Chapter 5

Optimisation of conditions for ChIP-chip when using cell types that are limiting in number

5.1. Introduction

As discussed in Chapters 1 and 3, the recent development of microarray technology coupled with chromatin immunoprecipitation (ChIP-chip) has accelerated our understanding and annotation of DNA-protein interactions. ChIP-chip methods were initially developed in yeast (Ren *et al*., 2000; Iyer *et al*., 2001) and subsequently applied to study DNA-protein interactions in other organisms including humans. However, the development of ChIP itself has its origins long before ChIP-chip was developed. In the 1980's a protocol for investigating DNA-protein interactions in living cells was developed (Solomon and Varshavsky, 1985; Solomon *et al*., 1988) and improved upon through the years to produce a ChIP method that is now widely used (Orlando, 2000). The fundamental steps performed in this ChIP procedure include cross-linking of DNA-protein interactions in living cells, extraction and shearing of chromatin, immuno-affinity 'pulldown' of chromatin bound by a protein of interest, and reversal of cross-links to isolate the interacting DNA (Figure 5.1). The isolated DNA can then be interrogated by microarray hybridisation. However, there are several technical parameters which need to be optimised to ensure the success of individual ChIP-chip experiments. These parameters are discussed below.

Preparation of cross-linked chromatin: Cross-linked chromatin can be prepared by the use of UV light or formaldehyde to covalently attach proteins to DNA sites *in vivo* (Orlando and Paro, 1993; Walter *et al*., 1994; Biggin, 1999). The most commonly used cross-linking agent is formaldehyde, which was first used in the late 1970's to investigate *in vivo* DNA-protein interactions (Varshavsky *et al*., 1979). It readily permeates the cell and nuclear membrane and generates a covalent cross-link between free amines on DNA, RNA or proteins, producing DNA-protein, RNA-protein, and protein-protein cross-links (Orlando *et al*., 1997). Other cross-linking agents have been used in combination with formaldehyde to detect indirect associations between DNA and transcription co-factors,

including dimethyl apidimidate (DMA) (Kurdistani *et al*., 2002), and disuccinimidyl glutarate (DSG) (Tian *et al*., 2005). The amount of cross-linking reagent and incubation time can affect the accessibility of the antibody to the antigen epitope (Orlando, 2000), the fragmentation of the chromatin (Orlando *et al*., 1997), and the efficiency by which weaker protein-DNA interactions are cross-linked. In general proteins with weaker DNA interactions require longer cross-linking times but it is important to optimize this as longer cross-linking times will increase the size of sonication fragments that can be obtained. Fragmentation of chromatin by sonication requires optimisation to ensure that the majority of the sonicated DNA is in the 300-1000bp range as larger DNA fragments are not efficiently immunoprecipitated (Orlando *et al*., 1997). Chromatin can also be prepared without the use of cross-linking reagents and in this method the 'native' chromatin is fragmented through nuclease digestion (NChIP) (O'Neill and Turner, 2003). This method can be used to detect modified histones, but cannot be used to detect the majority of non-histone interactions such as transcription factor binding events. Fragmentation of 'native' chromatin by micrococcal nuclease digestion also requires careful optimisation as over-digestion can result in sub-nucleosomal particles, which are not as efficiently precipitated as oligo-nucleosomes (O'Neill and Turner, 1995).

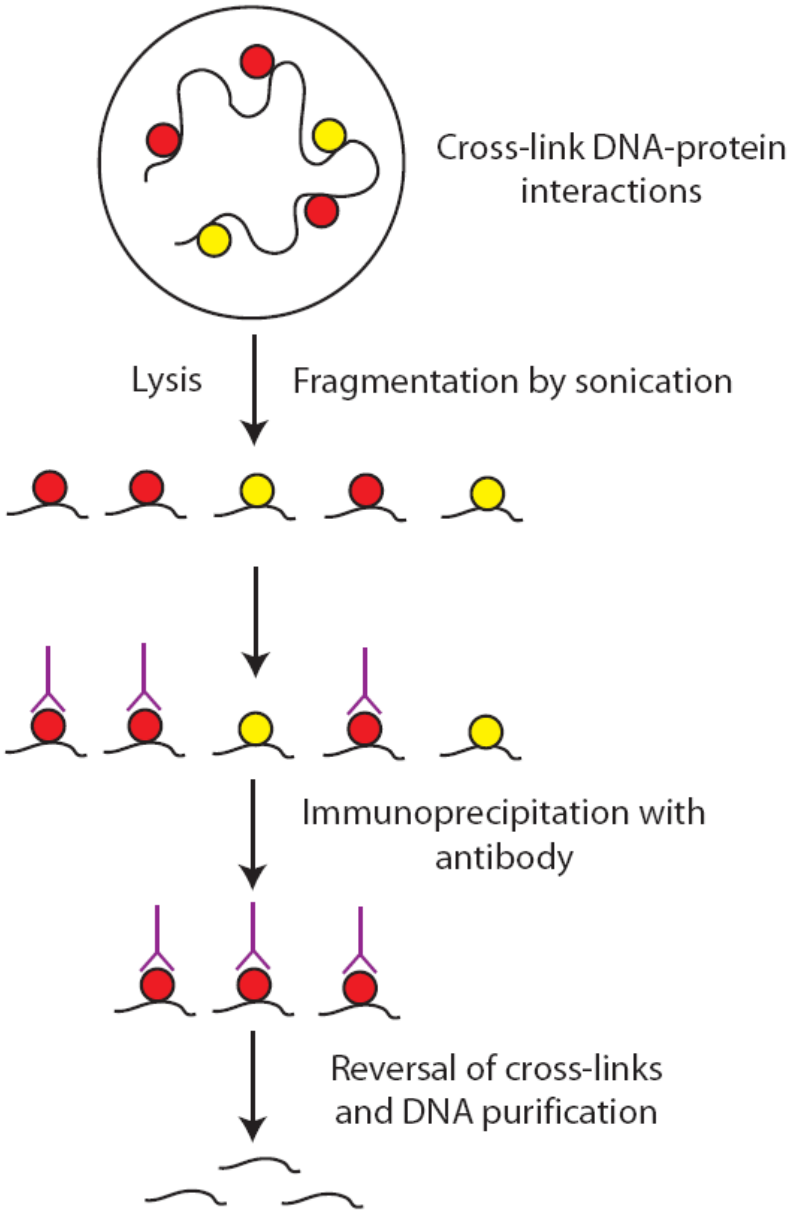

Figure 5.1 Principles of ChIP. The main principles of the ChIP procedure are outlined in the figure, beginning with the cross-linking of DNA-protein interactions in living cells followed by lysis and sonication of DNA. The DNA associated with a protein of interest (red circle) is immunoprecipitated by a specific antibody, the DNA-protein cross-links are reversed and the DNA which interacted with the protein is isolated.

Antibody specificity, affinity and epitope accessibility: The specificity and affinity of an antibody is key to the success of any ChIP-chip experiment. While western blotting and competiton assays with peptides containing the epitope of interest can highlight the

specificity of an antibody (Suka *et al*., 2001), these techniques cannot predict which antibodies will bind to the protein of interest *in vivo*. An epitope on a target protein may be masked by formaldehyde cross-linking, meaning that a number of antibodies raised to different epitopes are usually tested in ChIP-chip assays to identify one which can efficiently immunoprecipitate the target protein and its bound DNA (Horak *et al*., 2002). It is often important to use antibodies generated against C-terminus or the N-terminus epitopes, as these are often exposed on intact proteins and allow greater accessibility compared to epitopes buried within the protein. In addition, antibodies may cross-react with other proteins containing similar epitopes, which is often a problem when investigating members of the same protein family as they are often highly similar at the amino acid sequence level. Limitations in the number of highly specific antibodies for use in ChIP assays have led investigators to adopt alternative approaches for the study of DNA-protein interactions. An epitope tagging approach has been used successfully in yeast to map a number of DNA-protein interactions (Lee *et al*., 2002). In this approach, the DNA sequence for a defined epitope such as a c-Myc, TAP, His, FLAG, HA, V5 or Pk tag is introduced to a locus whose interactions are to be mapped (Puig *et al*., 2001). The tagged protein and its associated DNA binding sites can then be isolated using highly specific immuno-affinity approaches. This approach is more difficult in mammalian systems owing to the difficulty in introducing targeted genetic alterations and ensuring that the tagged transcription factor retains its normal DNA binding profile.

Amplification and labeling of ChIP DNA: The amount of DNA recovered from a ChIP assay is in a limiting quantity, typically in the range of 10 to 200 hundred nanograms. The limits of sensitivity of most array platforms require microgram amounts of target DNA to be labeled and hybridised to the array. Therefore a number of studies have combined multiple ChIP DNAs (Weinmann *et al*., 2002) or used DNA amplification to obtain enough DNA for microarray hybridisation. Three amplification methods have been reported: (i) ligation-mediated PCR (Ren *et al*., 2000) in which a double stranded oligo is ligated to the end of ChIP DNAs, which acts as a template for PCR amplification, (ii) random amplification in which a degenerate oligo sequence randomly anneals to DNA allowing the DNA to be amplified by PCR (Iyer *et al*., 2001), and (iii) T7-based linear amplification, in which poly dTs are added to the end of DNA by terminal

transferase, followed by a round of DNA synthesis using polyA-T7 sequences as primers (Liu *et al*., 2003; Bernstein *et al*.,2005). The synthesized DNA is then transcribed into RNA by T7 RNA polymerase. However, these methods can introduce sequence and length dependent biases during the amplification process, which can directly affect the ChIP-chip data. Following amplification, the DNA is fluorescently labeled by random priming using the Klenow fragment of *E. coli* DNA polymerase. Labeling by random priming can result in an additional amplification of 10-100 fold due to the strand displacement activity of the exonuclease deficient Klenow fragment (Walker *et al*., 1992). Klenow-mediated labeling of ChIP samples, without prior amplification of ChIP DNA samples, can be used to increase the amount of material in a non-biased way for ChIP-chip experiments (Koch *et al*., 2007, De Gobbi *et al*., 2007; Dhami, submitted).

Array platforms: A number of array platforms are available for ChIP-chip experiments, each of which has its own advantages and disadvantages. An array platform is judged on a number of features, which include (i) the density or number of features, (ii) the resolution of the array features, (ii) the percentage of genome sequence covered by the array and (iv) the sensitivity and reproducibility at reporting a wide dynamic range of enrichments in ChIP samples. Tiling path arrays which contain probes spanning large continuous regions of the human genome are commonly used for ChIP-chip studies and can be composed of oligonucleotide or PCR product probes (Bernstein *et al*., 2005; Koch *et al*., 2007). Microarrays containing oligonucleotide probes are used exclusively for very large scale studies as a much greater number of array features can be accommodated and the resolution is much higher. In addition, oligonucleotide probes can be synthesized directly onto the slide and avoid the necessity of PCR amplification. However, the use of PCR products as microarray probes may result in higher signal:noise ratios due to the length of the probe available for hybridisation. Therefore this type of array platform is often more sensitive and reliable in reporting enrichments compared to array platforms which use much smaller oligonucleotide probes (D. Vetrie, unpublished data).

Analysis of data: The analysis of 'raw' ChIP-chip data sets is a crucial step in identifying *bona fide* sites of DNA-protein interaction. Often the first step in the analysis procedure is normalisation. Normalisation is an important process in which any systematic and non-systematic biases in a ChIP-chip dataset can be accounted for and

adjusted accordingly to ensure that only true ChIP enrichments are reported in the dataset. Biases in datasets which can be accounted for by normalisation include low signal, high background in one channel (when performing two-channel hybridisation experiments), uneven signal in the two channels, and non-specific binding of antibodies. Standard array quantitation programs can be used to perform LOWESS normalization which allows for biases in ratio calculations to be accounted for in a signal-dependent manner (Quackenbush, 2002). This is a form of local normalization in which all data-points are not treated equally. On the other hand, total scaling normalization is a form of global normalization in which all datapoints are treated equally. A scaling factor is calculated to ensure that the median ratio of the data set is 1. This normalization method is usually used due to uneven signal intensity in the two microarray channels. Non-specific binding normalization can be used to correct for variations in non-specific binding as measured by a species-specific IgG control experiment (as discussed and performed in Chapters 3 and 4 of this thesis). This form of local normalization does not treat all data points equally; instead a normalized ratio is calculated by dividing the experimental ratio at individual data-points by the corresponding IgG ratio.

Once data sets have been appropriately normalized, a number of approaches can be used to determine which regions correspond to the binding of a protein of interest. A fold enrichment cut-off or a standard deviation cut-off can be used to define significant enrichments. However these methods make a number of assumptions regarding the dataset, namely that there is a normal Gaussian distribution of a dataset around the mean and that both negative and positive regions are present in the data set in similar proportions. These assumptions may not apply to all data sets and could result in false positive regions being identified. Several other more complex statistical analysis methods have been developed to analyse the likelihood of binding, which include single-array error model (Ren *et al*., 2000), rank-based target identification (Iyer *et al*., 2001) and linear regression models (Buck *et al*., 2005; Gibbons *et al*., 2005; Li *et al*., 2005). The single-array error model also assumes a normal distribution of the data and calculates a one-sided probability to determine which peaks of enrichments are significant, while rank-based identification does make any assumptions about the distribution of ChIP-chip data, instead it requires multiple replicates to determine whether an enrichment is

significant or not. Linear regression methods, such as ChIPOTle (discussed in Chapter 3), utilise two properties of ChIP-chip data to identify *bona fide* interactions. Namely, that enrichment of DNA fragments adjacent to sites of direct interaction is observed. Also the cluster of enriched elements surrounding a peak forms a distinct shape, with tailing off left and right as you move away from the site of direct interaction. Hidden Markov Models can also be used to identify significant peaks of enrichment in ChIP-chip data by portioning the data into locations that are either consistent or inconsistent with antibody binding and identifies protein interaction sites and centres with their probabilities (Koch *et al*., 2007). While these methods are useful for analysing discrete interactions, they are often unreliable when it comes to identifying low-level enrichments spread over large distances such as histone modifications. Furthermore, hierarchical clustering (Eisen *et al*., 1998) may be used to identify continous features.

Number of cells required to perform a ChIP reaction: While a number of the technical considerations discussed above have been overcome in recent years with the rapid development of microarray, antibody, and analysis technology, there is still one major caveat when performing ChIP experiments in combination with microarrays. This is the requirement for large numbers of cells, typically 10^7 per ChIP-chip assay. While this is usually not a problem when using cultured cell lines, it can be a difficult to obtain enough primary cells for ChIP-chip experiments. This is a particular problem in higher organisms, such as mammals, whose specialised cell types are often rare or difficult to access for experimentation.

There are numerous examples of biological questions which require ChIP-chip based approaches to help further our knowledge of fundamental issues relating to development and disease but which are also limited by the small amounts of tissue available for study. For example the epigenetic analysis of primary cells in many developmental systems and cancer stem cells (Reya *et al*., 2001) are examples where cell numbers are limiting as typically 3-4 orders of magnitude fewer cells are obtained than currently used in ChIP-chip protocols.

Recently, O'Neill and colleagues described a protocol which they called carrier ChIP (CChIP), in which 'native' or un-crosslinked chromatin immunoprecipitation (NChIP) procedures were used in combination with *Drosophila* 'carrier' chromatin to investigate

histone modifications in small populations of cells (O'Neill *et al*., 2006). Cultured mouse ES cells were used as a model system to test this new protocol and it was shown that CChIP could be used to analyse the distribution of histone modifications from 1,000 cells. The protocol was shown to be reproducible and was then applied to study histone modifications in samples containing 50-100 cells obtained from the inner cell mass and trophectoderm of mouse embryos. Fold enrichments were determined using radioactive PCR and phosphorimaging and the modest enrichments observed suggest that this procedure may not be applicable to microarray analysis. The presence of carrier chromatin which makes up the vast bulk of the DNA in the ChIP sample will invariably affect labeling efficiency of the material being assayed which makes up a very small proportion of the DNA in the sample. In addition the authors suggest that this CChIP protocol may not be applicable to formaldehyde cross-linked chromatin (XChIP), in which yields tends to be low when immunoprecipitating with antibodies raised to modified histones. This may be due to cross-linking of lysine-rich histone tails to nearby DNA, obscuring antibody access to its epitope (Robyr and Grunstein, 2003; Suka *et al*., 2001). Less than 1% of bound and unbound fractions are typically recovered in XChIP (compared to 10-20% yields in NChIP), which O'Neill and colleagues suggest may prevent the use of cross-linked chromatin in the analysis of histone modifcations in very small cell populations.

More recently two XChIP-based methods have been developed to study histone modifications in limited populations of cells, known as Q^2ChIP and miniChIP (Dahl and Collas, 2007; Attema *et al*., 2007). MiniChIP was used to analyse histone modifications in 50,000 cells using quantitative PCR while the Q^2ChIP method in combination with quantitative PCR showed that it was possible to analyse histone modifications in as few as 100 cells.

Thus, developing a robust ChIP-chip method for use with much fewer cells would allow us to understand global epigenetic events which may be specific to small populations of cells, which would be not be determined when studying mixed populations of cells. In addition, developing a ChIP protocol for use with fewer cells may have cost implications too. It would cost less to do whole-genome studies if lower numbers of cells and antibody

amounts could be used to generate the same data sets as experiments performed with more conventional cell numbers and antibody amounts.

The development of a highly sensitive PCR product-based array platform (Dhami *et al*., 2005) precludes the need for amplification of ChIP DNA prior to array hybridization (Koch *et al*., 2007). While this platform is sensitive enough to detect histone and transcription factor interactions from unamplified ChIP DNA derived from assays performed with 10^7 cells, it was not clear if this platform could be used to faithfully detect interactions from fewer cells. As the SCL locus tiling path array had been used to characterize in detail a number of histone modifications and transcription factor interactions from ChIP assays performed with 10^7 cells (Dhami, PhD Thesis, University of Cambridge, 2005) (Chaper3) it represented an ideal model system in which to develop a ChIP-chip protocol for the detection of DNA-protein interactions from reduced numbers of cells. The success of a low cell numbers ChIP-chip protocol could therefore be determined by comparisons with known profiles of ChIP enrichments across the SCL locus.

