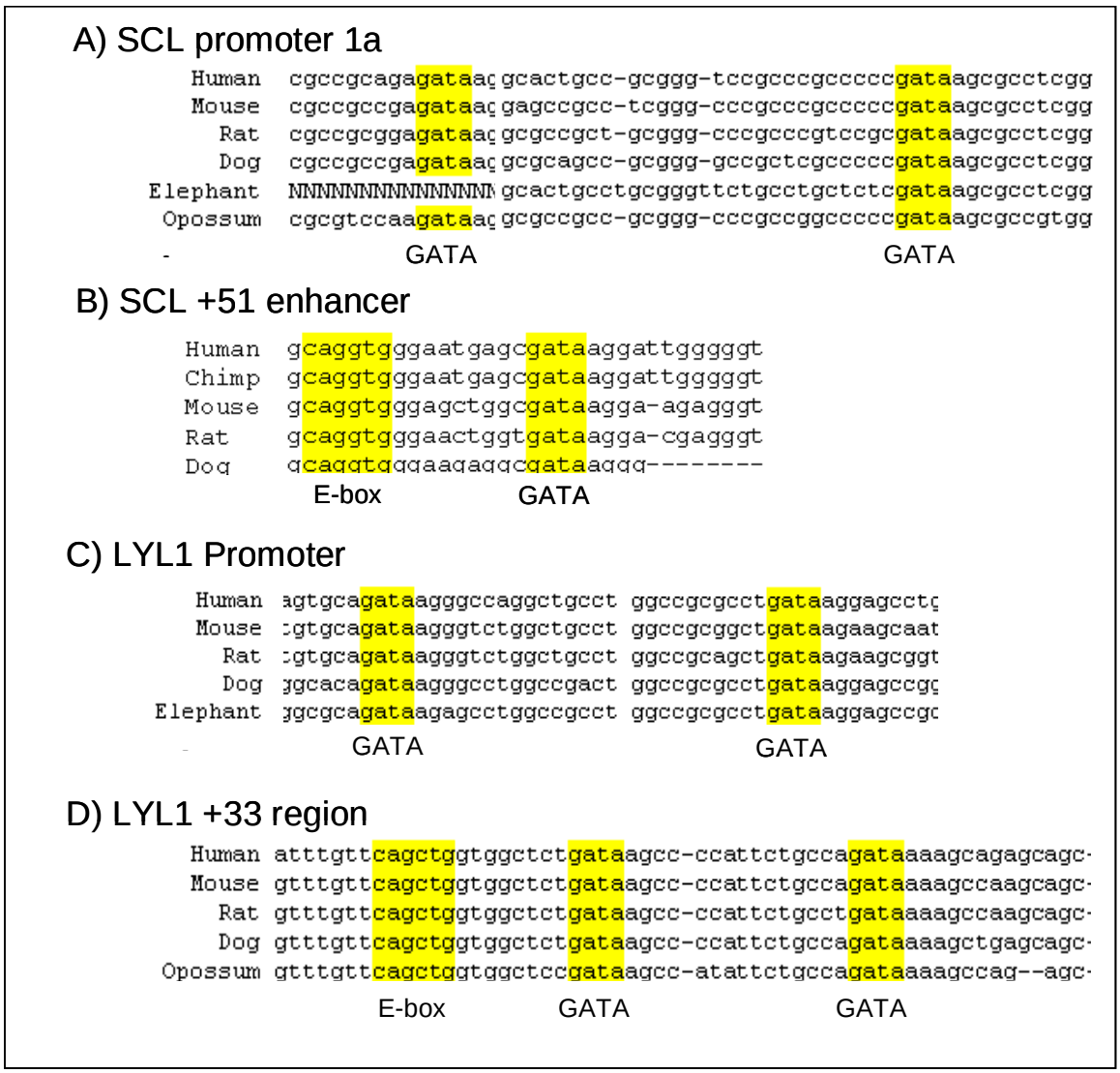


Figure 6.8. Multiple sequence alignments of transcription factor binding sites in the regulatory regions of SCL and LYL1. E-box and GATA motifs were identified by TESS and TFSearch and by viewing the conserved TFBS track on the UCSC genome browser. Multiple species sequence alignments were taken from the UCSC genome browser. Labelled in yellow are the conserved nucleotides across species in the E-box or GATA motifs. Panel A: SCL promoter 1a; panel B: SCL +51 enhancer; panel C: LYL1 putative promoter; panel D: LYL1 putative +33 enhancer.

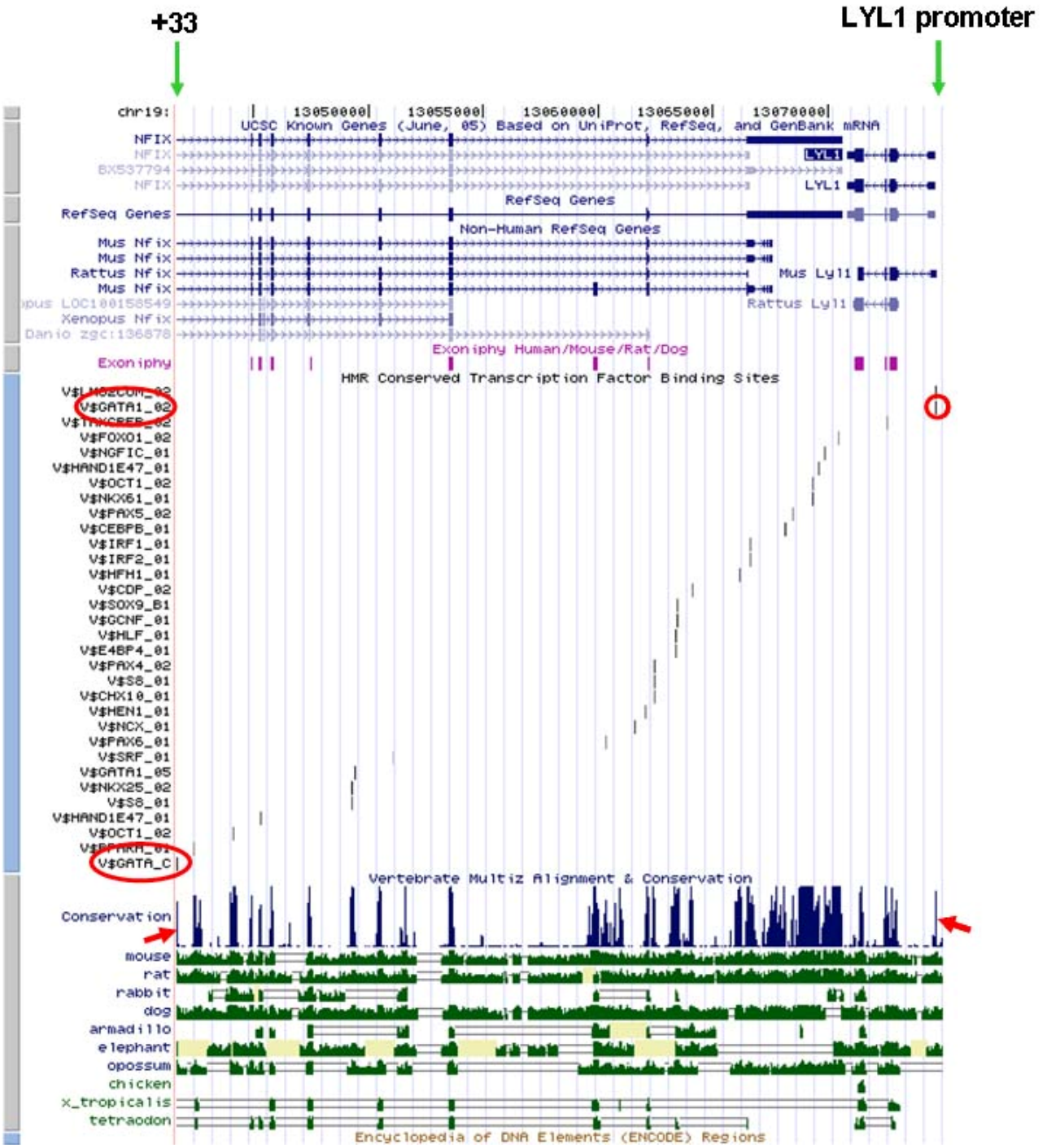


Figure 6.9. UCSC genome browser snapshot of the LYL1 promoter and putative enhancer regions. This diagram shows the LYL1 promoter and the +33 region in the UCSC genome browse. The green arrows indicate the position of the LYL1 promoter and the +33 region. The red circles show the conserved GATA sites identified by UCSC and the red arrows show the conservation of these GATA sites across species.

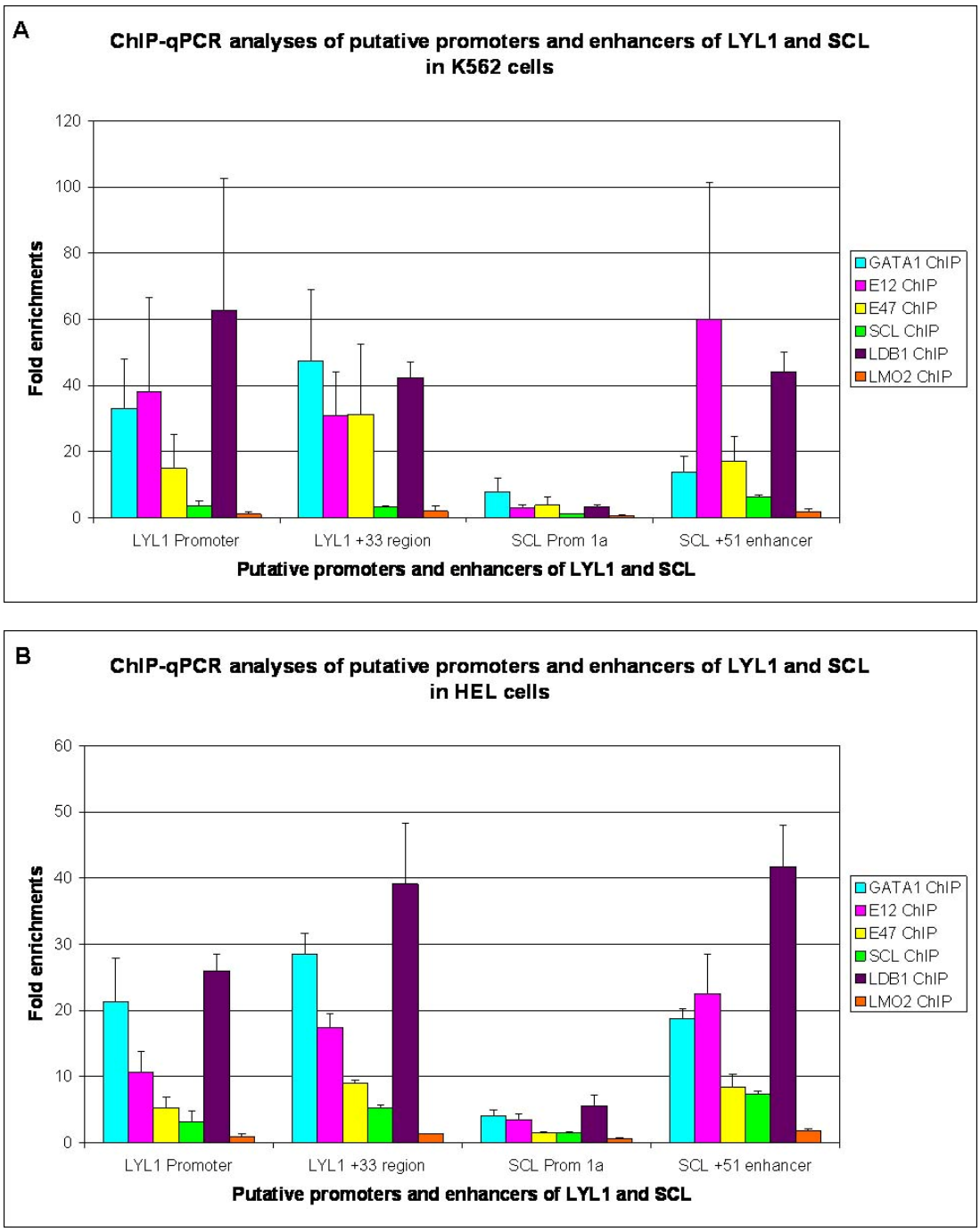


Figure 6.10. ChIP-qPCR analyses of promoters and distal regulatory elements of LYL1 and SCL in K562 and HEL cells. Histograms showed the fold enrichments of promoters and enhancers of LYL1 and SCL in ChIP-qPCR. Panel A: ChIP-qPCR in K562; panel B: ChIP-qPCR in HEL. Y-axes: fold enrichments above background. X-axes: regulatory regions of target genes. The colour-coded key for the various ChIP assays against members of the SCL erythroid complex is shown on the right of the panels. Error bars showed standard errors of two biological replicates.

6.3.2 Integration of expression and ChIP data: Derivation of transcriptional interaction networks in the erythropoeitic lineage.

The expression analyses on Affymetrix array and the ChIP-on-chip assays in the previous Chapters, together with the expression time-course studies of putative gene targets described in this Chapter, provide a wealth of biological information describing the cause and effects of regulatory interactions by members of the SCL erythroid complex in the human erythroid lineage. However, the vast amount of data produced from these experiments, and the variety of types/sources of data (array, qPCR, ChIP, two different cell lines, etc.,) makes overall interpretations difficult. To facilitate the integration of these datasets in a meaningful way, interaction network diagrams were generated using all or subsets of the data as discussed in the following sections.

6.3.2.1 Networks generation based on ChIP-on-chip data

The results obtained from the ChIP-on-chip experiments in Chapter 5 identified the promoters of 24 genes (Table 5.2) which were likely to be putative target for one or more members of the SCL erythroid complex. These 24 target genes, identified with both statistical methods A and B (summarised in Chapter 5, Table 5.2) were used for the generation of a network diagram using BioTapestry (Figure 6.11). This network diagram only shows the binding of each member of the SCL erythroid complex to the promoters of their putative target genes but does not integrate the mode of regulation involving these TFs at the transcriptional level (activation or repression).

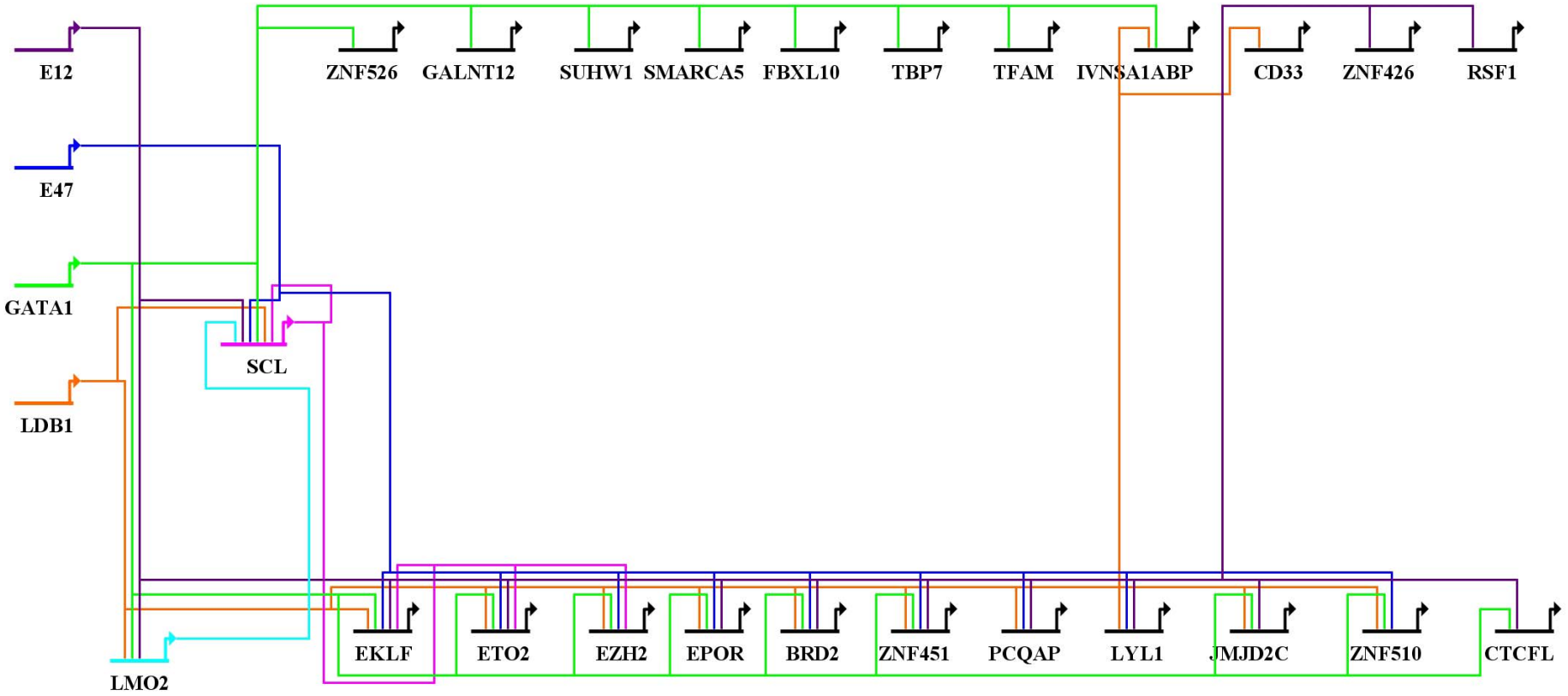
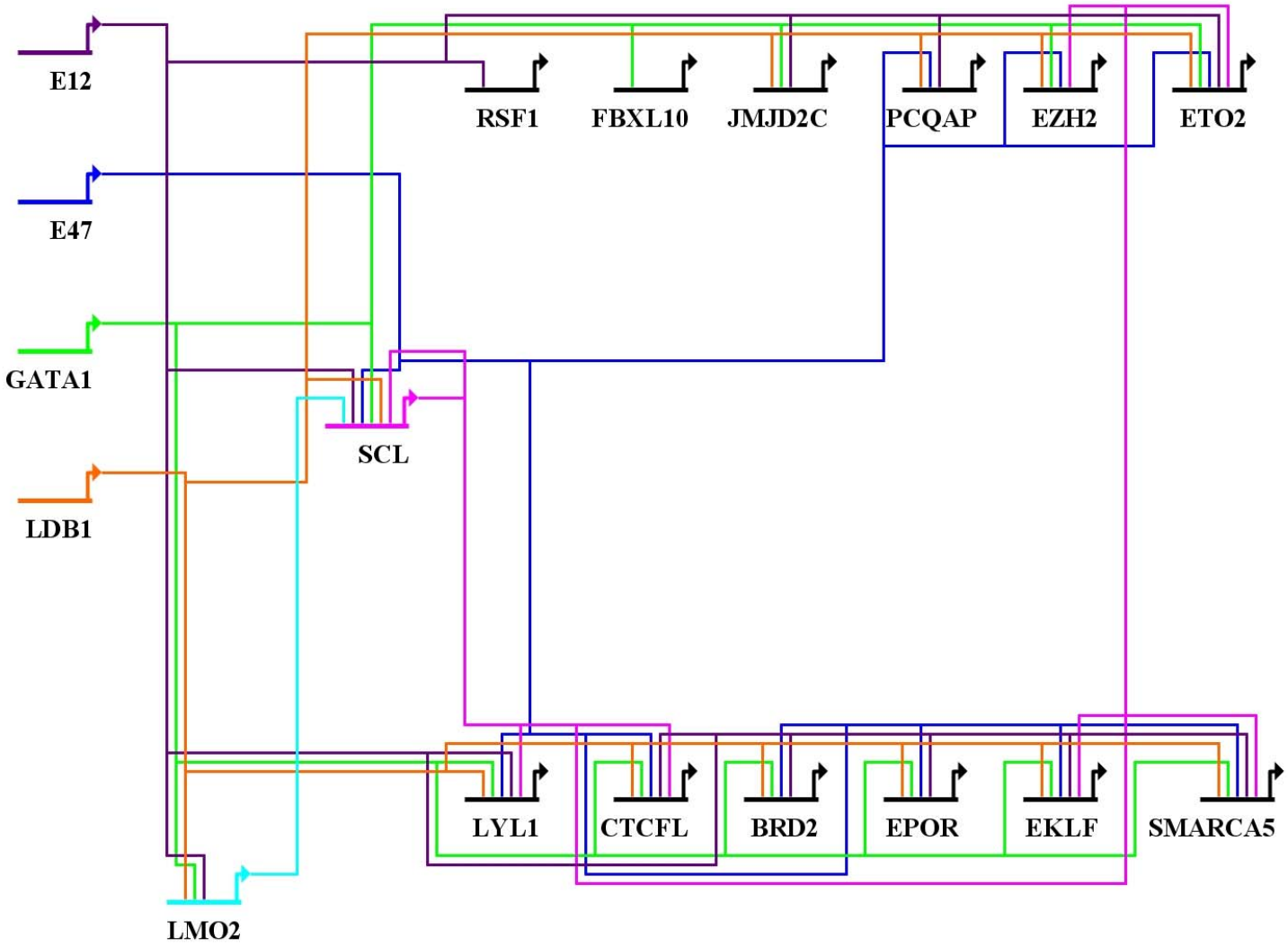


Figure 6.11. Network diagram of promoter-TF interactions of members of the SCL eyrthroid complex (GATA1, SCL, E12, E47, LDB1 and LMO2) in ChIP-on-chip analyses. Coloured bars on the left represent the 6 members (GATA1, SCL, E12, E47, LDB1 and LMO2) in the complex under study while the black bars on the right shows the targets. Note that LMO2 and SCL are also target genes. The coloured lines show the binding events between the TF and the target gene promoters; SCL: pink, GATA1: green, E12: violet, E47: blue, LDB1: orange, LMO2: aqua.

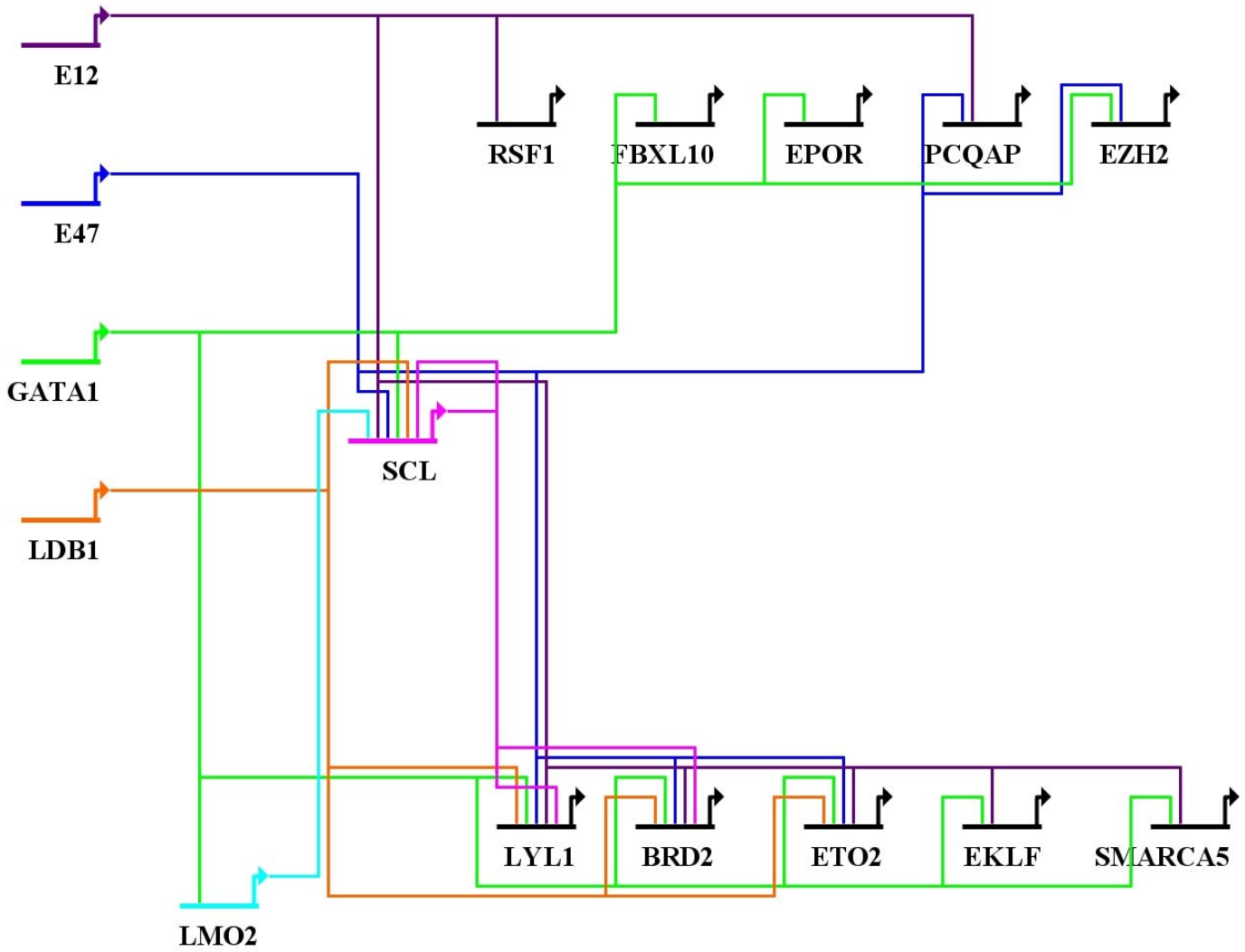
6.3.2.2 Integration of ChIP-qPCR data into networks based on ChIP-on-chip interactions

Fourteen of the target genes in the ChIP-on-chip analyses were further validated and characterised by ChIP-qPCR in K562 and HEL cells (Chapter 5, section 5.4.4.3). Additional TF binding events identified by ChIP-qPCR in K562 were incorporated into the network diagram for K562 (Figure 6.12 A). ChIP-qPCR data derived for HEL cells is shown in the network diagram in Figure 6.12 B. Since it was shown that approximately 60% of the TF-target interactions were in common between K562 and HEL (Chapter 5 section 5.4.4.3 D), the interactions detected were likely to be biologically relevant and reflect the erythropoietic lineage *in vivo*. Thus, a network diagram was also produced to reflect all of the interactions detected by ChIP-on-chip or ChIP-qPCR in either K562 or HEL as a representation of the erythroid lineage (Figure 6.12 C). The ChIP-qPCR provided additional information of transcription factor-promoter binding, than that obtained from ChIP-on-chip, as indicated by the number of novel linkages shown between the TFs and the targets in the diagrams of Figure 6.12.

A) Network based on ChIP-on-chip and ChIP-qPCR data in the K562 cell line



B) Network diagram based on ChIP-qPCR data in the HEL cell line



C) Network diagram based on ChIP-on-chip and ChIP-qPCR data in either K562 or HEL

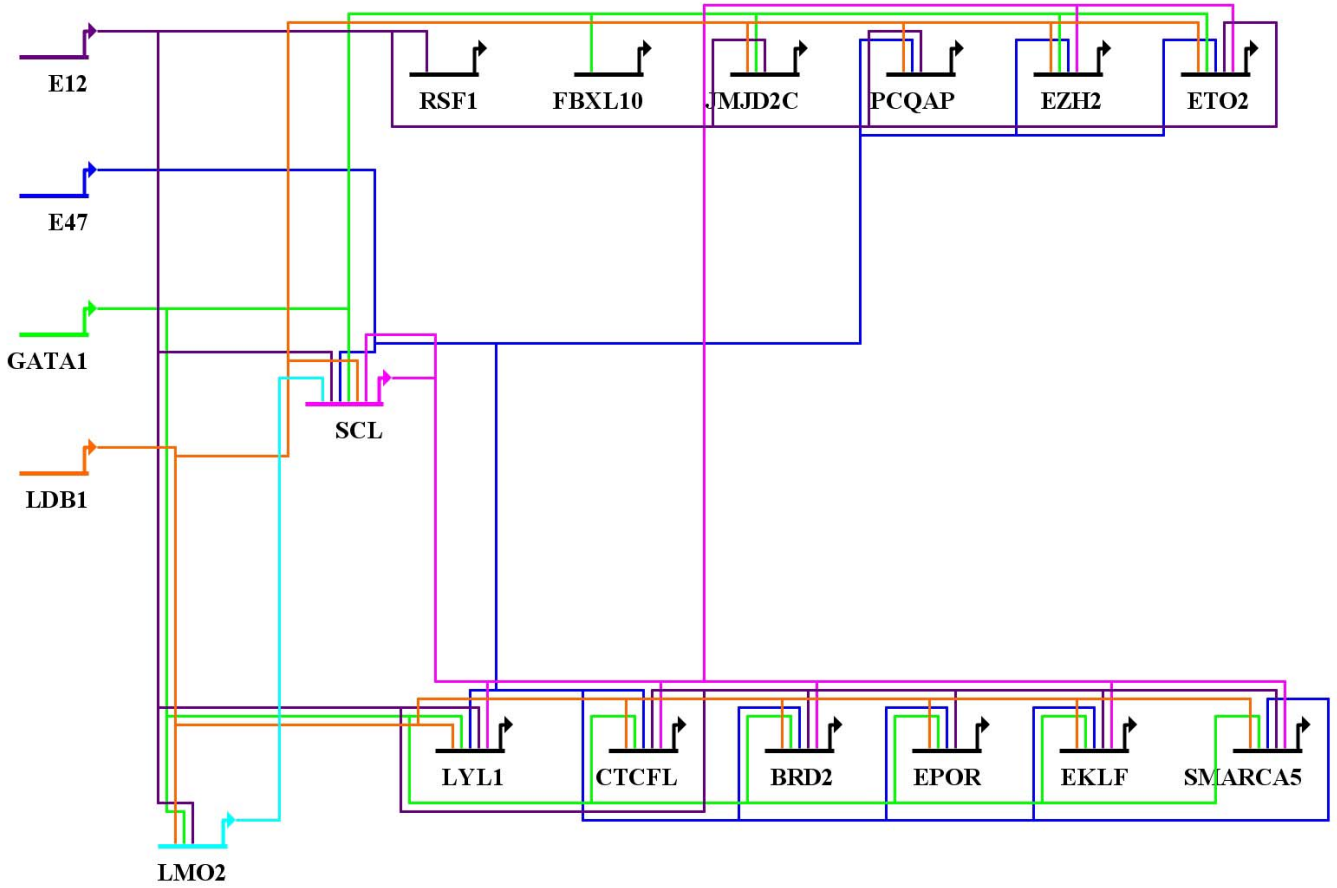


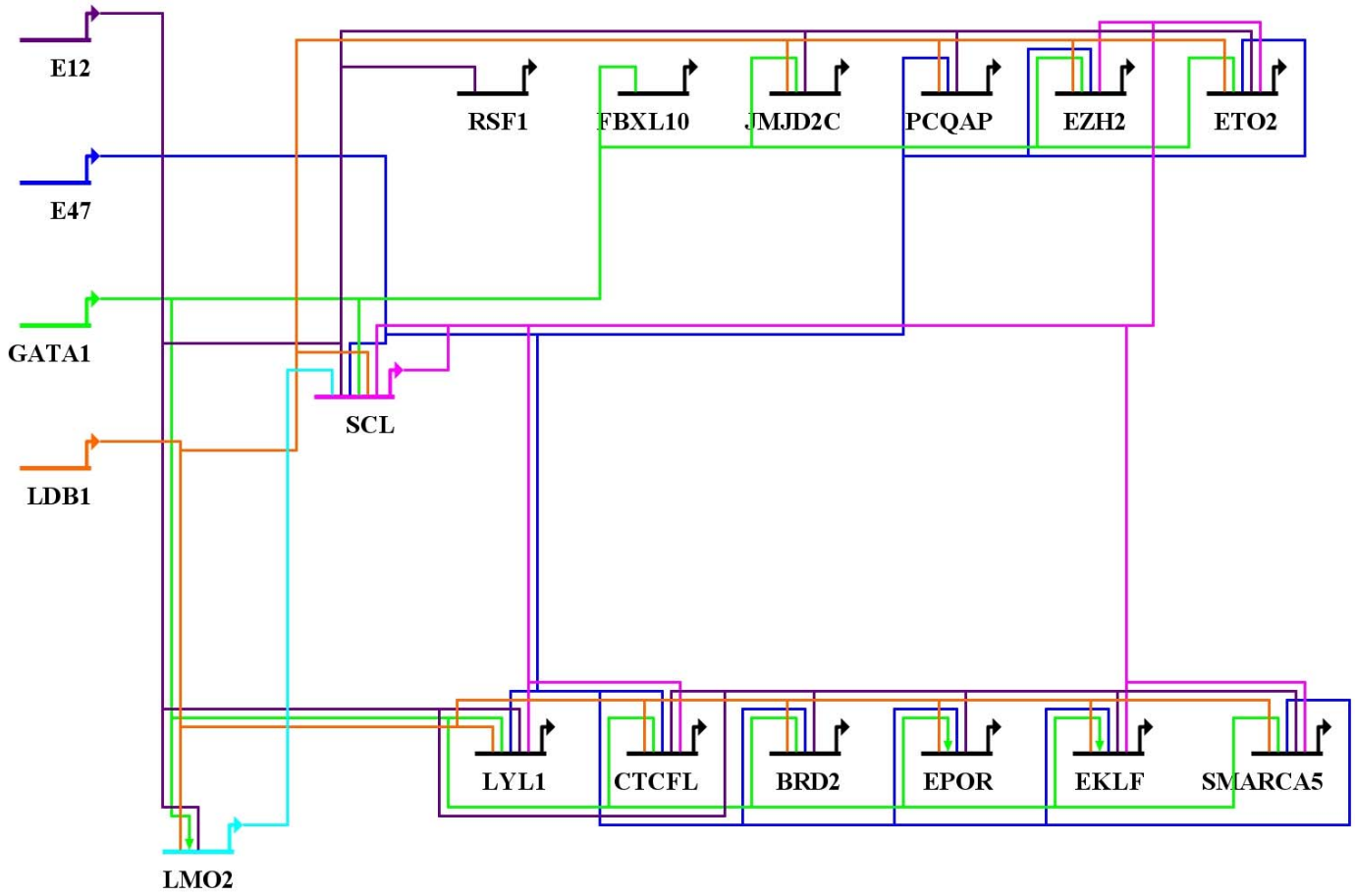
Figure 6.12. Network diagram of promoter-TF interactions of members of the SCL erythroid complex (GATA1, SCL, E12, E47, LDB1 and LMO2) based on ChIP-on-chip and ChIP-qPCR in the K562 and HEL cell lines. Coloured bars on the left represent the 6 members (GATA1, SCL, E12, E47, LDB1 and LMO2) in the complex under study while the black bars on the right shows the targets. Note that LMO2 and SCL are also target genes. The coloured lines show the binding events between the TF and the target gene promoters; SCL: pink, GATA1: green, E12: violet, E47: blue, LDB1: orange, LMO2: aqua. Panel A: the network in the K562 cell line; panel B: the network in the HEL cell line. Panel C: the putative network in the erythroid lineage (K562 and HEL cells data combined).

6.3.2.3 Integration of expression information into interaction networks

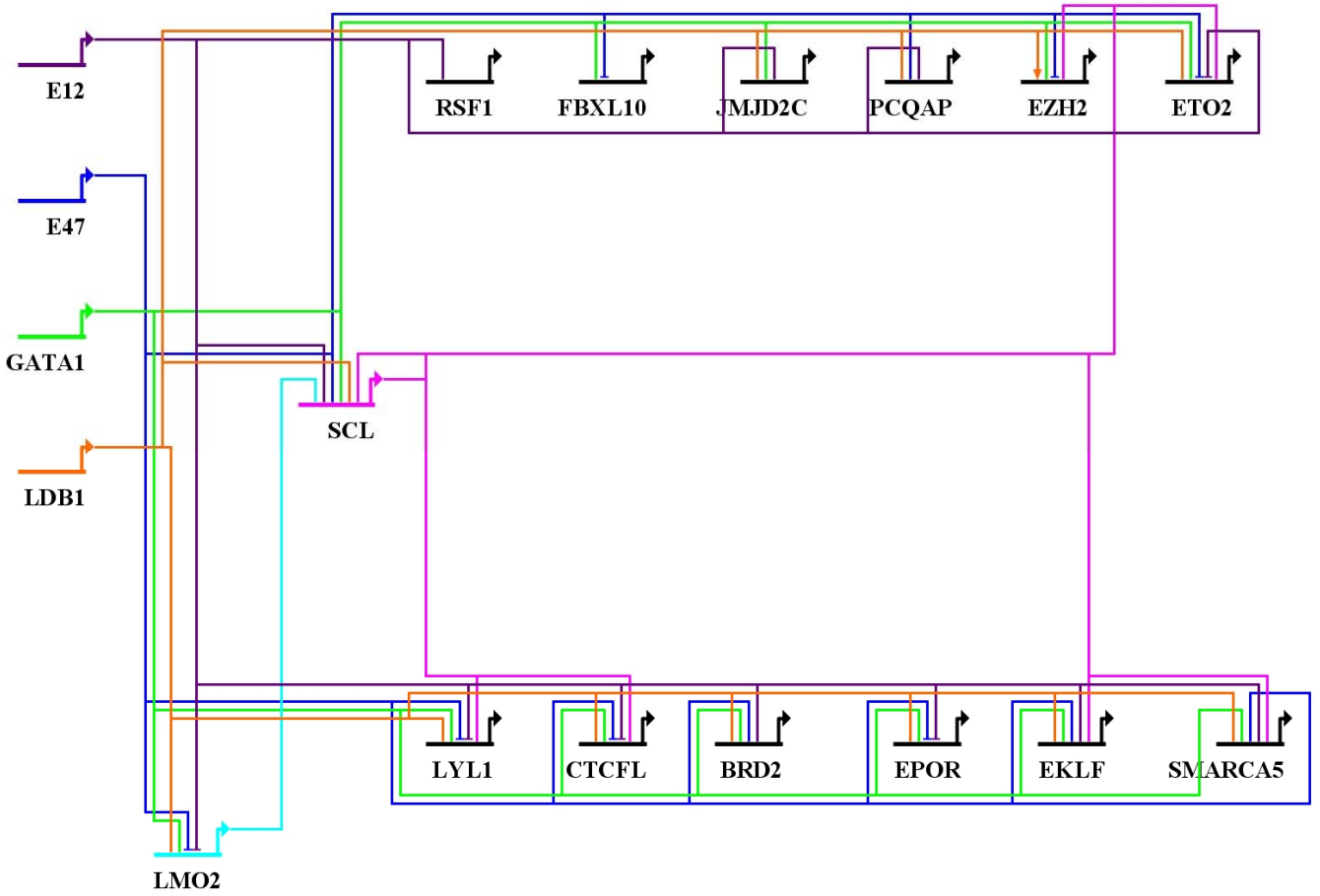
Since TF-target interactions had been confirmed at the level of ChIP-on-chip and ChIP-qPCR for the 14 targets of members of the SEC (see Chapter 5, section 5.4.4.4), it was possible to further elaborate the networks with data showing the mode of regulation of these targets (activation or repression). Thus, the siRNA-mediated TF knockdown data for these targets derived from both the Affymetrix GeneChip knockdown studies and the qPCR time-course studies (section 6.3.1.2) were incorporated into network analysis. This would provide a combined "cause and effect" network for the SCL erythroid complex, based on data described in this thesis. Initially, the Affymetrix GeneChip expression data for these target genes was integrated (Figure 6.13 A). This allowed the activation and repression information for each putative target gene to be determined based on the Affymetrix GeneChip study alone. It should also be noted that expression data was only included for direct TF-target interactions and downstream secondary effects manifested at the level of expression were not shown. Similarly, network diagrams were derived for these 14 target gene interactions with the incorporation of the qPCR time-course knockdown expression data at different time points (12, 24, 36, 48 hour) after knockdown (Figures 6.13 B, C, D, E). The cut-off fold increase or decrease in expression as described in section 6.3.1.2 was used to determine activation or repression status.

