Application of DNA Microarrays to Assess DNA Replication Timing and Chromosomal Aberrations.

Kathryn Woodfine
Darwin College

This dissertation is submitted for the Degree of Doctor of Philosophy

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration, except where specifically indicated in the text.

This Dissertation does not exceed the word limit set by the Biology Degree Committee.

Kathryn Woodfine.

Abstract

I have developed a directly quantitative method to assess the replication timing of sequences during the S phase of the cell cycle utilizing genomic clone DNA microarrays. This is achieved by the co-hybridisation of differentially labelled S and G1 phase DNA to the arrays. The genomic resolution of the replication timing measurements is limited only by the genomic clone size and density on the arrays.

I have demonstrated the power of this approach by constructing a genome wide map of replication timing in human lymphoblastoid cells using an array with clones spaced at 1 Mb intervals. I also constructed an array using chromosome 22 tile path clones and produced a high resolution replication timing map of 22q. Tile path resolution replication timing maps have also been produced for chromosomes 1 and 6.

I have shown a positive correlation, both genome wide and at a tiling path resolution, between replication timing and a range of genome parameters including GC content, gene density and transcriptional activity.

I have further developed the replication timing assay by using an array of PCR products spanning 4.5Mb at a resolution of 10kb, and an array spanning 20Kb using overlapping 500bp PCR products. This will allow the study of correlations with sequence features at a high resolution.

Using the Chromosome 22 tile path array I have also been able to show changes in replication timing in a cell line which contains a balanced translocation between chromosomes 17 and 22. I have also used the chromosome 22 tile path array to analyse deletions in DiGeorge patients and to detect VJ recombination at the immunoglobulin light chain λ lymphoblastoid cell lines.

Acknowledgments

I would like to thank all the people, who without which the work in this thesis would not be possible.

Firstly, all the external collaborators;
Silvana Debanardi and Bryan Young from the Molecular Oncology Unit, St Bartholomew's Hospital, Katrina Prescott from the Institute of Child Health, Richard Mott from the Wellcome Centre for Human Genetics and Charles Shaw-Smith from Addenbrookes Hospital.

The Human Genetics Department at The Wellcome Trust Sanger Institute. Especially; Carol Carder, Paul Hunt, Sean Humphray, Lisa French, Carol Scott, Rob Andrew, Pawendeep Dharmi, Cordelia Langford, Oliver Dovey and David Vetrie.

The Molecular Cytogenetics group (past and present) at the Sanger Institute, paticularly;
Heike Fiegler, Philippa Carr and Eleanor Douglas for the construction of the 1 Mb resolution genomic array, Shelia Clegg for teaching me FISH and Elena Prigmore for discussion and support.

The Molecular Genetics and Proteomics group at the Sanger Institute, especially; David Beare, for his immense help with sequence analysis of chromosome 22, John Collins for advice on analysis of low copy repeats and Owen McCann for help in the assembly of the chromosome 22 clone set

I would like to thank all the people who have given me a personal input into this thesis. In particular I would like to thank;
Gary Woodfine for his support, Andy Mungall, Lisa Rickman, and Deb Burford for proof reading and their suggestions and Susan Gribble for all her help, too great to mention.

Finally I would like to thank my two supervisors, Nigel Carter and Ian Dunham for all their help, support and patience.

Table of Contents

List of Figures:

List of Tables:

List of Abbreviations:

approx.	Approximately
ATP	Adenosine Triphosphate
BAC	Bacterial artificial chromosome
bp	base pair
BrdU	Bromodeoxyuridine
cdk	cyclin dependant kinase
cDNA	Complementary DNA
CpG	Cytosine and Guanosine dinucleotide
Cy	Cyanine dye
D. melanogaster	*Drosophila melanogaster*
dATP	2'-Deoxyadenosine 5'-triphosphate
dCTP	2'-Deoxycytosine 5'-triphosphate
dGTP	2'-Deoxyguanosine 5'-triphosphate
DMSO	Dimethyl sulphoxide
DNase	Deoxyribonuclease
dNTP	2'-Deoxynucleoside 5'-triphosphate
DOP	Degenerate Oligonucleotide Primer
dsDNA	Double stranded DNA
dTTP	2'-Deoxythymidine 5'-triphosphate
E. Coli	*Escherichia coli*
EDTA	Ethylenedinediaminetetraacetic acid
EST	Expressed sequence tag
Fig.	Figure
FISH	Florescence *in-situ* hybridisation
G	Giemsa
G1	Growth 1 phase of the cell cycle
G2	Growth 2 phase of the cell cycle
GC	Guanosine + Cytosine
H	Histone
HCl	Hydrochloric acid
HPLC	High
IgH	Immunoglobulin heavy chain
IgL	Immunoglobulin light chain
J	Joining region (IgL)
K	Lysine
Kb	kilobase
LB Agar	Luria-Bertani agar
log	logarithmic
M	molar
M	Mitosis phase of the cell cycle
Mb	Megabase
MCM	Mini chromosome maintainance

μg	microgram
mg	miligram
μl	microlitre
μM	micromolar
mM	milimolar
mRNA	Messenger RNA
NaAc	Sodium Acetate
NaCl	Sodium Chloride
nm	namometer
ORC	Origin Recognition Complex
Ori	Origin of Replication
PAC	P1-derived artificial chromosome
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PI	Propidium iodide
Pre-RC	Pre-Replication Complex
r.p.m	revolutions per minute
RNA	ribonucleic acid
RNase	Ribonuclease
S	Synthesis phase of the cell cycle
SDS	Sodium dodecyl sulphate
SSC	Sodium chloride/citrate solution
STS	Sequence tagged site
TDP	Timing Decision Point
TE	Tris (hydroxymethyl) aminomethane-Ethylenedinediaminetetraacetic acid
V	Variable region (IgL)

1: Introduction

For genetic information to be passed on from one generation of cells to the next the genome has to be replicated with high fidelity. This occurs during the synthesis (S) phase of the cell cycle. DNA replication is a temporally ordered process with different regions of the genome replicating at different times in S phase. The time at which DNA replication is initiated is a highly ordered process. Replication during the early part of S phase has been associated with chromatin conformation, epigenetic and sequence features of the genome and transcriptional activity. Each of these features are considered below. The main aim of the work presented in this thesis is to use genomic arrays to assess replication timing.

Conventional ways of assessing replication timing are laborious and can only assay small regions of the genome. These are described and compared to the use of microarrays for the assay of replication timing. The utilisation of microarrays to assess the replication timing of *Saccharomyces cerevisiae* and *Drosophila melanogaster* is explained. To assay the replication timing of the human genome an array sampling the genome at a 1Mb resolution was used. To examine replication timing of a whole chromosome a genomic clone microarray covering the q arm of chromosome 22 was constructed.

The arrays constructed to assess replication timing can also be used to detect copy number changes. The use of the arrays to identify deletions in 22q due to rearrangement of the immunoglobulin light chain λ and implicated in DiGeorge syndrome was also investigated.

1.1: Chromatin Conformation

1.1.1: Chromatin Condensation

Every human cell contains over two metres of DNA, packaged into a nucleus 5-20 μm in diameter. To achieve this, the DNA is packaged in a highly ordered process (Figure 1.1) This allows condensation of the DNA but still enables replication and

transcription machinery access. Several groups of proteins are involved in the condensation of DNA and its positioning within the interphase nuclei, and the combined structure of DNA and protein is termed chromatin.

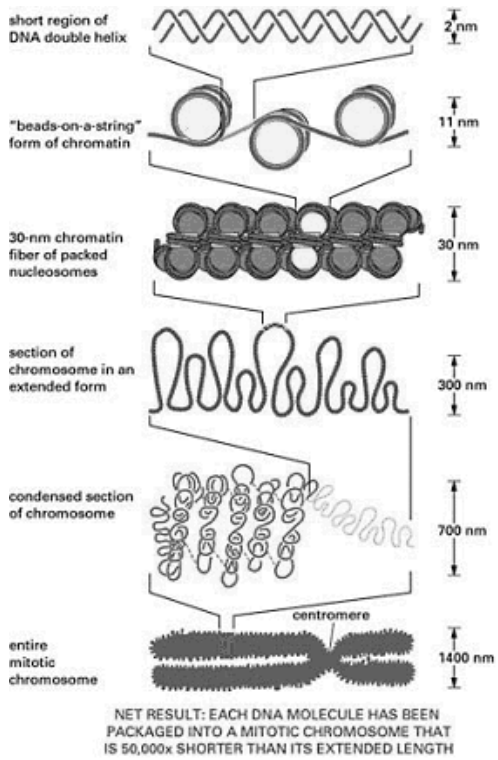


Figure 1.1: Levels of DNA condensation within a eukaryotic chromosome (Strachan 2001)

The first level of DNA condensation occurs when 146bp of DNA is wrapped around a nucleosome. The nucleosome is made of core histone proteins which are described in section 1.1.2. DNA is wrapped twice around a nucleosome to form the recognised beads-on-a-string level of chromatin packing. This produces a fibre approximately 11nm in width.

The second level of chromatin condensation is the arrangement of the nucleosomes into a 30nm chromatin fibre within which there are six nucleosomes per turn. The circular nucleosomes with their DNA wrapped around them can be represented as disc shaped and the discs align so their flat face is roughly parallel to the long axis of the chromatin fibre (See Figure 1.1).

The chromatin fibre is then arranged in loop domains each approximately 120Kb long (Munkel, Eils et al. 1999). These were first seen in scanning electron micrographs of lampbrush chromosomes of the *Urodela* (amphibian) oocyte (Miller 1965). These chromatin loops form rosettes. Each rosette is termed a multi loop sub-compartment (MLS) and comprises of six loops. A further 120Kb of DNA links each rosette. The rosette is attached at its centre to the nuclear matrix (Paulson and Laemmli 1977) and is associated with Histone H1 (Munkel, Eils et al. 1999). Two types of attachment regions have been described; attachment to the nuclear matrix of permanent regulatory regions containing non-transcribed DNA, and attachment to the nuclear skeleton of transient regions of DNA containing transcribing and replicating DNA (Craig, Boyle et al. 1997). This second class of attachment has been visualised by electron microscopy and shown to contain DNA replication and RNA transcription machinery (Hozak, Hassan et al. 1993; Hozak, Jackson et al. 1994).

The final level of chromosome compaction occurs prior to cell division in mitosis. Mitotic chromosomes have a diameter of 700nm. The rosettes are lined up and attached to a central chromosome scaffold of non-histone acidic proteins. The condensation of the metaphase chromosome is due, in part, to a protein called 13S condensin, which is thought to act in an ATP dependent manner and condense chromosomes by inducing a globally positive supercoil. (Kimura and Hirano 1997; Kimura, Rybenkov et al. 1999). The chromosomal scaffold attachment regions are very AT rich and are consequently twice as abundant in the gene poor regions of the chromosome (Craig, Boyle et al. 1997). This tight binding of the AT rich regions to the chromosomal scaffold may explain the banding pattern seen in giemsa stained metaphase chromosomes, with the tightly bound gene poor DNA producing densely stained regions. Metaphase chromosomal bands can therefore be classified into G light (GC rich) or G dark (AT rich) bands.

Chromatin has been classified into two categories, heterochromatin, identified by very intense giemsa staining and euchromatin. In heterochromatin, DNA is in a highly condensed state that restricts the access of additional proteins to the DNA. Heterochromatin was a term first used by the botanist Emil Heitz who identified parts of a moss karyotype that were more compact than others regions (Heitz 1928, Redi, Garagna et al. 2001). These ‘C’ bands were identified by a boiling technique called

'heitzen'. The compact regions were later identified in both animals and plants (Heitz 1930). Unlike its sister euchromatin, heterochromatin was identified as being transcriptionally inert (Ohno 1985). Heterochromatin is usually AT rich, late replicating and gene poor. The condensed state means access to the DNA by other proteins is restricted. Heterochromatin is very rich in repeat sequences such as satellite DNA sequences which are required for correct sister chromatid adhesion and chromatin separation during mitosis. Euchromatin is more loosely coiled than heterochromatin. It is gene rich (and transcriptionally active), and has a high GC content (See Table 1.1). Cimbora *et al*. (Cimbora, Schubeler et al. 2000) found that at the β-globin locus, which is located in euchromatin, the open chromatin state was necessary for early replication.

1.1.2: The Nucleosome and Epigenetic regulation.

Histones are the protein subunits that make up the nucleosomes. They are found in the chromatin of all eukaryotic cells and are highly conserved throughout evolution (Li 2002). The core histones that make up the nucleosome are H2A, H2B, H3 and H4. Two copies of each can be found within each nucleosome and are assembled as illustrated in Figure 1.2. A further class of Histone (H1) can be found associated with the DNA linking the individual nucleosomes.

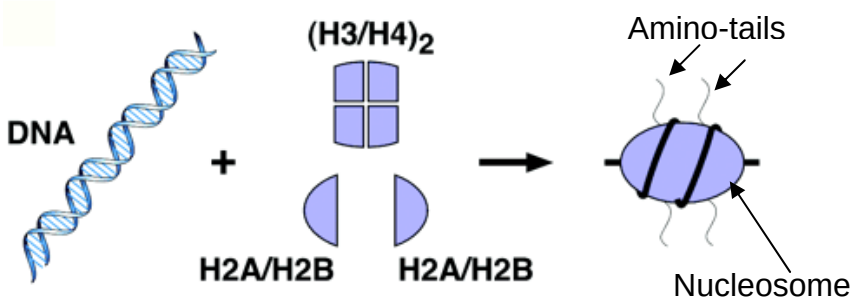


Figure 1.2: Composition of a nucleosome, the fundamental unit of chromatin. Adapted from Grewal *et al*, 2003. (Grewal and Moazed 2003). Each nucleosome contains 146bp of DNA wrapped around an octamer of core histone proteins. Histone proteins are arranged to have their amino tails protruding from the nucleosome core.

The extension of the amino terminus tails from the nucleosome core allow epigenetic regulation of the amino acid residues they contain. Covalent modification of the

amino acid residues is performed by chromatin re-modelling and chromatin modification enzymes. These enzymes mainly target the lysine residues of the amino acid tails. Modification is by acetylation, phosphorylation, methylation and ubiquitylation. The modifications are associated with the transcriptional activity of the associated DNA. Epigenetic regulation of the genome is a heritable feature, yet is independent of DNA sequence (Li 2002).

The amino termini of the H3 and H4 subunits are particularly involved in epigenetic regulation. Figure 1.3 shows the most prominent protein modifications that can occur at the lysine residues within the histone tails. The pattern of histone tail modification is called the histone code.

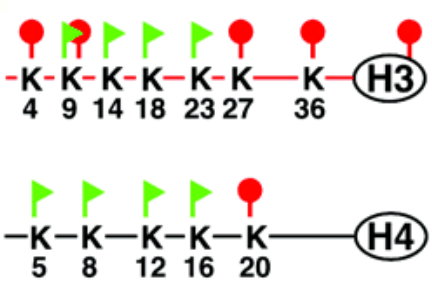


Figure 1.3: Modification at the lysine (K) residues in the H3 and H4 tails reproduced from Grewal *et al* 2003 (Grewal and Moazed 2003). A green flag signifies the lysine is subject to the addition of an acetyl group, whilst the red signal indicates that lysine is subject to the addition of a methyl group.

The addition of acetyl groups to the histone tails is associated with opening of the nucleosomes and a more transcriptionally active state of DNA. The de-compaction of the nucleosomes also makes the DNA more susceptible to DNAase1 activity (Kerem, Goitein et al. 1984). Conversely the removal of acetyl groups from the histone tails results in a closed conformation and transcriptional repression. A correlation between acetylation status and DNA replication timing has also been observed. Hyperacetylated DNA is early replicating, whilst hypoacetylated DNA is late replicating (Vogelauer, Rubbi et al. 2002).

Methylation of the histone tails has different effects, dependent on the location of the lysine residue that is modified. Methylation at the lysine in position 4 of the H3 (H3-K4) tail is associated with active gene expression whilst methylation at H3-K9 tail is involved in heterochromatin assembly. Methylation of H3-K9 is also thought to maintain DNA methylation (Grewal and Moazed 2003).

In *Arabidopsis thaliana*, H3-K9 methylation was shown to promote DNA methylation through heterochromatin protein 1 (HP1) (Jackson, Lindroth et al. 2002). The binding of HP1 to H3-K9 recruits a DNA methyltransferase, which in turn results in a covalent addition of a methyl group to cytosine molecules within the DNA double helix. Although this link is yet to be confirmed in mammals, H3-K9 methylation has been shown to recruit HP1 (Bannister, Zegerman et al. 2001; Lachner, O'Carroll et al. 2001). The down-stream interactions with DNA methylation are likely to be more complex due to the greater variety of methyltransferases in mammalian cells. The methylation of cytosine molecules is particularly important at CpG islands (clusters of the CpG dinucleotide). About 60% of human genes have CpG islands located at their 5' ends (Cross, Clark et al. 2000). Covalent addition of a methyl group to the cytosine of CpG dinucleotides in CpG islands renders their associated genes transcriptionally silent.

Table 1.1: A summary of chromatin and associated epigenetic features.

Feature	**Euchromatin**	**Heterochromatin**
Replication Time	Early	Late
Gene Density	Dense	Sparse
***Alu* Repeat content**	Dense	Sparse
Acetylation of histones	Hyperacetylated	Hypoacetylated
Methylation of CpG	Hypomethylated	Hypermethylated
DNAase sensitivity	Sensitive	Insensitive

In summary, the packaging of DNA into the interphase nucleus is a highly ordered process. This packaging is influenced by the epigenetic modification of the genome by histone acetylation and methylation, and methylation of CpG dinucleotides.

1.2: DNA Replication, the Eukaryotic Cell Cycle and Replication Origins

1.2.1: The Eukaryotic Cell Cycle

About three million cells are replaced and renewed in the human body every minute (http://www.nobel.se/medicine/educational/2001/). This is achieved by cell division. As this takes place the genetic material from the mother cell is divided into two daughter cells. To ensure no genetic material is lost or gained as cell division takes place the entire genome must be copied once and only once before cell division. The doubling of the genome takes place within the synthesis (S) phase of the cell cycle. The cell division and therefore halving of the genetic material takes place in the mitosis (M) phase of the cell cycle. These two phases are interspersed with two growth periods (G1 and G2), to complete the somatic cell cycle (Figure 1.4).

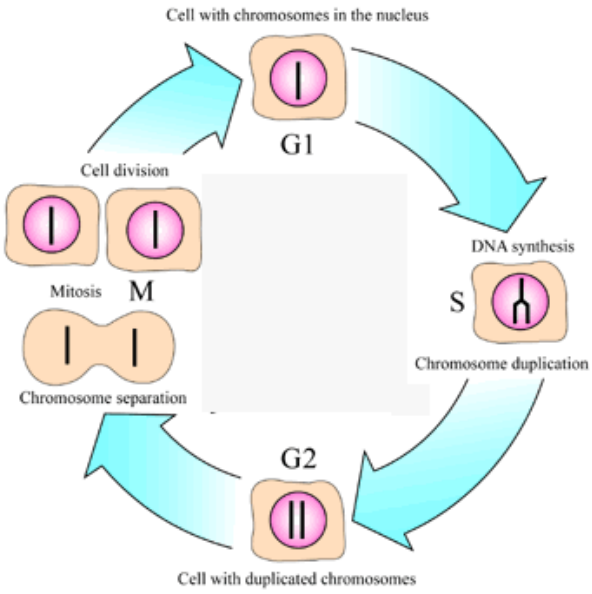


Figure 1.4: The Cell cycle.

1.2.2: DNA replication during the S phase of the cell cycle.

DNA replication is initiated at a specific site called the Replication Origin. As shown in Figure 1.5, a DNA helicase unwinds the DNA to produce a replication fork and DNA is then synthesised in a 5'-3' direction. DNA polymerase δ synthesises the DNA on the leading strand, whilst DNA polymerase ε synthesises short fragments (Okasaki fragments) of DNA using the lagging strand as a template. DNA ligase joins the newly synthesised lagging strand Okasaki fragments together to form a continuous DNA molecule. This continues in a bi-directional fashion until an entire replicon of 40-300Kb has been replicated (Natale, Li et al. 2000). DNA replication is said to be semi-conservative, with each newly synthesised DNA molecule containing one strand from the mother DNA molecule, and one newly synthesised daughter strand.

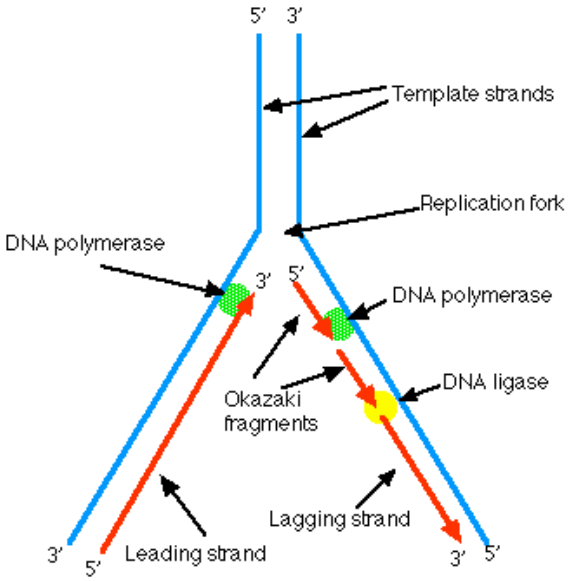


Figure 1.5: The DNA replication fork. For details see text. Figure taken from http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/D/DNAReplication.html

The temporal order of replication is strictly regulated, with some regions of the genome replicating much earlier than others. The replication of one replicon is triggered by the activation of one replication origin. As replication of a replicon nears completion the bubbles of nascent DNA formed from individual origins fuse to form two new DNA molecules (Figure 1.6)

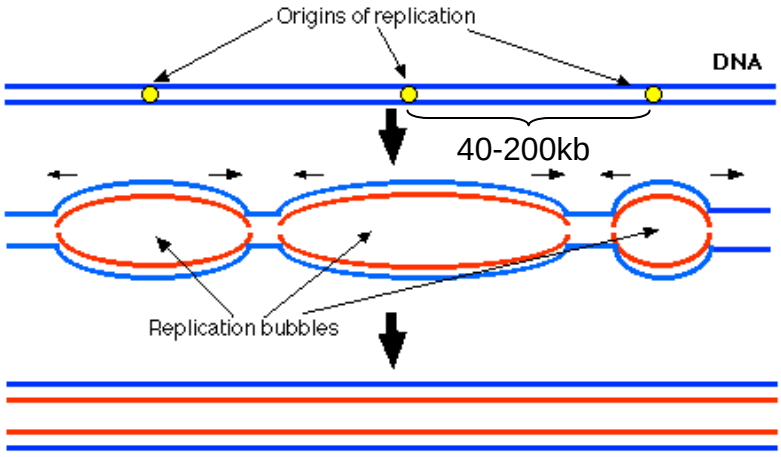


Figure 1.6: DNA replication from origins of replication. Figure taken from http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/D/DNAReplication.html.

In *Escherichia coli*, *Saccharomyces cerevisiae* and *Saccharomyces pombe* replication origins are sequence specific. Although the sequence is not conserved throughout these species it is always an AT rich sequence. However, in higher eukaryotes there is no consensus sequence at which DNA replication is initiated (Kelly and Brown 2000). In higher eukaryotes many sites have the potential to become origins. These sites have specific proteins bound to the chromatin throughout the entire cell cycle, called pre-replication complexes (Pre-RC's). Activation of a few of these sites by the binding of the proteins of an origin recognition complex (ORC) determines the site of replication initiation (Diffley and Labib 2002; Nishitani and Lygerou 2002). Very few replication origins and ORC binding sites have been defined in mammalian cells. Investigation by Natale *et al* showed that each mammalian cell contains 10^4-10^5 molecules of ORC. This suggests DNA replication is initiated once every 60-600Kb (Natale, Li et al. 2000).

The features that determine where the ORC subunits bind initially are still unknown in mammals. In common with the replication origins of lower eukaryotes, many replication origins contain tracts of AT rich sequence (Vashee, Cvetic et al. 2003). However *in vitro* binding studies using purified human ORC showed that ORC did not preferentially bind to these AT tracts (Vashee, Cvetic et al. 2003). In contrast the study of CpG island regions in four mammalian genes showed that clusters of the CpG dinucleotide were at the initiation sites for DNA replication. Short nascent

DNAs synthesised *in vivo*, were found to contain CpG islands suggesting the origins of replication were nested inside the CpG islands (Delgado, Gomez et al. 1998).

The origins that have been described in mammals sub-divide into two categories; loci at which replication is initiated at a discrete chromosomal location and loci at which the origins are found over a much larger zone of initiation (Gilbert 2001). The first category of origins includes the lamin B2 locus and the human β-globin locus. The lamin B2 locus was first mapped by Giacca (Giacca, Zentilin et al. 1994) to a 474bp region corresponding to the non-coding 3' end of the Lamin B2 gene. An evolutionary conserved AT rich region was also observed proximal to this region. The location of the lamin B2 origin was further refined and localised to a single nucleotide (Abdurashidova, Deganuto et al. 2000). The human β-globin origin was the first human origin to be mapped and localises to a defined region, less than 4Kb in size (Kitsberg, Selig et al. 1993) between the adult δ-globin and β-globin genes. Study of this region showed that this locus was used for replication in cells that both expressed and did not express β-globin, and that replication from the origin was bi-directional (Kitsberg, Selig et al. 1993). The β-globin origin was shown to be early replicating and it was proposed to be due to the open chromatin structure of the locus. Models of origin activation at this locus indicated long range control of the origin. Cis acting elements within the open chromatin regulate activation of origins over 50Kb away. (Cimbora, Schubeler et al. 2000).

The second class of replication origin consists of those with a large zone of initiation. These zones can be 10-50Kb regions of DNA within which replication begins from several sites. The Chinese hamster ovary cell DHFR locus is the best defined locus in this group. Two dimensional gel analyses of replication intermediates revealed a 55Kb region of initiation between the DHFR gene and the 2BE2121 gene (gene of unknown function) (Dijkwel and Hamlin 1995). This large region has been shown to contain a minimum of 20 origins, each with different efficiencies of activation (Dijkwel, Wang et al. 2002). Studies at this locus show that when the most active origin is deleted, adjacent origins increase or retain their initiation activity (Kalejta, Li et al. 1998). Other metazoan origins that have a dispersed zone of initiation include the *Drosophila melanogaster* Ori D locus (Ina, Sasaki et al. 2001), the rRNA genes

from human (Little, Platt et al. 1993) and the Chinese hamster Rhodopsin origin (Dijkwel, Mesner et al. 2000).

1.2.3: Cell cycle regulation and control checkpoints

DNA is licensed for replication during the G1 phase of the cell cycle and progression into the G2 phase of the cell cycle is prevented until all DNA has been replicated. This is achieved by a series of regulatory proteins and cell cycle checkpoints.

Early in the G1 phase of the cell cycle the nuclei pass through a 'timing decision point' (TDP). It is during this time that the replication timing program of the cell is established. This occurs before the licensing of the DNA for replication (Gilbert 2002). During the TDP, active DNA sequences are repositioned in clusters in the nucleus to provide a favourable environment for gene transcription. These clusters are related to chromosomal domains and contain a high concentration of replication regulators acting in trans (Dimitrova and Gilbert 2000; Li, Chen et al. 2001). After nuclear repositioning the replication timing program of the nuclei is determined. In the Chinese hamster β-globin locus, the determination of late replication within heterochromatin occurs coincidently with its repositioning at the nuclear periphery (Li, Chen et al. 2001). This provides a link between replication, nuclear position and a favourable environment for transcription (Dimitrova and Gilbert 1999). The importance for nuclear position at the TDP is confirmed by the analysis of sequences situated next to the nuclear periphery (associated with late replication, see section 1.3.3). If these loci are moved away from the nuclear periphery after the TDP they remain late replicating (Gilbert 2002).

To ensure that DNA is replicated once and only once, un-replicated DNA must be licensed to replicate. Only licensed DNA can undergo replication. Licensing occurs after the cells have undergone mitosis in the early G1 phase of the cell cycle. To license an origin for replication pre-replicative complexes (pre-RCs) are formed on the DNA (Nishitani and Lygerou 2002). The pre-RC's contain a six subunit protein, the origin recognition complex (ORC), Cdc6/18, Cdt1 and mini-chromosomal maintenance (MCM) proteins. All these proteins are highly conserved in eukaryotic

cells, highlighting their importance in replication. Some ORC subunits are present on the DNA throughout the cell cycle. The complete ORC is formed and then acts as a 'landing platform' for the other proteins involved in establishing the pre-RC. The Cdc6/188 and Cdt1 load first and recruit the binding of the hexameric MCM complex onto the chromatin (Figure 1.7). About 20 copies of the MCM hexamer are loaded onto the DNA at each Pre-RC (Diffley and Labib 2002). Once all proteins are loaded the origin is licensed for replication.

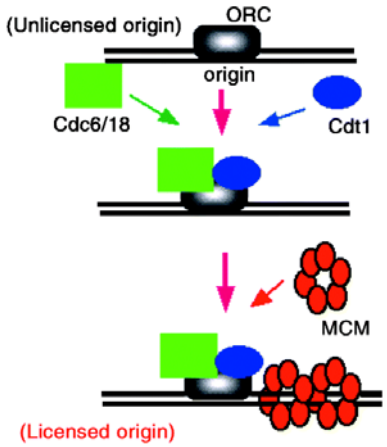


Figure 1.7: Formation of a pre-RC complex to license DNA for replication. Reproduced from Nishitani *et al* (Nishitani and Lygerou 2002).

Before the G1 cells move into S phase of the cell cycle they still have to pass a checkpoint (Restriction point – R) in late G1 phase (Ford and Pardee 1998). Before the restriction point is passed, cells can only continue to cycle if the nucleus receives extracellular signals. Once the restriction point is passed, cells are committed to the S phase of the cell cycle, and ultimately, division (Pollard 2002). This checkpoint is regulated by cyclin dependant kinases (cdk), which are, in turn, dependant on cyclins for their activation.

A number of proteins are involved in the activation of replication although the exact factors determining origin firing are poorly understood (Diffley and Labib 2002). Once an unknown signal to replicate is received, Cyclin D binds with Cdk4/6 and p21 in the cytoplasm to form an active kinase. This can pass through the nuclear envelope and phosphorylate pRB (retinoblastoma protein, a tumour suppressor gene). Passage of pRB through the nuclear envelope leads to progression through the restriction

point. In a second wave of activity, Cyclin E binds to cdk2 which also phosphorylates pRB and leads to progression into S phase (Pollard 2002).

After replication, the cdt1 proteins disassociate from the ORC and are degraded (DePamphilis 2003). This ensures origins that have been replicated no longer have their licence to replicate. The degradation of the licensing proteins also ensures that the cdt1 can not associate with any other ORCs forming a new origin of replication. The MCM and the cdk6/18 proteins also disassociate from newly replicated DNA. The MCMs drop off the DNA in front of the advancing replication fork. The soluble fraction of cdc6/18 is phosphorylated by cyclin dependant kinases and is exported to the cytoplasm after the onset of S phase (Nishitani and Lygerou 2002). The six subunit ORC protein also undergoes some disassociation. Some of the ORC subunits are removed from the chromatin, whilst others stay bound during all stages of the cell cycle. Many other replication control proteins have been identified in yeast and a few of these are conserved in humans. It appears that both cyclin dependant kinases (CDKs) and Skp1-cullin-F-Box protein (SCF) may be involved in re-replication control in human cells (Furstenthal, Swanson et al. 2001) however the exact mechanisms of control are unclear.

There is also a checkpoint control within S phase. Damage to the DNA or stalled replication forks activates a global intra-S integrity checkpoint (Diffley and Labib 2002). Activation of the intra-S checkpoint prevents the firing of origins late in S phase. Entry into mitosis is also prohibited. Experiments producing DNA damage with ionising radiation showed that the protein ATM kinase phosphorylates the proteins Nbs1 and Chk2 to trigger two individual branches of the intra-checkpoint control (Falck, Petrini et al. 2002).

To maintain genome integrity it is also important that no replication occurs in the G2 phase of the cell cycle. The presence of the protein germinin inhibits re-replication of DNA. Germinin binds to Cdt1 tightly; this binding prevents the loading of the MCM protein complex back onto the DNA. Without the loading of MCMs on to the DNA the pre-RC cannot be formed and DNA replication cannot take place (Nishitani and Lygerou 2002).

Finally, in the G2 phase of the cell cycle there is an additional checkpoint. The G2 checkpoint ensures that it is safe for the cell to enter mitotic division and ensures that all DNA replication and repair processes are complete. Rad and Hus (Hydroxyurea sensitive) proteins detect un-replicated DNA. The initiation of this checkpoint control is not fully understood, but a variety of genes are involved. These include Rad9, Rad1 and Hus1, which in human cells form a circular trimer that may detect damage as it slides along the DNA. This ensures only fully replicated undamaged DNA can be incorporated into daughter cells (Pollard 2002).

This evidence about the cell cycle shows that DNA replication is highly regulated. Regulation is applied not only while DNA replication takes place during S phase of the cell cycle, but during the G1 and G2 phases as well. Regulation in the G1 phase licences the DNA for replication and, possibly by organising nuclear position, determines when in S phase DNA replication occurs. Checkpoints in the G2 phase of the cell cycle ensure that all DNA is replicated before cell division takes place.

1.3: Conventional ways of assessing Replication Timing

Replication timing has been measured in a variety of ways since bromodeoxyuridine (BrdU, a thiamine analogue) was first used to assess replication timing by its incorporation into the DNA and assessment in metaphase chromosomes. Methods investigating the banding patterns on metaphase chromosomes have a limited resolution, imposed by chromosome morphology. Other methods, such as quantitative PCR or FISH based methods, assay small regions of the genome at a high resolution but are labour intensive and study of a large region is laborious

1.3.1: Assessment of replication timing by pulse labelling with Bromodeoxyuridine

The replication timing of the whole genome can be assessed at a very low resolution by pulse labelling cycling cells in culture with BrdU (Latt 1973; Yunis 1981; Drouin, Lemieux et al. 1990). The cells are harvested and those containing metaphase chromosomes are examined. The BrdU incorporation can then be used to determine the late and early replicating DNA as BrdU will only be intergrated into the DNA in a

specific fraction of S phase. In the example illustrated in Figure 1.8 the cells were pulse labelled for three hours. The S phase is 8 hours and G2 phase is 2 hours long in the example. In the top example (Figure 1.8), three hours have elapsed between the addition of BrdU to the culture media and the harvest of the metaphase chromosomes. DNA that has replicated in the last hour of S phase can be detected by BrdU incorporation. This is by using a monoclonal antibody to BrdU which is conjugated to a fluorescent reporter molecule. By the variation of the length of incubation with BrdU different banding patterns can be achieved. This relates to the replication timing of the chromosome.

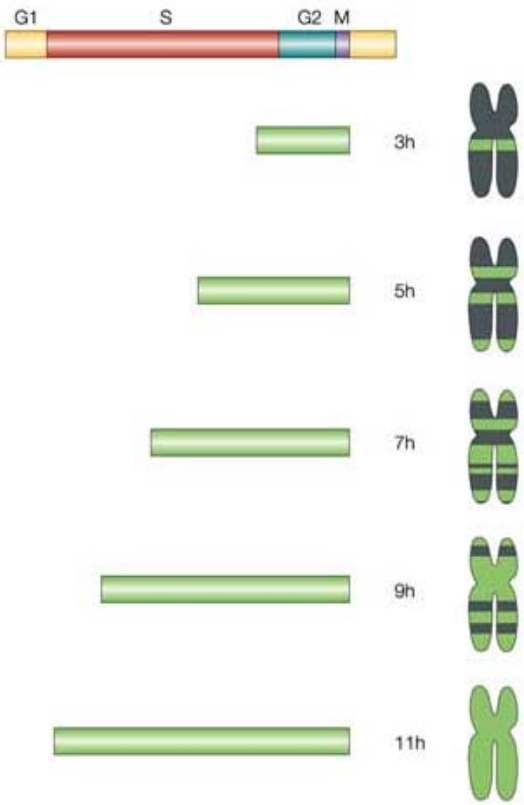


Figure 1.8: Figure from Goren and Cedar 2003. Pulse labelling of a cycling cell line with BrdU leads to banding patterns that report the replication time of the chromosomal bands. Green bands are those regions of DNA that have replicated the most recently before harvest at the end of the S phase. Black regions represent regions of the chromosome that have replicated before the BrdU was added. See text for full details.

The limit to this method is the resolution of the banding pattern of the metaphase chromosomes. Improved banding techniques have enhanced chromosome banding resolution to 1Mb (Yunis 1981; Drouin, Lemieux et al. 1990). However, this still only reports replication timing on a gross cytogenetic level. Using this methodology it is impossible to determine the replication timing of individual genes, or identify subtle changes in replication over small regions.

1.3.2: Assessment of replication timing by fluorescence in situ hybridisation.

An additional method of determining the replication timing of small regions of the genome is to perform a fluorescence *in situ* hybridisation assay (FISH - (Selig, Okumura et al. 1992), reviewed by(Boggs and Chinault 1997)). FISH is a technique that can be used for the mapping of sequences on metaphase chromosomes and the detection of DNA copy number changes within nuclei. DNA from the sequence of interest is labelled incorporation of a hapten which can be detected by conjugation with a flurochrome to make a probe. The DNA probe is hybridised (in the presence of Cot DNA to suppress hybridisation of common repeat sequences) to a DNA target affixed to a glass slide. A fluorescent signal confirms the hybridisation of the probe, and the presence of its complementary sequence within the target.

In the method to assay replication timing described by Selig (Selig, Okumura et al. 1992), an exponentially growing cell line was pulse labelled with BrdU to allow detection of interphase nuclei in the S phase of the cell cycle. The BrdU labelled nuclei were immunodetected with a monoclonal antibody to BrdU conjugated to FITC. The DNA probe was labelled by nick translation with biotin-dUTP and hybridised to the interphase nuclei.

There are three possible hybridisation signal combinations (Figure 1.9). Two singlet signals represent un-replicated DNA at both alleles (Figure 1.9 A), two doublet hybridisation signals represent replicated DNA at both alleles (Figure 1.9 B) and one singlet and one doublet represent one un-replicated and one replicated allele (Figure 1.9 C). This last scenario is usually present in the minority unless the replication

timing at the allele involved is asynchronous. This is further described in section 1.4.4.

Figure 1.9: Possible DNA replication patterns displayed by the fluorescence hybridisation assay. A: two alleles of un-replicated DNA, B: two alleles of replicated DNA and C: one allele of replicated DNA and one allele of un-replicated DNA. (Adapted from (Goren and Cedar 2003))

The ratio of nuclei with the pattern shown in Figure 1.9B versus the pattern shown in Figure 1.9A indicates the replication time of the loci being investigated. A probe producing a high proportion of doublets reports DNA that replicates early; conversely a probe producing a high proportion of singlets reports DNA that replicates late.

This method can give a wide variation in results, for example, the p53 locus data studied in a normal lymphoid cell line given by Amiel *et al* (Amiel, Litmanovitch et al. 1998) for three different individuals is reported as 47% double singlets (ss), 80%ss and 62%ss. There is also an inherent inaccuracy in the method as the assay relies on being able to resolve a doublet signal from a singlet. Therefore the DNA must separate enough after replication to resolve the doublet signal. Recent evidence suggests that the time between replication and sister chromatid separation is different at different loci (Azuara, Brown et al. 2003). This is due to the selective holding together of sister chromatids by specific protein complexes. Therefore this assay does not measure replication timing as previously thought, but measures sister chromatid separation. This raises the possibility that previous FISH based assays of replication timing may have overestimated the number of late replicating loci.

1.3.3: Replication Timing by flow sorting and PCR

One way to increase resolution is to compare the relative abundance of specific sequences of nascent DNA at different stages of the cell cycle (Gilbert 1986, Hassan and Cook 1993, Sinnett, Flint et al. 1993). Cell cultures in the exponential phase of their growth are pulse labelled with BrdU which is incorporated into the newly synthesised DNA. The nuclei are then stained with propidium iodide and equal numbers of nuclei are sorted into four S phase fractions (Figure 1.10 (Azuara, Brown et al. 2003)). Nascent DNA is then extracted by immunoprecipitation with anti BrdU. The quantity of newly synthesised DNA in each fraction is detected by semi-quantitative PCR with primers specific for the loci of interest. Fractions with the most nascent DNA have the most intense band when the PCR products are run on a gel. This will narrow down the time of replication to one of the four S phase fractions.

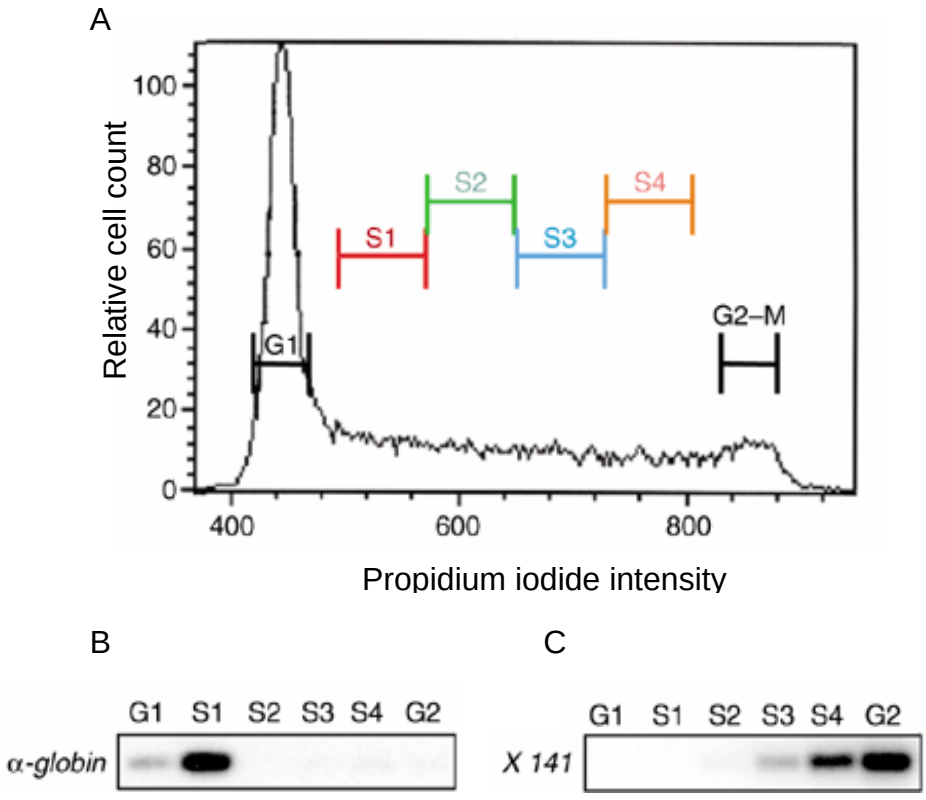


Figure 1.10: (Adapted from Azura et al, 2003) A: Cell cycle profile showing gate positions required to sort the nuclei into four separate S phase fractions. B. Gel photograph illustrating the enrichment in the S1 fraction for an early replicating locus. C. Gel photograph illustrating the enrichment in the S4/G2 fraction for an early replicating locus.

This method is limited by the number of fractions into which S phase can be accurately sorted. It will only therefore, give an approximation of the time of replication, placing the replication time for each locus within a specific quartile of S phase, or between two S phase fractions. This technique is also very labour intensive and is mainly used to screen small regions of the genome. However, recently the technique has been used to screen whole chromosome arms. Wanatabe and co-workers published a replication timing profile for chromosome arms 11q and 21q (Watanabe, Fujiyama et al. 2002). The average resolution sampled in the chromosome 11q data was 300Kb and in the chromosome 21q data this was increased to 200Kb (Figure 1.11). This study shows how the replication timing along the chromosome arm is linked to the GC content and cytogenetic banding. However the method is still limited as S phase was only sorted into four fractions so again the replication time given is an approximation.

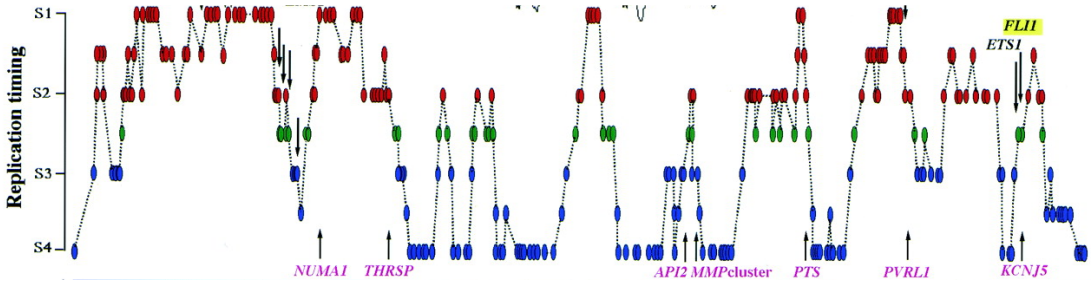


Figure 1.11: Replication timing profile of chromosome 11q using quantitative PCR. Adapted from Figure 3, (Watanabe, Fujiyama et al. 2002).

Ideally, methods for assessing replication timing would sample large regions of the genome at a high resolution and accuracy. To this end, I have investigated the use of genomic arrays to calculate the copy number change associated with DNA replication. (Section 1.6)

1.4: Temporal control of Replication Origin Activation

1.4.1: DNA Replication Timing and Correlation with Sequence Features.

Table 1.1 summarises how genome features such as GC content and gene density are related to the type of chromatin. The sequencing of the human genome (IHGSC 2001) allowed large scale analysis of previously hypothesised links between GC content, gene density, repeat content, cytogenetic banding and recombination rate. Publication of the human genome sequence (IHGSC 2001)of the human genome revealed that 98% of clones mapping to the darkest G-bands have a low GC content (average 37%), whereas more than 80% of clones located in the lightest G-bands are in regions of high GC content (average 45%).

Correlations have also been reported between GC content, cytogenetic banding and replication timing. Pulse labelling replicating cells with Bromodeoxyuridine) and examination of harvested metaphase cells (section 1.3) reveals a banding pattern of replicating DNA. The early replicating, BrdU incorporating, bands correlate with the cytogenetic G-light (GC-rich) bands (Latt 1973; Drouin, Lemieux et al. 1990).

Studies at individual G dark-G light boundaries correlate early replication with a high GC content. A high resolution study assessing replication timing using quantitative PCR analysed the boundary between the G light 13q14.3 and the G dark 13q21.1. The G light side of the boundary was shown to replicate early, whilst the G dark side of the boundary replicated late. Analysis using PCR primers spaced at approx 150Kb intervals showed that the switch in replication timing happened gradually from early-mid-late over a 1-2Mb region, rather than an abrupt change to coincide with the G light-G dark boundary (Strehl, LaSalle et al. 1997).