5.2. Aims of this Chapter

The aim of this chapter was to further develop existing ChIP-chip approaches in order to improve the sensitivity of the method when using cells which are limiting in number. This method will then be exploited to perform analyses of regulatory interactions in one such cell type in Chapter 6 of this thesis. Therefore, the aims of the work presented in this Chapter were:

1. To determine parameters important for developing low cell number ChIP-chip.
2. To determine the types of DNA interactions (histone modifications and transcription factors) that can be identified with these protocol developments.
3. To evaluate the sensitivity and reproducibility of the method at detecting known ChIP-chip profiles at the SCL locus. Quantitative PCR would be used to determine sensitivity while independent biological replicates would be used to confirm the reproducibility of the method.

5.3. Overall strategy

As discussed in Chapters 3 and 4, the human SCL locus has been extensively characterized for a number of histone modification and transcription factor interactions in the K562 cell line (Dhami, submitted). These results could be used to assess the limit of detection of known regulatory elements using low cell number ChIP-chip. The SCL tiling path array would then be used to develop a ChIP-chip protocol for the detection of histone acetyl and methyl modification patterns from a range of cell numbers. Various parameters were tested to optimize the ChIP protocol for use with low cell numbers and included antibody-chromatin ratio, amount of protein G agarose used to immunoprecipitate DNA-protein-antibody complexes, and the volume in which immunoprecipitations were performed. Statistical analysis would be used to identify the antibody amount that gave the optimal signal:noise ratio based on the analysis of known regulatory elements across the SCL locus. Given that signal:noise ratios may be compromised when performing experiments with low cell numbers, normalization methods would be investigated which improved signal:noise. The reproducibility of the low cell numbers protocol would be tested by performing replicates. Furthermore, histone modification enrichments reported by the SCL array in low cell number assays would be verified by SYBR green real-time PCR. Finally, the detection of transcription factor interactions in reduced cell numbers would also be investigated.

Results

5.4. The effect of titrating cell numbers in ChIP-chip and its effect on the detection of regulatory elements at the human SCL tiling locus

The SCL array had been previously shown to be a highly reproducible platform for the detection of known regulatory elements associated with histone H3 K9/K14 diacetylation (also referred to as histone H3 acetylation) in K562 cells (Dhami, submitted). Histone H3 acetylation is associated with active promoters and enhancers (Chapter 3). Initial experiments aimed at develop a ChIP-chip protocol for lower cell numbers focused on establishing the reproducible detection of known SCL regulatory elements in histone H3 acetylation ChIP assays whilst titrating cell numbers used.

Chromatin was prepared from 10^8 K562 cells and aliquots equivalent to 10^6, 10^5, and 10^4 cells were used to perform H3 K9/K14 acetylation ChIP assays as described in Chapter 2. All other conditions were the same as for ChIP assays performed with 10^7 cells. The only variable in these experiments was the amount of chromatin (i.e., cell number equivalents) used in each assay. Significant enrichments in ChIP-chip experiments were obtained by determining the mean ratios of two "background" regions represented on the SCL array which did not contain any known regulatory elements in K562 (Dhami, PhD Thesis, University of Cambridge, 2005). These regions spanned 47262287 bp to 47343557 bp, and 47424426 bp to 47489321 bp on chromosome 1. The significant enrichment threshold was three standard deviations above the mean background ratios. All known SCL regulatory regions associated with significant histone H3 acetylation in K562 cells were detected in three of the four assays (10^7, 10^6 and 10^5 cells) with varying degrees of enrichment (Table 5.1 and Figure 5.2). Some of these enrichments were located at or near the SCL, SIL and KCY promoters as these genes are expressed in K562.

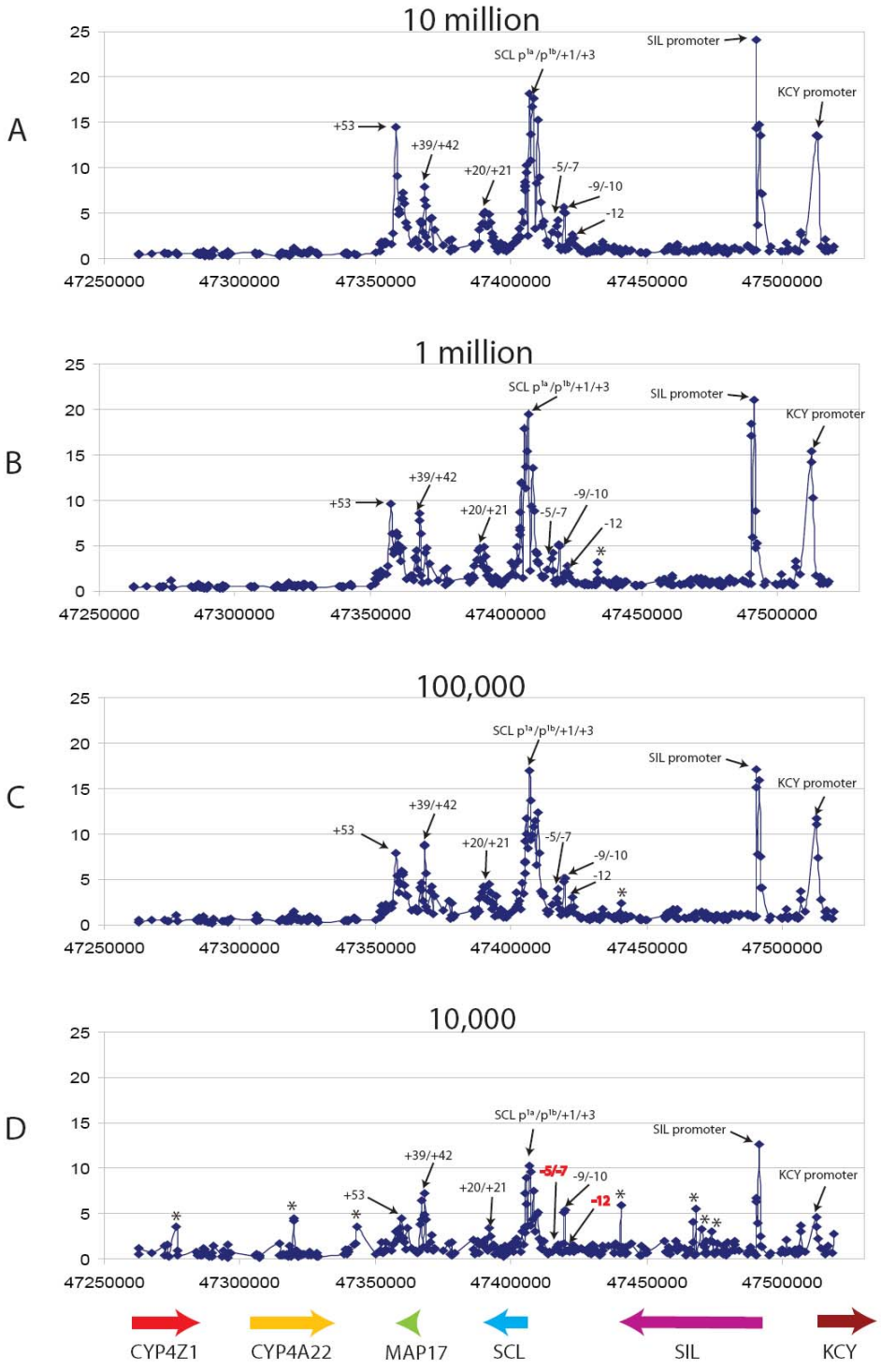

Figure 5.2: ChIP-chip profiles of H4 K9/K14 diacetylation across the human SCL locus in K562 for a range of cell numbers. Known regulatory elements showing significant enrichment for H3 K9/K14 diacetylation are indicated with black arrows. The nomenclature of regions denoted by +1, +3 etc. are based on their distance upstream (-) or downstream (+) in kilobases from the SCL promoter p^{1a}. The chromosome 1 coordinates are shown along the x-axis for each panel and the y-axis indicates the fold-enrichments. The

thick coloured arrows at the bottom of the figure represent the gene order and direction of transcription. Panel A represents data obtained from the ChIP-chip experiment conducted with 10 million cells and 10 μg of antibody. Panel B represents the 1 million cells experiment, panel C the 100,000 cells experiment, and finally panel D represents the data obtained from the experiment conducted with 10,000 cells. In panels B-D, significant enrichments that are not associated with regulatory elements are indicated with an asterices. In panel D, the -5/-7 and -12 regulatory elements are not detected as significant peaks and are indicated in red lettering.

Histone acetylation at the 5' end of the SCL gene extended 8 kb into the coding region and is found at regions such as the +1 and +3 regions which show DNase I hypersensitivity (Leroy-Viard *et al*., 1994; Follows *et al*., 2006). Enrichments were also found at the stem cell enhancer (+20/+21) (Gottgens *et al*., 2002), the MAP17 promoter (+42), the -9/-10 region (Gottgens *et al*., 1997) and the SCL erythroid enhancer (from +50 to +53) (Delabesse *et al*., 2005). In addition, significant enrichments were also observed at the -12, -5/-7 and +39 regions as described previously (Dhami, PhD Thesis, University of Cambridge, 2005). No significant enrichments were observed across the genomic region containing the CYP4Z1 and CYP4A22 genes (which are not expressed in K562) in assays performed with 10^7, 10^6 and 10^5 cells. However, both the 10^6 and 10^5 cell assays contained one significant peak of enrichment not associated with any known regulatory element.

Overall, the fold enrichment ratios reported for the 10^4 cell experiment were lower than the other three experiments. Whilst seven SCL regulatory regions were detected, two known elements (-5/-7 and -12) were not significantly enriched in this assay. However, unexpectedly, two significant peaks of enrichment were observed within the CYP4Z1 and CYP4A22 genes in an assay performed with 10^4 cells and seven other peaks of significant enrichment were also detected in this assay, which were not associated with known regulatory elements (indicated by asterisks in Figure 5.2).

Number of cells	No. of known regulatory Elements detected	Known regulatory elements detected	No. of elements not associated with known regulatory function	No. of known regulatory elements not detected	Mean CV of ratios (%)
10^7	9/9	KCY p, SILp, -12, -9/-10, -5/-7, SCLp, +20/+21, +39/+42, +53	0	0	6.45
10^6	9/9	KCY p, SILp, -12, -9/-10, -5/-7, SCLp, +20/+21, +39/+42, +53	1	0	8.01
10^5	9/9	KCY p, SILp, -12, -9/-10, -5/-7, SCLp, +20/+21, +39/+42, +53	1	0	8.85
10^4	7/9	KCY p, SIL p, -9/-10, SCLp, +20/+21, +39/+42, +53	7	2	22.42

Table 5.1: A summary of the performance of H3 K9/K14 diacetylation ChIP-chip assays performed with a standard range of cell numbers and 10 µg of antibody. The SCL tiling array was used to detect regulatory elements associated with H3 K9/K14 diacetylation in assays performed with 10^7, 10^6, 10^5, and 10^4 cells and 10ug of antibody. Assays performed with 10^7, 10^6, and 10^5 cells were associated with a low mean CV of ratios and all nine known regulatory elements were detected. Seven out of nine known regulatory elements were detected when this assay was performed with 10^4 cells and a high mean CV of ratios was calculated. The number of significantly enriched elements that were detected and not previously associated with known regulatory function was one for assays performed with 10^6 and 10^5 cells, while this increased to seven for the assay performed with 10^4 cells. Note KCYp , SILp and SCLp refer to the promoters of these genes.

The mean ratios and CVs of array elements spotted in triplicate were also calculated to determine if array elements were reporting reproducible ratios within individual assays. The reproducibility of the reported ratios between triplicate spots was high for assays performed with 10^5 to 10^7 cells as the mean CV of ratios were calculated to be between 6-8% (Table 5.1). These values were very similar to those previously described (Dhami 2005 thesis). However there was a significant increase in the mean CV of ratios for the experiment performed with 10^4 cells (22.4%).

Thus, it appears that as cell numbers are decreased in ChIP-chip assays, there was a concomitant increase in the detection of non-specific interactions, reduction in the detection of known regulatory sequences and decreased reproducibility in the ratios obtained. Factors which may attribute to these observations include (i) limitations in the amount of recovered DNA available for labeling and hybridization resulting in lower signal:noise ratio on the array and (ii) an excess of antibody during the immunoprecipitation step resulted in enrichment of non-specific interactions. Therefore while a ChIP-chip assay performed with 10^4 cells could detect the majority of known regulatory elements associated with H3 acetylation at the SCL locus, it was necessary to further optimize the procedure to ensure that the assays were as equally robust as conventional ChIP assays performed with 10^7 cells.

5.5. The effect of titrating antibody levels on low cell number ChIP

5.5.1. Histone H3 K9/K14 acetylation experiments

As mentioned above, an excess of H3 K9/K14 acetylation antibody in ChIP-chip assays using the chromatin equivalent from 10^4 cells may be responsible for the enrichment of DNA sequences not previously identified in assays performed with 10^7 cells. Therefore amount of antibody used in ChIP was titrated in series of experiments to identify the optimal antibody:chromatin ratio for a range of cell numbers. In addition, these antibody titration experiments was also performed for an antibody raised to H3K4me3, to ensure that results obtained with the H3 acetylation assays could also be achieved with other ChIP-chip assays.

Once again, ChIP experiments were performed using a panel of conditions with varying amounts of chromatin from K562 cells (10^7, 10^6, 10^5, and 10^4 cell equivalents). However, in these experiments, the amount of antibody was varied (10 μg, 5 μg, 2 μg, 1 μg, 0.5 μg, and 0.1 μg) across each constituent of the chromatin panel. In otherwords, for a given chromatin amount, 6 different amounts of antibody were tested in ChIP. This resulted in a set of results across a series of 24 conditions (4 chromatin amounts versus 6 antibody amounts). Figures 5.3 to 5.6 show the results of these ChIP-chip experiments for the detection of regulatory features across the SCL locus.

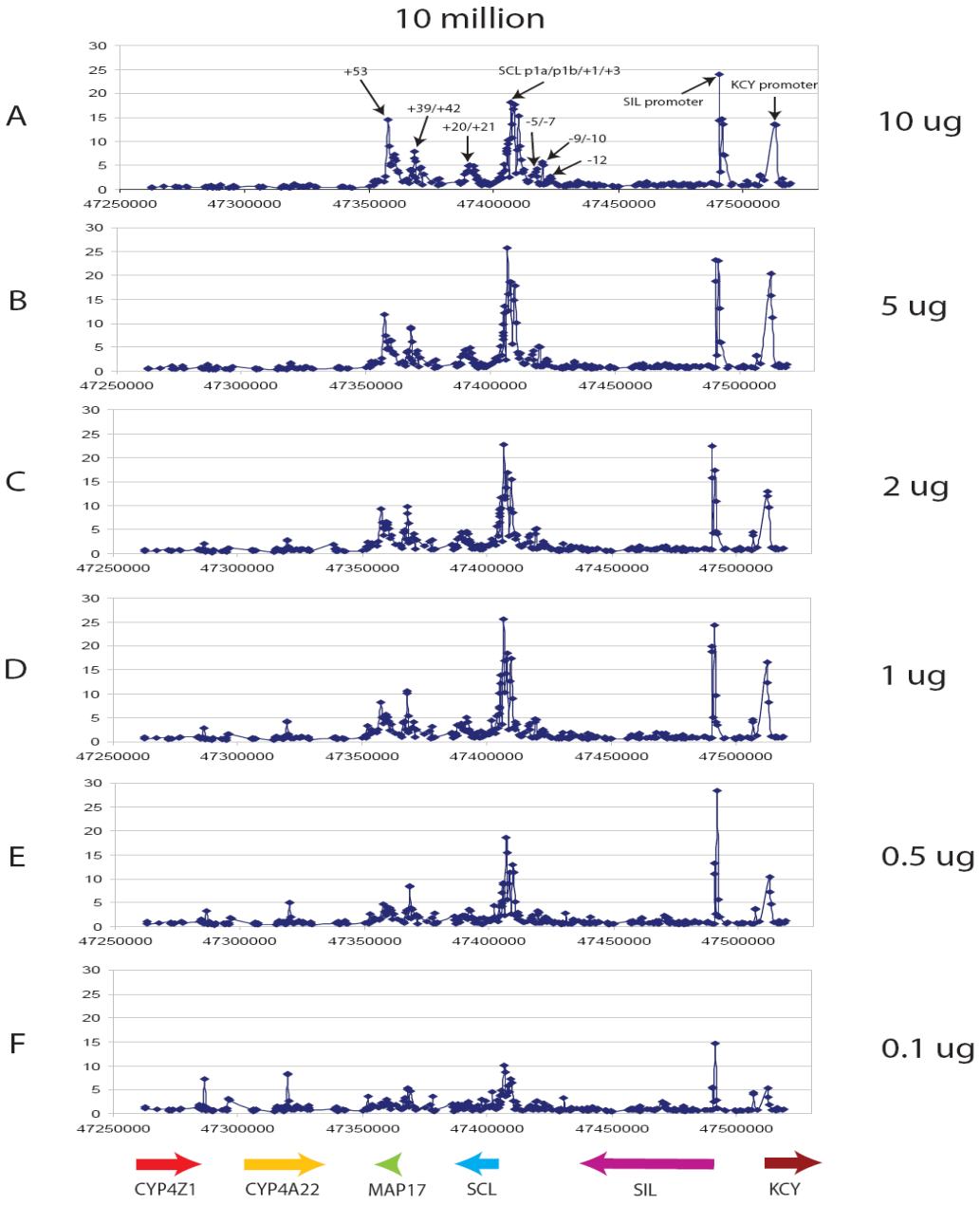

Figure 5.3: Titration of H3K9/K14 diacetylation antibody in ChIP assays performed with 10^7 K562 cells. The SCL ChIP-chip profiles for experiments performed with 10^7 K562 cells and the six antibody concentrations are shown. The known regulatory elements are indicated by black arrows in panel A. Variations in fold-enrichments are observed for regulatory elements across the six antibody concentrations in panels A-F. The detection of enriched regions not associated with known regulatory elements also varied with antibody concentration. Human chromosome 1 coordinates are indicated on the x-axis, with fold enrichments indicated on the y-axis. Each profile is scaled to 30 on the y-axis to allow for comparisons between titration series to be performed. The antibody amount is indicated to the right of each panel. The gene order and direction of transcription is indicated by thick coloured arrows at the bottom of the figure.