The networks generated using the Affymetrix GeneChip data and the expression time course data at the 24 hour time point were not the same (Figure 6.13 A and C). Only EKLF was found to be regulated in both studies. EPOR and LMO2 were shown to be activated by GATA1 in the Affymetrix GeneChip data but not in the expression time course. Conversely, EPOR, JMJD2C and ETO2 were shown to be activated by LDB1 while FBXL10 was shown to be repressed by GATA1 only in the expression time course.

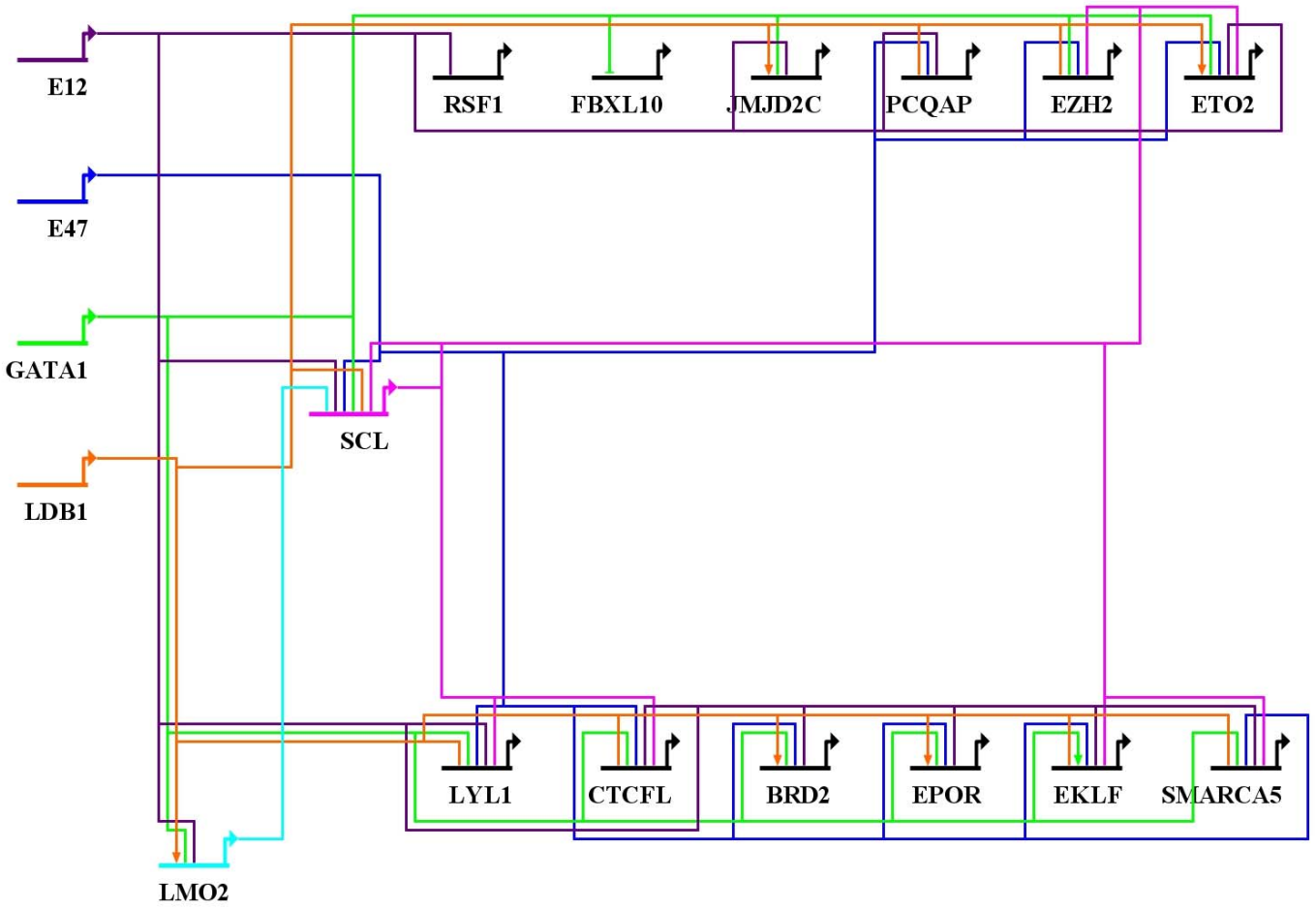
A) Network diagram based on ChIP-on-chip, ChIP-qPCR and Affymetrix expression study at the 24 hour time point in the K562 cell line



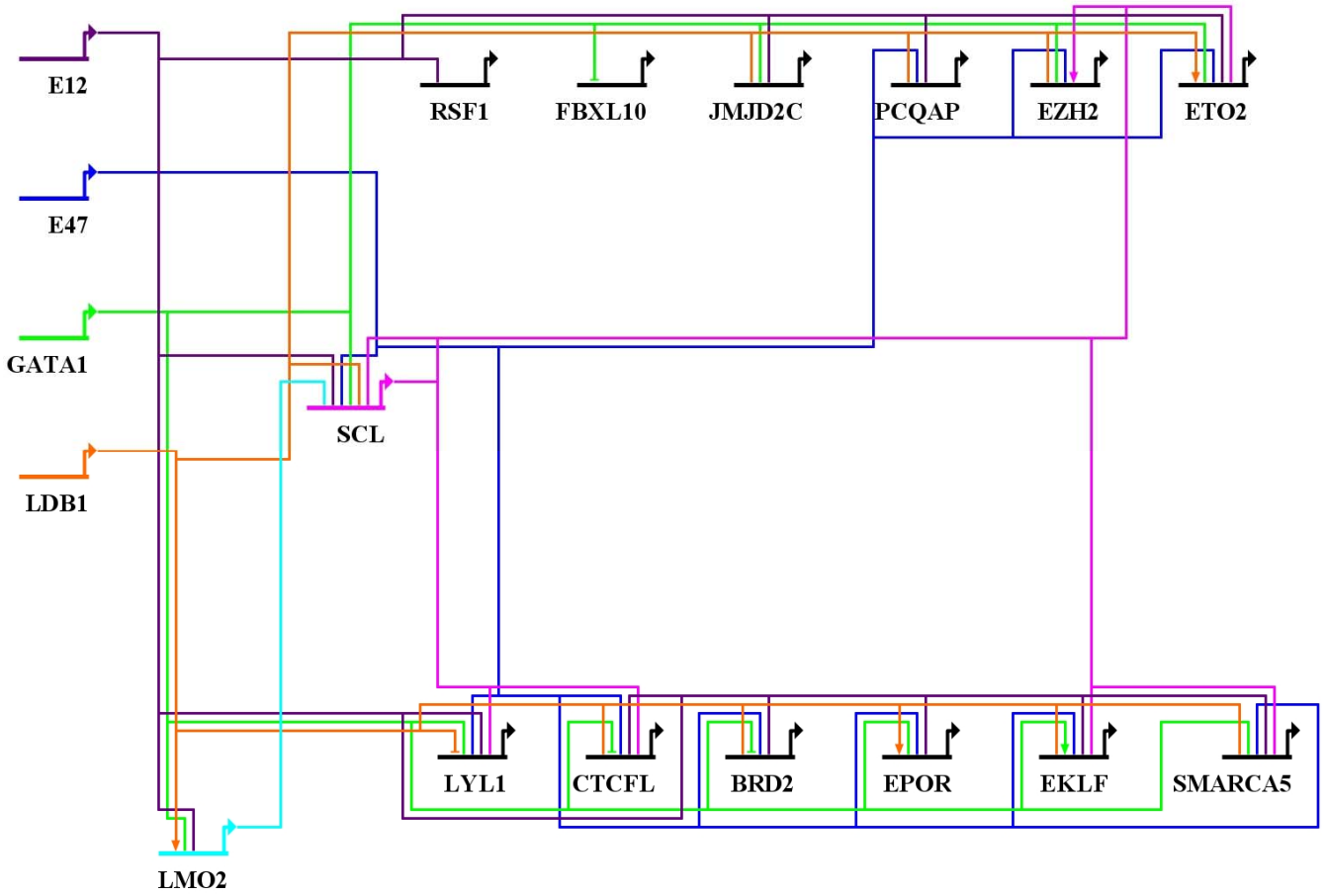
B) Network diagram based on ChIP-on-chip, ChIP-qPCR and time-course expression study at the 12 hour time point in the K562 cell line



C) Network diagram based on ChIP-on-chip, ChIP-qPCR and time-course expression study at the 24 hour time point in the K562 cell line



D) Network diagram based on ChIP-on-chip, ChIP-qPCR and time-course expression study at the 36 hour time point in the K562 cell line



E) Network diagram based on ChIP-on-chip, ChIP-qPCR and time-course expression study at the 48 hour time point in the K562 cell line

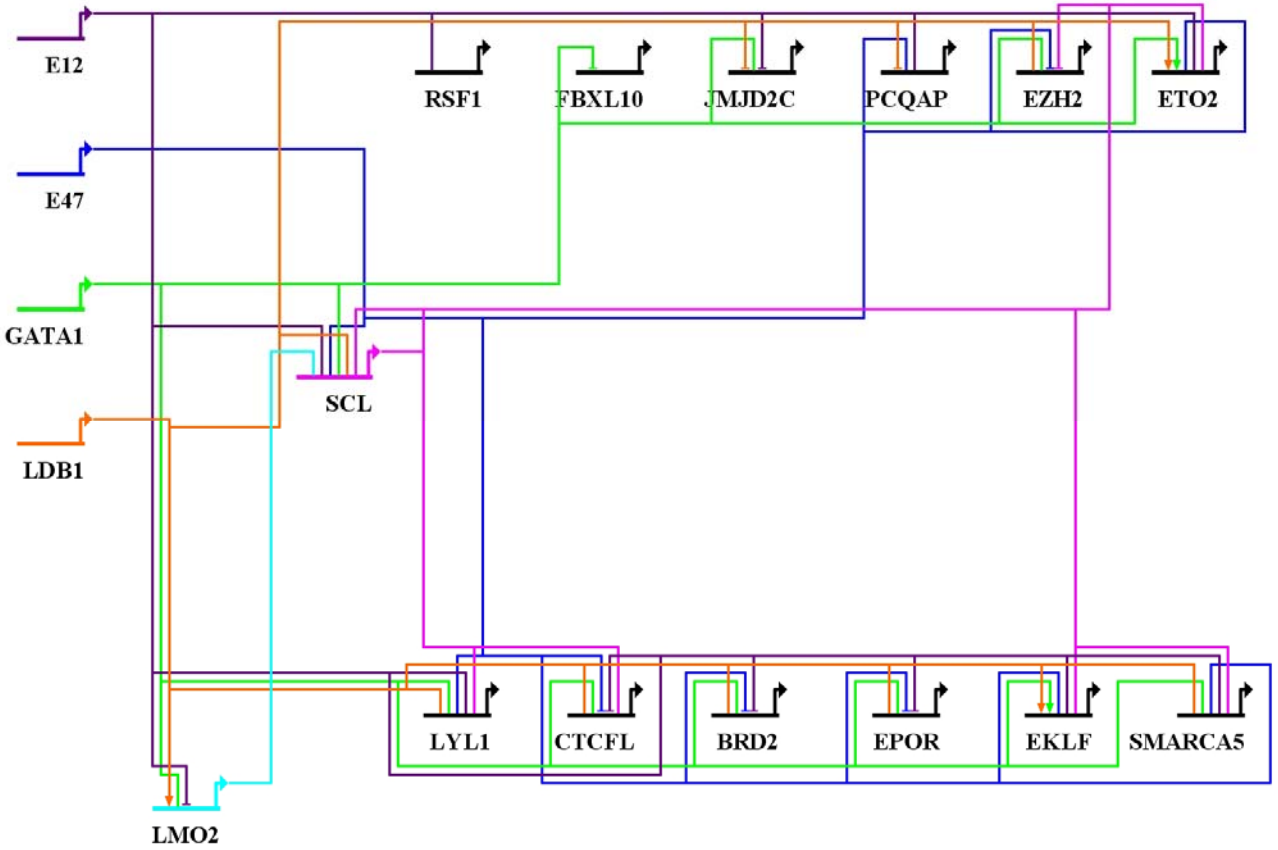


Figure 6.13. Network diagram of promoter-TF interactions of members of the SCL eyrthroid complex (GATA1, SCL, E12, E47, LDB1 and LMO2) based on ChIP-on-chip, ChIP-qPCR, Affymetrix GeneChip and expression time-course in the K562 cell line. Coloured bars on the left represent the 6 members (GATA1, SCL, E12, E47, LDB1 and LMO2) in the complex under study while the black bars on the right shows the targets. Note that LMO2 and SCL are also target genes. The coloured lines show the binding events between the TF and the target gene promoters; SCL: pink, GATA1: green, E12: violet, E47: blue, LDB1: orange, LMO2: aqua. The arrow head of these coloured lines indicate activation while the dash indicates repression. Panel A: network integrating ChIP-on-chip, ChIP-qPCR and Affymetrix GeneChip. Panel B: network integrating ChIP-on-chip, ChIP-qPCR and expression time-course at the 12 hour time point. Panel C: network integrating ChIP-on-chip, ChIP-qPCR and expression time-course at the 24 hour time point. Panel D: network integrating ChIP-on-chip, ChIP-qPCR and expression time-course at the 36 hour time point. Panel E: network integrating ChIP-on-chip, ChIP-qPCR and expression time-course at the 48 hour time point.

6.3.2.4 Networks generation with an integration of all experimental studies

To consolidate the network analysis for these 14 target genes in the K562 cell line, all ChIP interaction data (ChIP-on-chip and ChIP-qPCR) and all expression data (either from Affymetrix GeneChip or qPCR expression analysis) were integrated as shown in Figure 6.14. Direct TF-target interactions were defined as an interaction confirmed by either ChIP-on-chip or ChIP-qPCR. Up or down regulation status for any TF-target interaction was shown if an expression change above the cut-off was found in either the Affymetrix experiment or the qPCR time course study. Furthermore, expression information for this network was obtained from the 24 hour time points after siRNA-mediated knockdown of each TF. This would avoid incorporating effects related to transfection (at the 12 hour time-point) and off-target RNAi effects (36 or 48 hour time points) in the biological information used to elaborate the network. This network diagram clearly shows the multiple interactions obtained for the 8 genes (EZH2, ETO2, CTCFL, LYL1, BRD2, SCL, ELKF, SMARCA5) considered targets of the whole SCL erythroid complex.

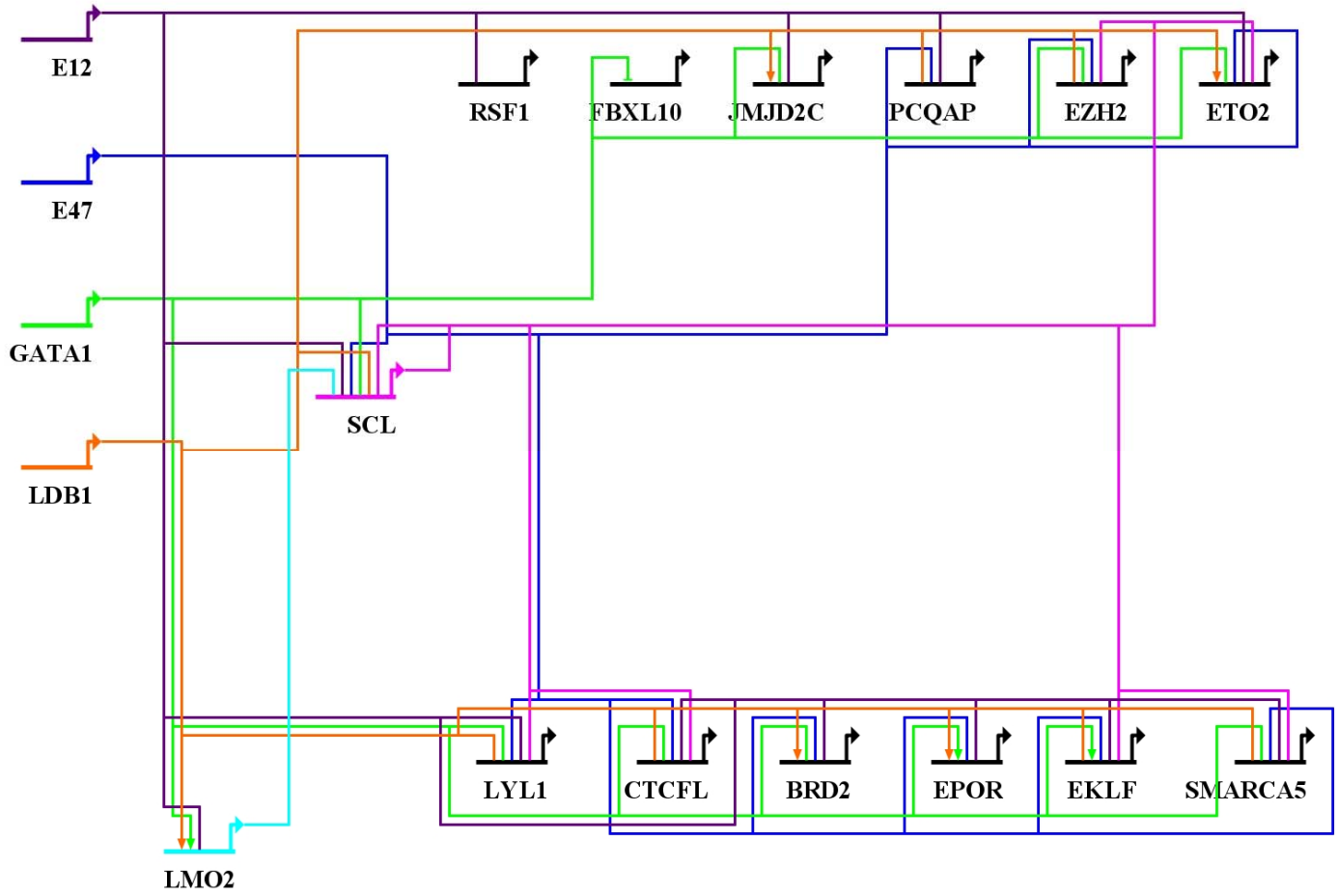


Figure 6.14. Network diagram of promoter-TF interactions of members of the SCL eyrthroid complex (GATA1, SCL, E12, E47, LDB1 and LMO2) based on ChIP-on-chip, ChIP-qPCR, Affymetrix GeneChip and expression time-course at the 24 hour time point in the K562 cells. Coloured bars on the left represent the 6 members (GATA1, SCL, E12, E47, LDB1 and LMO2) in the complex under study while the black bars on the right shows the targets. Note that LMO2 and SCL are also target genes. The coloured lines show the binding events between the TF and the target gene promoters; SCL: pink, GATA1: green, E12: violet, E47: blue, LDB1: orange, LMO2: aqua. The arrow head of these coloured lines indicates activation while the dash indicates repression.

6.3.2.5 Identification of network motifs

As mentioned in Chapter 1, a transcription network can consist of different types of network motifs which are combined and interlinked together to control gene expression (Chapter 1, section 1.2.1). In the transcription network centered around the regulation mediated by the SCL erythroid complex in K562 cells, a number of different types of these network motifs were identified based on the transcription factor-promoter binding events (Figure 6.15). These network motifs describe aspects of the overall network shown in Figure 6.14. Auto-regulation was observed for SCL (Figure 6.15 A). Feed forward loops were demonstrated for six genes. One example was that of GATA1 which regulated SCL which in turn regulated ETO2, EZH2, CTCFL, SMARCA5, EKLF and LYL1, while GATA1 itself also regulated these target genes (Figure 6.15 B). Fifteen regulator chains were identified - one example is illustrated by GATA1, LMO2 and SCL which were regulated in a series resulting in the regulation of ETO2 (Figure 6.15 C). A multiple input motif was observed where GATA1, SCL and LDB1 all worked together to control the expression of SMARCA5, LYL1 and ETO2 (Figure 6.15 D). A dense overlapping region was identified where different combinations of the 4 TFs GATA1, E12, LDB1 and E47 regulated FBXL10, JMJD2C, PCQAP and RSF1 (Figure 6.15 E). An example of single input motif was shown for GATA1 which activated SMARCA5, EPOR, ETO2, LYL1 and EKLF (Figure 6.15 F). Thus, the intricacies and multiple aspects of regulation which relate to the SCL erythroid complex, which are not apparent within any one dataset, can be deciphered by integrating datasets using network building software.

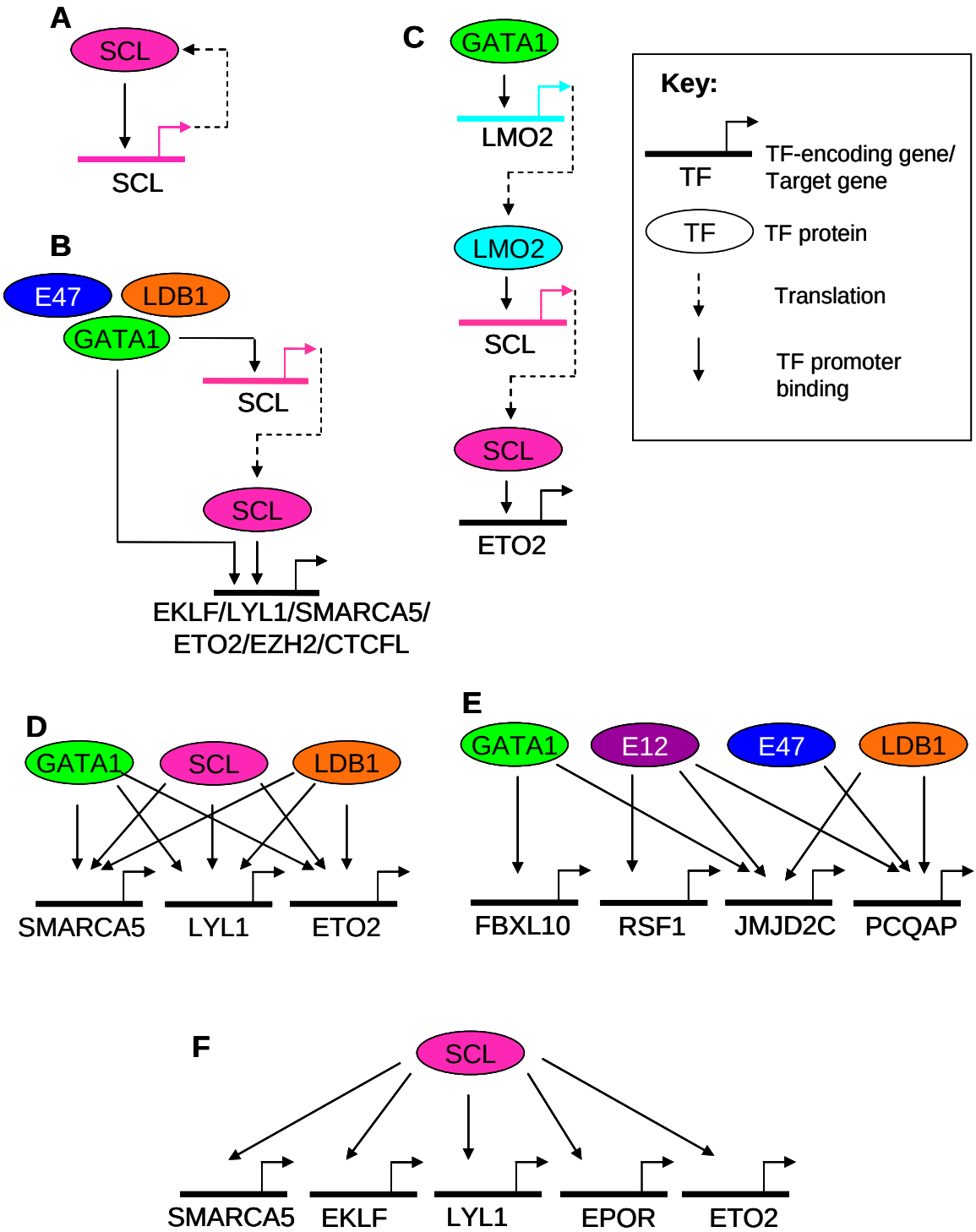


Figure 6.15. The complexities of regulation involving the SCL erythroid complex: Network motifs identified in the SCL erythroid complex transcription network in the K562 cell line. Combinations of these motifs regulate the expression patterns of target genes in the transcriptional network. The symbols and arrows descriptions are included in the key. A: auto-regulation motif. B: feed forward loop. C: regulator chain. D: multiple input motif. E: dense overlapping region. F: single input motif.

6.4 Discussion

The studies performed in this Chapter address a number of fundamental questions which describe the relationship between binding of the SCL erythroid complex to promoters and gene expression. Furthermore, the data in this Chapter helped in the interpretation of complex expression and TF binding datasets, as a means of understanding some of the issues associated with different types of data generated by different experimental approaches. The removal of transcription factor binding at promoters by siRNA knockdown was also studied to further characterise the relationship between the expression profiling by siRNA knockdown and transcription factor binding. Off-promoter binding study by transcription factors was another aspect of this Chapter to illustrate the fact that transcription factor binding is not limited to promoters. Auto-regulation of the SCL erythroid complex was further characterised in order to show that knockdown of any one member of the complex can impinge on the expression of the other members. Finally, the data obtained in this Chapter and previous Chapters were integrated to generate a network of transcriptional regulation in haematopoietic development by the SCL erythroid complex.

6.4.1 Low correlation rate between Affymetrix and ChIP-on-chip studies

A cross comparison between the data obtained in the Affymetrix analyses and the ChIP-on-chip study revealed that only a small portion of genes were found to be overlapped between these two studies (Section 6.3.1.1). A number of reasons may explain why there is low correlation between the datasets.

- **Time point studied**

The Affymetrix study only included the expression changes of one time point during the knockdown where an optimal silencing effect was observed. However, from the results obtained in the expression time-course study of the putative ChIP target genes during the knockdown of members of the SCL erythroid complex (section 6.3.1.2), many changes in expression level of the target genes occurred at other time points rather than the time point used for the Affymetrix analysis (only 27% of changes occurred at this time point). It is also possible that the change in expression of these target genes occurs immediately after the induction of knockdown and therefore they were not identified on the Affymetrix array.

- **Poised regulation**

The ChIP-on-chip analyses allowed us to identify the binding of a TF to the *cis*-regulatory elements but did not provide information about whether the factor actually regulates the genes nearby the regulatory elements. In fact, the recruitment of a TF may not necessarily correlates with the transcriptional control of its target genes. In a study of the mapping of NF-kappaB binding sites

along chromosome 22, binding was observed near a substantial number of genes whose expression was not regulated (Martone et al., 2003). These observations suggested that the recruitment of other transcription factors or co-factors, together with the chromatin modifiers, may be required to achieve a combinatorial effect on transcriptional regulation. Genes will only be expressed at a certain stage of development when all these co-factors and transcription factors are expressed and are recruited to the sites of regulation in the genome.

- **Auto-regulation of the SCL erythroid complex**

Various pieces of evidence from the Affymetrix analyses, ChIP-on-chip study and the expression time-course study all suggested auto-regulatory aspects for members of the SCL erythroid complex (see section 6.4.4). This adds to the complexity of the expression changes that may occur when one TF of the complex is perturbed – often resulting in perturbation or up-regulation of other members of the complex. This illustrates that expression of each member of the SCL erythroid complex is tightly controlled. One reason for this is that their target genes are maintained at a constant expression level even if one member of the complex is perturbed. As a result, no obvious expression changes may be detected.

- **Off-promoter regulation**

Since the TF promoter array used here had only coverage of 1 kb of the promoter regions/transcription start sites of genes, binding events occurring at other regulatory elements would not detected. The characterisation of binding events at the putative enhancers of SCL and LYL1 in section 6.3.1.5 illustrated that binding of a TF is not restricted to promoter regions. This was also shown in the mapping of NF-kappaB binding sites along chromosome 22 (Martone et al., 2003). Taken together, it is possible that a number of target genes picked up in the expression studies were regulated by TF of the SEC in *cis*-elements located outside of the promoters.

- **Indirect targets**

The Affymetrix expression analyses identified both direct target genes and secondary target genes downstream in transcriptional cascades whereas the ChIP-on-chip analyses identified only direct target genes. Depending on the role that the SCL eyrthoid complex has on transcriptional programmes (and the speed at which changes in these programmes can occur when the complex is perturbed), the number of secondary targets identified after 24 hours of siRNA-mediated knockdown may be very high. Furthermore, since magnitudes of expression changes do not correlate with whether the target is direct or secondary, top-scoring Affymetrix hits may not necessarily be direct targets. For both of these reasons, it is possible that a substantial proportion of top-scoring hits in the Affymetrix datasets could be secondary targets.

- **Effect of knockdown of TF on binding efficiency**

Although the siRNA-mediated knockdown of TFs was more than 70% of the original mRNA/protein levels in most cases, binding to DNA may not diminish in response to the knockdown - as demonstrated in the ChIP-on-chip study of GATA1. This would suggest that only a small proportion of the TF may be bound at any one time in the nucleus. In addition, the protein may need time to dissociate from their protein or DNA binding partners and degrade after the RNAi trigger is induced. The kinetics of these events may be somewhat different from the kinetics of knockdown of all of the unbound protein in the nucleus. Furthermore, there was not a correlation between GATA1 clearance from promoters and changes in gene expression – further supporting the idea that changes in GATA1 binding may not necessarily result in changes in expression of targets. Finally, electroporation itself was shown to perturb TF binding to some degree – this off-target effect would again reduce the correlation between GATA1-bound targets in wild type K562 cells, and expression changes observed in siRNA-mediated GATA1 knockdowns.

- **Microarray technology platform**

As described in Chapter 4 section 4.4.3.3, the validation rate of the differentially-expressed genes identified on the Affymetrix GeneChip by qPCR was only 37%. This indicates a high proportion of false positive targets may have been identified and is probably another reason why a low correlation was observed between the Affymetrix and ChIP-on-chip data.

- **Using siRNA knockdown to study expression**

A complete loss of the transcription factor cannot be achieved by transfecting siRNA into cells. Thus, variations in the amount of TF remaining in cells used in different replicate experiments may induce different effects on gene expression. Furthermore, as has been mentioned above, changes in target gene expression may be minimal when there is an incomplete gene knockdown due to binding of the remaining protein to its normal sites of regulation. Thus, a better way of studying downstream regulation by TFs may be to use an inducible knockout system.

6.4.2 The SCL complex transcription network

The whole SCL erythroid complex has previously been shown to regulate only three genes in human haematopoiesis: c-kit (Vitelli et al., 2000), α-globin (Anguita et al., 2004) and glycophorin A (Lahlil et al., 2004). Although researchers have tried to build a gene regulatory network of the erythroid lineage using published literature in mouse (Swiers et al., 2006), a genome-wide scale experimental study of the SCL complex in erythroid development in human has not been reported in the literature. The studies performed in this thesis provided further insights into the network of genes regulated by the SCL erythroid complex in K562 cells. Given that some of these targets were

also validated at the level of ChIP-qPCR in a second eyrthroid cell line (HEL), it is likely that this network is representative of events which occur in the human erythroid lineage. This project represents the first integrated approach to delineate the regulatory network controlled by the SCL erythroid complex on a genome-wide scale.

Eight additional direct target genes were identified for the whole SCL erythroid complex in this study where four of them are related to haematopoietic development (EKLF, ETO2, LYL1 and SCL) and the other four are related to chromatin structure, modification or remodelling (BRD2, CTCFL, SMARCA5 and EZH2) (see Chapter 5, section 5.5.5). According to the network diagram in Figure 6.15, EPOR, EKLF, BRD2 and ETO2 are likely to be activated while the mode of regulation for the others is unknown. The auto-regulatory role of the SCL erythroid complex as described in section 6.4.1 is likely to help modulate the expression of these target genes, which in turn control the expression of other erythroid-specific genes by direct transcription factor binding or chromatin remodelling in the erythroid lineage. The relationship between the network of genes controlled by the SCL erythroid complex and the wider network of gene regulation in erythroid development awaits further study.

6.5 Conclusions

The results presented in this Chapter illustrated the relationship between transcription factor binding and *cis*-regulatory elements, as well as the effect of knockdown on binding and release. The autoregulatory role of the SCL erythroid complex was further characterised. Despite the complexity of the datasets, a transcription network integrating all studies in this and previous Chapters was generated for the SCL erythroid complex for the first time in human erythroid cells.

Anguita, E., Hughes, J., Heyworth, C., Blobel, G.A., Wood, W.G., and Higgs, D.R. (2004). Globin gene activation during haemopoiesis is driven by protein complexes nucleated by GATA-1 and GATA-2. Embo J *23*, 2841-2852.
Chan, W.Y., Follows, G.A., Lacaud, G., Pimanda, J.E., Landry, J.R., Kinston, S., Knezevic, K., Piltz, S., Donaldson, I.J., Gambardella, L.*, et al.* (2007). The paralogous hematopoietic regulators Lyl1 and Scl are coregulated by Ets and GATA factors, but Lyl1 cannot rescue the early Scl-/- phenotype. Blood *109*, 1908-1916.
Lahlil, R., Lecuyer, E., Herblot, S., and Hoang, T. (2004). SCL assembles a multifactorial complex that determines glycophorin A expression. Mol Cell Biol *24*, 1439-1452.
Longabaugh, W.J., Davidson, E.H., and Bolouri, H. (2005). Computational representation of developmental genetic regulatory networks. Developmental biology *283*, 1-16.
Martone, R., Euskirchen, G., Bertone, P., Hartman, S., Royce, T.E., Luscombe, N.M., Rinn, J.L., Nelson, F.K., Miller, P., Gerstein, M.*, et al.* (2003). Distribution of NF-kappaB-binding sites across human chromosome 22. Proc Natl Acad Sci U S A *100*, 12247-12252.
Oda, K., Kimura, T., Matsuoka, Y., Funahashi, A., Muramatsu, M., and Kitano, H. (2004). Molecular Interaction Map of Macrophage. AfCS Research Reports *2*.
Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res *13*, 2498-2504.
Swiers, G., Patient, R., and Loose, M. (2006). Genetic regulatory networks programming hematopoietic stem cells and erythroid lineage specification. Dev Biol *294*, 525-540.
Valverde-Garduno, V., Guyot, B., Anguita, E., Hamlett, I., Porcher, C., and Vyas, P. (2004). Differences in the chromatin structure and cis-element organization of the human and mouse GATA1 loci: implications for cis-element identification. Blood *104*, 3106-3116.
Visvader, J., Begley, C.G., and Adams, J.M. (1991). Differential expression of the LYL, SCL and E2A helix-loop-helix genes within the hemopoietic system. Oncogene *6*, 187-194.
Vitelli, L., Condorelli, G., Lulli, V., Hoang, T., Luchetti, L., Croce, C.M., and Peschle, C. (2000). A pentamer transcriptional complex including tal-1 and retinoblastoma protein downmodulates c-kit expression in normal erythroblasts. Mol Cell Biol *20*, 5330-5342.

Chapter 7

Summary and future work

7.1 Summary of work presented in this thesis

With the completion of the sequencing of the human genome, studying how gene expression is regulated at the transcriptional level is fundamental in order to delineate biological pathways. Whilst haematopoiesis is one of the most well studied biological systems, the transcriptional control of the genes expressed therein is not fully understood. SCL, also known as TAL1, is the master regulator of haematopoietic development. It forms a multiprotein complex with GATA1, E2A, LDB1 and LMO2 (SCL erythroid complex) which binds to sequence-specific motifs in regulatory elements to regulate expression of its target genes. However, only three human genes have been identified as target genes for this complex: glycophorin, c-kit and α-globin. Therefore, the aim of the work presented in this thesis investigated the downstream regulatory network controlled by this SCL erythroid complex in human haematopoiesis. This was addressed using a combination of two independent approaches: (i) expression profiling to determine the effects of siRNA-mediated knockdown of the members of the SCL erythroid complex, and (ii) ChIP-on-chip studies to identify promoters bound by the complex. As a conclusion to this thesis, summaries of the principle findings of this thesis and possible avenues of future work will be discussed.

7.1.1 Developing working siRNAs for members of the SCL erythroid complex in K562 (Chapter 3)

In Chapter 3, siRNAs were used to induce knockdown of each of five members of the SCL erythroid complex. The siRNA assays were first characterised in the following aspects:

- The transfection efficiency and the delivery of siRNAs to the cells by electroporation were monitored by flow analyses using a FITC-labelled siRNA. This was done to investigate whether the transfection method was efficient for siRNA and to determine the correlation with knockdown efficiency. Using GATA1 knockdown as a model, it was shown that the transfection efficiency was over 90%. This was consistent with the level of GATA1 knockdown at both the mRNA and protein level.
- The growth pattern and cell morphology of K562 cells were investigated to study non-specific effects of electroporation and siRNA-mediated knockdown. It was shown that electroporation, but not siRNA, induced growth arrest and cell morphology changes within 48 hours after transfection.

The siRNA knockdown for each of five members of the SCL erythroid complex (GATA1, SCL, E2A, LDB1, LMO2) was further optimised through time-course studies. At least two siRNAs were shown to induce a knockdown efficiency resulting in less than 30% of the mRNA remaining at the 24 hour time point for each gene. To identify the protein levels during knockdown, antibodies were also tested and characterised for all of the members of the complex except LMO2.

The kinetics of knockdown was shown to be different for the different siRNAs and/or different genes being targeted in time-course study. It was found that 24 hour was the most optimal time point for all knockdowns apart from LDB1 which had an optimal knockdown at 36 hour after transfection. It was also noted that the mRNAs and proteins had different stability in some cases as the time points at which a maximum level of knockdown was induced for the mRNA differed from that of the protein.

7.1.2 Expression profiling analyses of siRNA knockdowns of the SCL erythroid complex (Chapter 4)

Following the characterisation and optimisation of the siRNA assays, the effect of knockdown of members of the SCL erythroid complex on the expression of other genes in the genome was studied in Chapter 4 using a human Affymetrix expression GeneChips. This array covers all of the annotated protein-coding genes in the entire human genome.

Samples prepared from each siRNA-mediated knockdown of the members of the SCL eyrthroid complex were hybridised to the Affymetrix GeneChips. Published known target genes for most transcription factors were identified by expression changes in siRNA assays, including glycophorin A (one of the published targets of the SCL erythroid complex), which was shown to change its level of expression in assays for SCL, GATA1 and E2A. This proves that the siRNA-induced knockdown in combination with Affymetrix expression analyses was able to identify published target genes of the transcription factors. However, the validation rate of genes which showed expression changes on the Affymetrix GeneChips using qPCR was only 37%. This demonstrates that the Affymetrix GeneChips may not be a reliable way (at least for this study) of expression profiling for the knockdown assays used here.

A disproportionate number of DNA binding proteins and haematopoietic-specific genes were identified for each member of the complex being studied. The DNA binding proteins include both transcription factors and some chromatin remodelling/modifying factors. This implied that the SCL erythroid complex plays an important role in the transcriptional regulation of genes related to blood development possibly through the control of other transcription factors and chromatin remodelling complexes. A number of genes were found to show expression changes induced by siRNA-

mediated knockdown of more than one member of the complex. In particular, 102 probes (92 genes) were shown to be regulated by GATA1, SCL and E2A – these, again, included a number of transcription factors. Taken together, these data confirm the idea that members of the SCL erythroid complex are likely work co-operatively to control gene expression. In addition, an auto-regulatory role of the SCL erythroid complex was observed i.e. one member of the complex regulating another. For example, E2A was activated by both SCL and GATA1. This provided further insight into the regulation of the SCL erythroid complex during haematopoiesis.