A further study over the major histocompatibility complex (MHC) region on chromosome 6p21.3 has also shown a change in replication with a progression from a GC poor to GC rich region. Figure 1.12 (Tenzen, Yamagata et al. 1997) shows a 450Kb region across the MHC Class II (GC-poor) and MHC Class III (GC-rich).

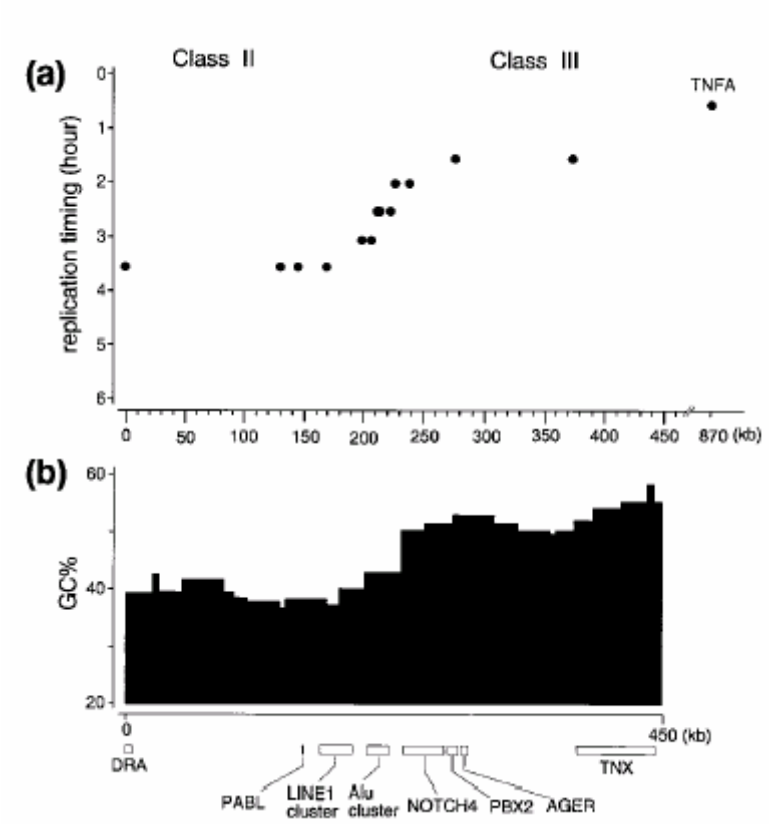


Figure 1.12: Change in Replication Timing (a) and GC content (b) across a 450Kb region of the MHC Class II and Class III taken from (Tenzen, Yamagata et al. 1997).

The Class III region of the MHC replicates about an hour and a half into S phase (after the cells have been removed from an aphedicolin block). The Class II region of the MHC replicates about two hours later (3.5 hours after release from an amphedicolin block). It can be seen that the transition between the change in replication timing and the change in GC content occur at the same point. The transition takes place within a zone of about 100kb, with loci in the transition zone undergoing replication at a mid time point.

The correlation between replication timing and GC content has also been observed over an entire chromosome arm (Watanabe, Fujiyama et al. 2002). Replication timing was assessed on chromosome 11q and chromosome 21q at a resolution of 300Kb (11q) and 200Kb (21q), using flow sorted S phase fractions and PCR (section 1.3). Wanatabe *et al* described a general correlation between replication timing and GC content on both chromosome arms. Zones of early replication were more GC rich than the late zones, although they did observe that the correlation reduced in atypical

regions of the chromosome arms (the pericentric and telomeric regions). The chromosome arm data was also used to study many transition zones between early and late replication (or vice versa). The data showed that the transitions in replication timing were identical, or very close to, regions showing a transition in GC content. The two chromosome arm profiles (Figure 1.13) show the correlation between replication timing and GC content.

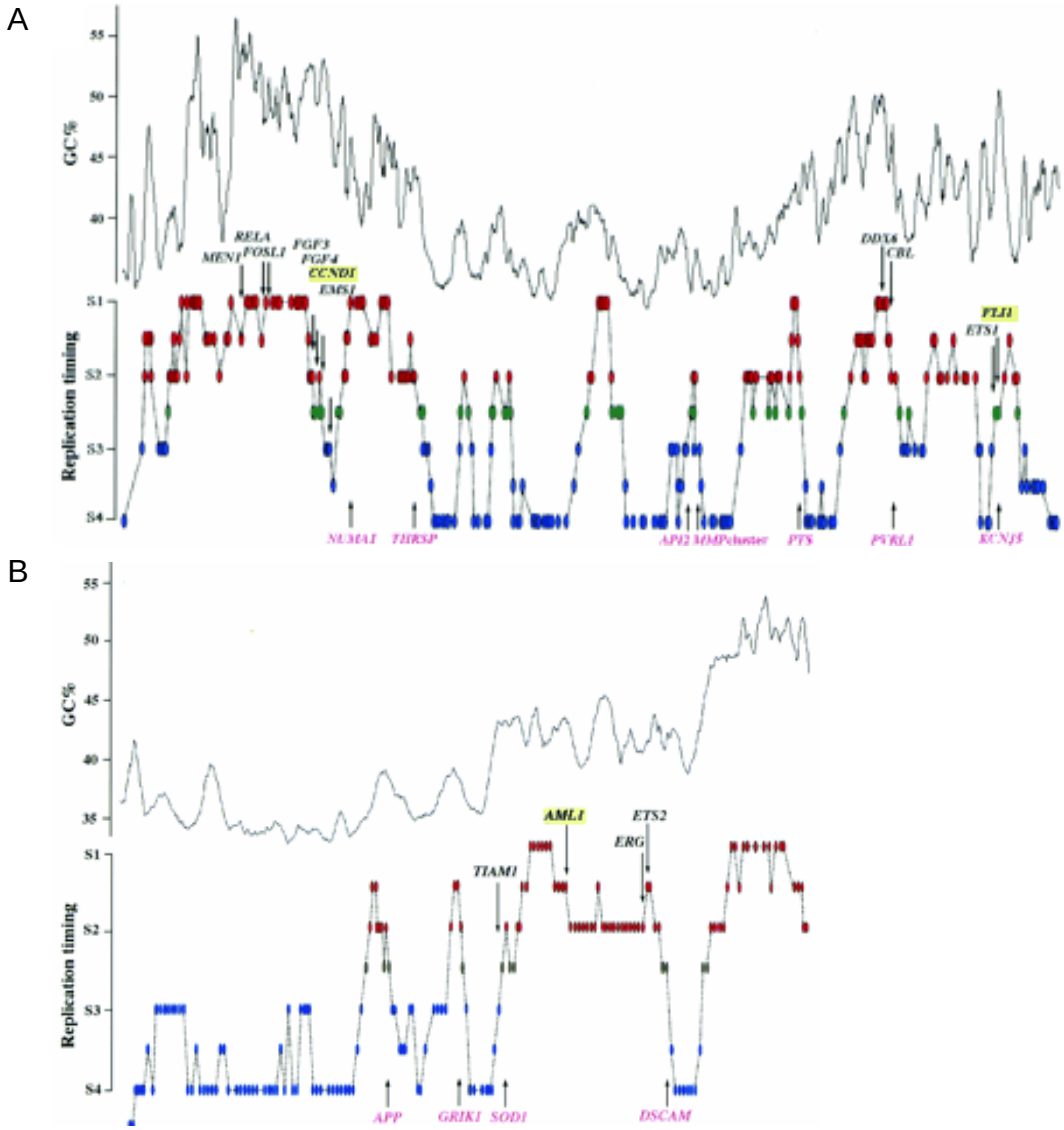


Figure 1.13: The correlation between GC content and replication timing profile on Chromosome 11q (A) and Chromosome 21q (B) (Watanabe, Fujiyama et al. 2002).

The overall correlation between replication timing and GC content can clearly be seen, however it can also be observed that the correlation is not absolute and changes in replication timing can be associated with small local change in GC content.

Many of the loci reporting an atypical relationship between GC content and replication timing are located at the centromeres and telomeres of chromosomes. Heterochromatic centromeric or telomeric DNA contains arrays of repetitive sequence. In *Saccharomyces cerevisiae* all telomeres have been shown to be late replicating. However, genome-wide analysis of the higher eukaryote *Drosophila melanogaster* showed that the euchromatin located close to either the centromere or the telomere was not found to be late replicating (Schubeler, Scalzo et al. 2002). Closer analysis of the centromere on the *Drosophila* chromosome 2L revealed that genes located in the β-heterochromatin of the centromere were early replicating, although they are not transcribed. Study of α-satellite DNA at the centromeres of human chromosomes reveals that although the centromeres replicate at slightly different times in the cell cycle, they all replicate in a narrow window during late S phase (Hultdin, Gronlund et al. 2001).

The investigation of human telomeres located on 22q and 16p13.3 have shown that in common with what was seen in yeast, some human telomeres are late replicating (Smith and Higgs 1999, Ofir, Wong et al. 1999). However, studies of other telomeric regions in the human showed that human telomeric sequences, like those in *Drosophila*, replicate at variable times (Hultdin, Gronlund et al. 2001, Ten Hagen, Gilbert et al. 1990).

Closer analysis of the replication timing of 325Kb of telomeric DNA from 16p13.3 using a FISH based assay (Selig, Okumura et al. 1992) showed the GC rich region lying in the most centromeric region of the 325Kb studied contained widely expressed genes and was early replicating. The subtelomeric 20Kb of the sequence was late replicating. Movement of early replicating DNA adjacent to the heterochromatic telomeric repeats delays the replication of the inserted sequence (Smith and Higgs 1999).

In summary, replication timing is correlated with GC content. The study of repetitive DNA sequence features reveals the heterochromatic telomeric DNA does not seem to

have a specific replication time in human cells. However, examination of the repetitive centromeric DNA shows a clear bias towards late replication.

1.4.2: DNA Replication Timing and Correlation with Chromatin and Epigenetic Features

The investigation of the replication timing in the human β-globin locus correlated replication timing with the open structure of the chromatin. This suggested that chromatin conformation was important in replication timing and that there is long range control of both origin choice and replication timing at the human β-globin locus (Cimbora, Schubeler et al. 2000).

Section 1.1 describes how chromatin condensation in the interphase nuclei is a highly ordered process. Chromatin conformation is associated with a variety of epigenetic features such as acetylation and methylation of histone proteins within the nucleosome and methylation of the CpG dinucleotide at CpG islands. The epigenetic status of the chromatin is reflected in its replication timing. It has long been acknowledged that the tightly condensed, epigenetically silenced, transcriptionally inert heterochromatin is late replicating (Holmquist 1987). In contrast the loosely coiled, transcriptionally active chromatin replicates early. This is particularly evident in the female mammalian X chromosomes and study of alleles that are asynchronously replicated (Discussed further in 1.4.4). The early replicating allele is usually hyperacetylated and the CpG islands are hypomethylated. In contrast the late replicating allele is transcriptionally silent, with methylated CpG islands and hypoacetylated histone proteins. The study of X inactivation in embryonic stem cells reveals that the change in replication timing (to late replication) of the inactive X chromosome is a relatively early event, taking place after the coating of the inactive X with Xist RNA but before changes in histone acetylation, or methylation of CpG islands at promoters occurs (for further details see 1.4.4 – (Avner and Heard 2001)).

The completion of S phase is vital for the complete condensation of the nuclei in M phase of the cell cycle. Replication mutants that are unable to complete S phase have condensation defects (Gatti and Baker 1989). ORC is just one of the proteins that are

important in both DNA replication and DNA condensation. The mitotic chromosomes of *Drosophila* ORC mutants are shorter and thicker than wild type chromosomes (Pflumm and Botchan 2001). Although some levels of chromosome condensation are possible without complete replication it is thought that complete replication is important for lengthwise compaction (Pflumm 2002). This is supported by the longer, less compact mitotic chromosomes observed during embryogenesis. During this time there are many more replication origins present to support rapid cell division. This suggests replication timing may alter in nuclei with chromatin compaction defects. During development, as the loop size between nuclear attachments is increased, the metaphase chromosomes become shorter and thicker (Wang, Castano et al. 2000).

1.4.3: DNA Replication Timing and correlation with Nuclear Position

As discussed previously in section 1.4.1, the correlation between replication timing and DNA sequence indicates there are similar regions of replication timing across the genome. Regions of similar replication timing are found in bands that correlate with the G-bands. There is a further correlation between replication timing and spatial relationship i.e. position within interphase nuclei.

In Figure 1.14, taken from Ferreira *et al* (Ferreira, Paolella et al. 1997). Nuclei were pulse labelled one hour before harvest and at different times after synchronous release into S phase, five different replication patterns are shown. Five different replication patterns can be seen (A-E). A and B show very early replicating DNA in foci dispersed throughout the internal interphase environment stained white on these micrographs. Little replication occurs towards the periphery of the interphase nuclei or adjacent to the nucleolus (The position of the nucleolus can be seen in the corresponding DIC images – F-J). As the interphase nuclei progress through S phase the pattern of replication changes. Figure 1.14 C shows some replication still occurs in internal foci, but most is localised adjacent to the nuclear membranes and the nucleolus. Finally in stages D and E it can be seen that no replication occurs within the internal nuclear environment, with all replication occurring adjacent to the nuclear membrane.

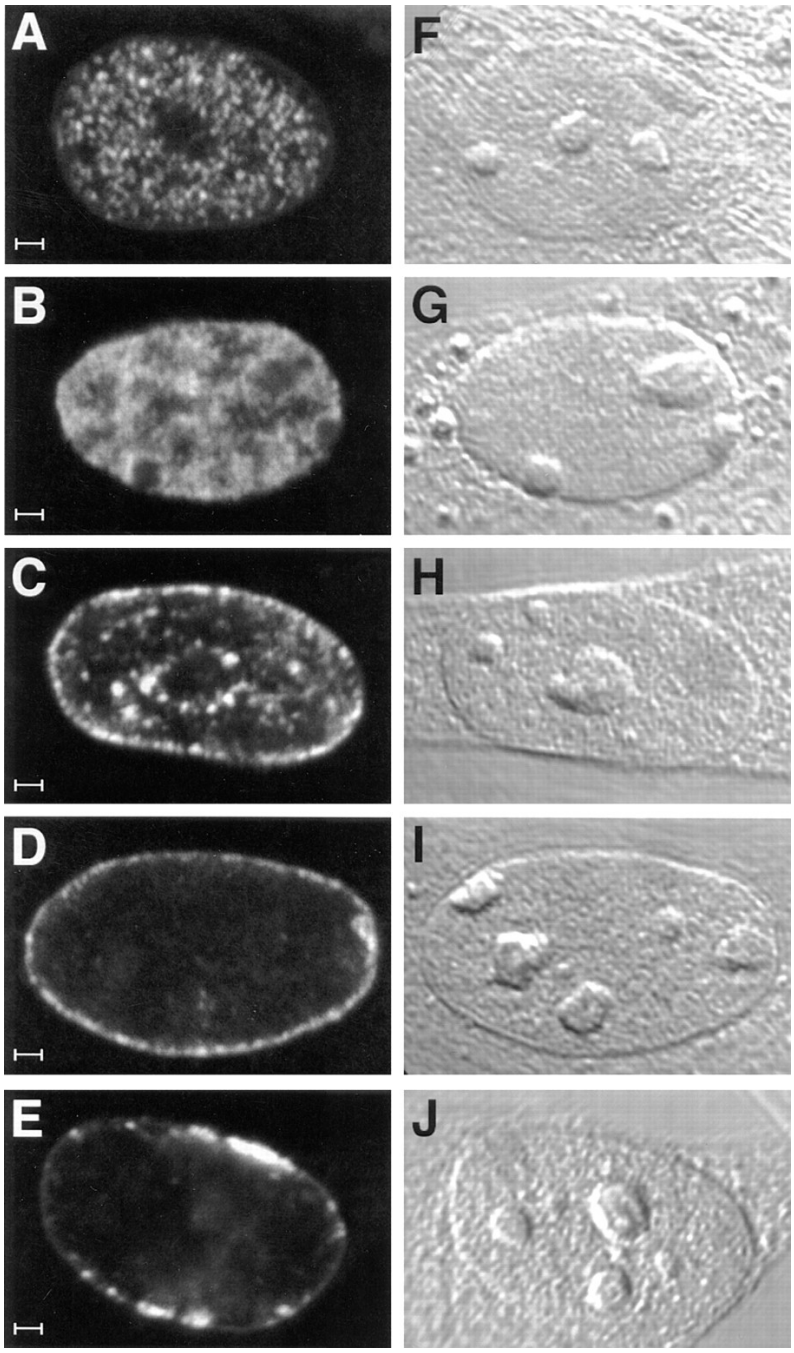


Figure 1.14: Different stages of replication in the interphase nucleus. Figure taken from (Ferreira, Paolella et al. 1997). For details see text.

The localisation of early replicating DNA in the internal nuclear environment and late replicating DNA to the periphery of interphase nuclei has also been visualised by differential pulse labelling of early and late replicating DNA (Schermelleh, Solovei et al. 2001). In Figure 1.15, early replicating DNA has been labelled in blue, whilst mid-late replicating DNA was labelled in red by the incorporation of differentially labelled nucleotides. Again it can be seen that early replication occurs in foci in the inside of

the interphase nuclei, whilst late replication occurs adjacent to the nucleolus and nuclear envelope. Each replication focus contains approximately 0.25 - 1.5 Mb of DNA and the replication machinery required. Each focus takes approximately one hour to replicate (Cremer and Cremer 2001). The foci that are late replicating and localise to the nuclei periphery have also been confirmed as containing AT-rich DNA located in G dark bands (Zink, Bornfleth et al. 1999). Early and late replicating DNA occupy distinct foci within the interphase nuclei. The median overlap between late and early DNA location being only 5-10% (Zink, Bornfleth et al. 1999).

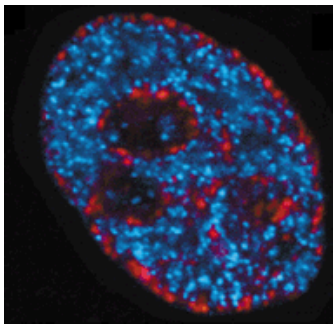


Figure 1.15: Localisation of early replicating DNA (blue) and late replicating DNA (red) in an interphase nuclei (Schermelleh, Solovei et al. 2001).

Similar labelling experiments were carried out on nuclei that were also stained for the protein lamin B. Lamin B is a component of the nuclear lamina and localises to the nucleoplasmic side of the nuclear envelope. It can be seen that late replicating DNA co-localises with lamin B at the nuclear envelope (Figure 1.16)

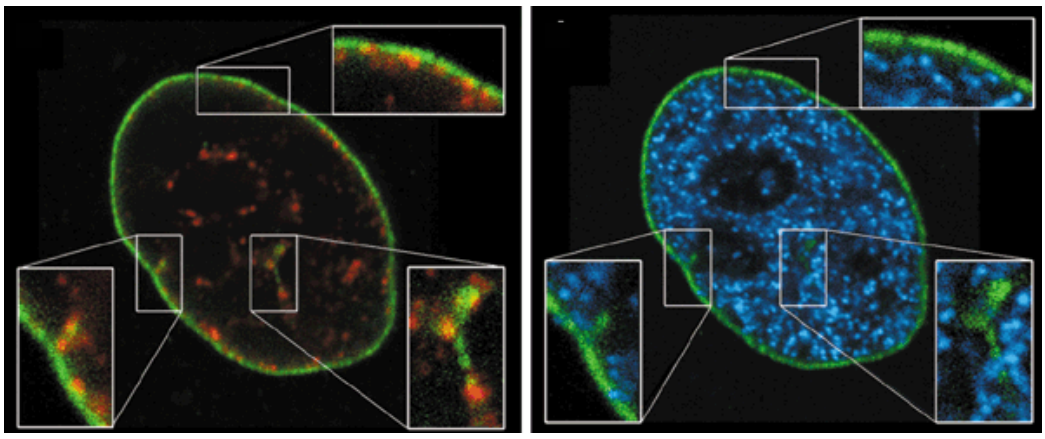


Figure 1.16: Mid-late replicating chromosome domains (red) associate with lamin B (green). Early replicating DNA (blue) does not (Schermelleh, Solovei et al. 2001).

The position of the chromatin within the interphase nuclei is established during early G1 phase of the cell cycle (Gilbert 2001). Studies in Chinese hamster ovary cells show that DNA attached to the nuclear matrix during the G1 phase of the cell cycle is enriched in replication origins (Djeliova, Russev et al. 2001). DNA replication occurs on the nuclear matrix and after replication has occurred the DNA disassociates. Replication origins are therefore transiently attached to the nuclear matrix, associating with replication origins in G1 and disassociating during the S phase. The establishment of the interphase position is co-incidental with the establishment of DNA replication timing as described in 1.2.3.

The position of the immunoglobulin heavy chain locus (IgH) in B cells shows that localisation in the interphase nuclei is dependant on replication timing and gene activity. During early stages of B cell development, the IgH locus is early replicating in both alleles and is maintained away from the nuclear periphery in the centre of the interphase nuclei. Once in the centre of the interphase nuclei VDJ recombination and germ line transcription of the IgH locus occur. In the later stages of the B cell development germ line transcription of the IgH locus is turned off. The entire locus no longer replicates early and the IgH locus is localised to the periphery of the interphase nuclei. The peripheral position of the IgH locus may ensure that the DNA can only replicate at the end of S phase (Zhou, Ermakova et al. 2002). Current data suggests that perinuclear position is indicative but not sufficient for late replication.

In summary, there is clearly an association between the position of DNA within the G1 and S phase nuclei and the time of replication within S phase. The position of the early replicating DNA away from the periphery of interphase nuclei is important. It is also significant that early replicating DNA is attached to the DNA matrix. This ensures the early replicating DNA is in a position favourable to replication when the cell enters S phase.

1.4.4: Asynchronous DNA Replication.

Loci at which expression occurs from both alleles replicate synchronously. Each allele on the two sister chromosomes replicate at the same point in the S phase of the cell cycle. This is not true for loci where expression is monoallelic, as these replicate asynchronously (Mostoslavsky, Singh et al. 2001), (Goren and Cedar 2003). Examples include imprinted regions, the X Chromosomes in females, immunoreceptor genes and genes encoding olfactory receptors (Singh, Ebrahimi et al. 2003).

Early evidence for the asynchronous replication timing in monoallelically expressed regions came from a study of the imprinted Prader-Willi syndrome critical region on 15q11.2 (Izumikawa, Naritomi et al. 1991). Replication banding studies (such as those described in 1.3.1) showed replication asynchrony between homologues of 15q11.2 in about 40% of individuals (Izumikawa, Naritomi et al. 1991).

The asynchronous replication timing of imprinted chromosomal regions seemed to be confirmed at a higher resolution by the characterisation of this region using the FISH assay described in 1.3.2 (Selig, Okumura et al. 1992). Imprinted regions showed a much higher proportion of nuclei displaying the ‘one singlet one doublet’ hybridisation signal (Figure 1. 9c (Goren and Cedar 2003), (Kitsberg, Selig et al. 1993). However, due to the limitations of this assay described by Azuara and colleagues (Azuara, Brown et al. 2003) the assessment of replication timing by the FISH assay may only show the difference between sister chromatid separation and not a difference in replication timing.

Nevertheless, replication asynchrony could be confirmed at a higher resolution using a different method at the human Igf2 loci, located on the imprinted region in chromosome 11p15 (Simon, Tenzen et al. 1999). Replication timing was assessed using quantitative PCR on flow sorted S phase fractions. Restriction site polymorphisms were used to distinguish the maternal and paternal alleles and one chromosomal copy was seen to replicate before the other.

Asynchronous replication is also witnessed in the female X chromosome in mammals. Females have two X chromosomes whereas males only have one X chromosome. To avoid any X chromosome gene dosage imbalance, one of the female X chromosomes is modified in the late blastocyst to become inactive. This inactivation involves chromosome-wide epigenetic changes, making the DNA chosen for inactivation transcriptionally inert. There is also a shift to a later replication time for the inactive X chromosome, whilst the active X chromosome retains its original replication time. This is one of the first developmental changes involved in X inactivation and precedes histone hypoactylation and DNA methylation. About 15% of genes on the X chromosome escape inactivation. It has been suggested that LINE repeats are involved in the propagation of X inactivation along the chromosome. Regions with a lower density of LINE repeats than the rest of the chromosome escape inactivation. (Avner and Heard 2001).

A third category of monoallelically expressed genes are found on autosomes, but unlike imprinted genes the pattern of expression is independent of the parent of origin. These genes include members of the family of olfactory receptors, or encode immuno-receptor genes. In both these systems, a wide range of receptors are encoded in the genome but it is important that only one is expressed in each individual cell (In olfactory receptors this is important for sensitivity to different aromas and in immunoreceptor genes this is important for the clonal development of B cells). An olfactory receptor neuron contains genes for more than 1,000 receptors found within clusters throughout the genome; however, only one is expressed on the cell's surface. To achieve this, the clusters not being expressed are epigenetically silenced, and only one parental allele is expressed in each cell. Part of this epigenetic silencing is reflected in a transition to late replication (Chess, Simon et al. 1994, Singh, Ebrahimi et al. 2003). The genes encoding antigen receptors are also monoallelically expressed. The late replication is randomly established within the early embryo and is maintained by the clonal development of these cells. In B cells it is predominantly the early replicating allele that undergoes rearrangement and ultimately, expression (Mostoslavsky, Singh et al. 2001).

In all three examples of monoallelic expression almost all cases show that the allele that replicates early is the allele that is expressed. This provides a possible link between replication timing and gene expression.

1.5: DNA Replication Timing and Correlation with Transcription

In higher eukaryotes DNA replication timing is thought to correlate with transcriptional activity. Studies of regions of the genome suggest that early replicating DNA is transcriptionally active. Conversely late replicating DNA is transcriptionally inert (Holmquist 1987). Experiments on HeLa cells have compared the patterns of transcription and replication (Hassan, Errington et al. 1994), (Hassan and Cook 1994). Optical sections were taken through the cell cycle of HeLa cells at different stages of the cell cycle. Transcription was indicated by labelling with Texas Red. Replication is indicated by labelling with fluorescein (green) as shown in Figure 1.17. During mitosis the cell is black as no transcription or replication occurs. At sites where replication and transcription occur together the signal ranges from purple to white, as intensities increase. (Figure 1.17) Transcriptionally active DNA was found to replicate in early S phase.

The few ubiquitously expressed housekeeping genes that have been assayed for replication timing have all been located within early replicating G light bands of chromosomes. On the other hand, tissue specific genes are located within G dark bands and almost always replicate late except when the tissue specific region of the genome is expressed; the gene then becomes early replicating. An example of this is the β-globin gene. This gene cluster is in a 200-300Kb stretch of hypoacteylated, DNase-1 resistant late replicating chromatin; however in erythroblasts, which can be induced to express β-globin, the 1Mb of chromatin surrounding the β-globin gene becomes early replicating (Gilbert 2002), (Cimbora, Schubeler et al. 2000).

This phenomenon is particularly evident in developmentally significant genes. Of those that have been studied, at the stage in development in which the gene is expressed it also replicates early. However, once the gene is no longer expressed the replication timing of the gene changes and is delayed (Nothias, Miranda et al. 1996).

It has been proposed that in higher eukaryotes replication timing is a developmentally regulated process that is closely associated with gene expression (Holmquist 1987).

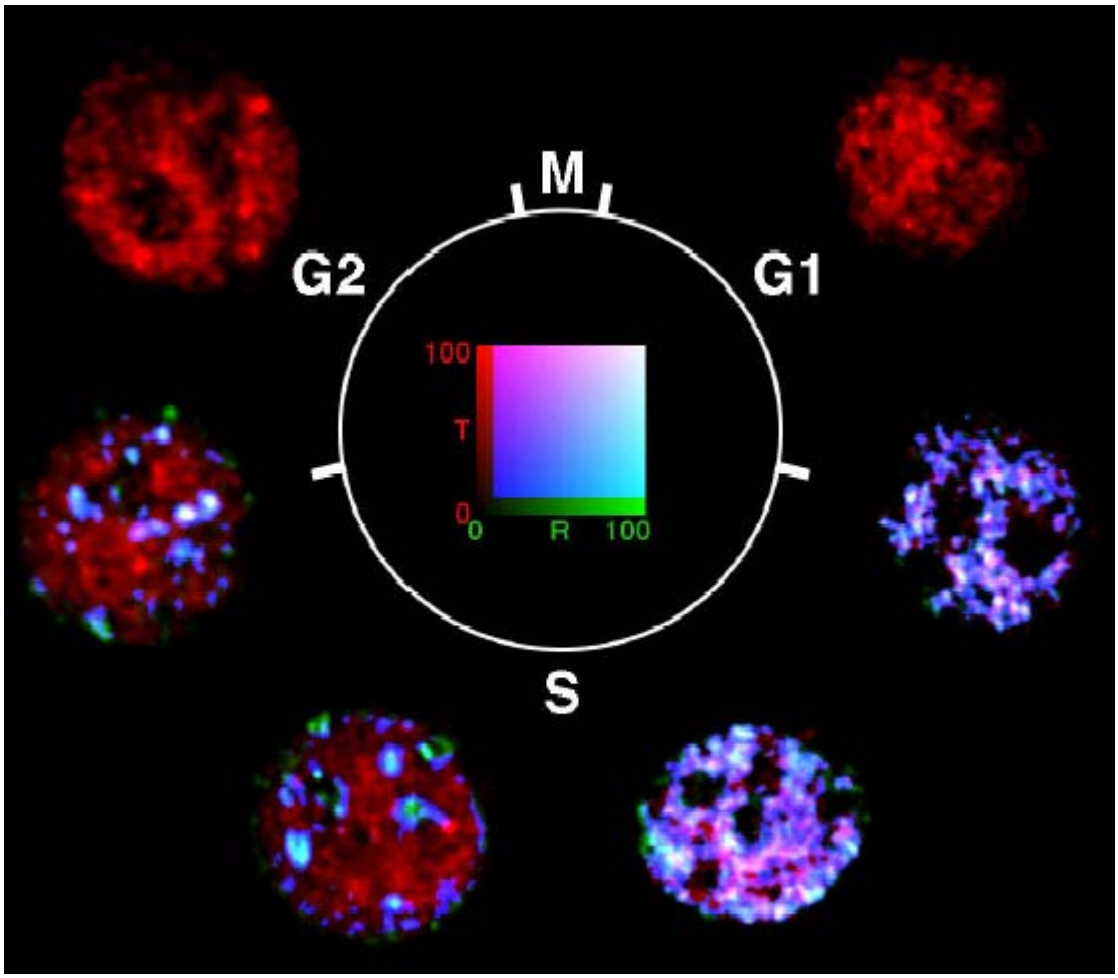


Figure 1.17: Co-localisation of DNA that is replicated in early S phase and transcriptional activity. For details see text (taken from (Cook 1999)).

The above studies have been restricted to the few loci that have been assayed. Further studies have addressed the question whether a correlation between gene transcription and replication timing observed at individual loci is observed at a global level across the entire genome using microarray technology, described in section 1.6.2. The study of the *Saccharomyces cerevisiae* genome in this way revealed no relationship between transcription and replication timing (Raghuraman, Winzeler et al. 2001). However, when a similar study was performed on the genome of the higher eukaryote *Drosophila melanogaster,* a correlation between replication timing and the probability of gene expression was found (Schubeler, Scalzo et al. 2002). If a gene was located within an early replicating region of the genome it is more likely to be expressed. This

relationship was found to be highly significant ($p = 10^{-48}$), but the correlation observed was not absolute; 20% of the earliest replicating genes are found to be transcriptionally silent, and conversely more than 20% of late replicating genes were found to be expressed in the cell line used for the study.

Replication timing of the genome is an important level of organisation in the eukaryote nucleus. The chromatin in higher eukaryotes is not as mobile as the chromatin in yeast (Gilbert 2002). As a result it is important that the transcriptionally active chromatin is available to the transcription machinery. The early replicating DNA must also be available to have access to the replication enzymes. As a result the early replicating, transcriptionally active DNA is found associated with open euchromatin. The question remains whether early replicating DNA leads to transcriptionally active chromatin, or vice-versa. Early replication is currently thought to be necessary, but not sufficient for gene transcription.

Reported genes injected into early S phase nuclei were found to be more than ten times more transcriptionally active than the same gene injected into late replicating S phase nuclei (Zhang, Xu et al. 2002). These transcriptional states remain stable as the cell continues cycling. The reporter genes injected into the early replicating DNA were packaged into chromatin containing acetylated histones, whereas late-injected genes were hypoacetylated (Zhang, Xu et al. 2002), (Goren and Cedar 2003). This reveals that the acetylation status of a region is correlated to the time of its replication.

Two models have been proposed to explain the relationship between early replication and transcriptional activity. Model 1 (Figure 1.18a) suggests that transcriptional potential is established by the early replication of DNA. During S phase, specific activating proteins are available for incorporation into the chromatin, aiding transcription, conversely during late S phase the proteins that are available produce transcriptional regression (Gilbert 2002), (Gilbert 1986). The second model (Figure 1.18b) suggests that closed, transcriptionally-inert chromatin and heterochromatin postpone the commencement of replication by confining the access of replication proteins to the chromatin in the early part of S phase (Gilbert 2002), (Stevenson and Gottschling 1999).

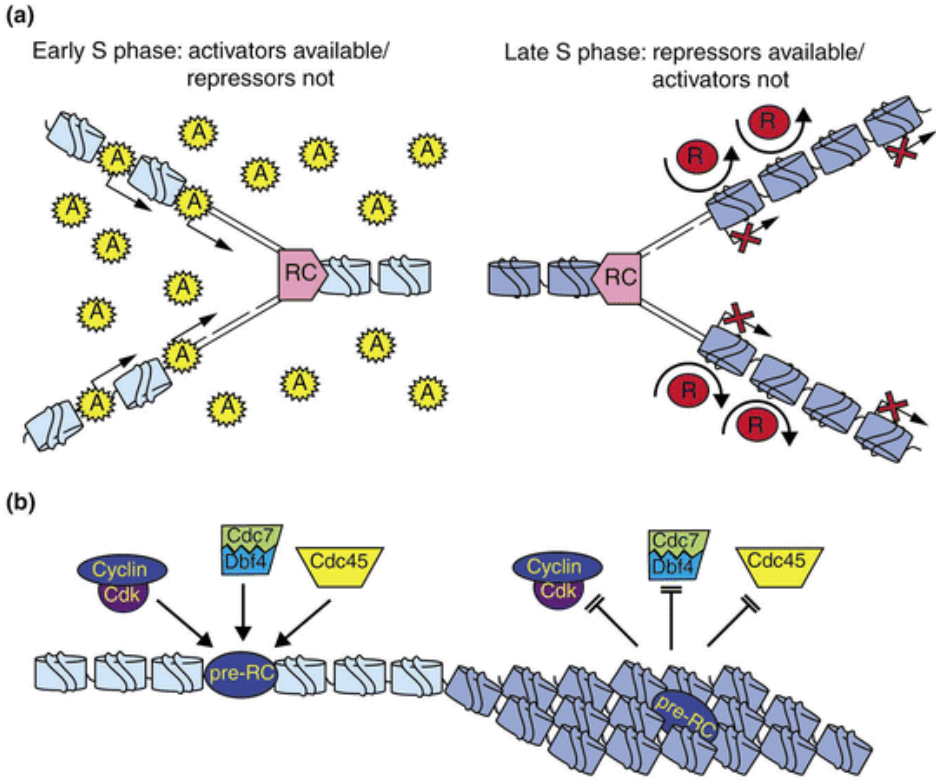


Figure 1.18: Models for linking transcription and replication. A: transcriptional activators are recruited to the chromatin during early S phase, conversely transcriptional repressors are recruited to the chromatin during late S phase. B: Replication initiators can access open, transcriptionally active chromatin first, in early S phase, whilst initiation of origins in heterochromatin is delayed due to inaccessibility of the replication origins (Gilbert 2002).

These models are not mutually exclusive, and data has been found to support both. The first model is supported by the work of Rountree et al (Rountree, Bachman et al. 2000). They illustrate that transcriptional repressors are recruited into the chromatin during late (but not early) S phase. DNA methyltransferase localises to sites of nascent DNA replication in late S phase and recruits histone deacetylases. The deacetylation of the chromatin promotes transcriptional repression.

The second model is supported by late replicating telomeres in yeast. Stevenson *et al* (Stevenson and Gottschling 1999) showed that the condensed telomeric DNA has an inhibitory affect on the replication timing of this region via the Sir-3 protein, which also has an affect on suppressing transcription at the telomeres.

The co-localisation of replication factories and transcriptional machinery in the interphase nuclei may also account for the correlation between early replication and transcriptional activity. Conventional thinking suggests that RNA and DNA polymerases move along the stationary DNA during replication and transcription. However experiments in rat fibroblasts in the S phase of the cell cycle show that the incorporation of BrdU into nascent DNA is not distributed evenly throughout the nucleus, but seems to be focused in approximately 150 distinct sites (Nakamura 1986), (Cook 1999). This supports results described in section 1.3.3, which also show replication foci in early S phase (Ferreira, Paolella et al. 1997; Schermelleh, Solovei et al. 2001). Evidence suggests that the newly synthesised DNA is attached to the nuclear matrix as these foci remain when chromatin is removed (Hassan and Cook 1993). Electron microscopy confirmed the attachment of large DNA replication factories to a diffuse nuclear cytoskeleton (Hozak, Hassan et al. 1993). Similar experiments revealed that RNA polymerases are also attached to the nuclear matrix (Hozak, Jackson et al. 1994), (Cook 1999).

In HeLa cells, fluorescence studies showed a near perfect overlap of sites of transcription and sites of replication (Hassan, Errington et al. 1994) Figure 1.17). One model proposed to describe the coincidence of transcription and replication factories on the nuclear cytoskeleton suggests that during the G1 phase of the cell cycle a replication factory is assembled around two transcription factories (Figure 1.19). This may bring an origin close to a factory so that it can attach and permit replication. In this way the DNA that is close to actively transcribed genes is already located close to replication factories so this DNA is synthesised early ((Hassan and Cook 1994), (Cook 1994; Cook 1995)).

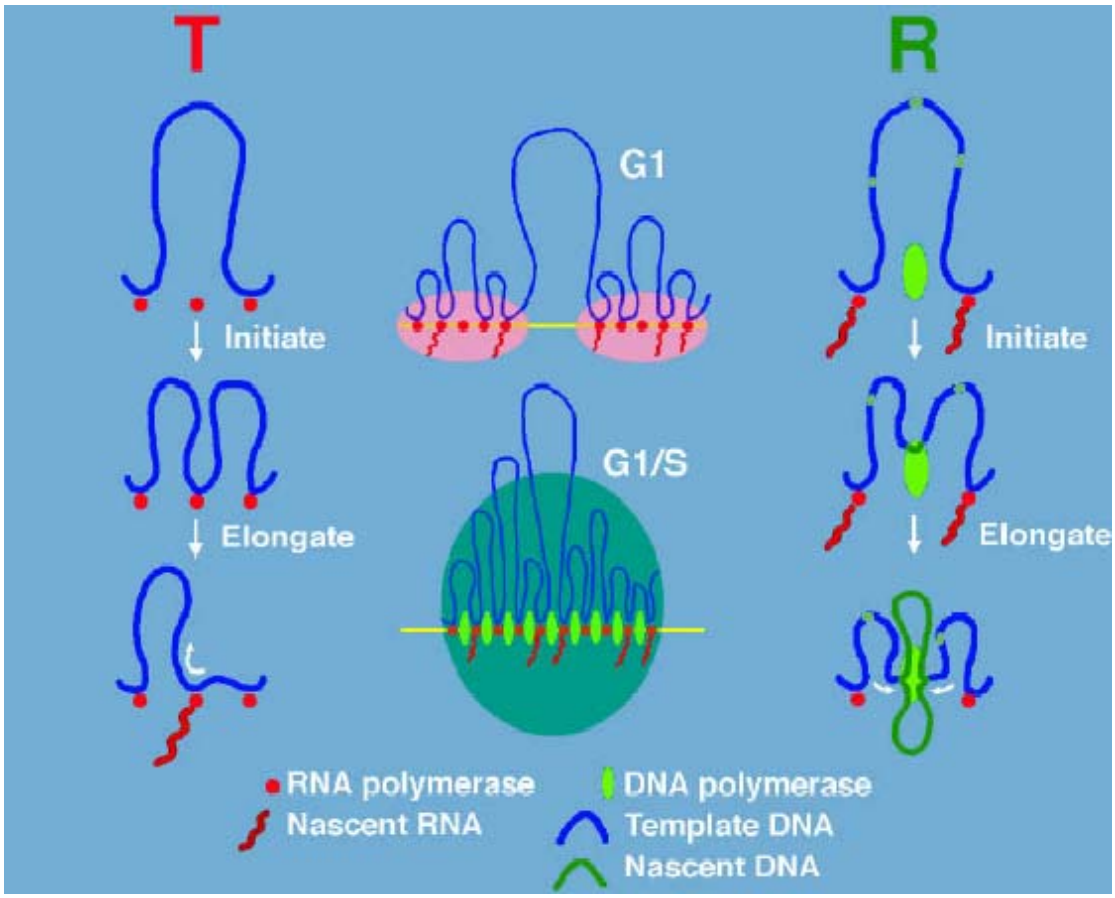


Figure 1.19: The initiation of replication from transcription factories. During transcription (left) DNA is looped to attach to DNA polymerases. The template slides past the fixed enzyme factory to continue transcription. Two of these transcription factories come together to form a replication focus at the G1/S phase boundary. DNA replication is then initiated as an origin of replication binds to a DNA polymerase (right). Template DNA slides along the DNA polymerase as loops of semi-conservative newly replicated DNA are formed (Cook 1994; Cook 1995).

From these studies, many questions remained unanswered concerning the association between replication timing and gene expression. For instance it is unknown if early replication drives the transcriptional activity of a region or vice versa. It is clear however, that other epigenetic factors have a role in the coincidence of early replication and transcriptional activity. The role of these epigenetic factors at sites of replication and transcription will have to be determined before it is possible to understand how DNA replication and RNA transcription interact.

1.6: Using Genomic Arrays to investigate Copy Number Changes.

As stated in section 1.3, an ideal way to assay replication timing would be to assay large regions with high accuracy. In this thesis I describe how replication timing can be assayed using genomic microarrays using their ability to quantify copy number differences between a test and a reference sample, as described below.

1.6.1: Using Genomic Arrays to investigate Chromosomal Copy Number Changes.

Comparative genomic hybridisation was developed in 1992 as a way to detect DNA copy number changes in DNA samples (Kallioniemi, Kallioniemi et al. 1992). The principle of this procedure is co-hybridisation of differentially labelled test and reference DNA onto normal metaphase target chromosomes and measurement of the test to reference fluorescence ratio along the chromosomes. Deviation from the 1:1 ratio of the intensities of test: reference DNA indicated either a copy number gain or copy number loss in the test DNA. However, the detection of copy number changes was limited to the resolution of the signals on the metaphase chromosomes. Low copy number gains and losses can only be resolved if larger than 3-5Mb (Kallioniemi, Kallioniemi et al. 1992).

This principle has more recently been combined with microarray technology to detect copy number changes at a higher resolution (Solinas-Toldo, Lampel et al. 1997; Pinkel, Segraves et al. 1998; Albertson, Ylstra et al. 2000). The metaphase chromosome targets are substituted with nucleic acid target sequences spotted in an array onto a glass slide. The target sequences are obtained from mapped and cloned DNA. As a result, the resolution of the arrays is only limited by the size of the clone and the number of clones represented on the array. The size of the targets that were originally used ranged from 40Kb for cosmid clones, to a maximum of 300Kb for bacterial artificial chromosomes (BACs) (Solinas-Toldo, Lampel et al. 1997). A genome-wide scanning array has been produced which samples the genome with an average resolution of 1.3Mb (Snijders, Nowak et al. 2001). This has been refined to an average resolution on 1Mb using DNA from BACS, PACS and cosmids (Fiegler, Carr et al. 2003). The advent of microarrays utilising overlapping golden path sequencing clones (IHGSC 2001) have allowed the study of whole chromosome arms

with the resolution limited only to the size of the clones and the extent of their overlap (Buckley, Mantripragada et al. 2002).

Originally, to extract enough DNA from clones to spot onto an array large amounts of clone culture needed to be grown (Pinkel, Segraves et al. 1998), (Solinas-Toldo, Lampel et al. 1997, Albertson, Ylstra et al. 2000). To avoid the logistical problems involved in growing large amounts of culture for each locus on the array, methods have been developed to amplify DNA from small amounts of clone DNA, yet still ensuring that the full sequence within the clone is covered. This has been achieved using rolling circle amplification, linker adapter PCR and by DOP-PCR. Rolling circle PCR amplification of the clone DNA utilises the proof reading phi 29 polymerase (Buckley, Mantripragada et al. 2002). Ligation-mediated PCR which produces representative amplification of the genome from just a single nucleus (Klein, Schmidt-Kittler et al. 1999; Snijders, Nowak et al. 2001). DOP-PCR uses amplification of clone DNA by three different, specifically designed degenerate oligonucleotide primers (DOP). This not only ensures the complete amplification of the clone DNA but also ensure that there is minimum contamination from the *E. coli* host vector DNA (Fiegler, Carr et al. 2003).

One problem with using clones from the golden path is that they can contain a high amount of repetitive DNA sequence. This can lead to cross hybridisation with other regions of the genome. This problem is negated by the inclusion of Cot 1 DNA in the hybridisation mix.

A second problem is the presence of low copy segmental duplications in the DNA represented on the array. This again results in cross hybridisation with other regions in the genome; however, these effects cannot be competed out with Cot 1. One way of resolving this problem is not to use DNA isolated from clones. A strategy using PCR products which eliminate segmentally duplicated and common repeat elements has been implemented which avoided problems caused by cross hybridisation to secondary areas of the genome (Buckley, Mantripragada et al. 2002).

Genomic arrays have been used for many applications, such as the detection of copy number changes in cancers (Albertson, Ylstra et al. 2000; Albertson 2003), congenital

microdeletion studies (Buckley, Mantripragada et al. 2002), cytogenetic chromosome rearrangement (e.g. at chromosomal breakpoints) (Fiegler, Gribble et al. 2003) and epigenetic studies (van Steensel and Henikoff 2003).

1.6.2: Using Genomic Arrays to assess Replication Timing.

The ability of microarrays to detect copy number changes has also been used as a novel method to assess replication timing. The principle is to assess changes in the amount of DNA present at a particular locus during S phase. As the change in DNA copy number is very small (a maximum two fold difference) it is essential that the technique is very precise.

The first organism to have its genome analysed in this way was the yeast *Saccharomyces cerevisiae* (Raghuraman, Winzeler et al. 2001). Figure 1.20 from their publication shows how this was achieved.