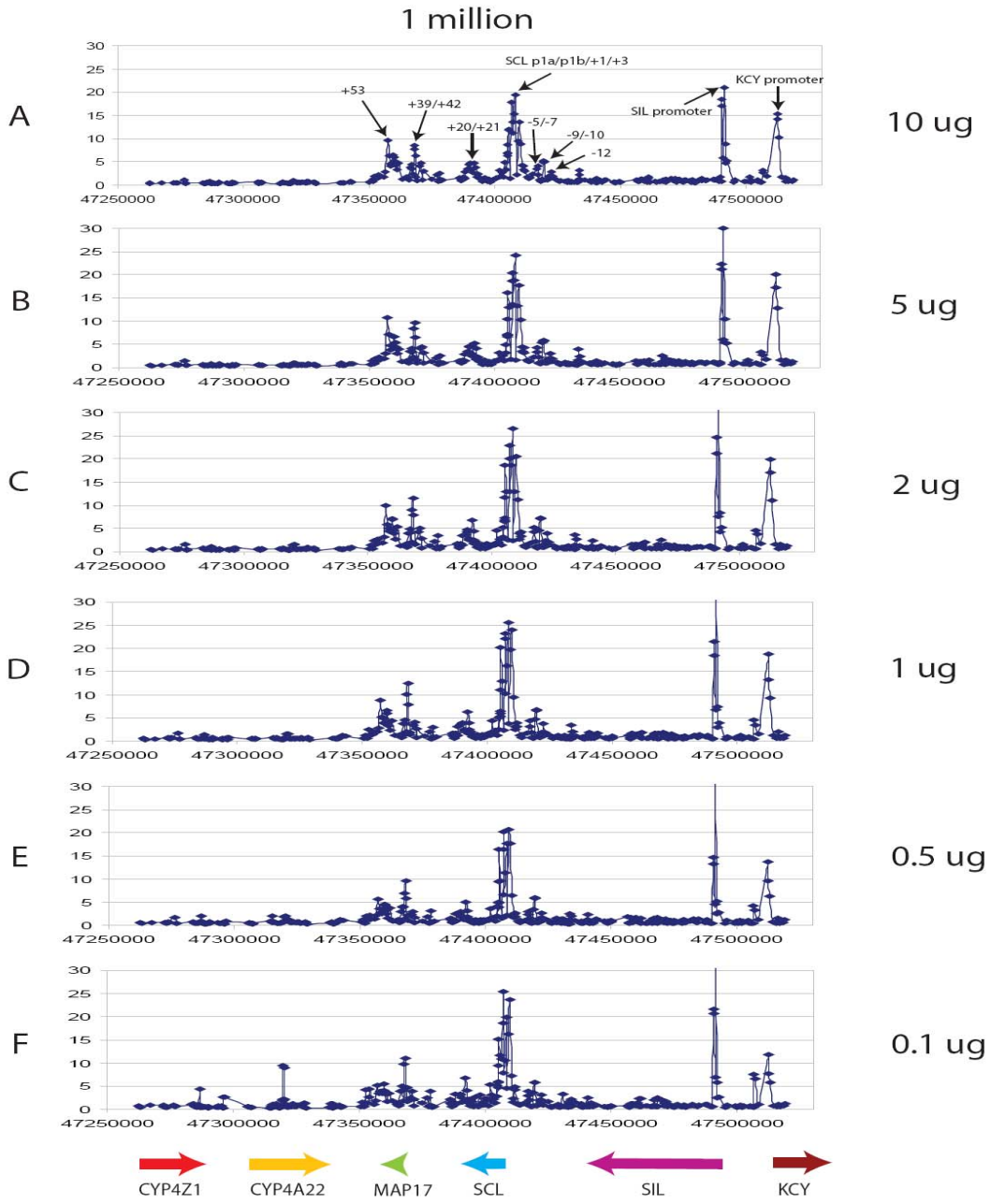

Figure 5.4: Titration of H3K9/K14 diacetylation antibody in ChIP assays performed with 10^6 K562 cells. The SCL ChIP-chip profiles for experiments performed with 10^6 K562 cells and the six antibody concentrations are shown. The known regulatory elements are indicated by black arrows in panel A. Variations in fold-enrichments are observed for regulatory elements across the six antibody concentrations in panels A-F. The detection of enriched regions not associated with known regulatory elements also varied with antibody concentration. Human chromosome 1 coordinates are indicated on the x-axis, with fold enrichments indicated on the y-axis. Each profile is scaled to 30 on the y-axis to allow for comparisons between titration series to be performed. The antibody amount is indicated to the right of each panel. The gene order and direction of transcription is indicated by thick coloured arrows at the bottom of the figure.

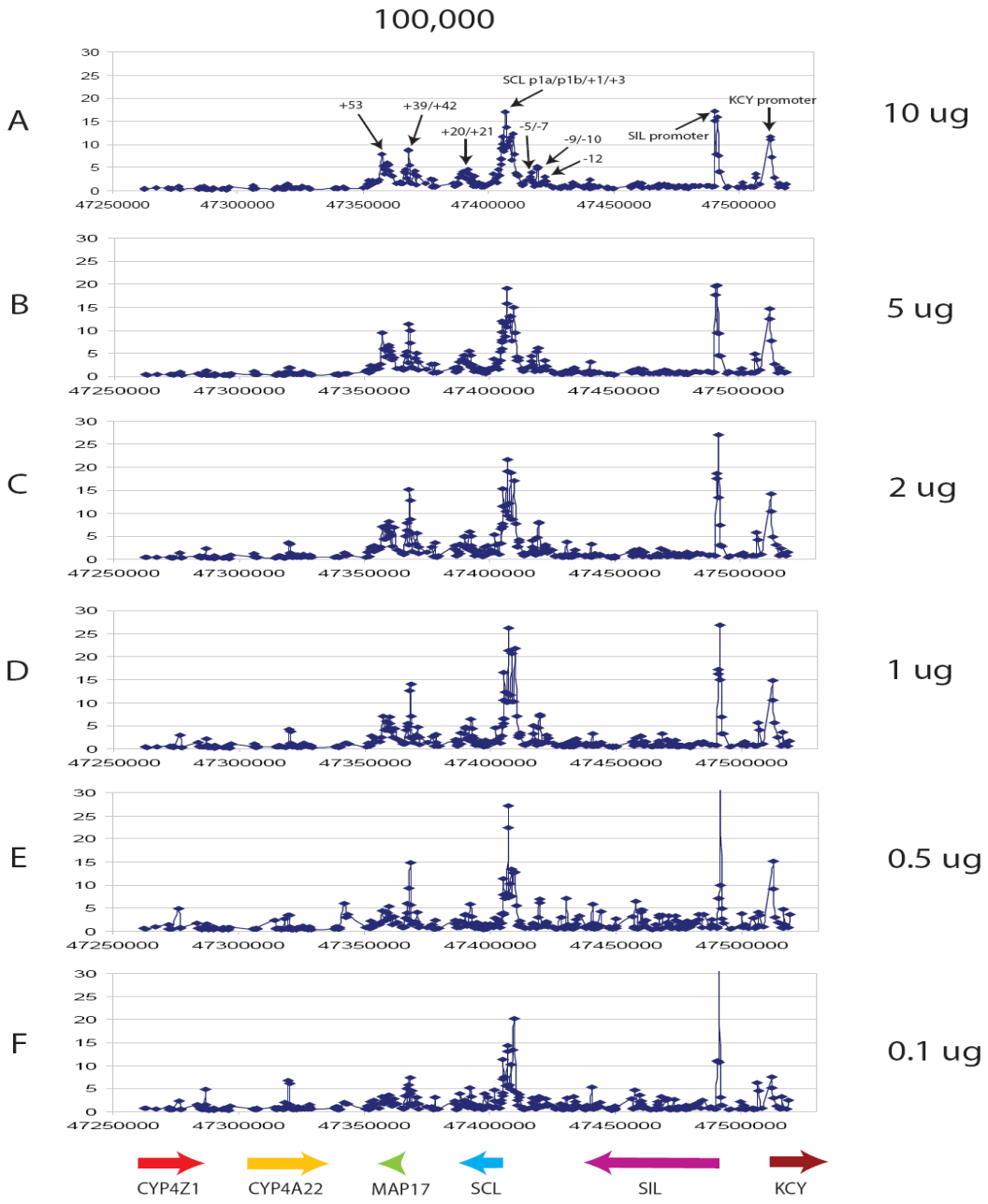

Figure 5.5: Titration of H3K9/K14 diacetylation antibody in ChIP assays performed with 10^5 K562 cells. The SCL ChIP-chip profiles for experiments performed with 10^5 K562 cells and the six antibody concentrations are shown. The known regulatory elements are indicated by black arrows in panel A. Variations in fold-enrichments are observed for regulatory elements across the six antibody concentrations in panels A-F. The detection of enriched regions not associated with known regulatory elements also varied with antibody concentration. Human chromosome 1 coordinates are indicated on the x-axis, with fold enrichments indicated on the y-axis. Each profile is scaled to 30 on the y-axis to allow for comparisons between titration series to be performed. The antibody amount is indicated to the right of each panel. The gene order and direction of transcription is indicated by thick coloured arrows at the bottom of the figure.

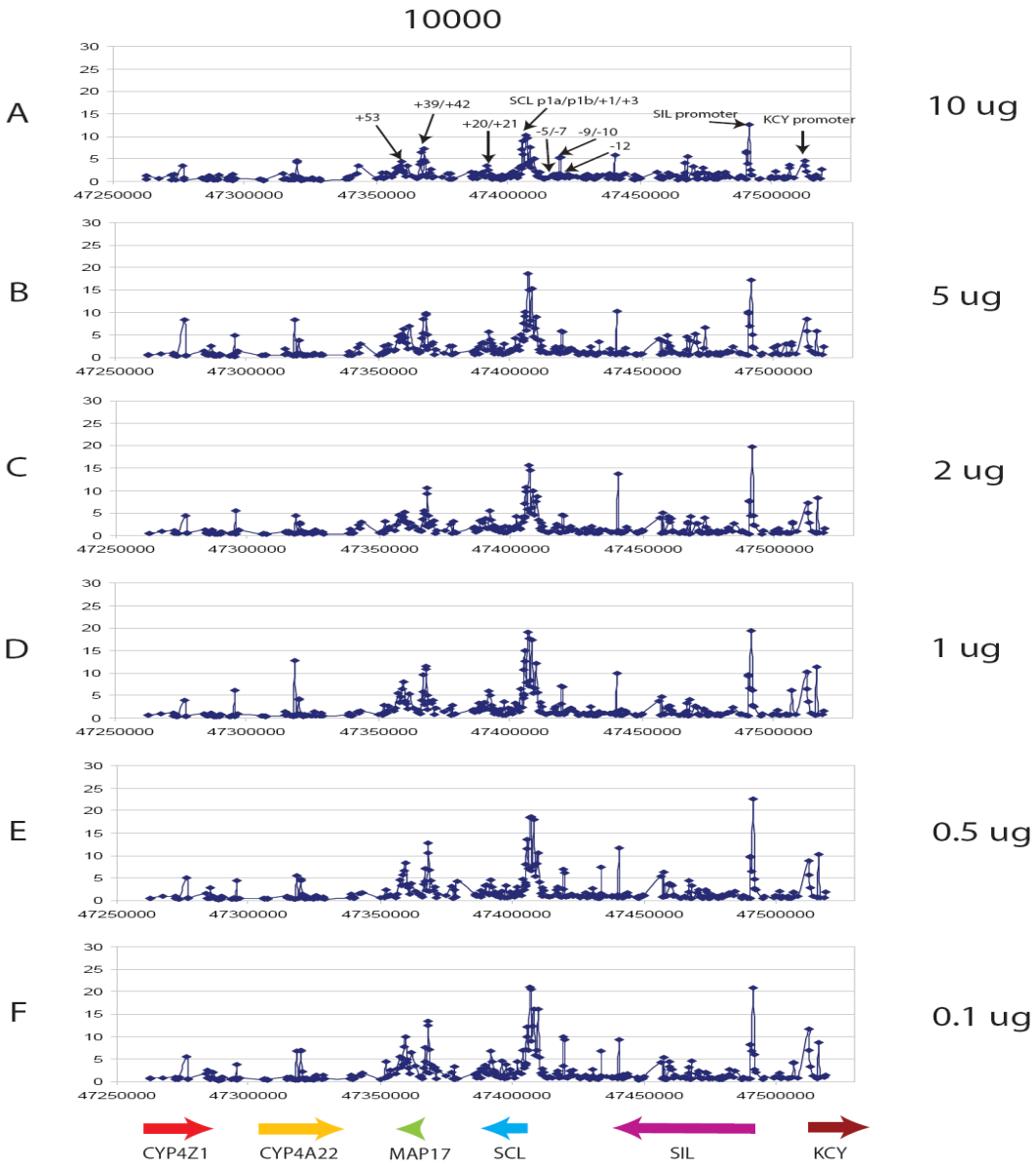

Figure 5.6: Titration of H3K9/K14 diacetylation antibody in ChIP assays performed with 10^4 K562 cells. The SCL ChIP-chip profiles for experiments performed with 10^4 K562 cells and the six antibody concentrations are shown. The known regulatory elements are indicated by black arrows in panel A. Variations in fold-enrichments are observed for regulatory elements across the six antibody concentrations in panels A-F. The detection of enriched regions not associated with known regulatory elements also varied with antibody concentration. Human chromosome 1 coordinates are indicated on the x-axis, with fold enrichments indicated on the y-axis. Each profile is scaled to 30 on the y-axis to allow for comparisonsbetween titration series to be performed. The antibody amount is indicated to the right of each panel. The gene order and direction of transcription is indicated by thick coloured arrows at the bottom of the figure.

To rigorously assess the results of these 24 experiments, the data were assessed as follows:

1. An optimal signal:noise ratio was determined for the titration series at each cell number. To do this, the data for 60 array elements on the SCL array which encompassed nine known regulatory elements detected in K562 cells were extracted and the average fold enrichment "signal" was calculated. The standard deviation of "background" regions represented on the SCL array (as defined in section 5.4) was determined. This represented a numerical value to describe 'noise' which would incorporate information obtained from non-specific enrichments. The average fold enrichment value was then divided by the background standard deviation value to obtain a "signal:noise" ratio (Table 5.2). This ratio would effectively determine how effective an experiment was at detecting *bona fide* regulatory elements across the SCL locus.
2. The total number of known regulatory regions at the SCL locus associated with significant enrichment by ChIP-chip was also assessed (Table 5.2). The significant enrichment threshold was set at three standard deviations above the mean background ratios as described previously.

Using these criteria, assays conducted with 10^7, 10^6 or 10^5 cells, showed the highest signal:noise ratios and detection of all known regulatory elements when using 5-10 μg of histone H3 K9/K14 antibody. However, as cell numbers were decreased to 10^4, the highest signal:noise was obtained when 0.1 μg of antibody was used; however, even at this antibody level, the signal:noise in these assays were more than 4-fold lower than those reported in assays with higher cell numbers, and only 7 of 9 known regulatory elements could be detected. These results demonstrate that antibody amount is crucial for increasing signal:noise in low cell number ChIP-chip experiments. However, it was still not possible to achieve the same sensitivity in low cell number assays, as that of conventional assays, for detecting all known regulatory elements across the SCL locus under the experimental conditions performed here.

Cell number	Antibody amount (μg)	Average fold enrichments (Signal)	Standard deviation of background regions (noise)	Signal:noise ratio	Significantly enriched known regulatory elements
10^7	10	7.82	0.28	27.08	KCYp, SILp, -12, -9/-10, -5/-7, SCLp, +20/+21, +39/+42, +53
	5	8.44	0.29	28.15	KCYp, SILp, -12, -9/-10, -5/-7, SCLp, +20/+21, +39/+42, +53
	2	7.35	0.36	20.26	KCYp, SILp, -12, -9/-10, -5/-7, SCLp, +20/+21, +39/+42, +53
	1	7.60	0.47	15.87	KCYp, SILp, -9/-10, -5/-7, SCLp, +20/+21, +39/+42, +53
	0.5	5.42	0.57	9.46	KCYp, SILp, -9/-10, SCLp, +20/+21, +39/+42, +53
	0.1	3.08	0.96	3.19	KCYp, SILp, SCLp, +39/+42,
10^6	10	7.40	0.35	21.08	KCYp, SILp, -12, -9/-10, -5/-7, SCLp, +20/+21, +39/+42, +53
	5	8.55	0.40	20.91	KCYp, SILp, -12, -9/-10, -5/-7, SCLp, +20/+21, +39/+42, +53
	2	8.98	0.43	20.50	KCYp, SILp, -12, -9/-10, -5/-7, SCLp, +20/+21, +39/+42, +53
	1	8.66	0.41	20.70	KCYp, SILp, -12, -9/-10, -5/-7, SCLp, +20/+21, +39/+42, +53
	0.5	6.59	0.41	15.77	KCYp, SILp, -12, -9/-10, -5/-7, SCLp, +20/+21, +39/+42, +53
	0.1	7.45	1.00	7.42	KCYp, SILp, -9/-10, SCLp, +20/+21, +39/+42, +53
10^5	10	6.81	0.33	20.32	KCYp, SILp, -12, -9/-10, -5/-7, SCLp, +20/+21, +39/+42, +53
	5	7.81	0.36	21.22	KCYp, SILp, -12, -9/-10, -5/-7, SCLp, +20/+21, +39/+42, +53
	2	8.21	0.57	14.40	KCYp, SILp, -12, -9/-10, -5/-7, SCLp, +20/+21, +39/+42, +53
	1	8.16	0.62	13.05	KCYp, SILp, -12, -9/-10, -5/-7, SCLp, +20/+21, +39/+42, +53
	0.5	6.30	1.22	5.15	KCYp, SILp, -9/-10, SCLp, +20/+21, +39/+42, +53
	0.1	5.17	0.94	5.45	KCYp, SILp, SCLp, +20/+21, +39/+42,
10^4	10	3.61	1.02	3.51	KCYp, SILp, -9/-10, SCLp, +20/+21, +39/+42, +53
	5	5.51	1.38	3.97	KCYp, SILp, -9/-10, SCLp, +20/+21, +39/+42, +53
	2	5.03	1.30	3.86	KCYp, SILp, -9/-10, SCLp, +20/+21, +39/+42, +53
	1	6.29	1.33	4.71	KCYp, SILp, -9/-10, SCLp, +20/+21, +39/+42, +53
	0.5	5.92	1.34	4.40	KCYp, SILp, -9/-10, SCLp, +20/+21, +39/+42, +53
	0.1	6.54	1.31	4.98	KCYp, SILp, -9/-10, SCLp, +20/+21, +39/+42, +53

Table 5.2: Statistical assessment of assay performance for the six antibody concentrations used in ChIP-chip assays performed with 10^7, 10^6, 10^5, and 10^4 cells and H3 K9/K14 diacetylation antibody. The average fold enrichments at known regulatory elements (signal) were calculated along with the standard deviation of background regions ('noise'). A signal:noise ratio for each antibody and cell amount was determined and the optimal ratio is indicated in red. Known regulatory regions associated with significant enrichments are also indicated for each assay in the titration series.