7.1.3 ChIP-on-chip analyses of the SCL erythroid complex (Chapter 5)

As a complimentary approach to identify targets of the SCL erythroid complex, chromatin immunoprecipitation in combination with a human transcription factor promoter array was used to confirm and identify putative target genes. The human transcription factor promoter array contained PCR products which covered 1 kb regions around the transcription start sites of the majority of human transcription factors and chromatin modifying/remodelling proteins. Multiple biological and technical replicates of ChIP-on-chip experiments were performed using the antibodies against all five members of SCL erythroid complex in wild type K562 cells. A number of transcription factors and chromatin modifying/remodelling factors were shown to bind one or more members of the complex in their promoter regions and were thus considered putative targets of members of the complex. Furthermore, each member of the complex also detected enrichments at the promoter, -10 enhancer and the +51 erythroid enhancer of SCL (where the consensus E-box GATA motif of the SCL erythroid complex had previously been identified). Conserved GATA and E-box motifs were also identified in the promoter regions of some of the novel putative targets. Binding to most of these promoter regions, by various members of the complex, were confirmed by ChIP-qPCR in K562 cells. Many of these binding events were also confirmed by ChIP-qPCR in another human erythroid cell line HEL, validating the biological relevance of the K562 targets. Consistent with the data obtained in the Affymetrix expression analyses, an auto-regulatory role for members of the SCL complex was observed. It was shown that LMO2, SCL and ETO2 (the latter has also been shown to be a member of the complex), were bound at their promoters by at least one other member. Taken all together, the data from this Chapter strongly suggested that eight new targets of the SCL erythroid complex had been identified: ETO2, LYL1, SMARCA5, CTCFL, BRD2, SCL, ELKF1, EZH2.

7.1.4 Further characterisation of putative target genes of members of the SCL erythroid complex (Chapter 6)

The Affymetrix expression analyses and the ChIP-on-chip analyses provided different information regarding the transcriptional regulation of the SCL erythroid complex. These data were compared in

Chapter 6 and it was demonstrated that only a small proportion of genes were identified to be putative targets by members of the SCL eytrhoid complex in both datasets. Experiments were described in this Chapter which provided some evidence as to why ChIP-on-chip and Affymetrix analysis yielded mainly different sets of targets.

Expression time-course studies of the putative target genes (obtained from ChIP-on-chip analysis) during siRNA-induced knockdown of each member of the complex revealed that target genes are affected at the level of expression in complex ways. Out of the fourteen putative target genes, four are repressed and two are activated by at least one member of the complex. Five are either activated or repressed by different members and the remaining three showed no significant change in expression. Similarly, it was also confirmed in these time-course studies that all members of the SCL erythroid complex are affected at the level of expression, when any one of the members is knocked down. Both of these experiments highlighted issues that may mean that the Affymetrix datasets may be difficult to interpret. Another reason to explain the differences in the Affymetrix and ChIP-on-chip datatsets was revealed in the ChIP-on-chip study during GATA1 knockdown. Enrichments in promoters and enhancers were shown to be not affected substantially in some cases when GATA1 was silenced. Electroporation was also shown to affect the binding of GATA1 to promoters when compared to the ChIP-on-chip study in the wild type cells. Thus, siRNA-mediated knockdown may not be sufficient to remove bound transcription factors from the promoters of their target genes. This would suggest that the Affymetrix analysis of the siRNA knockdown samples may not reveal detectable expression changes in some of its direct targets, where the relevant TF was still bound to its regulatory elements. The actual proportion of targets affected in this way is not known.

One limitation of using a promoter-based array in ChIP-on-chip studies is that binding of members of the SCL erythroid complex mediated at locations outside of promoters are not likely to be detected (unless promoters and other regulatory elements are in contact within the nucleus). However, for one target gene of the SCL erythroid complex, LYL1, it was possible to deduce binding events which lay outside the promoter region. Using known structural and functional similarities between LYL1 and SCL, it was shown that the both loci are bound by the complex in similar locations in both K562 and HEL cells. It was confirmed that the complex bound to a region containing a consensus E-box/GATA motif lying 33 kilobases downstream of the LYL1 promoter. This resulted in the identification of a putative LYL1 enhancer (named +33) which corresponds structurally to the SCL +51 erythroid enhancer. This provides evidence that SCL, through its erythroid complex, regulates its structural and functional paralogue, LYL1 in a similar manner to which it regulates its own expression in the erythroid lineage.

Based on the data obtained from the previous Chapters of this thesis and the expression time-course studies, transcriptional networks describing the relationship between the putative target genes and the transcription factors regulating them were generated. These network diagrams integrated all the information together to simulate the downstream regulatory network by the SCL erythroid complex in human erythroid cells. Many network motifs were also identified in these networks which were linked to the biology of erythroid development. Thus, the results from this thesis presented, for the first time, a regulatory network for the erythroid lineage, which was based on interactions involving the SCL erythroid complex.

7.2 Future work

The data describing the regulatory interactions involving the SCL erythroid complex in this thesis provide a useful benchmark for the elucidation of more complex regulatory networks of erythroid development in human cells. Possible avenues of future work are discussed below.

7.2.1 Confirmation of putative target genes in primary cells

All the studies performed for this thesis centered around an analysis of the human erythroid cell line K562. K562 is a well-established cell line, which is relatively easy to culture and transfect *in vitro* and grows relatively quickly with an average doubling time of 24 hours. Despite these advantages, this cell line was developed from a chronic myeloid leukaemia (CML) patient in blast crisis (Lozzio and Lozzio, 1977). It carries the BCR-ABL translocation where the BCR-ABL fusion protein is constitutively expressed in these cells. The BCR-ABL fusion is implicated in the suppression of apoptosis by activating a number of cell cycle genes and altering signalling pathways [reviewed in (Mughal and Goldman, 2006)]. Such a translocation undoubtedly has a number of undesirable effects on the expression of other genes in the genome, making it, to a certain degree, unrepresentative of the normal erythroid lineage. Thus, studies carried out with K562 should be taken with care and confirmed in other erythroid cell lines or, ideally, primary erythroid cells. HEL, an erythroid cell line derived from an erythroleukaemic patient, which does not contain the BCR-ABL translocation, was used to confirm the enrichments of promoters in the ChIP-qPCR studies. However, the HEL cell line is still a cancerous cell line and may not reflect normal erythroid development.

To confirm the regulation of putative target genes in normal human cells, primary erythroid cells (CD71+/GPA+) from normal individuals should be used for siRNA-induced knockdown expression studies and ChIP-on-chip study. The isolation of sufficient amount of primary cells for ChIP-on-chip studies is one of the obstacles hindering such future analyses. Moreover, primary cells are not

easy to transfect (Marodon et al., 2003). However, with the rapid advances of ChIP protocols and RNAi delivery, this is becoming possible. As mentioned previously in Chapter 1, retroviruses and lentiviruses have been successfully employed to deliver shRNAs into primary cells (Barton and Medzhitov, 2002; Stewart et al., 2003). Obviously, optimisation of transfection is required for different types of primary cells and cloning of shRNA to viral cassettes is necessary. As working siRNAs were already validated in Chapter 3 for members of the complex, cloning of these sequences into shRNAs should be possible, although it is not known whether these sequences will work well in shRNA systems. shRNAs must be processed into siRNA *in vivo* using the endogenous RNAi system. The flanking sequences, together with the sequences complementary to the target gene, must be carefully designed so that the shRNA can be processed *in vivo*. New technologies of ChIP study such as Carrier ChIP and MicroChIP have been used to map histone modifications with a limited number of cells (Chapter 5, section 5.1.1). Indeed, a fast carrier ChIP protocol has been developed to study transcription factor-DNA interaction in brain tissues (Hao et al., 2008). Such protocols would be required to make it possible to use limited numbers of primary erythroid cells for ChIP-on-chip or ChIP-qPCR analyses.

7.2.2 Investigation of histone modifications associated with the regulatory pattern

Another component of a transcription regulatory network is the chromatin domains and modifications associated with the genes in the network. As mentioned in Chapter 1, chromatin structure affects the binding of both sequence-specific and general transcription factors to regulatory element and thus plays a crucial role in transcriptional regulation of target genes. As a result, integrating the information for both chromatin modification and transcription factor binding can facilitate a more complete understanding of transcription.

Histone lysine acetylation has been shown to be associated with active transcription whereas deacetylation of histone subunits correlates with repression of transcription (Tse et al., 1998). Methylation of histone H3 lysine 4 is linked to active genes while methylation of histone H3 lysine 9 is linked to a repressed state (Lachner and Jenuwein, 2002) (Chapter 1, section 1.1.2.5). ChIP in combination with qPCR or microarray analysis could be used to study the modification status at the promoters and other regulatory elements of some of the putative target genes. This will allow us to determine the role of histone modifications in the regulation as well as to investigate the activation or repression of the genes. Furthermore, ChIP studies of these histone modifications during the knockdown of transcription factors in the SCL erythroid complex could be performed to study the significance of the transcription factor binding on the modifications status of relevant regulatory regions. This is because in some cases, chromatin-modifying enzymes are recruited to the regulatory elements of genes by transcription factors and co-factors that bind to these regions

(Brehm et al., 1998) (Chapter 1, section 1.1.2.3). Thus, studying how the modification status changes during knockdown can reveal important information about the precise role that transcription factors play in recruiting chromatin modifying complexes and how they facilitate the open chromatin structures which are required for RNA polymerase complexes to bind and initiate transcription or how they facilitate repressive chromatin configurations.

7.2.3 Identification of all possible regulatory elements bound by the SCL erythroid complex

As demonstrated in Chapter 6 section 6.3.1.5, the SCL erythroid complex has been shown to bind to the SCL +51 enhancer and a putative enhancer for LYL1. However, the microarray used in this thesis only contains promoter regions of human transcription factors. Thus, many regulation and binding events may be missed out in this experiment. To facilitate the identification of all possible binding sites within or outside the promoter regions of each member of the SCL erythroid complex, alternative methods should be employed. Tiling arrays covering a broader region of a particular gene locus of interest or whole genome tiling arrays can be used to study binding outside the promoter region (Horak et al., 2002; Lee et al., 2006). With the recent advances in the next generation sequencing technology, ChIP-seq can be performed for whole genome Solexa or 454 sequencing to map the binding sites in an unbiased manner (Barski et al., 2007; Johnson et al., 2007; Schones et al., 2008).

7.2.4 Knockout and ChIP studies of interesting putative target genes

The studies of transcriptional regulation in this thesis only focused on the SCL erythroid complex itself but not other transcription factors which are involved in erythroid development. From the ChIP-on-chip study, only a handful of direct target genes were identified for this complex. On the other hand, the Affymetrix analyses possibly identified a large number of putative direct and indirect target genes. One obvious area of interest would be to begin to assemble networks which incorporate all the appropriate links between direct and indirect targets of the SCL erythroid complex. One solution to this issue is to further investigate the transcriptional targets of the direct target genes of the SCL eryrthroid complex and identify their targets (so called “targets of the targets”). This again could be done by using a combination of complete gene knockout in mouse, expression analysis and ChIP-on-chip or ChIP-sequencing. Given that many problems are associated with the siRNA knockdown studies presented in this thesis, a better method of studying mode of regulation would be to generate conditional homozygous gene knockouts in mouse.

EKLF, LYL1 and ETO2 are interesting target genes which could be prioritised for knockouts as well as for use in ChIP-on-chip or ChIP-seq studies. Evidence was provided in this thesis that all

three of these TFs are targets of the SCL erythroid complex and their expression is regulated by the complex. Thus understanding the targets of these three TFs would be important for our understanding of erythropoiesis and regulatory events downstream of the SCL erythroid complex. EKLF is known to be a transcriptional activator of β-globin (Miller and Bieker, 1993). Furthermore, perturbation of EKLF in mouse ES cells demonstrated that EKLF is required for the final stages of definitive haematopoiesis (Nuez et al., 1995). ETO2 was identified as an interacting partner of the SCL erythroid complex and was shown to repress the transcription activator activities of the complex (Goardon et al., 2006). This was suggested as a developmental switch for expression of regulators related to terminal erythroid differentiation. The repressive function of ETO2 has been suggested to link to histone deacetylation as ETO2 associates with HDAC family members (Gelmetti et al., 1998; Wang et al., 1998). Therefore; studying the genes regulated by ETO2 and comparing them with the target genes of the SCL erythroid complex will provide insights into the dynamic changes in gene expression and repression controlled by the SCL erythroid complex, and the role of ETO2 within this complex, during erythroid differentiation. As mentioned previously in Chapter 1, LYL1 has a highly similar expression pattern and function as SCL and is considered to be its functional and structural paralogue. The identification of the LYL1 putative +33 enhancer (which resembles the +51 enhancer of SCL) also revealed a possible similar mode of regulation between the two during erythroid development. Thus, studying LYL1 in details allows us to further assess the role of this gene during erythropoiesis and would shed further light on its regulatory relationship with SCL.

7.2.5 Identification of other interacting partners in the SCL erythroid complex

In order to further characterise the SCL erythroid complex, studying the DNA elements to which it binds tells only one part of the story. Another aspect of understanding the complex is to characterise all of the protein components in the complex. Using CASTing and gel shift assay, the five initial components were identified to recognise and bind to the consensus E-box and GATA motifs (Wadman et al., 1997). Previous protein interaction studies using GST pull-down assays and co-immunoprecipitation have been used to identify other interacting partners of this complex. These include Sp1 (Lecuyer et al., 2002), and pRb (Vitelli et al., 2000). A larger-scale analyses of SCL interacting partners by expression of a biotin-tagged SCL protein followed by pull-down assays and mass-spectrometry identified ETO2 as a novel component (Goardon et al., 2006). However, a large-scale study to identify all possible components which bind specifically to the E-box/GATA composite DNA motif is still lacking. Immobilised DNA probes containing TFBS motifs have been used in matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF MS) (Nordhoff et al., 1999) to identify proteins which bind directly to DNA. Similar approaches

could be used to identify protein complexes which bind the E-box/GATA composite motifs or motifs nearby. However, this analysis would also allow us to examine protein components which do not directly bind DNA. Some known protein components of the SCL erythroid complex have been shown to recruit and interact with chromatin modifying or remodelling complexes which do not bind DNA directly. For example, SCL was shown to associate with HDACs (Goardon et al., 2006) whereas ETO2 was also shown to recruit HDAC family members (Gelmetti et al., 1998; Wang et al., 1998). In fact, the SCL complex has been demonstrated to associate with the SWI/SNF protein Brg1 which is a family of chromatin remodelling complexes which leads to transcriptional repression at the P4.2 protein in mouse (Xu et al., 2006). Thus, studying all the proteins by mass spectrometry would yield a better understanding of the role of all of the components of SCL erythroid complex and provide clues as to how this complex regulates gene transcription.

7.2.6 Functional assays for the putative LYL1 enhancer

Unlike SCL, where there have been many studies to identify the genomic sequences involved in its regulation (Chapter 1, section 1.4.2.1 E), little is known of the regulatory sequences which regulate LYL1 expression. Thus, the identification of a putative LYL1 enhancer (+33) which binds the SCL erythroid complex, gives one the opportunity to further characterise the regulation of LYL1. Previous analyses of the SCL +51 enhancer demonstrated it has enhancer activity in reporter assays (Dhami et al. submitted). To confirm the activity of the +33 region of LYL1, enhancer or promoter trap assays could be performed. This +33 region can be cloned into a reporter construct upstream of a luciferase reporter gene so that expression of the luciferase gene can be measured and compared to a control without the +33 region (Chapter 1, section 1.3.3.1 D). This will allow us to determine whether this region acts as an enhancer or promoter. However, similar to the situation with the +51 SCL enhancer, the +33 region of LYL1 lies closer to the promoter of another gene, NFIX than it does to the LYL1 promoter. Therefore, it would be necessary to determine whether this putative enhancer regulates expression of the LYL1 promoter and not the NFIX gene. One possible way would be to design the enhancer trap assays so that the reporter assay is driven under the control of either the LYL1 or NFIX promoters. However, given that enhancer traps are artificial constructs and do not take into account other regulatory aspects found *in vivo* or *ex vivo* (chromatin features, for example), this may not be sufficient to determine which gene the +33 region regulates *in vivo*. Therefore, one could generate mutations in the conserved E-box/GATA motif within the +33 element of an appropriate cell line (*ex vivo*) or as a mouse knockout (*in vivo*) and evaluate the effect this has on expression of NFIX and LYL1.

7.2.7 Studies of changes in expression or promoter binding of target genes at different stages of erythroid differentiation

As previously mentioned, the studies performed in this thesis focused on a cell line which may only reflect a specific timepoint in erythroid development. K562 cells can be differentiated into both erythroid and megakaryocytic lineages, suggesting that these cells are bipotential progenitors of both lineages. Thus, information obtained for this thesis, may not reflect aspects of early erythroid development alone, but also that of early megakaryocytic progenitors. The regulatory role of the SCL erythroid complex further downstream in erythroid development could be investigated by differentiating K562 cells. K562 can be differentiated into the erythroid lineage by the addition of hemin, and into the megakaryocytic lineage by the addition of phorbol myristate acetate (PMA), to culture media (Huo et al., 2006; Yi et al., 2004).

Indeed, hemin was used to demonstrate the differentiation of K562 in our laboratory (Figure 7.1). glycophorin A (GPA) is a cell surface marker of erythroid cells and a PE conjugated anti-GPA antibody was used to detect the cells expressing GPA in flow cytometry analysis. The highest GPA expression was observed at 72 hours after induction of differentiation. GPA expression diminished at 96 hour possibly due to the terminal differentiation of erythrocytes. This differentiation experiment demonstrated that K562 can be easily differentiated down the erythroid lineage and thus provides a useful platform for differentiation studies.

ChIP in combination with various analytical platforms (qPCR, arrays, or massively parallel sequencing) could be performed to study the changes in binding of DNA elements by the SCL complex during differentiation. This will allow us to determine which target genes are important at different stages of erythroid development.

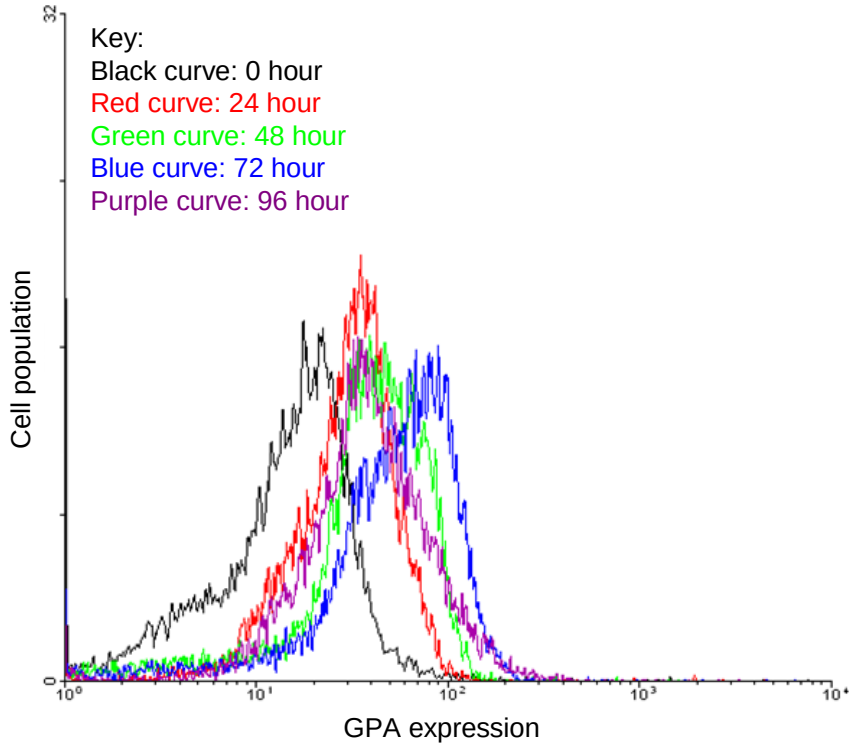


Figure 7.1. Differentiation of K562 cells by hemin. K562 cells were induced to differentiate down the erythroid lineage using hemin. The expression of the erythroid cell surface marker GPA was monitored by flow analysis using a PE conjugated anti-GPA antibody. Y-axis: number of cells in the population; x-axis: GPA expression. Black curve: 0 hour after induction of differentiation; red curve: 24 hour after induction of differentiation; green curve: 36 hour after induction of differentiation; blue curve: 48 hour after induction of differentiation; purple curve: 72 hour after induction of differentiation.

7.3 Final thoughts

The results presented in this thesis illustrated the downstream regulatory cascades of the SCL erythroid complex in the erythroid lineage. This is the initial effort of addressing its role in only one of the lineages in haematopoietic development. Given that there are more than 2000 transcription factors and more than 30000 human genes, understanding how various protein complexes control the expression of specific genes in haematopoietic development requires a huge effort. Advances in technology and computational tools are necessary to integrate all the data in a biologically meaningful way. The findings in this thesis provide a foundation and will make a valuable contribution towards this goal.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.
Barton, G.M., and Medzhitov, R. (2002). Retroviral delivery of small interfering RNA into primary cells. Proceedings of the National Academy of Sciences of the United States of America *99*, 14943-14945.
Brehm, A., Miska, E.A., McCance, D.J., Reid, J.L., Bannister, A.J., and Kouzarides, T. (1998). Retinoblastoma protein recruits histone deacetylase to repress transcription. Nature *391*, 597-601.
Gelmetti, V., Zhang, J., Fanelli, M., Minucci, S., Pelicci, P.G., and Lazar, M.A. (1998). Aberrant recruitment of the nuclear receptor corepressor-histone deacetylase complex by the acute myeloid leukemia fusion partner ETO. Molecular and cellular biology *18*, 7185-7191.
Goardon, N., Lambert, J.A., Rodriguez, P., Nissaire, P., Herblot, S., Thibault, P., Dumenil, D., Strouboulis, J., Romeo, P.H., and Hoang, T. (2006). ETO2 coordinates cellular proliferation and differentiation during erythropoiesis. Embo J *25*, 357-366.
Hao, H., Liu, H., Gonye, G., and Schwaber, J.S. (2008). A fast carrier chromatin immunoprecipitation method applicable to microdissected tissue samples. Journal of neuroscience methods *172*, 38-42.
Horak, C.E., Mahajan, M.C., Luscombe, N.M., Gerstein, M., Weissman, S.M., and Snyder, M. (2002). GATA-1 binding sites mapped in the beta-globin locus by using mammalian chIp-chip analysis. Proc Natl Acad Sci U S A *99*, 2924-2929.
Huo, X.F., Yu, J., Peng, H., Du, Z.W., Liu, X.L., Ma, Y.N., Zhang, X., Zhang, Y., Zhao, H.L., and Zhang, J.W. (2006). Differential expression changes in K562 cells during the hemin-induced erythroid differentiation and the phorbol myristate acetate (PMA)-induced megakaryocytic differentiation. Molecular and cellular biochemistry *292*, 155-167.
Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. Science *316*, 1497-1502.
Lachner, M., and Jenuwein, T. (2002). The many faces of histone lysine methylation. Curr Opin Cell Biol *14*, 286-298.
Lecuyer, E., Herblot, S., Saint-Denis, M., Martin, R., Begley, C.G., Porcher, C., Orkin, S.H., and Hoang, T. (2002). The SCL complex regulates c-kit expression in hematopoietic cells through functional interaction with Sp1. Blood *100*, 2430-2440.
Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K., *et al.* (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. Cell *125*, 301-313.
Lozzio, B.B., and Lozzio, C.B. (1977). Properties of the K562 cell line derived from a patient with chronic myeloid leukemia. Int J Cancer *19*, 136.
Marodon, G., Mouly, E., Blair, E.J., Frisen, C., Lemoine, F.M., and Klatzmann, D. (2003). Specific transgene expression in human and mouse CD4+ cells using lentiviral vectors with regulatory sequences from the CD4 gene. Blood *101*, 3416-3423.
Miller, I.J., and Bieker, J.J. (1993). A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Kruppel family of nuclear proteins. Molecular and cellular biology *13*, 2776-2786.
Mughal, T.I., and Goldman, J.M. (2006). Chronic myeloid leukemia: why does it evolve from chronic phase to blast transformation? Front Biosci *11*, 198-208.
Nordhoff, E., Krogsdam, A.M., Jorgensen, H.F., Kallipolitis, B.H., Clark, B.F., Roepstorff, P., and Kristiansen, K. (1999). Rapid identification of DNA-binding proteins by mass spectrometry. Nature biotechnology *17*, 884-888.
Nuez, B., Michalovich, D., Bygrave, A., Ploemacher, R., and Grosveld, F. (1995). Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene. Nature *375*, 316-318.
Schones, D.E., Cui, K., Cuddapah, S., Roh, T.Y., Barski, A., Wang, Z., Wei, G., and Zhao, K. (2008). Dynamic regulation of nucleosome positioning in the human genome. Cell *132*, 887-898.

Stewart, S.A., Dykxhoorn, D.M., Palliser, D., Mizuno, H., Yu, E.Y., An, D.S., Sabatini, D.M., Chen, I.S., Hahn, W.C., Sharp, P.A., *et al.* (2003). Lentivirus-delivered stable gene silencing by RNAi in primary cells. RNA (New York, NY *9*, 493-501.
Tse, C., Sera, T., Wolffe, A.P., and Hansen, J.C. (1998). Disruption of higher-order folding by core histone acetylation dramatically enhances transcription of nucleosomal arrays by RNA polymerase III. Mol Cell Biol *18*, 4629-4638.
Vitelli, L., Condorelli, G., Lulli, V., Hoang, T., Luchetti, L., Croce, C.M., and Peschle, C. (2000). A pentamer transcriptional complex including tal-1 and retinoblastoma protein downmodulates c-kit expression in normal erythroblasts. Mol Cell Biol *20*, 5330-5342.
Wadman, I.A., Osada, H., Grutz, G.G., Agulnick, A.D., Westphal, H., Forster, A., and Rabbitts, T.H. (1997). The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. Embo J *16*, 3145-3157.
Wang, J., Hoshino, T., Redner, R.L., Kajigaya, S., and Liu, J.M. (1998). ETO, fusion partner in t(8;21) acute myeloid leukemia, represses transcription by interaction with the human N-CoR/mSin3/HDAC1 complex. Proceedings of the National Academy of Sciences of the United States of America *95*, 10860-10865.
Xu, Z., Meng, X., Cai, Y., Koury, M.J., and Brandt, S.J. (2006). Recruitment of the SWI/SNF protein Brg1 by a multiprotein complex effects transcriptional repression in murine erythroid progenitors. The Biochemical journal *399*, 297-304.
Yi, Z.C., Wang, Z., Li, H.X., Liu, M.J., Wu, R.C., and Wang, X.H. (2004). Effects of chebulinic acid on differentiation of human leukemia K562 cells. Acta pharmacologica Sinica *25*, 231-238.

References

(2004). The ENCODE (ENCyclopedia Of DNA Elements) Project. Science (New York, NY 306, 636-640.

(2004). Finishing the euchromatic sequence of the human genome. Nature 431, 931-945.

(2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447, 799-816.

Acevedo, L.G., Iniguez, A.L., Holster, H.L., Zhang, X., Green, R., and Farnham, P.J. (2007). Genome-scale ChIP-chip analysis using 10,000 human cells. BioTechniques 43, 791-797.

Agulnick, A.D., Taira, M., Breen, J.J., Tanaka, T., Dawid, I.B., and Westphal, H. (1996). Interactions of the LIM-domain-binding factor Ldb1 with LIM homeodomain proteins. Nature 384, 270-272.

Akashi, K., Traver, D., Miyamoto, T., and Weissman, I.L. (2000). A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. Nature 404, 193-197.

Akiyama, Y. (1998). TFSEARCH: Searching Transcription Factor Binding Sites. http://wwwrwcporjp/papia/.

Allfrey, V.G., Faulkner, R., and Mirsky, A.E. (1964). Acetylation And Methylation Of Histones And Their Possible Role In The Regulation Of Rna Synthesis. Proceedings of the National Academy of Sciences of the United States of America 51, 786-794.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. Journal of molecular biology 215, 403-410.

Alwine, J.C., Kemp, D.J., and Stark, G.R. (1977). Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. Proceedings of the National Academy of Sciences of the United States of America 74, 5350-5354.

Anderson, K.P., Crable, S.C., and Lingrel, J.B. (1998). Multiple proteins binding to a GATA-E box-GATA motif regulate the erythroid Kruppel-like factor (EKLF) gene. The Journal of biological chemistry 273, 14347-14354.

Anderson, K.P., Crable, S.C., and Lingrel, J.B. (2000). The GATA-E box-GATA motif in the EKLF promoter is required for in vivo expression. Blood 95, 1652-1655.

Anguita, E., Hughes, J., Heyworth, C., Blobel, G.A., Wood, W.G., and Higgs, D.R. (2004). Globin gene activation during haemopoiesis is driven by protein complexes nucleated by GATA-1 and GATA-2. Embo J 23, 2841-2852.

Antes, T.J., Namciu, S.J., Fournier, R.E., and Levy-Wilson, B. (2001). The 5' boundary of the human apolipoprotein B chromatin domain in intestinal cells. Biochemistry 40, 6731-6742.

Aplan, P.D., Begley, C.G., Bertness, V., Nussmeier, M., Ezquerra, A., Coligan, J., and Kirsch, I.R. (1990). The SCL gene is formed from a transcriptionally complex locus. Molecular and cellular biology 10, 6426-6435.

Aplan, P.D., Lombardi, D.P., Ginsberg, A.M., Cossman, J., Bertness, V.L., and Kirsch, I.R. (1990). Disruption of the human SCL locus by "illegitimate" V-(D)-J recombinase activity. Science (New York, NY 250, 1426-1429.

Attema, J.L., Papathanasiou, P., Forsberg, E.C., Xu, J., Smale, S.T., and Weissman, I.L. (2007). Epigenetic characterization of hematopoietic stem cell differentiation using miniChIP and bisulfite sequencing analysis. Proceedings of the National Academy of Sciences of the United States of America 104, 12371-12376.

Aza-Blanc, P., Cooper, C.L., Wagner, K., Batalov, S., Deveraux, Q.L., and Cooke, M.P. (2003). Identification of modulators of TRAIL-induced apoptosis via RNAi-based phenotypic screening. Mol Cell 12, 627-637.

Baeuerle, P.A., and Baltimore, D. (1988). I kappa B: a specific inhibitor of the NF-kappa B transcription factor. Science (New York, NY 242, 540-546.

Bailey, T.L., and Elkan, C. (1995). The value of prior knowledge in discovering motifs with MEME. Proceedings / International Conference on Intelligent Systems for Molecular Biology ; ISMB 3, 21-29.

Bain, G., Maandag, E.C., Izon, D.J., Amsen, D., Kruisbeek, A.M., Weintraub, B.C., Krop, I., Schlissel, M.S., Feeney, A.J., van Roon, M., et al. (1994). E2A proteins are required for proper B cell development and initiation of immunoglobulin gene rearrangements. Cell 79, 885-892.

Bajic, V.B., Tan, S.L., Suzuki, Y., and Sugano, S. (2004). Promoter prediction analysis on the whole human genome. Nature biotechnology 22, 1467-1473.

Balduini, C.L., Pecci, A., Loffredo, G., Izzo, P., Noris, P., Grosso, M., Bergamaschi, G., Rosti, V., Magrini, U., Ceresa, I.F., et al. (2004). Effects of the R216Q mutation of GATA-1 on erythropoiesis and megakaryocytopoiesis. Thrombosis and haemostasis 91, 129-140.

Balhoff, J.P., and Wray, G.A. (2005). Evolutionary analysis of the well characterized endo16 promoter reveals substantial variation within functional sites. Proceedings of the National Academy of Sciences of the United States of America 102, 8591-8596.

Banerji, J., Olson, L., and Schaffner, W. (1983). A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. Cell 33, 729-740.

Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. Cell 27, 299-308.

Barndt, R.J., Dai, M., and Zhuang, Y. (2000). Functions of E2A-HEB heterodimers in T-cell development revealed by a dominant negative mutation of HEB. Molecular and cellular biology 20, 6677-6685.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell 129, 823-837.

Barski, A., and Frenkel, B. (2004). ChIP Display: novel method for identification of genomic targets of transcription factors. Nucleic acids research 32, e104.

Barton, G.M., and Medzhitov, R. (2002). Retroviral delivery of small interfering RNA into primary cells. Proceedings of the National Academy of Sciences of the United States of America 99, 14943-14945.

Basu, P., Morris, P.E., Haar, J.L., Wani, M.A., Lingrel, J.B., Gaensler, K.M., and Lloyd, J.A. (2005). KLF2 is essential for primitive erythropoiesis and regulates the human and murine embryonic beta-like globin genes in vivo. Blood 106, 2566-2571.

Becker, D.M., Fikes, J.D., and Guarente, L. (1991). A cDNA encoding a human CCAAT-binding protein cloned by functional complementation in yeast. Proceedings of the National Academy of Sciences of the United States of America 88, 1968-1972.

Begley, C.G., Aplan, P.D., Davey, M.P., Nakahara, K., Tchorz, K., Kurtzberg, J., Hershfield, M.S., Haynes, B.F., Cohen, D.I., Waldmann, T.A., et al. (1989). Chromosomal translocation in a human leukemic stem-cell line disrupts the T-cell antigen receptor delta-chain diversity region and results in a previously unreported fusion transcript. Proceedings of the National Academy of Sciences of the United States of America 86, 2031-2035.

Begley, C.G., Aplan, P.D., Denning, S.M., Haynes, B.F., Waldmann, T.A., and Kirsch, I.R. (1989). The gene SCL is expressed during early hematopoiesis and encodes a differentiation-related DNA-binding motif. Proceedings of the National Academy of Sciences of the United States of America 86, 10128-10132.

Begley, C.G., Visvader, J., Green, A.R., Aplan, P.D., Metcalf, D., Kirsch, I.R., and Gough, N.M. (1991). Molecular cloning and chromosomal localization of the murine homolog of the human helix-loop-helix gene SCL. Proceedings of the National Academy of Sciences of the United States of America 88, 869-873.

Bell, A.C., West, A.G., and Felsenfeld, G. (1999). The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. Cell 98, 387-396.

Bennett, S. (2004). Solexa Ltd. Pharmacogenomics 5, 433-438.

Bernard, O., Azogui, O., Lecointe, N., Mugneret, F., Berger, R., Larsen, C.J., and Mathieu-Mahul, D. (1992). A third tal-1 promoter is specifically used in human T cell leukemias. The Journal of experimental medicine 176, 919-925.

Bernard, O., Lecointe, N., Jonveaux, P., Souyri, M., Mauchauffe, M., Berger, R., Larsen, C.J., and Mathieu-Mahul, D. (1991). Two site-specific deletions and t(1;14) translocation restricted to human T-cell acute leukemias disrupt the 5' part of the tal-1 gene. Oncogene 6, 1477-1488.

Berns, K., Hijmans, E.M., Mullenders, J., Brummelkamp, T.R., Velds, A., Heimerikx, M., Kerkhoven, R.M., Madiredjo, M., Nijkamp, W., Weigelt, B., et al. (2004). A large-scale RNAi screen in human cells identifies new components of the p53 pathway. Nature 428, 431-437.

Bernstein, B.E., Humphrey, E.L., Erlich, R.L., Schneider, R., Bouman, P., Liu, J.S., Kouzarides, T., and Schreiber, S.L. (2002). Methylation of histone H3 Lys 4 in coding regions of active genes. Proceedings of the National Academy of Sciences of the United States of America 99, 8695-8700.

Bernstein, B.E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D.K., Huebert, D.J., McMahon, S., Karlsson, E.K., Kulbokas, E.J., 3rd, Gingeras, T.R., et al. (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. Cell 120, 169-181.

Birmingham, A., Anderson, E.M., Reynolds, A., Ilsley-Tyree, D., Leake, D., Fedorov, Y., Baskerville, S., Maksimova, E., Robinson, K., Karpilow, J., et al. (2006). 3' UTR seed matches, but not overall identity, are associated with RNAi off-targets. Nature methods 3, 199-204.

Blais, A., Tsikitis, M., Acosta-Alvear, D., Sharan, R., Kluger, Y., and Dynlacht, B.D. (2005). An initial blueprint for myogenic differentiation. Genes & development 19, 553-569.

Blanchette, M., and Tompa, M. (2003). FootPrinter: A program designed for phylogenetic footprinting. Nucleic acids research 31, 3840-3842.

Bochar, D.A., Savard, J., Wang, W., Lafleur, D.W., Moore, P., Cote, J., and Shiekhattar, R. (2000). A family of chromatin remodeling factors related to Williams syndrome transcription factor. Proceedings of the National Academy of Sciences of the United States of America 97, 1038-1043.

Bockamp, E.O., Fordham, J.L., Gottgens, B., Murrell, A.M., Sanchez, M.J., and Green, A.R. (1998). Transcriptional regulation of the stem cell leukemia gene by PU.1 and Elf-1. The Journal of biological chemistry 273, 29032-29042.

Bockamp, E.O., McLaughlin, F., Gottgens, B., Murrell, A.M., Elefanty, A.G., and Green, A.R. (1997). Distinct mechanisms direct SCL/tal-1 expression in erythroid cells and CD34 positive primitive myeloid cells. The Journal of biological chemistry 272, 8781-8790.

Bode, J., Benham, C., Knopp, A., and Mielke, C. (2000). Transcriptional augmentation: modulation of gene expression by scaffold/matrix-attached regions (S/MAR elements). Critical reviews in eukaryotic gene expression 10, 73-90.

Bode, J., Goetze, S., Heng, H., Krawetz, S.A., and Benham, C. (2003). From DNA structure to gene expression: mediators of nuclear compartmentalization and dynamics. Chromosome Res 11, 435-445.

Boden, D., Pusch, O., Lee, F., Tucker, L., and Ramratnam, B. (2004). Efficient gene transfer of HIV-1-specific short hairpin RNA into human lymphocytic cells using recombinant adeno-associated virus vectors. Mol Ther 9, 396-402.

Boden, D., Pusch, O., Silbermann, R., Lee, F., Tucker, L., and Ramratnam, B. (2004). Enhanced gene silencing of HIV-1 specific siRNA using microRNA designed hairpins. Nucleic acids research 32, 1154-1158.

Boehm, T., Foroni, L., Kaneko, Y., Perutz, M.F., and Rabbitts, T.H. (1991). The rhombotin family of cysteine-rich LIM-domain oncogenes: distinct members are involved in T-cell translocations to human chromosomes 11p15 and 11p13. Proceedings of the National Academy of Sciences of the United States of America 88, 4367-4371.

Bose, F., Fugazza, C., Casalgrandi, M., Capelli, A., Cunningham, J.M., Zhao, Q., Jane, S.M., Ottolenghi, S., and Ronchi, A. (2006). Functional interaction of CP2 with GATA-1 in the regulation of erythroid promoters. Molecular and cellular biology 26, 3942-3954.

Boutros, M., Kiger, A.A., Armknecht, S., Kerr, K., Hild, M., Koch, B., Haas, S.A., Paro, R., and Perrimon, N. (2004). Genome-wide RNAi analysis of growth and viability in Drosophila cells. Science (New York, NY 303, 832-835.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., et al. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell 122, 947-956.

Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K., et al. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature 441, 349-353.

Boyes, J., Byfield, P., Nakatani, Y., and Ogryzko, V. (1998). Regulation of activity of the transcription factor GATA-1 by acetylation. Nature 396, 594-598.

Boyle, E.I., Weng, S., Gollub, J., Jin, H., Botstein, D., Cherry, J.M., and Sherlock, G. (2004). GO::TermFinder--open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. Bioinformatics (Oxford, England) 20, 3710-3715.

Brandenberger, R., Khrebtukova, I., Thies, R.S., Miura, T., Jingli, C., Puri, R., Vasicek, T., Lebkowski, J., and Rao, M. (2004). MPSS profiling of human embryonic stem cells. BMC developmental biology 4, 10.

Brass, A.L., Dykxhoorn, D.M., Benita, Y., Yan, N., Engelman, A., Xavier, R.J., Lieberman, J., and Elledge, S.J. (2008). Identification of host proteins required for HIV infection through a functional genomic screen. Science (New York, NY 319, 921-926.

Brehm, A., Miska, E.A., McCance, D.J., Reid, J.L., Bannister, A.J., and Kouzarides, T. (1998). Retinoblastoma protein recruits histone deacetylase to repress transcription. Nature 391, 597-601.

Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D.H., Johnson, D., Luo, S., McCurdy, S., Foy, M., Ewan, M., et al. (2000). Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nature biotechnology 18, 630-634.