Newly replicated DNA was labelled with carbon and nitrogen isotopes; early replicating DNA was labelled with the light C^{12} and N^{14} isotopes (HL) whilst late replicating DNA was labelled with the heavy C^{13} and N^{15} isotopes (HH). After synchronisation of cells with α factor, samples were taken at specific time intervals and a restriction digest was performed. The DNA was fractionated by caesium chloride density centrifugation which separates DNA according to the density of the labels. The two DNA populations were then separately labelled with biotin and individually applied to a high density array containing probes covering the entire *Saccharomyces cerevisiae* genome. At each locus the hybridisation ratios of the separate experiments were compared and plotted against position on each chromosome.

The high density of the *Saccharomyces cerevisiae* array allowed replication origins to be mapped. These appeared as peaks on the replication timing profile. Slight differences in replication timing due to the progression of the replication fork were detected and the replication origins mapped. Fork migration rates were also calculated by determining the slope of the profile around the ori (Figure 1.20c). While the link between replication and transcription in *Saccharomyces cerevisiae* was investigated,

in general no correlation was found. This was puzzling as it had previously been clearly observed in higher eukaryotes. The only exception was a family of histone genes which replicated 10 minutes earlier than the genome average of 31mins after release into S phase.

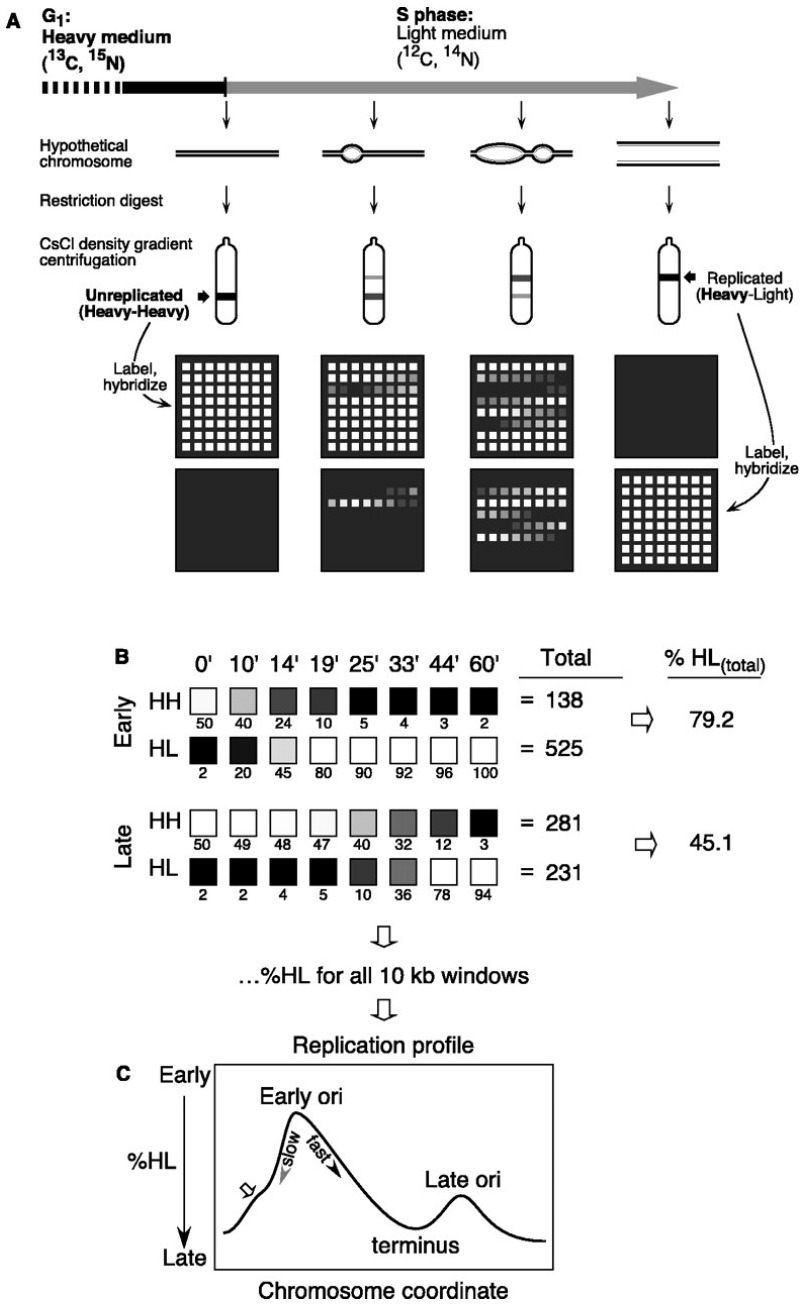


Figure 1.20: From (Raghuraman, Winzeler et al. 2001) illustrating how replication timing was assessed in *Saccharomyces cerevisiae.*

This study paved the way for the investigation of the replication timing of other organisms using microarrays. The measurement of replication timing on an array containing cDNA sequences from *Drosophila melanogaster* (Schubeler, Scalzo et al. 2002) allowed the correlation between replication timing and transcription to be assessed in a higher eukaryote. The method used was slightly different; BrdU was incorporated into the DNA of an unsynchronised *Drosophila* cell line. Flow sorting was then used to separate nuclei in early and late S phase using propidium iodide staining. The newly replicated DNA from each phase was isolated by BrdU immunoprecipitation. The two samples were differentially labelled using Cy3 and Cy5, before being co-hybridised to the same array (Figure 1.21).

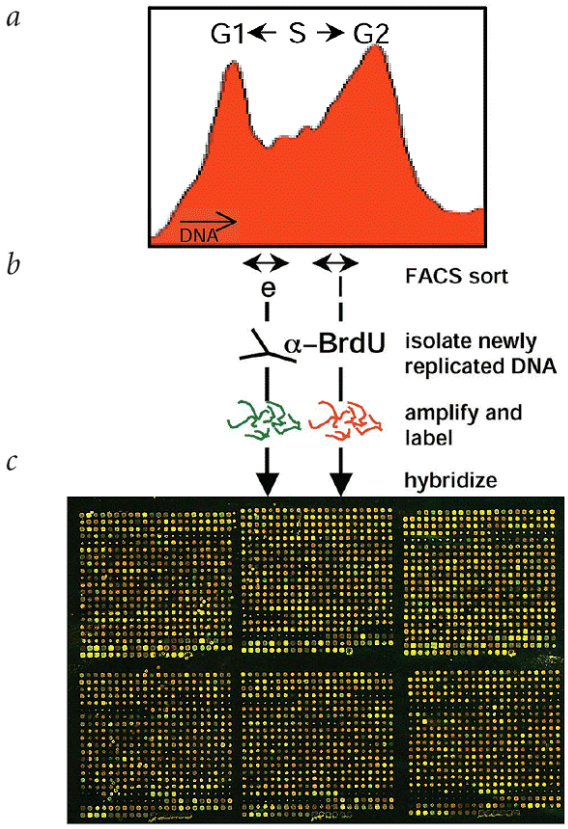


Figure 1.21: Reproduced from Schubeler et al 2002, illustrating how the replication timing of *Drosophila melanogaster* was assessed using microarray technology.

The array used contained 5,221 cDNAs giving an average sampling resolution of 20.5Kb. It also included probes containing retrotransposable elements, which map to blocks of repetitive DNA. The $\log_2$ ratio of late: early DNA was plotted against chromosomal position to produce a replication timing profile for each chromosome,

as shown in Figure 1.22. However, a higher resolution array would need to be used if origins were to be mapped, as in the yeast.

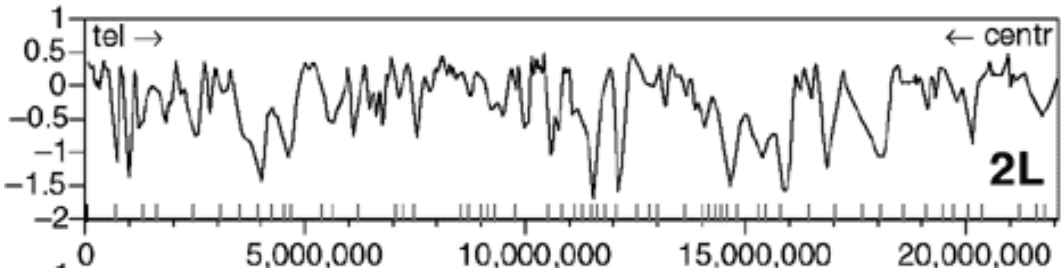


Figure 1.22: Replication profile of *Drosophila* chromosome arm 2L. The $\log_2$ transformation of the data meant that early replicating loci have more positive values, while late replicating regions have more negative values (Schubeler, Scalzo et al. 2002).

The cDNA array described was also used to measure transcriptional activity of the same *Drosophila* cell line. The use of microarray data ensured that there were enough data points to statistically correlate replication and transcription (Gilbert 2002). This study showed a correlation between the replication timing and the probability of gene expression. Early replication coincided with a higher likelihood of the gene being expressed and this correlation was highly significant.

The microarray probes used were derived from cDNA libraries and expressed sequence tags (EST's) which represented less than half the predicted number of *Drosophila* genes. This ensures that the analyses were conducted on coding DNA (McCune and Donaldson 2003). To determine the replication timing of non-coding regulatory regions and to understand how replication timing of non-coding DNA affects other characteristics of the genome such as transcriptional activity and the epigenetic code, any array used must contain representative genomic sequence. The arrays described in this thesis use cloned genomic DNA allowing correlations between replication timing and other features of the genome to be calculated for coding and non-coding DNA. In addition, large scale analysis of replication timing is carried out on a genome wide basis.

1.7 Aims of this Thesis

The main goal of this project was to use genomic microarrays to assess replication timing at a genome-wide level. The resolution of the replication timing map was then refined by the production of arrays from tile path clones.
A chromosome 22q genomic microarray was assembled. This was used to produce a replication timing map for chromosome 22q and to assess other features of the chromosome such as histone acetylation and copy number changes.

The project can be summarised with these aims.

1: A pilot study on a 4.5Mb region of chromosome 22. This involved the production of an array from clone DNA to cover the region chosen, the assessment of replication timing by co-hybridisation of differentially labelled S and G1 phase DNA to the array and the correlation of replication timing with GC content and gene density. (Chapter 3)

2: The production and verification of an array covering the whole of 22q using tilepath clones (average resolution 78Kb); the production of an array covering 4.5Mb of chromosome 22 with 500bp PCR products (average resolution 10Kb); and an array covering 200Kb of chromosome 22 with overlapping 500bp PCR products. (Chapter 4)

3: The production of a replication timing map of the whole genome in a lymphoblastoid cell line. This was performed at a 1Mb resolution and at a tile path resolution for chromosomes where tile path arrays are available. This data was then used for the assessment of the correlation between replication timing and sequence features of the genome; specifically GC content, gene density and density of common repeat elements. (Chapter 5)

4: The assay of the transcriptional activity in the lymphoblastoid cell line and the correlation of transcription with replication timing. The 22q tile path array was used to assess histone acetylation by chromatin immunoprecipitation and application to the

array. The acetylation status of chromosome 22 was correlated with replication timing (Chapter 6).

5: The assessment of microdeletions on chromosome 22 using the 22q tile path array (Chapter 7).

2: Materials & Methods

A new method to assess replication timing was developed by co-hybridisation of S and G1 phase DNA onto genomic clone microarrays. Initially an array covering 4.5 Mb of chromosome 22 was made to test this new method. The test array was constructed using tile path clones between CTA-415G2 and CTA-390B3 (approximately 17.5 – 21 Mb along the q arm of Chromosome 22). Subsequently a full tiling path was constructed using the methods optimised in the small tile path array. Replication Timing was assessed using the arrays made from chromosome 22 clones, an array sampling the whole genome a 1Mb intervals and a high density array sampling a region of Chromosome 22 (at a transition between a G light and G dark boundary) using high density PCR products.

Recipes for common buffers and reagents can be found in Appendix 1.

The strategy used for construction of a tiling path array is shown in Figure 2.1.

2.1 Construction of the 22 tile path array

2.1.1: Clone Selection and verification.

2.1.1.1: Clone selection

Clones were picked from the published 22q sequencing clones (Dunham et al 1999). In total 526 Chromosome 22 clones were picked (from the libraries detailed in Appendix 9b). Control clones were also picked from chromosome X (33 BACS and 62 PACS) to allow verification of copy number changes in male:female hybridisations; six *Drosophila* DNA clones acted as non-specific hybridisation controls.

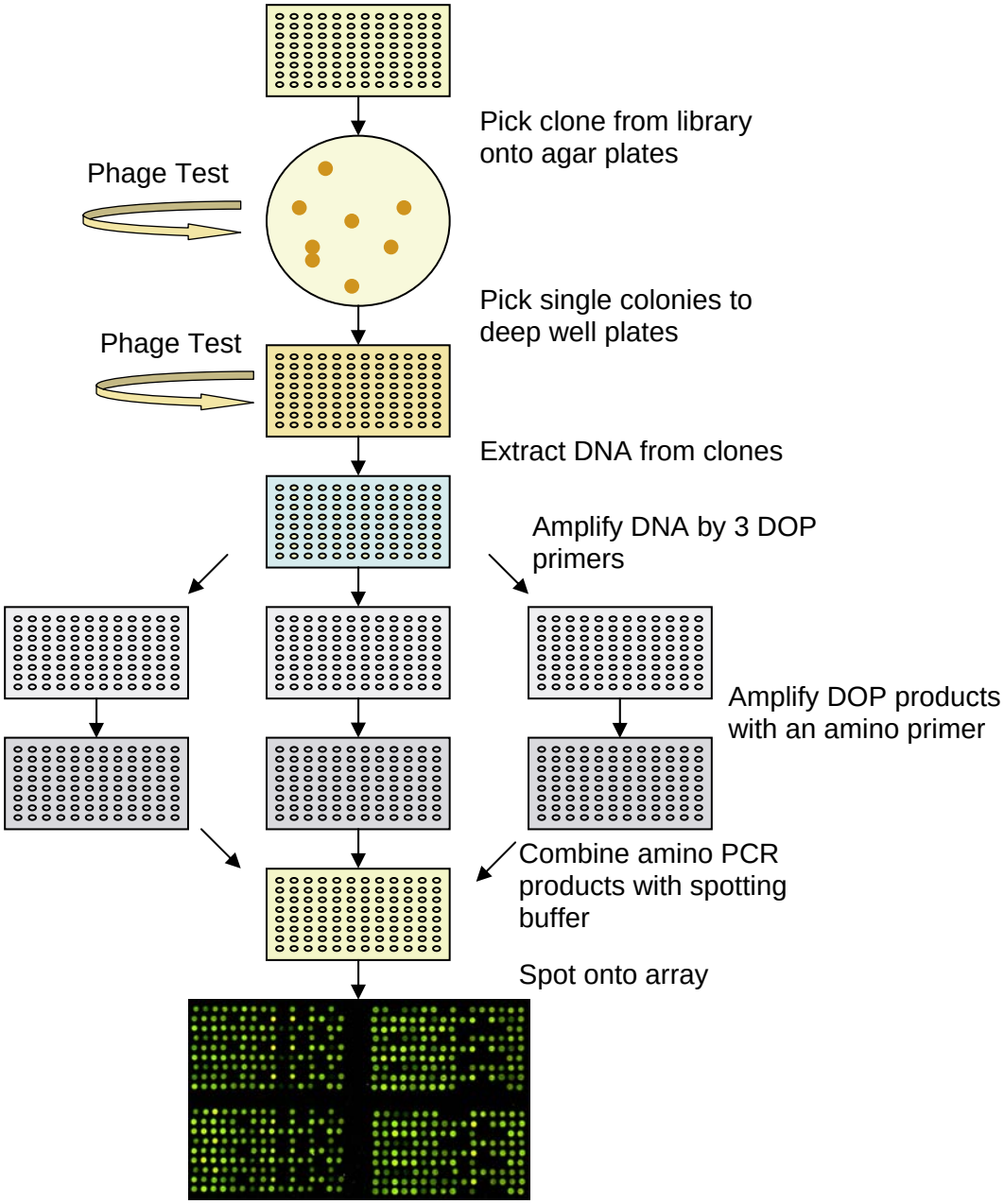


Figure 2.1: Flow diagram illustrating the construction of the tile path array

2.1.1.2: Preparation of glycerol clone stocks.

Clones were picked from libraries held at the Sanger Institute, or obtained as stabs from the University of Oklahoma and Research Genetics (Invitrogen). Clones were streaked onto LB Agar poured into Sterilin 10cm plates containing the appropriate antibiotic. Cosmids and PACs were streaked onto agar with a final kanamycin concentration of 30μg/ml. Fosmids and RP-11 BACs were picked onto agar with a final chloramphenicol concentration of 25μg/ml. CTA BACs (Appendix 9b) were

picked onto agar with a final chloramphenicol concentration of 12.5μg/ml. Clones were simultaneously streaked onto phage assay plates to test for bacteriophage infection. To prepare phage assay plates, LB agar containing 0.8% agarose was seeded with 2.25% DH10B phage-susceptible *E coli* grown in LB broth. This was poured onto a thin layer of LB agar, left to set and used on the same day. The clone plates and the phage assay plates were incubated overnight at 37^{o}C.

Phage assay plates were examined for DH10B growth, a clear lytic plaque indicating contamination of the corresponding clone with phage. If phage was found the whole agar plate was discarded and the remaining clones on the plate were re-streaked from the libraries.

Single colonies of clones that had passed phage testing were picked from the agar plates into deep well 96 well boxes (Costar) containing 1.5ml of LB broth containing 7.5% glycerol and the appropriate antibiotic for the clone picked (see above). Gas permeable plate sealers (Advanced Biotechnologies Ltd) were placed over the box. Clones were cultured at 37^{o}C for 16 hours, whilst shaking at 300rpm.

The deep well cultures were then tested for bacteriophage infection. A 96 pin hedgehog was dipped into a deep well plate before being stamped onto a phage assay plate. The hedgehog was rinsed in water, flamed with ethanol and left to cool in between each use. If phage was detected the whole 96 well plate was discarded and the remaining clones were repicked from the agar streaks. 200μl of culture from phage negative plates were alliquoted into 96 well flat bottom microtitre plates (Costar). The microtitre plates were frozen on dry ice and stored a -80^{o}C.

2.1.1.3: Clone Verification.

Initial clone verification was by HindIII digest fingerprinting. Bands were compared with a virtual digest performed on the published accession sequence for that clone. When the virtual fingerprint did not produce sufficient bands for reliable verification PCR was performed using primers designed to sequence tagged sites within the clone.

2.1.1.3.1: DNA preparation of Bacterial clones – Micro Prep.

Colonies were stamped from a glycerol stock (produced as described in 2.1.1.2) into a 96 deep well box (Costar) containing TY media and the appropriate antibiotic (as above) using a 96 pin hedgehog. The clones were cultured at 37^{o}C for 16 hours, shaking at 300 rpm.

250µl of the resulting cultures was alliquoted into a 96 well round-bottom microtitre plate (Corning). Plates were spun at 938g for 4 minutes. 25µl of GTE solution (50mM glucose, 10mM EDTA, 5mM Tris pH8) was added to resuspend each pellet in the individual wells. 25µl of 2mM NaOH/1% SDS was added to each well and the plate incubated for five minutes at room temperature. 25µl of 3M KOAc was added and incubated for five minutes at room temperature. The contents of each well were transferred into a 0.2µm costar 96 well filter-bottom plate. This was placed on top of a 96 well round bottom plate, each well containing 100ml of isopropanol and spun at 938g for 2 minutes.

The filter-bottom plate was discarded and the isopropanol plate was left at room temperature for 30 minutes. It was then spun at 1536g for 20 minutes, the supernatant was discarded and the plate inverted on tissue to dry the pellet. 100µl of 70% ethanol was added to each well and the plate spun at 1536g for ten minutes. The supernatant was discarded and the ethanol wash repeated. Again the supernatant was discarded and the pellet dried until transparent. The pellet was then resuspended in 5µl of T0.1E containing 50ng RNAse.

2.1.1.3.2: HindIII digest

A HindIII digestion mix was prepared and 4 μl was added to each well containing DNA. The plate was pulse centrifuged to 100g and the plate incubated at 37^{o}C for 2 hours. The plate was again pulse centrifuged to 100g and the digest reaction was terminated by the addition of 2μl of orangeG solution to each well. The plate was pulse centrifuged to 100g and 1μl of the digestion mix was run on a 1% agarose gel. The gel was loaded with a Promega marker lane every 5 wells. The gel was then run at 90 volts for 15 hours at 4^{o}C.

The gel was stained with Vista green for 45 minutes at room temperature, whilst shaking and the gel was washed in distilled water and scanned on a Typhoon 8600 scanner (Molecular Dynamics).

2.1.1.3.3: Analysis of Fingerprint gels

Gels were quantified using the image analysis program Image 3.10b. Lanes were first defined using the marker lanes which allows for gels that have not run straight. Individual bands were then analysed manually to define the presence and absence of every band (The intensities and definition of the band was studied in conjunction with the neighbouring bands to differentiate bands from background.) The marker lanes were co-aligned to produce an output image with the size of the sample represented by the position of the band.

The band positions obtained from the fingerprint gels were compared to virtual digests carried out on the sequenced portions of each clones. Although the HindIII digest was carried out on the whole clone the virtual digest was only carried out on the sequenced portion of the clone. It was therefore expected that the HindIII digest would contain more bands than the virtual digest. Each clone was manually examined against the virtual digest and called as pass or fail, by comparing the number and position of the bands in the actual and virtual profiles.

2.1.1.3.4: Verification of clones by detection of a Sequence Tagged Site by the Polymerase Chain Reaction (STS PCR)

Clones that required verification by STS PCR testing were streaked on LB Agar supplemented with the required antibiotic. Agar plates were incubated overnight at 37^{o}C. A single colony was taken from each plate and transferred to 100µl of T0.1E. 5µl of this was then taken to use as a template in the STS PCR.

The STS PCR was carried out in a volume of 15µl. To each well the following reagents were added. 1.5µl of 10x PCR Buffer (67mM $MgCl_2$, 670mM Tris-Cl, 167mM $(NH_4)_2SO_4$), 1.5µl of dNTP's (final conc. 5mM each), 0.5µl bovine serum albumin (5mg/ml – Sigma), 0.2µl β-mercaptoethanol (0.72M), 0.12µl Taq Polymerase (5units/ml), 0.75µl primer pair mix (Forward and Reverse primers both present at 100ng/ml), 5.425µl of T0.1E/cresol red/sucrose solution (T0.1E, sucrose (28% w/v), cresol red (0.008% w/v)). 5µl of the colony dissolved in T0.1E was added to each well as a template. Genomic DNA (5ng) was used as a positive control and water was used as a negative control.

Thermal cycling was performed using the following conditions. A single denatration step at 94^{o}C for five minutes was performed. 34 cycles were completed using 94^{o}C for 30 seconds, 58^{o}C for 30 seconds and 72^{o}C for 30 seconds. This was followed by a final elongation at 72^{o}C for five minutes.

After the PCR, 7.5µl of the reaction mixture was run on a 2.5% agarose gel made with 1xTBE and stained with ethidium bromide. When the gel was studied on a UV transiluminator the presence of a band at the anticipated size confirmed the success of the PCR and confirmed the correct clone was present in the template.

Of the 526 chromosome 22 clones tested, 470 passed verification and were included on the array. 467 of these were obtained directly from golden path sequencing clones, while the remaining 3 were identified by BAC end sequence and positioned on the chromosome 22 sequence.

2.1.2: Construction of the array from the Chromosome 22 clone set

DNA was prepared as described in 2.1.1.3.1. The DNA was dissolved in 195µl of T0.1E to give a final DNA concentration of 1ng/µl. This was amplified by three different degenerate oligonucleotide (DOP) primers. The fact the DOP primers anneal frequently to the clone DNA sequence ensures the whole clone is represented. It also provides a target for a second 'amino' primer on the 5' end of the product, which enables further amplification of the DNA and attachment of a covalently bound amino group to the PCR product. This in turn allowed covalent attachment of the DNA to the microarray slides.

2.1.2.1: DOP PCR amplification of clone DNA.

DNA was amplified using three different Degenrate Oligonucleotide PCR (DOP) primers. The primers were synthesised by Oswell DNA service and their sequence is as follows,

DOP 1: CCGACTCGAGNNNNNNCTAGAA
DOP 2: CCGACTCGAGNNNNNNTAGGAG
DOP 3: CCGACTCGAGNNNNNNTTCTAG

DOP-PCR was performed in 1x TAPS 2 buffer, 0.25% W1, 0.25mM dATP, 0.25mM dCTP, 0.25mM dTTP, 0.25mM dGTP, 2.5U AmpliTaq polymerase (Perkin-Elmer) 5µl of DNA prepared in protocol 2.1.1.3.1. was used as a template in a final reaction volume of 50µl.

The reactions were carried out on PTC-225 Tetrad thermocyclers (MJ Research). The DNA was denatured at 94^{o}C for three minutes followed by 10 cycles of 94^{o}C for 90 seconds, 30^{o}C for 150 seconds, ramping at 0.1^{o}C per second to 72^{o}C, then 72^{o}C for 180 seconds. This was followed by a further 30 cycles of PCR, 94^{o}C for 60 seconds, 62^{o}C for 90 seconds and 72^{o}C for 120 seconds. A final extension step at 72^{o}C for 480 seconds finished the reaction. Three different PCR's utilising the 3 different DOP primers (DOP 1, 2 and 3) were performed for each separate clone DNA sample.

5ml of each product was run on a 2.5% agarose gel made with 1xTBE containing ethidium bromide, to confirm the success of the PCR reaction.

2.1.2.2: Secondary DNA amplification using an amino-linked primer.

A secondary PCR is performed using the DOP PCR products as a template. This reaction utilised a 5' amine-modified primer. The primer was designed so that the 10 bases at the 3' end matched the 10 bases at the 5' end of the DOP primers. The primer sequence was GGAAACAGCCCGACTCGAG. The reaction was carried out in 1x Amino-linking buffer, 0.25mM dATP, 0.25mM dCTP, 0.25mM dTTP, 0.25mM dGTP, amino-linking primer (10ng/µl) and 1% AmpiTaq polymerase (Perkin-Elmer). 2µl of DOP amplified DNA prepared as in 2.1.2.1 as a template. The reaction was carried out in a volume of 60µl.

The PCR's were carried out on PTC-225 Tetrad thermocyclers (MJ Research). After initial denaturation at 95^{o}C for 600 seconds, 35 cycles were performed as follows, 95^{o}C for 60 seconds, 60^{o}C for 90 seconds and 72^{o}C for 420 seconds. This was followed by a final extension at 72^{o}C for 600 seconds.

5µl of each product was run on a 2.5% agarose gel as described previously to confirm the success of the PCR reaction.

2.1.2.3: Combining of amino products for spotting onto the array

For each clone DNA there were three different amino-linked DOP products. These were combined before spotting onto the array. 40µl of amino-linked DOP1, DOP2 and DOP3 product from the each clone DNA was alliquoted into the microtitre plate containing 39µl of 4x Spotting buffer (1M Sodium phosphate pH8.5, 0.001% sarkosyl). A 96 well 0.2µm costar filter plate was secured over a fresh 96 well microtitre plate. The contents of the first plate are transferred to the filter plate and it is centrifuged at 600g for five minutes. 15µl of the combined filtrate was transferred to 384 well plates in preparation for arraying.

2.1.3: Spotting of the Array

DNA from each clone was spotted in triplicate onto amine-binding glass microarray slides (3D link activated slides – Motorola) using a MicroGrid II array robot (Biorobotics). Slides were incubated in a humidity chamber with NaCl saturated water for 24-72 hours. The slides were then incubated in a 1% ammonium hydroxide solution for five minutes, followed by a five minute incubation in a 0.1% SDS solution. The DNA bound to the slides was denatured by incubation in distilled water at 95^{o}C for two minutes, before being plunged into ice cold water. Slides were dried by centrifugation at 150g for five minutes.

2.2: Construction of a High Resolution Arrays from PCR products.

2.2.1 Primer design

A high resolution array was constructed sampling at a resolution of 10Kb for a 4.5Mb region of 22q using approximately 500bp PCR products. The region chosen for analysis was between accession numbers AC005003 and AL079295 spanning bases 15398721 – 29982021 along the q arm of chromosome 22. In addition, overlapping 500bp products were designed to cover the central 200Kb of this region, allowing sampling at a very high resolution. This central 200Kb was positioned between 16495000-16695000 bp along the q arm of chromosome 22 and falls within the accession numbers Al021937-Z82246.

Primers were designed using the 'primer 3' program. The sequence chosen was repeat-masked and the primers blasted against the rest of the genome to ensure their target sequence was unique. The primer position was also weighted to be as close to the centre of each 10Kb region as possible. This allowed even sampling resolution. An amino linked tag sequence (5'-TGACCATG-3') was attached to the 5' end of the sense strand primer, to allow covalent binding to the slide. All primer sequences are in Appendix 2.

2.2.2. PCR amplification of 500bp products.

PCR reactions were performed using clone DNA as a template. The clone containing each primer was identified, picked and grown in 2xTY media containing the appropriate antibiotic as described in 2.1.1.2. The culture was then diluted 1:10 with sterile water for use as template. The PCR was performed in a 50µl reaction containing 12.5µl of the template in 1x amino-linking buffer, 0.05mM dATP, 0.05mM dCTP, 0.05mM dTTP, 0.05mM dGTP, 0.03125U/µl Taq polymerase (Perkin Elmer-Cetus), 5ng/µl sense-strand primer and 5ng/µl antisense-strand primer.

The reactions were carried out on PTC-225 Tetrad thermocyclers (MJ Research). The DNA was denatured at 94^{o}C for five minutes followed by a 30 cycles of 94^{o}C for 60 seconds, 62^{o}C for 90 seconds and 72^{o}C for 90 seconds. A final extension step at 72^{o}C for five minutes was performed. 5µl of the product was run on a 2.5% agarose gel. A successful PCR was denoted by a strong single band at approx 500bp. Reactions that were not successful at this stage were repeated using 10ng/µl genomic DNA as a template.

2.2.3 Preparation of products for spotting onto the array.

40µl of PCR product was aliquoted into the microtitre plate containing 13µl of 4x Spotting buffer (1M Sodium phosphate pH8.5, 0.001% sarkosyl). A 96 well 0.2µm costar filter plate was secured over a fresh 96 well microtitre plate. The contents of the first plate are transferred to the filter plate and were centrifuged at 600g for five minutes. 15µl of the combined filtrate was transferred to 384 well plates in preparation for arraying.

2.2.4 Spotting of arrays.

Arrays were spotted as described in section 2.1.3.

2.3: Acquisition of DNA for application to the array.

Several sources of DNA were used for array analysis. For array verification and male versus female studies total DNA was extracted from lymphoblastoid cell lines. For replication timing studies cells from cell lines were first sorted into G1 phase and S phase of the cell cycle before the DNA was extracted (2.3.2.3). Patient DNA was obtained from collaborators and control DNA used in these studies was a pool of 20 different male, or 20 different female DNAs obtained from ECACC (European Collection of Cell Cultures, UK). For Chromosome 22 add-in verification experiments individual chromosomes were flow sorted and the DNA extracted.

2.3.1: Extracting DNA from lymphoblastoid cell lines.

Five different lymphoblastoid cell lines (Table 2.1) were cultured and the DNA was extracted from logarithmically growing cell lines.

Table 2.1: Cell lines cultured for DNA extraction (names in brackets denote internal names)

Cell line name	**ECACC Number**	**Sex**
C0202-JAT (HRC 575)	94060845	Male
C0009-SAH (HRC 160)	93010702	Female
C0154-RA (HRC 193)	93012805	Male
C0020-RW (HRC 146)	92030511	Female
C0008-JH (HRC 159)	93010701	Female

2.3.1.1: Cell Culture of lymphoblastoid cell lines.

The Lymphoblastoid cell lines were cultured in RPMI 1640 media (Sigma) supplemented with 16% Foetal Bovine Serum (Gibco-BRL, Life Sciences), 2mM L-glutamine, 100units/ml penicillin and 10mg/ml streptomycin (all Sigma). They were incubated at 37^{o}C and split 1:2 every 3-4 days.

2.3.1.2: Extraction of DNA from lymphoblastoid cell line

From each culture, 5x10^{6} log phase cells were harvested and centrifuged at 300g for ten minutes. The cells were washed and resuspended in 0.5ml of Phosphate Buffered Saline (PBS). Genomic DNA was then extracted using a blood and cell culture mini kit (Quiagen) following the manufacture's instructions.

Briefly, cells were lysed by the addition of 0.5ml of lysis buffer (1.28M sucrose, 40mM Tris-Cl pH7.5, 20mM $MgCl_2$, 4% Triton X-100) and 1.5ml of ice-cold distilled water and incubation on ice for ten minutes. Lysed cells were centrifuged at 1300g for 15 minutes, the supernatant discarded and the pellet resuspended by vortexing in 0.25ml of lysis buffer and 0.75ml of ice-cold distilled water. The mix was centrifuged again for 15 minutes at 1300g and the supernatant discarded.

The nuclei were resuspended in 1ml of General lysis buffer (800mM guanidine HCl, 30mM Tris-Cl pH8.0, 30mM EDTA pH8.0, 5% Tween 20, 0.5% Triton X-100) by vortexing for 10-30 seconds. 25µl of Proteinase K (20mg/ml) stock solution was added and the nuclei incubated at 50^{o}C for 60 minutes.

Quiagen genomic tips (20/G) were equilibrated with 1ml of equilibration buffer (750mM NaCl, 50mM MOPS (3-(N-morpholino)propanesuphonic acid) pH7.0, 15% isopropanol, 0.15% Triton X-100). The buffer was allowed to flow through the tip by gravity. The prepared sample was vortexed for 10 seconds, the genomic tip placed over a 15ml falcon tube and the sample was applied to the resin in the genomic tip. The genomic tip was washed three times with 1ml of wash buffer (1.0M NaCl, 50mM MOPS pH7.0, 15% isopropanol), all solutions were allowed to move through the tip at 1g. The genomic tip was placed over a clean 15ml falcon tube and the DNA eluted

by the addition of 2x1ml of elution buffer (1.25M NaCl, 50mM Tris Cl pH8.5, 15% isopropanol). The eluted DNA was precipitated by the addition of 1.4ml of isopropanol and the falcon tube was inverted 10 times. The tube was then centrifuged at 5000g for 30 minutes and the supernatant removed. The DNA was resuspended in 100ul of T0.1E buffer.

The DNA concentration was measured using a TD-360 flurometer (Turner designs). The DNA sample was diluted 1:100 in Flurometer buffer (10mM Tris, 1mM EDTA, 0.2mM NaCl, Hoechst 33258 1μg/ml) and the DNA concentration measured against a 500ng/μl standard.

Table 2.2: Concentration of DNA extracted from cell lines

Cell line name	**DNA Conc. (ng/μl)**	**Amount in labelling reaction (μl)**
C0202-JAT (HRC 575)	195.5	2.30
C0009-SAH (HRC 160)	210.5	2.14
C0154-RA (HRC 193)	215.5	2.09
C0020-RW (HRC 146)	225.0	2.00
C0008-JH (HRC 159)	257.0	1.75

The DNA was then used in a Random prime labelling reaction (2.4.1)

2.3.2: Extracting DNA from sorted S phase and G phase nuclei.

Two different lymphoblastoid cell lines and one lymphoblastoid cell line with a Chromosome 17: Chromosome 22 translocation were cultured and the DNA extracted from sorted S and G phase nuclei.

Table 2.3: Cell lines cultured for S and G1 flow sorting (names in brackets denote internal names)

Cell line name	**ECACC Number**	**Cell type**
C0202-JAT (HRC 575)	94060845	Male Lymphoblastoid
C0009-SAH (HRC 160)	93010702	Female Lymphoblastoid
1274	N/A	t(17:22) Lymphoblastoid

2.3.2.1: Time course experiment to determine when the maximum number of cells were in S phase.

Prior to sorting, two time-course experiments were performed to assess the best time interval between splitting and harvesting lymphoblastoid and fibroblastoid cell lines to obtain the maximum number of cells in S phase when harvesting.

A 1ml sample was taken from the lymphoblastoid cell culture every two hours (during working hours) for 76 hours after subculture 1:2. Each sample was harvested as described in 2.3.2.3

2.3.2.2: Harvest of Cell for sorting into S and G1 phase.

Lymphoblastoid cell lines (HRC 575, HRC 160, & 1274) were sub-cultured 1:2 into 4 x 75cm^3 flasks and incubated for 26 hours and harvested by centrifugation at 300g for five minutes. HRC 575 Cells were washed in 5ml of PBS and spun at 300g for five minutes and the pellet resuspended in 75mM KCl and incubated room temperature for fifteen minutes. The cells were resuspended in PAB at a concentration of $7x10^6$/ml before sorting. Cells were stained with Hoechst 33258 at a concentration of 2μg/ml before incubation at room temperature for five minutes.

The other lymphoblastoid cell lines (HRC 160 & 1274) were washed in 5ml of PBS and centrifuged at 300g for five minutes. The cell pellet was resuspended in 0.5ml of PBS and 4.5ml of 70% ethanol. Samples were stored at 4^oC. Just before sorting the cells were centrifuged at 300g for 10 minutes and resuspended at a concentration of $3x10^6$ per ml in Trisodium Citrate buffer (1% Trisodium Citrate, 1% Triton X-100, 0.5mM Tris, 3.75μl/ml spermine). After incubation at 4^oC for ten minutes the cells were stained with Hoechst 33258 at a concentration of 2μg/ml before incubation at room temperature for five minutes.

2.3.2.3: Sorting of the nuclei.

G1 and S phase cells were sorted by a Coulter-Elite flow cytometer (HRC 575 & HRC 160) or a Mo-Flo flow cytometer (1274 and HRC 160 fractions)

Nuclei were sorted using gates in the positions indicated in Table 2.4 and Figure 2.2.

Table 2.4: Table to indicate where gates are positioned on cell profiles (Figure 2.2).

Cell line Sorted	**Profile Showing Gates**
HRC 575 (G1 and S phase)	A
HRC 575 (G1, 4 fractions of S phase & G2/M)	B
HRC 160 (G1 and S phase)	C
HRC 160 (G1, 4 fractions of S phase & G2/M)	D
1274 (t17:22) (G1 and S phase)	E

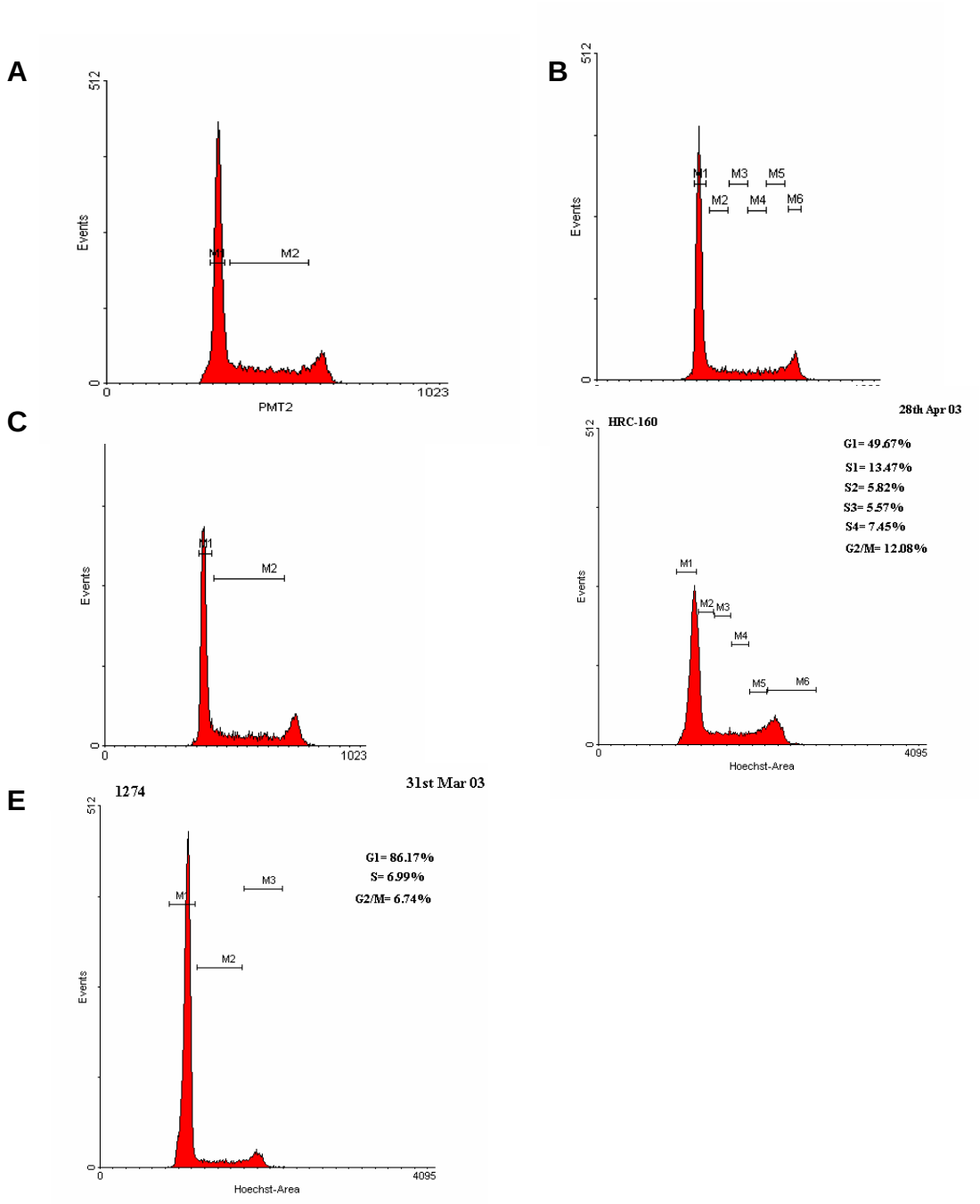


Figure 2.2: Flow sorter profiles and gate positions of sorted cell lines.

Cells were sorted into Sheath Buffer by their Hoechst intensity. To every 10^5 sorted nuclei, 25mM EDTA/1% sodium lauroyl sarcosine solution and 20µg/ml proteinase K (Gibco-BRL, Life Sciences) was added and the sample incubated overnight at 42^{o}C. The proteinase K was inactivated by the addition of 0.04mg/ml of Phenylmethylsulfonyl fluoride (Sigma) and incubation at room temperature for 40 minutes. The DNA was precipitated by the addition of 20µl 5M NaCl, 1µl of non-

fluorescent pellet paint (Novagen) and 1ml of absolute ethanol and incubated overnight at –20^{o}C. The precipitate was spun at 7700g for 15 minutes to pellet the DNA. The pellet was washed with 500µl of 70% ethanol then centrifuged at 7700g for 7 minutes. The pellet was resuspended in TE pH 8.0 and the DNA concentration was measured on the TD-360 flurometer as described previously and 1µl of DNA was run on a 1% agarose/TBE gel stained with ethidium bromide (Figure 2.3).

The DNA prepared above was used as the input for the random prime labelling reaction

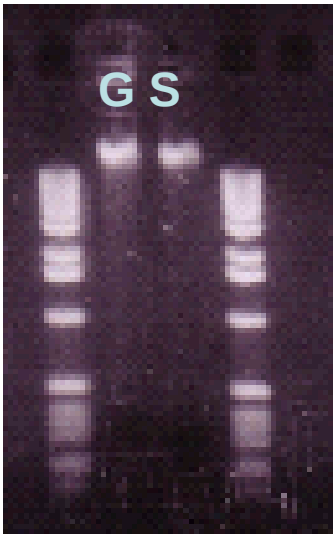


Figure 2.3: Purified S and G1 phase DNA run between two 1Kb markers.

2.3.3: Extraction of DNA from sorted Chromosomes

DNA was extracted from sorted chromosome 22s in the same way as DNA was extracted from nuclei. This is described in section 2.3.2

2.3.4: Male and Female control Pools.

Pools of 20 normal male DNA and 20 normal Female DNA were used as controls in the microdeletion studies. DNA was obtained from Human Random Collection at ECACC. The sex of each DNA was verified by PCR using chromosome Y specific primers. Each control DNA was made by pooling 20µg of DNA from each of 20

individuals and diluting the mixed DNAs to a concentration of 200ng/μl in distilled water.

2.3.5: Obtaining DNA for Microdeletion studies.

DNA from five patients with DiGeorge syndrome was obtained from collaborators at the Department of Clinical Genetics at Addenbrookes Hospital, Cambridge.

DNA from six patients with a DiGeorge-like phenotype, but that did not show a DiGeorge deletion on Chromosome 22 using standard fluorescence *in situ* hybridisation (FISH) probes was obtained by collaboration with the Molecular Medicine Unit, ICH.

2.4: Labelling of DNA and application to the array

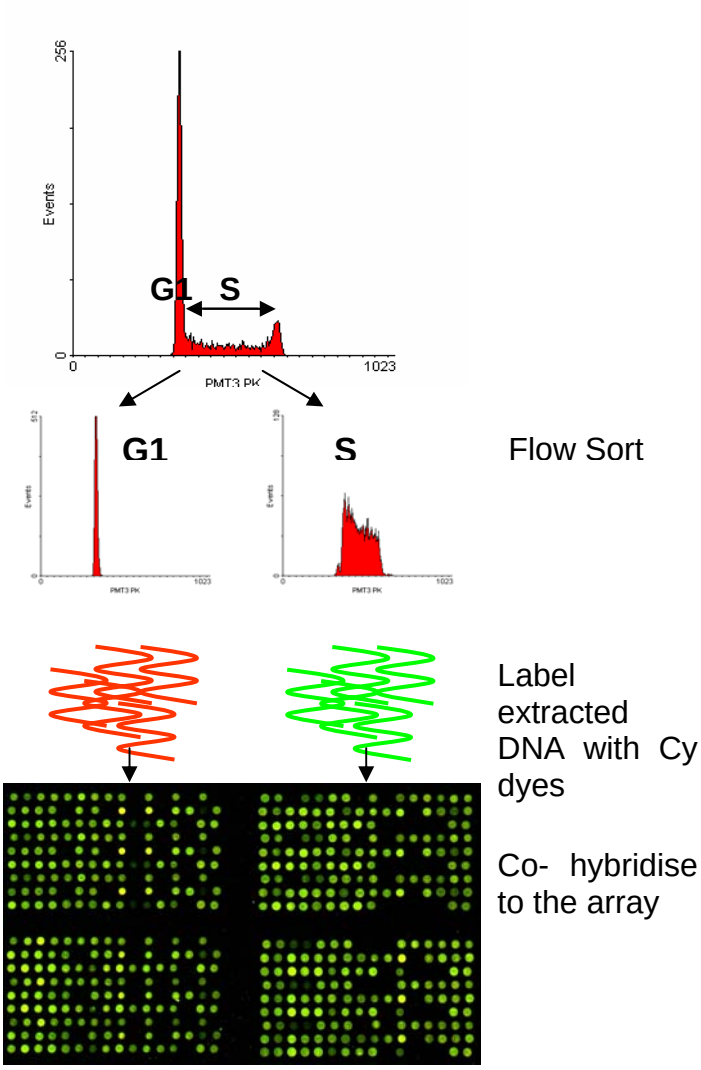


Figure 2.4: Flow diagram illustrating how DNA was applied to the constructed array.