5.5.2. Histone H3 K4 trimethylation experiments

In the previous section the optimal amount of antibody was determined for H3 K9/K14 diacetylation assays performed with a range of cell numbers. However, to ensure that the results obtained for H3 K9/K14 acteylation titration experiments were not related to the efficiency of only the H3 acetylation antibody, ChIP-chip titration experiments with an antibody raised to H3K4me3 were also performed under the same conditions as for H3 K9/K19 acetylation (Figures 5.7 to 5.10). The data was assessed using the same criteria as discussed above (Table 5.3).

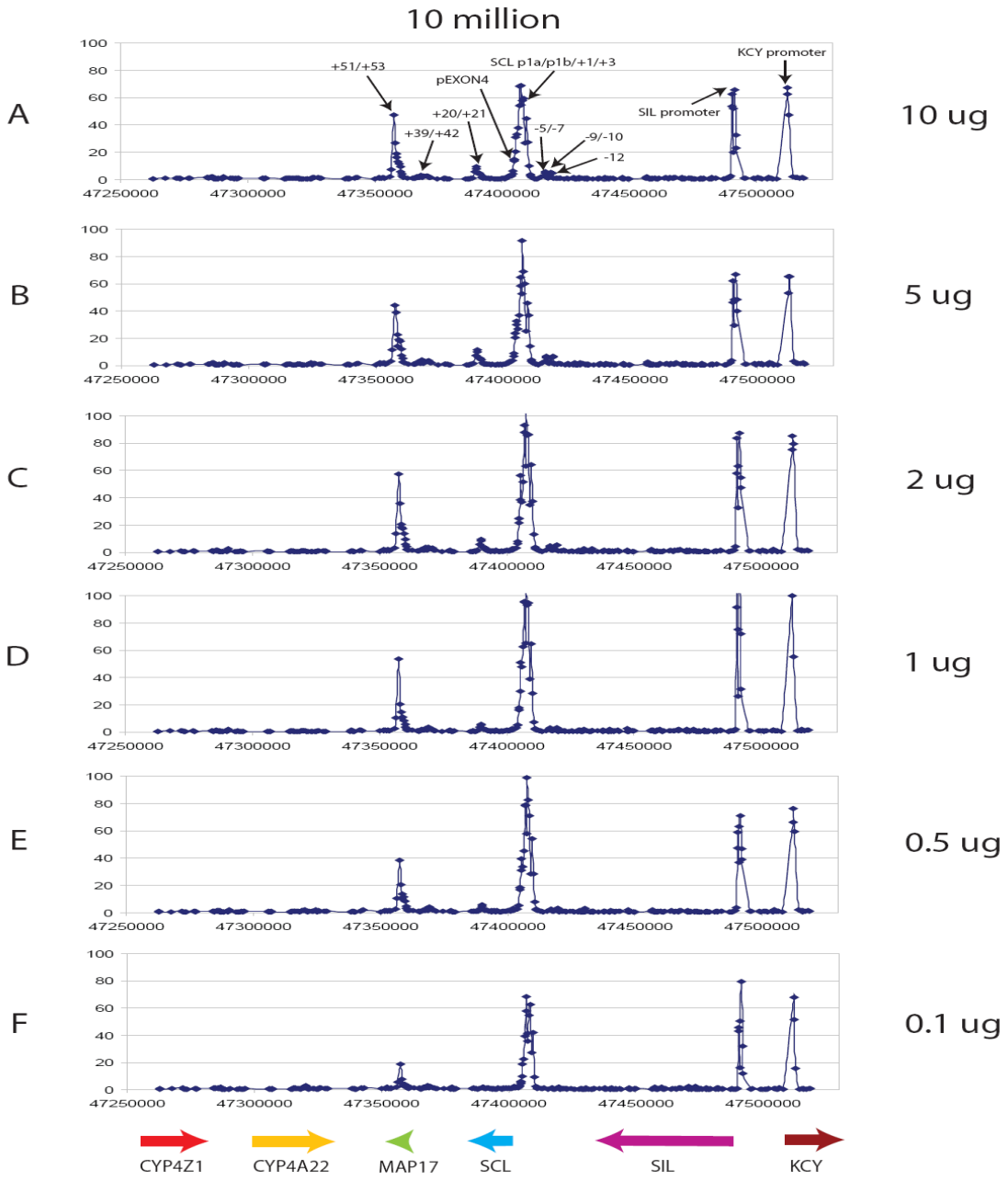

Figure 5.7: Titration of H3K4me3 antibody in ChIP assays performed with 10^7 K562 cells. The SCL ChIP-chip profiles for experiments performed with 10^7 K562 cells and the six antibody concentrations are shown. The known regulatory elements are indicated by black arrows in panel A. Variations in fold-enrichments are observed for regulatory elements across the six antibody concentrations in panels A-F. The detection of enriched regions not associated with known regulatory elements also varied with antibody concentration. Human chromosome 1 coordinates are indicated on the x-axis, with fold enrichments indicated on the y-axis. Each profile is scaled to 100 on the y-axis to allow for comparisons between titration series to be performed. The antibody amount is indicated to the right of each panel. The gene order and direction of transcription is indicated by thick coloured arrows at the bottom of the figure.

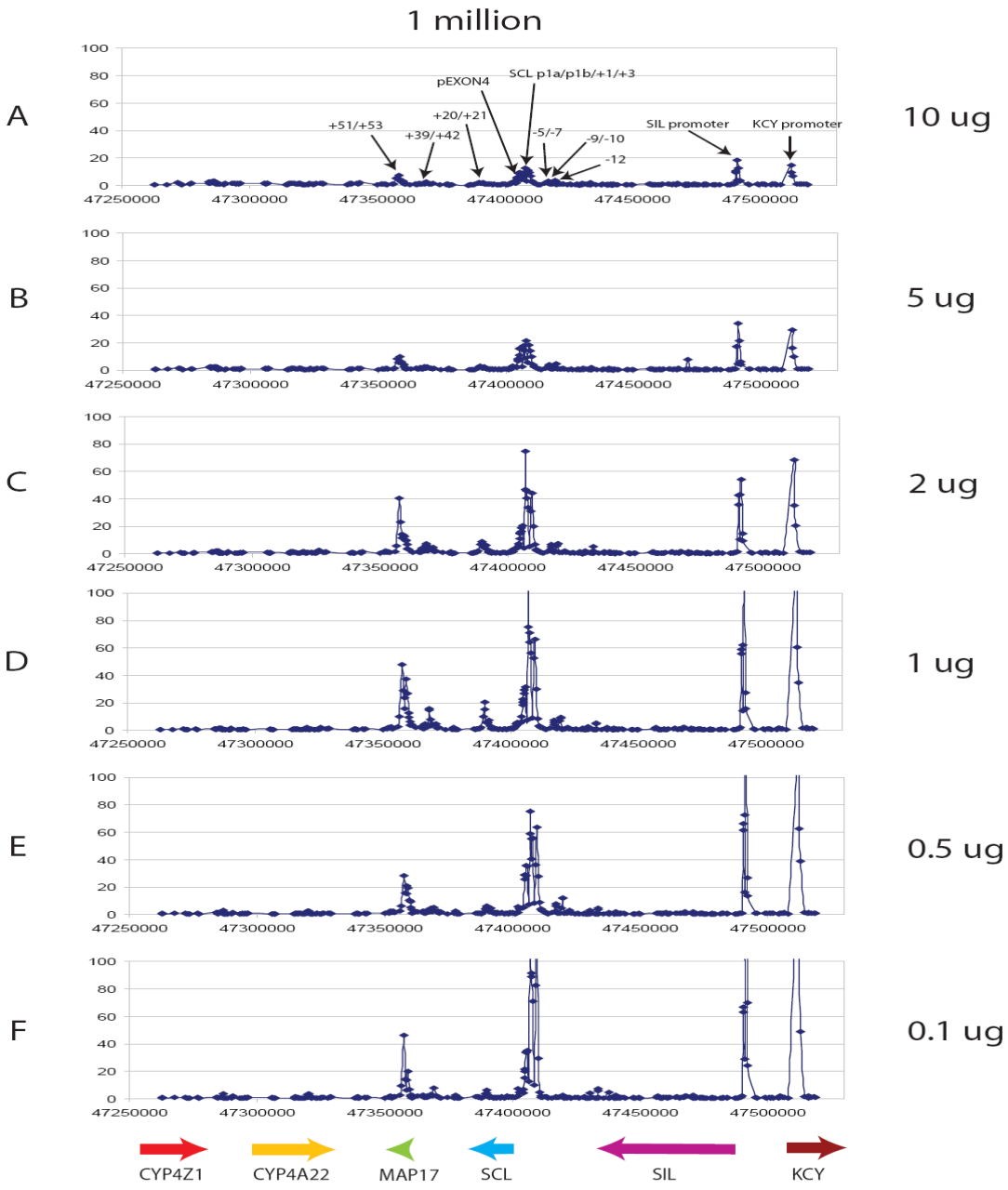

Figure 5.8: Titration of H3K4me3 antibody in ChIP assays performed with 10^6 K562 cells. The SCL ChIP-chip profiles for experiments performed with 10^6 K562 cells and the six antibody concentrations are shown. The known regulatory elements are indicated by black arrows in panel A. Variations in fold-enrichments are observed for regulatory elements across the six antibody concentrations in panels A-F. The detection of enriched regions not associated with known regulatory elements also varied with antibody concentration. Human chromosome 1 coordinates are indicated on the x-axis, with fold enrichments indicated on the y-axis. Each profile is scaled to 100 on the y-axis to allow for comparisons between titration series to be performed. The antibody amount is indicated to the right of each panel. The gene order and direction of transcription is indicated by thick coloured arrows at the bottom of the figure.

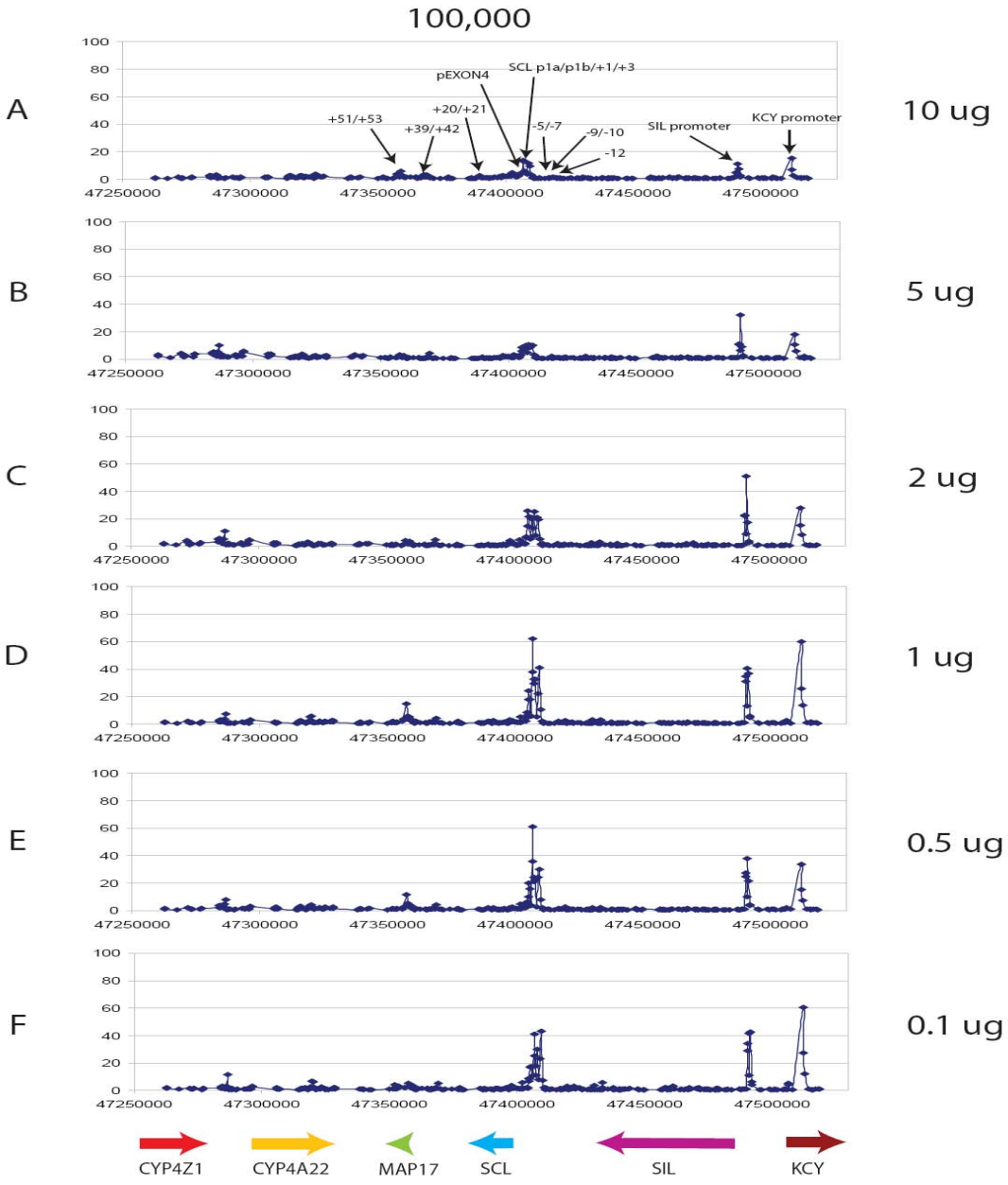

Figure 5.9: Titration of H3K4me3 antibody in ChIP assays performed with 10^5 K562 cells. The SCL ChIP-chip profiles for experiments performed with 10^5 K562 cells and the six antibody concentrations are shown. The known regulatory elements are indicated by black arrows in panel A. Variations in fold-enrichments are observed for regulatory elements across the six antibody concentrations in panels A-F. The detection of enriched regions not associated with known regulatory elements also varied with antibody concentration. Human chromosome 1 coordinates are indicated on the x-axis, with fold enrichments indicated on the y-axis. Each profile is scaled to 100 on the y-axis to allow for comparisons between titration series to be performed. The antibody amount is indicated to the right of each panel. The gene order and direction of transcription is indicated by thick coloured arrows at the bottom of the figure.

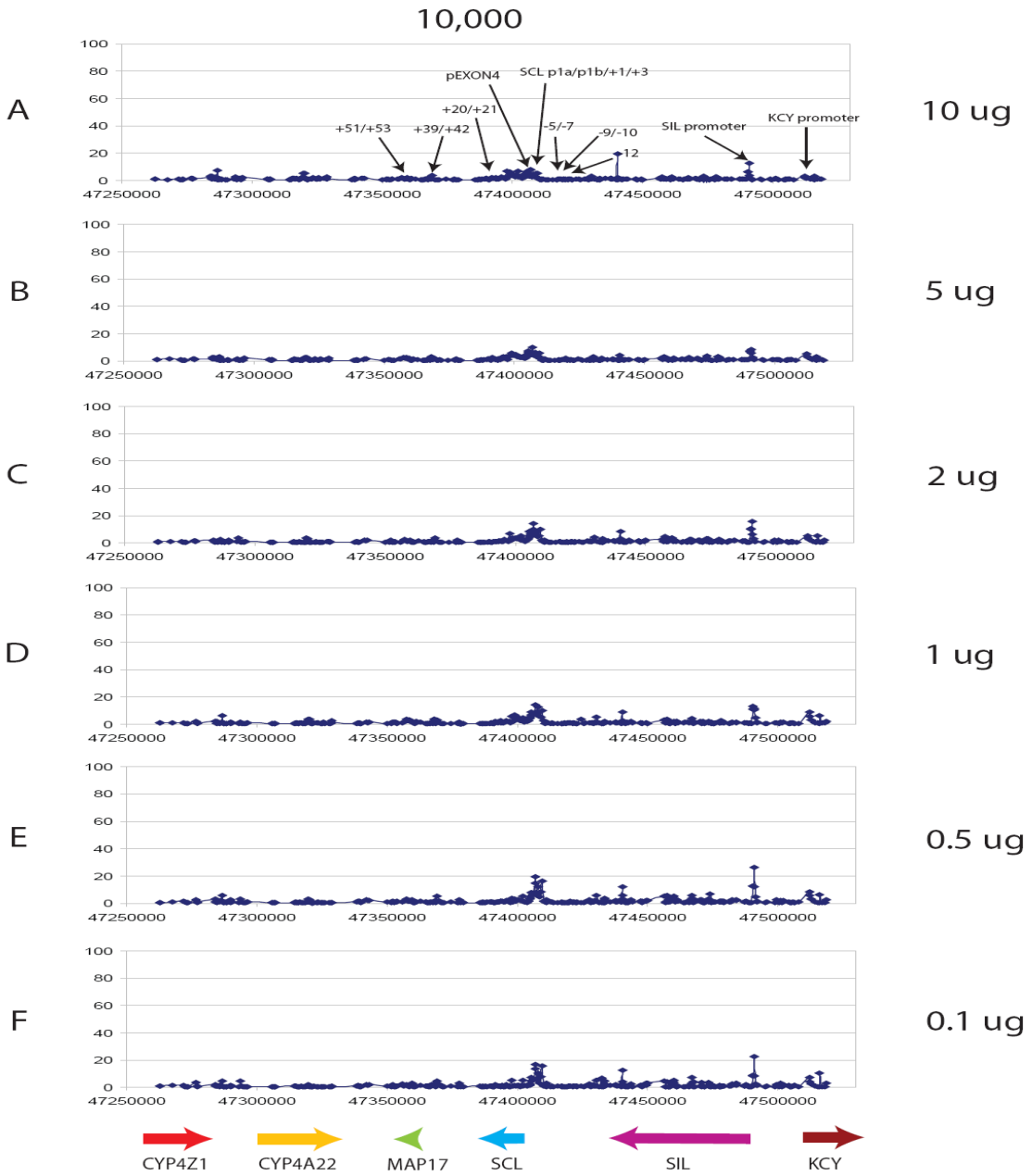

Figure 5.10: Titration of H3K4me3 antibody in ChIP assays performed with 10^4 K562 cells. The SCL ChIP-chip profiles for experiments performed with 10^4 K562 cells and the six antibody concentrations are shown. The known regulatory elements are indicated by black arrows in panel A. Variations in fold-enrichments are observed for regulatory elements across the six antibody concentrations in panels A-F. The detection of enriched regions not associated with known regulatory elements also varied with antibody concentration. Human chromosome 1 coordinates are indicated on the x-axis, with fold enrichments indicated on the y-axis. Each profile is scaled to 100 on the y-axis to allow for comparisons between titration series to be performed. The antibody amount is indicated to the right of each panel. The gene order and direction of transcription is indicated by thick coloured arrows at the bottom of the figure.