Brenner, S., Williams, S.R., Vermaas, E.H., Storck, T., Moon, K., McCollum, C., Mao, J.I., Luo, S., Kirchner, J.J., Eletr, S., et al. (2000). In vitro cloning of complex mixtures of DNA on microbeads: physical separation of differentially expressed cDNAs. Proceedings of the National Academy of Sciences of the United States of America 97, 1665-1670.

Brent, R., and Ptashne, M. (1985). A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. Cell 43, 729-736.

Bridge, A.J., Pebernard, S., Ducraux, A., Nicoulaz, A.L., and Iggo, R. (2003). Induction of an interferon response by RNAi vectors in mammalian cells. Nature genetics 34, 263-264.

Brudno, M., Do, C.B., Cooper, G.M., Kim, M.F., Davydov, E., Green, E.D., Sidow, A., and Batzoglou, S. (2003). LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. Genome research 13, 721-731.

Brummelkamp, T.R., Bernards, R., and Agami, R. (2002). A system for stable expression of short interfering RNAs in mammalian cells. Science (New York, NY 296, 550-553.

Brummelkamp, T.R., Nijman, S.M., Dirac, A.M., and Bernards, R. (2003). Loss of the cylindromatosis tumour suppressor inhibits apoptosis by activating NF-kappaB. Nature 424, 797-801.

Cai, J., Shin, S., Wright, L., Liu, Y., Zhou, D., Xue, H., Khrebtukova, I., Mattson, M.P., Svendsen, C.N., and Rao, M.S. (2006). Massively parallel signature sequencing profiling of fetal human neural precursor cells. Stem cells and development 15, 232-244.

Cao, H.B., Wang, A., Martin, B., Koehler, D.R., Zeitlin, P.L., Tanawell, A.K., and Hu, J. (2005). Down-regulation of IL-8 expression in human airway epithelial cells through helper-dependent adenoviral-mediated RNA interference. Cell research 15, 111-119.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science (New York, NY 298, 1039-1043.

Caplen, N.J., Parrish, S., Imani, F., Fire, A., and Morgan, R.A. (2001). Specific inhibition of gene expression by small double-stranded RNAs in invertebrate and vertebrate systems. Proceedings of the National Academy of Sciences of the United States of America 98, 9742-9747.

Capron C, Lécluse Y, Kaushik AL, Foudi A, Lacout C, Sekkai D, Godin I, Albagli O, Poullion I, Svinartchouk F, Schanze E, Vainchenker W, Sablitzky F, Bennaceur-Griscelli A, Duménil D. (2006). The SCL relative LYL-1 is required for fetal and adult hematopoietic stem cell function and B-cell differentiation. Blood 107, 4678-4686.

Cawley, S., Bekiranov, S., Ng, H.H., Kapranov, P., Sekinger, E.A., Kampa, D., Piccolboni, A., Sementchenko, V., Cheng, J., Williams, A.J., et al. (2004). Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. Cell 116, 499-509.

Chambeyron, S., and Bickmore, W.A. (2004). Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. Genes & development 18, 1119-1130.

Chan, W.C., and Nie, S. (1998). Quantum dot bioconjugates for ultrasensitive nonisotopic detection. Science (New York, NY 281, 2016-2018.

Chan, W.Y., Follows, G.A., Lacaud, G., Pimanda, J.E., Landry, J.R., Kinston, S., Knezevic, K., Piltz, S., Donaldson, I.J., Gambardella, L., et al. (2007). The paralogous hematopoietic regulators Lyl1 and Scl are

coregulated by Ets and GATA factors, but Lyl1 cannot rescue the early Scl-/- phenotype. Blood 109, 1908-1916.

Chen, A.A., Derfus, A.M., Khetani, S.R., and Bhatia, S.N. (2005). Quantum dots to monitor RNAi delivery and improve gene silencing. Nucleic acids research 33, e190.

Chen, C.C., Ko, T.M., Ma, H.I., Wu, H.L., Xiao, X., Li, J., Chang, C.M., Wu, P.Y., Chen, C.H., Han, J.M., et al. (2007). Long-term inhibition of hepatitis B virus in transgenic mice by double-stranded adeno-associated virus 8-delivered short hairpin RNA. Gene therapy 14, 11-19.

Chen, J.D., and Evans, R.M. (1995). A transcriptional co-repressor that interacts with nuclear hormone receptors. Nature 377, 454-457.

Chen, L., and Widom, J. (2005). Mechanism of transcriptional silencing in yeast. Cell 120, 37-48.

Chen, Y.T., Scanlan, M.J., Venditti, C.A., Chua, R., Theiler, G., Stevenson, B.J., Iseli, C., Gure, A.O., Vasicek, T., Strausberg, R.L., et al. (2005). Identification of cancer/testis-antigen genes by massively parallel signature sequencing. Proceedings of the National Academy of Sciences of the United States of America 102, 7940-7945.

Choi, I., Cho, B.R., Kim, D., Miyagawa, S., Kubo, T., Kim, J.Y., Park, C.G., Hwang, W.S., Lee, J.S., and Ahn, C. (2005). Choice of the adequate detection time for the accurate evaluation of the efficiency of siRNA-induced gene silencing. Journal of biotechnology 120, 251-261.

Choi, K., Kennedy, M., Kazarov, A., Papadimitriou, J.C., and Keller, G. (1998). A common precursor for hematopoietic and endothelial cells. Development (Cambridge, England) 125, 725-732.

Chung, Y.S., Zhang, W.J., Arentson, E., Kingsley, P.D., Palis, J., and Choi, K. (2002). Lineage analysis of the hemangioblast as defined by FLK1 and SCL expression. Development (Cambridge, England) 129, 5511-5520.

Clemens, J.C., Worby, C.A., Simonson-Leff, N., Muda, M., Maehama, T., Hemmings, B.A., and Dixon, J.E. (2000). Use of double-stranded RNA interference in Drosophila cell lines to dissect signal transduction pathways. Proceedings of the National Academy of Sciences of the United States of America 97, 6499-6503.

Cohen, B.A., Mitra, R.D., Hughes, J.D., and Church, G.M. (2000). A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. Nature genetics 26, 183-186.

Coller, H.A., Grandori, C., Tamayo, P., Colbert, T., Lander, E.S., Eisenman, R.N., and Golub, T.R. (2000). Expression analysis with oligonucleotide microarrays reveals that MYC regulates genes involved in growth, cell cycle, signaling, and adhesion. Proceedings of the National Academy of Sciences of the United States of America 97, 3260-3265.

Corey, D.R. (2007). Chemical modification: the key to clinical application of RNA interference? The Journal of clinical investigation 117, 3615-3622.

Courey, A.J., and Tjian, R. (1988). Analysis of Sp1 in vivo reveals multiple transcriptional domains, including a novel glutamine-rich activation motif. Cell 55, 887-898.

Crawford, G.E., Davis, S., Scacheri, P.C., Renaud, G., Halawi, M.J., Erdos, M.R., Green, R., Meltzer, P.S., Wolfsberg, T.G., and Collins, F.S. (2006). DNase-chip: a high-resolution method to identify DNase I hypersensitive sites using tiled microarrays. Nature methods 3, 503-509.

Crawford, G.E., Holt, I.E., Mullikin, J.C., Tai, D., Blakesley, R., Bouffard, G., Young, A., Masiello, C., Green, E.D., Wolfsberg, T.G., et al. (2004). Identifying gene regulatory elements by genome-wide recovery of DNase hypersensitive sites. Proceedings of the National Academy of Sciences of the United States of America 101, 992-997.

Crawford, G.E., Holt, I.E., Whittle, J., Webb, B.D., Tai, D., Davis, S., Margulies, E.H., Chen, Y., Bernat, J.A., Ginsburg, D., et al. (2006). Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). Genome research 16, 123-131.

Crossley, M., Tsang, A.P., Bieker, J.J., and Orkin, S.H. (1994). Regulation of the erythroid Kruppel-like factor (EKLF) gene promoter by the erythroid transcription factor GATA-1. The Journal of biological chemistry 269, 15440-15444.

Curtis, D.J., Hall, M.A., Van Stekelenburg, L.J., Robb, L., Jane, S.M., and Begley, C.G. (2004). SCL is required for normal function of short-term repopulating hematopoietic stem cells. Blood 103, 3342-3348.

Cuthbert, G.L., Daujat, S., Snowden, A.W., Erdjument-Bromage, H., Hagiwara, T., Yamada, M., Schneider, R., Gregory, P.D., Tempst, P., Bannister, A.J., et al. (2004). Histone deimination antagonizes arginine methylation. Cell 118, 545-553.

Dahl, J.A., and Collas, P. (2007). Q2ChIP, a quick and quantitative chromatin immunoprecipitation assay, unravels epigenetic dynamics of developmentally regulated genes in human carcinoma cells. Stem Cells 25, 1037-1046.

Dahmus, M.E. (1995). Phosphorylation of the C-terminal domain of RNA polymerase II. Biochimica et biophysica acta 1261, 171-182.

Daniel, J.A., Torok, M.S., Sun, Z.W., Schieltz, D., Allis, C.D., Yates, J.R., 3rd, and Grant, P.A. (2004). Deubiquitination of histone H2B by a yeast acetyltransferase complex regulates transcription. The Journal of biological chemistry 279, 1867-1871.

Das, P.M., Ramachandran, K., vanWert, J., and Singal, R. (2004). Chromatin immunoprecipitation assay. BioTechniques 37, 961-969.

Davuluri, R.V., Grosse, I., and Zhang, M.Q. (2001). Computational identification of promoters and first exons in the human genome. Nature genetics 29, 412-417.

De Gobbi, M., Anguita, E., Hughes, J., Sloane-Stanley, J.A., Sharpe, J.A., Koch, C.M., Dunham, I., Gibbons, R.J., Wood, W.G., and Higgs, D.R. (2007). Tissue-specific histone modification and transcription factor binding in alpha globin gene expression. Blood 110, 4503-4510.

de Jonge, J., Holtrop, M., Wilschut, J., and Huckriede, A. (2006). Reconstituted influenza virus envelopes as an efficient carrier system for cellular delivery of small-interfering RNAs. Gene therapy 13, 400-411.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. Science (New York, NY 295, 1306-1311.

Delabesse, E., Ogilvy, S., Chapman, M.A., Piltz, S.G., Gottgens, B., and Green, A.R. (2005). Transcriptional regulation of the SCL locus: identification of an enhancer that targets the primitive erythroid lineage in vivo. Molecular and cellular biology 25, 5215-5225.

Delassus, S., Titley, I., and Enver, T. (1999). Functional and molecular analysis of hematopoietic progenitors derived from the aorta-gonad-mesonephros region of the mouse embryo. Blood 94, 1495-1503.

DeRisi, J.L., Iyer, V.R., and Brown, P.O. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. Science (New York, NY 278, 680-686.

Dermitzakis, E.T., and Clark, A.G. (2002). Evolution of transcription factor binding sites in Mammalian gene regulatory regions: conservation and turnover. Molecular biology and evolution 19, 1114-1121.

Dhami, P. (2005). The SCL gene and transcriptional control of haematopoiesis. PhD Thesis, University of Cambridge.

Dhami, P., Coffey, A.J., Abbs, S., Vermeesch, J.R., Dumanski, J.P., Woodward, K.J., Andrews, R.M., Langford, C., and Vetrie, D. (2005). Exon array CGH: detection of copy-number changes at the resolution of individual exons in the human genome. Am J Hum Genet 76, 750-762.

Diakos, C., Krapf, G., Gerner, C., Inthal, A., Lemberger, C., Ban, J., Dohnal, A.M., and Panzer-Gruemayer, E.R. (2007). RNAi-mediated silencing of TEL/AML1 reveals a heat-shock protein- and survivin-dependent mechanism for survival. Blood 109, 2607-2610.

Dickins, R.A., Hemann, M.T., Zilfou, J.T., Simpson, D.R., Ibarra, I., Hannon, G.J., and Lowe, S.W. (2005). Probing tumor phenotypes using stable and regulated synthetic microRNA precursors. Nature genetics 37, 1289-1295.

Dobbin KK, Kawasaki ES, Petersen DW, Simon RM. (2005). Characterizing dye bias in microarray experiments. Bioinformatics 21, 2430-2437.

Dorschner, M.O., Hawrylycz, M., Humbert, R., Wallace, J.C., Shafer, A., Kawamoto, J., Mack, J., Hall, R., Goldy, J., Sabo, P.J., et al. (2004). High-throughput localization of functional elements by quantitative chromatin profiling. Nature methods 1, 219-225.

Down, T.A., and Hubbard, T.J. (2002). Computational detection and location of transcription start sites in mammalian genomic DNA. Genome research 12, 458-461.

Down, T.A., and Hubbard, T.J. (2005). NestedMICA: sensitive inference of over-represented motifs in nucleic acid sequence. Nucleic acids research 33, 1445-1453.

D'Souza, S.L., Elefanty, A.G., and Keller, G. (2005). SCL/Tal-1 is essential for hematopoietic commitment of the hemangioblast but not for its development. Blood 105, 3862-3870.

Echeverri, C.J., Beachy, P.A., Baum, B., Boutros, M., Buchholz, F., Chanda, S.K., Downward, J., Ellenberg, J., Fraser, A.G., Hacohen, N., et al. (2006). Minimizing the risk of reporting false positives in large-scale RNAi screens. Nature methods 3, 777-779.

Eisbacher, M., Holmes, M.L., Newton, A., Hogg, P.J., Khachigian, L.M., Crossley, M., and Chong, B.H. (2003). Protein-protein interaction between Fli-1 and GATA-1 mediates synergistic expression of megakaryocyte-specific genes through cooperative DNA binding. Molecular and cellular biology 23, 3427-3441.

Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. Proceedings of the National Academy of Sciences of the United States of America 95, 14863-14868.

Elbashir, S.M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., and Tuschl, T. (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature 411, 494-498.

Elbashir, S.M., Harborth, J., Weber, K., and Tuschl, T. (2002). Analysis of gene function in somatic mammalian cells using small interfering RNAs. Methods (San Diego, Calif 26, 199-213.

Elefanty, A.G., Begley, C.G., Hartley, L., Papaevangeliou, B., and Robb, L. (1999). SCL expression in the mouse embryo detected with a targeted lacZ reporter gene demonstrates its localization to hematopoietic, vascular, and neural tissues. Blood 94, 3754-3763.

Elefanty, A.G., Robb, L., Birner, R., and Begley, C.G. (1997). Hematopoietic-specific genes are not induced during in vitro differentiation of scl-null embryonic stem cells. Blood 90, 1435-1447.

Elkon, R., Rashi-Elkeles, S., Lerenthal, Y., Linhart, C., Tenne, T., Amariglio, N., Rechavi, G., Shamir, R., and Shiloh, Y. (2005). Dissection of a DNA-damage-induced transcriptional network using a combination of microarrays, RNA interference and computational promoter analysis. Genome biology 6, R43.

Elwood, N.J., Green, A.R., Melder, A., Begley, C.G., and Nicola, N. (1994). The SCL protein displays cell-specific heterogeneity in size. Leukemia 8, 106-114.

Endoh, M., Ogawa, M., Orkin, S., and Nishikawa, S. (2002). SCL/tal-1-dependent process determines a competence to select the definitive hematopoietic lineage prior to endothelial differentiation. Embo J 21, 6700-6708.

Euskirchen, G., Royce, T.E., Bertone, P., Martone, R., Rinn, J.L., Nelson, F.K., Sayward, F., Luscombe, N.M., Miller, P., Gerstein, M., et al. (2004). CREB binds to multiple loci on human chromosome 22. Molecular and cellular biology 24, 3804-3814.

Evans, T., Reitman, M., and Felsenfeld, G. (1988). An erythrocyte-specific DNA-binding factor recognizes a regulatory sequence common to all chicken globin genes. Proceedings of the National Academy of Sciences of the United States of America 85, 5976-5980.

Fedorov, Y., Anderson, E.M., Birmingham, A., Reynolds, A., Karpilow, J., Robinson, K., Leake, D., Marshall, W.S., and Khvorova, A. (2006). Off-target effects by siRNA can induce toxic phenotype. RNA (New York, NY 12, 1188-1196.

Ferrando, A.A., Neuberg, D.S., Staunton, J., Loh, M.L., Huard, C., Raimondi, S.C., Behm, F.G., Pui, C.H., Downing, J.R., Gilliland, D.G., et al. (2002). Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. Cancer cell 1, 75-87.

Ferreira, R., Ohneda, K., Yamamoto, M., and Philipsen, S. (2005). GATA1 function, a paradigm for transcription factors in hematopoiesis. Molecular and cellular biology 25, 1215-1227.

Fire, A., Albertson, D., Harrison, S.W., and Moerman, D.G. (1991). Production of antisense RNA leads to effective and specific inhibition of gene expression in C. elegans muscle. Development (Cambridge, England) 113, 503-514.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature 391, 806-811.

Follows, G.A., Dhami, P., Gottgens, B., Bruce, A.W., Campbell, P.J., Dillon, S.C., Smith, A.M., Koch, C., Donaldson, I.J., Scott, M.A., et al. (2006). Identifying gene regulatory elements by genomic microarray mapping of DNaseI hypersensitive sites. Genome research 16, 1310-1319.

Follows, G.A., Janes, M.E., Vallier, L., Green, A.R., and Gottgens, B. (2007). Real-time PCR mapping of DNaseI-hypersensitive sites using a novel ligation-mediated amplification technique. Nucleic acids research 35, e56.

Fordham, J.L., Gottgens, B., McLaughlin, F., and Green, A.R. (1999). Chromatin structure and transcriptional regulation of the stem cell leukaemia (SCL) gene in mast cells. Leukemia 13, 750-759.

Fox, A.H., Liew, C., Holmes, M., Kowalski, K., Mackay, J., and Crossley, M. (1999). Transcriptional cofactors of the FOG family interact with GATA proteins by means of multiple zinc fingers. Embo J 18, 2812-2822.

Freeman, L.A., and Garrard, W.T. (1992). DNA supercoiling in chromatin structure and gene expression. Critical reviews in eukaryotic gene expression 2, 165-209.

Freson, K., Devriendt, K., Matthijs, G., Van Hoof, A., De Vos, R., Thys, C., Minner, K., Hoylaerts, M.F., Vermylen, J., and Van Geet, C. (2001). Platelet characteristics in patients with X-linked macrothrombocytopenia because of a novel GATA1 mutation. Blood 98, 85-92.

Freson, K., Matthijs, G., Thys, C., Marien, P., Hoylaerts, M.F., Vermylen, J., and Van Geet, C. (2002). Different substitutions at residue D218 of the X-linked transcription factor GATA1 lead to altered clinical severity of macrothrombocytopenia and anemia and are associated with variable skewed X inactivation. Human molecular genetics 11, 147-152.

Fujiwara, Y., Browne, C.P., Cunniff, K., Goff, S.C., and Orkin, S.H. (1996). Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. Proceedings of the National Academy of Sciences of the United States of America 93, 12355-12358.

Galas, D.J., and Schmitz, A. (1978). DNAse footprinting: a simple method for the detection of protein-DNA binding specificity. Nucleic acids research 5, 3157-3170.

Gao, X., Cui, Y., Levenson, R.M., Chung, L.W., and Nie, S. (2004). In vivo cancer targeting and imaging with semiconductor quantum dots. Nature biotechnology 22, 969-976.

Garner, M.M., and Revzin, A. (1981). A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system. Nucleic acids research 9, 3047-3060.

Ge, Y., Dombkowski, A.A., LaFiura, K.M., Tatman, D., Yedidi, R.S., Stout, M.L., Buck, S.A., Massey, G., Becton, D.L., Weinstein, H.J., et al. (2006). Differential gene expression, GATA1 target genes, and the chemotherapy sensitivity of Down syndrome megakaryocytic leukemia. Blood 107, 1570-1581.

Gelmetti, V., Zhang, J., Fanelli, M., Minucci, S., Pelicci, P.G., and Lazar, M.A. (1998). Aberrant recruitment of the nuclear receptor corepressor-histone deacetylase complex by the acute myeloid leukemia fusion partner ETO. Molecular and cellular biology 18, 7185-7191.

Gershenzon, N.I., and Ioshikhes, I.P. (2005). Synergy of human Pol II core promoter elements revealed by statistical sequence analysis. Bioinformatics (Oxford, England) 21, 1295-1300.

Gerton, J.L., DeRisi, J., Shroff, R., Lichten, M., Brown, P.O., and Petes, T.D. (2000). Inaugural article: global mapping of meiotic recombination hotspots and coldspots in the yeast Saccharomyces cerevisiae. Proceedings of the National Academy of Sciences of the United States of America 97, 11383-11390.

Giroux, S., Kaushik, A.L., Capron, C., Jalil, A., Kelaidi, C., Sablitzky, F., Dumenil, D., Albagli, O., and Godin, I. (2007). lyl-1 and tal-1/scl, two genes encoding closely related bHLH transcription factors, display highly overlapping expression patterns during cardiovascular and hematopoietic ontogeny. Gene Expr Patterns 7, 215-226.

Goardon, N., Lambert, J.A., Rodriguez, P., Nissaire, P., Herblot, S., Thibault, P., Dumenil, D., Strouboulis, J., Romeo, P.H., and Hoang, T. (2006). ETO2 coordinates cellular proliferation and differentiation during erythropoiesis. Embo J 25, 357-366.

Godin, I., and Cumano, A. (2002). The hare and the tortoise: an embryonic haematopoietic race. Nat Rev Immunol 2, 593-604.

Godin, I., Garcia-Porrero, J.A., Dieterlen-Lievre, F., and Cumano, A. (1999). Stem cell emergence and hemopoietic activity are incompatible in mouse intraembryonic sites. The Journal of experimental medicine 190, 43-52.

Goldfarb, A.N., Goueli, S., Mickelson, D., and Greenberg, J.M. (1992). T-cell acute lymphoblastic leukemia--the associated gene SCL/tal codes for a 42-Kd nuclear phosphoprotein. Blood 80, 2858-2866.

Gonczy, P., Echeverri, C., Oegema, K., Coulson, A., Jones, S.J., Copley, R.R., Duperon, J., Oegema, J., Brehm, M., Cassin, E., et al. (2000). Functional genomic analysis of cell division in C. elegans using RNAi of genes on chromosome III. Nature 408, 331-336.

Gonzalez, G.A., and Montminy, M.R. (1989). Cyclic AMP stimulates somatostatin gene transcription by phosphorylation of CREB at serine 133. Cell 59, 675-680.

Gottgens, B., Barton, L.M., Gilbert, J.G., Bench, A.J., Sanchez, M.J., Bahn, S., Mistry, S., Grafham, D., McMurray, A., Vaudin, M., et al. (2000). Analysis of vertebrate SCL loci identifies conserved enhancers. Nature biotechnology 18, 181-186.

Gottgens, B., Broccardo, C., Sanchez, M.J., Deveaux, S., Murphy, G., Gothert, J.R., Kotsopoulou, E., Kinston, S., Delaney, L., Piltz, S., et al. (2004). The scl +18/19 stem cell enhancer is not required for hematopoiesis: identification of a 5' bifunctional hematopoietic-endothelial enhancer bound by Fli-1 and Elf-1. Molecular and cellular biology 24, 1870-1883.

Gottgens, B., McLaughlin, F., Bockamp, E.O., Fordham, J.L., Begley, C.G., Kosmopoulos, K., Elefanty, A.G., and Green, A.R. (1997). Transcription of the SCL gene in erythroid and CD34 positive primitive myeloid cells is controlled by a complex network of lineage-restricted chromatin-dependent and chromatin-independent regulatory elements. Oncogene 15, 2419-2428.

Gottgens, B., Nastos, A., Kinston, S., Piltz, S., Delabesse, E.C., Stanley, M., Sanchez, M.J., Ciau-Uitz, A., Patient, R., and Green, A.R. (2002). Establishing the transcriptional programme for blood: the SCL stem cell enhancer is regulated by a multiprotein complex containing Ets and GATA factors. Embo J 21, 3039-3050.

Gou, D., Jin, N., and Liu, L. (2003). Gene silencing in mammalian cells by PCR-based short hairpin RNA. FEBS letters 548, 113-118.

Grant, P.A., Eberharter, A., John, S., Cook, R.G., Turner, B.M., and Workman, J.L. (1999). Expanded lysine acetylation specificity of Gcn5 in native complexes. The Journal of biological chemistry 274, 5895-5900.

Grass, J.A., Boyer, M.E., Pal, S., Wu, J., Weiss, M.J., and Bresnick, E.H. (2003). GATA-1-dependent transcriptional repression of GATA-2 via disruption of positive autoregulation and domain-wide chromatin remodeling. Proceedings of the National Academy of Sciences of the United States of America 100, 8811-8816.

Green, A.R., Lints, T., Visvader, J., Harvey, R., and Begley, C.G. (1992). SCL is coexpressed with GATA-1 in hemopoietic cells but is also expressed in developing brain. Oncogene 7, 653-660.

Green, A.R., Rockman, S., DeLuca, E., and Begley, C.G. (1993). Induced myeloid differentiation of K562 cells with downregulation of erythroid and megakaryocytic transcription factors: a novel experimental model for hemopoietic lineage restriction. Exp Hematol 21, 525-531.

Green, A.R., Salvaris, E., and Begley, C.G. (1991). Erythroid expression of the 'helix-loop-helix' gene, SCL. Oncogene 6, 475-479.

Greenbaum, S., and Zhuang, Y. (2002). Identification of E2A target genes in B lymphocyte development by using a gene tagging-based chromatin immunoprecipitation system. Proceedings of the National Academy of Sciences of the United States of America 99, 15030-15035.

Gregory, R.C., Taxman, D.J., Seshasayee, D., Kensinger, M.H., Bieker, J.J., and Wojchowski, D.M. (1996). Functional interaction of GATA1 with erythroid Kruppel-like factor and Sp1 at defined erythroid promoters. Blood 87, 1793-1801.

Grimm, D., Streetz, K.L., Jopling, C.L., Storm, T.A., Pandey, K., Davis, C.R., Marion, P., Salazar, F., and Kay, M.A. (2006). Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways. Nature 441, 537-541.

Gross, D.S., and Garrard, W.T. (1988). Nuclease hypersensitive sites in chromatin. Annu Rev Biochem 57, 159-197.

Grosveld, F., van Assendelft, G.B., Greaves, D.R., and Kollias, G. (1987). Position-independent, high-level expression of the human beta-globin gene in transgenic mice. Cell 51, 975-985.

Grutz, G.G., Bucher, K., Lavenir, I., Larson, T., Larson, R., and Rabbitts, T.H. (1998). The oncogenic T cell LIM-protein Lmo2 forms part of a DNA-binding complex specifically in immature T cells. Embo J 17, 4594-4605.

Gu, W., and Roeder, R.G. (1997). Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain. Cell 90, 595-606.

Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. Cell 130, 77-88.

Gupta, S., Schoer, R.A., Egan, J.E., Hannon, G.J., and Mittal, V. (2004). Inducible, reversible, and stable RNA interference in mammalian cells. Proceedings of the National Academy of Sciences of the United States of America 101, 1927-1932.

Gyorffy, B., Surowiak, P., Kiesslich, O., Denkert, C., Schafer, R., Dietel, M., and Lage, H. (2005). Gene expression profiling of 30 cancer cell lines predicts resistance towards 11 anticancer drugs at clinically achieved concentrations. Int J Cancer.

Hahn, S. (2004). Structure and mechanism of the RNA polymerase II transcription machinery. Nat Struct Mol Biol 11, 394-403.

Hamlett I, Draper J, Strouboulis J, Iborra F, Porcher C, Vyas P. Characterization of megakaryocyte GATA1-interacting proteins: the corepressor ETO2 and GATA1 interact to regulate terminal megakaryocyte maturation. (2008). Blood 112, 2738-2749.

Hall, M.A., Curtis, D.J., Metcalf, D., Elefanty, A.G., Sourris, K., Robb, L., Gothert, J.R., Jane, S.M., and Begley, C.G. (2003). The critical regulator of embryonic hematopoiesis, SCL, is vital in the adult for megakaryopoiesis, erythropoiesis, and lineage choice in CFU-S12. Proceedings of the National Academy of Sciences of the United States of America 100, 992-997.

Hall, M.A., Slater, N.J., Begley, C.G., Salmon, J.M., Van Stekelenburg, L.J., McCormack, M.P., Jane, S.M., and Curtis, D.J. (2005). Functional but abnormal adult erythropoiesis in the absence of the stem cell leukemia gene. Molecular and cellular biology 25, 6355-6362.

Hao, H., Liu, H., Gonye, G., and Schwaber, J.S. (2008). A fast carrier chromatin immunoprecipitation method applicable to microdissected tissue samples. Journal of neuroscience methods 172, 38-42.

Harris, M.B., Mostecki, J., and Rothman, P.B. (2005). Repression of an interleukin-4-responsive promoter requires cooperative BCL-6 function. The Journal of biological chemistry 280, 13114-13121.

Hassa, P.O., Haenni, S.S., Elser, M., and Hottiger, M.O. (2006). Nuclear ADP-ribosylation reactions in mammalian cells: where are we today and where are we going? Microbiol Mol Biol Rev 70, 789-829.

Hassan, A.H., Neely, K.E., and Workman, J.L. (2001). Histone acetyltransferase complexes stabilize swi/snf binding to promoter nucleosomes. Cell 104, 817-827.

Heemskerk, M.H., Blom, B., Nolan, G., Stegmann, A.P., Bakker, A.Q., Weijer, K., Res, P.C., and Spits, H. (1997). Inhibition of T cell and promotion of natural killer cell development by the dominant negative helix loop helix factor Id3. The Journal of experimental medicine 186, 1597-1602.

Heinemeyer, T., Wingender, E., Reuter, I., Hermjakob, H., Kel, A.E., Kel, O.V., Ignatieva, E.V., Ananko, E.A., Podkolodnaya, O.A., Kolpakov, F.A., et al. (1998). Databases on transcriptional regulation: TRANSFAC, TRRD and COMPEL. Nucleic acids research 26, 362-367.

Heinzel, T., Lavinsky, R.M., Mullen, T.M., Soderstrom, M., Laherty, C.D., Torchia, J., Yang, W.M., Brard, G., Ngo, S.D., Davie, J.R., et al. (1997). A complex containing N-CoR, mSin3 and histone deacetylase mediates transcriptional repression. Nature 387, 43-48.

Hirasawa, R., Shimizu, R., Takahashi, S., Osawa, M., Takayanagi, S., Kato, Y., Onodera, M., Minegishi, N., Yamamoto, M., Fukao, K., et al. (2002). Essential and instructive roles of GATA factors in eosinophil development. The Journal of experimental medicine 195, 1379-1386.

Holen, T., Amarzguioui, M., Wiiger, M.T., Babaie, E., and Prydz, H. (2002). Positional effects of short interfering RNAs targeting the human coagulation trigger Tissue Factor. Nucleic acids research 30, 1757-1766.

Holmes, M., Turner, J., Fox, A., Chisholm, O., Crossley, M., and Chong, B. (1999). hFOG-2, a novel zinc finger protein, binds the co-repressor mCtBP2 and modulates GATA-mediated activation. The Journal of biological chemistry 274, 23491-23498.

Hope, I.A., Mahadevan, S., and Struhl, K. (1988). Structural and functional characterization of the short acidic transcriptional activation region of yeast GCN4 protein. Nature 333, 635-640.

Hoppe, K.L., and Francone, O.L. (1998). Binding and functional effects of transcription factors Sp1 and Sp3 on the proximal human lecithin:cholesterol acyltransferase promoter. Journal of lipid research 39, 969-977.

Horak, C.E., Mahajan, M.C., Luscombe, N.M., Gerstein, M., Weissman, S.M., and Snyder, M. (2002). GATA-1 binding sites mapped in the beta-globin locus by using mammalian chIp-chip analysis. Proceedings of the National Academy of Sciences of the United States of America 99, 2924-2929.

Horner, M.A., Quintin, S., Domeier, M.E., Kimble, J., Labouesse, M., and Mango, S.E. (1998). pha-4, an HNF-3 homolog, specifies pharyngeal organ identity in Caenorhabditis elegans. Genes & development 12, 1947-1952.

Hsieh, A.C., Bo, R., Manola, J., Vazquez, F., Bare, O., Khvorova, A., Scaringe, S., and Sellers, W.R. (2004). A library of siRNA duplexes targeting the phosphoinositide 3-kinase pathway: determinants of gene silencing for use in cell-based screens. Nucleic acids research 32, 893-901.

Hsu, H.L., Cheng, J.T., Chen, Q., and Baer, R. (1991). Enhancer-binding activity of the tal-1 oncoprotein in association with the E47/E12 helix-loop-helix proteins. Molecular and cellular biology 11, 3037-3042.

Hsu, H.L., Huang, L., Tsan, J.T., Funk, W., Wright, W.E., Hu, J.S., Kingston, R.E., and Baer, R. (1994). Preferred sequences for DNA recognition by the TAL1 helix-loop-helix proteins. Molecular and cellular biology 14, 1256-1265.

Hsu, H.L., Wadman, I., and Baer, R. (1994). Formation of in vivo complexes between the TAL1 and E2A polypeptides of leukemic T cells. Proceedings of the National Academy of Sciences of the United States of America 91, 3181-3185.

Hsu, H.L., Wadman, I., Tsan, J.T., and Baer, R. (1994). Positive and negative transcriptional control by the TAL1 helix-loop-helix protein. Proceedings of the National Academy of Sciences of the United States of America 91, 5947-5951.

Huang, D.Y., Kuo, Y.Y., and Chang, Z.F. (2005). GATA-1 mediates auto-regulation of Gfi-1B transcription in K562 cells. Nucleic acids research 33, 5331-5342.

Huang, D.Y., Kuo, Y.Y., Lai, J.S., Suzuki, Y., Sugano, S., and Chang, Z.F. (2004). GATA-1 and NF-Y cooperate to mediate erythroid-specific transcription of Gfi-1B gene. Nucleic acids research 32, 3935-3946.

Hughes, T.R., Mao, M., Jones, A.R., Burchard, J., Marton, M.J., Shannon, K.W., Lefkowitz, S.M., Ziman, M., Schelter, J.M., Meyer, M.R., et al. (2001). Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. Nature biotechnology 19, 342-347.

Huo, X.F., Yu, J., Peng, H., Du, Z.W., Liu, X.L., Ma, Y.N., Zhang, X., Zhang, Y., Zhao, H.L., and Zhang, J.W. (2006). Differential expression changes in K562 cells during the hemin-induced erythroid differentiation and the phorbol myristate acetate (PMA)-induced megakaryocytic differentiation. Molecular and cellular biochemistry 292, 155-167.

Hwang, L.Y., Siegelman, M., Davis, L., Oppenheimer-Marks, N., and Baer, R. (1993). Expression of the TAL1 proto-oncogene in cultured endothelial cells and blood vessels of the spleen. Oncogene 8, 3043-3046.

Ikawa, T., Kawamoto, H., Wright, L.Y., and Murre, C. (2004). Long-term cultured E2A-deficient hematopoietic progenitor cells are pluripotent. Immunity 20, 349-360.

Imagawa, S., Suzuki, N., Ohmine, K., Obara, N., Mukai, H.Y., Ozawa, K., Yamamoto, M., and Nagasawa, T. (2002). GATA suppresses erythropoietin gene expression through GATA site in mouse erythropoietin gene promoter. International journal of hematology 75, 376-381.

Imagawa, S., Yamamoto, M., and Miura, Y. (1997). Negative regulation of the erythropoietin gene expression by the GATA transcription factors. Blood 89, 1430-1439.

Inaba, T., Roberts, W.M., Shapiro, L.H., Jolly, K.W., Raimondi, S.C., Smith, S.D., and Look, A.T. (1992). Fusion of the leucine zipper gene HLF to the E2A gene in human acute B-lineage leukemia. Science (New York, NY 257, 531-534.

Irizarry, R.A., Bolstad, B.M., Collin, F., Cope, L.M., Hobbs, B., and Speed, T.P. (2003). Summaries of Affymetrix GeneChip probe level data. Nucleic acids research 31, e15.

Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U., and Speed, T.P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4, 249-264.

Iwasaki, H., Mizuno, S., Wells, R.A., Cantor, A.B., Watanabe, S., and Akashi, K. (2003). GATA-1 converts lymphoid and myelomonocytic progenitors into the megakaryocyte/erythrocyte lineages. Immunity 19, 451-462.

Iyer, V.R., Horak, C.E., Scafe, C.S., Botstein, D., Snyder, M., and Brown, P.O. (2001). Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. Nature 409, 533-538.

Jackson, A.L., Bartz, S.R., Schelter, J., Kobayashi, S.V., Burchard, J., Mao, M., Li, B., Cavet, G., and Linsley, P.S. (2003). Expression profiling reveals off-target gene regulation by RNAi. Nature biotechnology 21, 635-637.

Jackson, A.L., Burchard, J., Leake, D., Reynolds, A., Schelter, J., Guo, J., Johnson, J.M., Lim, L., Karpilow, J., Nichols, K., et al. (2006). Position-specific chemical modification of siRNAs reduces "off-target" transcript silencing. RNA (New York, NY 12, 1197-1205.

Jackson, A.L., Burchard, J., Schelter, J., Chau, B.N., Cleary, M., Lim, L., and Linsley, P.S. (2006). Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity. RNA (New York, NY 12, 1179-1187.

Jazag, A., Ijichi, H., Kanai, F., Imamura, T., Guleng, B., Ohta, M., Imamura, J., Tanaka, Y., Tateishi, K., Ikenoue, T., et al. (2005). Smad4 silencing in pancreatic cancer cell lines using stable RNA interference and gene expression profiles induced by transforming growth factor-beta. Oncogene 24, 662-671.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science (New York, NY 293, 1074-1080.

Jiang, J., Chan, Y.S., Loh, Y.H., Cai, J., Tong, G.Q., Lim, C.A., Robson, P., Zhong, S., and Ng, H.H. (2008). A core Klf circuitry regulates self-renewal of embryonic stem cells. Nature cell biology 10, 353-360.

Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. Science (New York, NY 316, 1497-1502.

Johnson, E.S. (2004). Protein modification by SUMO. Annu Rev Biochem 73, 355-382.

Johnson, K.D., Grass, J.A., Boyer, M.E., Kiekhaefer, C.M., Blobel, G.A., Weiss, M.J., and Bresnick, E.H. (2002). Cooperative activities of hematopoietic regulators recruit RNA polymerase II to a tissue-specific chromatin domain. Proceedings of the National Academy of Sciences of the United States of America 99, 11760-11765.

Jongeneel, C.V., Delorenzi, M., Iseli, C., Zhou, D., Haudenschild, C.D., Khrebtukova, I., Kuznetsov, D., Stevenson, B.J., Strausberg, R.L., Simpson, A.J., et al. (2005). An atlas of human gene expression from massively parallel signature sequencing (MPSS). Genome research 15, 1007-1014.

Judge, A.D., Sood, V., Shaw, J.R., Fang, D., McClintock, K., and MacLachlan, I. (2005). Sequence-dependent stimulation of the mammalian innate immune response by synthetic siRNA. Nature biotechnology 23, 457-462.

Jurata, L.W., and Gill, G.N. (1997). Functional analysis of the nuclear LIM domain interactor NLI. Molecular and cellular biology 17, 5688-5698.

Kallianpur, A.R., Jordan, J.E., and Brandt, S.J. (1994). The SCL/TAL-1 gene is expressed in progenitors of both the hematopoietic and vascular systems during embryogenesis. Blood 83, 1200-1208.

Kappel, A., Schlaeger, T.M., Flamme, I., Orkin, S.H., Risau, W., and Breier, G. (2000). Role of SCL/Tal-1, GATA, and ets transcription factor binding sites for the regulation of flk-1 expression during murine vascular development. Blood 96, 3078-3085.

Kawasaki, H., and Taira, K. (2004). Induction of DNA methylation and gene silencing by short interfering RNAs in human cells. Nature 431, 211-217.

Kee, B.L., and Murre, C. (1998). Induction of early B cell factor (EBF) and multiple B lineage genes by the basic helix-loop-helix transcription factor E12. The Journal of experimental medicine 188, 699-713.