The stages involved in the flow sorting, DNA labelling and hybridisation to the array is shown in Figure 2.4.

2.4.1: Labelling of DNA

450ng of each DNA sample was labelled using Cy-dye modified dCTP (for arrays less than 6cm^2). The DNA was added to water to give a final volume of 66µl. 60µl of random priming solution (Invitrogen – bioprime labelling kit) was added and the DNA denatured at 100^{o}C for 10 minutes. The DNA and primer mix was plunged into ice and 15 µl of 10x dNTP mix (2mM dATP, 2mM dGTP, 2mM dTTP and 0.5mM

dCTP in TE buffer), 6µl of 1M Cy3-dCTP or Cy5-dCTP (NEN Life Sciences) and 3µl of Klenow fragment (Invitrogen – bioprime labelling kit). The final reaction volume of 150µl was incubated at 37^{o}C overnight in the dark to prevent bleaching of the Cy dye.

The reaction was quenched by the addition of 15µl stop buffer (Invitrogen – bioprime labelling kit) and unincorporated nucleotides were removed by passage of the reaction mix through microspin G50 columns (Pharmacia). Columns were prepared following the manufacturers instructions and 55µl of the reaction mix was then loaded to the sloped surface of each of 3 spin columns, the columns placed in an empty 1.5ml eppendorfs to collect the labelled DNA and centrifuged at 735g for two minutes. Identical samples were pooled and 5µl was run on a 2.5% agarose gel made with TBE buffer and ethidium bromide. A smear indicated a successful labelling reaction (Figure 2.5).

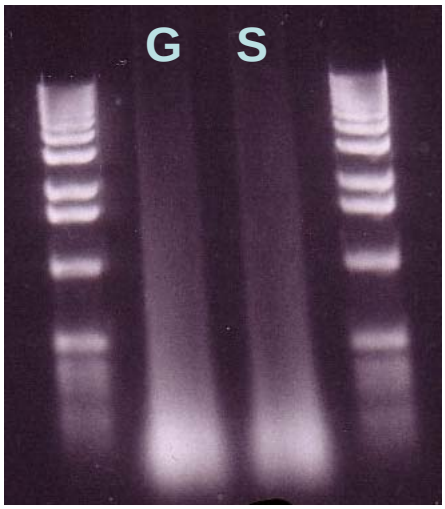


Figure 2.5: Labelled S and G1 phase DNA

2.4.2: Precipitation of Pre-hybridisaion and hybridisation DNA

For a 6cm^2 array, 180µl of Cy3 labelled DNA, 180µl of Cy5 labelled DNA, 135µl of human Cot1 DNA (Roche), 54µl of 3M sodium acetate pH 5.2 and 1000µl of 100% ethanol were added to a 1.5ml eppendorf tube for the hybridisation mix. To a separate 1.5ml eppendorf tube, a pre-hybridisation solution was prepared containing 80µl Herring sperm DNA (10mg/ml Sigma), 135µl human Cot1 DNA, 24µl of 3M sodium

acetate pH 5.2 and 900µl of 100% ethanol was added. The DNA was precipitated at –70^{o}C for 30 minutes or –20^{o}C overnight.

2.4.3: Application of the DNA to the array.

To the arrays produced as described in 2.1.3. a rubber cement wall was placed around the area of the array to create a well. When this first layer was dry a second layer was applied and left to dry.

The precipitated labelled and pre-hybridisation DNA mixes were centrifuged at 7700g for 15 minutes. The supernatant was tipped off and 500µl of 80% ethanol was added to wash the pellet. The samples were centrifuged at 7700g for 7 minutes and the supernatant was removed.

Hybridisation buffer was preheated to 70^{o}C. 160µl of the hybridisation buffer was added to the Herring sperm/Cot 1 DNA and the pellet resuspended (pre-hybridisation solution). The Cy labelled/Cot 1 DNA was resuspended in 60µl of hybridisation buffer and 6µl of yeast tRNA (100µg/µl, Invitrogen) added (hybridisation solution). The DNA/hybridisation buffer mixes were denatured at 70^{o}C for 10 minutes.

The pre-hybridisation solution was applied to the array in the centre of the rubber cement well. Care was taken to ensure that the pre-hybridisation solution covered the entire slide surface enclosed by rubber cement. The array was incubated with the pre-hybridisation solution in a square humidity chamber containing 3MM paper (Whatman) saturated with 2xSSC and 40% formamide and placed in an oven on a platform rocking at 5rpm at 37^{o}C for 60 minutes. The hybridisation solution was placed at 37^{o}C for 60 minutes in the dark.

After the pre-hybridisation incubation, as much pre-hybridisation solution as possible was removed from the slide. The hybridisation solution was then applied to the array and care was taken that the hybridisation solution was distributed over the area enclosed by the rubber cement. The slide was placed into a second small slide size hybridisation chamber containing 3mm paper (Whatmann) saturated with 2xSSC and

20% formamide. The chamber was sealed with parafilm and placed in an oven at 37^{o}C for 48 hours rocking at 5rpm.

2.4.4: Washing the array.

The slides were removed from the hybridisation chamber and the rubber cement wall was carefully removed using forceps. The slides were placed into a Hellendahl jar (Raymond Lamb) containing PBS/0.05% Tween 20 (BDH). The slides were then transferred into a trough containing PBS/0.05% Tween 20 and placed on a platform rocking at 70rpm for 10 minutes. The slides were transferred to a trough containing 50% formamide/2xSSC and placed in an oven at 42^{o}C whilst rocking for 30 minutes. The slides were transferred back into a trough containing fresh PBS/0.05% Tween 20 and washed for 10 minutes, rocking at 70rpm. The slides were dried by centrifugation at 150g for 5 minutes and stored in the dark before scanning.

2.5: Scanning and analysis of the array

2.5.1: Scanning of the slides

Slides were scanned on an Axon 4000B scanner (Axon instruments). The photon multiplier tube (PMT) levels used for detection were adjusted and tailored for each individual array. Arrays were analysed using Genepix 4.0. Software (Axon instruments). A scanned array image can be seen in Figure 2.6. Single image tiff files from Genepix 4.0 were imported into the 'Spot' analysis program (Jain, Tokuyasu et al. 2002) for analysis.

Figure 2.6: The 22q tile path array

2.5.2: Analysis of the slide

2.5.2.1: Description of the 1Mb and 22Tile path analysis programs

For analysis of the arrays, the Cy3 and Cy5 intensities were normalised. to adjust for any imbalance in the scanning of the flurochromes. For analysis of the replication

timing arrays the S phase ratios were scaled. The scaling factor was obtained from the S phase profile (Figure 2.2) and was determined by the fraction through S phase at which 50% of the DNA had replicated. The scaling factor was applied to the analysis of the co-hybridisation of S and G1 phase DNA performed on the 1Mb Chip. The average replication time of individual chromosomes was then used to normalise individual tile path arrays. The scaling factors for each array are shown in Table 2.5.

The 1Mb tiling path array was analysed using an Excel spreadsheet in the following way;

- Spots with less than twice the intensity of the average *Drosophila* spot intensity in either channel were rejected. Only spots with over twice the *Drosophila* background ratio in both channels were accepted and passed through to the next stage of analysis.
- A raw ratio of intensities was calculated by dividing the test (S phase) intensity by the reference (G1 phase) intensity.
- The raw ratios were normalised by dividing each individual ratio by the median of all ratios (for autosomes on the 1Mb array).
- The ratios were multiplied by the appropriate scaling factor to determine replication timing as shown in Table 2.5. This distributed all the ratios between 1 and 2.2.
- The median of the duplicates representing each locus was calculated.
- Spots that deviated more than 5% from this median are rejected and omitted from further analysis.
- The ratio taken for each locus was the median of accepted spots.
- The ratio of each locus was plotted against chromosome position.
- The average replication time of each chromosome was calculated to be used for the normalisation of individual tiling path arrays.

A second Excel spreadsheet was used to analyse the tiling path arrays;

- Spots with over twice the *Drosophila* background ratio in both channels were accepted and passed through to the next stage of analysis.

- A raw ratio of intensities was calculated by dividing the test (S phase) intensity by the reference (G1 phase) intensity.
- The raw ratios were normalised by dividing each individual ratio by the mean of all ratios (for the chromosome 22 clones only).
- The ratios were multiplied by the appropriate scaling factor to determine replication timing as shown in Table 2.5 (1.75 for replication timing arrays).
- The median of the triplicates representing each locus was calculated.
- Spots that deviate more than 5% from this median were rejected.
- If two or more spots are within 5% of the median, their average was taken as the final ratio at that locus. If two of more spots are not within 5% of the median all spots representing that locus are rejected.
- The ratio of each locus was plotted against chromosome position.

For the chromosome 22 tile path array ratios were then plotted on a graph against position of the midpoint of the clone on chromosome 22. The length and midpoints of clones were obtained from clones with end sequences by mapping the ends back against the published chromosome 22 sequence (Dunham, Shimizu et al. 1999). For clones where end sequences were unavailable the midpoint of the accessioned sequence was used.

2.5.2.2. Analysis of replication timing on arrays.

The analysis of replication timing experiments utilising the whole genome at a 1Mb resolution and a 22 tile path array are described in section 2.5.2.1 and normalised using the values reported in Table 2.5.

Tile path arrays for chromosome 6 and chromosome 1, produced in collaboration with our laboratory, were used to assess replication timing. The chromosome 1 array was constructed by Simon Gregory and Rachel Cooper, members of the chromosome 1 mapping group at the Sanger Institute. Replication timing experiments on the chromosome 1 tile path array were normalised to the average replication timing for chromosome 1 reported by the 1Mb resolution array i.e. 1.52. Array analysis was performed using a program written by Carol Scott at the Sanger Institute, available at http://intweb.sanger.ac.uk/cgi-bin/humace/1mbsetends.cgi.

The chromosome 6 tile path array was constructed in collaboration with Koichi Ichimura at Dept. of Pathology, University of Cambridge. The replication timing experiments performed on the chromosome 6 tile path array were normalised to 1.44. Analysis was performed using a spreadsheet provided by Koichi.

Analysis of the other types of array experiment was carried out using the same spreadsheet analysis as used for the replication timing arrays, described in section 2.5.2.1. For arrays assessing microdeletions, comparing cell lines or detecting breakpoints no scaling factor was used. For other arrays a scaling factor were applied as appropriate. (Table 2.5)

Table 2.5: Scaling factors for Microarray experiments

Array Experiment	Scaling Factor Applied
Replication Timing on 1Mb Array	1.44
Replication Timing on Chr 22 TP Array	1.75
Replication Timing on Chr 1 TP Array	1.52
Replication Timing on Chr 6 TP Array	1.44
Replication Timing on Chr 22 of t(17:22)	1.75

2.5.2.3: Analysis of chromosome 22 copy number change on the Chromosome 22 tile path arrays.

To verify that the chromosome 22 clones on the tile path array report the correct copy number changes, a series of experiments was performed adding a different amount of chromosome 22 into self:self hybridisations. To achieve this, copies of chromosome 22 were flow sorted to separate them from the rest of the genome. The DNA from the chromosomes was extracted as described in section 2.3.2.4. Four separate hybridisation experiments were performed as summarised in Table 2.6.

Table 2.6: Chromosome 22 add in experiments performed

Experiment	Cy3 labelled DNA	Cy5 labelled DNA
1	Genomic DNA	Genomic DNA
2	Genomic DNA + 1 copy of chromosome 22	Genomic DNA
3	Genomic DNA + 2 copies of chromosome 22	Genomic DNA
4	Genomic DNA + 4 copies of chromosome 22	Genomic DNA

Arrays were normalised on only the X clones to give a test:reference ratio of 1:1

2.5.3.4 Analysis of arrays reporting copy number change when different S phase fractions are hybridised onto the array.

S phase was sorted into four equal fractions based on the DNA content of the nuclei. Gates were placed on the cell cycle profile as shown in Figure 2.2b and nuclei into each of the S phase fractions, denoted S1, S2, S3 and S4 were sorted. G1 and G2/M nuclei were also sorted. DNA was extracted from the sorted nuclei as described in section 2.3. DNA from each fraction was hybridised against the G1 phase DNA. The arrays were normalised using the cell cycle profile by calculating the median value of each fraction as a proportion of total S phase.

The median value for each S phase fraction was expressed as a proportion of G1 to provide a scaling Figure between 1 and 2. These are summarised in Table 2.7.

Table 2.7: Scaling factor applied for the S phase fraction experiments

Array Hybridisation	Scaling Factor
G1:G1	1.00
S1:G1	1.17
S2:G1	1.36
S3:G1	1.53
S4:G1	1.71
G2/M:G1	2.00

2.5.3.5: Analysis of microdeletion studies.

All ratios obtained from arrays used in microdeletion studies were normalised so the average test:reference ratio was 1:1.

2.5.3.6: Analysis of DNA immunoprecipitated using antibodies against histone acetylation.

DNA from the lymphoblastoid cell line HRC 575 was assayed for histone acetylation on chromosome 22. This was done in collaboration with the Microarrays and Transcriptional Control group at the Sanger Institute. The chromatin immunoprecipitation was performed by Pawendeep Dhami using either of two antibodies, one for histone H3 acetylation and one for histone H4 acetylation. The Histone H3 antibody used was Anti-acetyl-Histone H3 (Upstate, USA). It is a rabbit polyclonal IgG antibody that recognises and is specific for acetylated human H3 of approx. 17kDa. The Histone H4 antibody used was Anti-acetyl-Histone H4, ChIP grade (Upstate, USA). It is extracted from rabbit antiserum. The antibody recognises acetylated histone proteins of aprox 10kDa. The antibody is known to cross react with acetylated histone H2B and may cross react with other acetylated proteins.

The arrays were normalised so that the average ratio of input DNA: immunoprecipitated DNA was 1:1.

2.6: Transcription analysis of a lymphoblastoid cell line.

2.6.1: Extraction of RNA from lymphoblastoid cell line.

Total RNA was extracted from the lymphoblastoid cell line HRC 575 using the Trizol purification method. Lymphoblastoid cells were cultured as described in 2.2.2.1 and harvested during the exponential stage of their growth. The cells were washed in PBS, quantified using a Haemocytometer and pelleted by centrifugation at 300g for 10 minutes. 1ml of Trizol (Gibco-BRL, Life Sciences) was added to every 10^7 cells and mixed thoroughly. The sample was incubated at room temperature for 5 minutes and

1ml samples were aliquoted into 2ml eppendorf tubes. 0.2ml of chloroform was added to each aliquot and mixed by vortexing at 15 seconds. Samples were incubated at room temperature for 3 minutes and centrifuged at 12000g for 15 minutes at 4^{o}C. The aqueous layer was transferred into a new 2ml eppendorf tube. 500µl of isopropanol was added and mixed by inversion. The sample was incubated at room temperature for 10 minutes and centrifuged at 12000g for 15 minutes at 4^{o}C. The supernatant was removed and the RNA pellet was washed with 1ml of 75% ethanol by centrifugation at 7500g for 5 minutes at 4^{o}C. The supernatant was removed and the pellet air dried. The pellet was resuspended in 50µl of HPLC water, 0.01% diethyl pyrocarbonate (DEPC) and incubated at 55^{o}C until the pellet was completely dissolved. The RNA was quantified using a spectrophotometer. 2µg was electrophoresed on a 1% agarose gel made up with TBE to assess the quality of the RNA. 1ml of 100% ethanol was added to the sample for storage at -70^{o}C.

2.6.2: Synthesis of cDNA

10µg of total RNA was incubated for 10 minutes at 65^{o}C with 100 pmol of a HPLC purified T7-(T)24 primer from the Superscript ds-cDNA Synthesis Kit (Gibco-BRL, Life Sciences). First strand buffer (1x concentration), dNTP's (10mM each) and DTT (final concentration 0.1M) was added to the RNA mix and incubated at 42^{o}C for 2 minutes. 200U/ml of superscript II reverse transcriptase was added, mixed and incubated at 42^{o}C for 1 hour. A second strand master mix was made (1x second strand buffer, dNTP's (10mM each) 10 U/µl *E. Coli* DNA Ligase, 10 U/µl *E. Coli* DNA Polymerase II, 10 U/µl *E. Coli* RNase H, made up to a reaction volume of 130µl with DEPC treated water). This was added to the first strand product, mixed thoroughly and incubated at 16^{o}C for 2 hours. 10 Units of T4 DNA Polymerase was added, the sample incubated at 16^{o}C for 5 minutes and the reaction quenched with EDTA pH8.0 (Final concentration 30mM).

Phase-lock tubes (eppendorf) were used to clean the cDNA. The phase lock tubes were prepared by centrifugation at 7700g for 30 seconds. An equal volume of room temperature buffer saturated phenol (65ml Alkalin buffer (10mM Tris HCl pH 8.0, 1mM EDTA) added to 1ml phenol:chloroform:isoamyl alcohol 25:24:1) was added to

the double stranded DNA and vortexed. The DNA mix was added to the phase-lock microfuge tube, spun at 7700g for 2 minutes and the aqueous top phase transferred to a fresh eppendorf tube. The DNA was precipitated in 0.5x vol 7.5M NH_4OAc, 4µl of glycogen (5 mg/ml) and 2.5x vol 100% ethanol. The cDNA was pelleted by centrifugation at 7700g for 20 minutes. The pellet was washed twice by the addition of 500µl of 80% ethanol to the pellet and centrifugation at 7700g for 5 minutes. The pellet was then air dried and resuspended in 12µl of DEPC treated water.

2.6.3: Production and labelling of cRNA and application to the array

Labelled cRNA was synthesised from the cDNA. Reagents from the Bioarray High Yield RNA transcript labelling Kit (Enzo Diagnostics) were used. The 12µl of ds-cDNA synthesised previously were used and the following reagents added; 10µl of DEPC treated water, 4µl 10x HY Reaction Buffer, 4µl Biotin-labelled ribonucleotides, 4µl DDT, 4µl RNase Inhibitor mix and 2µl T7 RNA Polymerase. This was mixed and incubated at 37^{o}C for 5 hours. The RNA was cleaned with the RNeasy mini kit (Qiagen). Samples with a yield greater than 40µg of cRNA were subsequently hybridised to Affymetrix U133 oligonucleotide arrays (Affymetrix). Hybridisation was at 45 °C for 16 hours.

2.6.4: Washing and analysis of array.

Arrays were washed and stained with streptavidin-phycoerythrin (SAPE, Molecular Probes) before signal amplification was performed using a biotinylated anti-streptavidin antibody (Vector Laboratories) following the recommended Affymetrix protocol for high density chips. Scans were carried out on a GeneArray scanner (Agilent Technologies). The fluorescence intensities of scanned arrays were analysed with Affymetrix GeneChip software. The Affymetrix Microarray Suite 5.0 was used for the quantification of gene expression levels. Global scaling was applied to the data to adjust the average recorded intensity to a target intensity of 100. Quantification data was exported from Affymetrix Microarray Suite 5.0 into Excel for further analysis. Presence or absence of gene expression was determined by a ‘present’ call,

in any of the oligos representing a gene, as determined by Affymetrix Microarray Suite 5.0.

2.7: FISH analysis of DiGeorge and VDJ recombination regions.

Deletions detected on the array were verified by fluorescence *in situ* hybridisation (FISH). The clones detailed in Tables 2.8 and 2.9 were picked for analysis.

Table 2.8: Clones picked from the VDJ recombination region;

Clone Name	**Accession Number**
cN9C5	D87023
cN9G6	D87020
cN22A12	D86999
cN24A12	D86998
cN29D3	D86991
cN31F3	D87002
cN35B9	D87010
cN48A11	D87007
cN50D10	D87011
cN52F2	D87006
cN61E11	D87014
cN63E9	D87013
cN68D6	D87015
cN75C12	D87017
cN81C12	AP000360
cN84E4	D87021
cN92H4	D87024
cN102D1	D86994

Table 2.9: Clones picked from the DiGeorge Region;

Clone Name	**Accession Number**
519d21	Ac008079
995o6	Ac008132
cN61D6	D87012
56c	Ac000080
Bac32	Ac007050
49c12	Ac000079
98c4	Ac000092
52f6	Ac005500
Pn_5	Ac002472
83c5	Ac000087

2.7.1: Mini Prep of Bacterial clone DNA

Clones chosen for verification were cultured in 10ml of TY media containing the appropriate antibiotic (see 2.1.1.2). The culture was grown at 37^{o}C for 16 hours whilst shaking at 200rpm. The culture was centrifuged at 2,000g for 10 minutes to pellet the bacteria. The pellet was resuspended in 200µl of Lysis buffer (10mM EDTA, 250mM Tris pH8.0, 50mM glucose, made with sterilised distilled water) and incubated at room temperature for 10 minutes. 400µl of 4M NaOH, 1% SDS was added and the preparation was incubated on ice for 5 minutes before 300µl of 3M Sodium acetate pH5.2 was added followed by a further 10 minute incubation on ice. The preparation was centrifuged at 7700g for 5 minutes and the supernatant was transferred into a fresh 1.5ml eppendorf tube. This was repeated until a clear supernatant was obtained. 600µl of isopropanol was added and mixed gently with the supernatant, this was incubated at –70^{o}C for 10 minutes.

The eppendorf tube was spun at 7700g for 5 minutes, the supernatant tipped off and the pellet resuspended in 200µl of 0.3M Sodium acetate pH7. 200µl of phenol:chloroform:water was added, mixed by vortexing and the eppendorf tube was spun at 7700g for 3 minutes. 150µl of the top aqueous layer was transferred into a fresh 1.5ml eppendorf tube. A further 50µl of 0.3M Sodium acetate pH 7.0 was added to the original phenol:chloroform:water containing tube; this was mixed by vortexing and the tubes were centrifuged at 7700g for 2 minutes. 50µl of the aqueous top layer was pooled with the first 150µl; 200µl of isopropanol was added, mixed by inversion and incubated at –70^{o}C for 10 minutes. The tubes were centrifuged at 7700g for 5 minutes and the supernatant discarded. The pellet was washed in 500µl of ice-cold 70% ethanol and spun at 7700g for 2 minutes. The supernatant was removed, the pellet dried and resuspended in 50µl of T0.1E containing 200mg/ml of RNAase A and incubated at 55^{o}C for 15 minutes. 1µl of each DNA preparation was run on a 1% agarose gel and the DNA was quantified using a TD-360 flurometer as previously described.

2.7.2: Nick Translation

Nick translation was performed to label the DNA with a dUTP conjugated to either the hapten biotin-16-dUTP (Roche) or digoxigenin-11-dUTP (Roche). 1μg of DNA prepared as described in 2.7.1 was used as the input DNA to the reaction. A reaction mix was made up using the input DNA, 1xNT buffer (50mM TrisHCl pH7.5, 10mM $MgSO_4$, 0.1mM dithiothretitol, 50μg/ml Bovine serum albumin), 10mM dATP, 10mM dCTP, 10mM dGTP, 30mM hapten-dUTP, 0.02 units/ml DNA polymerase 1 and 1μl Deoxyribonuclease 1 (DNAse 1 - Sigma.)

Concentration and incubation time was determined by a DNAase 1 titration for each separate vector (cosmid, fosmid, BAC and PAC). DNAse 1 was diluted in 50% glycerol, 100mM TrisHCl pH7.5, 20mM $MgSO_4$, 0.2mM dithiothretitol, 100μg/ml Bovine serum albumin to a concentration of 1μg/ml. If PAC or BAC DNA was used as input DNA the nick translation was incubated at 14^{o}C for 70 minutes. If cosmid or fosmid DNA was used as input DNA the nick translation was incubated at 14^{o}C for 40 minutes.

The reaction was stopped by the addition of 2.5μl of 0.5M EDTA. 2.5μl of 3M Sodium acetate pH7 was also added. The DNA was precipitated by the addition of 1000μl of absolute ethanol and incubation at –70^{o}C for 30 minutes. The pellet was washed in 500μl of 80% ethanol and the pellet resuspended in 10μl of T0.1E. 1μl was run on a 1% agarose gel to confirm a product size between 200-500bp.

2.7.3: Metaphase spread preparation

Metaphase preparations were made from two lymphoblastoid cell lines, HRC 575 (male) and HRC 160 (female). The lymphoblastoid cell lines were cultured as described in 2.3.3.1. Twenty-four hours after sub-culturing, BromodeoxyUridine (Roche) was added to a final concentration of 15μg/ml. The culture was incubated at 37^{o}C for three hours. Ethidium Bromide (Sigma) was then added to a concentration of 10μg/ml and the culture incubated at 37^{o}C for a further 75 minutes. Colcemid (Gibco-BRL, Life Sciences) was added to a final concentration of 0.1μg/ml and the culture

incubated for 45 minutes. The culture was transferred to 50ml Falcon tubes, centrifuged at 300g for 10 minutes and the supernatant was removed leaving the pellet as dry as possible. The tube was flicked to loosen the pellet and 10ml of 75mM KCl (pre-warmed to 37^{o}C) was added to resuspend the pellet. The suspension was transferred to a 15ml Falcon tube and incubated at 37^{o}C for 10 minutes. Ice cold fixative was prepared (3 parts dried methanol: 1 part glacial acetic acid- BDH) and 3ml was added to the cell suspension and mixed by gentle swirling. The suspension was centrifuged at 300g for 10 minutes. The supernatant was removed and the pellet resuspended in 10ml of ice-cold fix. Centrifugation and resuspension in fix was repeated twice more. After a final centrifugation at 300g for ten minutes the pellet was resuspended in enough fix to give a desired density of nuclei.

2.7.4: Hybridisation to Metaphase spreads

Slides were sonicated in 2% Decon solution, washed in 96% ethanol and polished dry using lint free tissue. A drop of metaphase suspension was applied to each end of the slide and allowed to spread and air dry. The slides were placed in a jar of fix (3 parts dried methanol: 1 part glacial acetic acid) at room temperature for 30 minutes. Slides were then dehydrated in a series of ethanols (70%, 70%, 90%, 90%, 100%) and air dried. The slides were fixed for 10 minutes in acetone and baked at 60^{o}C for two hours.

The nick translated probe was prepared for hybridisation to the slide. 0.5µl of probe was added to 1µl of Cot 1 DNA (Roche) and 11.5µl of Hybridisation buffer. This probe mix was denatured at 65^{o}C for 10 minutes before being pre-annealed at 37^{o}C for 30 minutes. Meanwhile the metaphase spread slides were denatured in 70% formamide, 0.6xSSC at 65^{o}C for 2 minutes. The denaturation was quenched in 70% ice cold ethanol and dehydrated through an ethanol series (70%, 70%, 90%, 90%, 100%) before air drying. The probe mix was applied to the metaphase spread and sealed under a 22x22mm coverslip with rubber cement. The slides were incubated overnight at 37^{o}C in a humid atmosphere.

After the overnight hybridisation the rubber cement was removed from the slide and the coverslips soaked off in 2xSSC. Slides were washed in 2xSSC, then 2 washes in

50% formamide, 1xSSC, followed by a wash in 2xSSC. All washes were performed at 42^{o}C for five minutes each. Slides were then mounted on a Cadenza immunostainer (Shannon).

2.7.5: Detection of labelled probes.

Three-layer detection was performed on the Cadenza. Antibodies for detection were diluted in blocking buffer (1% w/v blocking reagent (Boehringer), 0.05% Tween 20 (BDH), 1µl/ml Sodium azide, 4xSSC) and the Cadenza was used to incubate the antibody on the slide. Blocking buffer was applied to the slide before three-layer detection took place. The first layer was 4µg/ml avidin conjugated to Texas Red (Molecular Probes). The second layer was 4µg/ml biotinylated anti-avidin (Vector Labs) and/or 1:500 dilution of mouse anti-digoxin (Sigma). The third layer was 4µg/ml Avidin-Texas Red and/or 10µg/ml goat anti-mouse FITC (Sigma). In between incubation with the antibodies the slides were washed with 4xSSC, 0.05% Tween 20. After staining was completed the slides were removed from the Cadenza, washed in 2xSSC and stained with 0.08µg/ml 4,6-Diamidine-2-phenylindole dihydrochloride (DAPI - Roche) in 2xSSC. The slides were then rinsed in 2xSSC, dehydrated in an ethanol series (70%, 70%, 90%, 90%, 100%) and air dried. 20µl aliquots of anti-fade solution (Citifluor – Citifluor Ltd) were applied to each cell spot and a 22mm x 50mm coverslip was laid over the slide and fixed in place with clear nail polish.

2.7.6: Acquisition of FISH images.

Slides were studied on a Zeiss Axioskop epifluoroescence microscope. Metaphase spreads were located at x200 magnification using a DAPI filter (exitation λ362nm emission λ465nm). Probes on the metaphases were detected using either a FITC filter (excitation λ495nm emission λ520nm) or a Texas Red filter (excitation λ596nm emission λ620nm) and captured at x1000 magnification using the Cytovision capture suite.

2.8: Real-Time PCR analysis of S and G phase DNA.

Selected Early and Late replicating Clones were chosen to verify the array results using real-time (quantitative) PCR.

2.8.1: Primer design

Four clones were chosen from the array for assessment by real time PCR. The clones chosen were a single late replicating clone (cN69F4), two mid replicating clones (cE140F8 & cB13C9) and an early replicating clone (bK57G9). Short amplicons (approximately 150bp) were designed as they are optimal for the real time PCR reaction. A pair of PCR primers were placed approximately every 10Kb along the selected clone using the Taqman primer design program, part of the Primer Express Software (ABI). Primer sequences had a Tm between 58-60^{o}C, a GC content between 20-80% (no more than 3/5 GC's in the 5' end of the primer), and a short amplicon (approx 150bp).

Primer pairs were checked using the Primer Test program in the Primer Express software to ensure they did not have excessive secondary structure or primer-dimer formation. Primer sequences are shown in Appendix 3.

2.8.2: Real Time PCR on S and G1 phase DNA.

Real time PCR was carried out using the same S and G phase DNA sorted in 2.3.2.3 and used for replication timing analysis. The DNA was first diluted to a concentration of approximately 5ng/ml to ensure quantisation by real-time PCR fell within the scope of the real-time PCR machine.

For each reaction, 12.5µl of 2x Sybr Green Reaction mix (including Taq polymerase and dNTP mix - ABI), 1µl of template DNA, 50mM forward primer and 50mM reverse primer was used. The reaction mix was made up to 25µl with sterile water. Each reaction was carried out on each plate in triplicate. A control PCR on each 96 well plate comprising standard male DNA (ABI) and a control primer was also included to allow comparison between different PCR runs.

The reaction had an initial incubation at 50^{o}C for 2 minutes followed by holding at 95^{o}C for 10 minutes. A two-step PCR was then carried out with 40 cycles of a denaturation step at 95^{o}C for 15 sec, and a hybridisation and elongation step at 60^{o}C for 60 sec. During each cycle of the PCR the Sybr-Green intensity of each well was measured. The amount of DNA present in the original sample is proportional to the cycle of PCR at which Sybr-Green intensity first appears. After the PCR was completed a standard disassociation curve for Sybr-Green was performed. This ensured the PCR had been successful, only a single product had been produced, and that primer dimers had not been produced. Under these conditions the Sybr-Green intensity measured during the PCR cycles is due to the amplification of the target and is quantitative.

The Sybr-Green intensities collected during the PCR reaction were analysed using the Sequence Detection System Software (ABI). Background threshold levels were set at the number of cycles before any Sybr-Green fluorescence was detected. The detection level was set at the point where the increase in Sybr-Green level became exponential. The cycle number at which the detection level is set is a measure of the DNA concentration in the original sample. This was compared to an internal standard control curve (0.625ng, 1.25ng, 2.5ng, 5ng and 10ng of DNA) to determine the starting quantity of DNA present. The ratio of amount of S:G1 phase DNA for each set of primers designed and the average ratio of S:G1 phase DNA for each clone was then calculated.

3: Results 1
Pilot Replication Timing Studies Utilising a Genomic Array Representing 4.5Mb of Chromosome 22 Sequence.

3.1: Introduction.

Initially a small array was constructed to test the use of microarrays for the evaluation of replication timing. This array was designed to cover 4.5Mb of chromosome 22 spanning a boundary between a G dark and a G light band. A total of 83 overlapping clones, including cosmids, fosmids, PACS and BACS, were chosen between BACs CTA-415G2 and CTA-390B3 (inclusive) spanning the region from approximately 17Mb to 21.5Mb along the q arm of chromosome 22q.

3.2: Initial verification experiments on the 4.5Mb array.

To assess the systematic variation in measurements on the array a self:self hybridisation was performed. In this assay, DNA from the same source is used both as the test and the reference DNA and by definition all clones should report a 1:1 ratio. In this case, a self:self hybridisation was conducted on the 4.5Mb test array using differentially labelled DNA from the same sort of G1 nuclei. Analysis of the array rejected 11 out of the 83 points because the triplicate values were not within the rejection criteria described in section 2.5.2 . Briefly, to be included in the analysis, all loci were required to report intensities at least twice the values reported by the *Drosophila* clone DNA loci represented on the array. Triplicate spots were also required to report ratios within 5% of each other. The remaining 72 clones were available for further analysis. The mean G1:G1 ratio was 1.035 with a standard deviation of 0.0714. The G1:G1 ratios plotted against the mid point position of each clone on the array is shown in Figure 3.1. The distance between data points reflects the size of the clone used in the golden path of sequencing clones – a high proportion of cosmids were used between 17 and 18.5Mb while larger insert clone BAC and PACs predominate between 18.5 and 22.6Mb.

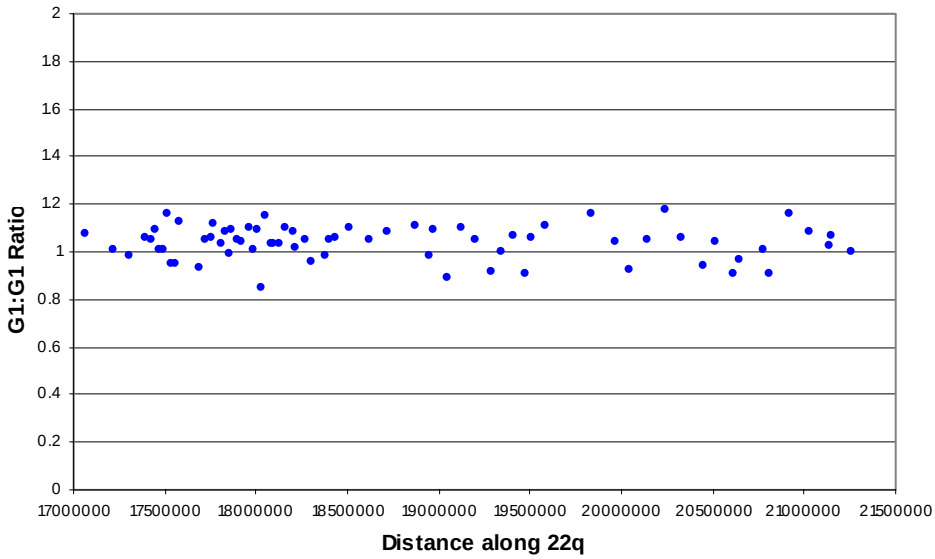


Figure 3.1: G1 self:self hybridisation performed on a 4.5Mb array. Data points on the X axis correspond to the position of the midpoint of each clone, and the Y axis shows the G1:G1 ratio.

3.3: S phase DNA: G1 phase DNA Hybridisation on the 4.5Mb Test Array.

Two S:G1 hybridisations were conducted on the 4.5Mb test array using differentially labelled S and G1 DNA sorted from the same preparation of nuclei. All 84 data points were included in the analysis for both replicates; no data points were rejected at the analysis stage. The mean ratio for each clone is shown in Figure 3.2a, vertical error bars representing one standard deviation on each clone, whilst horizontal error bars represent the extent size of the clone. Replicate experiments are shown in Figure 3.2b. Ratios close to 2:1 indicate early replicating regions whilst loci reporting ratios close to 1:1 replicate late.

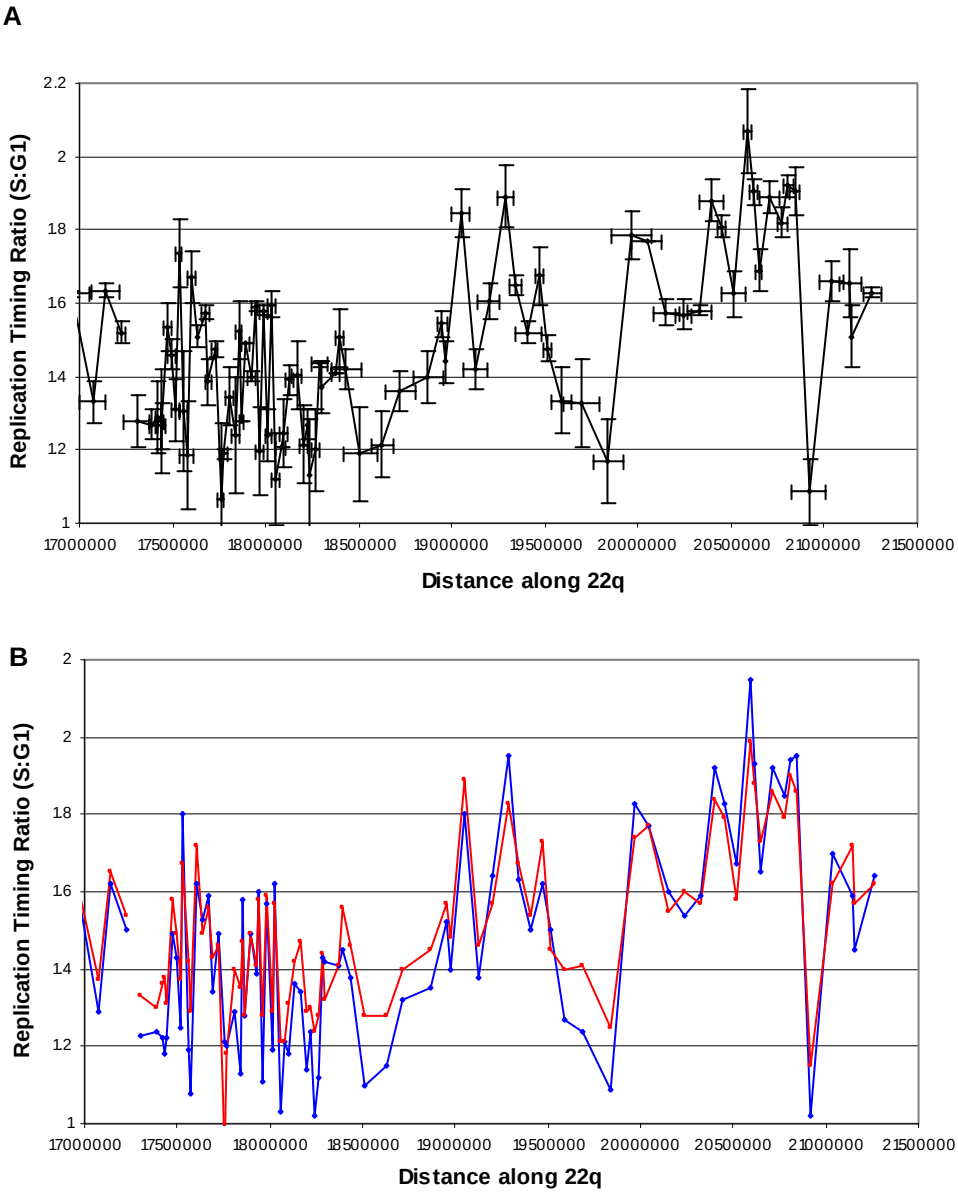


Figure 3.2: replication timing profiles for a 4.5 Mb region of chromosome 22q. A: Average S:G1 ratio of two arrays. A single standard deviation of the two arrays is indicated by the Y error bars. The X error bars represent the length of the clone and indicates the size of the overlap between clones. B: Replication profiles of two individual replicates.

3.4: Correlation between replication timing and sequence features.

The replication timing across the 4.5Mb region of chromosome 22 was also correlated with the guanine and cytosine (GC) content of each clone and the density of introns of genes within each clone.

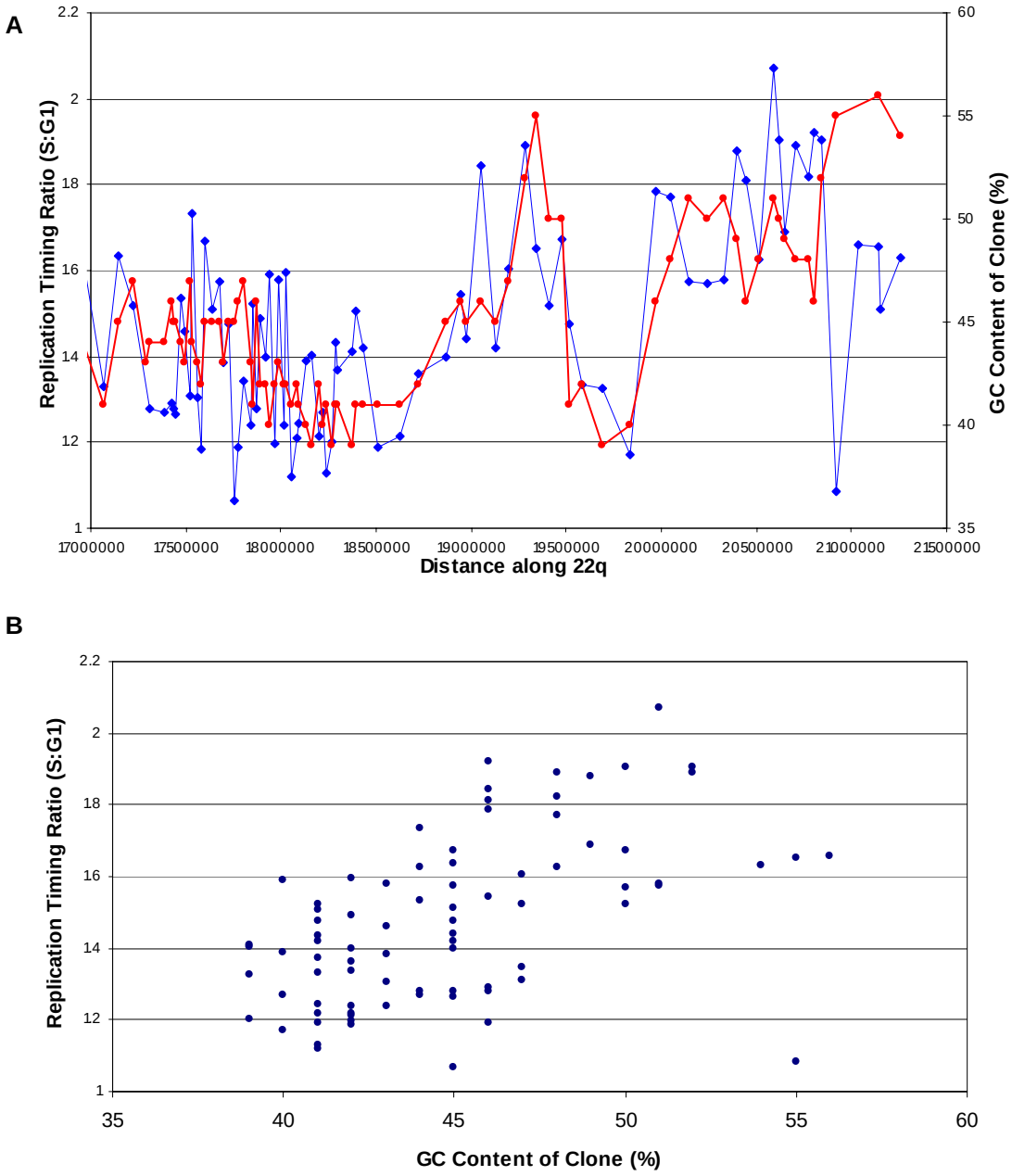


Figure 3.3: Correlation between replication timing and GC content over a 4.5 Mb region. A: Replication timing (blue) and GC content profile (red). B: Replication Timing versus GC Content. The equation of the best fit line through the data points is y = 0.03x + 0.127 with a correlation coefficient of 0.53.

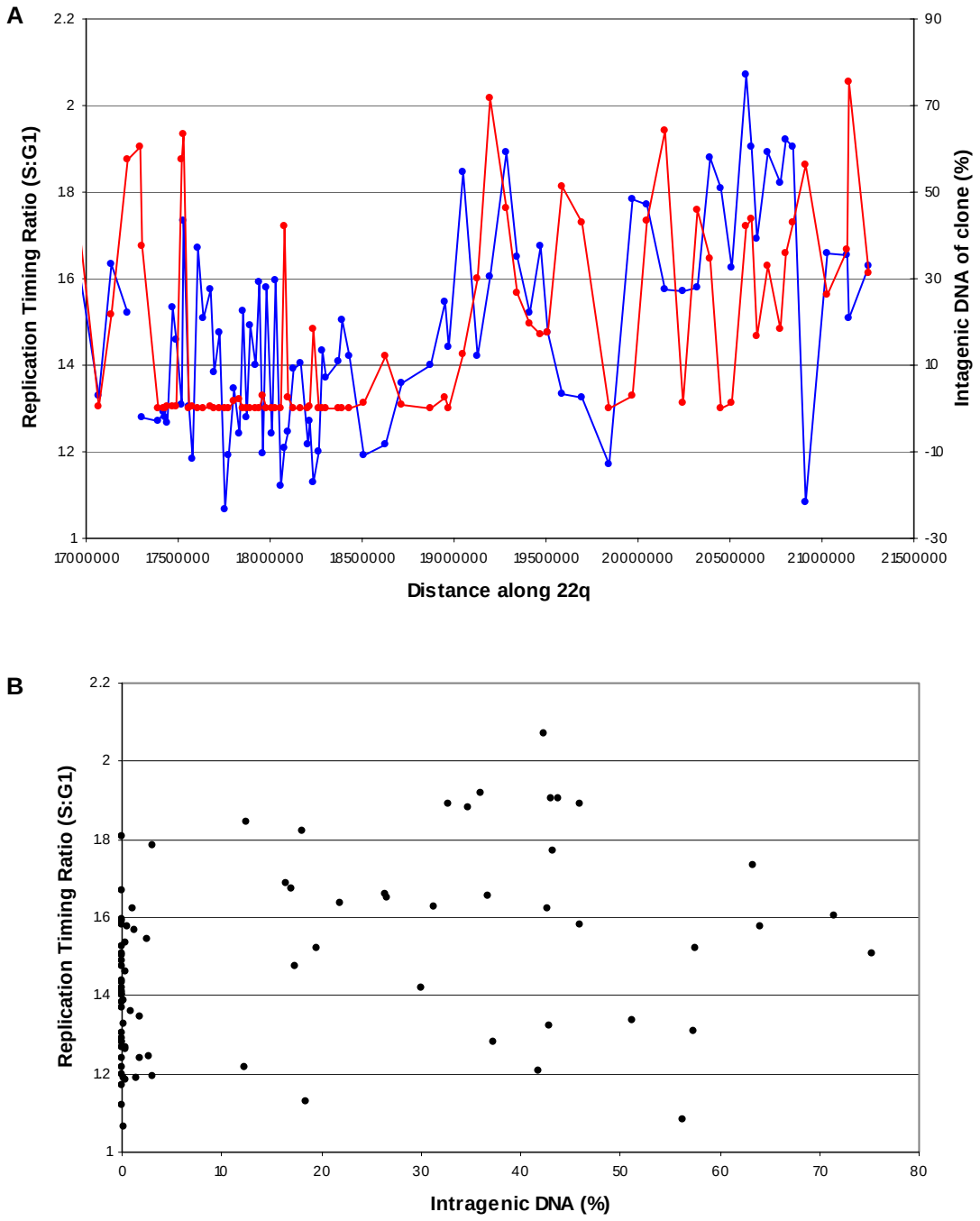


Figure 3.4: Correlation between replication timing and intragenic DNA over a 4.5 Mb region. A: Replication timing (blue) and intragenic DNA (red) profile. B: Replication timing versus intragenic DNA. The equation of the best fit line through the data points is y = 0.003x + 1.41. The correlation coefficient is 0.36

These preliminary experiments showed clear reproducible differences between the replication timing ratios reported for the different loci represented on the array.