As was found with the histone H3 K9/K14 acetylation assays, there were several significant peaks of enrichment detected at regions not associated with known regulatory for assays performed with 10^5 and 10^4 cells at all antibody concentration levels. Yet, the overall effect of H3K4me3 antibody concentration on signal:noise across the panel of chromatin from 10^7, 10^6, 10^5 and 10^4 cell equivalents was markedly different than that obtained for the H3 K9/K14 acetylation. The signal:noise levels for experiments which used chromatin from 10^7, 10^6, and 10^5 cells was highest at antibody levels of 1 μg. However, signal:noise decreased more than 6-fold across this range, and resulted in fewer known SCL regulatory regions being detected at 10^5 cell levels. Furthermore, the highest signal:noise level for the H3K4me3 ChIP-chip assay with 10^4 cells was achieved at an antibody concentration of 5 μg – however, signal:noise was more than 20-fold reduced and only 4 known SCL regulatory elements were detected (when compared to the data obatined for H3K4me3 in conventional ChIP-chip assays with 10^7 cells). Thus, as cell numbers were reduced in this series of experiments, the most effective antibody level was higher than for assays with higher cell numbers – this was in marked contrast to the results achieved with the histone H3 K9/K14 titration series.

Cell number	Antibody amount (µg)	Average fold enrichments (Signal)	Standard deviation of background regions (noise)	Signal: noise ratio	Known regulatory elements significantly Enriched
10^7	10	23.14	0.36	63.84	KCYp, SILp, SCLp, pExon4, -12, -9/-10, -5/-7, +20/+21, +39/+42, +51/+53
	5	26.00	0.42	61.21	KCYp, SILp, SCLp, pExon4, -12, -9/-10, -5/-7, +20/+21, +39/+42, +51/+53
	2	31.53	0.40	78.64	KCYp, SILp, SCLp, pExon4, -12, -9/-10, -5/-7, +20/+21, +39/+42, +51/+53
	1	32.08	0.38	82.52	KCYp, SILp, SCLp, pExon4, -12, -9/-10, -5/-7, +20/+21, +39/+42, +51/+53
	0.5	25.73	0.38	66.56	KCYp, SILp, SCLp, pExon4, -9/-10, +20/+21, +39/+42, +51/+53
	0.1	16.99	0.41	41.28	KCYp, SILp, SCLp, pExon4, +51/+53
10^6	10	4.95	0.55	8.93	KCYp, SILp, SCLp, pExon4, -9/-10, -5/-7, +51/+53
	5	7.93	0.68	11.57	KCYp, SILp, SCLp, pExon4, -9/-10, -5/-7, +51/+53
	2	17.76	0.54	32.30	KCYp, SILp, SCLp, pExon4, -9/-10, -5/-7, +20/+21, +39/+42, +51/+53
	1	27.77	0.58	47.21	KCYp, SILp, SCLp, pExon4, -9/-10, -5/-7, +20/+21, +39/+42, +51/+53
	0.5	24.53	0.55	43.99	KCYp, SILp, SCLp, pExon4, -9/-10, -5/-7, +20/+21, +39/+42, +51/+53
	0.1	33.72	0.84	39.93	KCYp, SILp, SCLp, pExon4, +20/+21, +39/+42, +51/+53
10^5	10	3.59	0.59	5.99	KCYp, SILp, SCLp, pExon4, +51/+53
	5	4.40	1.19	3.68	KCYp, SILp, SCLp, pExon4,
	2	7.67	1.13	6.76	KCYp, SILp, SCLp, pExon4,
	1	11.91	0.88	13.48	KCYp, SILp, SCLp, pExon4, +51/+53
	0.5	9.28	0.88	10.46	KCYp, SILp, SCLp, pExon4, +51/+53
	0.1	10.29	1.15	8.89	KCYp, SILp, SCLp, pExon4
10^4	10	2.50	1.65	1.51	KCYp, SILp, SCLp, pExon4,
	5	2.97	0.66	4.50	KCYp, SILp, SCLp, pExon4,
	2	3.41	0.94	3.61	KCYp, SILp, SCLp, pExon4,
	1	4.06	1.01	4.02	KCYp, SILp, SCLp, pExon4,
	0.5	4.21	1.42	2.96	KCYp, SILp, SCLp, pExon4,
	0.1	3.65	1.47	2.48	KCYp, SILp, SCLp, pExon4,

Table 5.3: Statistical assessment of assay performance for the six antibody concentrations used in ChIP-chip assays performed with 10^7, 10^6, 10^5, and 10^4 cells and H3K4me3 antibody. The average fold enrichments at known regulatory elements (signal) were calculated along with the standard deviation of background regions ('noise'). A signal:noise ratio for each antibody and cell amount was determined and

the optimal ratio is indicated in red. Known regulatory regions associated with significant enrichments are also indicated for each assay in the titration series.

To more clearly describe the differences in the behaviour of the two histone modification assays used in the titration series described above, the relationship between antibody amount and signal:noise was plotted for each panel of cell amounts for both the H3K9/K14 and H3K3me3 assays (Figure 5.11). This figure showed that the general trend between antibody amount and signal:noise displayed different patterns for three of the four cell amounts (10^7, 10^6 and 10^5), whilst the pattern for experiments performed with 10^4 cells was more similar between the two histone modification assays. In general, signal:noise was greatest for lower H3K4me3 antibody amounts while the reverse was true for experiments performed with the H3 acetylation antibody. Yet, while optimal antibody levels differed between the two antibodies at the various cell equivalents, this analysis showed that it was possible to identify concentrations at which both antibodies performed well. Thus, titration of antibody and cell levels in ChIP-chip experiments are crucial issues that should be addressed when optimizing conditions for low cell numbers.

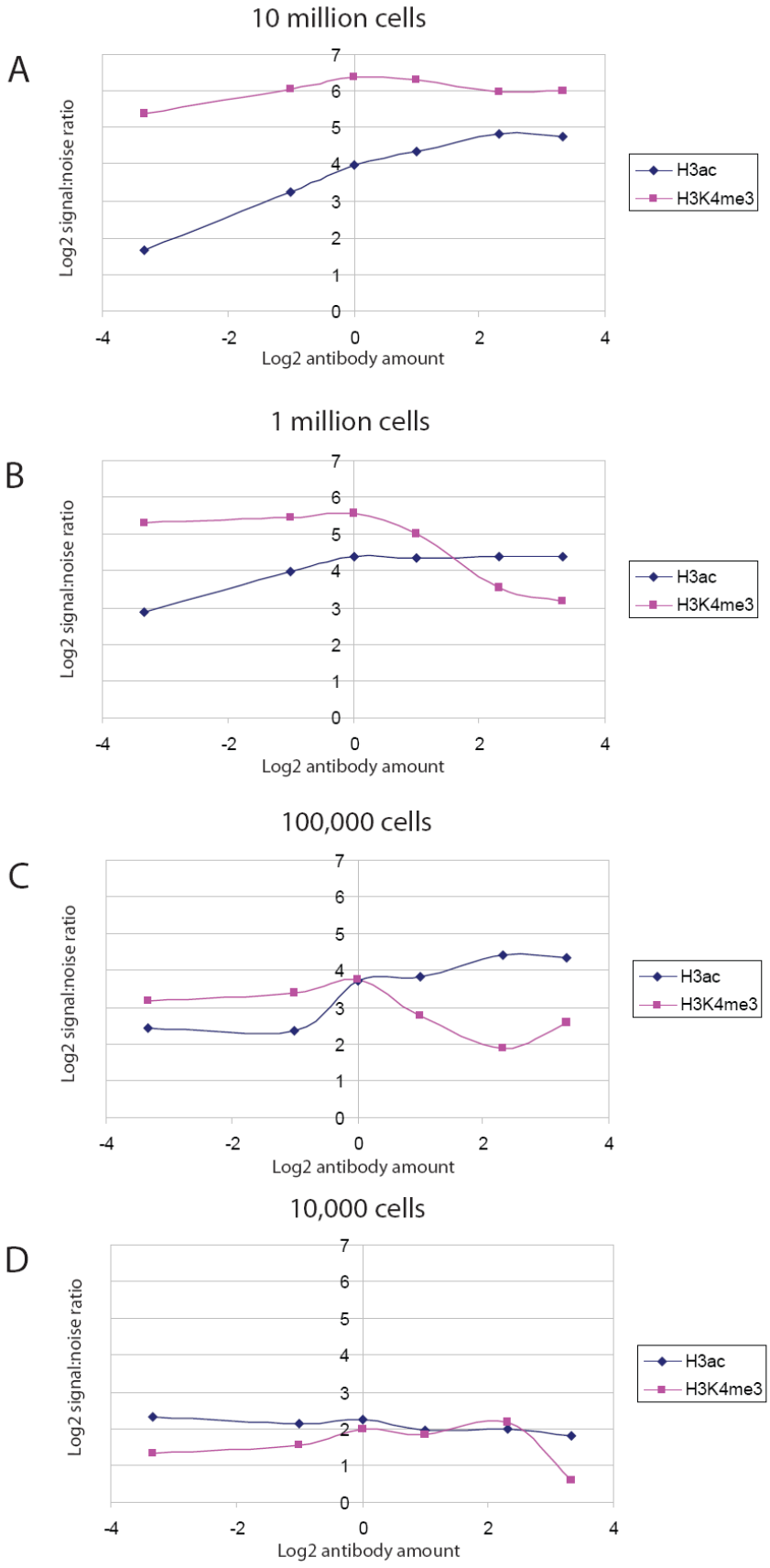

Figure 5.11: A comparison of the optimal antibody concentrations for titration assays performed with H3 acetylation and H3K4me3 antibodies. The $\log_2$ values of the six antibody amounts (displayed on the x-axes) were plotted against the $\log_2$ signal:noise ratios (displayed on the y-axes) for assays performed with H3 acetylation and H3K4me3 antibodies across the range of standard cell numbers. The plots in Panels A-D represent the values obtained for antibody titration assays performed with 10^7, 10^6, 10^5, and 10^4 cells respectively with H3ac plots indicated in blue while H3K4me3 plots are in pink.

5.6. Examining the effect of protein-G agarose concentration on ChIP-chip experiments performed with low cell numbers

It was previously reported that low cell number ChIP-chip identified a number of enrichments at regions of the SCL locus (in assays with 10^5 and 10^4 cell equivalents - see sections, 5.4 and 5.5) for which there was no evidence to support them as being *bona fide* regulatory elements [(such as high levels of non-coding DNA sequence conservation or annotation with features which would normally be associated with regulatory function – i.e., promoters, CpG islands). These regions were considered to be false-positives or regions that enriched non-specifically in ChIP-chip. The presence of these enrichments in assays for histone modifications using low cell numbers precipitated the need to optimize other conditions in the ChIP-chip procedure to remove or account for any technical artifacts of the method.

It was hypothesized that these "false" enrichments may be present in ChIP experiments due to an excess of protein-G agarose. Protein-G agarose contains recombinant protein-G covalently bound to agarose beads and is used in the immunoprecipitation of DNA-protein-antibody complexes as protein-G has a high affinity for IgGs. However, protein-G agarose can also bind non-specifically to DNA during the immunoprecipitation and some protocols use sheared herring or salmon sperm DNA in the immunoprecipitation step to block this non-specific binding (Weinmann *et al*., 2001). Therefore, the effect of reducing protein-G agarose was examined for the 10^5 cell experiment performed with 0.5 µg of H3 acetylation antibody which was associated with the highest standard deviation of background regions ('noise') for the 10^5 cell titration series (Table 5.2). 50 µl of protein-G agarose was ordinarily used in immunoprecipitation reactions so the effect of using less (25 µl and 10 µl) protein-G agarose was examined (Figure 5.12).

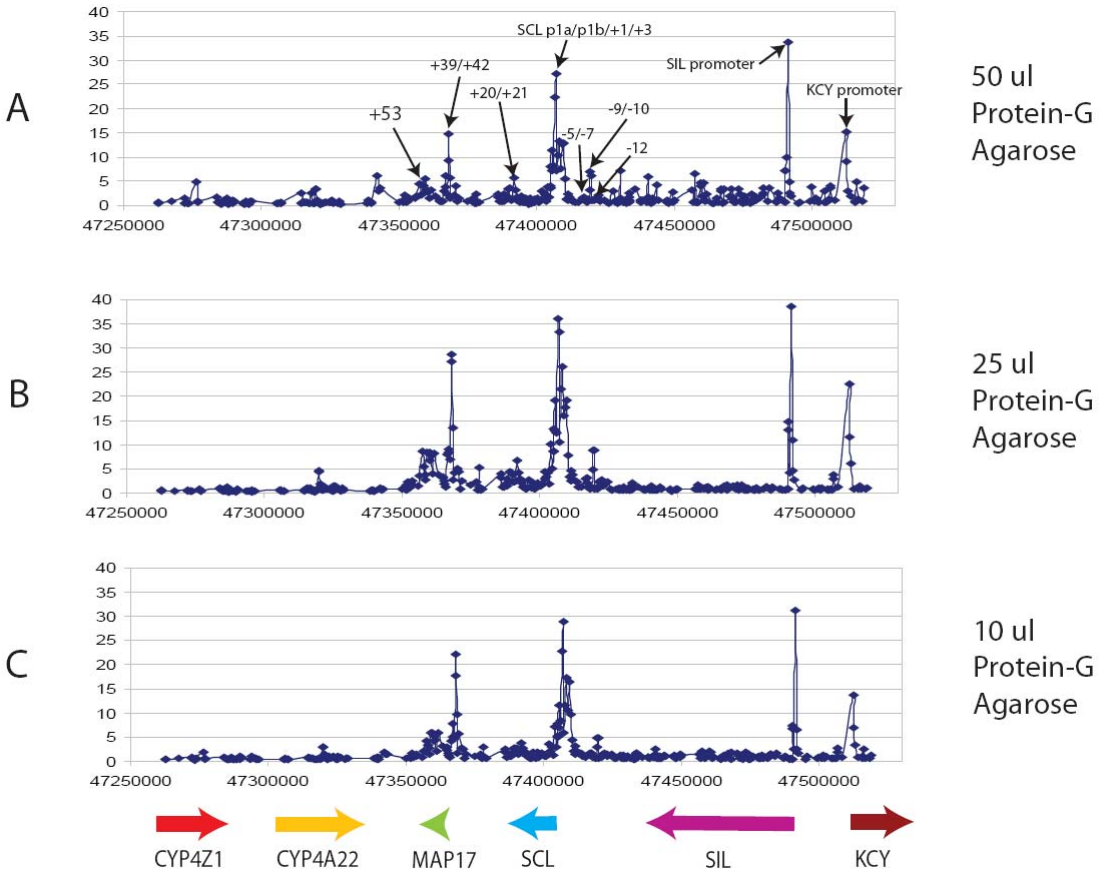

Figure 5.12: Titration of protein-G agarose in ChIP assays performed with 10^5 K562 cells and 0.5 µg of H3 acetylation antibody. The SCL ChIP-chip profiles for experiments performed with 10^5 K562 cells and 50 µl, 25 µl, and 10 µl of protein-G agarose are shown in panels A-C respectively. The known regulatory elements are indicated by black arrows in panel A. Variations in fold-enrichments are observed for regulatory elements across the three protein-G agarose concentrations. The detection of enriched regions not associated with known regulatory elements also varied with protein-G agarose concentration. Human chromosome 1 coordinates are indicated on the x-axis, with fold enrichments indicated on the y-axis. The gene order and direction of transcription is indicated by thick coloured arrows at the bottom of the figure.

It was clear from this targeted titration series that many of the regions associated with no known regulatory function were reduced to background levels when 25 µl or 10 µl of protein-G agarose was used in the immunoprecipitation step instead of 50 µl. This may be because the protein-G agarose was no longer in such excess resulting in less non-specific DNA being immunoprecipated. The effect of using less protein-G agarose was quantified in terms of average enrichment at known regulatory elements (signal) and standard

deviation of background regions (noise) as described before. The average signal was calculated to be 8.19 for the experiment performed with 25 µl of protein-G agarose and 7.02 for the experiment performed with 10 µl of protein-G agarose. Thus, the use of less protein-G agarose resulted in an increase in average signal when compared to the experiment performed with 50 µl of protein-G agarose (average signal of 6.30). Noise was also reduced in background regions when less protein G-agarose was used, as it was calculated to 0.64 and 0.68 for experiments performed with 25 µl and 10 µl of protein-G agarose respectively (compared to 1.22 for the 50 µl of protein-G agarose experiment). The signal:noise ratios were calculated to be 12.79 and 10.32 for the 25 µl and 10 µl experiments respectively. Thus it seems that 25 µl of protein-G agarose may be a more optimal protein-G agarose concentration for H3 acetylation assays performed with 10^5 cells for removing apparent "false" positive enrichments. However, it cannot be ruled out that these "false" enrichments do, in fact, mark the location of *bona fide* regulatory elements which are only identified in ChIP-chip assays with lower cell numbers. This issue is dealt with in more detail in section 5.9 of this Chapter.

5.7. Assessing reproducibility of the limited cell numbers ChIP-chip method

With the results of these experiments in mind, the reproducibility of performing ChIP-chip with low cell numbers was assessed. An additional biological replicate (biological replicate refers to a ChIP experiment that was conducted with an independent preparation of chromatin) was performed with the optimal histone H3 K9/K14 acetylation antibody concentration with 10^4 cells. Figure 5.13 illustrates the results obtained when comparing two independent biological replicate experiments.