Kel, A.E., Gossling, E., Reuter, I., Cheremushkin, E., Kel-Margoulis, O.V., and Wingender, E. (2003). MATCH: A tool for searching transcription factor binding sites in DNA sequences. Nucleic acids research 31, 3576-3579.

Kennedy, M., Firpo, M., Choi, K., Wall, C., Robertson, S., Kabrun, N., and Keller, G. (1997). A common precursor for primitive erythropoiesis and definitive haematopoiesis. Nature 386, 488-493.

Kennerdell, J.R., and Carthew, R.W. (1998). Use of dsRNA-mediated genetic interference to demonstrate that frizzled and frizzled 2 act in the wingless pathway. Cell 95, 1017-1026.

Kennerdell, J.R., and Carthew, R.W. (2000). Heritable gene silencing in Drosophila using double-stranded RNA. Nature biotechnology 18, 896-898.

Kim, J., Bhinge, A.A., Morgan, X.C., and Iyer, V.R. (2005). Mapping DNA-protein interactions in large genomes by sequence tag analysis of genomic enrichment. Nature methods 2, 47-53.

Kim, J., and Iyer, V.R. (2004). Global role of TATA box-binding protein recruitment to promoters in mediating gene expression profiles. Molecular and cellular biology 24, 8104-8112.

Kim, J.B., Porreca, G.J., Song, L., Greenway, S.C., Gorham, J.M., Church, G.M., Seidman, C.E., and Seidman, J.G. (2007). Polony multiplex analysis of gene expression (PMAGE) in mouse hypertrophic cardiomyopathy. Science (New York, NY 316, 1481-1484.

Kim, T.H., Barrera, L.O., Qu, C., Van Calcar, S., Trinklein, N.D., Cooper, S.J., Luna, R.M., Glass, C.K., Rosenfeld, M.G., Myers, R.M., et al. (2005). Direct isolation and identification of promoters in the human genome. Genome research 15, 830-839.

Kim, T.H., Barrera, L.O., Zheng, M., Qu, C., Singer, M.A., Richmond, T.A., Wu, Y., Green, R.D., and Ren, B. (2005). A high-resolution map of active promoters in the human genome. Nature 436, 876-880.

Kittler, R., Pelletier, L., Heninger, A.K., Slabicki, M., Theis, M., Miroslaw, L., Poser, I., Lawo, S., Grabner, H., Kozak, K., et al. (2007). Genome-scale RNAi profiling of cell division in human tissue culture cells. Nature cell biology 9, 1401-1412.

Kittler, R., Putz, G., Pelletier, L., Poser, I., Heninger, A.K., Drechsel, D., Fischer, S., Konstantinova, I., Habermann, B., Grabner, H., et al. (2004). An endoribonuclease-prepared siRNA screen in human cells identifies genes essential for cell division. Nature 432, 1036-1040.

Kloosterman, W.P., and Plasterk, R.H. (2006). The diverse functions of microRNAs in animal development and disease. Developmental cell 11, 441-450.

Kobayashi, M., Nishikawa, K., and Yamamoto, M. (2001). Hematopoietic regulatory domain of gata1 gene is positively regulated by GATA1 protein in zebrafish embryos. Development (Cambridge, England) 128, 2341-2350.

Koch, C.M., Andrews, R.M., Flicek, P., Dillon, S.C., Karaoz, U., Clelland, G.K., Wilcox, S., Beare, D.M., Fowler, J.C., Couttet, P., et al. (2007). The landscape of histone modifications across 1% of the human genome in five human cell lines. Genome research 17, 691-707.

Kodzius, R., Kojima, M., Nishiyori, H., Nakamura, M., Fukuda, S., Tagami, M., Sasaki, D., Imamura, K., Kai, C., Harbers, M., et al. (2006). CAGE: cap analysis of gene expression. Nature methods 3, 211-222.

Koizumi, K., Higashida, H., Yoo, S., Islam, M.S., Ivanov, A.I., Guo, V., Pozzi, P., Yu, S.H., Rovescalli, A.C., Tang, D., et al. (2007). RNA interference screen to identify genes required for Drosophila embryonic nervous system development. Proceedings of the National Academy of Sciences of the United States of America 104, 5626-5631.

Kolfschoten, I.G., van Leeuwen, B., Berns, K., Mullenders, J., Beijersbergen, R.L., Bernards, R., Voorhoeve, P.M., and Agami, R. (2005). A genetic screen identifies PITX1 as a suppressor of RAS activity and tumorigenicity. Cell 121, 849-858.

Krosl, G., He, G., Lefrancois, M., Charron, F., Romeo, P.H., Jolicoeur, P., Kirsch, I.R., Nemer, M., and Hoang, T. (1998). Transcription factor SCL is required for c-kit expression and c-Kit function in hemopoietic cells. The Journal of experimental medicine 188, 439-450.

Kuhl, C., Atzberger, A., Iborra, F., Nieswandt, B., Porcher, C., and Vyas, P. (2005). GATA1-mediated megakaryocyte differentiation and growth control can be uncoupled and mapped to different domains in GATA1. Molecular and cellular biology 25, 8592-8606.

Kuhn, K., Baker, S.C., Chudin, E., Lieu, M.H., Oeser, S., Bennett, H., Rigault, P., Barker, D., McDaniel, T.K., and Chee, M.S. (2004). A novel, high-performance random array platform for quantitative gene expression profiling. Genome research 14, 2347-2356.

Kulessa, H., Frampton, J., and Graf, T. (1995). GATA-1 reprograms avian myelomonocytic cell lines into eosinophils, thromboblasts, and erythroblasts. Genes & development 9, 1250-1262.

Kumaravelu, P., Hook, L., Morrison, A.M., Ure, J., Zhao, S., Zuyev, S., Ansell, J., and Medvinsky, A. (2002). Quantitative developmental anatomy of definitive haematopoietic stem cells/long-term repopulating units (HSC/RUs): role of the aorta-gonad-mesonephros (AGM) region and the yolk sac in colonisation of the mouse embryonic liver. Development (Cambridge, England) 129, 4891-4899.

Kurdistani, S.K., Tavazoie, S., and Grunstein, M. (2004). Mapping global histone acetylation patterns to gene expression. Cell 117, 721-733.

Labastie, M.C., Cortes, F., Romeo, P.H., Dulac, C., and Peault, B. (1998). Molecular identity of hematopoietic precursor cells emerging in the human embryo. Blood 92, 3624-3635.

Lachner, M., and Jenuwein, T. (2002). The many faces of histone lysine methylation. Curr Opin Cell Biol 14, 286-298.

Lahlil, R., Lecuyer, E., Herblot, S., and Hoang, T. (2004). SCL assembles a multifactorial complex that determines glycophorin A expression. Molecular and cellular biology 24, 1439-1452.

Lamond, A.I., and Earnshaw, W.C. (1998). Structure and function in the nucleus. Science (New York, NY 280, 547-553.

Landry, J.R., Kinston, S., Knezevic, K., de Bruijn, M.F., Wilson, N., Nottingham, W.T., Peitz, M., Edenhofer, F., Pimanda, J.E., Ottersbach, K., et al. (2008). Runx genes are direct targets of Scl/Tal1 in the yolk sac and fetal liver. Blood 111, 3005-3014.

Larson, R.C., Lavenir, I., Larson, T.A., Baer, R., Warren, A.J., Wadman, I., Nottage, K., and Rabbitts, T.H. (1996). Protein dimerization between Lmo2 (Rbtn2) and Tal1 alters thymocyte development and potentiates T cell tumorigenesis in transgenic mice. Embo J 15, 1021-1027.

Lassar, A.B., Davis, R.L., Wright, W.E., Kadesch, T., Murre, C., Voronova, A., Baltimore, D., and Weintraub, H. (1991). Functional activity of myogenic HLH proteins requires hetero-oligomerization with E12/E47-like proteins in vivo. Cell 66, 305-315.

Lazrak, M., Deleuze, V., Noel, D., Haouzi, D., Chalhoub, E., Dohet, C., Robbins, I., and Mathieu, D. (2004). The bHLH TAL-1/SCL regulates endothelial cell migration and morphogenesis. Journal of cell science 117, 1161-1171.

Lecointe, N., Bernard, O., Naert, K., Joulin, V., Larsen, C.J., Romeo, P.H., and Mathieu-Mahul, D. (1994). GATA-and SP1-binding sites are required for the full activity of the tissue-specific promoter of the tal-1 gene. Oncogene 9, 2623-2632.

Lecuyer, E., Herblot, S., Saint-Denis, M., Martin, R., Begley, C.G., Porcher, C., Orkin, S.H., and Hoang, T. (2002). The SCL complex regulates c-kit expression in hematopoietic cells through functional interaction with Sp1. Blood 100, 2430-2440.

Lee, D.H., Gershenzon, N., Gupta, M., Ioshikhes, I.P., Reinberg, D., and Lewis, B.A. (2005). Functional characterization of core promoter elements: the downstream core element is recognized by TAF1. Molecular and cellular biology 25, 9674-9686.

Lee, D.Y., Teyssier, C., Strahl, B.D., and Stallcup, M.R. (2005). Role of protein methylation in regulation of transcription. Endocrine reviews 26, 147-170.

Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 75, 843-854.

Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K., et al. (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. Cell 125, 301-313.

Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.T., Thompson, C.M., Simon, I., et al. (2002). Transcriptional regulatory networks in Saccharomyces cerevisiae. Science (New York, NY 298, 799-804.

Lee, T.I., and Young, R.A. (2000). Transcription of eukaryotic protein-coding genes. Annual review of genetics 34, 77-137.

LeRoy, G., Rickards, B., and Flint, S.J. (2008). The double bromodomain proteins Brd2 and Brd3 couple histone acetylation to transcription. Mol Cell 30, 51-60.

Leroy-Viard, K., Vinit, M.A., Lecointe, N., Mathieu-Mahul, D., and Romeo, P.H. (1994). Distinct DNase-I hypersensitive sites are associated with TAL-1 transcription in erythroid and T-cell lines. Blood 84, 3819-3827.

Letting, D.L., Chen, Y.Y., Rakowski, C., Reedy, S., and Blobel, G.A. (2004). Context-dependent regulation of GATA-1 by friend of GATA-1. Proceedings of the National Academy of Sciences of the United States of America 101, 476-481.

Leuther, K.K., Salmeron, J.M., and Johnston, S.A. (1993). Genetic evidence that an activation domain of GAL4 does not require acidity and may form a beta sheet. Cell 72, 575-585.

Li, M., Li, H., and Rossi, J.J. (2006). RNAi in combination with a ribozyme and TAR decoy for treatment of HIV infection in hematopoietic cell gene therapy. Annals of the New York Academy of Sciences 1082, 172-179.

Li, Q., Peterson, K.R., Fang, X., and Stamatoyannopoulos, G. (2002). Locus control regions. Blood 100, 3077-3086.

Li, Z., Van Calcar, S., Qu, C., Cavenee, W.K., Zhang, M.Q., and Ren, B. (2003). A global transcriptional regulatory role for c-Myc in Burkitt's lymphoma cells. Proceedings of the National Academy of Sciences of the United States of America 100, 8164-8169.

Liang, P., and Pardee, A.B. (1992). Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science (New York, NY 257, 967-971.

Lieb, J.D., Liu, X., Botstein, D., and Brown, P.O. (2001). Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. Nature genetics 28, 327-334.

Lieu, P.T., Jozsi, P., Gilles, P., and Peterson, T. (2005). Development of a DNA-labeling system for array-based comparative genomic hybridization. J Biomol Tech 16, 104-111.

Lim, C.Y., Santoso, B., Boulay, T., Dong, E., Ohler, U., and Kadonaga, J.T. (2004). The MTE, a new core promoter element for transcription by RNA polymerase II. Genes & development 18, 1606-1617.

Lin, Y.W., and Aplan, P.D. (2007). Gene expression profiling of precursor T-cell lymphoblastic leukemia/lymphoma identifies oncogenic pathways that are potential therapeutic targets. Leukemia 21, 1276-1284.

Lipshutz, R.J., Fodor, S.P., Gingeras, T.R., and Lockhart, D.J. (1999). High density synthetic oligonucleotide arrays. Nature genetics 21, 20-24.

Liu, X., Noll, D.M., Lieb, J.D., and Clarke, N.D. (2005). DIP-chip: rapid and accurate determination of DNA-binding specificity. Genome research 15, 421-427.

Liu, X.S., Brutlag, D.L., and Liu, J.S. (2002). An algorithm for finding protein-DNA binding sites with applications to chromatin-immunoprecipitation microarray experiments. Nature biotechnology 20, 835-839.

Loh, Y.H., Wu, Q., Chew, J.L., Vega, V.B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J., et al. (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nature genetics 38, 431-440.

Longabaugh, W.J., Davidson, E.H., and Bolouri, H. (2005). Computational representation of developmental genetic regulatory networks. Developmental biology 283, 1-16.

Lorch, Y., Zhang, M., and Kornberg, R.D. (1999). Histone octamer transfer by a chromatin-remodeling complex. Cell 96, 389-392.

Louie, M.C., Yang, H.Q., Ma, A.H., Xu, W., Zou, J.X., Kung, H.J., and Chen, H.W. (2003). Androgen-induced recruitment of RNA polymerase II to a nuclear receptor-p160 coactivator complex. Proceedings of the National Academy of Sciences of the United States of America 100, 2226-2230.

Loukinov, D.I., Pugacheva, E., Vatolin, S., Pack, S.D., Moon, H., Chernukhin, I., Mannan, P., Larsson, E., Kanduri, C., Vostrov, A.A., et al. (2002). BORIS, a novel male germ-line-specific protein associated with epigenetic reprogramming events, shares the same 11-zinc-finger domain with CTCF, the insulator protein involved in reading imprinting marks in the soma. Proceedings of the National Academy of Sciences of the United States of America 99, 6806-6811.

Loyola, A., Huang, J.Y., LeRoy, G., Hu, S., Wang, Y.H., Donnelly, R.J., Lane, W.S., Lee, S.C., and Reinberg, D. (2003). Functional analysis of the subunits of the chromatin assembly factor RSF. Molecular and cellular biology 23, 6759-6768.

Lozzio, B.B., and Lozzio, C.B. (1977). Properties of the K562 cell line derived from a patient with chronic myeloid leukemia. Int J Cancer 19, 136.

Lozzio, B.B., Lozzio, C.B., Bamberger, E.G., and Feliu, A.S. (1981). A multipotential leukemia cell line (K-562) of human origin. Proceedings of the Society for Experimental Biology and Medicine Society for Experimental Biology and Medicine (New York, NY 166, 546-550.

Lugus, J.J., Chung, Y.S., Mills, J.C., Kim, S.I., Grass, J., Kyba, M., Doherty, J.M., Bresnick, E.H., and Choi, K. (2007). GATA2 functions at multiple steps in hemangioblast development and differentiation. Development (Cambridge, England) 134, 393-405.

Lum, L., Yao, S., Mozer, B., Rovescalli, A., Von Kessler, D., Nirenberg, M., and Beachy, P.A. (2003). Identification of Hedgehog pathway components by RNAi in Drosophila cultured cells. Science (New York, NY 299, 2039-2045.

Lupberger, J., Kreuzer, K.A., Baskaynak, G., Peters, U.R., le Coutre, P., and Schmidt, C.A. (2002). Quantitative analysis of beta-actin, beta-2-microglobulin and porphobilinogen deaminase mRNA and their comparison as control transcripts for RT-PCR. Molecular and cellular probes 16, 25-30.

Mackay, J.P., Kowalski, K., Fox, A.H., Czolij, R., King, G.F., and Crossley, M. (1998). Involvement of the N-finger in the self-association of GATA-1. The Journal of biological chemistry 273, 30560-30567.

MacKeigan, J.P., Murphy, L.O., and Blenis, J. (2005). Sensitized RNAi screen of human kinases and phosphatases identifies new regulators of apoptosis and chemoresistance. Nature cell biology 7, 591-600.

Mahadevan, L.C., Willis, A.C., and Barratt, M.J. (1991). Rapid histone H3 phosphorylation in response to growth factors, phorbol esters, okadaic acid, and protein synthesis inhibitors. Cell 65, 775-783.

Mansson, R., Tsapogas, P., Akerlund, M., Lagergren, A., Gisler, R., and Sigvardsson, M. (2004). Pearson correlation analysis of microarray data allows for the identification of genetic targets for early B-cell factor. The Journal of biological chemistry 279, 17905-17913.

Mao, D.Y., Watson, J.D., Yan, P.S., Barsyte-Lovejoy, D., Khosravi, F., Wong, W.W., Farnham, P.J., Huang, T.H., and Penn, L.Z. (2003). Analysis of Myc bound loci identified by CpG island arrays shows that Max is essential for Myc-dependent repression. Curr Biol 13, 882-886.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. Nature 437, 376-380.

Marodon, G., Mouly, E., Blair, E.J., Frisen, C., Lemoine, F.M., and Klatzmann, D. (2003). Specific transgene expression in human and mouse CD4+ cells using lentiviral vectors with regulatory sequences from the CD4 gene. Blood 101, 3416-3423.

Marques, J.T., Devosse, T., Wang, D., Zamanian-Daryoush, M., Serbinowski, P., Hartmann, R., Fujita, T., Behlke, M.A., and Williams, B.R. (2006). A structural basis for discriminating between self and nonself double-stranded RNAs in mammalian cells. Nature biotechnology 24, 559-565.

Martin, D.I., and Orkin, S.H. (1990). Transcriptional activation and DNA binding by the erythroid factor GF-1/NF-E1/Eryf 1. Genes & development 4, 1886-1898.

Martin, D.I., Zon, L.I., Mutter, G., and Orkin, S.H. (1990). Expression of an erythroid transcription factor in megakaryocytic and mast cell lineages. Nature 344, 444-447.

Martin, P., and Papayannopoulou, T. (1982). HEL cells: a new human erythroleukemia cell line with spontaneous and induced globin expression. Science (New York, NY 216, 1233-1235.

Marton, M.J., DeRisi, J.L., Bennett, H.A., Iyer, V.R., Meyer, M.R., Roberts, C.J., Stoughton, R., Burchard, J., Slade, D., Dai, H., et al. (1998). Drug target validation and identification of secondary drug target effects using DNA microarrays. Nature medicine 4, 1293-1301.

Martone, R., Euskirchen, G., Bertone, P., Hartman, S., Royce, T.E., Luscombe, N.M., Rinn, J.L., Nelson, F.K., Miller, P., Gerstein, M., et al. (2003). Distribution of NF-kappaB-binding sites across human chromosome 22. Proceedings of the National Academy of Sciences of the United States of America 100, 12247-12252.

Massari, M.E., Grant, P.A., Pray-Grant, M.G., Berger, S.L., Workman, J.L., and Murre, C. (1999). A conserved motif present in a class of helix-loop-helix proteins activates transcription by direct recruitment of the SAGA complex. Mol Cell 4, 63-73.

Matsukura, S., Jones, P.A., and Takai, D. (2003). Establishment of conditional vectors for hairpin siRNA knockdowns. Nucleic acids research 31, e77.

Matsumoto, N., Kubo, A., Liu, H., Akita, K., Laub, F., Ramirez, F., Keller, G., and Friedman, S.L. (2006). Developmental regulation of yolk sac hematopoiesis by Kruppel-like factor 6. Blood 107, 1357-1365.

Matsumura, I., Kawasaki, A., Tanaka, H., Sonoyama, J., Ezoe, S., Minegishi, N., Nakajima, K., Yamamoto, M., and Kanakura, Y. (2000). Biologic significance of GATA-1 activities in Ras-mediated megakaryocytic differentiation of hematopoietic cell lines. Blood 96, 2440-2450.

McArthur, M., Gerum, S., and Stamatoyannopoulos, G. (2001). Quantification of DNaseI-sensitivity by real-time PCR: quantitative analysis of DNaseI-hypersensitivity of the mouse beta-globin LCR. Journal of molecular biology 313, 27-34.

Medina, K.L., Pongubala, J.M., Reddy, K.L., Lancki, D.W., Dekoter, R., Kieslinger, M., Grosschedl, R., and Singh, H. (2004). Assembling a gene regulatory network for specification of the B cell fate. Developmental cell 7, 607-617.

Medvinsky, A., and Dzierzak, E. (1996). Definitive hematopoiesis is autonomously initiated by the AGM region. Cell 86, 897-906.

Meersseman, G., Pennings, S., and Bradbury, E.M. (1992). Mobile nucleosomes--a general behavior. Embo J 11, 2951-2959.

Mehaffey, M.G., Newton, A.L., Gandhi, M.J., Crossley, M., and Drachman, J.G. (2001). X-linked thrombocytopenia caused by a novel mutation of GATA-1. Blood 98, 2681-2688.

Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., and Tuschl, T. (2004). Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. Mol Cell 15, 185-197.

Mellentin, J.D., Smith, S.D., and Cleary, M.L. (1989). lyl-1, a novel gene altered by chromosomal translocation in T cell leukemia, codes for a protein with a helix-loop-helix DNA binding motif. Cell 58, 77-83.

Meng, A., Tang, H., Yuan, B., Ong, B.A., Long, Q., and Lin, S. (1999). Positive and negative cis-acting elements are required for hematopoietic expression of zebrafish GATA-1. Blood 93, 500-508.

Merika, M., and Orkin, S.H. (1995). Functional synergy and physical interactions of the erythroid transcription factor GATA-1 with the Kruppel family proteins Sp1 and EKLF. Molecular and cellular biology 15, 2437-2447.

Mermod, N., O'Neill, E.A., Kelly, T.J., and Tjian, R. (1989). The proline-rich transcriptional activator of CTF/NF-I is distinct from the replication and DNA binding domain. Cell 58, 741-753.

Metzger, E., Wissmann, M., Yin, N., Muller, J.M., Schneider, R., Peters, A.H., Gunther, T., Buettner, R., and Schule, R. (2005). LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. Nature 437, 436-439.

Migliaccio, A.R., Rana, R.A., Sanchez, M., Lorenzini, R., Centurione, L., Bianchi, L., Vannucchi, A.M., Migliaccio, G., and Orkin, S.H. (2003). GATA-1 as a regulator of mast cell differentiation revealed by the phenotype of the GATA-1low mouse mutant. The Journal of experimental medicine 197, 281-296.

Mikkola, H.K., Klintman, J., Yang, H., Hock, H., Schlaeger, T.M., Fujiwara, Y., and Orkin, S.H. (2003). Haematopoietic stem cells retain long-term repopulating activity and multipotency in the absence of stem-cell leukaemia SCL/tal-1 gene. Nature 421, 547-551.

Miller, I.J., and Bieker, J.J. (1993). A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Kruppel family of nuclear proteins. Molecular and cellular biology 13, 2776-2786.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002). Network motifs: simple building blocks of complex networks. Science (New York, NY 298, 824-827.

Milyavsky, M., Tabach, Y., Shats, I., Erez, N., Cohen, Y., Tang, X., Kalis, M., Kogan, I., Buganim, Y., Goldfinger, N., et al. (2005). Transcriptional programs following genetic alterations in p53, INK4A, and H-Ras genes along defined stages of malignant transformation. Cancer research 65, 4530-4543.

Mineno, J., Okamoto, S., Ando, T., Sato, M., Chono, H., Izu, H., Takayama, M., Asada, K., Mirochnitchenko, O., Inouye, M., et al. (2006). The expression profile of microRNAs in mouse embryos. Nucleic acids research 34, 1765-1771.

Moffat, J., Grueneberg, D.A., Yang, X., Kim, S.Y., Kloepfer, A.M., Hinkle, G., Piqani, B., Eisenhaure, T.M., Luo, B., Grenier, J.K., et al. (2006). A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. Cell 124, 1283-1298.

Molkentin, J.D. (2000). The zinc finger-containing transcription factors GATA-4, -5, and -6. Ubiquitously expressed regulators of tissue-specific gene expression. The Journal of biological chemistry 275, 38949-38952.

Moore, M.A., and Metcalf, D. (1970). Ontogeny of the haemopoietic system: yolk sac origin of in vivo and in vitro colony forming cells in the developing mouse embryo. British journal of haematology 18, 279-296.

Morin, R.D., O'Connor, M.D., Griffith, M., Kuchenbauer, F., Delaney, A., Prabhu, A.L., Zhao, Y., McDonald, H., Zeng, T., Hirst, M., et al. (2008). Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome research.

Moroni, E., Mastrangelo, T., Razzini, R., Cairns, L., Moi, P., Ottolenghi, S., and Giglioni, B. (2000). Regulation of mouse p45 NF-E2 transcription by an erythroid-specific GATA-dependent intronic alternative promoter. The Journal of biological chemistry 275, 10567-10576.

Morris, K.V., Chan, S.W., Jacobsen, S.E., and Looney, D.J. (2004). Small interfering RNA-induced transcriptional gene silencing in human cells. Science (New York, NY 305, 1289-1292.

Morrison, T.B., Weis, J.J., and Wittwer, C.T. (1998). Quantification of low-copy transcripts by continuous SYBR Green I monitoring during amplification. BioTechniques 24, 954-958, 960, 962.

Mouthon, M.A., Bernard, O., Mitjavila, M.T., Romeo, P.H., Vainchenker, W., and Mathieu-Mahul, D. (1993). Expression of tal-1 and GATA-binding proteins during human hematopoiesis. Blood 81, 647-655.

Mughal, T.I., and Goldman, J.M. (2006). Chronic myeloid leukemia: why does it evolve from chronic phase to blast transformation? Front Biosci 11, 198-208.

Mukhopadhyay, M., Teufel, A., Yamashita, T., Agulnick, A.D., Chen, L., Downs, K.M., Schindler, A., Grinberg, A., Huang, S.P., Dorward, D., et al. (2003). Functional ablation of the mouse Ldb1 gene results in severe patterning defects during gastrulation. Development (Cambridge, England) 130, 495-505.

Muntean, A.G., and Crispino, J.D. (2005). Differential requirements for the activation domain and FOG-interaction surface of GATA-1 in megakaryocyte gene expression and development. Blood 106, 1223-1231.

Murre, C., McCaw, P.S., and Baltimore, D. (1989). A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. Cell 56, 777-783.

Myers, J.W., Jones, J.T., Meyer, T., and Ferrell, J.E., Jr. (2003). Recombinant Dicer efficiently converts large dsRNAs into siRNAs suitable for gene silencing. Nature biotechnology 21, 324-328.

Myers, L.C., and Kornberg, R.D. (2000). Mediator of transcriptional regulation. Annu Rev Biochem 69, 729-749.

Nagy, P.L., Griesenbeck, J., Kornberg, R.D., and Cleary, M.L. (2002). A trithorax-group complex purified from Saccharomyces cerevisiae is required for methylation of histone H3. Proceedings of the National Academy of Sciences of the United States of America 99, 90-94.

Nakamura, Y., Umehara, T., Nakano, K., Jang, M.K., Shirouzu, M., Morita, S., Uda-Tochio, H., Hamana, H., Terada, T., Adachi, N., et al. (2007). Crystal structure of the human BRD2 bromodomain: insights into dimerization and recognition of acetylated histone H4. The Journal of biological chemistry 282, 4193-4201.

Neale, G.A., Rehg, J.E., and Goorha, R.M. (1995). Ectopic expression of rhombotin-2 causes selective expansion of CD4-CD8- lymphocytes in the thymus and T-cell tumors in transgenic mice. Blood 86, 3060-3071.

Nelson, C.J., Santos-Rosa, H., and Kouzarides, T. (2006). Proline isomerization of histone H3 regulates lysine methylation and gene expression. Cell 126, 905-916.

Ngo, V.N., Davis, R.E., Lamy, L., Yu, X., Zhao, H., Lenz, G., Lam, L.T., Dave, S., Yang, L., Powell, J., et al. (2006). A loss-of-function RNA interference screen for molecular targets in cancer. Nature 441, 106-110.

Nichols, K.E., Crispino, J.D., Poncz, M., White, J.G., Orkin, S.H., Maris, J.M., and Weiss, M.J. (2000). Familial dyserythropoietic anaemia and thrombocytopenia due to an inherited mutation in GATA1. Nature genetics 24, 266-270.

Nicolis, S., Bertini, C., Ronchi, A., Crotta, S., Lanfranco, L., Moroni, E., Giglioni, B., and Ottolenghi, S. (1991). An erythroid specific enhancer upstream to the gene encoding the cell-type specific transcription factor GATA-1. Nucleic acids research 19, 5285-5291.

Nilsson, R., Bajic, V.B., Suzuki, H., di Bernardo, D., Bjorkegren, J., Katayama, S., Reid, J.F., Sweet, M.J., Gariboldi, M., Carninci, P., et al. (2006). Transcriptional network dynamics in macrophage activation. Genomics 88, 133-142.

Nishikawa, K., Kobayashi, M., Masumi, A., Lyons, S.E., Weinstein, B.M., Liu, P.P., and Yamamoto, M. (2003). Self-association of Gata1 enhances transcriptional activity in vivo in zebra fish embryos. Molecular and cellular biology 23, 8295-8305.

Nishimura, S., Takahashi, S., Kuroha, T., Suwabe, N., Nagasawa, T., Trainor, C., and Yamamoto, M. (2000). A GATA box in the GATA-1 gene hematopoietic enhancer is a critical element in the network of GATA factors and sites that regulate this gene. Molecular and cellular biology 20, 713-723.

Noble, W.S., Kuehn, S., Thurman, R., Yu, M., and Stamatoyannopoulos, J. (2005). Predicting the in vivo signature of human gene regulatory sequences. Bioinformatics (Oxford, England) 21 Suppl 1, i338-343.

Nordhoff, E., Krogsdam, A.M., Jorgensen, H.F., Kallipolitis, B.H., Clark, B.F., Roepstorff, P., and Kristiansen, K. (1999). Rapid identification of DNA-binding proteins by mass spectrometry. Nature biotechnology 17, 884-888.

Nottingham, W.T., Jarratt, A., Burgess, M., Speck, C.L., Cheng, J.F., Prabhakar, S., Rubin, E.M., Li, P.S., Sloane-Stanley, J., Kong, A.S.J., et al. (2007). Runx1-mediated hematopoietic stem-cell emergence is controlled by a Gata/Ets/SCL-regulated enhancer. Blood 110, 4188-4197.

Nourse, J., Mellentin, J.D., Galili, N., Wilkinson, J., Stanbridge, E., Smith, S.D., and Cleary, M.L. (1990). Chromosomal translocation t(1;19) results in synthesis of a homeobox fusion mRNA that codes for a potential chimeric transcription factor. Cell 60, 535-545.

Nuez, B., Michalovich, D., Bygrave, A., Ploemacher, R., and Grosveld, F. (1995). Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene. Nature 375, 316-318.

Nunez, E., Kwon, Y.S., Hutt, K.R., Hu, Q., Cardamone, M.D., Ohgi, K.A., Garcia-Bassets, I., Rose, D.W., Glass, C.K., Rosenfeld, M.G., et al. (2008). Nuclear receptor-enhanced transcription requires motor- and LSD1-dependent gene networking in interchromatin granules. Cell 134, 189.

Oberley, M.J., Inman, D.R., and Farnham, P.J. (2003). E2F6 negatively regulates BRCA1 in human cancer cells without methylation of histone H3 on lysine 9. The Journal of biological chemistry 278, 42466-42476.

Oda, K., Kimura, T., Matsuoka, Y., Funahashi, A., Muramatsu, M., and Kitano, H. (2004). Molecular Interaction Map of Macrophage. AfCS Research Reports 2.

Odom, D.T., Dowell, R.D., Jacobsen, E.S., Gordon, W., Danford, T.W., MacIsaac, K.D., Rolfe, P.A., Conboy, C.M., Gifford, D.K., and Fraenkel, E. (2007). Tissue-specific transcriptional regulation has diverged significantly between human and mouse. Nature genetics 39, 730-732.

Odom, D.T., Zizlsperger, N., Gordon, D.B., Bell, G.W., Rinaldi, N.J., Murray, H.L., Volkert, T.L., Schreiber, J., Rolfe, P.A., Gifford, D.K., et al. (2004). Control of pancreas and liver gene expression by HNF transcription factors. Science (New York, NY 303, 1378-1381.

Ogilvy, S., Ferreira, R., Piltz, S.G., Bowen, J.M., Gottgens, B., and Green, A.R. (2007). The SCL +40 enhancer targets the midbrain together with primitive and definitive hematopoiesis and is regulated by SCL and GATA proteins. Molecular and cellular biology 27, 7206-7219.

Oliva, A., Rosebrock, A., Ferrezuelo, F., Pyne, S., Chen, H., Skiena, S., Futcher, B., and Leatherwood, J. (2005). The cell cycle-regulated genes of Schizosaccharomyces pombe. PLoS biology 3, e225.

Olsen, P.H., and Ambros, V. (1999). The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation. Developmental biology 216, 671-680.

O'Neil, J., Shank, J., Cusson, N., Murre, C., and Kelliher, M. (2004). TAL1/SCL induces leukemia by inhibiting the transcriptional activity of E47/HEB. Cancer cell 5, 587-596.

O'Neill, L.P., and Turner, B.M. (1996). Immunoprecipitation of chromatin. Methods in enzymology 274, 189-197.

O'Neill, L.P., VerMilyea, M.D., and Turner, B.M. (2006). Epigenetic characterization of the early embryo with a chromatin immunoprecipitation protocol applicable to small cell populations. Nature genetics 38, 835-841.

Onizuka, T., Endo, S., Hirano, M., Kanai, S., and Akiyama, H. (2002). Design of a fluorescent electrophoretic mobility shift assay improved for the quantitative and multiple analysis of protein-DNA complexes. Bioscience, biotechnology, and biochemistry 66, 2732-2734.

Ono, Y., Fukuhara, N., and Yoshie, O. (1997). Transcriptional activity of TAL1 in T cell acute lymphoblastic leukemia (T-ALL) requires RBTN1 or -2 and induces TALLA1, a highly specific tumor marker of T-ALL. The Journal of biological chemistry 272, 4576-4581.

Ono, Y., Fukuhara, N., and Yoshie, O. (1998). TAL1 and LIM-only proteins synergistically induce retinaldehyde dehydrogenase 2 expression in T-cell acute lymphoblastic leukemia by acting as cofactors for GATA3. Molecular and cellular biology 18, 6939-6950.

Onodera, K., Takahashi, S., Nishimura, S., Ohta, J., Motohashi, H., Yomogida, K., Hayashi, N., Engel, J.D., and Yamamoto, M. (1997). GATA-1 transcription is controlled by distinct regulatory mechanisms during primitive and definitive erythropoiesis. Proceedings of the National Academy of Sciences of the United States of America 94, 4487-4492.

Onodera, K., Yomogida, K., Suwabe, N., Takahashi, S., Muraosa, Y., Hayashi, N., Ito, E., Gu, L., Rassoulzadegan, M., Engel, J.D., et al. (1997). Conserved structure, regulatory elements, and transcriptional regulation from the GATA-1 gene testis promoter. Journal of biochemistry 121, 251-263.

Orian, A., van Steensel, B., Delrow, J., Bussemaker, H.J., Li, L., Sawado, T., Williams, E., Loo, L.W., Cowley, S.M., Yost, C., et al. (2003). Genomic binding by the Drosophila Myc, Max, Mad/Mnt transcription factor network. Genes & development 17, 1101-1114.

Orlando, V., Strutt, H., and Paro, R. (1997). Analysis of chromatin structure by in vivo formaldehyde cross-linking. Methods (San Diego, Calif 11, 205-214.

Pabo, C.O., and Sauer, R.T. (1992). Transcription factors: structural families and principles of DNA recognition. Annu Rev Biochem 61, 1053-1095.

Paddison, P.J., Caudy, A.A., Bernstein, E., Hannon, G.J., and Conklin, D.S. (2002). Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. Genes & development 16, 948-958.

Pal, S., Cantor, A.B., Johnson, K.D., Moran, T.B., Boyer, M.E., Orkin, S.H., and Bresnick, E.H. (2004). Coregulator-dependent facilitation of chromatin occupancy by GATA-1. Proceedings of the National Academy of Sciences of the United States of America 101, 980-985.

Palomero, T., Odom, D.T., O'Neil, J., Ferrando, A.A., Margolin, A., Neuberg, D.S., Winter, S.S., Larson, R.S., Li, W., Liu, X.S., et al. (2006). Transcriptional regulatory networks downstream of TAL1/SCL in T-cell acute lymphoblastic leukemia. Blood 108, 986-992.

Parry, D.H., Xu, J., and Ruvkun, G. (2007). A whole-genome RNAi Screen for C. elegans miRNA pathway genes. Curr Biol 17, 2013-2022.

Pebernard, S., and Iggo, R.D. (2004). Determinants of interferon-stimulated gene induction by RNAi vectors. Differentiation; research in biological diversity 72, 103-111.

Pennacchio, L.A., Ahituv, N., Moses, A.M., Prabhakar, S., Nobrega, M.A., Shoukry, M., Minovitsky, S., Dubchak, I., Holt, A., Lewis, K.D., et al. (2006). In vivo enhancer analysis of human conserved non-coding sequences. Nature 444, 499-502.

Pevny, L., Lin, C.S., D'Agati, V., Simon, M.C., Orkin, S.H., and Costantini, F. (1995). Development of hematopoietic cells lacking transcription factor GATA-1. Development (Cambridge, England) 121, 163-172.

Pevny, L., Simon, M.C., Robertson, E., Klein, W.H., Tsai, S.F., D'Agati, V., Orkin, S.H., and Costantini, F. (1991). Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. Nature 349, 257-260.

Pokholok, D.K., Hannett, N.M., and Young, R.A. (2002). Exchange of RNA polymerase II initiation and elongation factors during gene expression in vivo. Mol Cell 9, 799-809.

Pokholok, D.K., Harbison, C.T., Levine, S., Cole, M., Hannett, N.M., Lee, T.I., Bell, G.W., Walker, K., Rolfe, P.A., Herbolsheimer, E., et al. (2005). Genome-wide map of nucleosome acetylation and methylation in yeast. Cell 122, 517-527.

Poot, R.A., Dellaire, G., Hulsmann, B.B., Grimaldi, M.A., Corona, D.F., Becker, P.B., Bickmore, W.A., and Varga-Weisz, P.D. (2000). HuCHRAC, a human ISWI chromatin remodelling complex contains hACF1 and two novel histone-fold proteins. Embo J 19, 3377-3387.

Porcher, C., Swat, W., Rockwell, K., Fujiwara, Y., Alt, F.W., and Orkin, S.H. (1996). The T cell leukemia oncoprotein SCL/tal-1 is essential for development of all hematopoietic lineages. Cell 86, 47-57.

Privalsky, M.L. (2004). The role of corepressors in transcriptional regulation by nuclear hormone receptors. Annual review of physiology 66, 315-360.

Ptashne, M., and Gann, A. (1997). Transcriptional activation by recruitment. Nature 386, 569-577.

Pulford, K., Lecointe, N., Leroy-Viard, K., Jones, M., Mathieu-Mahul, D., and Mason, D.Y. (1995). Expression of TAL-1 proteins in human tissues. Blood 85, 675-684.

Qiu, Y., Sharma, A., and Stein, R. (1998). p300 mediates transcriptional stimulation by the basic helix-loop-helix activators of the insulin gene. Molecular and cellular biology 18, 2957-2964.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B.D., Sun, Z.W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D., et al. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature 406, 593-599.

Rekhtman, N., Radparvar, F., Evans, T., and Skoultchi, A.I. (1999). Direct interaction of hematopoietic transcription factors PU.1 and GATA-1: functional antagonism in erythroid cells. Genes & development 13, 1398-1411.

Ren, B., Cam, H., Takahashi, Y., Volkert, T., Terragni, J., Young, R.A., and Dynlacht, B.D. (2002). E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints. Genes & development 16, 245-256.

Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., et al. (2000). Genome-wide location and function of DNA binding proteins. Science (New York, NY 290, 2306-2309.

Reynaud, D., Ravet, E., Titeux, M., Mazurier, F., Renia, L., Dubart-Kupperschmitt, A., Romeo, P.H., and Pflumio, F. (2005). SCL/TAL1 expression level regulates human hematopoietic stem cell self-renewal and engraftment. Blood 106, 2318-2328.