3.5: Discussion.

To verify the assessment of replication timing on arrays a small region of chromosome 22 was chosen for study. A microarray was constructed from sequencing clone DNAs across a 4.5Mb region located 17.5 - 21Mb along chromosome 22. The 4.5Mb clone array was initially verified with a G1 self:self hybridisation. A 1:1 ratio should be reported on all clones. An average ratio of 1.035 was reported on the array. A low standard deviation of 0.0714 was observed indicating that all the clones hybridise in a similar fashion.

An early replicating region of the genome will contain twice as much DNA throughout S phase as it will during G1 phase and so the ratio reported will be 2:1. Conversely, a late replicating region of the genome will not duplicate its DNA until the end of S phase so that in this assay a ratio of close to 1:1 will be generated. In this way, the replication timing can be reported as a ratio of S:G1 DNA which should vary between 1 and 2. On this test array, all the ratios (except one) were between 1 and 2. The clones with mid points between 17-18.5 Mb along the q arm of chromosome 22 replicated later than clones with midpoints between 18.5-21Mb. This is consistent with the fact the proximal region corresponds to a G dark band and the distal region corresponds to a G light band (Strehl, LaSalle et al. 1997).

Comparison of the S:G1 ratio with GC content showed that GC rich clones generally reported earlier replication timing ratios than GC poor clones. This is in agreement with of previous studies (Tenzen, Yamagata et al. 1997; Watanabe, Fujiyama et al. 2002). The correlation with density of intragenic DNA was less clear. However a positive correlation was still observed as has been reported previously (Strehl, LaSalle et al. 1997; Cook 1999; Gilbert 2002).

These initial experiments on the 4.5Mb region validated this assay as having the potential to accurately assess replication timing and it was decided to expand this approach by developing a microarray covering the whole of human chromosome 22q.

4. Results 2
Preparation and Verification of the Genomic Microarrays

4.1: Introduction

To investigate replication timing and to correlate this with genome features at high resolution, a tile path genomic array using large insert clone DNA was constructed to cover the whole of 22q. The tile path resolution utilised overlapping sequencing clones giving an average resolution of 78Kb. After construction the array was verified extensively to assess reproducibility and response to copy number changes.

The replication timing assay entails the flow sorting of nuclei from the G1 and S phases of the cell cycle. To allow rapid sorting, the number of S phase cells within the population of nuclei to be sorted was optimised by adjusting the time of growth between sub-culture and harvest of the cells. This is described in section 4.2.

The construction and verification of the 22q tile path array is described in section 4.3. Array verification experiments were also performed on pre-constructed arrays assessing the entire genome at a 1Mb resolution. This is described in section 4.4.

At a later stage in the project, an array was also constructed, with a 10Kb resolution over a 4.5Mb region of chromosome 22q with 500bp PCR products. This array also contained a 200Kb region covered with overlapping 500bp PCR products. The verification of this array is described in section 4.5.

A further stage of array verification was to test whether each loci on the 22q tile path array responded to chromosome 22 copy number change. This was achieved by adding DNA from flow sorted chromosome 22 to one half of a G1:G1 hybridisation. This is described in section 4.6.

4.2: Optimisation of S phase fractions

The assessment of replication timing in cells developed in this study is dependant on the ability to flow sort S and G1 phase nuclei. In any unsynchronised cell population the majority of the cells are in the G1 phase. The time taken to sort the S phase fraction is thus a limiting factor. To minimise the number of cells and the time required for flow sorting, the optimum time to yield the maximum number of cells in S phase after subculture was assessed.

The male lymphoblastoid cell line HRC 575 (46, XY) was harvested at different intervals after subculture and passed through a flow sorter to obtain a cell cycle profile as described in 2.3.2.2. The percentage of cells in S phase was plotted against the time between sub culture and harvest (Figure 4.1).

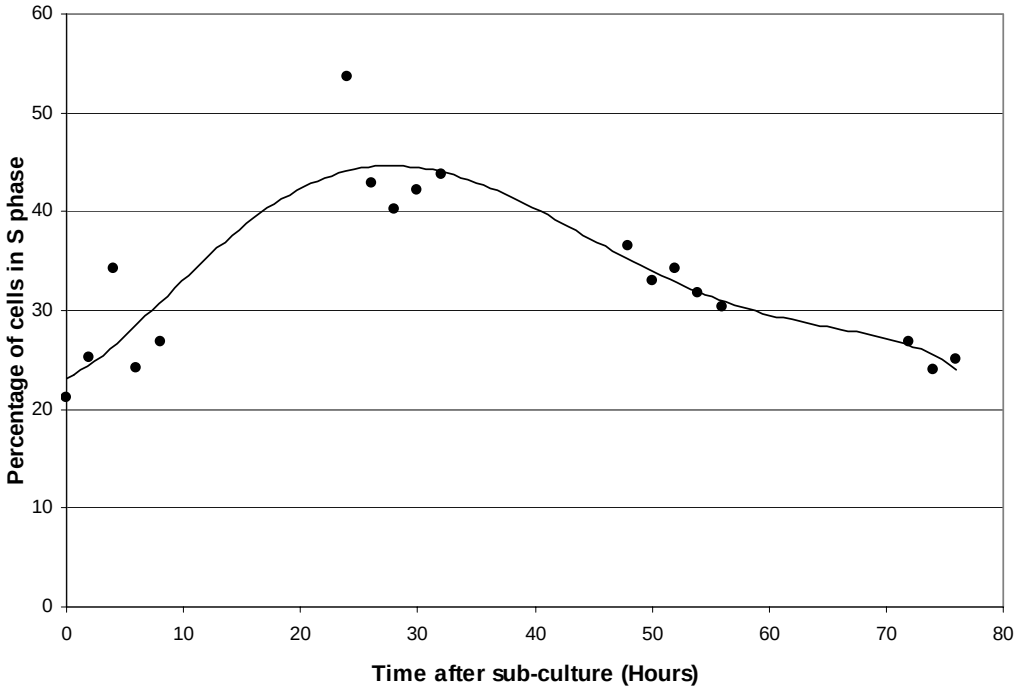


Figure 4.1: The change in the proportion of the cells in S phase at times after subculture for a lymphoblastoid cell line.

The optimal time between sub-culture and harvest of the lymphoblastoid cell line for a maximal S phase fraction was approximately 26 hours. The flow sort profile 26

hours after subculture is shown in Figure 4.2. This shows a high proportion of cells in S phase.

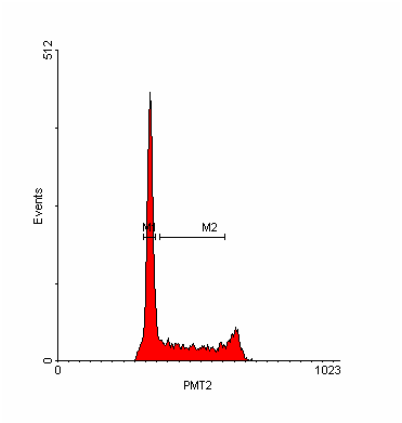


Figure 4.2: Lymphoblastoid nuclei flow sort profile after harvest 26 hours from subculture.

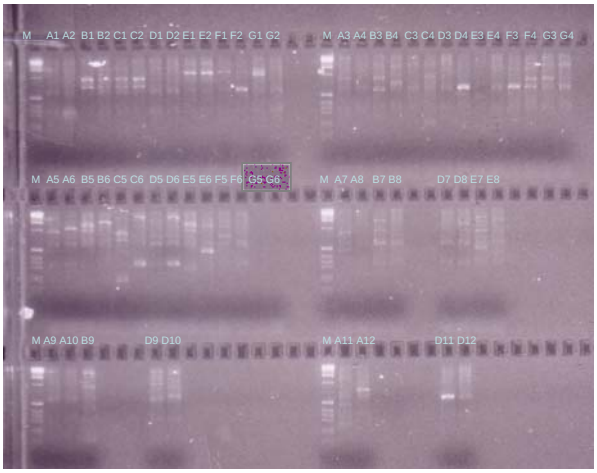
4.3: Preparation and initial verification of the 22q tile path array.

A 22q tile path array was constructed as described in 2.1. This comprised 470 clones, including cosmids, fosmids, PACS and BACS and covered the whole of the q arm of chromosome 22. 95 Chromosome X clones were also spotted onto the array. These were used as an intrinsic control to measure single copy number changes in male versus female DNA hybridisations.

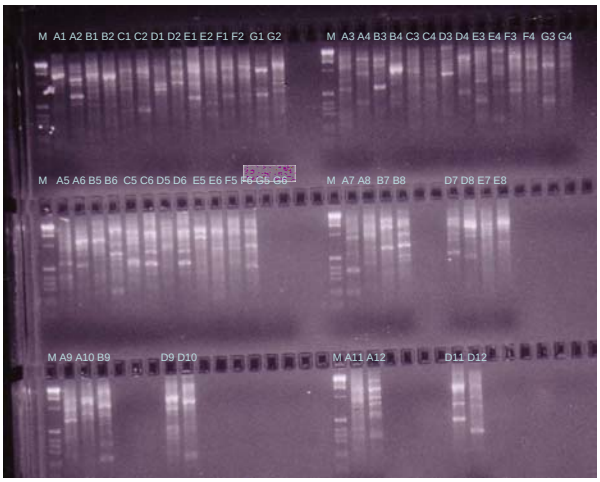
4.3.1: Amplification of chromosome 22 tile path clones.

The clones were first amplified by degenerate oligonucleotide primed PCR (DOP-PCR) using three different primers as described in 2.1.2.1. To ensure the PCRs had been successful and that no contamination had taken place, 5µl of the PCR product was assessed by gel electrophoresis as shown in Figure 4.3.

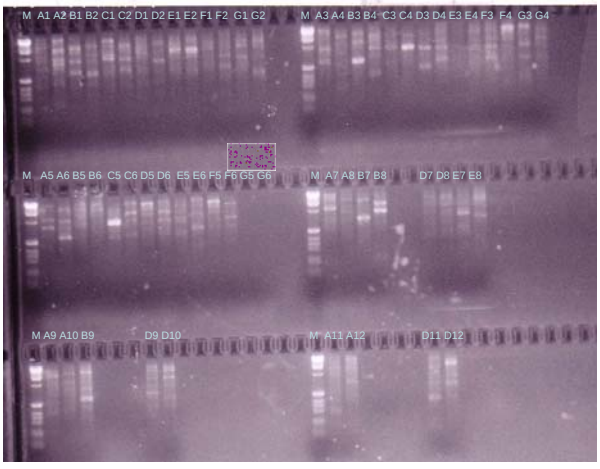
Clones were then amplified by a second round amino-linking PCR as described in 2.1.2.2 (see Figure 4.4).



DOP 1



DOP 2

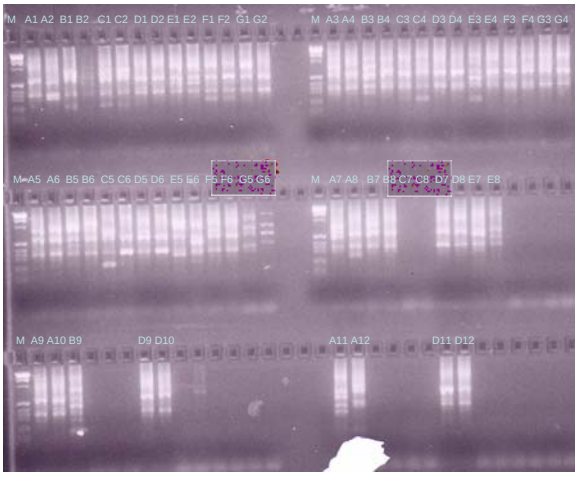


DOP 3

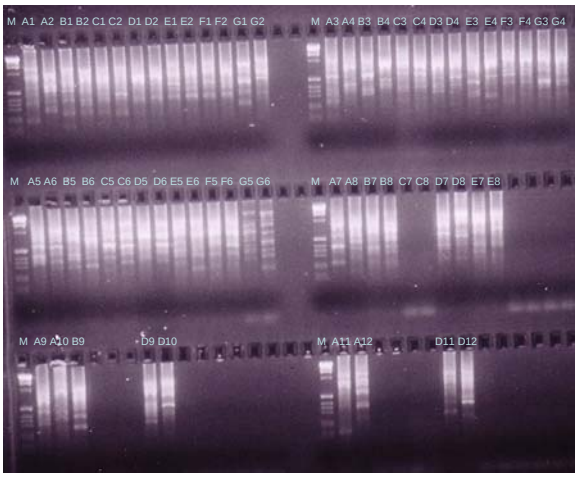
Key:

A1	p393	C1	100h	E4	bK217C2
A2	dJ1172A22	C2	91c	E5	bA140I15
A3	dJ127B20	C3	8C	E6	bK358H9
A4	dJ366L4	C4	cN81C12	E7	bK1109B5
A5	dJ403E2	C5	cN74G7	E8	kB63E7
A6	dJ20O8	C6	cB23F1	F1	cN61N10
A7	dJ101G11	D1	b445	F2	444P20
A8	dJ697G8	D2	b9j16	F3	174C04
A9	dJ437O22	D3	kB1027C1	F4	188L13
A10	dJ1014D13	D4	kB1269D1	F5	61P17
A11	dJ319F24	D5	kB1572G7	F6	73P10
A12	dJ340K22	D6	kB113H7	G1	bA9F11
B1	dJ796I17	D7	kB1839H6	G2	bA354I12
B2	dJ549K18	D8	kB1125A3	G3	bA255N20
B3	dJ477H23	D9	bK322B1	G4	fF96H12
B4	dJ162H14	D10	kB1561E1	G5	negative ctl
B5	dJ37M3	D11	kB1674E1	G6	negative ctl
B6	dJ477J10	D12	kB1896H1		
B7	dJ100G10	E1	kB1195A5		
B8	dJ34P24	E2	bK437G10		
B9	dJ566L20	E3	bK414D7		

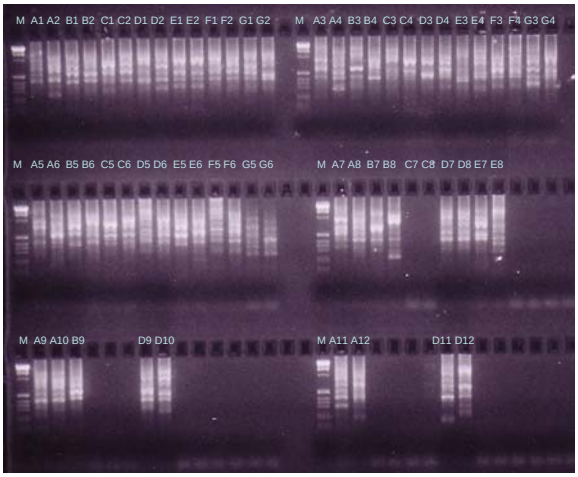
Figure 4.3: DOP-PCR amplification of a selection of chromosome 22 tile path clones, as indicated in the key.



Amino 1



Amino 2



Amino 3

Key:

A1	p393	C1	100h	E2	bK437G10
A2	dJ1172A22	C2	91c	E3	bK414D7
A3	dJ127B20	C3	8C	E4	bK217C2
A4	dJ366L4	C4	cN81C12	E5	bA140I15
A5	dJ403E2	C5	cN74G7	E6	bK358H9
A6	dJ20O8	C6	cB23F1	E7	bK1109B5
A7	dJ101G11	C7	negative ctl	E8	kB63E7
A8	dJ697G8	C8	negative ctl	F1	cN61N10
A9	dJ437O22	D1	b445	F2	444P20
A10	dJ1014D13	D2	b9j16	F3	174C04
A11	dJ319F24	D3	kB1027C1	F4	188L13
A12	dJ340K22	D4	kB1269D1	F5	61P17
B1	dJ796I17	D5	kB1572G7	F6	73P10
B2	dJ549K18	D6	kB113H7	G1	bA9F11
B3	dJ477H23	D7	kB1839H6	G2	bA354I12
B4	dJ162H14	D8	kB1125A3	G3	bA255N20
B5	dJ37M3	D9	bK322B1	G4	fF96H12
B6	dJ477J10	D10	kB1561E1	G5	negative ctl
B7	dJ100G10	D11	kB1674E1	from DOP	
B8	dJ34P24	D12	kB1896H1	G6	negative ctl
B9	dJ566L20	E1	kB1195A5	from DOP	

Figure 4.4: Amino-linking PCR amplification of a selection of Chromosome 22 tile path clones as indicated in the key.

4.3.2: Male:male hybridisation onto the chromosome 22 tile path array.

A male self:self hybridisation was carried out using DNA extracted from HRC 575 lymphoblastoid cell line to assess the background variation in measurements.

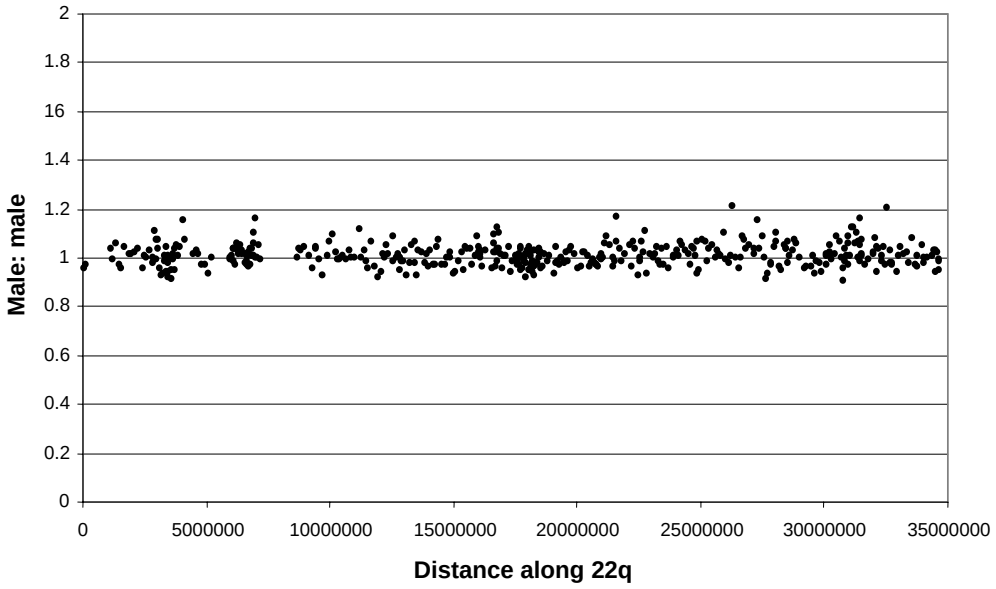


Figure 4.5: Male self:self hybridisation on the chromosome 22q array

The self:self hybrisiation was performed in duplicate. Four of the 470 clones represented on this array were excluded by the analysis program, because the clone intensities were not sufficiently above the *Drosophila* BAC clone background level, or because the triplicate spots were not all within 5% of the mean for that triplicate (for details of the analysis see section 2.5.2.). The average ratio of all the chromosome 22 clones was 1.00 with a standard deviation of 0.04.

4.3.3: Male:female hybridisation onto the array.

A male:female hybridisation was carried out using differentially labelled DNA extracted from a male lymphoblastoid cell line (HRC 575) and a female lymphoblastoid cell line (HRC 160). The aim of this experiment was to verify that X clones on the array accurately reported a single copy number difference between the male and female DNA (i.e. a ratio of 0.5).

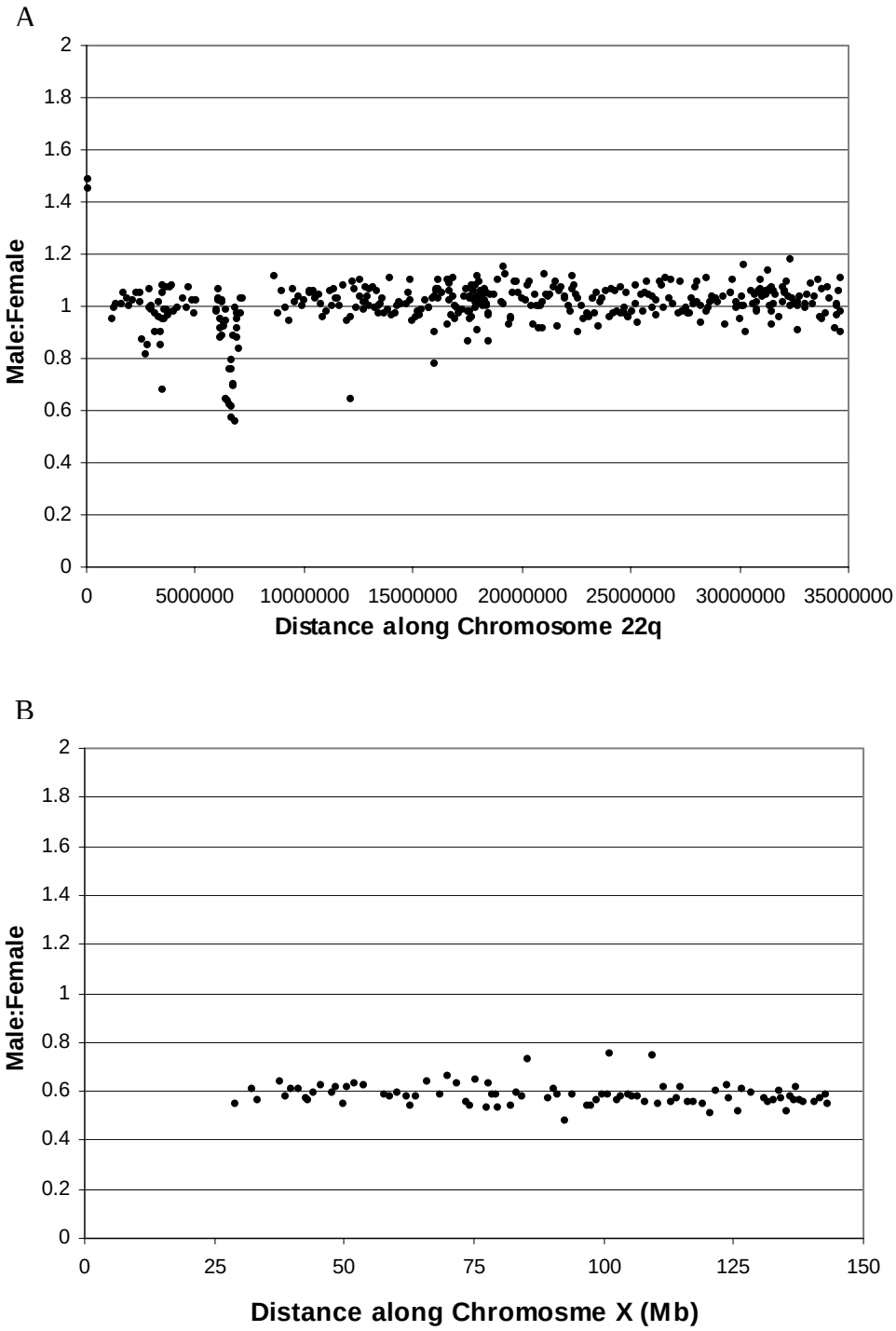


Figure 4.6: Male:female hybridisation on the constructed 22q array. A: Male:Female ratio on the 22q clones. B: Male:Female ratio on the X clones on the 22 array.

Fifty of the 470 clones were rejected during the analysis stage. The average male:female ratio on the chromosome 22 clones was 1.00 with a standard deviation of 0.09 (Figure 4.6). The region 6433945 – 6823353bp along the q arm of chromosome 22 shows clone ratios that could be interpreted as either a single copy

deletion within the male cell line, or a gain in the female cell line. This was further investigated (detailed in section 7.3) and revealed a deletion in the male cell line. Omitting this region from the statistical analysis, the standard deviation of the data points reduces to 0.07. Other clones on the chromosome 22q tile path array also reported unexpected ratios and are summarised in Table 4.1.

Table 4.1: Clones showing unexpected ratios in a male:female hybridisation

Clone	**Clone position (bp)**	**Possible reason for aberrant ratio**
cN14H11	99514	Centromeric clone
cN64E9	114958	Centromeric clone
59f	3467897	Rich in low copy repeats
995o6	2710127	Rich in low copy repeats
699j1	2822641	Rich in low copy repeats
519d21	2577096	Rich in low copy repeats
83e8	3443824	Rich in low copy repeats
dJ477H23	12110340	Clone not verified
cN113A11	16010331	Clone not verified

Seven of the 93 chromosome X clones were rejected by the analysis criteria described in section 2.5.2. The average male:female ratio on the chromosome X clones was 0.58 with a standard deviation of 0.04.

4.3.4: G1 self:self phase DNA Hybridisation onto the 22q tile path array.

A G1 self:self hybridisation was carried out to assess whether extraction from cell sorted nuclei affected ratio measurement variance. DNA was obtained from the G1 phase of the cell cycle as described in 2.3.2, differentially labelled and hybridised to the 22q array.

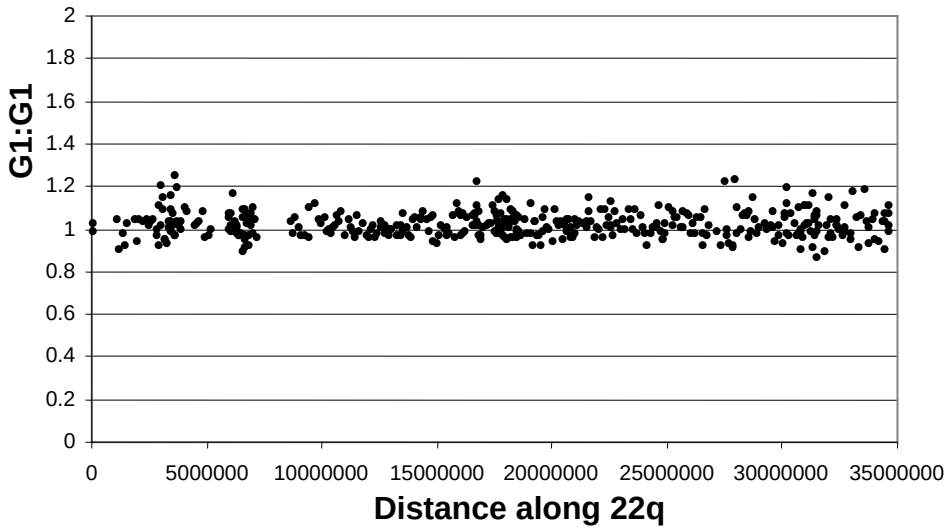


Figure 4.7: G1:G1 hybridisation on the 22q array.

Eighteen of the 444 clones were excluded from the analysis according to the criteria described in 2.5.2. The mean ratio reported was 1.00. The standard deviation of the ratios was 0.06.

4.4: Control Hybridisations on the 1Mb array

Similar verification experiments as detailed above were performed on an array sampling the whole genome at a 1Mb resolution.

4.4.1: Male:male hybridisation on the 1Mb array

A single male:male hybridisation was carried out. Of the 3126 clones on the array 82 were excluded at the analysis stage. The mean ratio reported by the remaining clones was 1.00 with a standard deviation of 0.039. The ratio profiles for all chromosomes can be seen in Appendix 4.

4.4.2: Male:female hybridisation on the 1Mb array

A male:female hybridisation was carried out on the 1Mb array. Of the 2955 autosomal clones on the 1Mb array, 256 were excluded at the analysis stage. The average ratio reported was 1.00 with a standard deviation was 0.10. Of the 150 Chromosome X clones on the array, 17 were excluded at the analysis stage. The average ratio reported was 0.75 and the standard deviation was 0.051. The chromosome profiles for all chromosomes can also be seen in Appendix 5.

4.5: Production of a high resolution array from PCR products.

A high resolution array was constructed sampling a 4.5Mb region of chromosome 22, 15398721 – 19982021bp along the q arm at a resolution of one approximately 500bp PCR product every 10Kb. In addition, overlapping 500bp products were designed to cover the region 16495000-16695000bp along the q arm of chromosome 22. The design of primers is described in section 2.2.1.

The first round of amplification was performed using clone DNA as template. Products from each PCR plate were analysed by gel electrophoresis using a 2.5% agarose gel. A successful PCR was indicated by a single band with a product size of approximately 500bp (Figure 4.8)

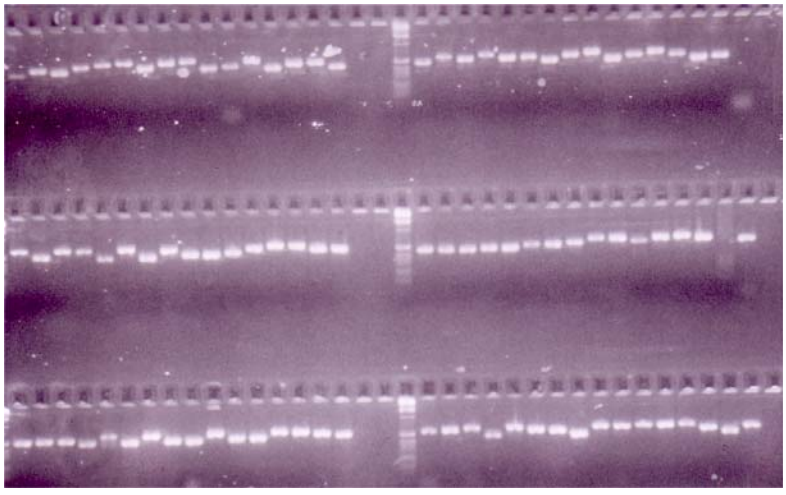


Figure 4.8: PCR products obtained from the amplification of primers STSG 495474 – STSG 495569 in a 96 well format as detailed in Appendix 2b.

A strong, clean amplification product was observed for 599 of the 714 primer pairs tested. A further 16 primer pairs produced a weak product whilst 99 produced no product. Primers producing a weak product or no product were re-amplified using genomic DNA as a template. Of the 115 primer pairs re-amplified, 68 gave a strong product, 16 gave a weak product, 30 produced no product and 1 generated a double band, suggesting amplification of more than one region of the genome, although this was not confirmed.

Each PCR product was spotted on the array in quadruplicate. As with the tile path arrays self:self and male:female hybridisations were used for array verification as shown in Figure 4.9 and Figure 4.10.

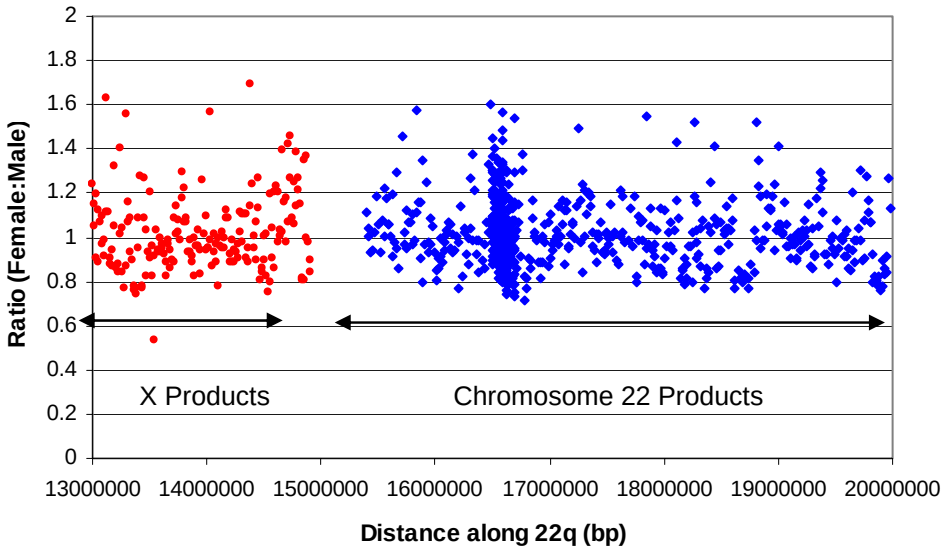


Figure 4.9: A G1:G1 hybridisation on the high resolution PCR product array.

For the G1:G1 hybridisation, 50 of the 714 Chromosome 22 PCR products failed the analysis criteria due to the criteria given in 2.5.2. Of the remaining 664 clones, the mean ratio reported was 1.00 and the standard deviation was 0.15.

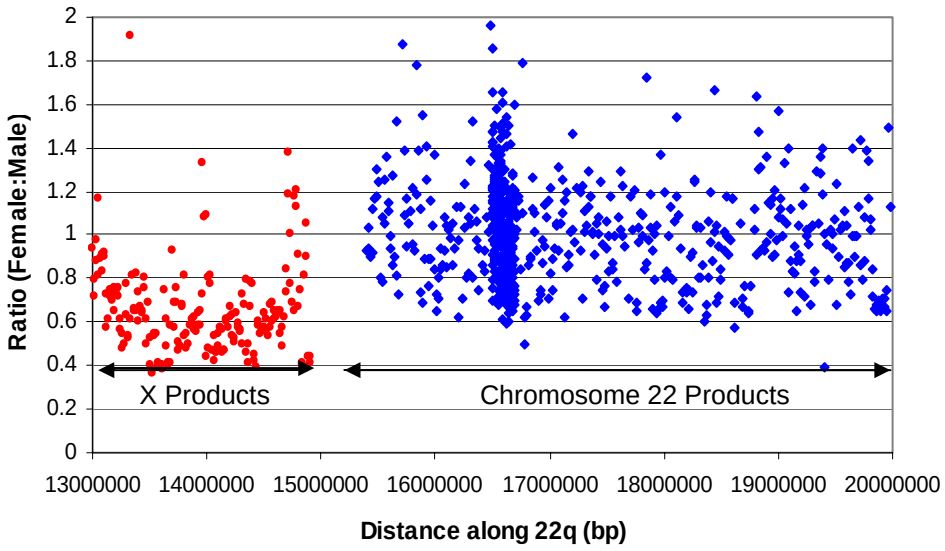


Figure 4.10: A male:female hybridisation on the high resolution PCR product array.

Analysis of the male:female hybridisation revealed forty of the 714 chromosome 22 PCR products on the array failed the analysis criteria described in 2.5.2. The standard deviation of the remaining loci was 0.23. The chromosome X PCR products were analysed, and gave an average male: female ratio of 0.67. The average standard deviation of the chromosome X loci on the array was 0.26.

4.6: Detection of chromosome 22 copy number changes on clone arrays

4.6.1: Detection of chromosome 22 copy number change on the 1 Mb tile path array.

The reporting of a copy number change by a clone, in response to a chromosome 22 sequence in the hybridisation mix, was assessed by the addition of flow sorted chromosome 22 DNA to a self:self hybridisation utilising genomic DNA. This is described in section 2.5.2.3. Results for the 1Mb resolution genomic array are shown in Figure 4.11.

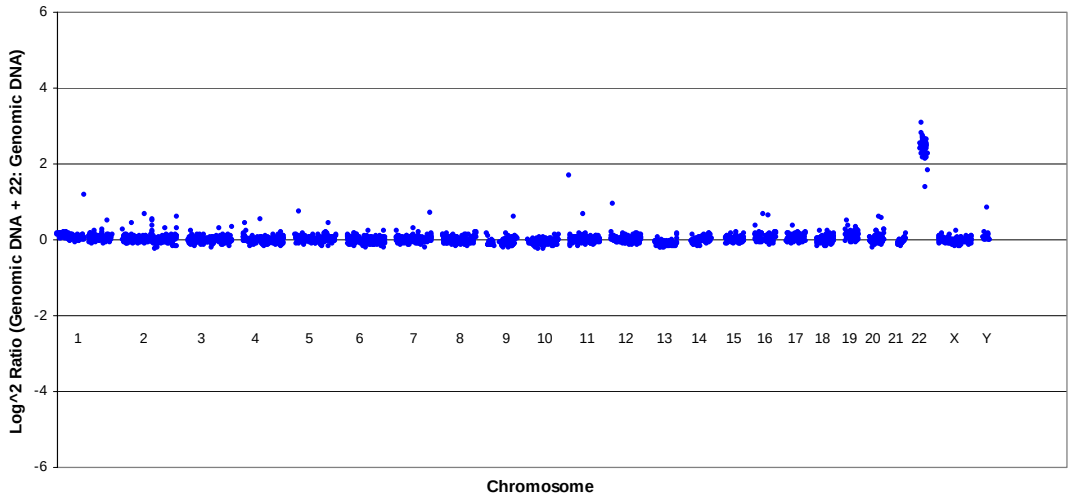


Figure 4.11: A genomic DNA + Chr 22:genomic DNA hybridisation on the 1 Mb array.

All chromosome 22 clones on the 1Mb array responded to the addition of five copies of chromosome 22 into the hybridisation mix by showing a copy number gain. However some clones on other chromosomes also report a copy number gain.

Examining the chromosome 22 clones in detail, the average ratio reported was 5.57 with a standard deviation of 0.94 (see Figure 4.12).

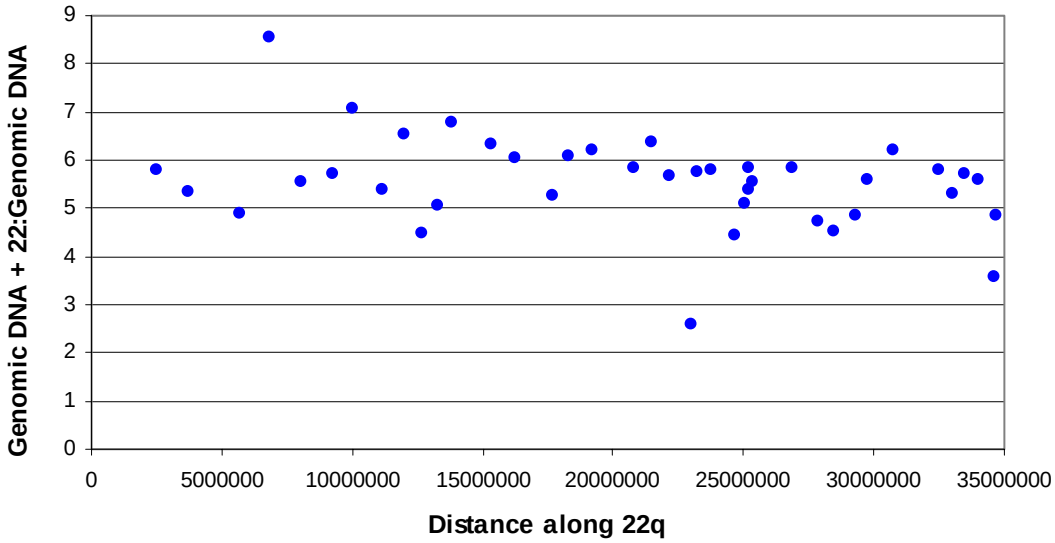


Figure 4.12: Response of the chromosome 22 clones to a chromosome 22 add-in experiment.

One clone, RP11-50L23 located 6.8 Mb along the q arm of chromosome 22, reported a particularly high ratio of 8.41.

The chromosome 22 clone reporting the lowest ratio (2.66) was CTA-150C2. However this ratio is still significantly above all the ratios reported on clones from other chromosomes, except the chromosome 11 clone CTC-908H22 (discussed below) and so this clone still reports a change in chromosome 22 copy number.

Several clones in the rest of the genome reported high ratios indicating that they too report a response to the increased amount of chromosome 22 in the hybridisation mix. Clones that reported a ratio above the 99% confidence interval for the mean ratio of modal clones are detailed in Table 4.2.

The clone showing the largest response to the chromosome 22 DNA is a clone located at the 11p telomere (CTC-908H22). This is illustrated in Figure 4.13.

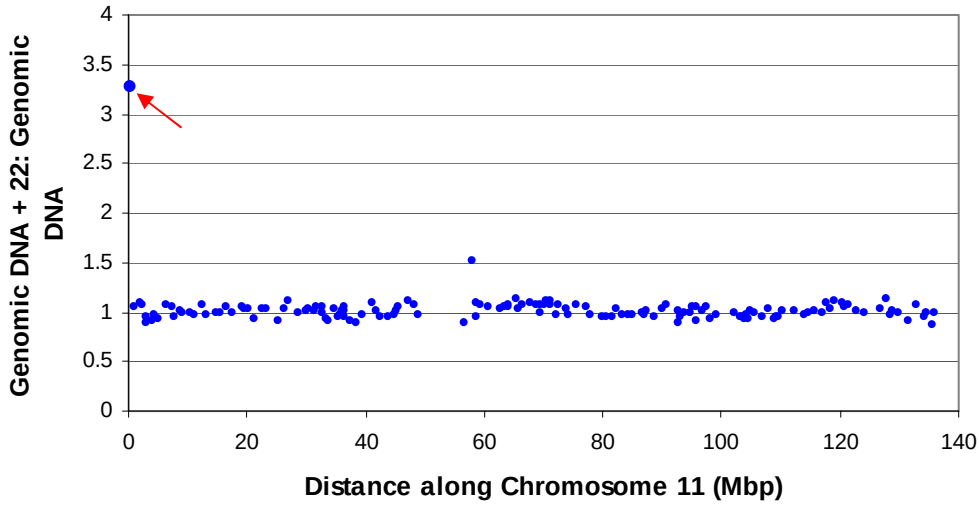


Figure 4.13: Hybridisation ratios reported by chromosome 11 clones after a genomic DNA + 22: genomic DNA hybridisation

4.6.2: Detection of chromosome copy number changes on the 22 tile path array.

A similar experiment was performed on the 22 tile path arrays to test the responsiveness of array loci to chromosome 22 copy number change. Arrays were performed with an estimated 1 additional, 2 additional and four additional copies of

chromosome 22 in the hybridisation mix. A G1 self:self and a G2:G1 hybridisation was also performed within the same batch of arrays. Arrays were normalised against the chromosome X clones, which should report no copy number change. The mean copy number change for the 22 clone was calculated.

The response of clones to copy number changes are shown in Figure 4.14 where the ratio is plotted against the approximate number of extra copies of chromosome 22 added to the hybridisation mix. Hyper-responsive clones plotted on Figure 4.14 are p87O8, pac699j1, dJ293L6, and cN69E4. Clones under reporting copy number change are, b444p24, cN61D6, cN20A6 and cN21F1. Four clones reporting a correct response were also included for comparison. These clones, chosen at random were not located within the first 9Mb of the q arm, known to contain a considerable segmental duplication. These clones are dJ127L4, bK282F2, fF4G12 and bK126B4.

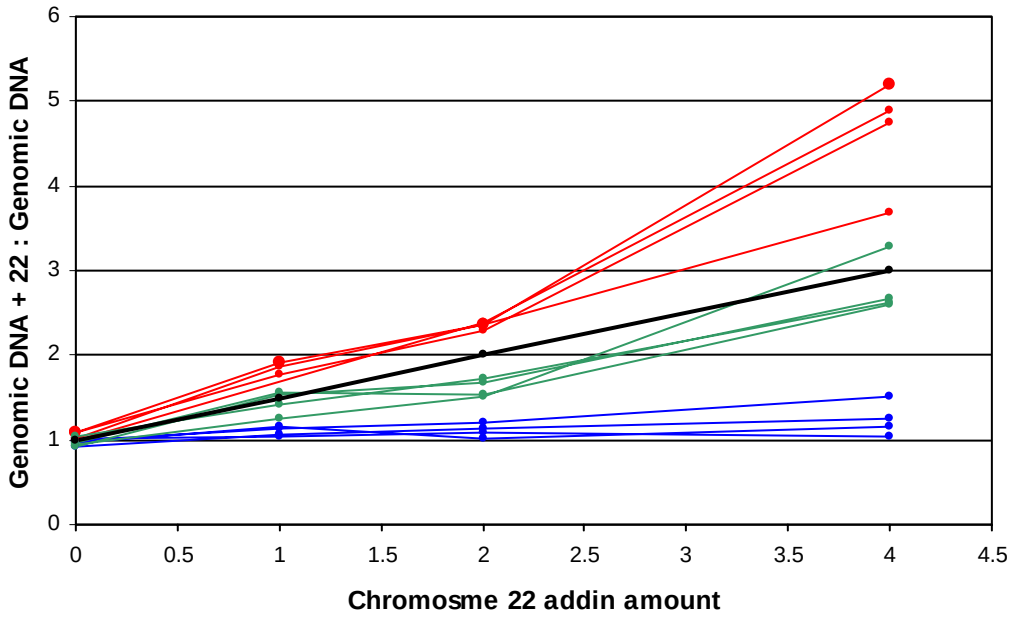


Figure 4.14: Ratios reported when different amounts of chromosome 22 are added into the hybridisation mix. Red: Clones that are hyper-responsive to addition of chromosome 22. Blue: Clones that are not responsive to the addition of chromosome 22. Green: Clones that report a normal response to chromosome 22. Black: Ideal copy number change reported.

Of the 470 clones on the chromosome 22 array, only twenty clones over or under-reported the response to chromosome 22 DNA within the hybridisation mix. This

indicates that 96% of the chromosome 22 clones report copy number changes accurately.

4.7: Discussion

4.7.1: Control hybridisations performed on the clone arrays.

A control self:self hybridisation was performed on the clone DNA arrays and average expected 1:1 ratios were reported by both the 1Mb and 22 tile path arrays. The standard deviations reported by the 22q tile path array and the 1Mb array were comparable, showing the reproducibility of the method when constructing an array from large insert clone DNA using DOP PCR.