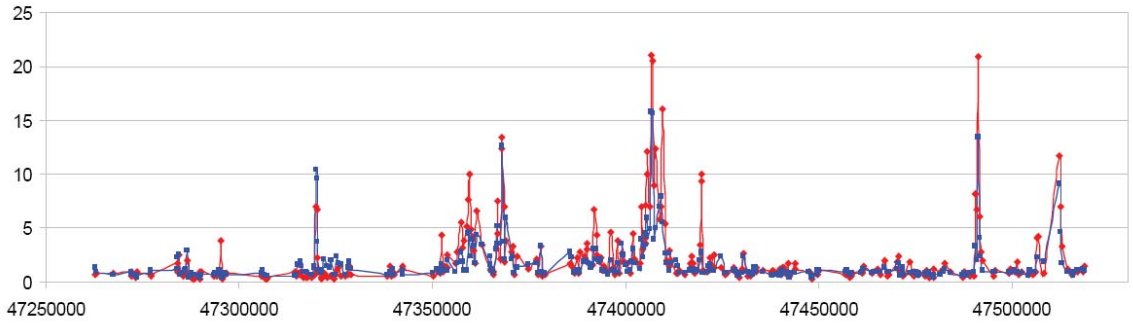

Figure 5.13: Biological replicates for ChIP-chip assays perfomed with 10^4 cells. Biological replicate experiments performed with 10^4 K562 cells and 0.1 μg of H3 acetylation antibody are presented. The red profile represents the data derived from biological replicate 1, while the blue profile represents the data from biological replicate 2. The human chromosome 1 coordinates are indicated along the x-axis and fold enrichments are indicated on the y-axis. The coloured arrows at the bottom of the figure represent the gene order and direction of transcription.

The reproducibility of performing ChIP-chip with limited cell numbers was assessed by calculating the coefficient of variation (CV) across the two biological replicates performed for each assay. This CV was then expressed as a percentage of the mean value for each tile in the SCL tile-path. The mean CV (%) across all tiles for the two H3 acetylation biological replicates was 25.7. While this is a large mean CV, a Pearson correlation coefficient of 0.89 was calculated when corresponding array elements were compared. Pearson's correlation reflects the degree of linear relationship between two variables. A correlation of +1 means that there is a perfect positive linear relationship between two variables, while a correlation of -1 means that there is a perfect negative linear relationship between two variables and a correlation of 0 means there is no linear relationship between the two variables. Therefore a correlation coefficient of 0.89 indicates that there is a high positive linear relationship between fold enrichments reported by the two biological replicates. This suggests that the reduced cell numbers ChIP method could be used to reproducibly identify known regulatory elements when as few as 10^4 cells are used per ChIP assay.

5.8. Histone H3K4me3 microarray data correlates with real-time quantitative PCR of ChIP material derived from 10^5 and 10^4 cells

In order to verify that enrichments observed in ChIP-chip assays perfomed with 10^5 and 10^4 cells were representative of DNA enrichments in the ChIP material, real-time quantitative PCR was performed (See Appendix 1 for sequences of primer pairs). ChIP DNAs obtained from assays performed with 10^5 and 10^4 cells with 1 μg of H3K4me3 antibody were used for this investigation. A summary of the results are shown in Figure 5.14. The fold enrichments obtained from microarray hybridizations performed with these samples are shown alongside those obtained with quantitative PCR for the 14 SCL

regions investigated. These regions included 7 known regulatory regions which enrich to various levels in ChIP-chip assays and 7 negative controls which show low/no enrichment in ChIP-chip assays. The array elements which showed significant H3K4me3 enrichments in microarray hybridization experiments with 10^5 and 10^4 cells also showed high levels of enrichment by quantitative PCR. In most cases, the array data and quantitative data were similar within comparable experiments. Array elements corresponding to the SCL+1 and +3 regions and the KCY+1 region reported some of the highest enrichments in ChIP-chip assays and showed the highest enrichments in quantitative PCR of 10^5 cell samples. The difference in fold enrichments reported by microarray hybridisation and quantitative PCR can be explained by the fact that quantitation of array elements which show high enrichment are no longer linear when pixel values reach saturation. This is not the case for quantitative PCR as quantitation is linear with DNA copy number in the ChIP sample. Negative control regions were associated with little or no enrichment in quantitative PCR assays performed with material derived from 10^5 and 10^4 cells.

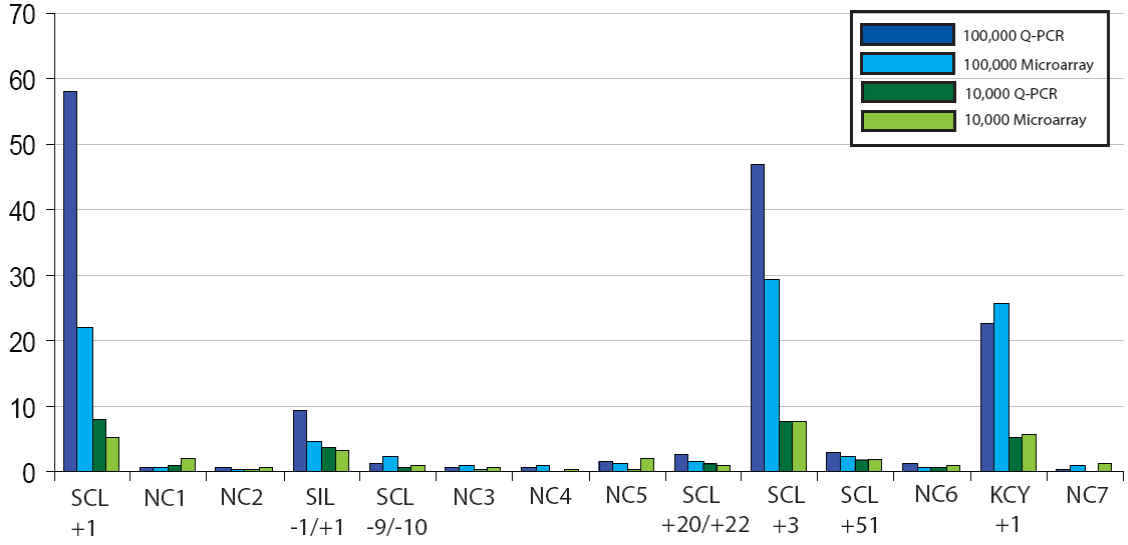

Figure 5.14: SYBR green real-time quantitative PCR of H3K4me3 ChIP material from reduced cell numbers. Enrichments reported by microarray hybridization with material derived from a ChIP performed with 10^5 K562 cells and 1 µg of H3K4me3 antibody (blue bar) and those reported by quantitative PCR (red bar) are shown side by side for each region. Similarly enrichments reported by microarray hybridization with material derved from a ChIP performed with 10^4 K562 cells and 1 µg of H3K4me3 antibody (yellow

bar) and those reported by quantitative PCR (green bar) are shown side by side for each region. The relevant genomic regions located across the SCL locus are indicated below the x-axis, and fold enrichments are indicated along the y-axis. NC = negative control.

5.9. Assessing the impact of normalising ChIP-chip data with normal rabbit IgG data

As discussed in section 5.5, a number of non-specific enrichments regions were identified as being significantly enriched in ChIP-chip assays performed with 10^5 and 10^4 cells. Whilst optimizing levels of protein-G agarose can reduce these "false" enrichments (see section 5.6), the use of mock antisera normalization was also investigated with the low cell number ChIP-chip data. It has previously been shown that using mock antisera to normalize histone modification and transcription factor data can effectively remove the effect of non-specific enrichments in conventional (10^6 cell) assays from ChIP-chip data (see Chapters 3 and 4).

In order to further investigate whether these peaks were (i) non novel regulatory elements only observed when ChIP-chip assays were performed with lower cell numbers, or (ii) non-specific interactions/artifact of performing the assays with fewer cells which could be normalised out of the final data sets, a series of 'mock' ChIP experiments were performed. As the H3 acetylation and H3K4me3 antibodies were raised in rabbit, mock ChIP experiments were performed with pre-immune antisera from rabbit (normal rabbit IgG). A series of titration experiments where performed with 10^5 and 10^4 K562 cells with 10 μg, 5 μg, 2 μg, 1 μg, 0.5 μg, and 0.1 μg of pre-immune antisera for the normalisation of corresponding H3 acetylation and H3K4me3 data sets. These experiments would simulate some of the titration conditions used to detect histone modifications in lower cell numbers assays. Thus, by matching up the data for the mock ChIP-chips with their respective experimental datasets, the effect of this type of normalization could be determined. The impact of normalising experimental ChIP-chip data with normal rabbit IgG data was assessed by plotting the $\log_2$ values for normal rabbit IgG ChIP data against $\log_2$ values for experimental ChIP data before (green plot) and after normalization (red plot) (Figure 5.15). Data points with low enrichment in the mock ChIP and a high experimental enrichment (known regulatory elements) are located in the top left corner of the scatter plot. Normalisation of these data points with the

corresponding normal rabbit IgG ChIP value increases the fold enrichment (see boxed data points in top left corner for an example of the data before and after normalisation). Data points with high enrichment in the normal rabbit IgG ChIP and also showing high experimental enrichments (non-specific enrichments) are located in the top right corner of the scatter plot. Normalisation of these data points with the normal rabbit IgG ChIP value decreases fold enrichment values to background levels (see boxed data points on right side of scatter plot for an example of this). Furthermore, no additional peaks of enrichment were introduced in the data by this method of normalisation.

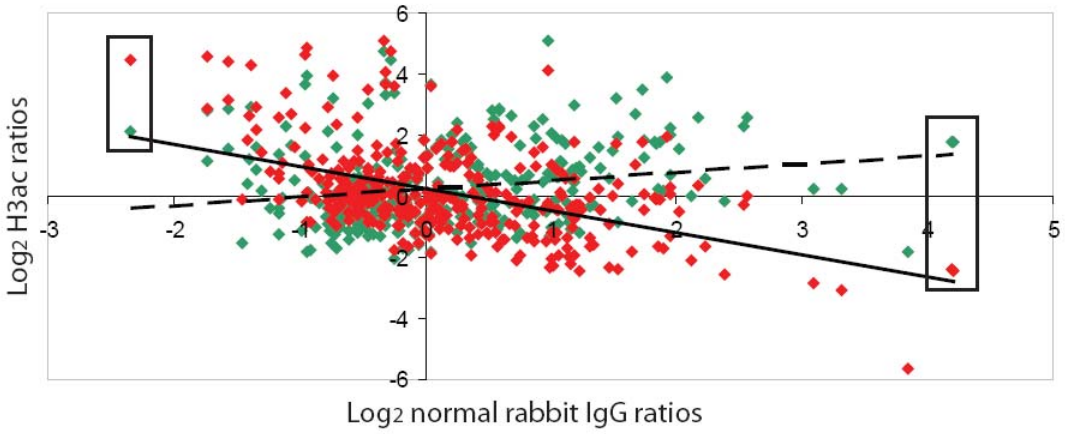

Figure 5.15: How normalisation with a normal rabbit IgG affects the spread of data for a ChIP-chip experiment performed with 10^5 cells. A scatter plot of Log_2 mock antibody ChIP chip values versus Log_2 experimental values before (green data points) and after normalisation (red data points) with normal rabbit IgG data illustrates the impact of normalisation on a data series. The data from a normal rabbit IgG mock ChIP performed with 10^5 cells and 0.5 µg of normal rabbit IgG was plotted against the values obtained from a ChIP-chip experiment performed with 10^5 cells and 0.5 µg of H3 acetylation antibody. Data points with low enrichment in the mock ChIP and a high experimental enrichment (known regulatory elements) are located in the top left corner of the scatter plot. Normalisation with the corresponding normal rabbit IgG ChIP value increases the fold enrichment for these data points (see boxed data points in top left corner for an example of the data before and after normalisation). Data points with high enrichment in the normal rabbit IgG ChIP and also showing high experimental enrichments (non-specific enrichments) are located in the top right corner of the scatter plot. Normalisation with the normal rabbit IgG ChIP value decreases the fold enrichment for these data points (see boxed data points on right side of scatter plot for an example of this). The slope of the trendline for the unnormalised data (dashed line) and the normalized data (solid

black line) indicate that those data points enriched in the H3ac data set and the normal rabbit IgG data set are reduced following normalisation.

The data from the H3 acetylation experiment performed with 10^5 cells and 0.5 µg of antibody was then plotted across the SCL region to examine the impact of normalisation on fold enrichments at known regulatory regions and those enrichments suspected to be non-specific interactions (Figure 5.16). The Figure shows that rabbit IgG normalization had the effect of increasing fold enrichments at the KCY promoter, SCL promoter and the +53 region, while fold-enrichments at other known regulatory elements such as the +39/+42 region, the +20/+21 stem cell enhancer and -9/-10 region were reduced following normalisation. These regions may seem to be more susceptible to non-specific enrichment in rabbit IgG but the reasons for this are not known. More importantly the effect of rabbit IgG normalisation was to reduce to background levels many of the regions associated with significant enrichment that were hypothesized to be non-specific interactions.

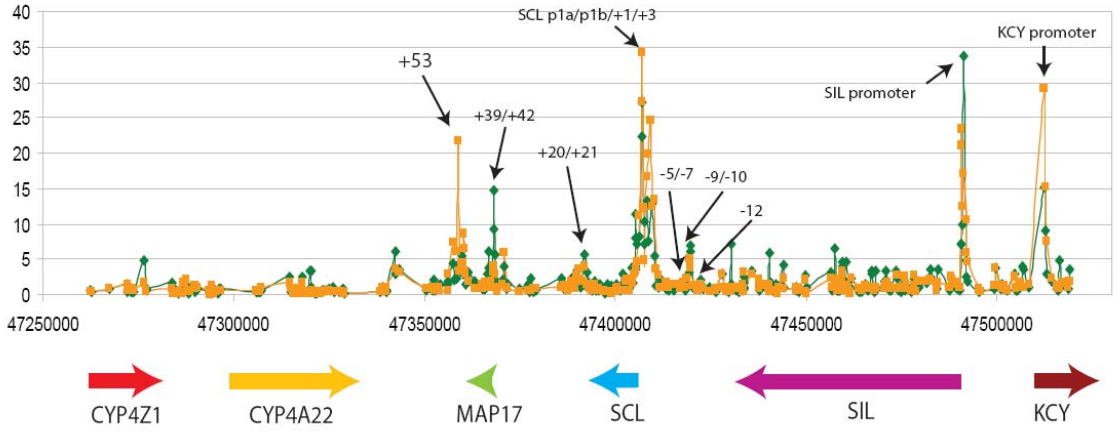

Figure 5.16: Rabbit IgG normalisation of a ChIP-chip experiment performed with 10^5 cells. The data from a ChIP-chip experiment performed with 10^5 cells and 0.5 µg of H3K4me3 is presented before and after rabbit IgG normalisation. The green profile represents the unnormalised data and the orange profile represents the rabbit IgG normalised data. Gene order and direction of transcription is indicated by thick coloured arrows at the bottom of the figure. Human chromosome 1 coordinates are indicated on the x-axis and fold enrichments are indicated on the y-axis.

The normal rabbit IgG titration series performed with 10^4 cells gave very weak signals in ChIP-chip experiments and it was not possible to quantitate spot enrichments to perform the appropriate normalization of data derived from experiments performed with 10^4 cells. However, the effect of using normal rabbit IgG data derived from experiments performed with 10^5 cells was examined for normalising H3 acetylation and H3K4me3 experiments performed with 10^4 cells. The effect of this normalisation on experiments performed with 10^4 cells and 0.1 μg of H3 acetylation antibody and 1 μg of H3K4me3 antibody (both of which gave the highest signal:noise ratios following rabbit IgG normalisation) is presented in Figure 5.17. The data is plotted for each antibody before and after normalisation with normal rabbit IgG data. The peaks of enrichment observed for those interactions not found at known regulatory elements in the unnormalised data sets are reduced to background levels following normalisation. In addition, fold enrichments observed at the majority of known regulatory elements were increased following rabbit IgG normalisation. The data plotted in panel A showed that the fold enrichments at the +53 region, the SCL promoter and the KCY promoter doubled following normalisation of the H3 acetylation data. However, enrichments at lower enriched regions such as the +39/+42 and -9/-10 regulatory elements were reduced following normalisation but remained above the significance threshold. Fold enrichments at the SIL promoter and the +20/+21 enhancer were unaffected by rabbit IgG normalisation. In panel B, fold enrichments associated with the SCL promoter, the KCY promoter and the +53 region were also increased following rabbit IgG normalization of the H3K4me3 data. Fold enrichments associated with the +39/+42 and +20/+21 regions were also reduced following normalisation, while enrichment at the -9/-10 region was found to be significant following normalisation.