Rippe, K., von Hippel, P.H., and Langowski, J. (1995). Action at a distance: DNA-looping and initiation of transcription. Trends in biochemical sciences 20, 500-506.

Robb, L., Elwood, N.J., Elefanty, A.G., Kontgen, F., Li, R., Barnett, L.D., and Begley, C.G. (1996). The scl gene product is required for the generation of all hematopoietic lineages in the adult mouse. Embo J 15, 4123-4129.

Robb, L., Lyons, I., Li, R., Hartley, L., Kontgen, F., Harvey, R.P., Metcalf, D., and Begley, C.G. (1995). Absence of yolk sac hematopoiesis from mice with a targeted disruption of the scl gene. Proceedings of the National Academy of Sciences of the United States of America 92, 7075-7079.

Robert, F., Pokholok, D.K., Hannett, N.M., Rinaldi, N.J., Chandy, M., Rolfe, A., Workman, J.L., Gifford, D.K., and Young, R.A. (2004). Global position and recruitment of HATs and HDACs in the yeast genome. Mol Cell 16, 199-209.

Robertson, S.M., Kennedy, M., Shannon, J.M., and Keller, G. (2000). A transitional stage in the commitment of mesoderm to hematopoiesis requiring the transcription factor SCL/tal-1. Development (Cambridge, England) 127, 2447-2459.

Robyr, D., Suka, Y., Xenarios, I., Kurdistani, S.K., Wang, A., Suka, N., and Grunstein, M. (2002). Microarray deacetylation maps determine genome-wide functions for yeast histone deacetylases. Cell 109, 437-446.

Robzyk, K., Recht, J., and Osley, M.A. (2000). Rad6-dependent ubiquitination of histone H2B in yeast. Science (New York, NY 287, 501-504.

Roth, F.P., Hughes, J.D., Estep, P.W., and Church, G.M. (1998). Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation. Nature biotechnology 16, 939-945.

Rylski, M., Welch, J.J., Chen, Y.Y., Letting, D.L., Diehl, J.A., Chodosh, L.A., Blobel, G.A., and Weiss, M.J. (2003). GATA-1-mediated proliferation arrest during erythroid maturation. Molecular and cellular biology 23, 5031-5042.

Saha, A., Wittmeyer, J., and Cairns, B.R. (2006). Chromatin remodelling: the industrial revolution of DNA around histones. Nature reviews 7, 437-447.

Salghetti, S.E., Caudy, A.A., Chenoweth, J.G., and Tansey, W.P. (2001). Regulation of transcriptional activation domain function by ubiquitin. Science (New York, NY 293, 1651-1653.

Sanchez, M., Gottgens, B., Sinclair, A.M., Stanley, M., Begley, C.G., Hunter, S., and Green, A.R. (1999). An SCL 3' enhancer targets developing endothelium together with embryonic and adult haematopoietic progenitors. Development (Cambridge, England) 126, 3891-3904.

Sanchez, M.J., Bockamp, E.O., Miller, J., Gambardella, L., and Green, A.R. (2001). Selective rescue of early haematopoietic progenitors in Scl(-/-) mice by expressing Scl under the control of a stem cell enhancer. Development (Cambridge, England) 128, 4815-4827.

Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W.W., and Lenhard, B. (2004). JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic acids research 32, D91-94.

Sassone-Corsi, P., Mizzen, C.A., Cheung, P., Crosio, C., Monaco, L., Jacquot, S., Hanauer, A., and Allis, C.D. (1999). Requirement of Rsk-2 for epidermal growth factor-activated phosphorylation of histone H3. Science (New York, NY 285, 886-891.

Scacheri, P.C., Rozenblatt-Rosen, O., Caplen, N.J., Wolfsberg, T.G., Umayam, L., Lee, J.C., Hughes, C.M., Shanmugam, K.S., Bhattacharjee, A., Meyerson, M., et al. (2004). Short interfering RNAs can induce unexpected and divergent changes in the levels of untargeted proteins in mammalian cells. Proceedings of the National Academy of Sciences of the United States of America 101, 1892-1897.

Schena, M., Shalon, D., Davis, R.W., and Brown, P.O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science (New York, NY 270, 467-470.

Scherf, M., Klingenhoff, A., and Werner, T. (2000). Highly specific localization of promoter regions in large genomic sequences by PromoterInspector: a novel context analysis approach. Journal of molecular biology 297, 599-606.

Schiebel, W., Pelissier, T., Riedel, L., Thalmeir, S., Schiebel, R., Kempe, D., Lottspeich, F., Sanger, H.L., and Wassenegger, M. (1998). Isolation of an RNA-directed RNA polymerase-specific cDNA clone from tomato. The Plant cell 10, 2087-2101.

Schlabach, M.R., Luo, J., Solimini, N.L., Hu, G., Xu, Q., Li, M.Z., Zhao, Z., Smogorzewska, A., Sowa, M.E., Ang, X.L., et al. (2008). Cancer proliferation gene discovery through functional genomics. Science (New York, NY 319, 620-624.

Schmitter, D., Filkowski, J., Sewer, A., Pillai, R.S., Oakeley, E.J., Zavolan, M., Svoboda, P., and Filipowicz, W. (2006). Effects of Dicer and Argonaute down-regulation on mRNA levels in human HEK293 cells. Nucleic acids research 34, 4801-4815.

Schones, D.E., Cui, K., Cuddapah, S., Roh, T.Y., Barski, A., Wang, Z., Wei, G., and Zhao, K. (2008). Dynamic regulation of nucleosome positioning in the human genome. Cell 132, 887-898.

Schug, J., Overton, G.C. (1997). TESS: Transcriptional Element Search Software on the WWW. Technical Report CBIL-TR-1997-1001-v0.0. . Computational Biology and Informatics Laboratory, School of Medicine, University of Pennsylvania.

Schuh, A.H., Tipping, A.J., Clark, A.J., Hamlett, I., Guyot, B., Iborra, F.J., Rodriguez, P., Strouboulis, J., Enver, T., Vyas, P., et al. (2005). ETO-2 associates with SCL in erythroid cells and megakaryocytes and provides repressor functions in erythropoiesis. Molecular and cellular biology 25, 10235-10250.

Segal, E., Shapira, M., Regev, A., Pe'er, D., Botstein, D., Koller, D., and Friedman, N. (2003). Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. Nature genetics 34, 166-176.

Semizarov, D., Frost, L., Sarthy, A., Kroeger, P., Halbert, D.N., and Fesik, S.W. (2003). Specificity of short interfering RNA determined through gene expression signatures. Proceedings of the National Academy of Sciences of the United States of America 100, 6347-6352.

Sen, G., Wehrman, T.S., Myers, J.W., and Blau, H.M. (2004). Restriction enzyme-generated siRNA (REGS) vectors and libraries. Nature genetics 36, 183-189.

Sertil, O., Kapoor, R., Cohen, B.D., Abramova, N., and Lowry, C.V. (2003). Synergistic repression of anaerobic genes by Mot3 and Rox1 in Saccharomyces cerevisiae. Nucleic acids research 31, 5831-5837.

Shabalina, S.A., and Spiridonov, N.A. (2004). The mammalian transcriptome and the function of non-coding DNA sequences. Genome biology 5, 105.

Shannon, M.F., and Rao, S. (2002). Transcription. Of chips and ChIPs. Science (New York, NY 296, 666-669.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research 13, 2498-2504.

Shen-Orr, S.S., Milo, R., Mangan, S., and Alon, U. (2002). Network motifs in the transcriptional regulation network of Escherichia coli. Nature genetics 31, 64-68.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., and Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell 119, 941-953.

Shi, Y.J., Matson, C., Lan, F., Iwase, S., Baba, T., and Shi, Y. (2005). Regulation of LSD1 histone demethylase activity by its associated factors. Mol Cell 19, 857-864.

Shimizu, R., Trainor, C.D., Nishikawa, K., Kobayashi, M., Ohneda, K., and Yamamoto, M. (2007). GATA-1 self-association controls erythroid development in vivo. The Journal of biological chemistry 282, 15862-15871.

Shipp, M.A., Ross, K.N., Tamayo, P., Weng, A.P., Kutok, J.L., Aguiar, R.C., Gaasenbeek, M., Angelo, M., Reich, M., Pinkus, G.S., et al. (2002). Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. Nature medicine 8, 68-74.

Shiraki, T., Kondo, S., Katayama, S., Waki, K., Kasukawa, T., Kawaji, H., Kodzius, R., Watahiki, A., Nakamura, M., Arakawa, T., et al. (2003). Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. Proceedings of the National Academy of Sciences of the United States of America 100, 15776-15781.

Shivdasani, R.A., Fujiwara, Y., McDevitt, M.A., and Orkin, S.H. (1997). A lineage-selective knockout establishes the critical role of transcription factor GATA-1 in megakaryocyte growth and platelet development. Embo J 16, 3965-3973.

Shivdasani, R.A., Mayer, E.L., and Orkin, S.H. (1995). Absence of blood formation in mice lacking the T-cell leukaemia oncoprotein tal-1/SCL. Nature 373, 432-434.

Siddiqui, A.S., Delaney, A.D., Schnerch, A., Griffith, O.L., Jones, S.J., and Marra, M.A. (2006). Sequence biases in large scale gene expression profiling data. Nucleic acids research 34, e83.

Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S., et al. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome research 15, 1034-1050.

Sijen, T., Fleenor, J., Simmer, F., Thijssen, K.L., Parrish, S., Timmons, L., Plasterk, R.H., and Fire, A. (2001). On the role of RNA amplification in dsRNA-triggered gene silencing. Cell 107, 465-476.

Silberstein, L., Sanchez, M.J., Socolovsky, M., Liu, Y., Hoffman, G., Kinston, S., Piltz, S., Bowen, M., Gambardella, L., Green, A.R., et al. (2005). Transgenic analysis of the stem cell leukemia +19 stem cell enhancer in adult and embryonic hematopoietic and endothelial cells. Stem Cells 23, 1378-1388.

Silva, J.M., Li, M.Z., Chang, K., Ge, W., Golding, M.C., Rickles, R.J., Siolas, D., Hu, G., Paddison, P.J., Schlabach, M.R., et al. (2005). Second-generation shRNA libraries covering the mouse and human genomes. Nature genetics 37, 1281-1288.

Silva, J.M., Marran, K., Parker, J.S., Silva, J., Golding, M., Schlabach, M.R., Elledge, S.J., Hannon, G.J., and Chang, K. (2008). Profiling essential genes in human mammary cells by multiplex RNAi screening. Science (New York, NY 319, 617-620.

Silver, L., and Palis, J. (1997). Initiation of murine embryonic erythropoiesis: a spatial analysis. Blood 89, 1154-1164.

Singh-Gasson, S., Green, R.D., Yue, Y., Nelson, C., Blattner, F., Sussman, M.R., and Cerrina, F. (1999). Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. Nature biotechnology 17, 974-978.

Sledz, C.A., Holko, M., de Veer, M.J., Silverman, R.H., and Williams, B.R. (2003). Activation of the interferon system by short-interfering RNAs. Nature cell biology 5, 834-839.

Smale, S.T., and Kadonaga, J.T. (2003). The RNA polymerase II core promoter. Annu Rev Biochem 72, 449-479.

Smardon, A., Spoerke, J.M., Stacey, S.C., Klein, M.E., Mackin, N., and Maine, E.M. (2000). EGO-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in C. elegans. Curr Biol 10, 169-178.

Smith, E.M., Gisler, R., and Sigvardsson, M. (2002). Cloning and characterization of a promoter flanking the early B cell factor (EBF) gene indicates roles for E-proteins and autoregulation in the control of EBF expression. J Immunol 169, 261-270.

Song, E., Zhu, P., Lee, S.K., Chowdhury, D., Kussman, S., Dykxhoorn, D.M., Feng, Y., Palliser, D., Weiner, D.B., Shankar, P., et al. (2005). Antibody mediated in vivo delivery of small interfering RNAs via cell-surface receptors. Nature biotechnology 23, 709-717.

Song, S.H., Hou, C., and Dean, A. (2007). A positive role for NLI/Ldb1 in long-range beta-globin locus control region function. Mol Cell 28, 810-822.

Sonnichsen, B., Koski, L.B., Walsh, A., Marschall, P., Neumann, B., Brehm, M., Alleaume, A.M., Artelt, J., Bettencourt, P., Cassin, E., et al. (2005). Full-genome RNAi profiling of early embryogenesis in Caenorhabditis elegans. Nature 434, 462-469.

Spicuglia, S., Kumar, S., Yeh, J.H., Vachez, E., Chasson, L., Gorbatch, S., Cautres, J., and Ferrier, P. (2002). Promoter activation by enhancer-dependent and -independent loading of activator and coactivator complexes. Mol Cell 10, 1479-1487.

Spilianakis, C.G., Lalioti, M.D., Town, T., Lee, G.R., and Flavell, R.A. (2005). Interchromosomal associations between alternatively expressed loci. Nature 435, 637-645.

Stewart, S.A., Dykxhoorn, D.M., Palliser, D., Mizuno, H., Yu, E.Y., An, D.S., Sabatini, D.M., Chen, I.S., Hahn, W.C., Sharp, P.A., et al. (2003). Lentivirus-delivered stable gene silencing by RNAi in primary cells. RNA (New York, NY 9, 493-501.

Stielow, B., Sapetschnig, A., Kruger, I., Kunert, N., Brehm, A., Boutros, M., and Suske, G. (2008). Identification of SUMO-dependent chromatin-associated transcriptional repression components by a genome-wide RNAi screen. Mol Cell 29, 742-754.

Stopka, T., Amanatullah, D.F., Papetti, M., and Skoultchi, A.I. (2005). PU.1 inhibits the erythroid program by binding to GATA-1 on DNA and creating a repressive chromatin structure. Embo J 24, 3712-3723.

Stopka T, Skoultchi AI. The ISWI ATPase Snf2h is required for early mouse development. (2003). Proc Natl Acad Sci U S A. 100, 14097-102.

Stopka T, Zakova D, Fuchs O, Kubrova O, Blafkova J, Jelinek J, Necas E, Zivny J. (2000). Chromatin remodeling gene SMARCA5 is dysregulated in primitive hematopoietic cells of acute leukemia. Leukemia 14, 1247-1252.

Studitsky, V.M., Clark, D.J., and Felsenfeld, G. (1994). A histone octamer can step around a transcribing polymerase without leaving the template. Cell 76, 371-382.

Sun, X.H., and Baltimore, D. (1991). An inhibitory domain of E12 transcription factor prevents DNA binding in E12 homodimers but not in E12 heterodimers. Cell 64, 459-470.

Suzuki, Y., Tsunoda, T., Sese, J., Taira, H., Mizushima-Sugano, J., Hata, H., Ota, T., Isogai, T., Tanaka, T., Nakamura, Y., et al. (2001). Identification and characterization of the potential promoter regions of 1031 kinds of human genes. Genome research 11, 677-684.

Swiers, G., Patient, R., and Loose, M. (2006). Genetic regulatory networks programming hematopoietic stem cells and erythroid lineage specification. Developmental biology 294, 525-540.

Szekvolgyi, L., Balint, B.L., Imre, L., Goda, K., Szabo, M., Nagy, L., and Szabo, G. (2006). Chip-on-beads: flow-cytometric evaluation of chromatin immunoprecipitation. Cytometry A 69, 1086-1091.

Tabara, H., Grishok, A., and Mello, C.C. (1998). RNAi in C. elegans: soaking in the genome sequence. Science (New York, NY 282, 430-431.

Tavazoie, S., Hughes, J.D., Campbell, M.J., Cho, R.J., and Church, G.M. (1999). Systematic determination of genetic network architecture. Nature genetics 22, 281-285.

Tavernarakis, N., Wang, S.L., Dorovkov, M., Ryazanov, A., and Driscoll, M. (2000). Heritable and inducible genetic interference by double-stranded RNA encoded by transgenes. Nature genetics 24, 180-183.

Thomson, S., Clayton, A.L., Hazzalin, C.A., Rose, S., Barratt, M.J., and Mahadevan, L.C. (1999). The nucleosomal response associated with immediate-early gene induction is mediated via alternative MAP kinase cascades: MSK1 as a potential histone H3/HMG-14 kinase. Embo J 18, 4779-4793.

Thut, C.J., Chen, J.L., Klemm, R., and Tjian, R. (1995). p53 transcriptional activation mediated by coactivators TAFII40 and TAFII60. Science (New York, NY 267, 100-104.

Timmons, L., Court, D.L., and Fire, A. (2001). Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in Caenorhabditis elegans. Gene 263, 103-112.

Timmons, L., and Fire, A. (1998). Specific interference by ingested dsRNA. Nature 395, 854.

Torres, T.T., Metta, M., Ottenwalder, B., and Schlotterer, C. (2008). Gene expression profiling by massively parallel sequencing. Genome research 18, 172-177.

Tsai, F.Y., and Orkin, S.H. (1997). Transcription factor GATA-2 is required for proliferation/survival of early hematopoietic cells and mast cell formation, but not for erythroid and myeloid terminal differentiation. Blood 89, 3636-3643.

Tsai, S.F., Martin, D.I., Zon, L.I., D'Andrea, A.D., Wong, G.G., and Orkin, S.H. (1989). Cloning of cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells. Nature 339, 446-451.

Tsang, A.P., Visvader, J.E., Turner, C.A., Fujiwara, Y., Yu, C., Weiss, M.J., Crossley, M., and Orkin, S.H. (1997). FOG, a multitype zinc finger protein, acts as a cofactor for transcription factor GATA-1 in erythroid and megakaryocytic differentiation. Cell 90, 109-119.

Tse, C., Sera, T., Wolffe, A.P., and Hansen, J.C. (1998). Disruption of higher-order folding by core histone acetylation dramatically enhances transcription of nucleosomal arrays by RNA polymerase III. Molecular and cellular biology 18, 4629-4638.

Tuerk, C., and Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science (New York, NY 249, 505-510.

Tusher, V.G., Tibshirani, R., and Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. Proceedings of the National Academy of Sciences of the United States of America 98, 5116-5121.

Utley, R.T., Ikeda, K., Grant, P.A., Cote, J., Steger, D.J., Eberharter, A., John, S., and Workman, J.L. (1998). Transcriptional activators direct histone acetyltransferase complexes to nucleosomes. Nature 394, 498-502.

Valverde-Garduno, V., Guyot, B., Anguita, E., Hamlett, I., Porcher, C., and Vyas, P. (2004). Differences in the chromatin structure and cis-element organization of the human and mouse GATA1 loci: implications for cis-element identification. Blood 104, 3106-3116.

van de Wetering, M., Oving, I., Muncan, V., Pon Fong, M.T., Brantjes, H., van Leenen, D., Holstege, F.C., Brummelkamp, T.R., Agami, R., and Clevers, H. (2003). Specific inhibition of gene expression using a stably integrated, inducible small-interfering-RNA vector. EMBO reports 4, 609-615.

van Eekelen, J.A., Bradley, C.K., Gothert, J.R., Robb, L., Elefanty, A.G., Begley, C.G., and Harvey, A.R. (2003). Expression pattern of the stem cell leukaemia gene in the CNS of the embryonic and adult mouse. Neuroscience 122, 421-436.

van Haaften, G., Romeijn, R., Pothof, J., Koole, W., Mullenders, L.H., Pastink, A., Plasterk, R.H., and Tijsterman, M. (2006). Identification of conserved pathways of DNA-damage response and radiation protection by genome-wide RNAi. Curr Biol 16, 1344-1350.

van Steensel, B., and Henikoff, S. (2000). Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. Nature biotechnology 18, 424-428.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., and Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome biology 3, RESEARCH0034.

Velculescu, V.E., Zhang, L., Vogelstein, B., and Kinzler, K.W. (1995). Serial analysis of gene expression. Science (New York, NY 270, 484-487.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., et al. (2001). The sequence of the human genome. Science (New York, NY 291, 1304-1351.

Vernimmen, D., De Gobbi, M., Sloane-Stanley, J.A., Wood, W.G., and Higgs, D.R. (2007). Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. Embo J 26, 2041-2051.

Viens, A., Mechold, U., Lehrmann, H., Harel-Bellan, A., and Ogryzko, V. (2004). Use of protein biotinylation in vivo for chromatin immunoprecipitation. Analytical biochemistry 325, 68-76.

Vilar, J.M., and Saiz, L. (2005). DNA looping in gene regulation: from the assembly of macromolecular complexes to the control of transcriptional noise. Current opinion in genetics & development 15, 136-144.

Visel, A., Minovitsky, S., Dubchak, I., and Pennacchio, L.A. (2007). VISTA Enhancer Browser--a database of tissue-specific human enhancers. Nucleic acids research 35, D88-92.

Visvader, J., Begley, C.G., and Adams, J.M. (1991). Differential expression of the LYL, SCL and E2A helix-loop-helix genes within the hemopoietic system. Oncogene 6, 187-194.

Visvader, J.E., Fujiwara, Y., and Orkin, S.H. (1998). Unsuspected role for the T-cell leukemia protein SCL/tal-1 in vascular development. Genes & development 12, 473-479.

Visvader, J.E., Mao, X., Fujiwara, Y., Hahm, K., and Orkin, S.H. (1997). The LIM-domain binding protein Ldb1 and its partner LMO2 act as negative regulators of erythroid differentiation. Proceedings of the National Academy of Sciences of the United States of America 94, 13707-13712.

Vitelli, L., Condorelli, G., Lulli, V., Hoang, T., Luchetti, L., Croce, C.M., and Peschle, C. (2000). A pentamer transcriptional complex including tal-1 and retinoblastoma protein downmodulates c-kit expression in normal erythroblasts. Molecular and cellular biology 20, 5330-5342.

Vogel, M.J., Guelen, L., de Wit, E., Peric-Hupkes, D., Loden, M., Talhout, W., Feenstra, M., Abbas, B., Classen, A.K., and van Steensel, B. (2006). Human heterochromatin proteins form large domains containing KRAB-ZNF genes. Genome research 16, 1493-1504.

Vogel, M.J., Peric-Hupkes, D., and van Steensel, B. (2007). Detection of in vivo protein-DNA interactions using DamID in mammalian cells. Nature protocols 2, 1467-1478.

Vokes, S.A., Ji, H., McCuine, S., Tenzen, T., Giles, S., Zhong, S., Longabaugh, W.J., Davidson, E.H., Wong, W.H., and McMahon, A.P. (2007). Genomic characterization of Gli-activator targets in sonic hedgehog-mediated neural patterning. Development (Cambridge, England) 134, 1977-1989.

Vyas, P., Ault, K., Jackson, C.W., Orkin, S.H., and Shivdasani, R.A. (1999). Consequences of GATA-1 deficiency in megakaryocytes and platelets. Blood 93, 2867-2875.

Vyas, P., McDevitt, M.A., Cantor, A.B., Katz, S.G., Fujiwara, Y., and Orkin, S.H. (1999). Different sequence requirements for expression in erythroid and megakaryocytic cells within a regulatory element upstream of the GATA-1 gene. Development (Cambridge, England) 126, 2799-2811.

Wadman, I., Li, J., Bash, R.O., Forster, A., Osada, H., Rabbitts, T.H., and Baer, R. (1994). Specific in vivo association between the bHLH and LIM proteins implicated in human T cell leukemia. Embo J 13, 4831-4839.

Wadman, I.A., Osada, H., Grutz, G.G., Agulnick, A.D., Westphal, H., Forster, A., and Rabbitts, T.H. (1997). The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. Embo J 16, 3145-3157.

Wakabayashi, J., Yomogida, K., Nakajima, O., Yoh, K., Takahashi, S., Engel, J.D., Ohneda, K., and Yamamoto, M. (2003). GATA-1 testis activation region is essential for Sertoli cell-specific expression of GATA-1 gene in transgenic mouse. Genes Cells 8, 619-630.

Walker, G.T. (1993). Empirical aspects of strand displacement amplification. PCR methods and applications 3, 1-6.

Wall, L., deBoer, E., and Grosveld, F. (1988). The human beta-globin gene 3' enhancer contains multiple binding sites for an erythroid-specific protein. Genes & development 2, 1089-1100.

Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R.S., and Zhang, Y. (2004). Role of histone H2A ubiquitination in Polycomb silencing. Nature 431, 873-878.

Wang, J., Hoshino, T., Redner, R.L., Kajigaya, S., and Liu, J.M. (1998). ETO, fusion partner in t(8;21) acute myeloid leukemia, represses transcription by interaction with the human N-CoR/mSin3/HDAC1 complex. Proceedings of the National Academy of Sciences of the United States of America 95, 10860-10865.

Wang, Y., Wysocka, J., Sayegh, J., Lee, Y.H., Perlin, J.R., Leonelli, L., Sonbuchner, L.S., McDonald, C.H., Cook, R.G., Dou, Y., et al. (2004). Human PAD4 regulates histone arginine methylation levels via demethylimination. Science (New York, NY 306, 279-283.

Warren, A.J., Colledge, W.H., Carlton, M.B., Evans, M.J., Smith, A.J., and Rabbitts, T.H. (1994). The oncogenic cysteine-rich LIM domain protein rbtn2 is essential for erythroid development. Cell 78, 45-57.

Wasserman, W.W., and Fickett, J.W. (1998). Identification of regulatory regions which confer muscle-specific gene expression. Journal of molecular biology 278, 167-181.

Waterhouse, P.M., Wang, M.B., and Lough, T. (2001). Gene silencing as an adaptive defence against viruses. Nature 411, 834-842.

Weber, F., de Villiers, J., and Schaffner, W. (1984). An SV40 "enhancer trap" incorporates exogenous enhancers or generates enhancers from its own sequences. Cell 36, 983-992.

Wechsler, J., Greene, M., McDevitt, M.A., Anastasi, J., Karp, J.E., Le Beau, M.M., and Crispino, J.D. (2002). Acquired mutations in GATA1 in the megakaryoblastic leukemia of Down syndrome. Nature genetics 32, 148-152.

Wei, C.L., Wu, Q., Vega, V.B., Chiu, K.P., Ng, P., Zhang, T., Shahab, A., Yong, H.C., Fu, Y., Weng, Z., et al. (2006). A global map of p53 transcription-factor binding sites in the human genome. Cell 124, 207-219.

Weinberg, M.S., Villeneuve, L.M., Ehsani, A., Amarzguioui, M., Aagaard, L., Chen, Z.X., Riggs, A.D., Rossi, J.J., and Morris, K.V. (2006). The antisense strand of small interfering RNAs directs histone methylation and transcriptional gene silencing in human cells. RNA (New York, NY 12, 256-262.

Weinmann, A.S., Bartley, S.M., Zhang, T., Zhang, M.Q., and Farnham, P.J. (2001). Use of chromatin immunoprecipitation to clone novel E2F target promoters. Molecular and cellular biology 21, 6820-6832.

Weinmann, A.S., Yan, P.S., Oberley, M.J., Huang, T.H., and Farnham, P.J. (2002). Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis. Genes & development 16, 235-244.

Weiss, M.J., Keller, G., and Orkin, S.H. (1994). Novel insights into erythroid development revealed through in vitro differentiation of GATA-1 embryonic stem cells. Genes & development 8, 1184-1197.

Weiss, M.J., and Orkin, S.H. (1995). Transcription factor GATA-1 permits survival and maturation of erythroid precursors by preventing apoptosis. Proceedings of the National Academy of Sciences of the United States of America 92, 9623-9627.

Welch, J.J., Watts, J.A., Vakoc, C.R., Yao, Y., Wang, H., Hardison, R.C., Blobel, G.A., Chodosh, L.A., and Weiss, M.J. (2004). Global regulation of erythroid gene expression by transcription factor GATA-1. Blood 104, 3136-3147.

Wells, J., Yan, P.S., Cechvala, M., Huang, T., and Farnham, P.J. (2003). Identification of novel pRb binding sites using CpG microarrays suggests that E2F recruits pRb to specific genomic sites during S phase. Oncogene 22, 1445-1460.

West, A.G., Gaszner, M., and Felsenfeld, G. (2002). Insulators: many functions, many mechanisms. Genes & development 16, 271-288.

Westbrook, T.F., Martin, E.S., Schlabach, M.R., Leng, Y., Liang, A.C., Feng, B., Zhao, J.J., Roberts, T.M., Mandel, G., Hannon, G.J., et al. (2005). A genetic screen for candidate tumor suppressors identifies REST. Cell 121, 837-848.

White, K.P., Rifkin, S.A., Hurban, P., and Hogness, D.S. (1999). Microarray analysis of Drosophila development during metamorphosis. Science (New York, NY 286, 2179-2184.

Wightman, B., Ha, I., and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. Cell 75, 855-862.

Williams, A.O., Isaacs, R.J., and Stowell, K.M. (2007). Down-regulation of human topoisomerase IIalpha expression correlates with relative amounts of specificity factors Sp1 and Sp3 bound at proximal and distal promoter regions. BMC molecular biology 8, 36.

Williams, B.R. (1997). Role of the double-stranded RNA-activated protein kinase (PKR) in cell regulation. Biochem Soc Trans 25, 509-513.

Williams, R.R., Azuara, V., Perry, P., Sauer, S., Dvorkina, M., Jorgensen, H., Roix, J., McQueen, P., Misteli, T., Merkenschlager, M., et al. (2006). Neural induction promotes large-scale chromatin reorganisation of the Mash1 locus. Journal of cell science 119, 132-140.

Wolfrum, C., Shi, S., Jayaprakash, K.N., Jayaraman, M., Wang, G., Pandey, R.K., Rajeev, K.G., Nakayama, T., Charrise, K., Ndungo, E.M., et al. (2007). Mechanisms and optimization of in vivo delivery of lipophilic siRNAs. Nature biotechnology 25, 1149-1157.

Wolstein, O., Silkov, A., Revach, M., and Dikstein, R. (2000). Specific interaction of TAFII105 with OCA-B is involved in activation of octamer-dependent transcription. The Journal of biological chemistry 275, 16459-16465.

Wong, J.M., and Bateman, E. (1994). TBP-DNA interactions in the minor groove discriminate between A:T and T:A base pairs. Nucleic acids research 22, 1890-1896.

Wood, A., Krogan, N.J., Dover, J., Schneider, J., Heidt, J., Boateng, M.A., Dean, K., Golshani, A., Zhang, Y., Greenblatt, J.F., et al. (2003). Bre1, an E3 ubiquitin ligase required for recruitment and substrate selection of Rad6 at a promoter. Mol Cell 11, 267-274.

Woolfe, A., Goodson, M., Goode, D.K., Snell, P., McEwen, G.K., Vavouri, T., Smith, S.F., North, P., Callaway, H., Kelly, K., et al. (2005). Highly conserved non-coding sequences are associated with vertebrate development. PLoS biology 3, e7.

Wright, W.E., Binder, M., and Funk, W. (1991). Cyclic amplification and selection of targets (CASTing) for the myogenin consensus binding site. Molecular and cellular biology 11, 4104-4110.

Wu, C., Wong, Y.C., and Elgin, S.C. (1979). The chromatin structure of specific genes: II. Disruption of chromatin structure during gene activity. Cell 16, 807-814.

Wu, Z., and Irizarry, R.A. (2005). Stochastic models inspired by hybridization theory for short oligonucleotide arrays. J Comput Biol 12, 882-893.

Wyrick, J.J., Aparicio, J.G., Chen, T., Barnett, J.D., Jennings, E.G., Young, R.A., Bell, S.P., and Aparicio, O.M. (2001). Genome-wide distribution of ORC and MCM proteins in S. cerevisiae: high-resolution mapping of replication origins. Science (New York, NY 294, 2357-2360.

Xu, Z., Huang, S., Chang, L.S., Agulnick, A.D., and Brandt, S.J. (2003). Identification of a TAL1 target gene reveals a positive role for the LIM domain-binding protein Ldb1 in erythroid gene expression and differentiation. Molecular and cellular biology 23, 7585-7599.

Xu, Z., Meng, X., Cai, Y., Koury, M.J., and Brandt, S.J. (2006). Recruitment of the SWI/SNF protein Brg1 by a multiprotein complex effects transcriptional repression in murine erythroid progenitors. The Biochemical journal 399, 297-304.

Yamada, Y., Pannell, R., Forster, A., and Rabbitts, T.H. (2000). The oncogenic LIM-only transcription factor Lmo2 regulates angiogenesis but not vasculogenesis in mice. Proceedings of the National Academy of Sciences of the United States of America 97, 320-324.

Yamada, Y., Warren, A.J., Dobson, C., Forster, A., Pannell, R., and Rabbitts, T.H. (1998). The T cell leukemia LIM protein Lmo2 is necessary for adult mouse hematopoiesis. Proceedings of the National Academy of Sciences of the United States of America 95, 3890-3895.

Yamaguchi, Y., Ackerman, S.J., Minegishi, N., Takiguchi, M., Yamamoto, M., and Suda, T. (1998). Mechanisms of transcription in eosinophils: GATA-1, but not GATA-2, transactivates the promoter of the eosinophil granule major basic protein gene. Blood 91, 3447-3458.

Yamaguchi, Y., Nishio, H., Kishi, K., Ackerman, S.J., and Suda, T. (1999). C/EBPbeta and GATA-1 synergistically regulate activity of the eosinophil granule major basic protein promoter: implication for C/EBPbeta activity in eosinophil gene expression. Blood 94, 1429-1439.

Yang, D., Buchholz, F., Huang, Z., Goga, A., Chen, C.Y., Brodsky, F.M., and Bishop, J.M. (2002). Short RNA duplexes produced by hydrolysis with Escherichia coli RNase III mediate effective RNA interference in mammalian cells. Proceedings of the National Academy of Sciences of the United States of America 99, 9942-9947.

Yi, Z.C., Wang, Z., Li, H.X., Liu, M.J., Wu, R.C., and Wang, X.H. (2004). Effects of chebulinic acid on differentiation of human leukemia K562 cells. Acta pharmacologica Sinica 25, 231-238.

Yoder, M.C., Hiatt, K., and Mukherjee, P. (1997). In vivo repopulating hematopoietic stem cells are present in the murine yolk sac at day 9.0 postcoitus. Proceedings of the National Academy of Sciences of the United States of America 94, 6776-6780.

Yokomizo, T., Takahashi, S., Mochizuki, N., Kuroha, T., Ema, M., Wakamatsu, A., Shimizu, R., Ohneda, O., Osato, M., Okada, H., et al. (2007). Characterization of GATA-1(+) hemangioblastic cells in the mouse embryo. Embo J 26, 184-196.

Yomogida, K., Ohtani, H., Harigae, H., Ito, E., Nishimune, Y., Engel, J.D., and Yamamoto, M. (1994). Developmental stage- and spermatogenic cycle-specific expression of transcription factor GATA-1 in mouse Sertoli cells. Development (Cambridge, England) 120, 1759-1766.

Yoo, J., Herman, L.E., Li, C., Krantz, S.B., and Tuan, D. (1996). Dynamic changes in the locus control region of erythroid progenitor cells demonstrated by polymerase chain reaction. Blood 87, 2558-2567.

Yoo, J.Y., Kim, J.H., Kim, J., Huang, J.H., Zhang, S.N., Kang, Y.A., Kim, H., and Yun, C.O. (2008). Short hairpin RNA-expressing oncolytic adenovirus-mediated inhibition of IL-8: effects on antiangiogenesis and tumor growth inhibition. Gene therapy 15, 635-651.

Yu, C., Cantor, A.B., Yang, H., Browne, C., Wells, R.A., Fujiwara, Y., and Orkin, S.H. (2002). Targeted deletion of a high-affinity GATA-binding site in the GATA-1 promoter leads to selective loss of the eosinophil lineage in vivo. The Journal of experimental medicine 195, 1387-1395.

Yu, C., Niakan, K.K., Matsushita, M., Stamatoyannopoulos, G., Orkin, S.H., and Raskind, W.H. (2002). X-linked thrombocytopenia with thalassemia from a mutation in the amino finger of GATA-1 affecting DNA binding rather than FOG-1 interaction. Blood 100, 2040-2045.

Zeller, K.I., Zhao, X., Lee, C.W., Chiu, K.P., Yao, F., Yustein, J.T., Ooi, H.S., Orlov, Y.L., Shahab, A., Yong, H.C., et al. (2006). Global mapping of c-Myc binding sites and target gene networks in human B cells. Proceedings of the National Academy of Sciences of the United States of America 103, 17834-17839.

Zeng, P.Y., Vakoc, C.R., Chen, Z.C., Blobel, G.A., and Berger, S.L. (2006). In vivo dual cross-linking for identification of indirect DNA-associated proteins by chromatin immunoprecipitation. BioTechniques 41, 694, 696, 698.

Zeng, Y., Wagner, E.J., and Cullen, B.R. (2002). Both natural and designed micro RNAs can inhibit the expression of cognate mRNAs when expressed in human cells. Mol Cell 9, 1327-1333.

Zenzie-Gregory, B., Khachi, A., Garraway, I.P., and Smale, S.T. (1993). Mechanism of initiator-mediated transcription: evidence for a functional interaction between the TATA-binding protein and DNA in the absence of a specific recognition sequence. Molecular and cellular biology 13, 3841-3849.

Zhang, P., Zhang, X., Iwama, A., Yu, C., Smith, K.A., Mueller, B.U., Narravula, S., Torbett, B.E., Orkin, S.H., and Tenen, D.G. (2000). PU.1 inhibits GATA-1 function and erythroid differentiation by blocking GATA-1 DNA binding. Blood 96, 2641-2648.

Zhang, W., Yang, H., Kong, X., Mohapatra, S., San Juan-Vergara, H., Hellermann, G., Behera, S., Singam, R., Lockey, R.F., and Mohapatra, S.S. (2005). Inhibition of respiratory syncytial virus infection with intranasal siRNA nanoparticles targeting the viral NS1 gene. Nature medicine 11, 56-62.

Zhang, Y., Payne, K.J., Zhu, Y., Price, M.A., Parrish, Y.K., Zielinska, E., Barsky, L.W., and Crooks, G.M. (2005). SCL expression at critical points in human hematopoietic lineage commitment. Stem Cells 23, 852-860.

Zhao, H., and Dean, A. (2004). An insulator blocks spreading of histone acetylation and interferes with RNA polymerase II transfer between an enhancer and gene. Nucleic acids research 32, 4903-4919.

Zhao, Z., Tavoosidana, G., Sjolinder, M., Gondor, A., Mariano, P., Wang, S., Kanduri, C., Lezcano, M., Sandhu, K.S., Singh, U., et al. (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. Nature genetics 38, 1341-1347.

Zhu, B., Zheng, Y., Pham, A.D., Mandal, S.S., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. (2005). Monoubiquitination of human histone H2B: the factors involved and their roles in HOX gene regulation. Mol Cell 20, 601-611.

Zhu, P., Zhou, W., Wang, J., Puc, J., Ohgi, K.A., Erdjument-Bromage, H., Tempst, P., Glass, C.K., and Rosenfeld, M.G. (2007). A histone H2A deubiquitinase complex coordinating histone acetylation and H1 dissociation in transcriptional regulation. Mol Cell 27, 609-621.

Zhuang, Y., Jackson, A., Pan, L., Shen, K., and Dai, M. (2004). Regulation of E2A gene expression in B-lymphocyte development. Molecular immunology 40, 1165-1177.

Zhuang, Y., Soriano, P., and Weintraub, H. (1994). The helix-loop-helix gene E2A is required for B cell formation. Cell 79, 875-884.

Zimmermann, T.S., Lee, A.C., Akinc, A., Bramlage, B., Bumcrot, D., Fedoruk, M.N., Harborth, J., Heyes, J.A., Jeffs, L.B., John, M., et al. (2006). RNAi-mediated gene silencing in non-human primates. Nature 441, 111-114.

Zon, L.I., Tsai, S.F., Burgess, S., Matsudaira, P., Bruns, G.A., and Orkin, S.H. (1990). The major human erythroid DNA-binding protein (GF-1): primary sequence and localization of the gene to the X chromosome. Proceedings of the National Academy of Sciences of the United States of America 87, 668-672.

Zon, L.I., Yamaguchi, Y., Yee, K., Albee, E.A., Kimura, A., Bennett, J.C., Orkin, S.H., and Ackerman, S.J. (1993). Expression of mRNA for the GATA-binding proteins in human eosinophils and basophils: potential role in gene transcription. Blood 81, 3234-3241.