Clones from chromosome X were included on each type of array constructed. These clones provide an intrinsic control to allow simple verification of copy number changes using male:female hybridisations. A male:female hybridisation should report a ratio of 1:1 on loci derived from autosome sequence but a ratio of 0.5:1 on loci representing chromosome X due to the X chromosome copy number difference between males and females. A ratio 0.5:1 was not reported on by any of the X loci represented on the arrays. The lowest ratio reported was 0.58:1 on the 22q tile path array and the highest ratio reported was 0.75:1 on the 1Mb array. This could be due the representation of different X clones on the two arrays. There are 46 more X clones on the 1Mb resolution array than there are on the chromosome 22q array. This under reporting of the copy number difference on chromosome X has been reported previously (Pinkel, Segraves et al. 1998; Fiegler, Gribble et al. 2003). Possible reasons for this underestimate could be an under-representation of chromosome X sequences in Cot 1 DNA leading to incomplete suppression of repeats on the chromosome X loci and cross-hybridisation of other regions of the genome with high sequence homology. Chromosome X has been identified as being paticulary rich in LINE repeats (IHGSC 2001). However, assessment of segmental duplications throughout the whole genome (Bailey, Gu et al. 2002) shows that chromosome X is relatively sparse in interchromsomal repeats.

Another possibility for the underestimate of the copy number change reported by the chromosome X clones is that, unlike autosomes, the two copies of chromosome X in the female DNA are not the same but differ epigenetically. In females one X chromosome is epigenetically silenced, rendering it transcriptionally inactive. This is to ensure that there is the same dosage of genes encoded on chromosome X in males and females (for review see (Avner and Heard 2001)). This epigenetic silencing involves the tight condensation of the chromatin into an inactive barr body. Inactivation makes the DNA within the inactive X chromosome very inaccessible which may affect DNA labelling such that Cy dye is not incorporated into the inactive chromosome X with the same efficiency as it is into active chromosomes. This means that after a male:female hybridisation a full 1:2 ratio would not be reported on the X clones.

4.7.2: Verification of the 1 Mb resolution and chromosome 22 Tile path arrays.

Further verification on the chromosome 22 and 1Mb arrays were performed with a series of experiments utilising different amounts of additional chromosome 22 DNA in the hybridisation mix. On the 1Mb array, one clone from chromosome 22 (RP11-50L23) can be seen to be hyper-responsive to the chromosome 22 DNA. This clone is located 6.8 Mb along the q arm of chromosome 22 within the locus encoding the immunoglobin light chain λ region. During lymphoblastoid development this region undergoes rearrangement and deletion. The control cell line, from which the genomic DNA was extracted (HRC 575), has been shown to have a deletion in this region (see 4.3.3 and 7.3). It is therefore likely that the hyper-sensitivity of this clone is due to the presence of only one copy of chromosome 22 in the cell line the genomic DNA was extracted from. Calculations reported in section 4.6.2 assumed two copies of chromosome 22 in the genomic DNA.

Chromosome add-in experiments on the 22q tile path array showed over 96% of loci reported the correct response to increased dosage of chromosome 22. The linear response reported by representative clones of this majority group (see Figure 4.14) confirm that the clones responded appropriately to the extra copies of chromosome 22 added.

Several clones did not report the correct response to additional copies of chromosome 22. On the 1Mb array the chromosome 22 clones adjacent to the telomere reported a depressed response to the addition of chromosome 22 to the hybridisation mix. This is unsurprising as the telomeric region contains a large amount of genome repeats (see section 1.3.1). Therefore it is likely that these clones will cross hybridise with other regions of the genome. Table 4.2 details clones not mapping to chromosome 22 represented on the 1Mb array that responded to the additional copies of chromosome 22 in the hybridisation mix.

Table 4.2: Clone not mapped to chromosome 22 that responded to extra chromosome 22 in the hybridisation mix

Clone	Chr.	position	Ratio	End sequence	Segmental Duplications
RP11-114F20	3	197298109	1.25	Match	None
RP5-1107C24	20	60245780.5	1.26	Match	None
CTD-3113P16	19	244656.5	1.28	multiple, none on 22	19,21,4,5,8,6,22
RP11-260J21	2	236147059	1.30	Match	None
CTD-2547N9	19	9002070.5	1.31	Match	None
RP11-278G12	2	38037637	1.33	Match	None
RP11-260A9	17	27226356.5	1.36	Match	None
RP11-565I3	4	7435638.5	1.38	Maps to Chr 14	None for 4
RP11-1E1	4	78240481	1.38	Match	18,9
RP11-24O13	2	130447729	1.38	multiple, On 22	None
RP11-276J4	1	223457232.5	1.44	multiple, none on 22	9,13,10,1,5
RP11-30F17	19	6552559	1.44	No end sequence	19
RP11-205K6	9	120945187	1.47	Match	9
RP11-100N3	11	58074075	1.52	Match	None
RP4-724E16	20	51861416.5	1.53	Match	None
GS1-172I13	2	241706787	1.53	No end sequence	1,2,21
RP11-209H16	2	129390774.5	1.54	multiple, On 22	2
RP11-71B7	2	93952867	1.57	multiple, none on 22	None
RP11-122O1	20	37014297	1.59	No end sequence	4,1,7,11,14,3,12,9,X
RP1-29O12	5	14786570.5	1.59	No end sequence	5
RP11-408D2	16	35065398.5	1.61	multiple, On 22	16,6
RP11-165M2	16	55986775	1.66	Match	16
RP11-208G20	7	150257256.5	1.70	multiple, none on 22	7
RP11-434F12	24	18960023	1.78	multiple, none on 22	Y,12,3,UL
RP3-467F14	12	6148320.5	1.97	multiple, On 22	15,4
CTC-908H22	11	175000	3.26	multiple, On 22	1,4,11

End sequence match= sequence from end sequencing of the chromosome matched their location in Ensembl. UL= Unlocated, contig not mapped to any chromosome

The 26 1Mb array clones that cross hybridised with chromosome 22 were analysed in two different ways to see if the cross hybridisation could be explained. All the clones

in the 1Mb clone set had been end sequenced and compared to the genome sequence to verify position, and locate other regions of similarity. The study of this database (http://intweb.sanger.ac.uk/cgi-bin/humace/1mbsetends.cgi) showed that five clones had end sequences that contained a significant amount of homology to chromosome 22 sequence. This could either indicate a mixed well when the clone was picked or a segmental duplication within the DNA that was end sequenced. A mixed well would lead to representation of more than one region of the genome on the array such that the reporting of copy number changes at this locus would be inaccurate. The presence of segmental duplications within the clone results in cross hybridisation of other regions of the genome. The end sequences of one clone, RP11-565I3, mapped to chromosome 14, not chromosome 4 as previously thought.

Clones were also analysed using the segmental duplication track on the UCSC genome sequencing database (http://humanparalogy.gene.cwru.edu). The segmental duplications were identified as described by Bailey et al (Bailey, Gu et al. 2002). This analysis revealed a further clone with homology to chromosome 22. However it should be noted that not all clones showing a homology to chromosome 22 by their end sequence are detected on this database. This confirms the incomplete status of this database and the human genome sequence at the time of analysis (Bailey, Gu et al. 2002), (IHGSC 2001).

A further 11 clones had end sequences that mapped to more than one chromosome, or segmental duplications involving chromosomes other than 22. Although this does not explain the cross hybridisation with sequences from chromosome 22, it does indicate that these clones contain repetitive DNA. Inefficient blocking by Cot 1 DNA, or the presence of chromosome 22 segmental duplications that were not identified by Bailey *et al* (Bailey, Yavor et al. 2002) may explain the cross hybridisation with chromosome 22.

The remaining eight clones had end sequences that match their positions assigned on the 1Mb profiles (Appendix 4) and no duplications within chromosome 22. However, most of these clones have ratios toward the lower end of those identified in Table 3.2. The statistical analysis used to identify clones with a significant response to the additional chromosome 22 DNA uses the 99% confidence level of modal

values. On a purely statistical basis, on an array containing 3,500 clones, 35 clones would be expected to report a ratio over the 1.24 cut-off identified.

To be classified as an atypical reporting clone the clone had to report a copy number change with a standard deviation outside the 99% confidence intervals in two of the three arrays. These clones are summarised in Table 4.3.

Table 4.3: Clones not responding with the correct copy number change when chromosome add-in experiments were performed on the tiling path arrays.

Clone	Accession No.	Position	No. of arrays	Comments	Segmental duplication*
Clones under reporting copy no. changes					
cN64E9	AP000526	114958	2	centromeric	1,2,9,10,14,16,22,UL
p87o8	AC007064	1248002	3	Seg dup	None
pac699j1	AC008103	2822641	3	Seg dup	1,4,5,6,13,20,22,UL
56c	AC000080	3878158	2	Seg dup	None
2H8	D87003	6336208	3	Seg dup	1,2,4,15,16,22,UL
bA541J16	AL080241	12558437	2		None
bA329J7	AL118497	12578094	2		None
cE78G1	Z70288	17685472	3		None
dJ293L6	AL049749	20707056	3		None
dJ591N18	AL031594	24555326	2		None
dJ408N23	Z98048	24844579	2		None
cN69F4	Z72006	32361681	3	telomeric	22
n1g3	AC002055	34687355	3	telomeric	2,22
Clones over reporting copy no. changes					
b444p24	AC007663	4165628	2	seg dup	22, UL
cN61D6	D87012	5997695	3	VJ region	None
cN75C12	D87017	6963826	2		22
cN20A6	Z69713	17796756	2		None
bK299D3	Z84468	32481166	2	-	None
cN21F1	Z94162	33107523	3	-	None
66C4	AC000050	34627340	2	telomeric	None

* as reported by CWRU browser: Segmental duplication database on the UCSC website. Seg dup = clone contains a segmental duplication, UL= Unlocated, Contig not mapped to any chromosome

Several of the clones under-reporting copy number changes contained segmental duplications. Regions with homology on other chromosomes will cross hybridise with DNA from other chromosomes. This cross hybridisation will depress the ratios reported. For example, if all the sequence within a clone is duplicated on another

chromosome there will be four copies present within the genomic DNA. Addition of an extra copy of chromosome 22 into the hybridisation mix will result in an 5:4 ratio as compared to the 3:2 ratio if the clone contained unique sequence. In this way copy number changes will be underestimated for regions of segmental duplication involving other chromosomes. Other clones under reporting copy number change were located adjacent to the centromere or telomere and contain an abnormal amount of common repeat elements. The incomplete suppression of common repeat elements by Cot 1 may lead to the under or over reporting of chromosome 22 DNA copy number.

One clone that was hypersensitive to the chromosome 22 DNA was the clone cN61D6 (Accession no. D87012). This is located in the region encoding the immunoglobulin light chain λ. This has been shown to be deleted in some lymphoblastoid cell lines (4.3.3 and 7.3). The increased ratio reported in response to the additional copies of chromosome 22 may therefore be due to the single copy of chromosome 22 for this region within the genomic DNA used for the hybridisation.

This is an intrinsic problem with an array containing DNA representing an entire chromosome. More detail about regions of the array that contain a significant amount of segmental duplication are detailed in section 7.4.1. Clones with a high quantity of common repeats may not be fully blocked by the inclusion of Cot 1 in the hybridisation mix. Again this would result in cross hybridisation with other regions of the genome, under reporting the response to chromosome 22 DNA. This should be taken into account when reporting data from these clones.

4.7.3: Control hybridisations on the 500bp PCR product array

The standard deviation reported by a self:self hybridisation on the high resolution PCR product array was 2.5 times the standard deviation of the 22 tile path array and reflects the excessive noise shown around the 1:1 ratio. The standard deviation reported by a male:female hybridisation was also 2.5 times the standard deviation observed for the chromosome 22 array. The ratio reported by the chromosome X clones on the array did not represent the expected ratio for a full single copy number loss. The average standard deviation reported by the chromosome X clones was also

much greater than the standard deviation of the chromosome X clones on the chromosome 22 tiling path array.

It was noted that the intensities of the Cy3 and Cy5 signals from the PCR product arrays were considerably reduced compared to the large insert arrays. On average the signal intensities were 100 times less than those reported for the arrays spotted from DOP amplified clone products. The signal:background ratio of each spot on the array was therefore higher than for the clone-based arrays reducing the sensitivity and reproducibility of the PCR-based arrays The reduced intensities could be due to the PCR products spotted onto the array being smaller than those used for the clone arrays, or because they undergo one less round of amplification before spotting onto the array so that the final concentration of DNA in the spotting buffer is decreased. One way to increase the DNA concentration may be to include an extra round PCR amplification, prior to spotting of the products onto the array.

The self:self hybridisation on the PCR product array also revealed regions which show reduced ratios. This can be seen on Figure 4.9 located 18.62-18.76 and 19.82-19.90 Mb along 22q. This could be due to a labelling bias, where Cy3 and Cy5 are incorporated with different efficiency into GC or AT rich DNA. To test this hypothesis, the correlation between the GC content of the PCR product and the ratio reported by the self: self hybridisation, for a random selection of loci, was plotted. As seen on Figure 4.15 there is no correlation between GC content of sequence and ratio reported, so it is unlikely a labelling bias is responsible for the high standard deviations of the ratios observed.

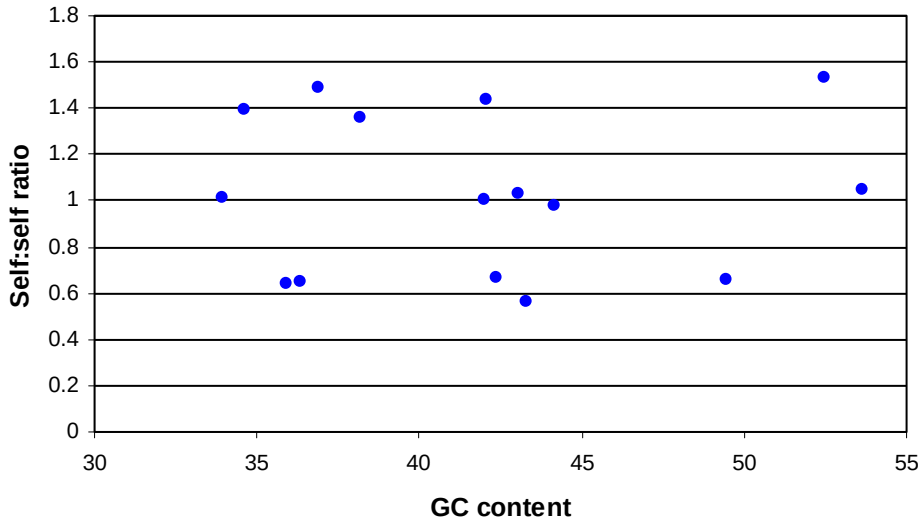


Figure 4.15: Correlation between GC content of PCR product and the ratio reported by a self:self hybridisation.

The average male:female hybridisation ratios reported on the X loci was 0.67. Again the standard deviation of ratios were larger than those seen for the clone arrays, but the average ratio reported for the X clones was comparable.

4.7.4: Summary

In summary, the clone array verification experiments showed that the reporting of copy number change by the constructed 22q tile path array was accurate. The chromosome add-in experiments showed that a vast majority of the clones represented on the array reported the expected response to additional copies of chromosome 22. It was concluded that the arrays were suitable for detecting the small copy number changes necessary for the assay of replication timing.

The verification experiments performed on the PCR product array showed wide variation in the ratios reported by control experiments indicating that these arrays would be less sensitive to replication timing differences compared to clone based arrays.

5. Results 3
Using Genomic Microarrays to assess Replication Timing in a Human Cell line and correlation with sequence features

5.1: Introduction

DNA microarrays have been used to assess replication timing in the yeast *Saccharomyces cerevisiae* (Raghuraman, Winzeler et al. 2001) and in *Drosophila melanogaster* (Schubeler, Scalzo et al. 2002). Assessment of replication involves timing measuring the difference in copy number in an S phase population, between sequences that replicate early in S phase (and are therefore present in a high copy number) and those that replicate late in S phase. This Chapter describes how genomic arrays have been used for the first time to assess human replication timing. In these experiments differentially labelled S and G1 phase DNA is co-hybridised to the arrays described below. Loci reporting a ratio close to 2:1 are early replicating; conversely loci reporting a ratio close to 1:1 are late replicating.

Initially the hybridisations were carried out on an array representing the entire genome with clones at a 1Mb resolution. Section 5.2 describes the replication timing profiles obtained for each chromosome at this resolution. Correlations between the replication timing and sequence features of the genome at this resolution are also shown.

The S:G1 hybridisations were also performed on arrays constructed at tile path resolution. A large, a medium and a small chromosome are examined at this resolution, using tiling path arrays for chromosomes 1, 6 and 22. Analysis of replication timing at this higher resolution enables more accurate mapping of zones in which a transition in replication time occurs. The tile path arrays and the correlations with sequence features at this level are described in section 5.3.

Finally, the replication timing of a small region of chromosome 22 is assessed at a very high resolution. PCR products were used to cover a 4.5Mb region of chromosome 22 at 10Kb resolution. A region of 200Kb was analysed at an even

higher resolution using overlapping 500bp PCR products. This array is described in 5.4.

The replication timing method was substantiated by comparison with alternative ways of assessment of replication timing. The replication timing of 11q was compared with published data for this region and the results are described in section 5.5.1. Assessment of the replication timing of clones within chromosome 22, described by the array as early, mid or late replicating was performed by real-time PCR and the results correlated with the array data. This is reported in section 5.5.2.

During DNA replication, copy number increases from one to two at each locus. This should occur early in S phase for early replicating loci, but in later fractions of S phase for later replicating loci. This can be detected by the hybridisation of fractions of S phase to the array. S phase nuclei were sorted into four fractions and co-hybridised with G1 phase DNA on the 22 tile path arrays. This allowed comparison of the array analysis with assay of replication timing using quantitative PCR, in which S phase is conventionally divided into four fractions and the fraction in which a predetermined loci replicates is resolved. This is described in section 5.7.

5.2: Assessment of Replication Timing on the 1Mb array

5.2.1: Obtaining the Average Replication Timing of Individual Chromosomes.

The S:G1 ratios and standard deviation of each locus in 4 replicate 1Mb genome wide arrays were calculated and the average co-efficient of variation of all loci in the four replicates was found to be 5.5%.

The ratios for all the loci contained within the same chromosome were averaged and are shown in Figure 5.1 and Table 5.1.

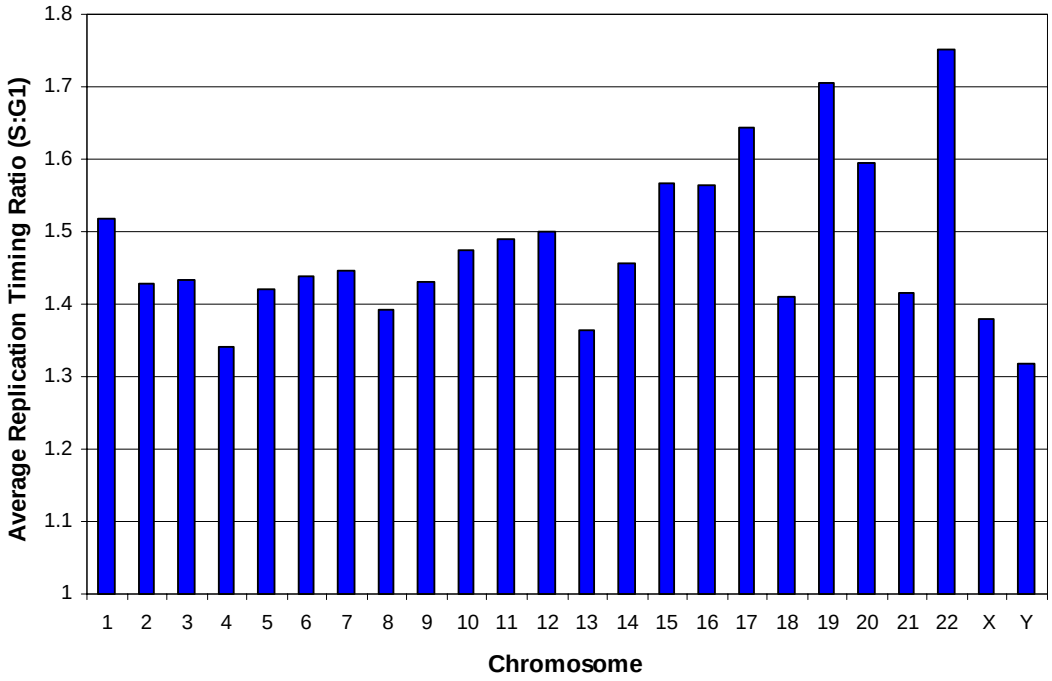


Figure 5.1: The average replication times of all 24 chromosomes.

Table 5.1: The average replication times of all 24 chromosomes (Early-late)

Chromosome	**Mean Replication Timing Ratio**
22	1.75
19	1.72
17	1.64
20	1.60
15	1.57
16	1.56
1	1.52
12	1.50
11	1.49
10	1.49
14	1.46
7	1.45
6	1.44
9	1.44
3	1.43
2	1.43
5	1.42
18	1.42
21	1.42
8	1.39
X	1.38
13	1.36
4	1.34
Y	1.32

The average replication timing for each chromosome was used to normalise the individual chromosome tiling path arrays.

5.2.2: Correlating Chromosomal Replication Timing with Sequence Features of the Genome.

The chromosome wide individual sequence features were plotted against the replication timing of each chromosome as shown in Figure 5.2.

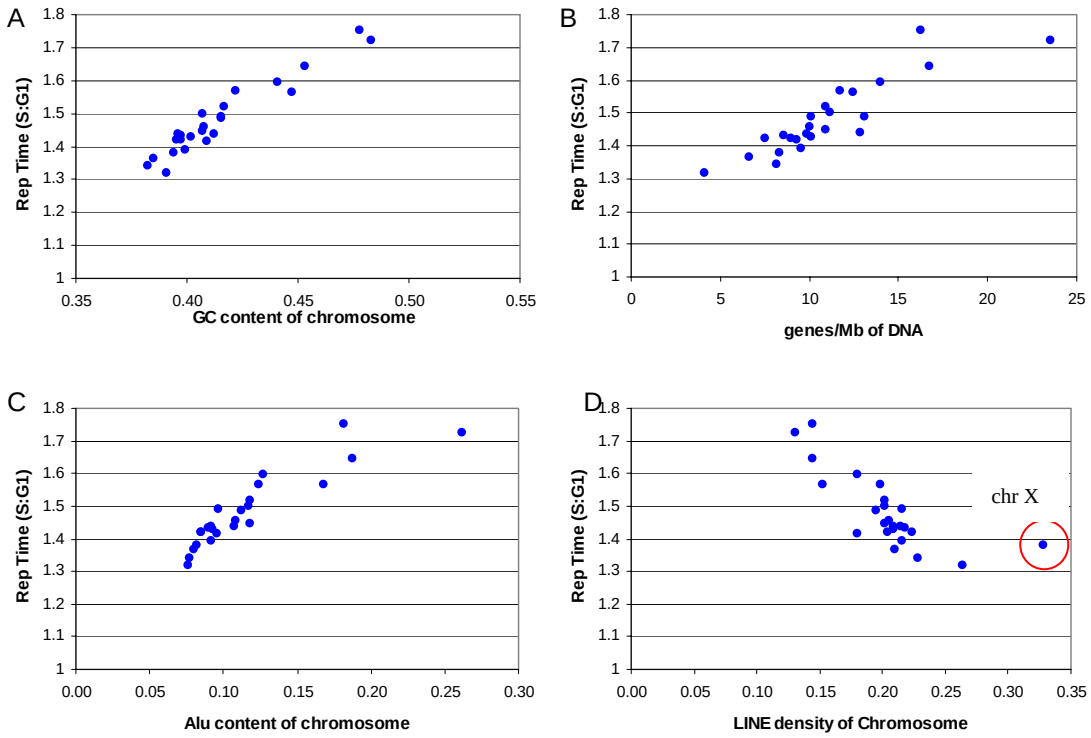


Figure 5.2: Correlation between Replication Timing and Sequence Features of the Genome. A: Correlation with GC content. B: Correlation with Gene Density. C: Correlation with *Alu* repeat content. D: Correlation with LINE density.

Linear regression was then performed on these data to assess the correlation between replication timing and sequence features. The linear regression statistics are shown in Table 5.2.

Table 5.2: Linear regression performed between replication timing and genome statistics.

Genome Feature	**Regression Coefficient (x)**	**Intercept (y)**	**Correlation Coefficient (r)**
GC Content	0.04	0.17	0.96
Gene Density	0.02	1.20	0.89
***Alu* Repeat Content**	0.02	1.21	0.9
LINE Repeat Content	-0.02	1.91	0.72

Significant correlations were found for all sequence features although the best correlation with the S:G1 ratio was found with GC content. A very strong positive correlation was seen with *Alu* repeat content and a strong positive correlation was observed with gene density. A strong negative correlation was observed between

LINE repeat content and replication timing; however the correlation coefficient is not as strong as those observed with other features. Definitions of the strength of correlations is found in Appendix 13.

5.2.3: Assessing Replication Timing at a 1 Mb resolution.

Data from the same array experiments could also be used to assess replication timing at a 1Mb resolution. Each clone on the array was positioned according to the NCBI 31 assembly of the human genome as represented in Ensembl. The replication timing ratio was then plotted against position on the chromosome to produce a replication timing profile for each chromosome (Appendix 6). Example replication timing profiles for two chromosomes are shown in Figure 5.3

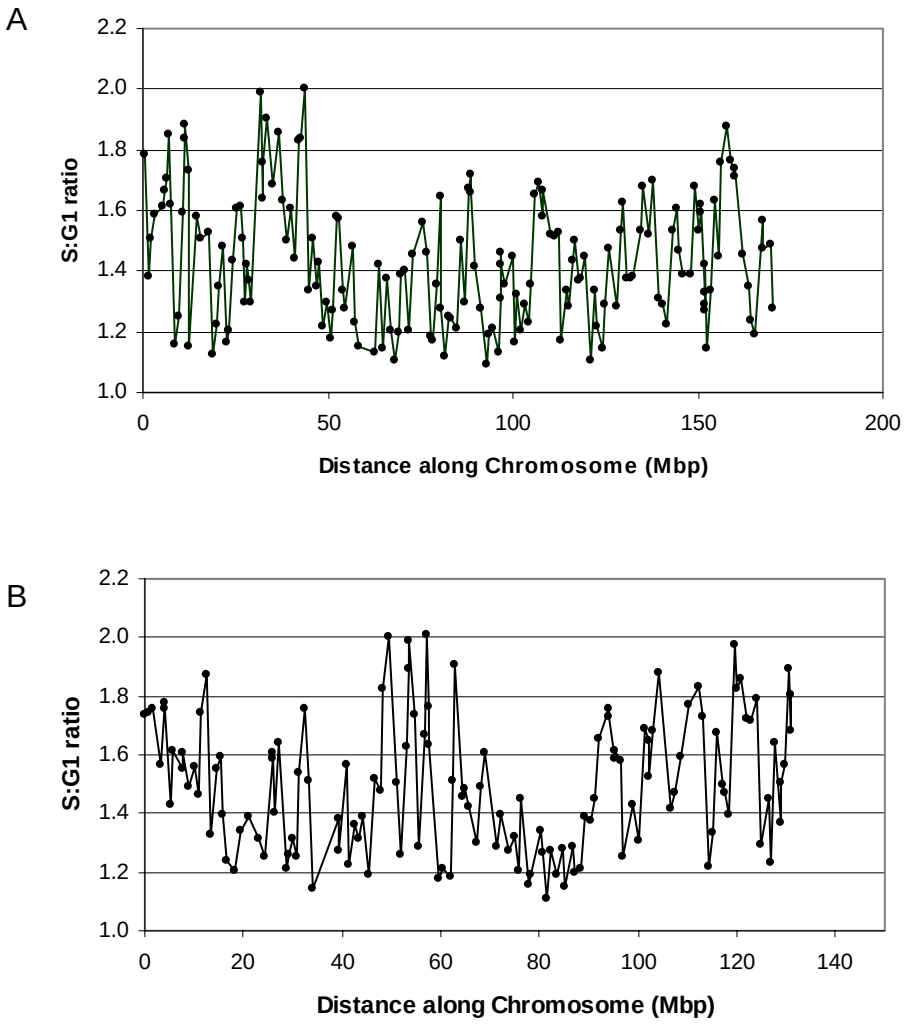


Figure 5.3: Replication timing profiles of; A: chromosomes 6 and B: chromosome 12

Correlations between replication timing and sequence features were also performed with the whole genome sampled at a 1Mb resolution. The sequence features used for the correlation were for just the clone sequence represented on the array, not the whole 1Mb window represented by that clone. The correlations are shown in Figure 5.4.

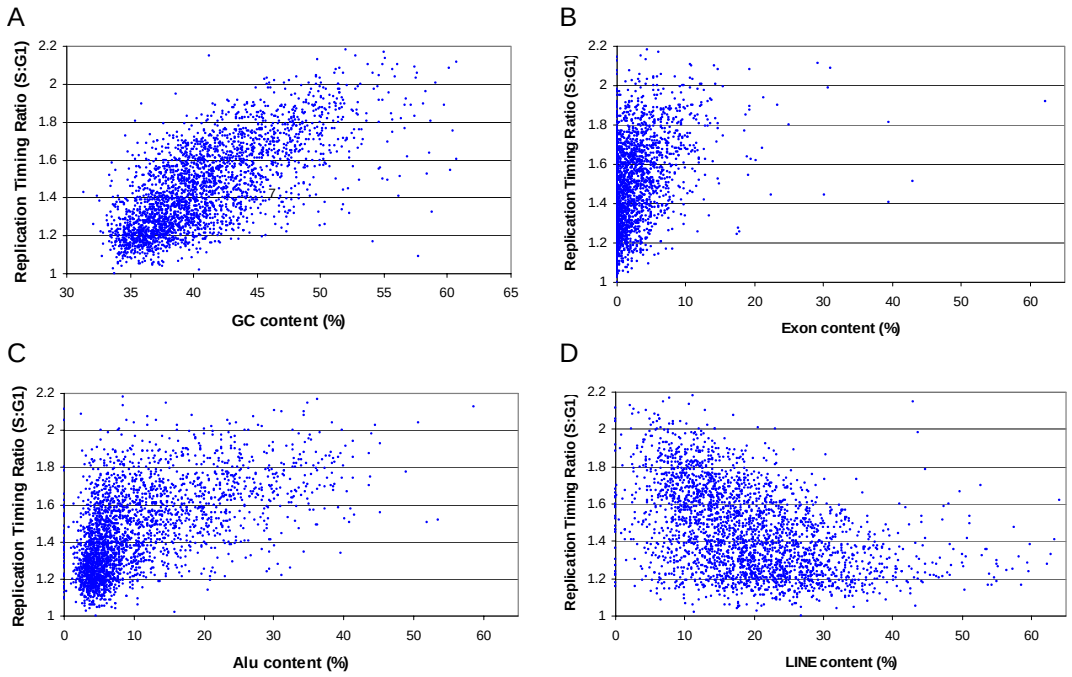


Figure 5.4: Correlation between Replication Timing and Sequence Features of the Genome. A: Correlation with GC content. B: Correlation with Exon Density. C: Correlation with *Alu* repeat content. D: Correlation with LINE density.

Linear regression was then performed on this data to assess the correlation between replication timing and sequence features. The linear regression statistics are shown in Table 5.3.

Table 5.3: Linear regression performed between replication timing and genome statistics at a 1 Mb resolution.

Genome Feature	**Regression Coefficient (x)**	**Intercept (y)**	**Correlation Coefficient (r)**
GC Content	0.032	0.12	0.70
Gene Density	0.002	1.37	0.35
Exon Density	0.023	1.40	0.42
***Alu* Repeat Content**	0.014	1.30	0.56
LINE Repeat Content	-0.008	1.62	0.40

The best correlation was again seen with GC content. Positive correlations were also observed with *Alu* repeat content, exon density and gene density. A negative correlation was observed with LINE repeat density.

5.3: Assessment of Replication Timing at Tile path Resolution

Replication timing was assayed using tiling path arrays for chromosomes 1, 6 and 22. These arrays were constructed using DNA extracted from sequencing clones. This ensured the data obtained provided complete coverage for chromosome 22q, chromosome 6 and chromosome 1.

5.3.1: The Replication Timing of Chromosome 22.

Replication Timing was assayed on the chromosome 22q tiling path array. The S:G1 hybridisation was carried out on four separate arrays to ensure reproducibility of the data. The average co-efficient of variation between the four replicates was found to be 5.5%. The replication timing of chromosome 22q is shown in Figure 5.5.

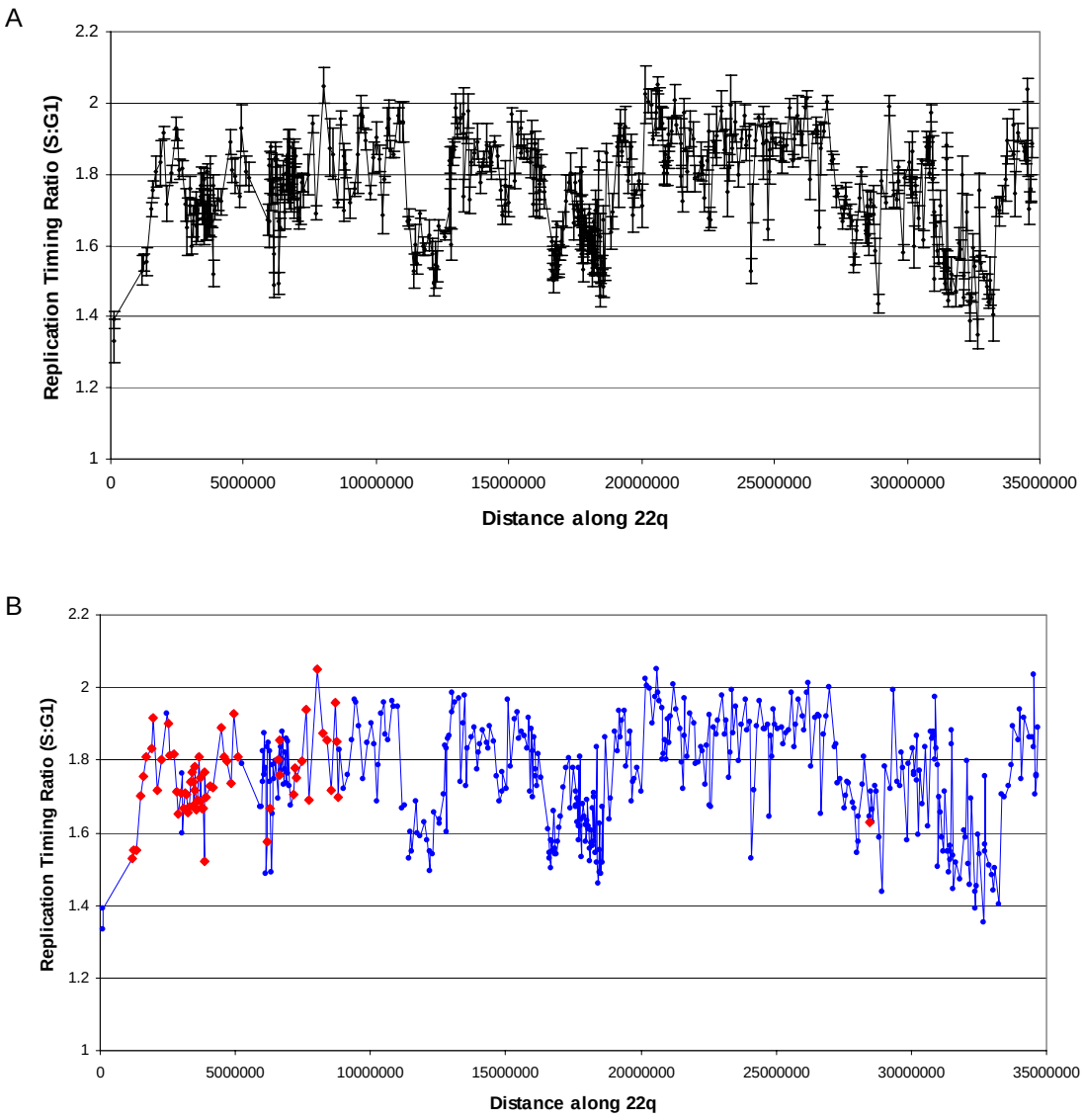


Figure 5.5: Replication Timing profile of Chromosome 22. A: Replication Timing profile of chromosome 22 showing the standard error of each loci (n=4) on the Y error bars. B: Replication Timing profile of chromosome 22 with clones containing significant amounts of segmental duplication highlighted in red (Buckley, Mantripragada et al. 2002).

Chromosome 22 contains considerable amounts of segmental duplication at 22q11. This has a significant impact on the ability of clones on the array to accurately report locus-specific copy number changes as sequences with homology to these regions will cross hybridise with the DNA on the array. These clones are highlighted on Figure 5.5B. As a result the replication timing of the arrays reported by these regions will be inaccurate and a composite of the replication timing of all the regions sharing the

homology. As a result these clones are not included in further analyses assessing the correlation of replication timing with other sequence features. Replication timing was plotted against sequence features of the genome such as gene density, GC content and common repeat content. These correlations are shown in Figure 5.6.

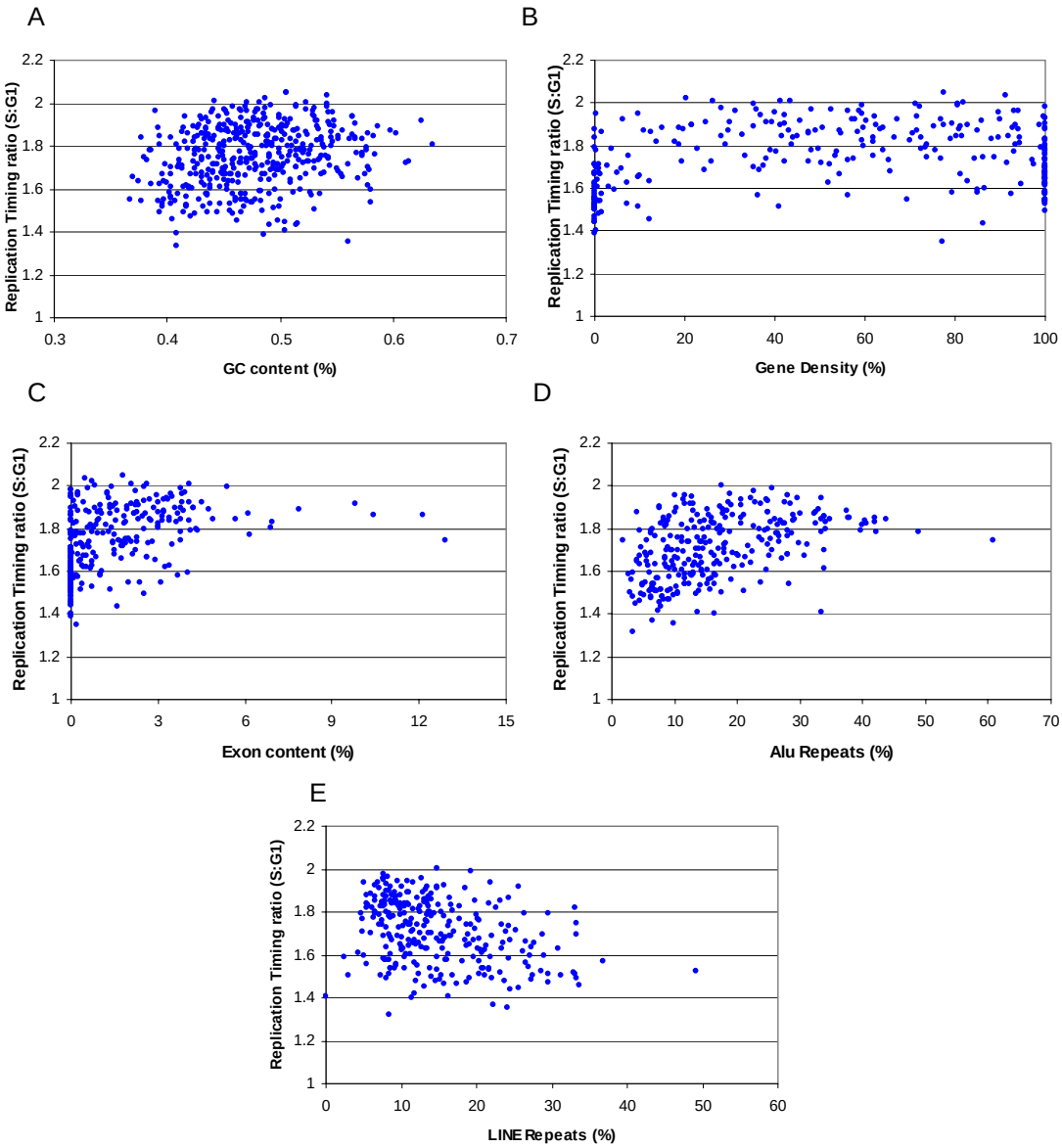


Figure 5.6: Correlation between replication timing and other genome features on the 22 tile path array. A: Correlation with GC content. B: Correlation with gene density. C: Correlation with exon density. D: Correlation with *Alu* repeat content. E: Correlation with LINE repeat content.

Linear regression was then performed on this data to assess the correlation between replication timing and sequence features. The linear regression statistics are shown in Table 5.4.

Table 5.4: Linear regression performed between replication timing and genome statistics at a 78Kb resolution on the 22 tile path array.

Genome Feature	**Regression Coefficient (x)**	**Intercept (y)**	**Correlation Coefficient (r)**
GC Content	0.63	1.45	0.22
Gene Density	0.001	1.71	0.19
Exon density	0.03	1.70	0.39
***Alu* Repeat Content**	0.007	1.59	0.50
LINE Repeat Content	-0.007	1.81	0.34

The genome features were plotted against chromosome position (Figure 5.7) together with replication timing to allow comparison of areas where the correlation between replication timing and other features is strong, and regions where correlation was weak.

Figure 5.7 shows that replication timing follows the general patterns of change in the various different genome features. It can be seen that regions such 30-34 Mb along 22q show a very poor correlation with GC content, yet show very good correlations with exon density and *Alu* repeat content.

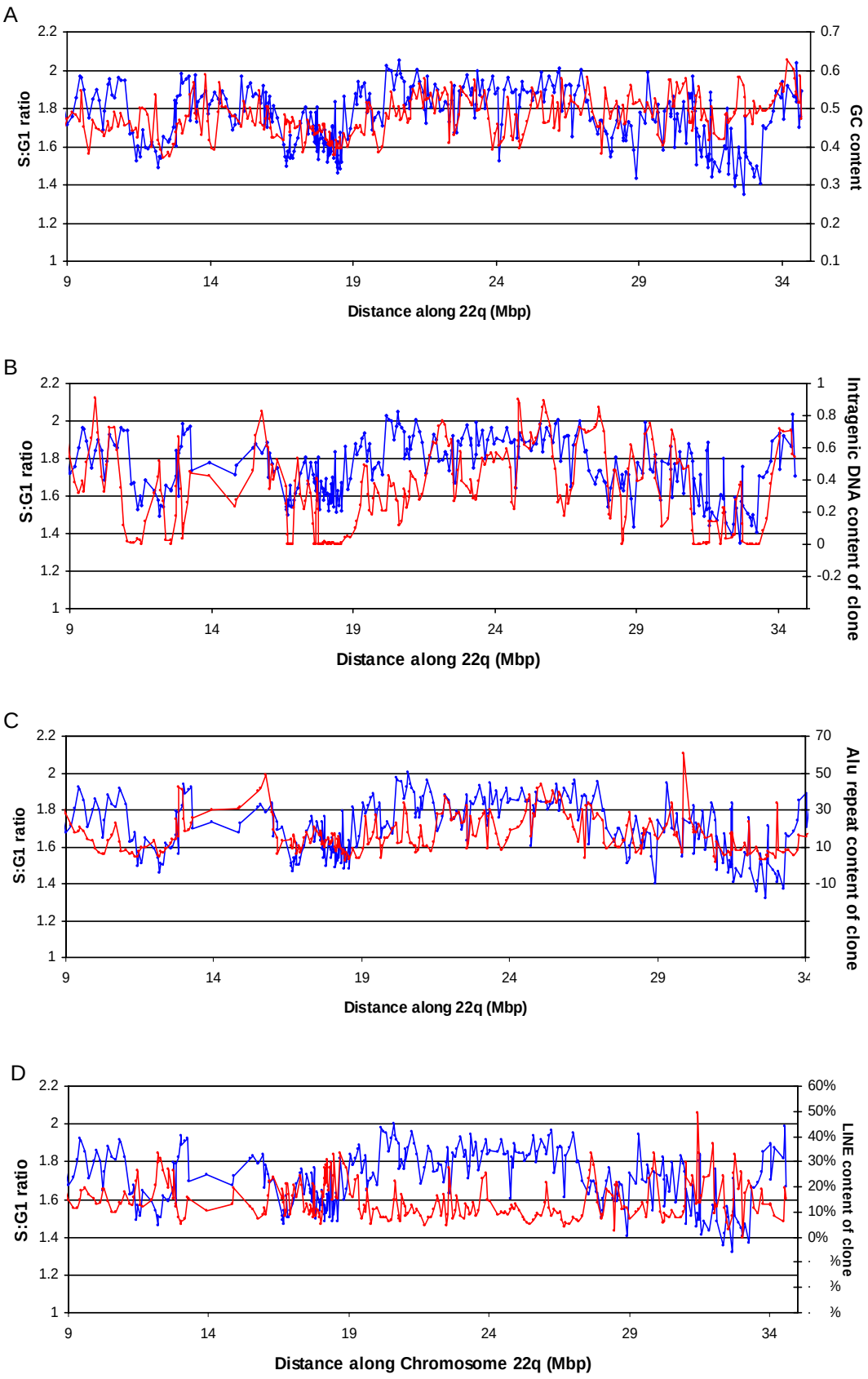


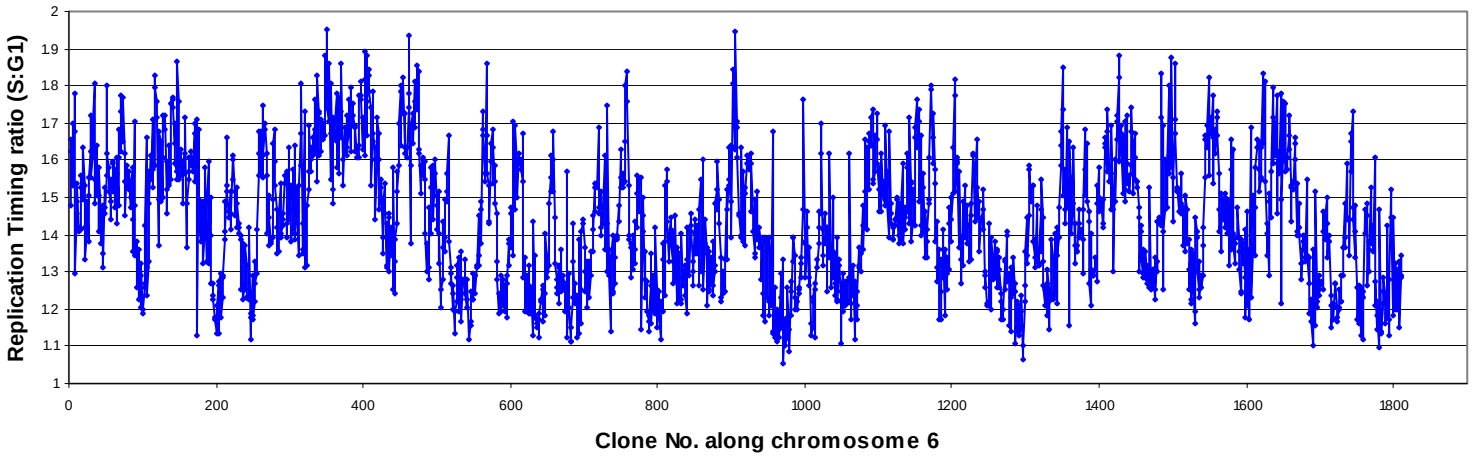
Figure 5.7: Replication Timing profile of chromosome 22 with other genome features. A: GC content, B: Exon density C: *Alu* repeat content D: LINE repeat content.