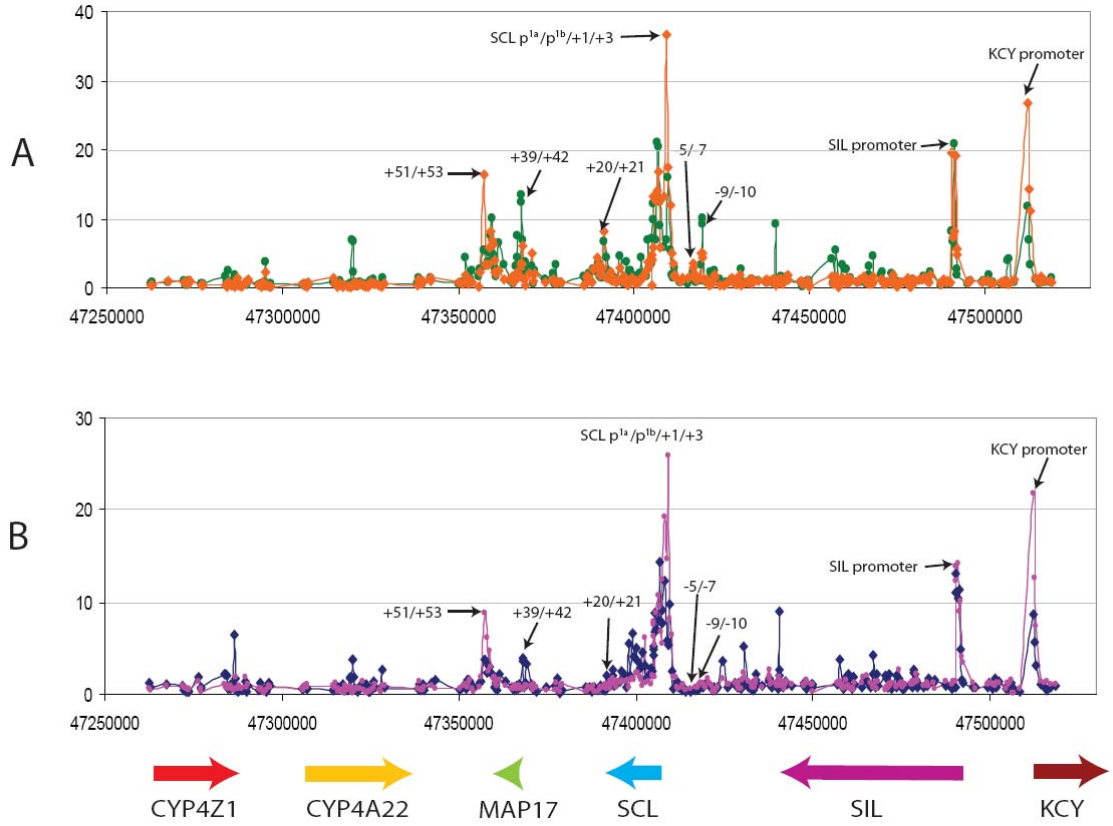

Figure 5.17: Rabbit IgG normalisation of data derived from ChIP-chip experiments performed with 10^4 cells. The data for two ChIP-chip experiments performed with 10^4 cells is presented before and after rabbit IgG normalisation. Panel A represents data from a ChIP-chip experiment performed with 10^4 K562 cells and 0.1 µg of H3 acetylation antibody. The green profile represents the unnormalised data and the orange profile represents the rabbit IgG normalized data. Panel B represents data from a ChIP-chip experiment performed with 10^4 K562 cells and 1 µg of H3K4me3 antibody. The blue profile represents the unnormalised data and the purple profile represents the rabbit IgG normalised data. Gene order and direction of transcription is indicated by thick coloured arrows at the bottom of the figure. Human chromosome 1 coordinates are indicated on the x-axis and fold enrichments are indicated on the y-axis.

In order to empirically define the effect of rabbit IgG on the low cell number data sets, the 24 data sets were analysed to determine the effect on average fold enrichment (signal) and standard deviation of background values (noise) as described previously. The data was then compared with unnormalised data and the findings are summarised in Table 5.4.

Cell number used in ChIP assay	Antibody	Antibody amount (µg)	Average Signal before and after normalisation		Standard deviation of background regions (noise) before and after normalisation		Signal/noise Before(left) and after Normalization (right)	
10^5	H3ac	10	6.81	8.97	0.33	0.46	20.32	19.27
		5	7.81	10.90	0.36	0.42	21.22	25.40
		2	8.21	9.05	0.57	0.37	14.40	24.15
		1	8.16	10.36	0.62	0.36	13.05	28.38
		0.5	6.30	8.00	1.22	0.54	5.15	14.62
		0.1	5.17	7.24	0.94	0.31	5.45	23.23
10^5	H3K4me3	10	3.59	4.65	0.59	0.55	5.99	8.33
		5	4.40	6.68	1.19	0.88	3.68	7.52
		2	7.67	9.77	1.13	0.68	6.76	14.17
		1	11.91	17.36	0.88	0.49	13.48	34.81
		0.5	9.28	13.69	0.88	0.64	10.46	21.26
		0.1	10.29	17.75	1.15	0.45	8.89	38.90
10^4	H3ac	10	3.61	4.24	1.02	0.42	3.51	9.92
		5	5.51	6.91	1.38	0.47	3.97	14.56
		2	5.03	5.37	1.30	0.64	3.86	8.39
		1	6.29	7.66	1.33	0.42	4.71	17.87
		0.5	5.92	7.68	1.34	0.41	4.40	18.50
		0.1	6.54	8.30	1.31	0.33	4.98	24.76
10^4	H3K4me3	10	2.50	2.96	1.65	0.92	1.51	3.20
		5	2.97	3.81	0.66	0.50	4.50	7.61
		2	3.41	3.86	0.94	0.51	3.61	7.45
		1	4.06	5.38	1.01	0.46	4.02	11.46
		0.5	4.21	5.96	1.42	0.97	2.96	6.14
		0.1	3.65	5.13	1.47	1.01	2.48	5.07

Table 5.4: Comparisons of 10^5 and 10^4 ChIP-chip data before and after rabbit IgG normalisation. The average fold enrichment (signal) values associated with known regulatory elements were calculated before and after rabbit IgG normalisation. The standard deviation of known background regions (noise) was also calculated before and after normalization. The optimal so-called signal:noise ratio is indicated in red for each titration series before and after normalisation.

As can be seen from this Table, rabbit IgG normalisation had a profound impact on the signal:noise for the 10^4 and 10^5 cells data series in two ways. Firstly, three of the four optimal antibody concentrations for maximum enrichments in each series were reduced following normalization compared to those concentrations obtained previously (see section 5.5). Taking normalization into account, the optimal concentration for ChIPs

performed with low cell numbers (10^5 or 10^4) is reduced somewhere in the range of 10-100 fold when compared to the optimal antibody concentration for conventional ChIPs performed with 10^7 cells. Second, the average signal for each of the 24 hybridisations increased following normalisation (by between 22-48% following normalisation), while standard deviations in the background regions were decreased in 22 of the 24 experiments (by between 36-64% following normalization). Overall, this resulted in an increase signal:noise in 23 of the 24 experiments (between 70%-270% following normalisation); the one remaining experiment (10^5 cells and 10 μg H3 acetylation antibody) showed a small reduction in the signal:noise ratio following normalisation which was due to a larger increase in the standard deviation of background regions relative to the increase in average signal.

Cell number	Histone modification	Mean series signal before and after normalisation		% increase/ decrease after normalisation	Mean series noise before and after normalisation		% increase/ decrease after normalisation	Mean series signal: noise before/after normalisation		% increase/ decrease after normalisation
10^5	H3ac	7.07	9.08	+28.40	0.67	0.41	-39.10	13.625	22.50	+69.68
10^5	H3K4me3	7.85	11.65	+48.28	0.97	0.615	-36.59	8.21	20.83	+153.73
10^4	H3ac	5.48	6.69	+22.06	1.28	0.44	-64.79	4.23	15.66	+269.64
10^4	H3K4me3	3.46	4.51	+30.28	1.19	0.72	-38.88	3.18	6.82	+114.51

Table 5.5: Assessing the impact of normalisation on the average titration series signal, average series standard deviation and average series signal:noise values for experiments performed with 10^5 and 10^4 cells. The mean titration series 'signal' for all six antibody concentrations was calculated for experiments performed with 10^4 and 10^5 cells with H3 acetylation (H3ac) and H3K4me3 antibodies . Mean signal values were calculated before and after normalisation. The mean titration series 'noise' was also calculated before and after normalization. This allowed for a mean series signal:noise ratio to be calculated before and after normalisation and the percentage increase in signal:noise following normalisation is presented in the final column.

5.10. Testing of other histone modification antibodies with the modified ChIP-chip method

Finally, to further explore the utility of the lower cell ChIP-chip method to examine other aspects of regulatory function associated with histone modifications. Lower cell number ChIP-chip was validated for three other histone methylation modifications (H3K27me3,

H3K36me3 and H3K79me3). Because of time limitations, comprehensive titration panels could not be performed for these assays. However, based on the results of this Chapter, optimal antibody and cell number conditions were used in these additional assays as those determined for the H3K4me3. The resulting datasets obtained with assays using 10^6 and 10^5 K562 cells were compared to experiments performed with 10^7 K562 cells. Figure 5.18 illustrates the application of the modified ChIP-chip method to detect regions of the SCL locus associated with H3K27me3 in K562 cells. This showed that the CYP4Z1 and CYP4A22 genes (which are not expressed in K562) are associated with elevated levels of H3K27me3 (i.e., H3K72me3 is a mark of repressed gene expression) in assays performed with 10^6 and 10^5 cells, albeit to slightly different levels as that obtained with conventional 10^7 cell assays. Therefore this assay could be used to reliably detect regions enriched for H3K27me3 from as few as 10^5 cells.

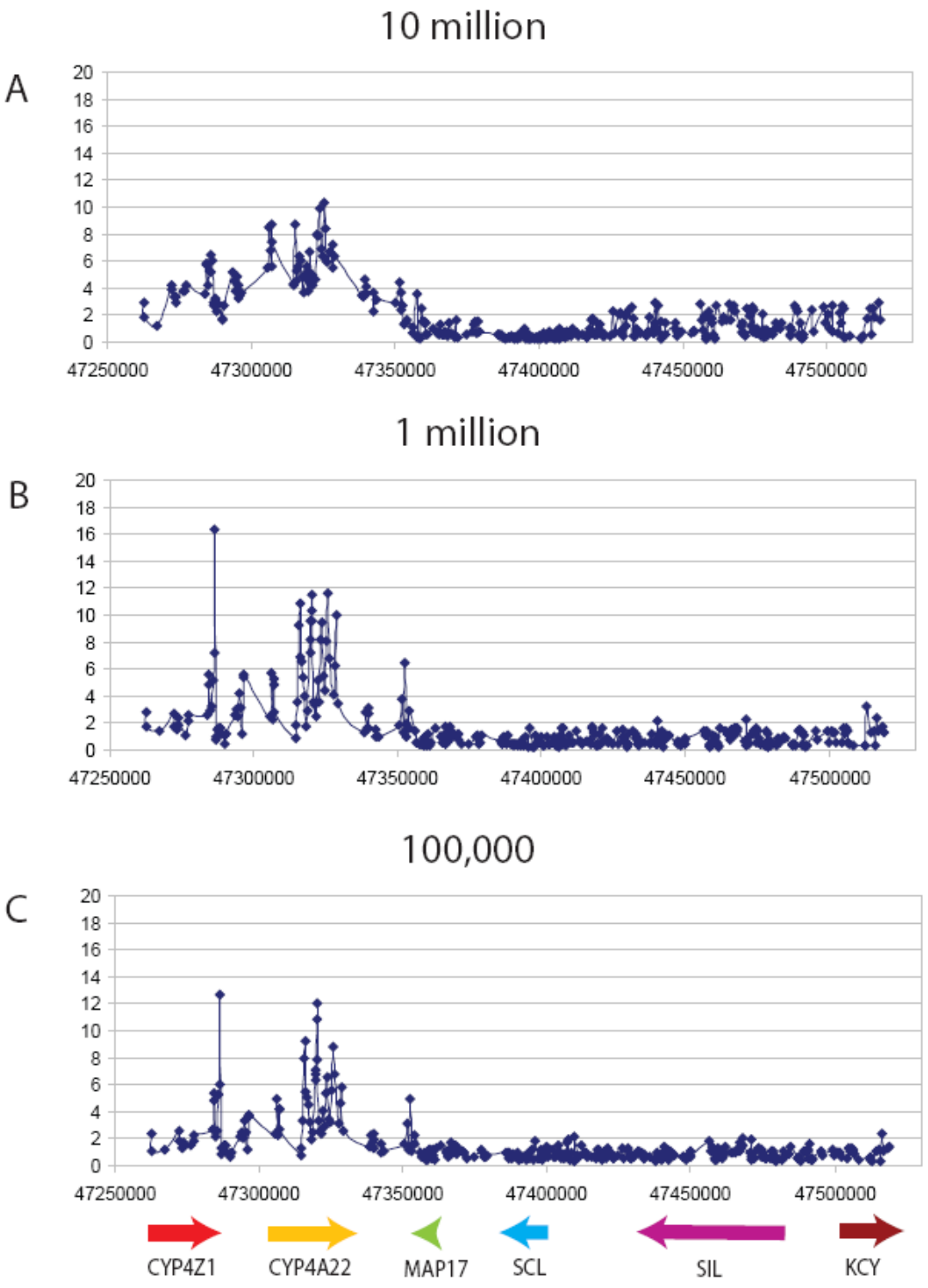

Figure 5.18: Testing the modified ChIP-chip method to detect regions associated with H3K27me3. ChIP-chip data obtained from a experiments performed with a range of K562 cell numbers and an antibody raised to H3K27me3. Panel A: ChIP-chip experiment performed with 10^7 cells and 10 µg H3K27me3.

Panel B: A ChIP-chip experiment performed with 10^6 cells and 1 µg H3K27me3, Panel C: 10^5 cells, 0.5 µg H3K27me3. Fold enrichment values are presented on the y-axis and the human chromosome 1 coordinates are indicated on the x-axis. The location of genes in the region and direction of transcription is indicated by coloured arrows.

An antibody raised to H3K36me3 was also tested with the modified ChIP-chip method. H3K36me3 is known to be associated with the transcribed portion of active genes (Bannister *et al*., 2005 b) and was found to be associated with the transcribed portion of the SCL and SIL genes (Figure 5.19) in the K562 cell line. Once again, this assay was found to perform similarly when 10^7, 10^6 or 10^5 cells were used.

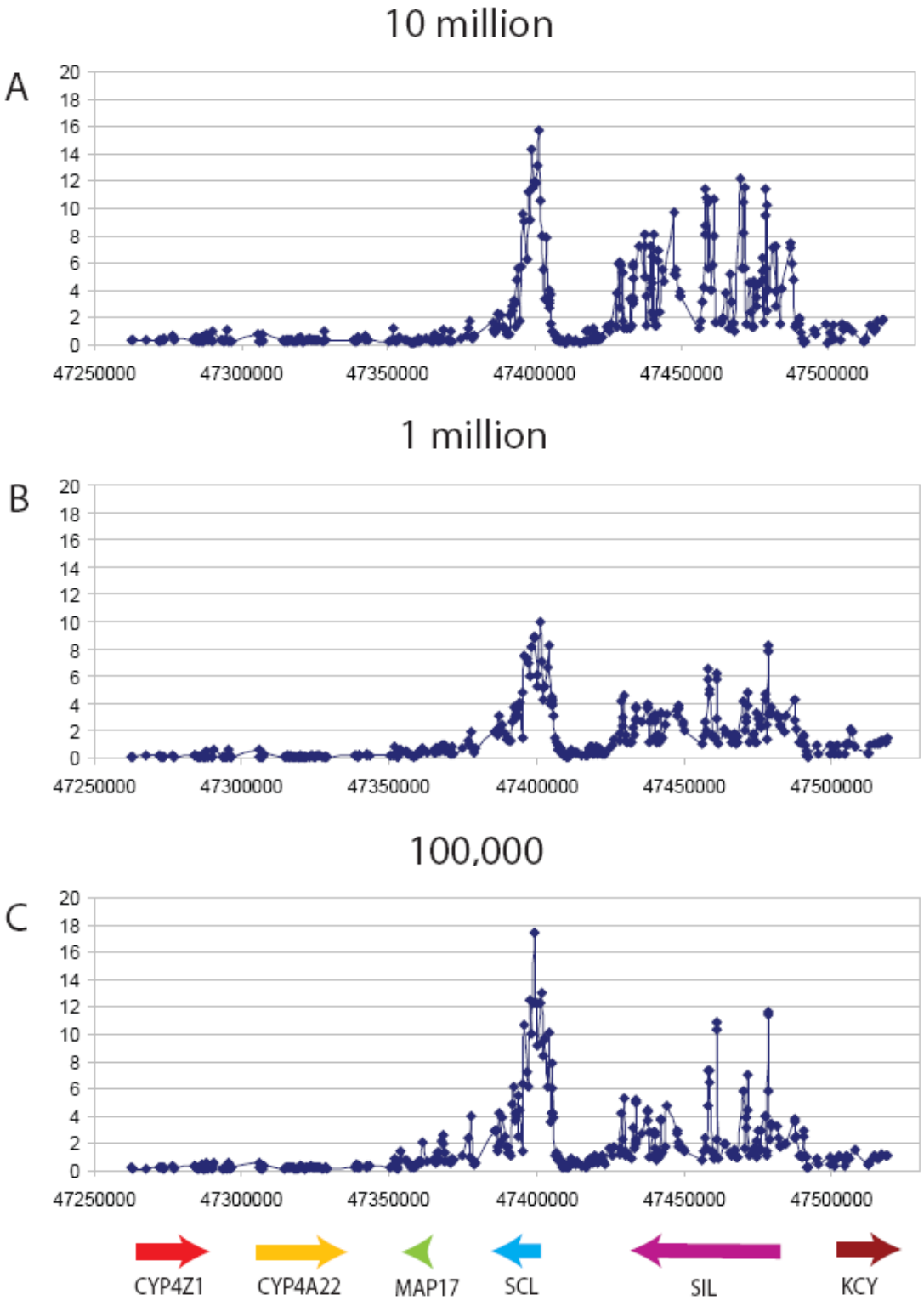

Figure 5.19: Testing the modified ChIP-chip method to detect regions associated with H3K36me3. ChIP-chip data obtained from a experiments performed with a range of K562 cell numbers and an antibody raised to H3K36me3. Panel A: ChIP-chip experiment performed with 10^7 cells and 10 µg H3K36me3.

Panel B: A ChIP-chip experiment performed with 10^6 cells and 1 µg H3K36me3, Panel C: 10^5 cells, 0.5 µg H3K36me3. Fold enrichment values are presented on the y-axis and the human chromosome 1 coordinates are indicated on the x-axis. The location of genes in the region and direction of transcription is indicated by coloured arrows.

Lastly, ChIP-chip assays for histone H3K79me3 (Figure 5.20) were performed. The role of this modification in mammalian cells is not well understood, but here it was found to associate with the transcribed portion of SCL, SIL and KCY genes immediately downstream of the TSS. All three of these genes are expressed in K562. This pattern of enrichment was again similar across the range of cell numbers studied.