Zon, L.I., Youssoufian, H., Mather, C., Lodish, H.F., and Orkin, S.H. (1991). Activation of the erythropoietin receptor promoter by transcription factor GATA-1. Proceedings of the National Academy of Sciences of the United States of America 88, 10638-10641.

Zweidler-Mckay, P.A., Grimes, H.L., Flubacher, M.M., and Tsichlis, P.N. (1996). Gfi-1 encodes a nuclear zinc finger protein that binds DNA and functions as a transcriptional repressor. Molecular and cellular biology 16, 4024-4034.

Appendix 1A

Sequences of primer pairs used in qPCR to analyse expression of the SCL complex during siRNA knockdown assays

Gene	Forward primer sequences (5'→3')	Reverse primer sequences (5'→3')
GATA1	CAAGCTACACCAGGTGAACCG	AGCTGGTCCTTCGGCTGC
SCL	TTTTGTGAAGACGGCACGG	TGAGAGCTGACAACCCCAGG
E12	CAACTGCACCTCAACAGCGAG	GCCGTTTCAAACAGGCTGC
E47	AGGTGCTGTCCCTGGAGGAG	CCGACTTGAGGTGCATCTGG
LDB1	CCAGCTAGCACCTTCGCC	GTCGTCAATGCCGTTGGC
LMO2	CGGCGCCTCTACTACAAACTG	CATCTCATAGGCACGAATCCG

Appendix 1B

Sequences of primer pairs used in qPCR to analyse expression of housekeeping genes during siRNA knockdown assays

Gene	Forward primer sequences (5'→3')	Reverse primer sequences (5'→3')
β-actin	AGAAGGAGATCACTGCCCTGG	CACATCTGCTGGAAGGTGGAC
GAPDH	AGGTCCACCACTGACACGTTG	AGCTGAACGGGAAGCTCACT
RPL16	GGCTTGCCTCTAGTGTCCTC	CTGATCTGCTGACGGGAGTT
β-tubulin	GCAGATGCTTAACGTGCAGA	CAATGAAGGTGACTGCCATC

Appendix 1C

Sequences of primer pairs used in qPCR to analyse expression of ChIP-on-chip target genes during time-course siRNA knockdown assays

Gene	Forward primer sequences (5'→3')	Reverse primer sequences (5'→3')
BRD2	TAGGCCCTTCTGGCTTTGGA	CATGGGCCTGCTCTCTTCCT
CTCFL	AAGCTGCGAAGGGATGGAAG	TCTCTGCAGGCGACAGGAAA
EKLF	CACCAAGAGCTCCCACCTGA	CCCCGTGTGTTTCCGGTAGT
EPOR	CGAGCCCAGAGAGCGAGTTT	AGGACTTCCAGGGAAGCAGGT
ETO2	CGGTCATCAACCAGCAGGAG	CTTCTCCCAGTCCCGATGCT
EZH2	TGGGAAAGTACACGGGGATA	CAGGATCGTCTCCATCATCA
FBXL10	AGCAGACAGAAGCCACGAAC	TGTGGAAGTCGGTGAAACAA
JMJD2C	TGACTGGCCTTATGTGGTGA	GTTTGACCCACGGAAATGAC

LMO2	CGGCGCCTCTACTACAAACTG	CATCTCATAGGCACGAATCCG
LYL1	CATCTTCCCTAGCAGCCGGTTG	GTTGGTGAACACGCGCCG
PCQAP	AAGCCCGCTGTGGATAGACC	GTGTTGAGCAAGGCGGTGAC
RSF1	AAGGCGAGTACACAAGCGAAGA	GCTTTCGAACTGACCGCTTTG
SCL	TTTTGTGAAGACGGCACGG	TGAGAGCTGACAACCCCAGG
SMARCA5	AAGGCAGAGAAAAAGAAACGAGGA	TTTCTTCCTCGACCATCAGGTG

Appendix 1D

Sequences of primer pairs used in ChIP-qPCR to analyse enrichments of ChIP-on-chip target genes

Region name	Forward primer sequences (5'→3')	Reverse primer sequences (5'→3')
BRD2	GGTTCACGGCACGGAGAGAT	AGGGAGTTTCTCGCCCTCGT
CTCFL	GGGTTGAAGTGGATGAGGAA	CCAGTATCTCAGTGCCTCCTG
EKLF (1)	TGCAACCCTTCTTCCCCTGT	GCGGCAAGAGCTACACCAAG
EKLF (2)	GGCTGCCTCGTGAACTCTGA	GGCCAACGTGAAGTTTGTGC
EPOR	GGGACGCGATCAGGAGTCTT	GCCATGCCTGTTTCTGGACTT
ETO2	CAGGCTGGGGAAGGTCTCC	GGCCCCAGATGGTTCCTGT
EZH2 (1)	GGGGAACTGGTTCCAGGACA	TCTTATATCCCAGGAAGCCAGGTAAA
EZH2 (2)	GAGAGGAAAGGAGAAATTGTTCATTGTT	GTGTTTTTGTATTATTTGAATGTGGGAAAC
EZH2 (3)	GAGTCCTGAGGCCAATGGGTA	TTCGCCCCTTTGTTACAGCTC
FBXL10	GCGCTGGGTCACACAGTACA	CGCCCTCTGGAAACTGACCT
JMJD2C (1)	CTTCCGGGCAAGGTTCTGTG	CGCGCTGTGGTTAACTTAGGC
JMJD2C (2)	CCCGTTAGCCTTAGCTCAATTAATCA	TTGGGCTAATATGCTGAATTTTCTGTT
JMJD2C (3)	CTCGACGGGAGGGTGAGG	TTGGGAGACTTGTTCCGCACT
LMO2 (1)	TGGGCTAATTGCCTGCTTTG	CCTGCCCTCAGCCGTTAAGT
LMO2 (2)	CACTGGAGCCAATGAGGGAAG	GCAGACTTCCGACTGGCAGA
LYL1 (1)	CCCGGTTTCCTCCCTCTCAC	TGGTTTCCTCCGGGGTCAG
LYL1 (2)	GGGCCTCAGGGCAGGAAG	GTAGCCCCCACGTGTCTTCG
LYL1 +33 region	GGGCCTGCGAACAGGAGATA	CCTCGTGGCTGCTCTGCTTT
PCQAP (1)	GGGGTCTGGATCTCAGGGACTA	ACACACGCCCCTCTTCCAGT
PCQAP (2)	AAAATTAGCTGGGCTTCGTAGCAG	TGTTGCCCAGACTGGAGTGC
PCQAP (3)	CTCCTGAGGCCCATGTTGGT	AGTTCTGCCTCCTTGGAACTATGG
PCQAP (4)	CCTCACCACCGACTGCTTGT	TTTTCGGACTCAGCCCACCT
RSF1	CAGCGGCACCAGAGAGAGAA	GTCGTCTCCCTCCCATTTGC
SCL (Promoter 1a)	CGCCGCAGAGATAAGGCACT	CCCACTCCCTCCGGTGAAAT

SCL +51 enhancer	TGACCTTACAGCCCTTCACCC	AGCTCCCTGCTCCCAGCAC
SMARCA5 (1)	TTCTTCTTCCCGGTCCTTGC	GGGCCTTCTCTCCCTTCACA
SMARCA5 (2)	TCAACTCTCGAAAAATGTGTCTCCTT	TGTAGTCTGTTTGATAACGAATCTGCAT
SMARCA5 (3)	AAGTCGGTGGCGTAGGTCGT	GTTGGGGGAAGGGAAAAGGA

Appendix 1E

Sequences of primer pairs for the negative control regions used in ChIP-qPCR

Region name	Amplicon name	Forward primer sequences (5'→3')	Reverse primer sequences (5'→3')
NC i	HSSIL/M10Aq	TCTCTTTGAACACAGGGCAATG	TATTAGTCTAGGTGTACTGGCAGTTG
NC ii	HSSIL/M51Bq	TGAATGCTTCCCTTGTGATG	GTAATGTTTCCTTACTGGTTAGCAAC
NC iii	HSSCL/M15Bq	GTGCCCTTGAGAGCCTAGGG	CCTCAACAGCCTGTCTTATAATTG
NC iv	HSTAL.138q	CATCACCTGCAAAATGGAGG	TAAGCTGAGGCAGGCATTGTC
NC v	HSTAL.108q	GGATTGAGGAGAGGGCATGTG	GCACGGCTGTGGAGCTATG
NC vi	HSTAL.106q	CAGCAGAGGTCCCAAAGCC	CAGTACTCCCAGCTTGCTTCC
NC vii	HSTAL.77q	TTCTGTACCTGCCAGCCAAG	CCCGACGAGCGTTATGTAAG
NC viii	HSSIL/M55Aq	TCATGATGATATTTAGCATACTCAGCAAAG	GGAGAATGATAACTTGTGTCAGGC
NC ix	HSSCL/M137Aq	TCTCTGGAAGTCATAAATACAACA	AATCTGCTCATCAAGTAATACG
NC x	HSSCL/M182Aq	TTTGCAGTGCCCTGTTCTTAG	TGTTGGCTACCTTGATCATGTG
NC xi	HSTAL.7q	TCATGCCATTTCCGTTGTAC	TTGAACACTTGGAGATGATGATG
NC i	HSSIL/M10Aq	TCTCTTTGAACACAGGGCAATG	TATTAGTCTAGGTGTACTGGCAGTTG
NC ii	HSSIL/M51Bq	TGAATGCTTCCCTTGTGATG	GTAATGTTTCCTTACTGGTTAGCAAC
NC iii	HSSCL/M15Bq	GTGCCCTTGAGAGCCTAGGG	CCTCAACAGCCTGTCTTATAATTG
NC iv	HSTAL.138q	CATCACCTGCAAAATGGAGG	TAAGCTGAGGCAGGCATTGTC
NC v	HSTAL.108q	GGATTGAGGAGAGGGCATGTG	GCACGGCTGTGGAGCTATG
NC vi	HSTAL.106q	CAGCAGAGGTCCCAAAGCC	CAGTACTCCCAGCTTGCTTCC
NC vii	HSTAL.77q	TTCTGTACCTGCCAGCCAAG	CCCGACGAGCGTTATGTAAG
NC viii	HSSIL/M55Aq	TCATGATGATATTTAGCATACTCAGCAAAG	GGAGAATGATAACTTGTGTCAGGC
NC ix	HSSCL/M137Aq	TCTCTGGAAGTCATAAATACAACA	AATCTGCTCATCAAGTAATACG
NC x	HSSCL/M182Aq	TTTGCAGTGCCCTGTTCTTAG	TGTTGGCTACCTTGATCATGTG

Appendix 2

Gene Ontology classification of differentially expressed genes in Affymetric GeneChip analysis of siRNA knockdown study

Down-regulated genes in GATA1 knockdown

GO ID	GO term	P-value	% in gene list	% in human genome
Biological process				
GO:0006783	heme biosynthetic process	0.001	1.526718	16
GO:0042168	heme metabolic process	0.003211	1.526718	21
GO:0006366	transcription from RNA polymerase II promoter	0.00412	5.725191	637
GO:0046688	response to copper ion	0.005321	1.145038	8
GO:0030005	cellular di-, tri-valent inorganic cation homeostasis	0.007385	3.053435	182
GO:0006779	porphyrin biosynthetic process	0.007833	1.526718	26
GO:0055066	di-, tri-valent inorganic cation homeostasis	0.008977	3.053435	187
GO:0016568	chromatin modification	0.009466	3.435115	246
GO:0006357	regulation of transcription from RNA polymerase II promoter	0.00988	4.580153	450
GO:0016070	RNA metabolic process	0.009884	18.32061	4323
Molecular function				
GO:0005515	protein binding	9.86E-05	34.35115	9352
GO:0030528	transcription regulator activity	0.000571	12.21374	2181
GO:0008134	transcription factor binding	0.002783	4.580153	436
GO:0004998	transferrin receptor activity	0.006204	0.763359	2
GO:0003723	RNA binding	0.008275	6.870229	1008
Cellular component				
GO:0044424	intracellular part	7.08E-05	45.41985	13765
GO:0043231	intracellular membrane-bounded organelle	7.27E-05	34.35115	9375
GO:0043227	membrane-bounded organelle	7.34E-05	34.35115	9377
GO:0043229	intracellular organelle	0.000935	38.16794	11422
GO:0043226	organelle	0.000951	38.16794	11426
GO:0031974	membrane-enclosed lumen	0.001469	7.251908	1002
GO:0043233	organelle lumen	0.001469	7.251908	1002
GO:0031410	cytoplasmic vesicle	0.002717	4.580153	454
GO:0031982	vesicle	0.003221	4.580153	462
GO:0005737	cytoplasm	0.00622	28.24427	8035

GO:0016023	cytoplasmic membrane-bounded vesicle	0.009675	3.816794	365

Up-regulated genes in GATA1 knockdown

GO ID	GO term	P-value	% in gene list	% in human genome
Biological process				
GO:0048519	negative regulation of biological process	3.36E-13	9.04685	3.532632
GO:0048523	negative regulation of cellular process	7.32E-12	8.400646	3.327346
GO:0032502	developmental process	1.20E-10	15.34733	9.56006
GO:0008219	cell death	8.00E-10	6.946688	2.685827
GO:0016265	death	8.00E-10	6.946688	2.685827
GO:0050793	regulation of developmental process	2.50E-09	6.300485	2.329427
GO:0050793	regulation of developmental process	2.90E-09	6.300485	2.340832
GO:0048522	positive regulation of cellular process	4.93E-09	7.26979	3.076441
GO:0006915	apoptosis	6.31E-09	6.462036	2.514755
GO:0012501	programmed cell death	8.31E-09	6.462036	2.537565
GO:0042981	regulation of apoptosis	8.87E-09	5.169628	1.679354
GO:0043067	regulation of programmed cell death	1.20E-08	5.169628	1.699313
GO:0043067	regulation of programmed cell death	1.20E-08	5.169628	1.699313
GO:0048518	positive regulation of biological process	1.29E-08	7.592892	3.412882
GO:0007243	protein kinase cascade	2.31E-06	3.71567	1.134776
GO:0048856	anatomical structure development	3.30E-06	9.208401	5.465743
GO:0007154	cell communication	2.65E-05	19.063	16.10926
GO:0043066	negative regulation of apoptosis	5.13E-05	2.584814	0.658626
GO:0043069	negative regulation of programmed cell death	6.13E-05	2.584814	0.66718
GO:0007275	multicellular organismal development	6.53E-05	10.01616	6.737376
GO:0051093	negative regulation of developmental process	8.12E-05	3.069467	0.960853
GO:0007165	signal transduction	0.000114	17.60905	14.89465
GO:0006917	induction of apoptosis	0.00012	2.423263	0.613007
GO:0048731	system development	0.000127	7.754443	4.715878
GO:0012502	induction of programmed cell death	0.000128	2.423263	0.615858
GO:0050789	regulation of biological process	0.000194	28.43296	27.95028
GO:0006916	anti-apoptosis	0.000201	2.100162	0.470447
GO:0042127	regulation of cell proliferation	0.000486	3.392569	1.297294
GO:0051094	positive regulation of developmental process	0.000489	3.069467	1.080603
GO:0050794	regulation of cellular process	0.00049	27.6252	27.30876
GO:0042127	regulation of cell proliferation	0.000495	3.55412	1.411342
GO:0065007	biological regulation	0.000523	29.07916	29.15633

GO:0048513	organ development	0.000732	5.977383	3.392923
GO:0010627	regulation of protein kinase cascade	0.000859	1.938611	0.450489
GO:0010627	regulation of protein kinase cascade	0.000859	1.938611	0.450489
GO:0009966	regulation of signal transduction	0.001032	3.71567	1.593819
GO:0008283	cell proliferation	0.001048	4.523425	2.218231
GO:0007167	enzyme linked receptor protein signaling pathway	0.001211	2.746365	0.935192
GO:0009966	regulation of signal transduction	0.001429	3.55412	1.505432
GO:0045321	leukocyte activation	0.001507	2.100162	0.561686
GO:0032501	multicellular organismal process	0.001972	12.92407	10.52947
GO:0043065	positive regulation of apoptosis	0.002138	2.423263	0.769823
GO:0043068	positive regulation of programmed cell death	0.002447	2.423263	0.778377
GO:0006928	cell motion	0.002569	2.907916	1.09771
GO:0051674	localization of cell	0.002569	2.907916	1.09771
GO:0007242	intracellular signaling cascade	0.002836	8.239095	5.742309
GO:0008285	negative regulation of cell proliferation	0.003013	2.261712	0.692841
Molecular pathway				
GO:0005515	protein binding	7.13E-09	29.88691	13.54603
GO:0019899	enzyme binding	2.99E-07	3.392569	26.66439
GO:0019210	kinase inhibitor activity	1.91E-05	1.292407	0.895276
GO:0019207	kinase regulator activity	2.80E-05	1.938611	0.122601
GO:0005159	insulin-like growth factor receptor binding	0.000112	0.807754	0.362102
GO:0004860	protein kinase inhibitor activity	0.000237	1.130856	0.037066
GO:0019900	kinase binding	0.000835	1.453958	0.116899
GO:0019887	protein kinase regulator activity	0.003639	1.453958	0.26231
GO:0030234	enzyme regulator activity	0.006298	5.977383	0.313632
Cellular component				
GO:0031226	intrinsic to plasma membrane	5.36E-05	6.462036	3.666638
GO:0005737	cytoplasm	5.51E-05	24.23263	22.90936
GO:0005887	integral to plasma membrane	0.000104	6.300485	3.621019
GO:0005886	plasma membrane	0.000178	13.89338	11.35346
GO:0044459	plasma membrane part	0.000493	9.208401	6.671799

Down-regulated genes in SCL knockdown

GO ID	GO term	p-value	% in gene list	% in human genome
Biological process				
GO:0048519	negative regulation of biological process	2.02E-10	8.352941	3.532632

GO:0048523	negative regulation of cellular process	3.13E-10	8	3.327346
GO:0033036	macromolecule localization	5.39E-08	7.294118	3.236108
GO:0016043	cellular component organization and biogenesis	1.37E-07	11.76471	6.706013
GO:0008104	protein localization	1.50E-07	6.941176	3.073589
GO:0050793	regulation of developmental process	2.86E-07	5.764706	2.329427
GO:0050793	regulation of developmental process	3.37E-07	5.764706	2.340832
GO:0015031	protein transport	1.51E-06	6.352941	2.851196
GO:0045184	establishment of protein localization	1.56E-06	6.352941	2.854047
GO:0016071	mRNA metabolic process	2.40E-06	3.176471	0.900978
GO:0032502	developmental process	4.05E-06	14.58824	9.56006
GO:0031324	negative regulation of cellular metabolic process	6.61E-06	3.882353	1.345765
GO:0048522	positive regulation of cellular process	7.93E-06	6.470588	3.076441
GO:0009892	negative regulation of metabolic process	9.00E-06	3.882353	1.362872
GO:0008380	RNA splicing	1.21E-05	2.588235	0.664329
GO:0007049	cell cycle	1.33E-05	5.764706	2.620249
GO:0010629	negative regulation of gene expression	1.63E-05	3.058824	0.920936
GO:0048518	positive regulation of biological process	1.84E-05	6.823529	3.412882
GO:0007243	protein kinase cascade	2.24E-05	3.411765	1.134776
GO:0006915	apoptosis	2.70E-05	5.529412	2.514755
GO:0016070	RNA metabolic process	2.77E-05	17.29412	12.32572
GO:0010605	negative regulation of macromolecule metabolic process	2.90E-05	3.529412	1.217461
GO:0012501	programmed cell death	3.55E-05	5.529412	2.537565
GO:0008219	cell death	7.45E-05	5.647059	2.685827
GO:0016265	death	7.45E-05	5.647059	2.685827
GO:0006469	negative regulation of protein kinase activity	0.000102	1.294118	0.176774
GO:0033673	negative regulation of kinase activity	0.000102	1.294118	0.176774
GO:0065009	regulation of molecular function	0.000114	4.705882	2.061415
GO:0042981	regulation of apoptosis	0.000133	4.117647	1.679354
GO:0051348	negative regulation of transferase activity	0.000144	1.294118	0.182477
GO:0045934	negative regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	0.000167	2.941176	0.966555
GO:0043067	regulation of programmed cell death	0.000177	4.117647	1.699313
GO:0043067	regulation of programmed cell death	0.000177	4.117647	1.699313
GO:0051128	regulation of cellular component organization and biogenesis	0.000204	2.588235	0.778377
GO:0016044	membrane organization and biogenesis	0.000248	2.705882	0.852508
GO:0006139	nucleobase, nucleoside, nucleotide and nucleic acid metabolic	0.000288	21.05882	16.37727

	process			
GO:0016481	negative regulation of transcription	0.000374	2.705882	0.872466
GO:0019222	regulation of metabolic process	0.000414	16.70588	12.33427
GO:0060255	regulation of macromolecule metabolic process	0.000499	16.35294	12.0463
GO:0060255	regulation of macromolecule metabolic process	0.000499	16.35294	12.0463
GO:0008285	negative regulation of cell proliferation	0.000522	2.352941	0.692841
GO:0051641	cellular localization	0.000574	6.470588	3.506971
GO:0009890	negative regulation of biosynthetic process	0.000603	2.941176	1.034984
GO:0016192	vesicle-mediated transport	0.000625	4.117647	1.793402
GO:0065008	regulation of biological quality	0.000634	6.352941	3.427138
GO:0043170	macromolecule metabolic process	0.000683	33.88235	29.49847
GO:0006397	mRNA processing	0.00072	2.352941	0.707097
GO:0008283	cell proliferation	0.000747	4.705882	2.218231
GO:0006886	intracellular protein transport	0.000868	3.882353	1.659396
GO:0010558	negative regulation of macromolecule biosynthetic process	0.001039	2.823529	0.995067
GO:0010558	negative regulation of macromolecule biosynthetic process	0.001039	2.823529	0.995067
GO:0042127	regulation of cell proliferation	0.001227	3.294118	1.297294
GO:0006996	organelle organization and biogenesis	0.001376	7.882353	4.74439
GO:0050790	regulation of catalytic activity	0.001629	4.117647	1.870385
GO:0043283	biopolymer metabolic process	0.001638	29.05882	24.80255
GO:0051246	regulation of protein metabolic process	0.001858	3.058824	1.174693
GO:0043086	negative regulation of catalytic activity	0.001969	1.529412	0.33359
GO:0051649	establishment of localization in cell	0.002042	6.117647	3.378667
GO:0042127	regulation of cell proliferation	0.002116	3.411765	1.411342
GO:0007249	I-kappaB kinase/NF-kappaB cascade	0.002187	1.647059	0.390614
GO:0006366	transcription from RNA polymerase II promoter	0.002275	4	1.816212
GO:0031323	regulation of cellular metabolic process	0.002485	15.88235	11.96077
GO:0016568	chromatin modification	0.002594	2.235294	0.701394
GO:0048869	cellular developmental process	0.003053	6.352941	3.612465
GO:0006897	endocytosis	0.003148	2	0.581644
GO:0010324	membrane invagination	0.003148	2	0.581644
GO:0043066	negative regulation of apoptosis	0.0042	2.117647	0.658626
GO:0010467	gene expression	0.004986	19.64706	15.7329
GO:0043069	negative regulation of programmed cell death	0.005032	2.117647	0.66718
GO:0046907	intracellular transport	0.00511	5.294118	2.845494
GO:0007242	intracellular signaling cascade	0.005441	8.823529	5.742309

GO:0022402	cell cycle process	0.005698	3.294118	1.40564
GO:0051093	negative regulation of developmental process	0.006743	2.588235	0.960853
GO:0006325	establishment and/or maintenance of chromatin architecture	0.007632	3.176471	1.348616
GO:0051726	regulation of cell cycle	0.007893	2.352941	0.826847
GO:0009966	regulation of signal transduction	0.008083	3.529412	1.593819
GO:0065007	biological regulation	0.009721	32.70588	29.15633
GO:0044238	primary metabolic process	0.009831	37.64706	34.37402
GO:0051276	chromosome organization and biogenesis	0.009934	3.529412	1.610926
Molecular function				
GO:0005515	protein binding	1.49E-20	38.70588	26.66439
GO:0003723	RNA binding	1.50E-06	6.235294	2.874006
GO:0019899	enzyme binding	4.32E-05	2.823529	0.895276
GO:0042802	identical protein binding	0.000306	3.176471	1.21461
GO:0019900	kinase binding	0.001686	1.294118	0.26231
GO:0016563	transcription activator activity	0.007084	2.352941	0.895276
GO:0004842	ubiquitin-protein ligase activity	0.009739	1.647059	0.496108
Cellular component				
GO:0043231	intracellular membrane-bounded organelle	2.34E-29	41.64706	26.72996
GO:0043227	membrane-bounded organelle	2.45E-29	41.64706	26.73567
GO:0044424	intracellular part	9.88E-26	52.82353	39.24671
GO:0043229	intracellular organelle	8.29E-24	45.76471	32.56636
GO:0043226	organelle	8.94E-24	45.76471	32.57777
GO:0005634	nucleus	9.26E-20	29.41176	18.29898
GO:0005737	cytoplasm	3.59E-19	34.23529	22.90936
GO:0005622	intracellular	3.80E-17	54.58824	44.60696
GO:0044446	intracellular organelle part	5.81E-15	21.29412	12.65931
GO:0044422	organelle part	7.28E-15	21.29412	12.68782
GO:0044428	nuclear part	8.03E-10	7.764706	3.350155
GO:0012505	endomembrane system	1.29E-07	7.411765	3.521227
GO:0031410	cytoplasmic vesicle	5.56E-07	3.882353	1.294443
GO:0031982	vesicle	8.57E-07	3.882353	1.317253
GO:0031974	membrane-enclosed lumen	9.56E-07	6.235294	2.856898
GO:0043233	organelle lumen	9.56E-07	6.235294	2.856898
GO:0044444	cytoplasmic part	1.36E-06	19.29412	13.87392
GO:0031090	organelle membrane	4.54E-06	8.705882	4.869843
GO:0016023	cytoplasmic membrane-bounded vesicle	1.11E-05	3.176471	1.040687

GO:0031988	membrane-bounded vesicle	1.46E-05	3.176471	1.054943
GO:0031981	nuclear lumen	1.60E-05	4.705882	2.0272
GO:0005794	Golgi apparatus	0.000159	4.705882	2.215379
GO:0005654	nucleoplasm	0.000342	3.647059	1.545348
GO:0030530	heterogeneous nuclear ribonucleoprotein complex	0.000384	0.705882	0.051322
GO:0005694	chromosome	0.001655	3.529412	1.588116
GO:0044451	nucleoplasm part	0.001808	3.176471	1.354318
GO:0000776	kinetochore	0.002892	0.941176	0.14256
GO:0042470	melanosome	0.003264	1.176471	0.236649
GO:0048770	pigment granule	0.003264	1.176471	0.236649
GO:0005783	endoplasmic reticulum	0.003991	4.941176	2.71719
GO:0044431	Golgi apparatus part	0.004245	3.058824	1.340062
GO:0044432	endoplasmic reticulum part	0.004425	3.647059	1.756337
GO:0005815	microtubule organizing center	0.005121	1.647059	0.479001
GO:0005813	centrosome	0.005351	1.529412	0.419126
GO:0005829	cytosol	0.006926	3.411765	1.628033
GO:0005635	nuclear envelope	0.007784	1.882353	0.630114
GO:0031252	leading edge	0.008146	1.058824	0.210989

Up-regulated genes in SCL knockdown

GO ID	**GO term**	**P-value**	**% in gene list**	**% in genome**
Biological process				
GO:0032501	multicellular organismal process	1.22E-10	15.65558	10.52947
GO:0048856	anatomical structure development	1.98E-09	9.328115	5.465743
GO:0007275	multicellular organismal development	3.80E-09	10.82844	6.737376
GO:0032502	developmental process	1.95E-08	13.95956	9.56006
GO:0048731	system development	1.00E-07	8.023483	4.715878
GO:0048519	negative regulation of biological process	3.84E-07	6.392694	3.532632
GO:0048518	positive regulation of biological process	6.09E-07	6.196999	3.412882
GO:0048522	positive regulation of cellular process	1.58E-06	5.675147	3.076441
GO:0048523	negative regulation of cellular process	1.62E-06	6.001305	3.327346
GO:0048513	organ development	2.03E-06	6.066536	3.392923
GO:0048869	cellular developmental process	4.23E-05	6.066536	3.612465
GO:0006928	cell motion	9.01E-05	2.609263	1.09771
GO:0051674	localization of cell	9.01E-05	2.609263	1.09771
GO:0030154	cell differentiation	0.000163	5.479452	3.247512

GO:0009653	anatomical structure morphogenesis	0.000219	4.631442	2.603142
GO:0040011	locomotion	0.000319	2.08741	0.812591
GO:0048870	cell motility	0.00072	2.022179	0.801186
GO:0016477	cell migration	0.000739	1.956947	0.761269
GO:0050793	regulation of developmental process	0.001436	4.109589	2.329427
GO:0050793	regulation of developmental process	0.001689	4.109589	2.340832
GO:0048741	skeletal muscle fiber development	0.00203	0.652316	0.105494
GO:0048747	muscle fiber development	0.002655	0.652316	0.108345
GO:0014706	striated muscle development	0.002777	1.043705	0.279417
GO:0048468	cell development	0.00429	2.4788	1.183246
GO:0045445	myoblast differentiation	0.005172	0.456621	0.051322
GO:0051094	positive regulation of developmental process	0.008574	2.283105	1.080603
Cellular component				
GO:0031226	intrinsic to plasma membrane	1.74E-05	6.066536	3.666638
GO:0005887	integral to plasma membrane	3.66E-05	5.936073	3.621019
GO:0044459	plasma membrane part	0.000294	9.262883	6.671799

Down-regulated genes in E2A knockdown

GO ID	GO term	P-value	% in gene list	% in human genome
Biological process				
GO:0015031	protein transport	1.47E-06	8.78187	2.851196
GO:0045184	establishment of protein localization	1.51E-06	8.78187	2.854047
GO:0033036	macromolecule localization	2.11E-06	9.348442	3.236108
GO:0008104	protein localization	2.27E-06	9.065156	3.073589
GO:0008283	cell proliferation	6.94E-06	7.365439	2.218231
GO:0016070	RNA metabolic process	2.79E-05	20.67989	12.32572
GO:0007049	cell cycle	5.03E-05	7.648725	2.620249
GO:0006366	transcription from RNA polymerase II promoter	5.58E-05	6.232295	1.816212
GO:0044249	cellular biosynthetic process	6.35E-05	24.07932	15.58749
GO:0048522	positive regulation of cellular process	0.000106	8.215297	3.076441
GO:0016043	cellular component organization and biogenesis	0.000117	13.31445	6.706013
GO:0051726	regulation of cell cycle	0.000276	3.966006	0.826847
GO:0048518	positive regulation of biological process	0.000282	8.498584	3.412882
GO:0060255	regulation of macromolecule metabolic process	0.000551	19.26346	12.0463
GO:0060255	regulation of macromolecule metabolic process	0.000551	19.26346	12.0463
GO:0048519	negative regulation of biological process	0.000575	8.498584	3.532632

GO:0009058	biosynthetic process	0.000735	26.34561	18.62116
GO:0010604	positive regulation of macromolecule metabolic process	0.00123	4.532578	1.220312
GO:0019222	regulation of metabolic process	0.001276	19.26346	12.33427
GO:0010467	gene expression	0.001516	22.94618	15.7329
GO:0048523	negative regulation of cellular process	0.001618	7.932011	3.327346
GO:0031323	regulation of cellular metabolic process	0.001832	18.69688	11.96077
GO:0006357	regulation of transcription from RNA polymerase II promoter	0.002334	4.532578	1.283038
GO:0045893	positive regulation of transcription, DNA-dependent	0.002432	3.399433	0.727055
GO:0051254	positive regulation of RNA metabolic process	0.002635	3.399433	0.732757
GO:0044237	cellular metabolic process	0.002762	41.92635	35.0355
GO:0010468	regulation of gene expression	0.0029	17.84703	11.37342
GO:0031325	positive regulation of cellular metabolic process	0.002999	4.532578	1.308699
GO:0016192	vesicle-mediated transport	0.003074	5.382436	1.793402
GO:0009893	positive regulation of metabolic process	0.003829	4.532578	1.33436
GO:0006139	nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	0.003832	23.22946	16.37727
GO:0045941	positive regulation of transcription	0.005429	3.68272	0.923788
GO:0010468	regulation of gene expression	0.005664	17.56374	11.36202
GO:0010628	positive regulation of gene expression	0.006818	3.68272	0.943746
GO:0006351	transcription, DNA-dependent	0.007216	16.14731	10.20443
GO:0032774	RNA biosynthetic process	0.00758	16.14731	10.22154
GO:0042127	regulation of cell proliferation	0.007689	4.532578	1.411342
GO:0006355	regulation of transcription, DNA-dependent	0.007733	15.86402	9.982037
GO:0051252	regulation of RNA metabolic process	0.008538	15.86402	10.01625
GO:0010556	regulation of macromolecule biosynthetic process	0.009549	17.28045	11.30214
GO:0010556	regulation of macromolecule biosynthetic process	0.009549	17.28045	11.30214
Molecular function				
GO:0005515	protein binding	1.76E-15	43.62606	26.66439
GO:0003723	RNA binding	2.11E-06	8.498584	2.874006
GO:0008134	transcription factor binding	0.000101	4.815864	1.243121
GO:0003700	transcription factor activity	0.000224	9.631728	4.365181
GO:0030528	transcription regulator activity	0.001101	11.61473	6.218459
GO:0016563	transcription activator activity	0.001199	3.68272	0.895276
GO:0017111	nucleoside-triphosphatase activity	0.001385	7.082153	2.928178
GO:0016462	pyrophosphatase activity	0.002483	7.082153	3.030821
GO:0016818	hydrolase activity, acting on acid anhydrides, in phosphorus-	0.00295	7.082153	3.062185

	containing anhydrides			
GO:0016817	hydrolase activity, acting on acid anhydrides	0.003139	7.082153	3.073589
GO:0003712	transcription cofactor activity	0.003381	3.399433	0.843954
GO:0019899	enzyme binding	0.006038	3.399433	0.895276
GO:0003713	transcription coactivator activity	0.008901	2.549575	0.521769
Cellular component				
GO:0043231	intracellular membrane-bounded organelle	3.18E-24	48.72521	26.72996
GO:0043227	membrane-bounded organelle	3.27E-24	48.72521	26.73567
GO:0043229	intracellular organelle	3.57E-21	52.9745	32.56636
GO:0043226	organelle	3.73E-21	52.9745	32.57777
GO:0044424	intracellular part	1.23E-19	58.35694	39.24671
GO:0005622	intracellular	1.27E-15	60.62323	44.60696
GO:0044446	intracellular organelle part	1.30E-13	26.34561	12.65931
GO:0044422	organelle part	1.50E-13	26.34561	12.68782
GO:0005634	nucleus	3.85E-13	32.86119	18.29898
GO:0005737	cytoplasm	1.11E-12	37.67705	22.90936
GO:0031090	organelle membrane	3.61E-09	13.03116	4.869843
GO:0012505	endomembrane system	5.59E-09	10.76487	3.521227
GO:0044428	nuclear part	8.35E-08	9.915014	3.350155
GO:0016023	cytoplasmic membrane-bounded vesicle	1.55E-07	5.382436	1.040687
GO:0031410	cytoplasmic vesicle	1.69E-07	5.949008	1.294443
GO:0031988	membrane-bounded vesicle	1.94E-07	5.382436	1.054943
GO:0031982	vesicle	2.30E-07	5.949008	1.317253
GO:0044444	cytoplasmic part	3.28E-07	23.51275	13.87392
GO:0031974	membrane-enclosed lumen	1.39E-06	8.498584	2.856898
GO:0043233	organelle lumen	1.39E-06	8.498584	2.856898
GO:0030530	heterogeneous nuclear ribonucleoprotein complex	6.73E-05	1.416431	0.051322
GO:0031981	nuclear lumen	8.36E-05	6.232295	2.0272
GO:0031965	nuclear membrane	0.000109	3.116147	0.510364
GO:0042470	melanosome	0.000127	2.266289	0.236649
GO:0048770	pigment granule	0.000127	2.266289	0.236649
GO:0005792	microsome	0.000444	2.832861	0.47615
GO:0044432	endoplasmic reticulum part	0.000529	5.382436	1.756337
GO:0042598	vesicular fraction	0.000641	2.832861	0.496108
GO:0005635	nuclear envelope	0.000847	3.116147	0.630114
GO:0044431	Golgi apparatus part	0.000938	4.532578	1.340062

GO:0005624	membrane fraction	0.000971	5.09915	1.665099
GO:0005626	insoluble fraction	0.001338	5.09915	1.705015
GO:0005654	nucleoplasm	0.001405	4.815864	1.545348
GO:0005783	endoplasmic reticulum	0.002811	6.515581	2.71719
GO:0000267	cell fraction	0.003407	5.665722	2.186867
GO:0005794	Golgi apparatus	0.00409	5.665722	2.215379
GO:0044451	nucleoplasm part	0.004387	4.249292	1.354318
GO:0005789	endoplasmic reticulum membrane	0.006661	4.532578	1.57386
GO:0005813	centrosome	0.00877	2.266289	0.419126
GO:0042175	nuclear envelope-endoplasmic reticulum network	0.008948	4.532578	1.613777

Up-regulated genes in E2A knockdown

GO ID	GO term	P-value	% in gene list	% in human genome
Molecular function				
GO:0005515	protein binding	0.00035	33.41523	26.66439
Cellular component				
GO:0005634	nucleus	0.003608	23.83292	18.29898

Down-regulated genes in LDB1 knockdown

GO ID	GO term	P-value	% in gene list	% in human genome
Biological process				
GO:0006396	RNA processing	1.96E-09	4.86618	1.551051
GO:0007049	cell cycle	2.13E-09	6.569343	2.620249
GO:0022403	cell cycle phase	5.05E-09	4.014599	1.12052
GO:0043933	macromolecular complex subunit organization	7.71E-09	6.326034	2.546118
GO:0016071	mRNA metabolic process	1.15E-08	3.527981	0.900978
GO:0008380	RNA splicing	6.23E-08	2.919708	0.664329
GO:0051301	cell division	1.02E-07	3.041363	0.73846
GO:0022402	cell cycle process	1.23E-07	4.257908	1.40564
GO:0065003	macromolecular complex assembly	1.59E-07	5.717762	2.349386
GO:0006397	mRNA processing	2.27E-07	2.919708	0.707097
GO:0044237	cellular metabolic process	4.28E-07	39.41606	35.0355
GO:0000279	M phase	5.83E-07	3.284672	0.92949
GO:0000278	mitotic cell cycle	6.59E-07	3.406326	1.00077
GO:0043283	biopolymer metabolic process	7.91E-07	29.80535	24.80255
GO:0044238	primary metabolic process	1.88E-06	38.44282	34.37402