5.3.2: The Replication Timing of Chromosome 6.

Replication timing was assayed using a chromosome 6 tiling path array. In light of the reproducibility of measurements on the 1Mb and 22 tile path arrays, the number of replicates assayed for chromosome 6 was reduced to two. The average co-efficient of variation between the two replicates is 1.54%. The replication timing profile of chromosome 6 is shown in Figure 5.8. Replication timing is plotted against the order of the clone on the chromosome 6 tile path rather than absolute position because at the time of this work clones for some regions, for example in the MHC locus, had not been mapped onto the finished chromosome 6 sequence.

Linear regression was then performed on this data to assess the correlation between replication timing and sequence features. The relationship between replication timing and sequence features of the genome are shown in Figure 5.9. The linear regression statistics are shown in Table 5.5.

A


Replication Timing ratio (S:G1)
2
1.9
1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1
0
200
400
600
800
1000
1200
1400
1600
1800
Clone No. along chromosome 6

B

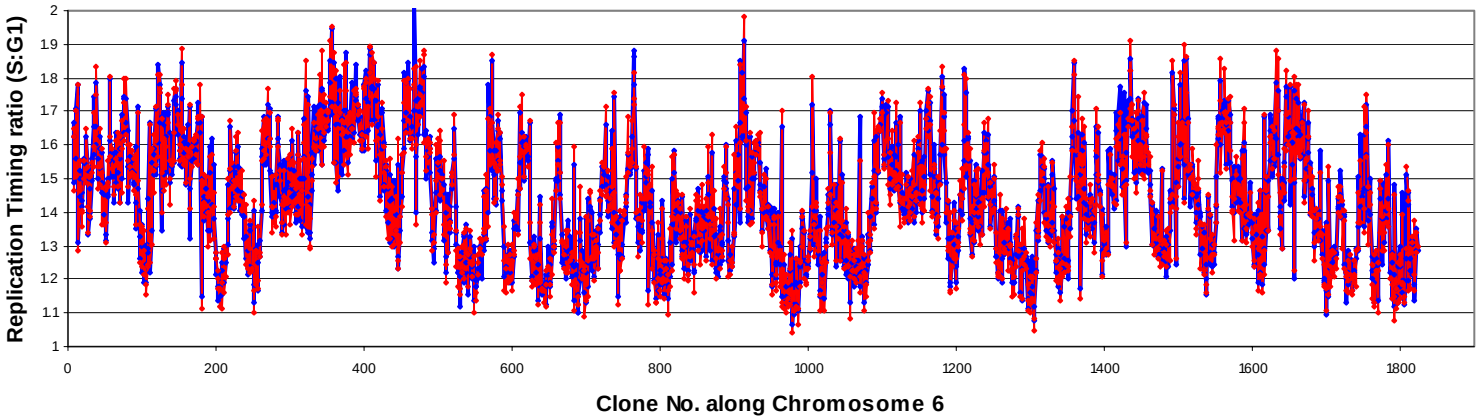

Replication Timing ratio (S:G1)
2
1.9
1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1
0
200
400
600
800
1000
1200
1400
1600
1800
Clone No. along Chromosome 6

Figure 5.8: (Previous page) The replication timing profile of chromosome 6. A: Average replication timing profile of two array experiments. B: Replication timing profile of chromosome 6 showing reproducibility of the data. Blue: replicate 1, Red: replicate 2.

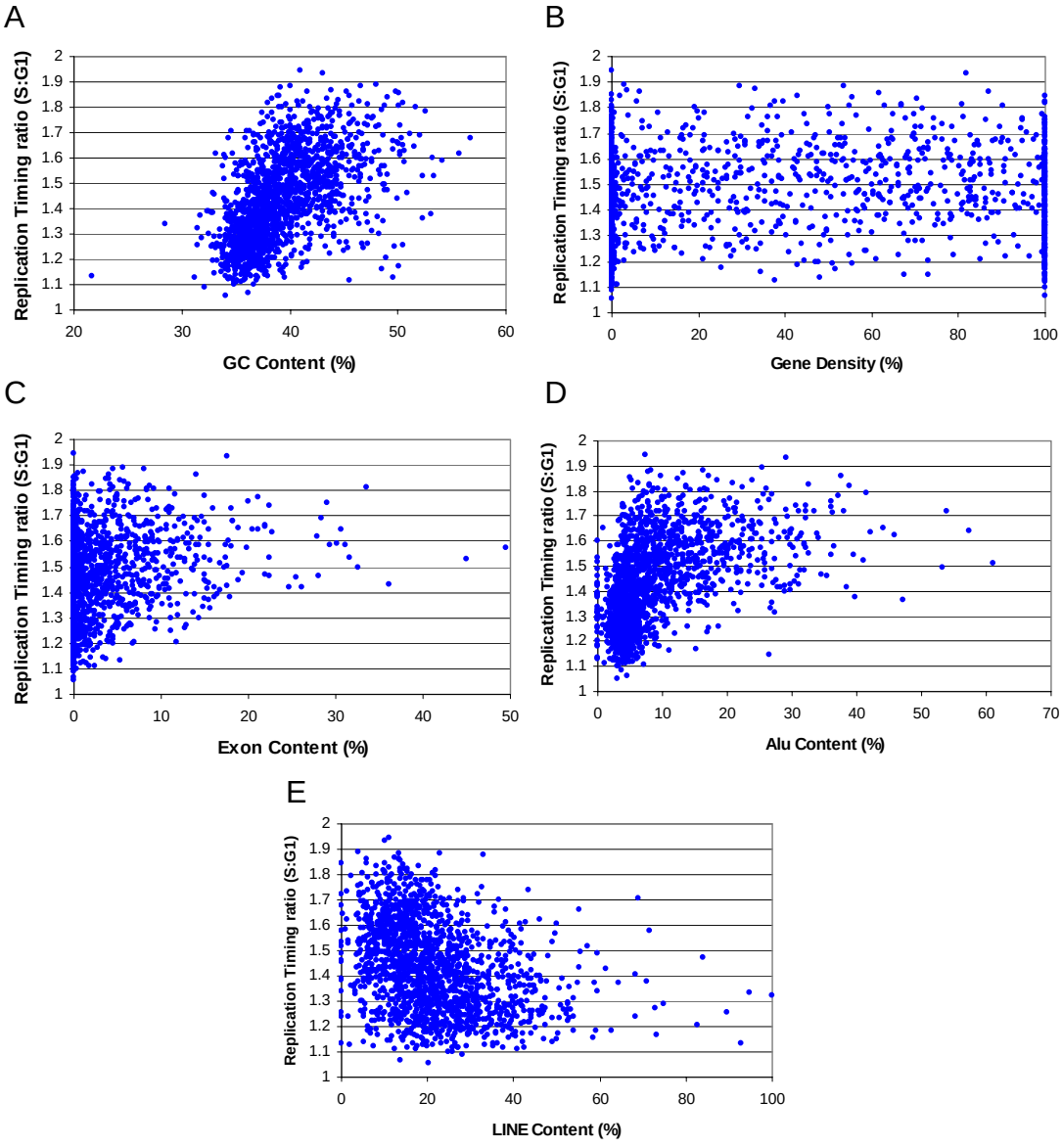


Figure 5.9: Correlation between replication timing and other genome features on the chromosome 6 tile path array. A: Correlation with GC content. B: Correlation with gene density. C: Correlation with exon density. D: Correlation with *Alu* repeat content. E: Correlation with LINE repeat content.

Table 5.5: Linear regression performed between replication timing and genome statistics at a 94Kb resolution on the chromosome 6 tile path array.

Genome Feature	**Regression Coefficient (x)**	**Intercept (y)**	**Correlation Coefficient (r)**
GC Content	0.024	0.48	0.54
Gene Density	0.001	1.40	0.22
Exon density	0.010	1.41	0.30
***Alu* Repeat Content**	0.010	1.34	0.48
LINE Repeat Content	-0.005	1.53	0.34

Unlike the chromosome 22 tile path data, but in common with the correlations seen at a 1Mb resolution and on a chromosome wide analysis, the best correlation was observed with GC content. The worst correlation observed was with gene density, with a correlation co-efficient of just 0.22.

5.3.3: The Replication Timing of Chromosome 1

Duplicate chromosome 1 tiling path arrays were used to assay the replication timing of this chromosome. The average co-efficient of variation between the two replicates was 2.24%. The replication timing profile of chromosome 1 is shown in Figure 5.10.

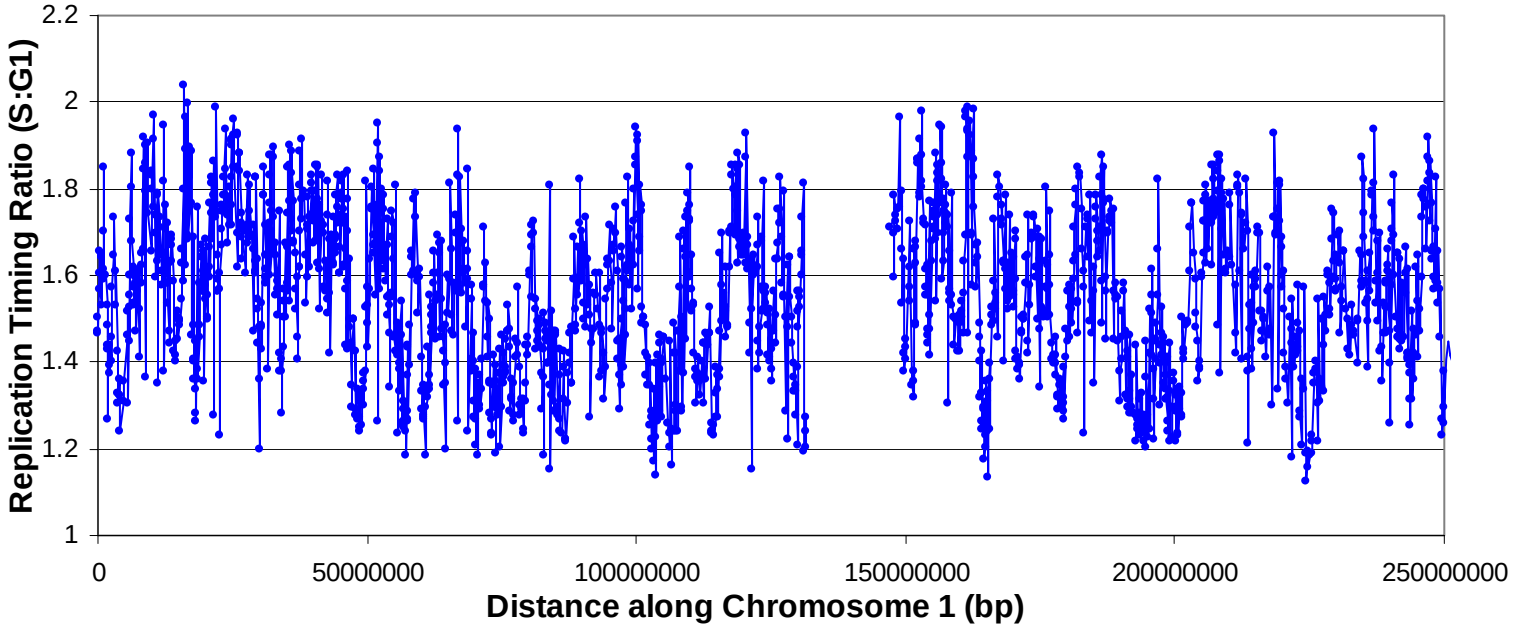


Figure 5.10: The replication timing profile of chromosome 1. The ratios reported are the average of two arrays.

Linear regression was then performed on this data to assess the correlation between replication timing and sequence features. The relationship between replication timing and sequence features of the genome are shown in Figure 5.11. The linear regression statistics are shown in Table 5.6.

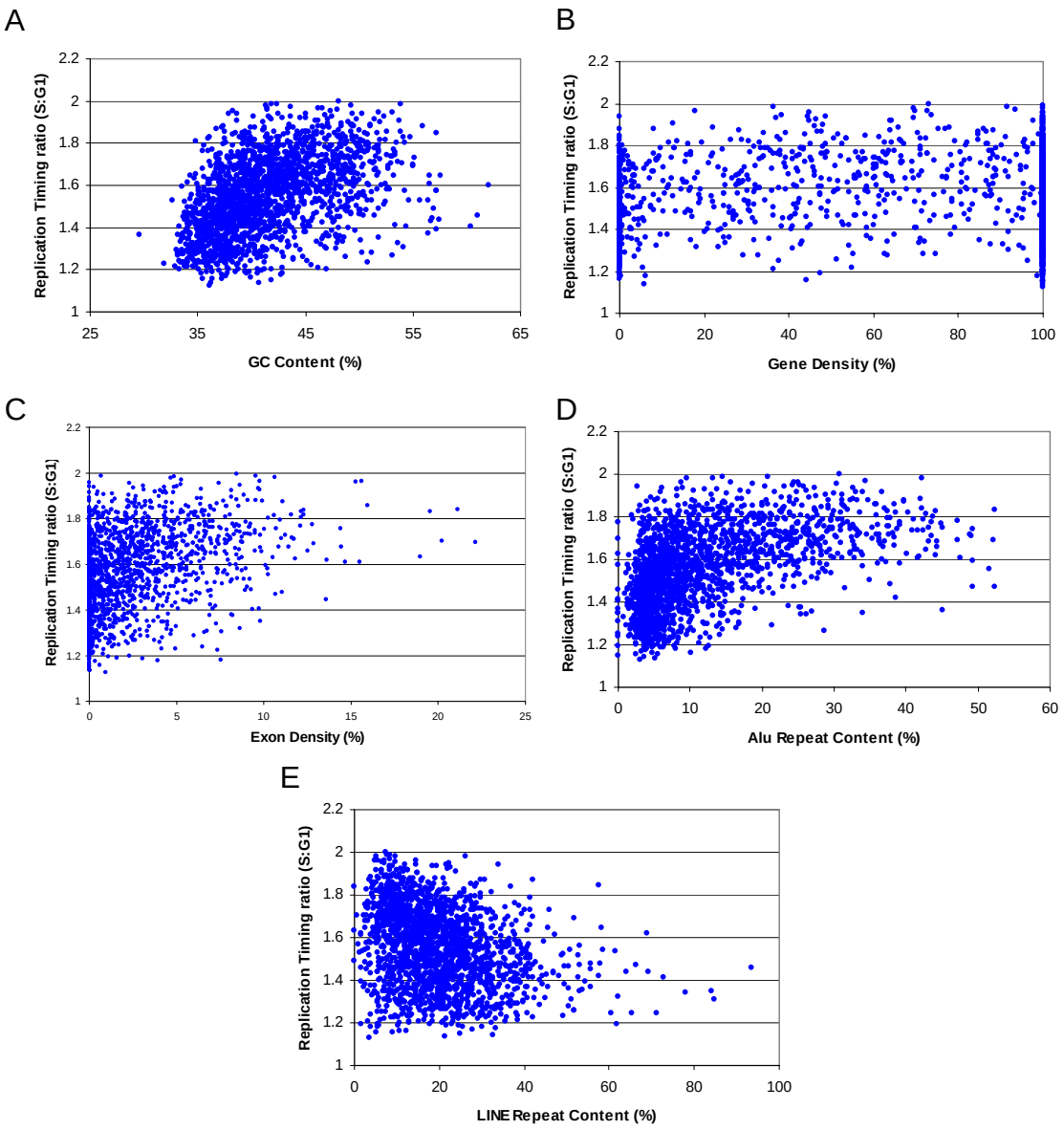


Figure 5.11: Correlation between replication timing and other genome features on the chromosome 1 tile path array. A: Correlation with GC content. B: Correlation with gene density. C: Correlation with exon density. D: Correlation with *Alu* repeat content. E: Correlation with LINE repeat content.

Table 5.6: Linear regression performed between replication timing and genome statistics at a 94Kb resolution on the chromosome 1 tile path array.

Genome Feature	Regression Coefficient (x)	Intercept (y)	Correlation Coefficient (r)
GC Content	0.016	0.88	0.45
Gene Density	0.0004	1.52	0.08
Exon density	0.024	1.50	0.40
***Alu* Repeat Content**	0.009	1.44	0.51
LINE Repeat Content	-0.005	1.64	0.30

As reported by the chromosome 22 array, the best correlation was seen with *Alu* repeat content, and again a poor correlation with gene density was found.

5.3.4 Comparison of Replication timing between two different lymphoblastoid cell lines.

The replication timing of two different lymphoblastoid cell lines was examined. S and G1 phase DNA was flow sorted from a HRC 575 (male) cell line and a HRC 160 (female) cell line. The S and G1 phase DNA from each sort was differentially labelled. The DNA from HRC 575 and HRC 160 was then hybridised to individual chromosome 22 tile path arrays.

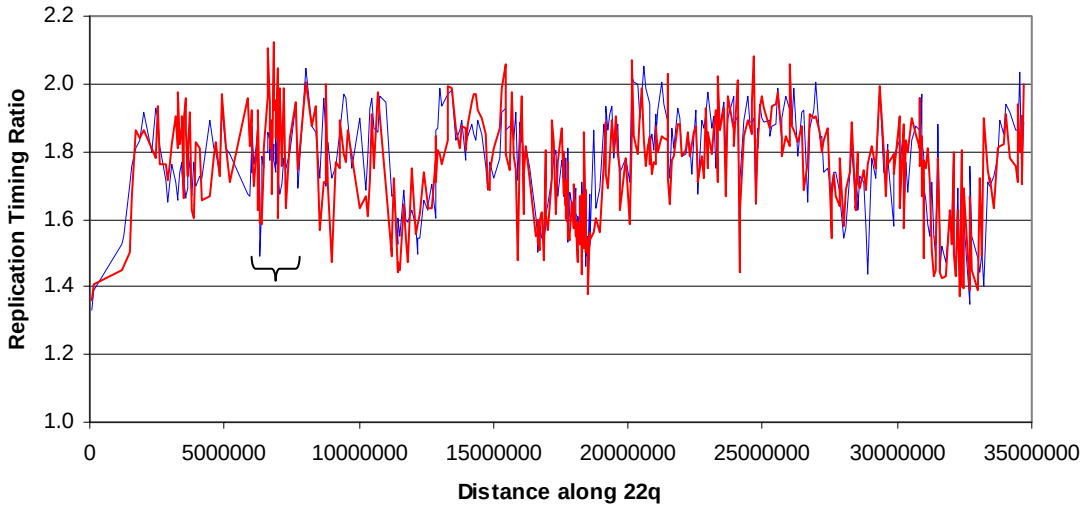


Figure 5.12: Replication timing profiles of two different lymphoblastoid cell lines. Blue: Male lymphoblastoid cell line (HRC 575). This cell line has a deletion in the

immunoglobulin light chain λ region, detailed in 4.2 and marked by the black scroll. Red: Female lymphoblastoid (HRC 160) cell line, with no deletion.

Few regions of replication timing difference can be seen outside the experimental variation of 0.1. The exception is the immunoglobulin light chain λ region, which has been shown to have a different copy number in each cell line, and a region 3Mb from the start of the q arm sequence. Regression analysis was performed on this data to compare the reproducibility of replication timing in different cell lines of lymphoid origin. The correlation coefficient was found to be 0.73 if regression was performed on the whole of 22q, and 0.79 if performed on the last 25Mb used for the sequence feature correlations.

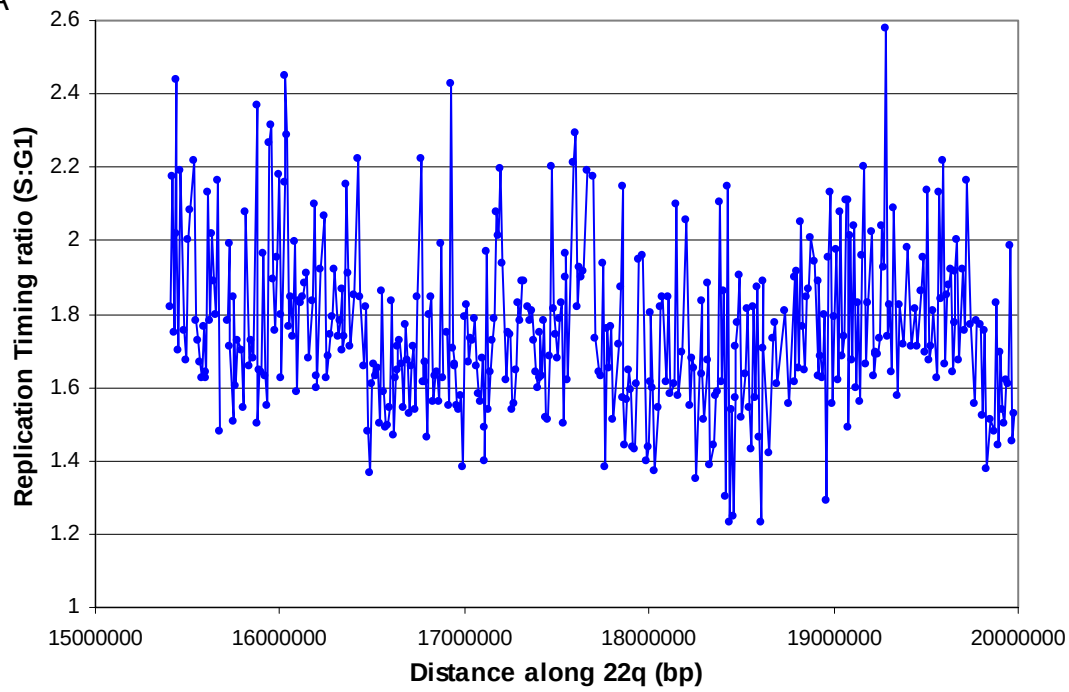
5.4: Assessment of Replication Timing at High Resolution Using an Array constructed with 500bp PCR Products.

An S:G1 hybridisation was performed on an array consisting of high resolution PCR products. In common with the self:self hybridisations described in section 4.6. the noise observed on the PCR product arrays showed a greater amplitude compared to the clone product arrays. The S:G1 hybridisations were repeated six times on separate arrays. The average coefficient of variation for each locus on the array was 14.47%. This is very similar to the coefficient of variation of 14.65% produced by the self:self hybridisation by the chromosome 22 products on the high resolution array reported in section 4.6.

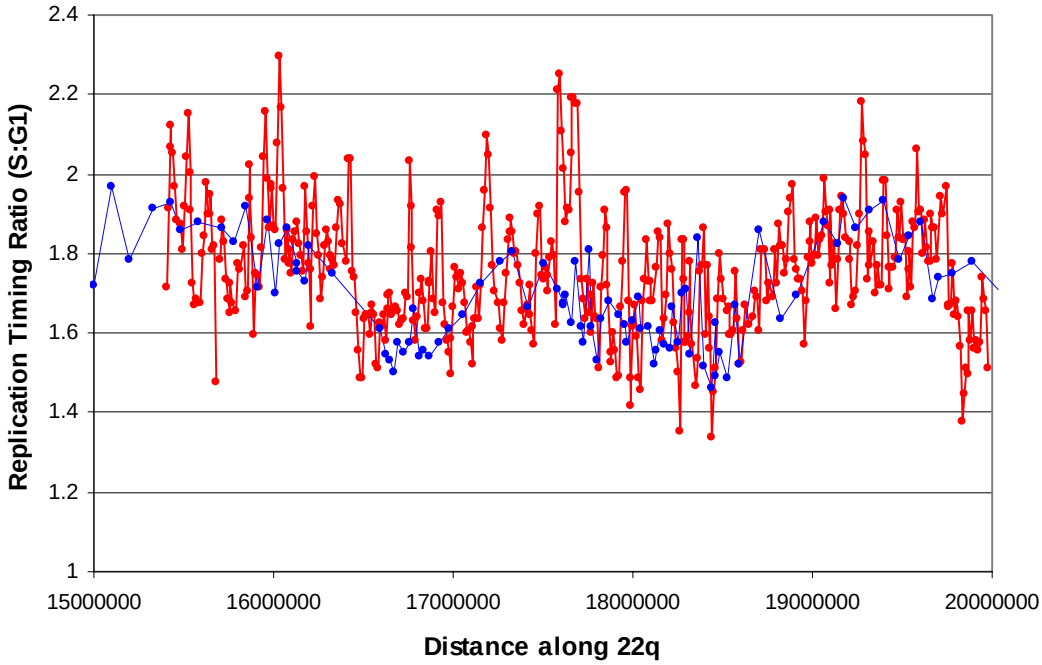
5.4.1: PCR Product array at 10Kb resolution.

The region of chromosome 22 15.4 - 20Mb along the q arm was assayed at a 10Kb resolution. This region was chosen because it included a transition from late-early replication when the timing was assayed on the 22q tile path array. The transition was identified as being approx 16.4Mb along 22q and was between clones dJ90G24 and cN38H9. The array was normalised to the average replication timing observed for the 15.5-20Mb region on the 22q tile path array (i.e 1.668). The replication timing profile for this region is shown in Figure 5.13.

A


Replication Timing ratio (S:G1)
2.6
2.4
2.2
2
1.8
1.6
1.4
1.2
1
15000000
16000000
17000000
18000000
19000000
20000000
Distance along 22q (bp)

B


Replication Timing Ratio (S:G1)
2.4
2.2
2
1.8
1.6
1.4
1.2
1
15000000
16000000
17000000
18000000
19000000
20000000
Distance along 22q

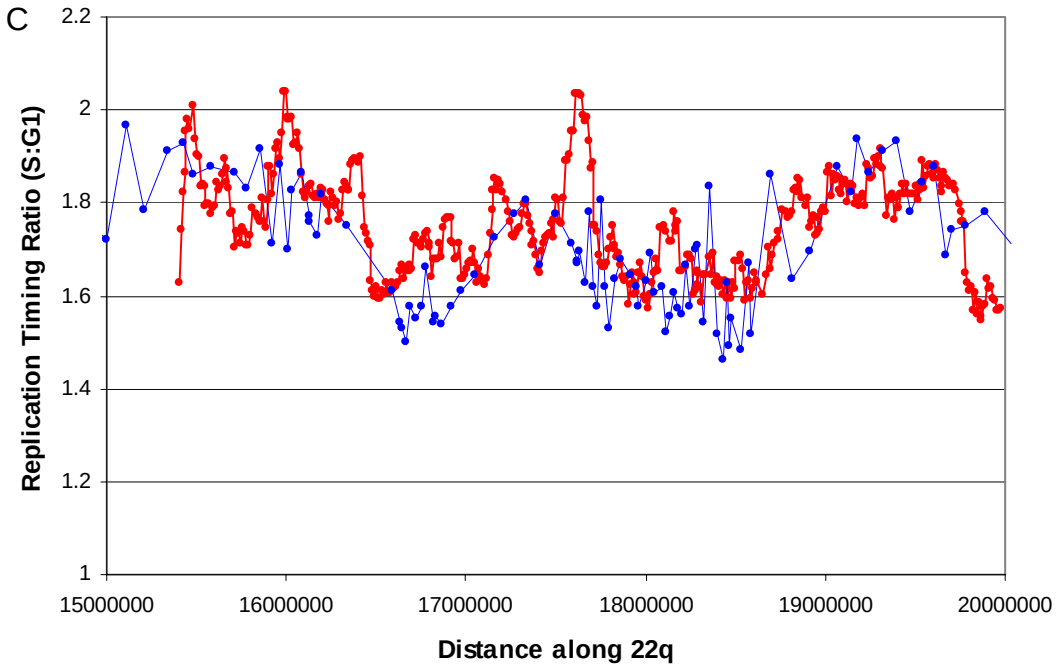


Figure 5.13: Replication timing profile of the region 15.4-20Kb along 22q at a 10Kb resolution. A: Replication Timing sampling all loci. B: Average replication timing utilising a 30Kb moving window. Red: Data from 500bp PCR product arrays. Blue:Data from the 22q clone array. C: Average replication timing utilising a 100Kb moving window. Red: Data from 500bp PCR product arrays. Blue: Data from the 22q clone array.

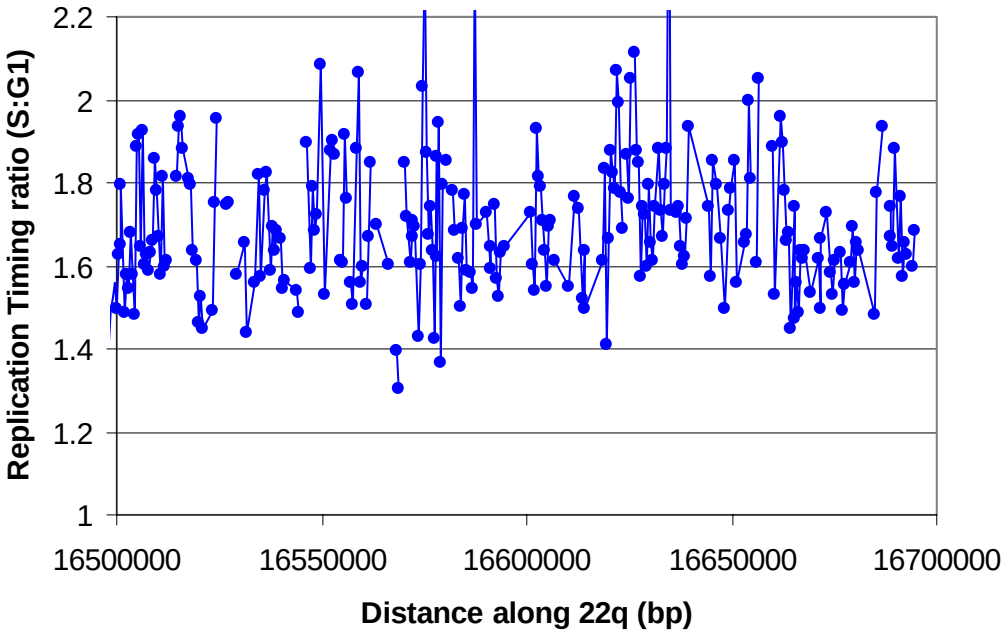
A transition in replication timing, from an early region to a late region can be observed 16.43-16.48 Mb along chromosome 22q. This shows a parallel with what is observed on the 22q tile path array, but narrows down the transition region from the 16.33-16.59Mb along chromosome 22q seen on the lower resolution tile path array. However transitions not observed on the 22q tile path array can be seen on this higher resolution array. A comparison between the profiles observed on the tile path and high resolution array can be seen on Figure 5.13.

The replication timing profile on the high resolution arrays shows a general correlation with those reported by the 22 tile path array. However there are also some inconsistencies, for instance, the early replicating region observed on the high resolution array between 17.8-17.9Mb is not detected by the 22q tile path clones within this region.

Chromosome X PCR products were also spotted onto the array for copy number change verification. The average replication timing of chromosome X is significantly later replicating than chromosome 22 (mean S:G1 ratio 1.38 as opposed to 1.75). The PCR products derived from chromosome X should therefore show a later replication than those from chromosome 22. This was found, with the average replication timing of the chromosome X products being 0.76, compared to 1.67 on chromosome 22. The reporting of a ratio less than 1, the high co-efficient of variation and the reporting of ratios above 2:1 on the chromosome 22 products, reflects that there remain problems with the detection of replication timing on these PCR product arrays. This is discussed further in section 5.7.7.

5.4.2: PCR product array utilising overlapping 500bp products.

The region 16.5-16.7Mb along chromosome 22q was sampled using overlapping 500bp products. The array was normalised as described in 4.2. The replication timing profile for this region is shown in Figure 5.14.



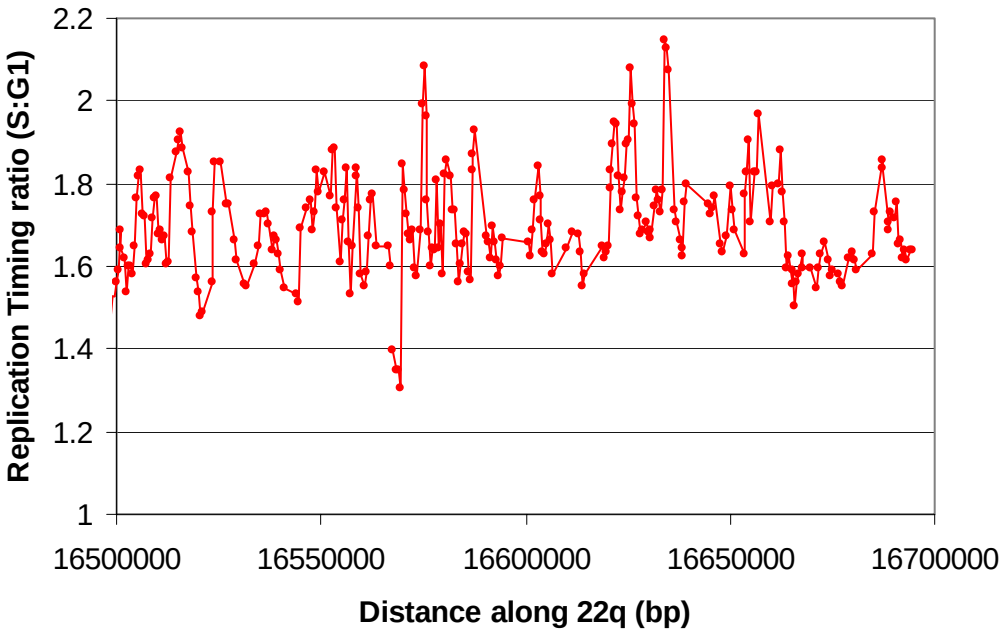


Figure 5.14: Replication Timing Profile of a region of 22q represented on the array by overlapping 500bp PCR products. A: Replication Timing sampling all loci. B: Average replication timing utilising a 1500bp moving window.

The profiles plotted in Figure 5.14 show gaps where the replication timing ratio is not reported. This is because sequences in these regions are not represented on the array

as insufficient unique sequence was available to allow the design of specific PCR primers.

A comparison between the profiles observed on the tile path, 10Kb resolution and a 500bp array can be seen in Figure 5.15.

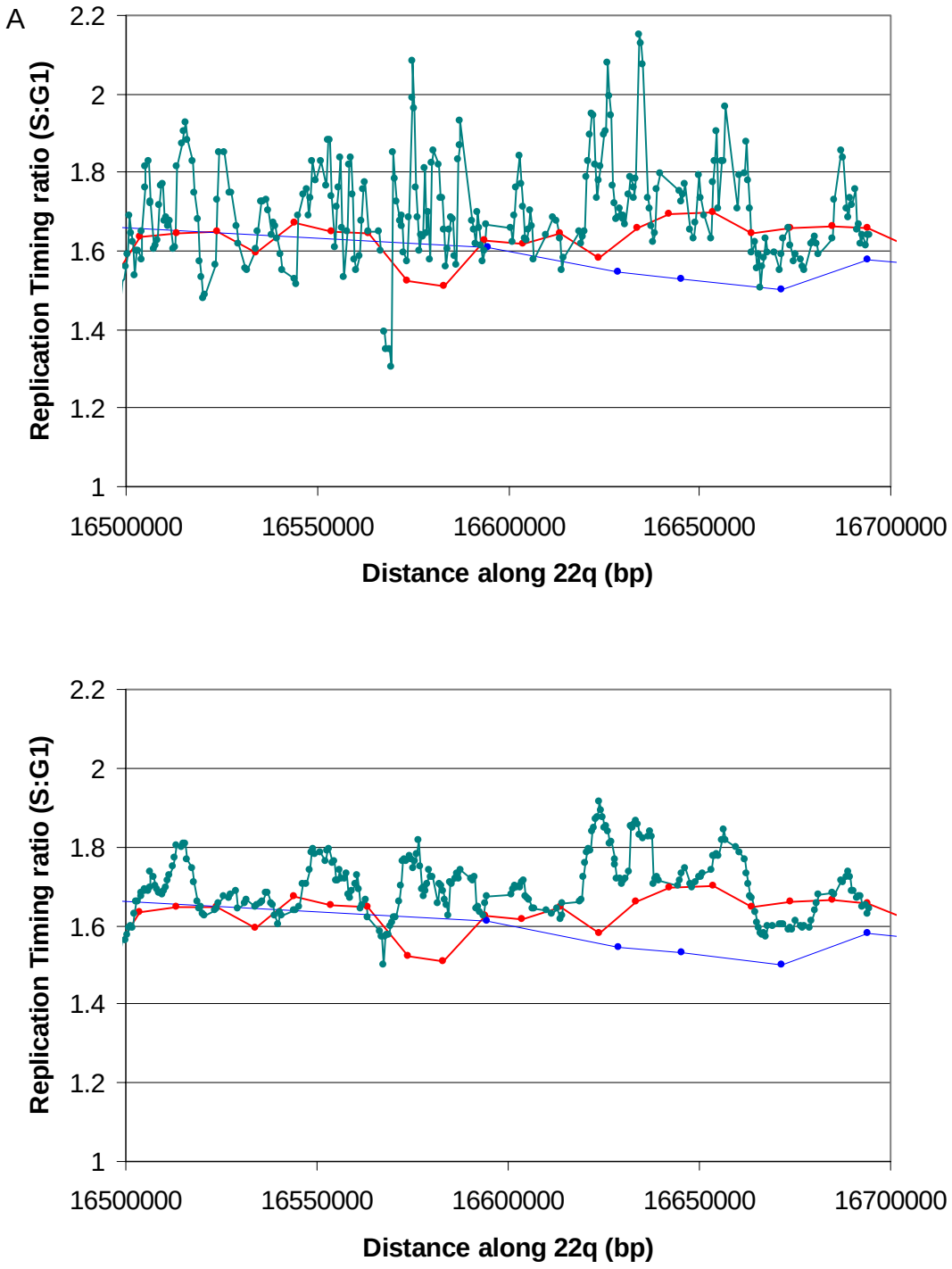


Figure 5.15: (Previous page) A: Comparison of the replication timing profile obtained on the 22 tile path array (blue) and a 10Kb resolution PCR product array (red) and a 500bp resolution array (green). B: Comparison of the replication timing profile obtained on the 22 tile path array (blue) and a 10Kb resolution PCR product array (red) and moving 5000bp window from the 500bp array (green).

The 500bp products show a profile that is different from that obtained with the 10Kb resolution array. However the 10Kb resolution array shows a good correlation with high resolution arrays at the points where 500bp products coincide with regions represented on the 500bp resolution array.

5.5: Correlation of assessment of Replication Timing by arrays with Replication Timing assessed by Quantitative PCR.

In order to validate my replication timing approach, I was able to compare our replication timing assay results from the 1Mb resolution array with a previously published independent analysis of chromosome arm 11q (Watanabe, Fujiyama et al. 2002).

Further corroboration of the method was performed by selecting clones represented on the 22q tile path array and calculating the difference in copy number between sorted S and G1 phase fractions by real time PCR.

5.5.1 Correlation with published quantitative PCR data on Chromosome 11q.

Wantanabe *et al* (Watanabe, Fujiyama et al. 2002) separated nuclei from a monocytic leukaemia cell line (46, XY) by flow sorting into 4 S phase fractions, extracted nascent DNA and then used semi-quantitative PCR to identify fractions enriched for specific STSs across the chromosome arm. DNA replicated in the 4th S phase fraction in the Wantanabe *et al* data would correspond to a late replication timing ratio on the array of 1.25:1 or below. Like-wise, replication in the 3rd S phase fraction would correspond to a replication timing ratio of between 1.25 and 1.5, *et cetera*.

The STSs on 11q were sequenced and remapped according to Build 31 of the human genome on the University of California, Santa Cruz website. The average spacing of each STS used in the Watanabe *et al* data is 300kb (Watanabe, Fujiyama et al. 2002), which is a resolution higher than that obtained with the 1Mb array.

The replication profile for both methods is shown plotted against chromosome 11 position in Figure 5.16. The replication profile of the two methods is very similar. Slight discrepancies can be seen at approximately 67, 95 and 111Mb along chromosome 11, where the replication timing reported by the arrays is later than that reported by the quantitative PCR method. A possible explanation for this discrepancy is that the two studies use a different cell line. The array data is obtained from a lymphoblastoid cell line with a normal karyotype and the Q-PCR data is obtained from a monocytic leukaemia cell line.

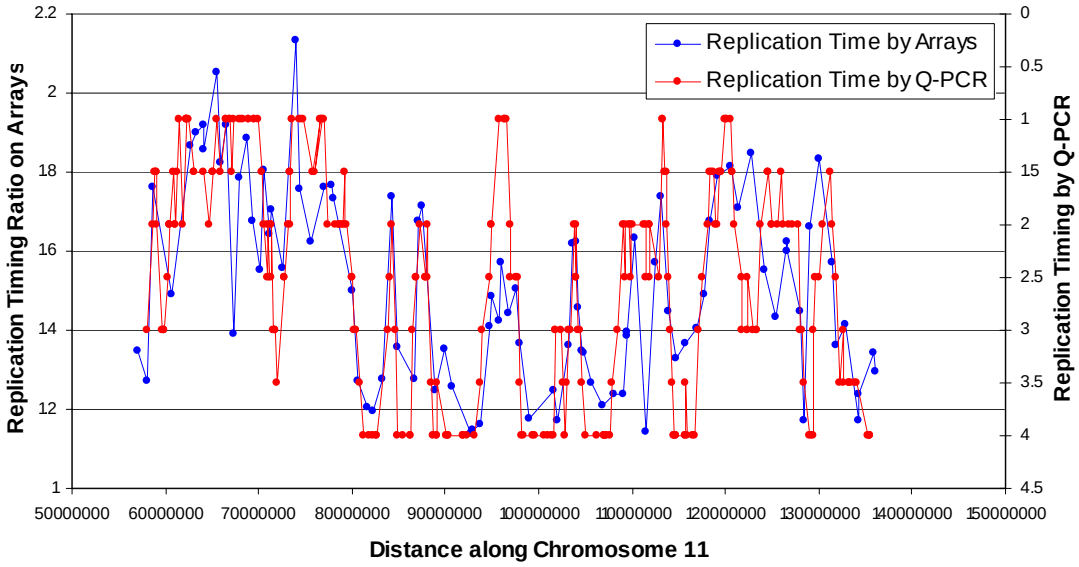


Figure 5.16: Replication Timing on 11q. Blue: Replication Timing reported by the 1 Mb array. Red: Replication Time reported by Wantanabe *et al.*

Data points within 100Kb of each other, from the two different methods were correlated as shown in Figure 5.17.

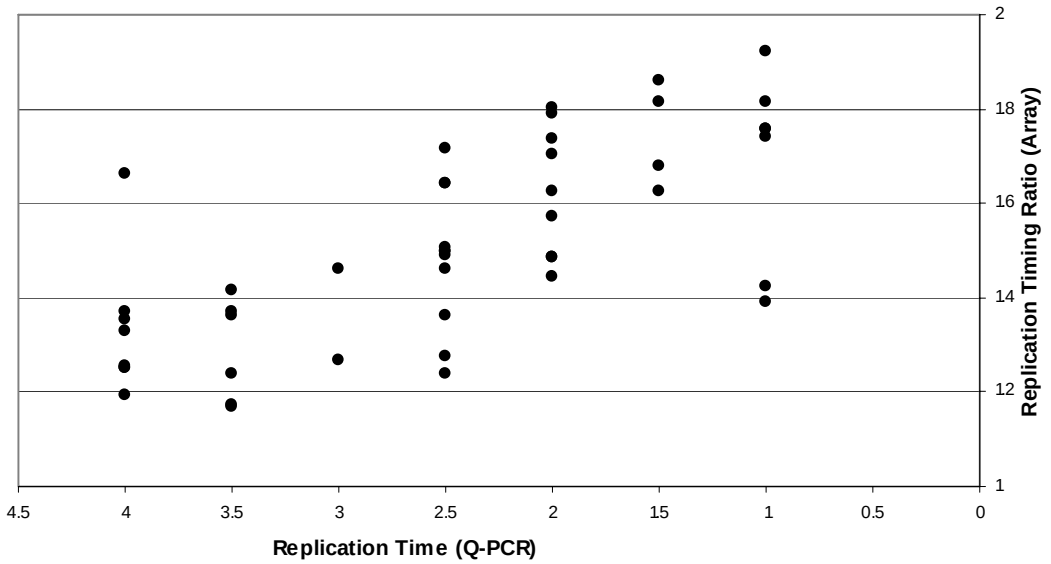


Figure 5.17: Correlation between quantitative PCR data and array data of loci within 100Kb of each other (y = -0.148x + 1.88 by linear regression).

A moderate-strong correlation (r=0.69) was found between the two methods. It should be noted that this is despite the use of two different cell types, albeit both lymphoid in origin.

The MHC region is located on chromosome 6 so replication timing data generated with the chromosome tiling path array can be used for comparison with previously published studies at this locus.

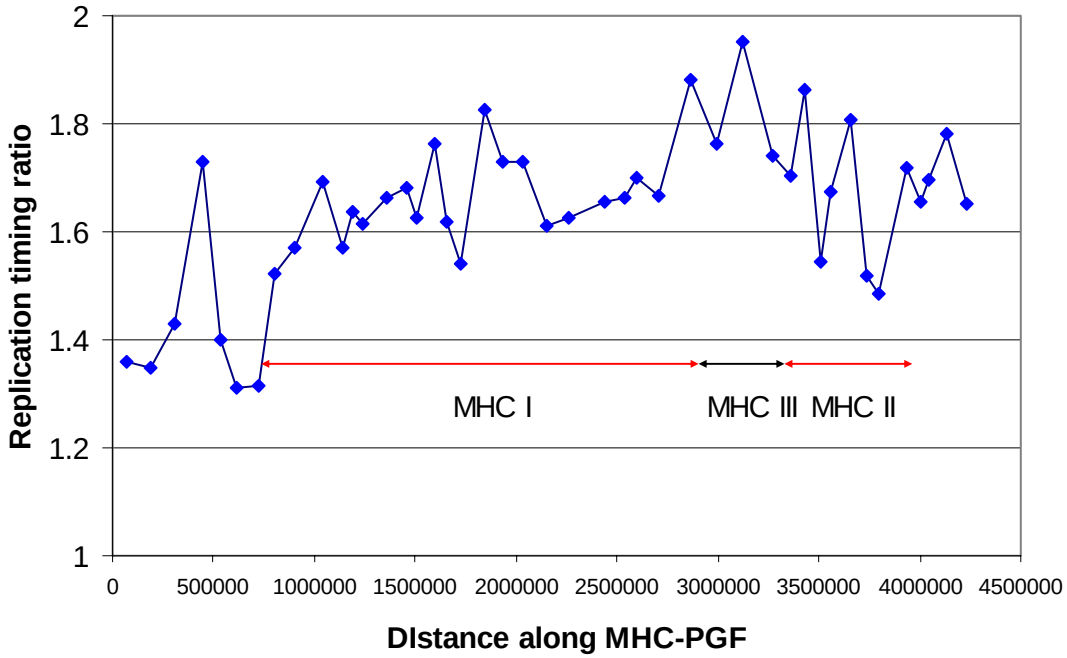


Figure 5.18: Replication timing data of the MHC region collected from the chromosome 6 tile path array

The replication timing data obtained by the arrays show a transition from early – late replication between MHC III and MHC II. This confirms what has been observed when this region was studied at high resolution by Tenzen *et al* (Tenzen, Yamagata et al. 1997) using a PCR based method.