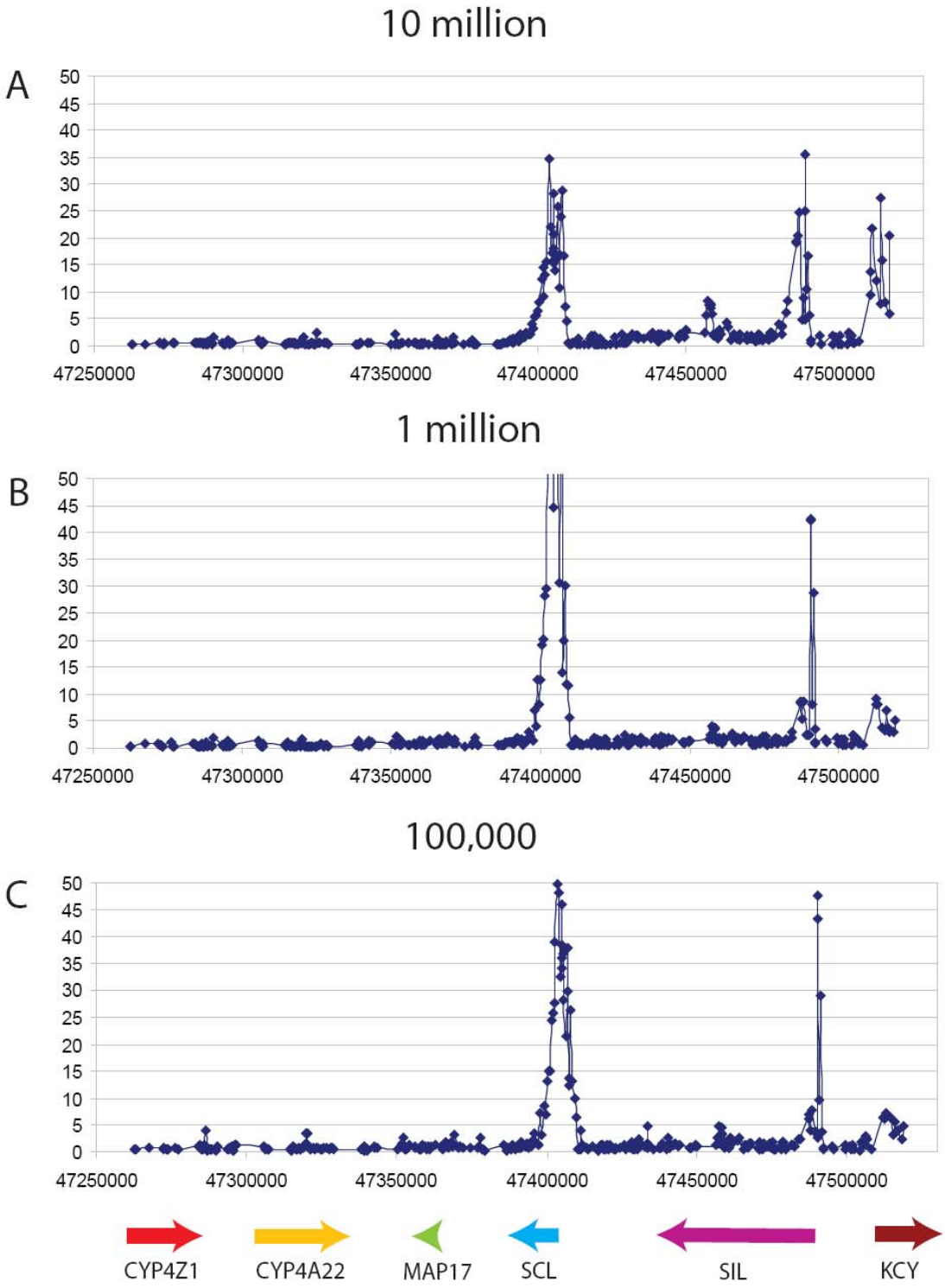

Figure 5.20: Testing the modified ChIP-chip method to detect regions associated with H3K79me3. ChIP-chip data obtained from a experiments performed with a range of K562 cell numbers and an antibody raised to H3K79me3. Panel A: ChIP-chip experiment performed with 10^7 cells and 10 µg H3K79me3. Panel B: A ChIP-chip experiment performed with 10^6 cells and 1 µg H3K79me3, Panel C: 10^5 cells, 0.5 µg

H3K79me3. Fold enrichment values are presented on the y-axis and the human chromosome 1 coordinates are indicated on the x-axis. The location of genes in the region and direction of transcription is indicated by coloured arrows.

5.11. Detection of transcription factor interactions using low cell number ChIP-chip

While the focus of the modified ChIP-chip method was to investigate histone modifications in low numbers of cells, it was important to determine if ChIP-chip could be used to study transcription factor interactions in low cell numbers. Many transcription factors are expressed at specific stages of development often in discrete populations of cells – therefore the development of a method to study transcription factor interactions in low numbers of cells would be of great benefit. A number of GATA-1 interactions have been previously characterised in K562 cells at the SCL locus (Dhami, submitted). ChIP-chip experiments were performed with 10^7, 10^6, 10^5 and 10^4 cells for the detection of GATA-1 interactions. Because of time limitations, comprehensive titration panels could not be performed for these assays and only one antibody concentration was chosen at each level of cell number. Based on the optimization experiments conducted for histone modifications, experiment performed with 10^6 cells used 1 ug of GATA-1 antibody and experiments performed with 10^5 cells and 10^4 cells used 0.5 ug of GATA-1 antibody. The results of these experiments are presented in Figure 5.21.

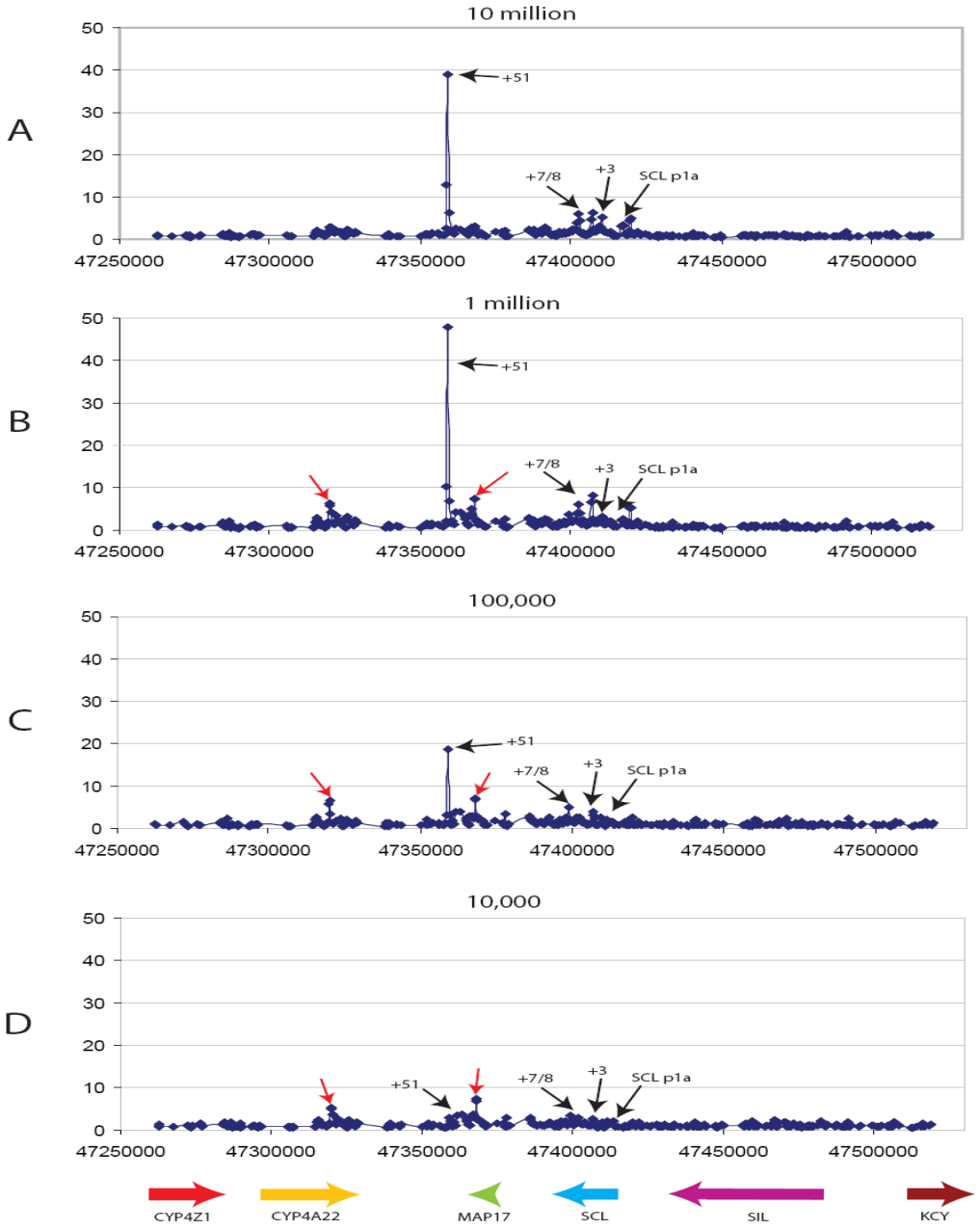

Figure 5.21: Detection of GATA-1 enrichments in reduced numbers of K562 cells. 10^7, 10^6, 10^5, and 10^4 cells were used in ChIP-chip experiments for the detection of GATA-1 enrichments. Panel A represents the experiment performed with 10^7 cells and 10 µg of antibody. Panel B represents the experiment performed with 10^6 cells and 1 µg of antibody. Panel C represents the experiment performed with 10^5 cells and 0.5 µg of antibody and panel D represents the experiment performed with 10^4 cells and 0.5 µg of antibody. The gene order and direction of transcription is indicated below panel D. The human chromosome 1 coordinates are indicated along the x-axes, while the values on the y-axes represent fold enrichments observed in ChIP-chip assays. Known GATA-1 interacting regions are indicated by black arrows while red arrows indicate significant enrichments not associated with known GATA-1 interactions.

In the experiment performed with 10^7 cells the genomic regions which displayed significant GATA-1 enrichments are shown. These regions include the SCL erythroid enhancer (+51) which displays the highest fold enrichment, and lower enrichments are observed at the SCL p^{1a}, the +3, and +7/+8 regions. These regions are located at or within close proximity to regions of histone H3 acetylation and known regulatory activity (Dhami 2007, submitted). These regions also contain highly conserved GATA family consensus binding sequences, providing further support that they are genuine GATA-1 binding sites in K562. The +51 element can is significantly enriched in assays performed with 10^6 cells and 10^5 cells, while no enrichment is observed for 10^4 cells. While the +51 region is still detected in 10^5 cell assays, there is only one array element which is significantly enriched as opposed to three neighbouring array elements which are enriched in assays performed with 10^7 and 10^6 cells. A number of significantly enriched peaks are observed as cell numbers are reduced (indicated by red arrow-heads), which do not contain a conserved GATA family consensus sequence (data not shown) and may represent an artifact of reducing the number of cells used in a ChIP-chip assay. The subtly enriched +3 region is not detected as significantly enriched in the assays performed with 10^6 or 10^4 cells, while GATA-1 interaction at SCL p^{1a} are not detected as significant when assays were performed with 10^5 or 10^4 cells. The data suggests that ChIP-chip assays can be used for the detection of highly enriched transcription factor interactions when assays are performed with 10^7-10^5 cells. Low enrichments, however, do not reproduce through to low cell number ChIP-chip experiments as the overall level of enrichments decrease and they are lost in the 'noise'.

5.12. Discussion

This Chapter describes experiments aimed at optimizing the conditions necessary to perform ChIP-chip for the detection of regulatory elements in the human genome when using reduced numbers of cells. The SCL genomic tiling path array was used in ChIP-chip to detect statistically significant H3 K9/K14 acetylation and H3K4me3 events associated with known regulatory elements in ChIP assays performed with 10^4 - 10^7 K562 cells. It was demonstrated that SCL array platform was sensitive enough to detect these

histone modifications from as few as 10^4 cells and could reproducibly detect and quantify ChIP enrichments, which were confirmed by quantitative PCR.

5.12.1. The development of a ChIP-chip method applicable to small cell populations

The main focus of this Chapter was the development of a ChIP-chip method for use with low numbers of cells. Three parameters were investigated for their effect on the efficiency of the ChIP procedure, namely chromatin amount, antibody concentration, and protein-G agarose concentration. It was shown that chromatin aliquots equivalent to 10^4 cells could be used to identify known regulatory elements in the SCL locus that were associated with H3 acetylation and H3K4me3. As chromatin amounts were reduced in ChIP assays it became apparent that antibody concentration was an important factor in identifying known regulatory elements. Generally when using less chromatin in a ChIP assay, it was found to be more beneficial to use reduced antibody concentrations, although the optimal antibody to chromatin ratio may be assay-specific. However, that said, the modified procedure was also demonstrated to perform well for H3K27me3, H3K36me3, and H3K79me3 assays, by extrapolating the conditions used for H3K4me3 and H3K9.K14ac. Thus, it may not always be necessary to perform complete titration series for all assays in order to determine a good working range of conditions for low cell number ChIP-chip.

Protein-G agarose concentration in the immunoprecipitation step was also examined for its effect on ChIP efficiency and how it effects non-specific enrichments in the resultant ChIP sample. It was observed that using less protein-G agarose (25µl) may be more beneficial for experiments performed with 10^5 cells as too much protein-G agarose may be contributing to 'noise' due to non-specific interaction with DNA. This effect was not examined for experiments performed with 10^4 cells but it is expected that a similar relationship may also exist for ChIP experiments performed with this number of cells. However, in practical terms, the use of low amounts of protein-G agarose may be limited in low cell ChIP-chip assays, as it is difficult to visualize the protein-G agarose pellets during the ChIP wash and elution steps as it is used in diminishingly small quantities.

The normalization of low cell number ChIP-chip data sets with rabbit "mock" IgG data was also examined and resulted in profound improvements in signal:noise levels across

the SCL locus for the detection of H3K4me3 and H3K9/14ac histone modifications. Many other laboratories routinely publish ChIP-chip data without considering the effect that non-specific interactions can contribute to the overall dataset. However, given that the vast majority of ChIP-chip data is currently performed using conventional ChIP, the contribution that this type of noise has on these datasets is not known. However, as more biological studies move towards low cell number ChIP-chip, the need to consider non-specific noise will become increasingly important.

5.12.2. Further optimisation of the modified ChIP-chip method

In this study, all aliquots of chromatin used for cell equivalent experiments were derived from batches of chromatin prepared from 10^8 cells. Therefore, optimising the preparation of chromatin from fewer numbers of cells is an important consideration. For example, when cross-linking low numbers of cells, it may be important to perform the cross-linking step in smaller volumes directly in tubes (as apposed to large flasks) to minimise the loss of material associated with transferring cross-linked cells from flasks to tubes. Cell and nuclei pellets may need to be coloured with a dye when preparing chromatin from low numbers of cells to minimize loss of material. The sonication of smaller numbers of cells may also need to be optimised to generate fragments of the correct size. Finally it may be beneficial to coat tubes in a bovine serum albumin to prevent chromatin proteins adhering to the sides of tubes during the immunoprecipiation step. This may enhance the recovery of precipitated histones. It may also be possible to use the modified ChIP method to investigate histone modifications in fewer cells as random amplification methods (discussed in the introduction to this Chapter) could be used to amplify ChIP DNA samples and provide a quantifiable signal on microarrays. Amplification of ChIP DNA may mean that investigating histone modifications in 10^3 or fewer cells may be possible using microarrays.

While the focus of this study was to develop a method for the study of histone modifications in human cell types limiting in number, a ChIP-chip assay for the detection of the transcription factor GATA-1 enrichment in low numbers of cells was also investigated. Proof-of-principle experiments showed that it was possible to detect known transcription factor interactions from as few as 10^5 cells. However, given that a number

of parameters (chromatin amount, antibody concentration and protein-G agarose concentration) where shown to be important for detecting histone modifications, the optimal concentrations for detecting transcription may be different to those used for detecting histone modifications and require further detailed investigations.

5.12.3. Comparisons with other methods

A number of ChIP-based methods have been developed recently for investigating histone modifications in small populations of cells as discussed in the introduction of this Chapter (O'Neill *et al*., 2006; Attema *et al*., 2007; Dahl and Collas, 2007). The carrier-ChIP method developed by O'Neill and colleagues used *Drosophila* carrier chromatin and PCR amplification of ChIP DNAs to analyse histone modifications in as few as 100 cells (O'Neill *et al*., 2006). The miniChIP method was developed by Attema and colleagues for the analysis of $5x10^4$ cells in early haematopoesis (Attema *et al*., 2007). The method was developed by altering a number of steps in the ChIP procedure which included the formaldehyde cross-linking step, sonication, pre-clearing and antibody immunoprecipitation conditions. Finally Q^2ChIP was developed to analyse histone modifications in as few as 100 cells (Dahl and Collas, 2007). Alterations in a number of ChIP steps were used to increase efficiency over conventional ChIP. These included cross-linking cells in suspension rather than in flasks to enhance cell recovery, while a HDAC inhibitor was used during the cross-linking stage as opposed to after cross-linking to maximise the amount of acetylated histones that could be precipitated. However, none of these studies demonstrated that their modified ChIP method could be used for the characterisation of histone modification patterns in a high-throughput manner on microarrays. The carrier ChIP, miniChIP, and Q^2ChIP methods all used quantitative PCR to determine histone modification enrichments at selected loci and the use of quantitative PCR is limited to investigating smaller genomic regions. In addition it may not be possible to use the carrier ChIP method for microarray analysis of histone modifications as the vast majority of the ChIP DNA is composed of *Drosophila* DNA, which would affect labeling efficiency of the human DNA. Thus the method described in this Chapter is the first known report of microarrays being used to detect histone modifications from as few as 10^4 cells.

However, with the recent developments in sequencing technology (Bentley, 2006) such as the Solexa method, it may be possible to sequence ChIP samples from experiments performed with low numbers of cells and provide a direct read-out of DNA sequence enrichment in ChIP samples. In the Solexa method, millions of individual DNA molecules in a ChIP sample can be linked using adapters to the surface of a glass cell and then amplified to create a cluster of DNA molecules with identical tags. The DNA molecules in each cluster are then simultaneously sequenced using the sequencing by synthesis chemistry which can be used to assess abundance of a particular DNA sequence in a ChIP sample (Barski *et al*., 2007; Mikkelsen *et al*., 2007). This sequencing method requires only nanogram quantities of starting material and could be used to provide whole-genome epigenetic maps for limited cell populations.

5.13. Conclusions

The work described in this Chapter showed how empirically-tested modifications to the conventional ChIP-chip protocol can be used to reproducibly detect histone modifications in reduced numbers of cells. Optimal antibody concentrations for ChIP assays established here across a range of cell numbers provide a working range of concentrations that may be applicable for the study of different cell types. These optimisations may be used for the study of histone modifications in cell types which previously would not have been amenable to ChIP-chip analysis. To this end, the following Chapter describes the application of these modifications to study several histone modifications in human primary monocytes and human embryonic stem cells to understand chromatin state during differentiation.