GO:0006139	nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	2.60E-06	21.28954	16.37727
GO:0007067	mitosis	2.76E-06	2.676399	0.675733
GO:0033036	macromolecule localization	4.60E-06	6.569343	3.236108
GO:0000087	M phase of mitotic cell cycle	5.92E-06	2.676399	0.704245
GO:0006974	response to DNA damage stimulus	1.07E-05	3.77129	1.357169
GO:0048519	negative regulation of biological process	1.41E-05	6.812652	3.532632
GO:0008104	protein localization	1.49E-05	6.20438	3.073589
GO:0048523	negative regulation of cellular process	2.96E-05	6.447689	3.327346
GO:0043170	macromolecule metabolic process	3.64E-05	33.21168	29.49847
GO:0034621	cellular macromolecular complex subunit organization	0.000136	4.501217	2.021498
GO:0006461	protein complex assembly	0.000168	3.284672	1.217461
GO:0015031	protein transport	0.000171	5.596107	2.851196
GO:0045184	establishment of protein localization	0.000176	5.596107	2.854047
GO:0016043	cellular component organization and biogenesis	0.000323	10.09732	6.706013
GO:0008152	metabolic process	0.000325	41.36253	39.29233
GO:0016070	RNA metabolic process	0.000616	15.93674	12.32572
GO:0006259	DNA metabolic process	0.000831	4.622871	2.269552
GO:0006281	DNA repair	0.001054	3.041363	1.177544
GO:0051726	regulation of cell cycle	0.001733	2.43309	0.826847
GO:0034622	cellular macromolecular complex assembly	0.002029	3.892944	1.813361
GO:0022613	ribonucleoprotein complex biogenesis and assembly	0.002078	2.311436	0.764121
GO:0006986	response to unfolded protein	0.002976	1.094891	0.173923
GO:0051789	response to protein stimulus	0.002976	1.094891	0.173923
GO:0006950	response to stress	0.004932	7.055961	4.459271
GO:0046500	S-adenosylmethionine metabolic process	0.007132	0.364964	0.008554
GO:0010467	gene expression	0.009963	18.61314	15.7329
Molecular function				
GO:0005515	protein binding	1.39E-21	37.7129	26.66439
GO:0003723	RNA binding	2.58E-08	6.569343	2.874006
GO:0000287	magnesium ion binding	7.02E-06	3.649635	1.314401
GO:0003729	mRNA binding	0.00044	0.973236	0.116899
GO:0003712	transcription cofactor activity	0.003273	2.311436	0.843954
GO:0003724	RNA helicase activity	0.005444	0.729927	0.079833
Cellular component				
GO:0043231	intracellular membrane-bounded organelle	5.39E-38	42.70073	26.72996

GO:0043227	membrane-bounded organelle	5.66E-38	42.70073	26.73567
GO:0044424	intracellular part	4.21E-35	53.64964	39.24671
GO:0043226	organelle	1.29E-26	45.01217	32.57777
GO:0043229	intracellular organelle	2.83E-26	44.89051	32.56636
GO:0005737	cytoplasm	7.01E-26	35.27981	22.90936
GO:0044428	nuclear part	2.59E-24	10.7056	3.350155
GO:0044422	organelle part	1.05E-23	23.35766	12.68782
GO:0044446	intracellular organelle part	2.14E-23	23.23601	12.65931
GO:0005622	intracellular	2.24E-23	54.62287	44.60696
GO:0005634	nucleus	1.75E-19	28.3455	18.29898
GO:0031974	membrane-enclosed lumen	1.53E-15	8.150852	2.856898
GO:0043233	organelle lumen	1.53E-15	8.150852	2.856898
GO:0031981	nuclear lumen	4.04E-12	6.082725	2.0272
GO:0005654	nucleoplasm	2.13E-11	5.109489	1.545348
GO:0044444	cytoplasmic part	7.15E-11	20.43796	13.87392
GO:0044451	nucleoplasm part	5.47E-10	4.501217	1.354318
GO:0042470	melanosome	2.12E-07	1.703163	0.236649
GO:0048770	pigment granule	2.12E-07	1.703163	0.236649
GO:0031090	organelle membrane	2.53E-06	8.515815	4.869843
GO:0005739	mitochondrion	4.78E-06	6.082725	3.016565
GO:0016604	nuclear body	2.20E-05	1.703163	0.336441
GO:0043234	protein complex	3.43E-05	12.0438	8.3426
GO:0012505	endomembrane system	0.000211	6.20438	3.521227
GO:0032991	macromolecular complex	0.000295	15.32847	11.90659
GO:0044464	cell part	0.000637	62.40876	65.22681
GO:0005623	cell	0.000642	62.40876	65.22966
GO:0031967	organelle envelope	0.00087	4.014599	1.961623
GO:0031975	envelope	0.000924	4.014599	1.967325
GO:0005681	spliceosome	0.001196	1.581509	0.40487
GO:0016607	nuclear speck	0.001393	1.216545	0.2395
GO:0031982	vesicle	0.001686	3.041363	1.317253
GO:0044429	mitochondrial part	0.002319	3.649635	1.781998
GO:0005819	spindle	0.002593	1.216545	0.256608
GO:0016023	cytoplasmic membrane-bounded vesicle	0.003552	2.554745	1.040687
GO:0031410	cytoplasmic vesicle	0.003716	2.919708	1.294443
GO:0031988	membrane-bounded vesicle	0.004344	2.554745	1.054943

GO:0005876	spindle microtubule	0.005795	0.608273	0.054173
GO:0000793	condensed chromosome	0.007003	1.094891	0.230947
GO:0005815	microtubule organizing center	0.007258	1.581509	0.479001
GO:0005813	centrosome	0.008522	1.459854	0.419126

Up-regulated genes in LDB1 knockdown

GO ID	GO term	P-value	% in gene list	% in human genome
Biological process				
GO:0032502	developmental process	5.35E-07	16.56051	9.56006
GO:0048856	anatomical structure development	8.34E-07	11.1465	5.465743
GO:0007275	multicellular organismal development	3.46E-06	12.57962	6.737376
GO:0048731	system development	1.82E-05	9.55414	4.715878
GO:0007243	protein kinase cascade	7.36E-05	3.821656	1.134776
GO:0032501	multicellular organismal process	8.16E-05	16.56051	10.52947
GO:0007242	intracellular signaling cascade	0.000153	10.50955	5.742309
Molecular function				
GO:0005515	protein binding	1.08E-05	34.55414	26.66439
GO:0004674	protein serine/threonine kinase activity	0.003902	4.617834	2.024349
GO:0005201	extracellular matrix structural constituent	0.007076	1.592357	0.325036
Cellular component				
GO:0005737	cytoplasm	0.000852	28.98089	22.90936

Down-regulated genes in LMO2 knockdown

GO ID	GO term	P-value	% in gene list	% in human genome
Molecular function				
GO:0043566	structure-specific DNA binding	0.004711	8.163265	0.47615
Cellular component				
GO:0044428	nuclear part	0.002122	18.36735	3.350155
GO:0031974	membrane-enclosed lumen	0.004614	16.32653	2.856898
GO:0043233	organelle lumen	0.004614	16.32653	2.856898
GO:0031253	cell projection membrane	0.00838	4.081633	0.034214

Up-regulated genes in LMO2 knockdown

GO ID	GO term	P-value	% in gene list	% in human genome
Biological process				
GO:0048856	anatomical structure development	3.55E-05	9.529148	5.465743

GO:0032502	developmental process	4.35E-05	14.46188	9.56006
GO:0007610	behavior	0.000154	2.802691	0.886722
GO:0048519	negative regulation of biological process	0.000231	6.726457	3.532632
GO:0048523	negative regulation of cellular process	0.000353	6.390135	3.327346
GO:0048731	system development	0.00041	8.183857	4.715878
GO:0007275	multicellular organismal development	0.001617	10.42601	6.737376
GO:0048869	cellular developmental process	0.00216	6.502242	3.612465
GO:0030154	cell differentiation	0.007777	5.829596	3.247512

Appendix 3A

Characterisation of antibodies for the SCL erythroid complex for western blot

TF	Western blot (Fig. 3.3)	Name of antibody	Origin	Source	Poly- or mono-clonal	Target epitope	Catalog no.	Specificity and signal detection	Titrated dilution
SCL	A	Anti-TAL1	Rabbit	Abcam	Polyclonal	Amino acid 7-21 of human SCL protein	ab12115	No predicted protein bands identified	1:200
SCL	B	TAL1	Rabbit	Active Motif	Polyclonal	Amino acid 7-21 of human SCL protein	39066	Predicted protein bands identified	1:1000
SCL	C	Anti-TAL1 3BTL73	Mouse	Collaborator	Monoclonal	Full length human SCL protein	---	No predicted protein bands identified	1:100
E2A	D	Anti-TCF3	Rabbit	Abcam	Polyclonal	Amino acid 517-531 of human E2A protein	ab54462	No predicted protein bands identified	1:200
E2A	E	TCF3 M20	Goat	Santa Cruz	Polyclonal	C-terminus of mouse origin	sc-8635	No predicted protein bands identified	1:200
E2A	F	Anti-E2A	Mouse	BD Biosciences	Monoclonal	Amino acid 463-483 of human E2A protein	554102	No predicted protein bands identified	1:250
E47	G	E47 N649	Rabbit	Santa Cruz	Polyclonal	Full-length human E47 protein	sc-763	Predicted protein bands identified	1:5000
E12	H	E12 H208	Rabbit	Santa Cruz	Polyclonal	bHLH domain of human E12 protein	sc-762	Predicted protein bands identified	1:2000
E47	I	E47	Rabbit	Active Motif	Polyclonal	Amino acid 517-531 of human E47 protein	39314	No predicted protein bands identified	1:100
E47	J	E47 (Ab-1)	Rabbit	Merck	Polyclonal	Amino acid 517-531 of human E47 protein	PC695	No predicted protein bands identified	1:500
GATA1	K	GATA1 M20	Goat	Santa Cruz	Polyclonal	C-terminus of mouse origin	sc-1234	Predicted protein bands identified	1:1000
LDB1	L	CLIM-2 N18	Goat	Santa Cruz	Polyclonal	N-terminus of human origin	sc-11198	Predicted protein bands identified	1:1000
LMO2	M	LMO2 N16	Goat	Santa Cruz	Polyclonal	N-terminus of human origin	sc-10497	No predicted protein bands identified	1:100
LMO2	N	LMO2 G16	Goat	Santa Cruz	Polyclonal	C-terminus of human origin	sc-10499	No predicted protein bands identified	1:100
LMO2	O	LMO2	Sheep	Abcam	Polyclonal	Full-length human LMO2 protein	ab16132	No predicted protein bands identified	1:100

Secondary antibodies used in western blotting

Name of antibody	Origin	Source	Poly- or mono-clonal	Catalog no.	Titrated dilution
Anti-goat IgG	Donkey	Santa Cruz	Polyclonal	sc-2020	1:20000
Anti-rabbit IgG	Goat	Santa Cruz	Polyclonal	sc-2004	1:20000
Anti-mouse IgG	Goat	Santa Cruz	Polyclonal	sc-2005	1:20000
Anti-sheep IgG	Rabbit	Abcam	Polyclonal	ab6747	1:20000

Appendix 3B

Characterisation of antibodies for the SCL erythroid complex for ChIP-on-chip

TF	Name of antibody	Origin	Source	Poly- or mono-clonal	Target epitope	Catalogue no.	ChIP-on-chip results
SCL	TAL1	Rabbit	Active Motif	Polyclonal	Amino acid 7-21 of human SCL protein	39066	Substantial enrichments at +51 enhancer. Selected for subsequent analyses.
SCL	Anti-TAL1 3BTL73	Mouse	D. Mathieu	Monoclonal	Full length human SCL protein	---	Substantial enrichments at +51 enhancer.
E2A	Anti-TCF3	Rabbit	Abcam	Polyclonal	Amino acid 517-531 of human E2A protein	ab54462	No substantial enrichments at any enhancers or promoters.
E2A	Anti-E2A	Mouse	BD Biosciences	Monoclonal	Amino acid 463-483 of human E2A protein	554102	Substantial enrichments at +51 enhancer.
E47	E47 N649	Rabbit	Santa Cruz	Polyclonal	Full-length human E47 protein	sc-763	Substantial enrichments at +51, +3, -9/-10 enhancers and promoter 1a. Selected for subsequent analyses.
E12	E12 H208	Rabbit	Santa Cruz	Polyclonal	bHLH domain of human E12 protein	sc-762	Substantial enrichments at +51, +3, -9/-10 enhancers and promoter 1a. Selected for subsequent analyses.
GATA1	GATA1 M20	Goat	Santa Cruz	Polyclonal	C-terminus of mouse origin	sc-1234	Substantial enrichments at +51, +3, -9/-10 enhancers and promoter 1a. Selected for subsequent analyses.
LDB1	CLIM-2 N18	Goat	Santa Cruz	Polyclonal	N-terminus of human origin	sc-11198	Substantial enrichments at +51, +3, -9/-10 enhancers and promoter 1a. Selected for subsequent analyses.
LMO2	LMO2 N16	Goat	Santa Cruz	Polyclonal	N-terminus of human origin	sc-10497	Enrichments at +51 enhancer. Selected for subsequent analyses.
LMO2	LMO2 G16	Goat	Santa Cruz	Polyclonal	C-terminus of human origin	sc-10499	Enrichments at +51 enhancer.

Appendix 4

Multiple sequence alignments of transcription factor binding sites at promoter regions of selected putative target genes

E-box (E12, E47 and TAL1) and GATA motifs were identified by TESS and TFSEARCH in a 4 kb window spanning the transcription start sites of the 14 promoter regions of selected putative target genes. Multiple species sequence alignments showed one or more relatively conserved E-box or GATA motifs identified (named from 1 to 4 in brackets). Shaded in yellow are the conserved nucleotides for motifs found across a number of species. The motifs labelled with an asterisk (*) were shown to have significant enrichments in the ChIP-qPCR studies (section 5.4.4 3). Panels A to N show the alignments for the 14 selected target genes. For SCL, shown in panel K, the sequence of the composite E-box/GATA motif in the +51 enhancer is shown.

A) BRD2

```
Human       aagtccacatctgcggtccgcaccttatctctcc
Mouse       acgt--------------caaatc----------
Rat         acgt--------------caaacc----------
Rabbit      acgcccacgtctacagcgcgcacctcctcgcccc
Dog         acgccccatctaccgttctcagctcatttcccg
Armadillo   atttccacatctaccgttcgcacagcccaaatcc
Elephant    aagctcacttcttccgtccgtttctcatctcccc
                      E12 E47          GATA
```

B) CTCFL

```
Human       tgagggggaatggataagg----------aggggag-----gagggccc-------acaggaggcactga
Mouse       -------gctggagtcctgcg-------gagaagat-----gacggac--------agaggggcgctgg
Rat         caacaaggcaagagtgctg----------ggagag-----gaaggtcct------gcaagagtcgagga
Dog         ------------------------------------------------------------------
Armadillo   ==================================================================
Elephant    ==================================================================
Opossum     cgaggggcgccgggaaaa----------aggagggccc--gaggtgcg-------tcgggggaaaagga
Chicken     tgacggcacgaggatgtggtgagggcacaaggagggcatggggagggca-------caaggatgtgctga
X. tropicalis tgggatagcaggtatagta----------gggagag-----atggtgcctatagtagcagtgggggataa
Tetraodon   ctgggtggaggtgatccaa----------cggtgga-----ggtgtgcag------ccagaccacaccca
                          GATA
```

C) EKLF

```
(1)
Human       cgccccctcacctgtgtgcgtgc---------gcagatgcgcc--------ttcaggtgggagctcttgg
Rhesus      caccccctcacctgtgtgcgtgc---------gcagatgcgcc--------ttcaggtgggagctcttgg
Mouse       cccgcccttacccgtgtgcgtgc---------gcaggtgcgcc--------ttgaggtgcgagctcttgg
Dog         cgcgccctcacctgtgtgcgtgc---------gcagatgagcc--------ttcaggtgtgagctcttgg
Horse       ==================================================================
Armadillo   ==================================================================
Opossum     taggctctcacctgtgtgggttc---------ttagatgggcc--------ttgaggtgggaactcttgg
Platypus    ctcctcctcaccggtgtgcgtcc---------tcagatgcgcc--------ttgagatgcgaactcttgg
Lizard      -----ctttacctgtgtgggtgc---------gcagatgggcc--------tttagatgggagctcttgg
Chicken     -----ccttacccgtgtgcgtcc---------gcatgtgcgcc--------ttgaggtgggagctcttgg
X. tropicalis cgtctgtttatgtgtattcccgctcctacaatacagatgccccctcgtattccattcgcgtgtattctg
Stickleback ----------------------------------------------------aggtgggagctcttgg
                      E12 E47 TAL1              E12 E47          E12 TAL1

(2*)
Human       ctgtgatagcccc---ttcgagggctc---ctctctgtccttagctga--ttggctgcagcct-ctgata
Rhesus      ctgtgatagcccc---ttcgagggctc---ctctctgcccttagctga--ttggctgcagcct-ctgata
Mouse       ctaggaaagtttc---ttggagggctc---ctgtctgcccacatctga--ttggctgtcttcc-ctgata
Dog         caccaatggcccc---tcagagggctc---ctctctgctctcagctga--ttggctgcagccc-ctgata
Horse       ctccagtggcccc---ttggagggctc---ctctctgccctcagctga--ttggctgcggccc-ctgata
                    GATA                                                  GATA
```

D) EPOR

```
Human      cagtccacagctgg----gtcagcagctgcctc--cgccggacgcagctgac---
Mouse      tagtccacagctgg----gtcagcagcagc-----tgccggacacagctgac---
Rat        tagtccacagctgg----gtcagcagcagc-----tgccggacacagctgac---
Rabbit     gggtccccagctgg----gccagcagctgcgtccgccccggacacagctggc---
Dog        cagtccacagctgg----gtcggcagctggctc--ctccagactcagctgac---
Armadillo  ctgtctgtggccgg----cgcagcaggtggttc--tgctgggcacagcgggc---
Elephant   NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
Opossum    gaggctacagctggagcaggcagcagctgcagt--atccggacacagctggcagc
                    E12 E47 TAL1        E12 E47 TAL1          E12 E47 TAL1
```

E) ETO2

```
Human    gcagccagctgtgtccccgtgataatgc
Mouse    gcagccagctgtgtcacagtgataatgc
Rat      gcagccagctgtgtcgccgtgataatgc
Dog      gcagccagctgtgtcaccgtgatagtac
Opossum  gcagccagctgtggctcggtgataatgc
              E12 E47 TAL1        GATA
```

F) EZH2

(1)

```
Human      ctcggataccaaaatccacagatgctcaaaagtccgtg
Rhesus     ctcggaaaccaaaatccacagatgctcaaaagtccctg
Mouse      ========================================
Dog        ========================================
Horse      ========================================
Armadillo  ----------------------------------------
Opossum    ========================================
Platypus   ========================================
Lizard     tgaagttatccatatacatggatgtcc--aggccccttt
Chicken    ========================================
                    GATA          E12 E47
```

(2)

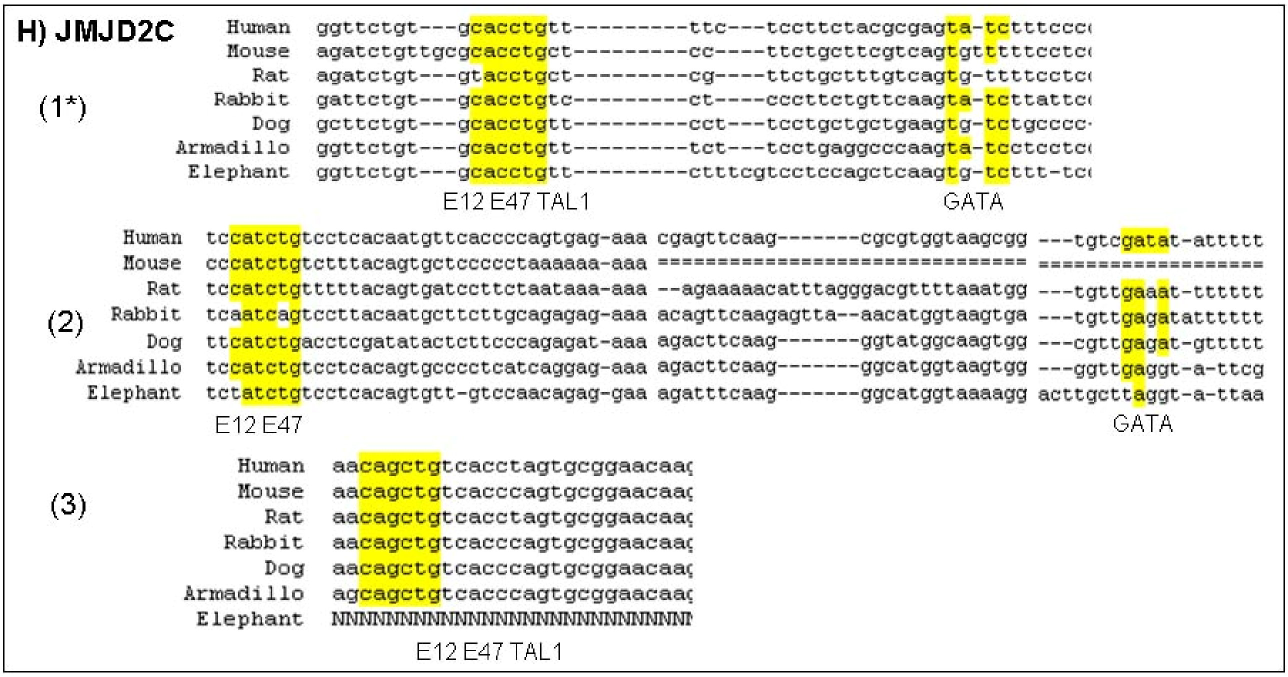
```
Human      -----tataacacctgtaaagcaggtt----aaaaatctagtg--tatcctcaa--
Rhesus     -----tgtaacacctctaaagctggtt----aaaaatctagtg--tattctcaa--
Mouse      -----tgtagcatctttaaagccggtt-----aaaatctagtg--actcaccag--
Dog        -----tgtagtatctctaatgcaagtt-----aacatctagtg--atttctcaa--
Horse      -----tgtaatatctctacagcagatt-----aaaatctagta--atttctcaa--
Armadillo  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
Opossum    -----agcaatattcctaa-----gtt----gtcaatgaaatt--aattct-aa--
Platypus   ----------catct----------tt-----tcaatcacgtggcagctcccta--
Lizard     -----atccacacacttttatcaagtc--------atttagcg--a-tcatcaa--
Chicken    -----actcacatgct-------------------------------------a--
X. tropicalis  gacaacttgagaccacccaacccagctcatgaaatgtatagtg--ttatacaaaac
Stickleback    ========================================================
                          E12 E47 TAL1                          GATA
```

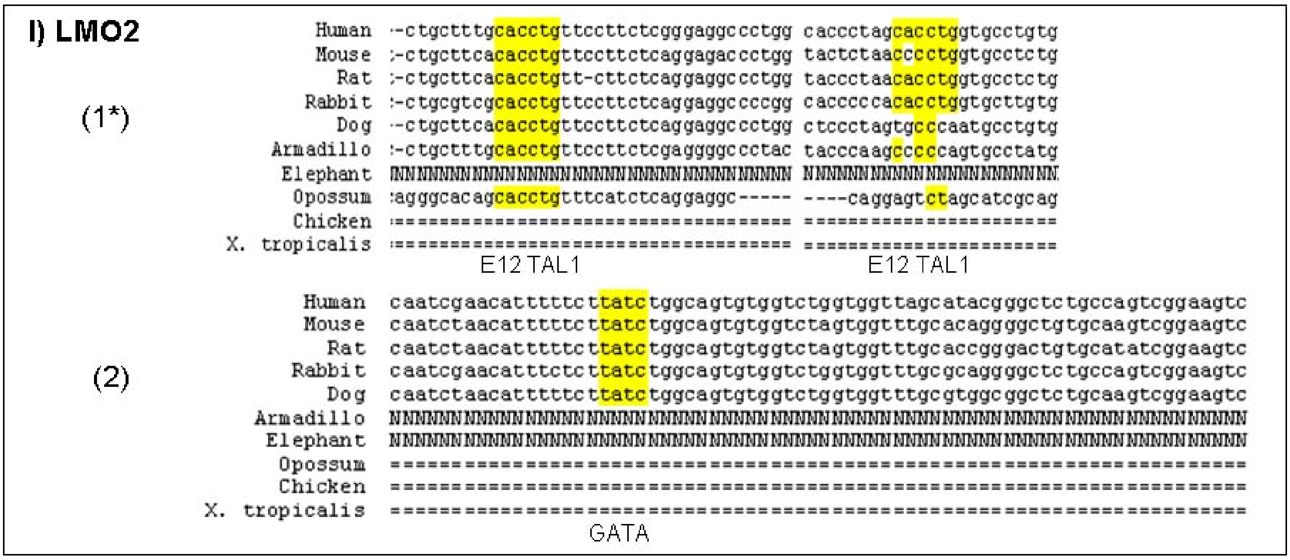
(3)

```
Human      aaccaccaacagatgacagag-
Rhesus     aatcaccaacaggtgacagag-
Mouse      ======================
Dog        agcctcccaccagagacaagg-
Horse      aaccacccaacaatgataaag-
Armadillo  tgcaaccaa-----aggtgag-
Opossum    ======================
Platypus   ======================
Chicken    ======================
                    E12 E47
```

G) FBXL10

```
Human      gctgggtcacacagtacacgtatctctccaggacataccagcacatctcatagtagaaggggtaacggaatttgggctgcacctgaaagcaaag
Rhesus     gctgggtcacacagtacacgtatctctccaggacataccagcacatctcatagtagaaggggtaacggaatttgggctgcacctgaaaacaaag
Mouse      gctgggtcacacagtacacatatctctccaagacataccagcacatctcatagtagaaggggtaacggaacttgggctgaacctgaaagcagag
Dog        gctgggtcacacagtacacgtatctttccaagacataccagcacatctcatagtagaaggggtaacgaaatttgggctgcacctgaaagcaaag
Horse      gctgggtcacacagtacacgtatctctccaagacataccagcacatctcatagtagaaggggtaacggaatttgggctgcacctgaaagcaaag
Armadillo  tctgggtcacacagtacacgtatcgctccaggacataccagcacatctcatagtagaaggggtagcgaaatttgggctgcacctgcaaccaaag
Opossum    gctgggtcacgcagtgtacgtatctctccaaaacataccaacacatctcatagtagaaagggtaccggaatttggcttgtacctaagagca-ac
Platypus   ggtgggtcagacagtgcacgtacctctccaggacataccagcacatctcatagtagaaagggtaacggaatttggcctgcacctgaaagca-ag
Lizard     gctgggcaacgcaaaagacgtacctctctaggacataccaacacatctcatagtagaacgggtagcggaacttggcctgcacctgcgagcaaa-
Chicken    gctgggtcacacaggacacgtatctctccagaacataccagcacatttcatagtagaaaggataacggaatttggcatgcacctgaaagcaag-
X. tropicalis  gtctggtcaaagaacagacatacctctctaggacataccaacacatctcataataaaaaggatagcggaacttggcatgtacctgtaagtat-g
Stickleback    gcttggtgaggcagtgcaggtatctctccaggacataccagcatatctcaaagtagaaggggaaacggaacttggaatggacctggagacaaag
                          GATA                                          E12 E47 TAL1
```


H) JMJD2C
(1*)
Human ggttctgt---gcacctgtt---------ttc---tccttctacgcgagta-tctttccc
Mouse agatctgttgcgcacctgct---------cc----ttctgcttcgtcagtgttttcctc
Rat agatctgt---gtacctgct---------cg----ttctgctttgtcagtg-ttttcctc
Rabbit gattctgt---gcacctgtc---------ct----cccttctgttcaagta-tcttattc
Dog gcttctgt---gcacctgtt---------cct---tcctgctgctgaagtg-tctgcccc-
Armadillo ggttctgt---gcacctgtt---------tct---tcctgaggcccaagta-tcctcctc
Elephant ggttctgt---gcacctgtt---------ctttcgtcctccagctcaagtg-tcttt-tc
E12 E47 TAL1
GATA
(2)
Human tccatctgtcctcacaatgttcaccccagtgag-aaa cgagttcaag-------cgcgtggtaagcgg ---tgtcgatat-attttt
Mouse cccatctgtctttacagtgctcccccctaaaaaa-aaa ================================ ==================
Rat tccatctgtttttacagtgatccttctaataaa-aaa --agaaaaacatttagggacgttttaaatgg ---tgttgaaat-tttttt
Rabbit tcaatcagtccttacaatgcttcttgcagagag-aaa acagttcaagagtta--aacatggtaagtga ---tgttgagatattttttt
Dog ttcatctgacctcgatatactcttcccagagat-aaa agacttcaag-------ggtatggcaagtgg ---cgttgagat-gttttt
Armadillo tccatctgtcctcacagtgcccctcatcaggag-aaa agacttcaag-------ggcatggtaagtgg ---ggttgaggt-a-ttcg
Elephant tctatctgtcctcacagtgtt-gtccaacagag-gaa agatttcaag-------ggcatggtaaaggg acttgcttaggt-a-ttaa
E12 E47
GATA
(3)
Human aacagctgtcacctagtgcggaacaac
Mouse aacagctgtcacccagtgcggaacaac
Rat aacagctgtcacctagtgcggaacaac
Rabbit aacagctgtcacccagtgcggaacaac
Dog aacagctgtcacccagtgcggaacaac
Armadillo agcagctgtcacccagtgcggaacaac
Elephant NNNNNNNNNNNNNNNNNNNNNNNNNNN
E12 E47 TAL1


I) LMO2
(1*)
Human --ctgctttgcacctgttccttctcgggaggccctgg caccctagcacctggtgcctgtg
Mouse --ctgcttcacacctgttccttctcaggagaccctgg tactctaaccccctggtgcctctg
Rat --ctgcttcacacctgtt-cttctcaggaggccctgg taccctaacacctggtgcctctg
Rabbit --ctgcgtcgcacctgttccttctcaggaggcccgg cacccccacacctggtgcttgtg
Dog --ctgcttcacacctgttccttctcaggaggccctgg ctccctagtgcccaatgcctgtg
Armadillo --ctgctttgcacctgttccttctcgaggggccctac tacccaagccccagtgcctatg
Elephant NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN NNNNNNNNNNNNNNNNNNNNNNN
Opossum -agggcacagcacctgtttcatctcaggaggc----- ----caggagtctagcatcgcag
Chicken ==================================== =======================
X. tropicalis ==================================== =======================
E12 TAL1
E12 TAL1
(2)
Human caatcgaacatttttcttatctggcagtgtggtctggtggttagcatacgggctctgccagtcggaagtc
Mouse caatctaacatttttcttatctggcagtgtggtctagtggtttgcacaggggctgtgcaagtcggaagtc
Rat caatctaacatttttcttatctggcagtgtggtctagtggtttgcaccgggactgtgcatatcggaagtc
Rabbit caatcgaacatttctcttatctggcagtgtggtctggtggtttgcgcaggggctctgccagtcggaagtc
Dog caatctaacatttttcttatctggcagtgtggtctggtggtttgcgtggcggctctgcaagtcggaagtc
Armadillo NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
Elephant NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
Opossum ==
Chicken ==
X. tropicalis ==
GATA


J) LYL1
(1*)
Human agtgcagataagggccaggctgcct ggccgcgcctgataaggagcctg
Mouse cgtgcagataagggtctggctgcct ggccgcggctgataagaagcaat
Rat cgtgcagataagggtctggctgcct ggccgcagctgataagaagcggt
Dog ggcacagataagggcctggccgact ggccgcgcctgataaggagccgg
Elephant ggcgcagataagagcctggccgcct ggccgcgcctgataaggagccgc
Opossum ========================= =======================
GATA
GATA
(2)
Human ccagggcaggaagcacctgggtctggagcggatgcacagggc
Mouse -caggacaggaagcacctgggtctggagcggatgcacagaat
Rat -caggacaggaagcacctgggtctggagcggatgcacagaat
Dog gcggagcaggaagcacctgggtctggagcggatgcaccaggc
Elephant actgggcaggaagcacctgggtctggaacggatgcacagggc
Opossum ccggggccgagagggcgtgcactaaagttgggctccaaacc
E12 E47 TAL1

K) SCL

```
Human     gcaggtgggaatgagcgataaggattgggggt
Chimp     gcaggtgggaatgagcgataaggattgggggt
Mouse     gcaggtgggagctggcgataagga-agagggt
Rat       gcaggtgggaactggtgataagga-cgagggt
Dog       gcaggtgggaagaggcgataaggg--------
          . . . . . . . . . . . . . . . . .
            E-box       GATA
```

L) SMARCA5

(1)

```
Human  ggatggccgagtgcgcagtggggataagcggg-gcgcagctgc
Mouse  ggagggccgagtgcgcagtggggataagcggg-gcgcagctgc
Rat    ggagggccgaatgcgcagtggggataagcggg-gcgcagctgc
Rabbit ggatggccgagtgcgcagtggggatgagctggcgcgcagctgc
Dog    ggaaggccgagtgcgcagtggggataagcggg-gcgcagctgc
Armadillo -gggggccgagtgcgcagtggggataaacggg-gcgcagctgc
Elephant ggatggc--gatgcgcagtggggatag--cgg-gcgcagctgc
Opossum  ==========================================
Chicken  ==========================================
X. tropicalis ======================================
                                      GATA      E12 E47 TAL1
```

(2)

```
Human  :tcgttatcaaacagactaca ctttactc-------aacaaacatttaga----gaatactgaatatgaggaataaaaaaagataaaaaa
Mouse  :--------gaatggaaca-- .ttttattc-------aacataggt---------ggggactaaata-atgggattaattaagat---aaa
Rat    -------------------- .ttttattc-------aatataggt---------ggggacgaaata-acgagcgggattaagatgacaaa
Rabbit NNNNNNNNNNNNNNNNNNNN NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
Dog    ==================== ======================================================================
Armadillo :atgttctcaactggaataca .ctttat---------aacactgatt--------gaataccggacatggggtt---acaaagattaacaa
Opossum  ==================== ======================================================================
Chicken  ==================== ======================================================================
X. tropicalis ============== ======================================================================
              GATA                                                                        GATA
```

(3*)

```
Human  :-gctggcgagatagatgttatcttcc-
Mouse  :ggctggaga---agagaatccactcc-
Rat    :ggttggaga---agataatctactcc-
Rabbit :-gctggcgagctagatgatatctgcc-
Dog    :-gctggtgagatagatgatttttttcc
Armadillo :-gctggcgagataggtattatattcc-
Elephant :-gctggcgacatagatgatatctgac-
Opossum  :-ggtggcgaggagggtgagaagcggc-
Chicken  :============================
X. tropicalis :=======================
                        GATA    GATA
```

M) PCQAP

(1)

```
Human  --ggagcacctgccgtg
Mouse  ==================
Rat    ==================
Rabbit --accgcccttgcc-tg
Dog    ctggaattcccgtcccg
Armadillo ==================
Opossum   ==================
Chicken   ==================
X. tropicalis ==================
```

E12 E47 TAL1

(2)

```
Human  cacctgtaatcccagctacttgggagg
Mouse  ==========================
Rat    ==========================
Dog    --------------------------
Opossum   ==========================
Chicken   ==========================
X. tropicalis ==========================
```

E12 E47 TAL1

(3)

```
Human ccacc-------tggtgcttattgaccacttgggctccagatttt tcctcacgtcacctccct-------c---- cccagggtagcagataaaggc
Mouse ========================================== ============================ ====================
Rat   ========================================== ============================ ====================
Dog   tcact-------tggcacttgggaaccatctaggtttaggattt ttgccacctacccccacccc aaaac---- ccctgaggcacagacaagggc
Armadillo ccacc----------ttcttgtgaaccactggggttcagggttt ttgtcacctgccggcctt-------tcttc cccagggacacaggaaacggt
Opossum   ========================================== ============================ ====================
Chicken   ========================================== ============================ ====================
X. tropicalis ========================================== ============================ ====================
```

E12 TAL1 GATA

(4*)

```
Human  gggtgcaggtgggctgagtccgaaaagagagtcagcgaagggagatagggggtggggtcgttttataggat
Mouse  ------------------------------------------------------------------------
Rat    ------------------------------------------------------------------------
Rabbit ========================================================================
Dog    ========================================================================
Armadillo ========================================================================
Opossum   ========================================================================
Chicken   ========================================================================
X. tropicalis ========================================================================
```

E12 E47 TAL1 GATA

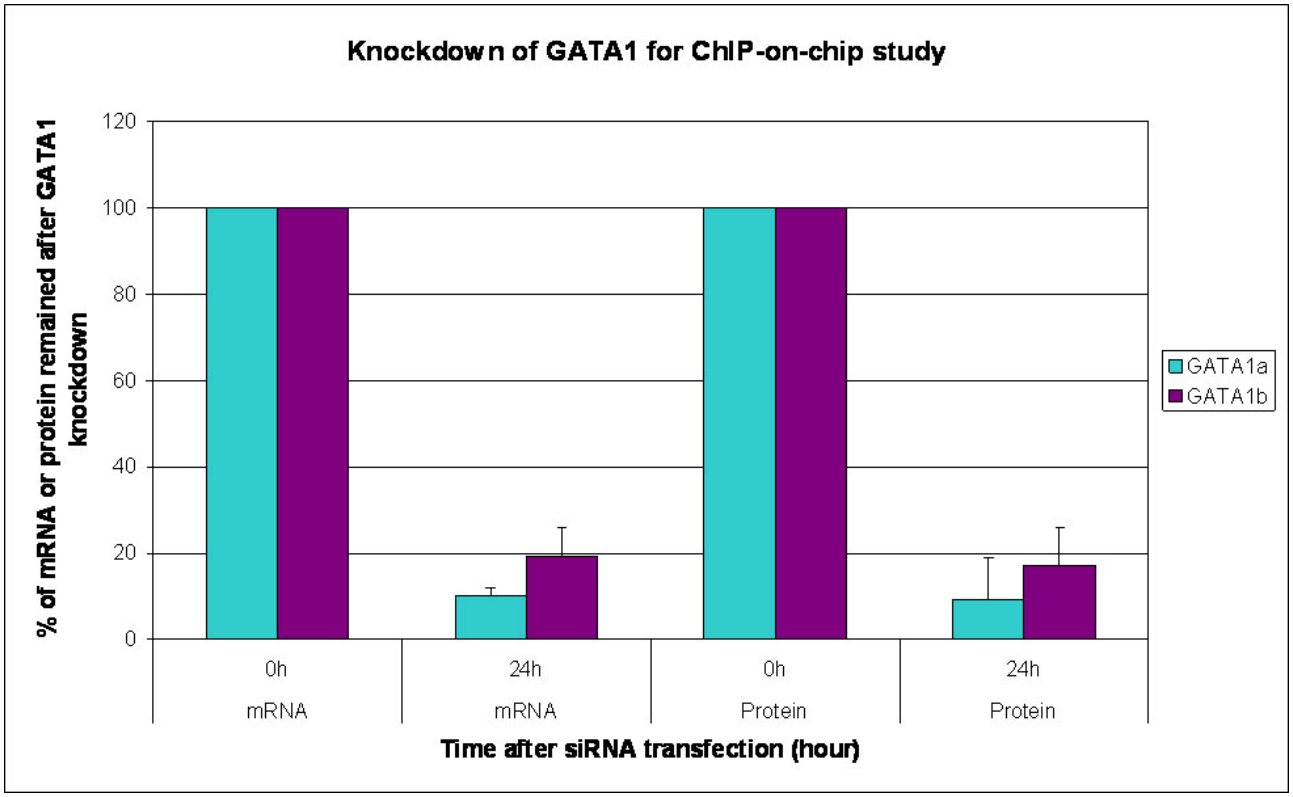
N) RSF1

```
Human agcaccttcttca-------actgtt
Mouse agcaccttcttca-------actgtt
Rat   agcaccttcttca-------actgtt
Rabbit NNNNNNNNNNNNNNNNNNNNNNNNN
Dog   agcaccttcttca-------actgtt
Armadillo agcaccttcttca-------actgtt
Elephant agcacttccttc--------acagtt
Opossum agcaccttcttct-------gctgtc
```

E12 E47

Appendix 5

Confirmation of GATA1 knockdown by qPCR and western blotting in GATA1 ChIP-on-chip study



Two siRNAs directed against GATA1 were used: GATA1a and GATA1b. Knockdown of GATA1 at the mRNA level was quantified by quantitative PCR as described in Chapter 3. Knockdown of GATA1 at the protein level was quantified by densitometry of bands as determined by immuno-detection of the western blotting. Bar chart showed the mRNA or protein level of GATA1 remaining (y-axis) after siRNA transfection relative to luciferase siRNA transfection across the time points (x-axis) Blue bars indicate mRNA levels in the GATA1a condition while the purple bars indicate mRNA levels in the GATA1b condition. Error bars show the standard deviation across three independent biological replicates.