5.5.2: Verification of replication timing by arrays by analysis by Quantitative PCR

To further verify our approach, four clones from chromosome 22 were chosen for analysis by real time PCR. These were one late replicating clone (cN69F4, position on X axis=1.38), two mid replicating clones (cE140F8 X=1.71 & cB13C9 X=1.64) and an early replicating clone (bK57G9 X=1.97). Primer pairs were designed every 10Kb along the clone. Each primer pair was assayed by real time PCR in quadruplicate. The average coefficient of variation was 7.2%. The S:G1 ratio for each primer pair was calculated and compared to the ratio obtained for the entire clone by array analysis (detailed in section 5.3.1) as shown in Figure 5.19.

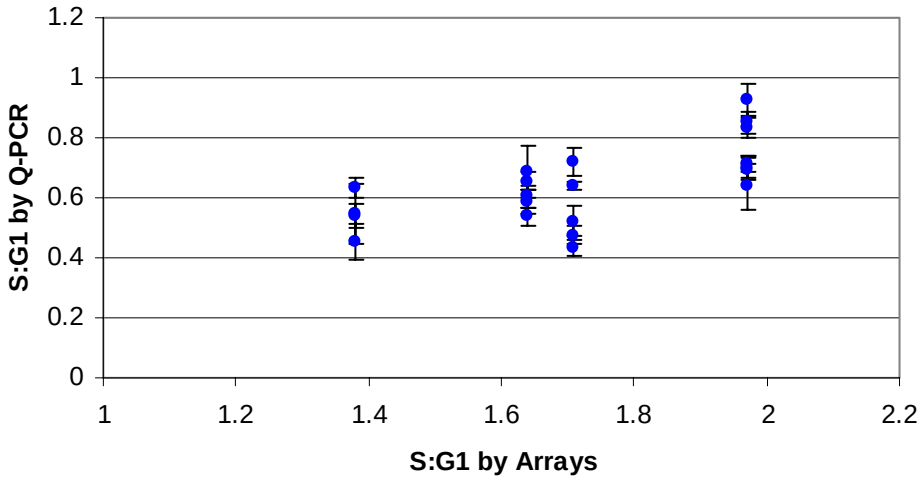


Figure 5.19: Comparison of S:G1 as determined by the replication timing arrays (X axis) and by quantitative PCR (Y axis). Y error bars show the standard error of each quadruplicate for the quantitative PCR experiment.

The PCR data was averaged over a clone (to make them more comparable with the array data) and the correlation co-efficient was calculated as 0.87. This supports the data presented in section 5.5.1 and reveals that using microarrays to assess replication timing produces comparable data to that produced by a method utilising real time PCR.

5.6: Replication time in flow sorted S phase fractions.

S phase was sorted into five different fractions as shown in Figure 2.2 and DNA was extracted. The four fractions were co-hybridised against G1 in four separate experiments. G1:G1 and G2/M:G1 hybridisations were also performed. These were normalised as described in Table 2.7.

The five earliest replicating, five latest replicating and five mid replicating clones were selected from the replication timing profile shown in Figure 5.5. The ratio obtained for each fraction is plotted in Figure 5.20.

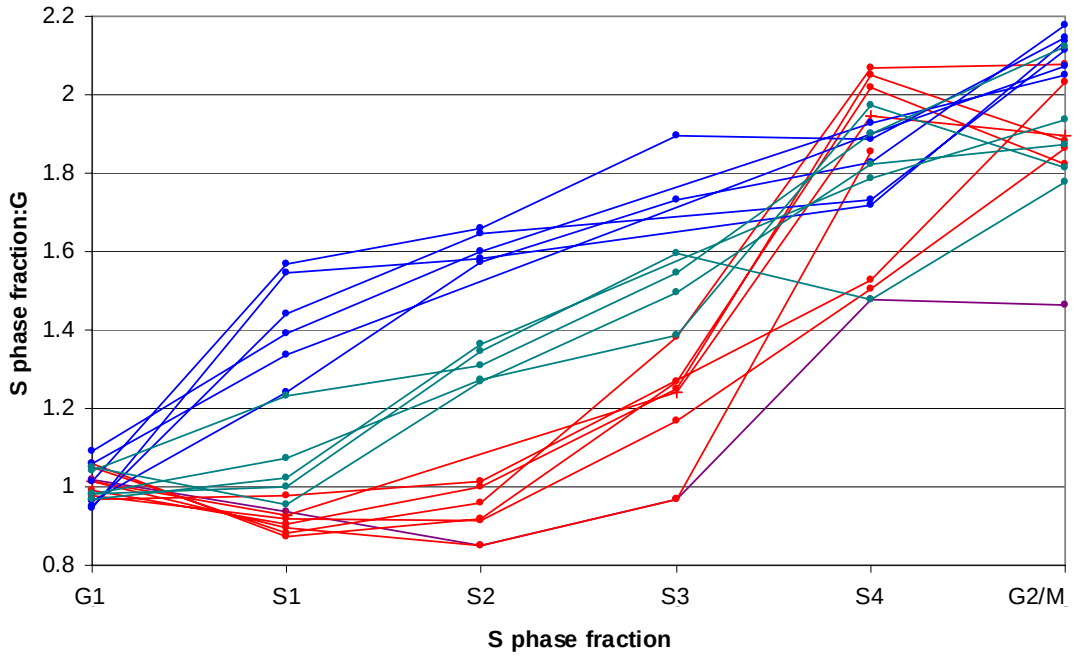


Figure 5.20: Ratio obtained for each S phase fraction when hybridised against G1. Four classes of clones are show. Blue: Early replicating, Green: Mid replicating, Red: Late replicating, Purple: centromeric, containing a large amount of common repeat elements. The clones used in this analysis are detailed in Table 5.7.

Table 5.7: Clones used for the analysis displayed in Figure 5.20

Class	Clones
Early replicating	dJ355C18, bK212A2, bK57G9, cN84E4, dJ1119A7, bK221G9
Mid replicating	dJ90G24, cE140F8, dJ15I23, bK243E7, cB13C9
Late replicating	cN2H8, cN29F4, cN22D1, cN69F4, bK262A13, bA191L9, cN129H9
Centromeric	cN14H11

Early replicating clones will double in copy number within the first fraction of the S phase sort (S1). They will then remain at a double copy number throughout the rest of S phase fractions. Conversely, late replicating clones will remain at a single copy number throughout early S phase fractions and double in copy number in the S4 fraction. Mid replicating clones will double in copy number in the S2 or S3 fractions. All clones should have replicated by the G2/M fraction

This is reflected in Figure 5.21. Early replicating clones increase in copy number in early S phase fractions, while late replicating clones increase in copy number in late S phase. Most clones show an average ratio of 2:1 when the G2/M fraction is ratioed

against G1. One clone in which this is not the case is the clone cN14H11 (purple on Figure 5.21). This clone is the most centromeric sequence clone of the q arm of chromosome 22 and is rich in common repeat elements. The incomplete suppression of these repeat elements may explain why the ratio reported for the G2/M:G1 hybridisation is only 1.46, instead of 2:1.

A region that showed a transition between an early and a late replicating DNA was also analysed. The region chosen was 26.8-28.0Mb along 22q. The results can be seen in Figure 5.21.

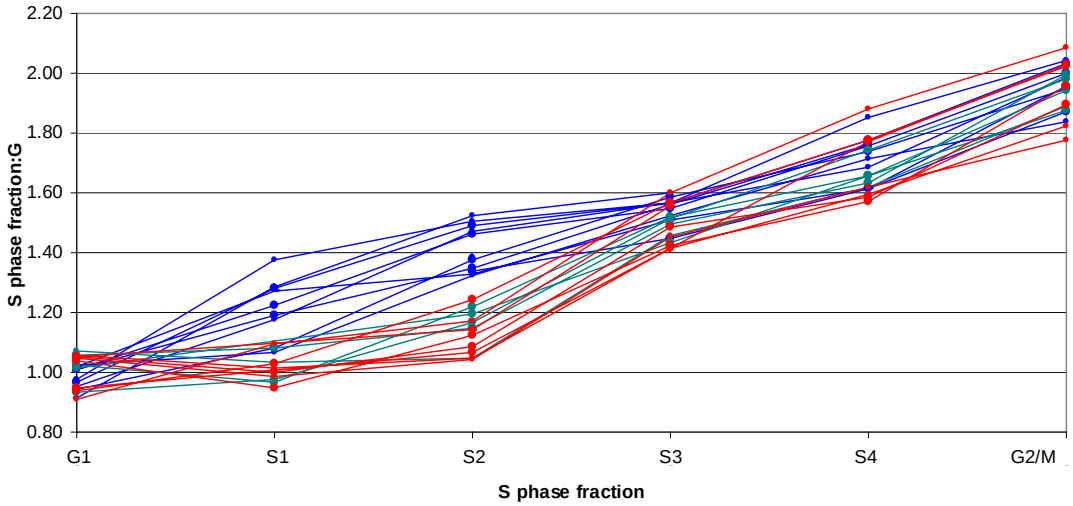


Figure 5.21: Ratio obtained for each S phase fraction when hybridised against G1. Blue: Early replicating side of transition, Green: Clones within the transition of replication timing, Red: Late replicating side of transition.

This analysis of consecutive overlapping clones confirms what is seen in Figure 5.22. Early replicating clones increase in copy number in early S phase fractions whilst the later replicating clones increase in copy number in later S phase fractions.

Replication timing analysis of the whole genome (section 5.2.1) shows that chromosome 22 is an early replicating chromosome, with an average replication timing ratio of 1.75:1. This is confirmed by a great majority of the chromosome 22 clones increasing in copy number in the S1 or S2 fraction of S phase. Conversely chromosome X is late replicating with a replication timing ratio of 1.38:1. Analysis of chromosome X clones on the array reveals that all clones representing chromosome X

increase copy number within fractions representing the latter half of S phase (Figure 5.22).

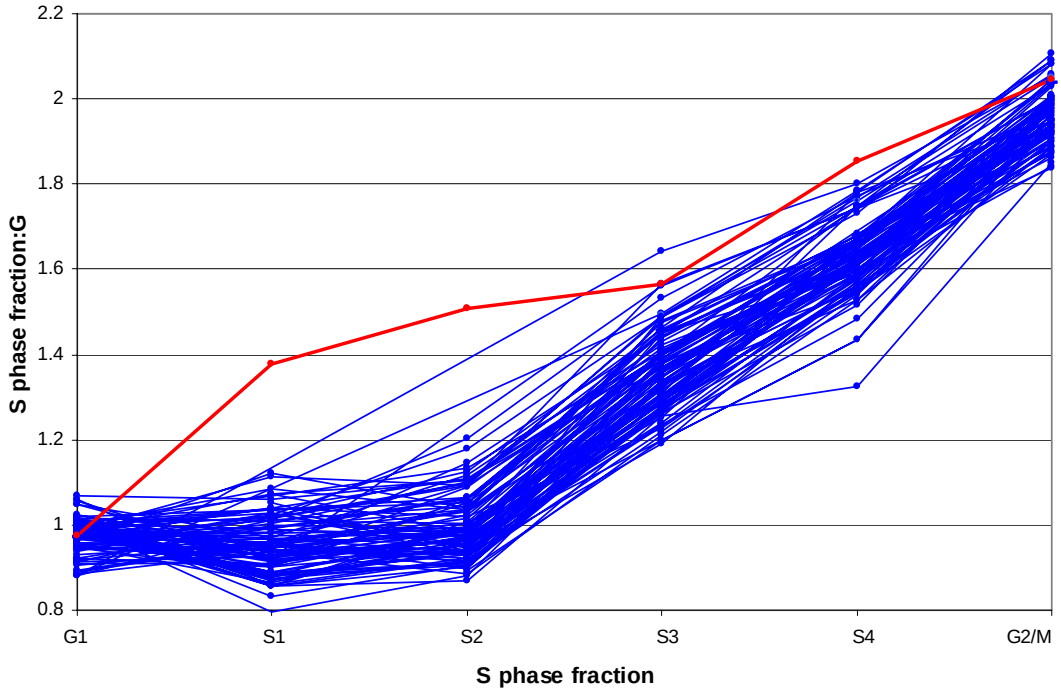


Figure 5.22: Ratio obtained on loci representing chromosome X for each S phase fraction when hybridised against G1 (Blue: chromosome X clones. Red: a typical chromosome 22 clone).

5.7: Discussion

The data presented in Sections 5.2-5.4 demonstrate how microarrays sampling genome sequence can be used to assess replication timing. Unlike conventional methods of assaying replication timing, genomic arrays report the replication timing of large genomic regions with a high accuracy. The spatial resolution of the method is only limited by the clone size and density of clones represented on the array. The work in this Chapter describes, for the first time, a high resolution replication timing analysis of a mammalian genome.

The replication timing was correlated with sequence features of the genome. Correlation coefficients for each feature are as reported in sections 5.2 and 5.3 and summarised in Table 5.8. Regions of early replication map to G light chromosomal bands, whilst regions of late replication map to G dark chromosomal bands.

Table 5.8: Regression co-efficients for correlations between replication timing and sequence features of the genome. (TP = tile path array). The strongest individual correlations are highlighted in red, whilst the weakest correlations are highlighted in blue. The best correlations are seen when all sequence features are combined, and are highlighted in green.

Genome Feature	**Chromosome Wide**	**1 Mb Chip**	**22 TP**	**6 TP**	**1 TP**
GC Content	0.96	0.7	0.22	0.54	0.49
Gene Density	0.89	0.35	0.19	0.22	0.08
Exon density	Not done	0.42	0.39	0.15	0.41
***Alu* Repeat Content**	0.9	0.56	0.45	0.48	0.51
LINE Repeat Content	0.72	0.4	0.34	0.34	0.3
Multiple Regression Analysis	Not done	0.76	0.57	Not done	Not done

5.7.1: Correlation between Replication Timing and Sequence Features.

Initial analysis of replication timing on a chromosome wide level shows that chromosomes of similar sizes exhibit very different replication timings. For example chromosomes 18 and 21 were very late replicating, whilst chromosomes 19 and 22 were early replicating. Chromosome 18 has already been shown to be late replicating and very gene poor. In contrast chromosome 19 is early replicating and gene dense (Zink, Bornfleth et al. 1999; Cross, Clark et al. 2000). Chromosome X which is abnormally rich in LINE repeats (IHGSC 2001), and the Y chromosome which includes a large amount of heterochromatin on the q arm, were both shown to be late replicating.

This initial large scale analysis also revealed that the gene deserts on chromosomes 13 (48-89Mb) and 14 (79-86Mb) defined by sequence analysis of the genome (IHGSC 2001) are late replicating. This is shown in Figure 5.23.

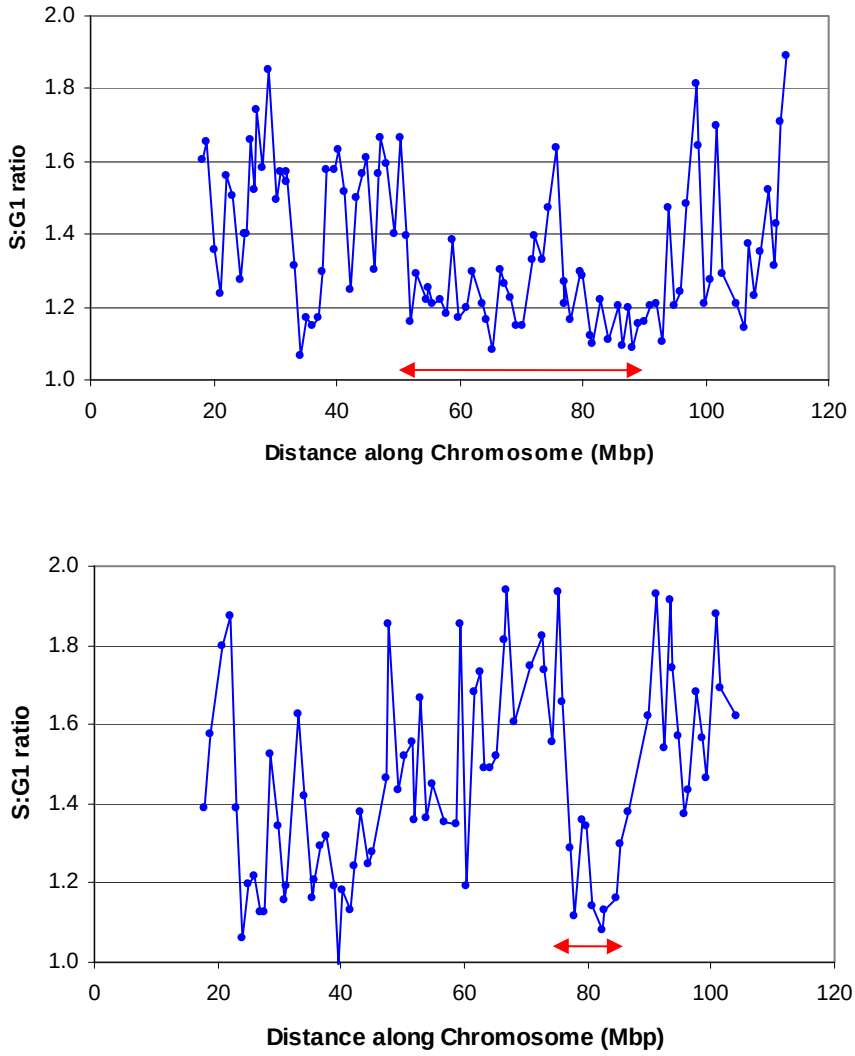


Figure 5.23: Replication timing of chromosomes 13 and 14. Gene deserts are marked with the red arrow.

Initial inspection of the regression coefficients shows that as the sampling resolution of the genome increases, the correlations with sequence features decrease. Therefore the chromosome wide correlations are much better than those using information from the tiling path arrays. As all the features are interrelated, as shown in Table 1.1, it is difficult to determine individually which features drive early replication. The effect of genome features on replication timing will be considered in turn below.

These results demonstrate that replication timing can be assayed at a tiling path level and that the replication timing can be correlated with other genome features. However the average resolution of the 22q tiling path array is only 78Kb and the average

replicon is thought to be approx. 40-100Kb (Nakamura 1986; Natale, Li et al. 2000) so to assay replication timing at the level of the replicon a higher resolution array needs to be used.

5.7.1.1: Correlation between Replication Timing and GC Content.

The best correlation on a chromosome wide basis was seen with GC content. This was also the case when the whole genome was sampled at a 1Mb resolution. The correlation between replication timing and GC content has been previously reported, including across an R/G band boundary (Strehl, LaSalle et al. 1997) and along entire chromosome arms (Watanabe, Fujiyama et al. 2002).

Analysis of the correlation with GC content on a tile path level gives disparate results. The correlation on chromosome 22 was weak, with a correlation coefficient of 0.22. The only feature on chromosome 22 showing a weaker correlation was with gene density. However on chromosome 6 GC content shows the strongest correlation with replication timing with a correlation coefficient of 0.54. Analysis of the replication timing and GC content of chromosome 22, in relation to chromosomal position (Figure 5.7), shows that, generally, the replication timing does follow the GC content of this chromosome. However, replication timing and GC content become uncorrelated at 22q13. This region is unusual in that while it is GC rich it is gene and *Alu* repeat poor. In this region it appears that gene density rather than GC content may drive replication timing.

The correlation with GC content on the chromosome 6 and 1 tile path arrays is quite good. It should be noted that chromosome 22 is very small. Small regions of difference between GC content and replication timing at 22q13 have a large effect on the correlation performed on all of 22q. The results in Table 5.8 show that GC content is important and may influence replication timing.

5.7.1.2: Correlation with Gene Density.

A weaker correlation was observed with measures of gene density. Two different measures of gene density were assayed; gene density (all intragenic DNA) and exon density (Exonic DNA only). Gene density was defined as the percentage of exonic and intronic DNA that is found within each clone. Exon density was defined as the percentage of exonic DNA (not that within introns) found within a clone. The correlation with gene density on a chromosome wide level was strong, with a correlation coefficient of 0.89 as reported in Table 5.8. However the correlation coefficient on the other arrays was poor. The correlation with gene density was the weakest seen on three of the four arrays analysed, with the correlation on chromosome 1 being just 0.08. On the chromosome 6 tile path array the correlation with gene density was the second weakest with a correlation coefficient of 0.22. In the case of the chromosome 6 tile path the weakest correlation was with exon density. In this case the correlation coefficient was 0.15.

The correlations with exon density are also weaker than those observed with other genome features. The correlation seen on the tile path arrays were variable, as with GC content. The correlation on the chromosome 1 tile path array was modest with a correlation coefficient of 0.41, however the correlation on the chromosome 6 tile path was very weak with a correlation coefficient of 0.15.

This analysis shows that that the correlation between replication timing and measures of gene density were weak in comparison to those with GC content.

5.7.1.3: Correlation with Common Repeat Elements.

Replication timing was also correlated with two types of common repeat element, *Alu* repeats and LINE repeats. In line with the other genome features the best correlations were seen at a chromosome wide level.

The correlations with *Alu* repeats were strong. The correlation between replication timing and *Alu* repeat content exhibited either the strongest or second-strongest correlation at all resolutions tested.

The correlation between replication timing and LINE repeats was the poorest of all the sequence features investigated at a genome wide level. The correlations at a 1Mb and tile path resolution were all somewhat similar with the coefficient correlations ranging from 0.3-0.4.

When looking at the correlation with LINE content on a chromosome wide level an outlier can be identified. (Figure 5.2D at 32.8, 1.38); this is the locus that corresponds to chromosome X. Chromosome X is known to be unusually rich in LINE repeat elements (IHGSC 2001). If this point is removed from the analysis the correlation co-efficient is 0.88, which is similar to the correlation co-efficient of the other features.

Analysis of the correlation between replication timing and repeat content shows a strong positive correlation, whilst a negative correlation was observed with LINE repeat content. This is consistent with the characteristics of active and inert chromatin as documented in Table 1.1.

5.7.1.4: Inter-correlation between replication timing and sequence features.

Statistical analysis performed in collaboration with Richard Mott (Wellcome Trust Centre for Human Genetics, Oxford) showed, by multiple regression analysis, that the genome features investigated were highly correlated with each other.

The multiple regression analysis showed that the correlation coefficient between replication timing and all the sequence features explored was 0.75 when the genome was sampled at a 1Mb resolution, a small but highly statistically significant ($P<10^{-16}$) improvement over the 0.70 correlation with GC content data alone. Multiple regression on chromosome 22 revealed a correlation with all sequence features of 0.57. This a considerable improvement over the correlation with the best single sequence feature (*Alu* – 0.45, $P<10^{-16}$).

The correlations reported, are all indicative of transcriptionally active, open forms of chromatin being important for early replication. A further investigation between

replication timing and euchromatic, transcriptionally active chromatin is reported in section 6.2 and discussed in section 6.5.1.

5.7.2: Correlation between Replication Timing and chromosomal bands.

5.7.2.1: Correlation with Giemsa banding.

It is widely acknowledged that R bands (GC rich) in mammalian chromosomes replicate in the first half of S phase and G bands (GC poor) replicate late (Ganner and Evans 1971; Dutrillaux, Couturier et al. 1976; Holmquist, Gray et al. 1982). The correlation of replication timing with high resolution G banding is shown in Figure 5.24. The replication timing profile at a 1Mb level of chromosome 6 shows by visual inspection that G dark regions generally replicate late, such as those 48-51Mb and 93-96Mb along the chromosome, and G light bands replicate early, such as those 27-46Mb and 105-113Mb along the chromosome. These correlations can also be seen at a tile path resolution on chromosome 22 with the G dark bands 16.6-18.8Mb and 112.2-12.5Mb along the chromosome replicating late, conversely G light bands located 33.6-34.7Mb and 12.9-15.9Mb along the chromosome replicate early. The associations with G banding cannot be exact due to the different condensation levels of G dark and G light regions of metaphase chromosomes.

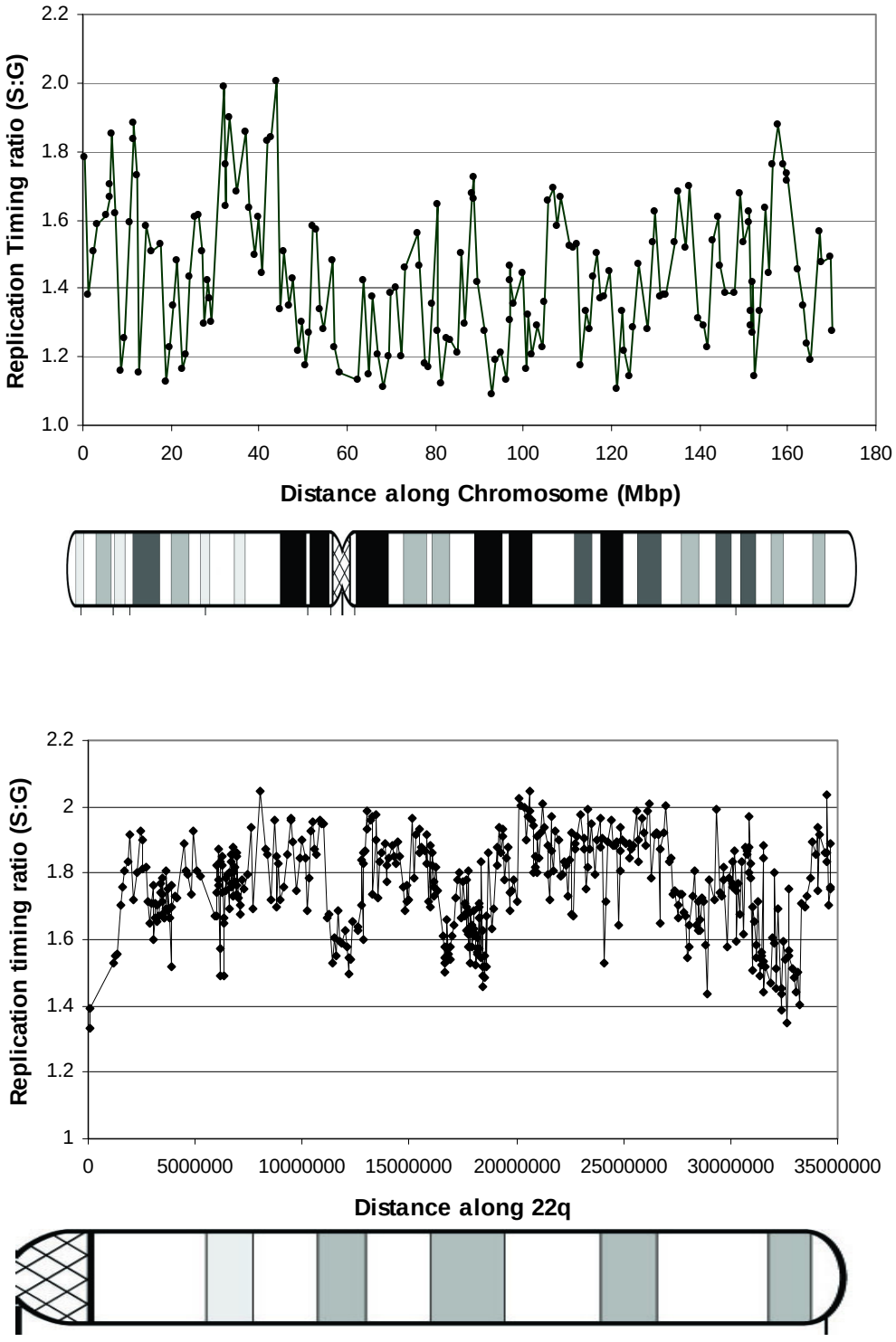


Figure 5.24: Comparison between replication timing ratio and high resolution giemsa banding of chromosomes (resolution = 850 bands, (Francke 1994)). (A) Chromosome 6 at a 1Mb resolution (B) Chromosome 22q at a tile path resolution

Statistical analysis on the replication timing ratios of clones that map to dark and light bands reveal there is a significant difference between the replication timing of sequence located in dark bands and sequence located in light bands.

An unpaired T test with Welch correction (to allow for the difference in means in the two populations when the variances are unequal) was performed on the chromosome 6 data obtained from the 1Mb array. Each locus on the array was assigned in either a dark or light band. The average replication timing of loci in light bands was 1.534 (standard deviation = 0.195) whilst the average replication timing of loci in dark bands was 1.334 (standard deviation = 0.164). The difference in replication timing observed was highly statistically significant with a P value less than 0.0001.

The same analysis was performed on the chromosome 22 tile path array data. The average replication timing of loci in light bands was 1.808 (standard deviation = 0.125) whilst the average replication timing of loci in dark bands was 1.677 (standard deviation = 0.153). The difference in replication timing observed was also highly statistically significant with a P value less than 0.0001.

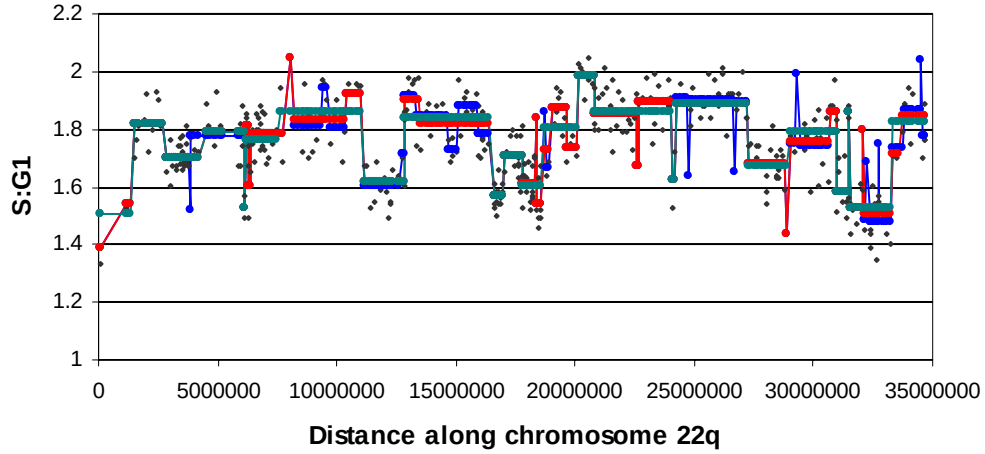
5.7.2.2: Regions of co-ordinated replication.

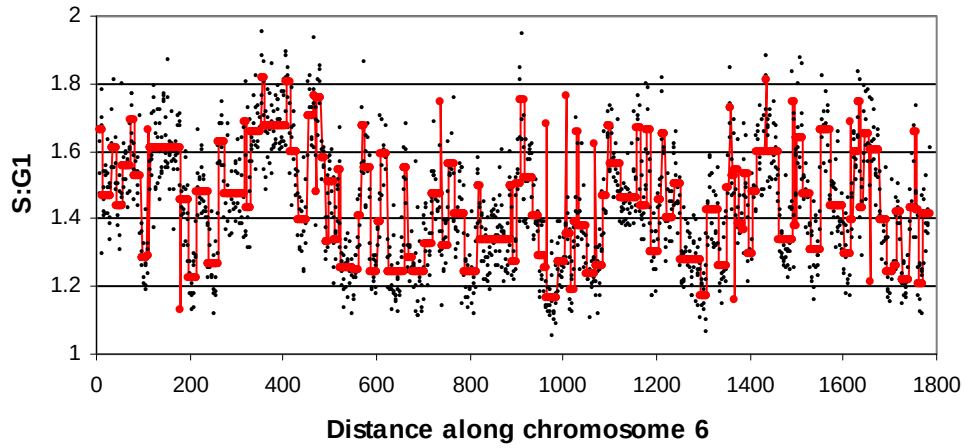
In order to assess the patterns of replication timing observed in the plot of the tile path data, we attempted to identify regions of similar replication timing and regions which differed significantly in replication timing from adjacent stretches in chromosomes 22, 6 and 1. A perl script was purpose written by Richard Mott at the Wellcome Trust Centre for Human Genetics for analysis of the replication timing data produced by the arrays. The program (detailed in Appendix 7) was used to find the optimal segmentation of the chromosome tile path data. Although the degree of segmentation observed can be adjusted by altering the segmentation penalty values, *B*, and it is not completely clear what a biologically meaningful value of this parameter should be; the analysis has the effect of delineating the patterns that are indicated by visual inspection.

This analysis was performed altering the segmentation penalty values (B), This was executed only on the 22q tile path data and indicated that the patterns of segmentation in the data was highly non-random, with $P < 0.001$.

Results of this analysis for a series of representative values of *B* are shown in Figure 5.25a for chromosome 22q. This illustrates that chromosome 22 has clear segments of consistently very early replicating DNA stretching over several megabases. Interspersed within these are megabase sized segments of later replicating DNA. Transitions between segments of early and late replicating areas of chromosome 22 (and vice-versa) are observed between data points whose midpoints are less than 160Kb apart (e.g. at ~11100000bp and ~12700000bp) suggesting disparate replication timing of adjacent replicons.

The statistical analysis described above was also performed on the data obtained from the chromosome 6 and chromosome 1 tile path arrays, using an intermediate segmentation penalty value. Consistent with the analysis of the chromosome 22 data, megabase sized segments of early, or late, replicating DNA could be identified. This is shown in Figure 5.25b and 5.25c.





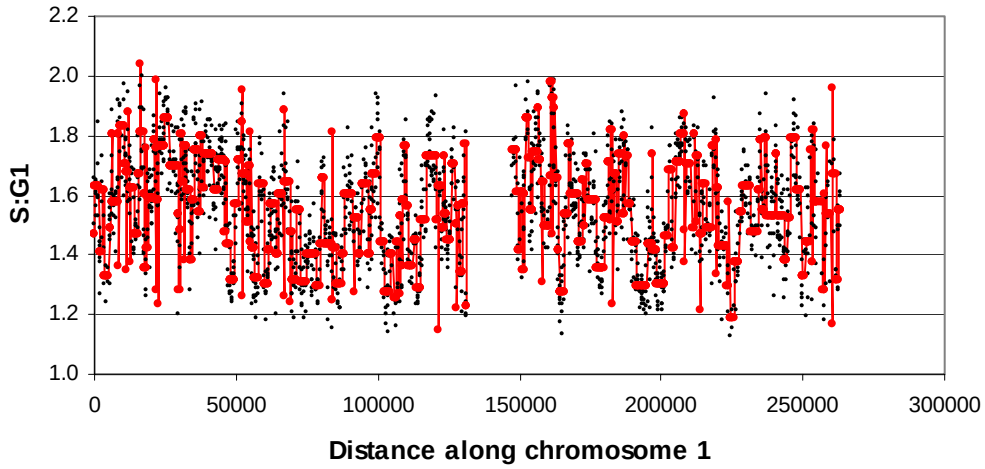


Figure 5.25: Statistical analysis to identify regions of the genome with similar replication timing. A: The graph shows the results of three runs of segmentation on the chromosome 22q data using representative segmentation penalty score (B) of 0.02 (blue), 0.04 (red) and 0.06 (green). Segmentation runs are plotted on top of the raw

replication timing data (black circles). B: Chromosome 6 (B = 0.04) C: Chromosome 1 (B = 0.04)

The comparison of the replication timing profiles with the banding patterns of chromosomes and the statistical analysis described above detailing the optimal segmentation of the replication timing data, showed that regions of the chromosome stretching over several megabases, replicate at similar times. The correlation with the giemsa banding of the chromosome suggests a link between replication timing and GC content. The identification of regions of several megabases that replicate at the same time suggest that groups of adjacent replicons replicate together. This is consistent with the observations made when chromosomes were studied by pulse labelling with BrdU (Dutrillaux, Couturier et al. 1976; Drouin, Lemieux et al. 1990; Cohen, Cobb et al. 1998), as described in section 1.3.

5.7.3: Rate of Replication

The replication timing data obtained from the 1Mb array can be used to assess the rate of genome replication. For this, S phase was divided into centiles based on S:G1 ratio. The number of loci replicating in each centile was counted and the cumulative number of loci replicated was plotted against the proportion of S phase completed (Figure. 5.26). Replication appears to start slowly, but increases to a linear rate of replication at about a third of the way through S phase finally again appearing to slow at the end of S phase. The slow initial rate of replication is supported by the shape of the distribution of S phase as measured on the flow cytometer (see S phase sorted fraction in Figure. 2.2) where there is a higher frequency of nuclei with lower DNA content. This implies that the DNA content of nuclei increases more slowly at the start of S phase and we can infer that either the frequency of the initiation of replication and/or the length of replicons are reduced during this period. The rate of replication then increases to a linear rate at about a third of the way through S phase. During this linear stage approximately 14% of the genome is replicated during each tenth of S phase. The replication rate slows towards the end of S phase. The slow rate of replication at the end of S phase cannot be explained from the cell cycle profile which displays a relatively even frequency of nuclei with increasing DNA content from the middle of S phase onwards. As most heterochromatin will replicate during this late

stage, and heterochromatic regions are not represented on the 1Mb genome wide arrays, the rate of replication for this final part of S phase is likely to be underestimated.

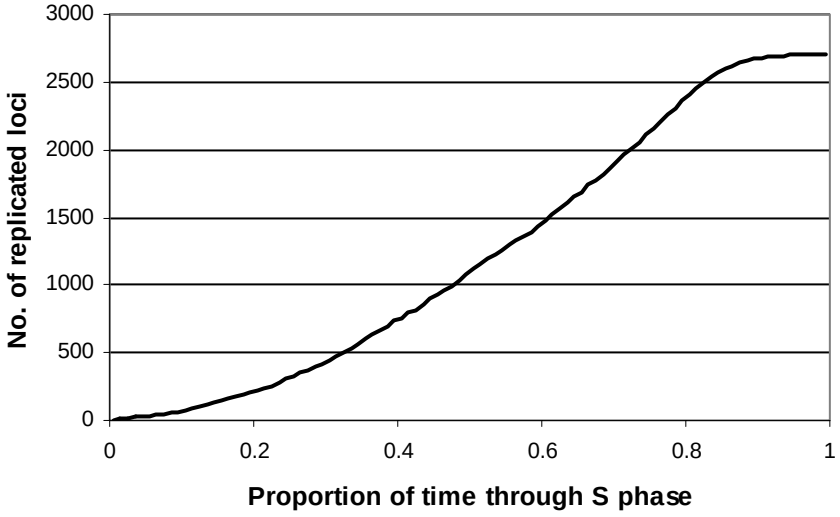


Figure 5.26: The rate of replication during the S phase of the cell cycle. Rate of replication is indicated by the slope of the curve plotted.

5.7.4: Comparison with other arrays assessing replication timing and limitations of the method.

Microarrays have previously been used to assess replication timing in the yeast (Raghuraman, Winzeler et al. 2001) and in *Drosophila melanogaster* (Schubeler, Scalzo et al. 2002). However the method used here is subtly different in a number of ways.

Firstly, the two previous studies have focused solely on coding regions of the genome. The study on yeast used the high density oligonucleotide array produced by Affymetrix. This chip represents each *Saccharomyces cerevisiae* open reading frame with up to 20 oligonucleotide sequences on the array. No regions outside the open reading frame were included (Raghuraman, Winzeler et al. 2001). The study on *Drosophila* used a cDNA array. The array was constructed to represent 5,543 expressed sequence tags from *D. melanogaster*. Both previous studies have therefore not assayed any non-coding regions of the genome, however, in yeast and *Drosophila* there is less non-transcribed DNA than is found in the human genome. The study

described in sections 5.2-5.3 assayed the human genome with DNA prepared from sequencing clones. This ensures both coding and non-coding regions of the genome are sampled. The representation of non-coding regions of the genome enable correlations with other sequence features such as GC content and repeat content to be performed. The inclusion of non-coding sequence also ensures there is an unbiased representation of sequences found in open and closed chromatin. Studies that only assay the replication timing for coding regions of the genome will not assay the replication timing of transcriptionally inert 'closed' chromatin and therefore any conclusions drawn are going to be biased towards what is found in transcriptionally active 'open' chromatin. To produce a complete understanding of replication timing both coding and non-coding regions of the genome should be sampled.

Secondly, the way replicating DNA is identified and extracted for application onto the array differs in each different assay. The methods used are described in 1.6.2. Briefly, for the yeast experiment, newly synthesised DNA was labelled with light carbon and nitrogen isotopes, in a background of DNA labelled with heavy isotopes. Post synchronisation samples were collected throughout S phase and a caesium chloride density gradient was used to separate newly replicated DNA from non replicated DNA. These were differentially labelled and co-hybridised to the array. The *Drosophila* experiment utilised cells pulse labelled with BrdU. The BrdU was incorporated into nascent DNA. The nuclei were then stained with propidium iodide and flow sorted by their PI intensity into an early S phase fraction and a late S phase fraction. The newly replicated DNA from each fraction was isolated by immunoprecipitation, amplified and differentially labelled using PCR.

Both these methods have the limitation that to obtain the replication timing ratio, one S phase fraction is ratioed against a different S phase fraction. This means that both methods may under-report very early replicating DNA sequences. A mean of the early S phase fraction is used as the early replicating reference point. DNA that replicates before this point will not be detected by the array. Also early replicating DNA may not be sufficiently labelled with BrdU to be detected. Using the method described in this thesis, the ratioing of S phase DNA against G1 phase DNA enables early replicating DNA sequences to be detected. This separation of S from G1 phase could not be achieved using the published *Drosophila* cell sort profile as the coefficient of

variation for the G1 and G2 peaks were too high. The S phase sort would therefore be contaminated from DNA from both the G1 and G2/M fractions of the cell cycle. This would affect the replication timing ratio reported. A comparison between the two flow sort profiles can be seen in Figure 5.27A

One limitation of the assessment of replication timing on arrays is the purity of the sort. Any contamination of S phase cells in the G1 fraction means that DNA from very early replicating loci may be present in the G1 fraction hybridised to the array. As a result, the true replication timing ratio of 2:1 may not be reported. Contamination of the S phase fraction with G1 phase cells will reduce all the S:G1 phase ratios. To verify the flow sorting purity, flow sorted fractions were passed back through the flow sorter and the degree of contamination was measured. This is shown in Figure 5.27B.

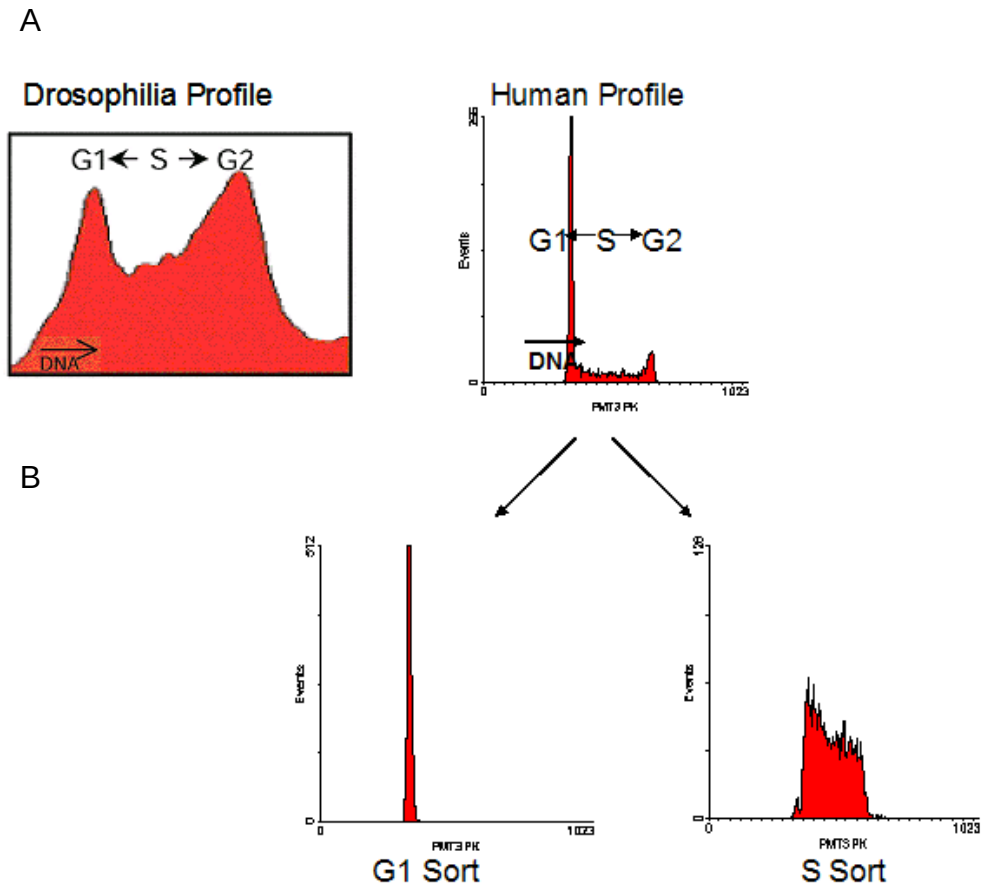


Figure 5.27: A: Comparison between *Drosophila* flow sort profile from (Schubeler, Scalzo et al. 2002) and the human lymphoblastoid flow sort profile obtained from sorting HRC575. B: Purity of the flow sort showing the G1 and S phase fractions.

Figure 5.27B shows that the flow sort is very pure, however there is a small amount of contamination and this will mask high S:G ratios of DNA which replicates very early in S phase and on the boundaries of G1 and S phase.

This contamination may also lead to inaccuracies in the normalisation of the array. The application of a curve fitting model to the cell cycle profile obtained in Figure 2.2 allowed extraction of best fit approximations of the distributions of G1 and S phase. The cell cycle analysis program Cylchred (Ormerod, Payne et al. 1987; Watson, Chambers et al. 1987) was used to estimate that within the S phase sorting window there are no contaminating G1 nuclei. However within the G1 phase sorting window there are approximately 2% of nuclei in very early S phase. In addition, approximately 4% of the earliest S phase nuclei are not represented within the S phase fraction. The consequence of this level of inaccuracy of sorting is that for a few very early replicating loci the theoretical replication timing maxima of 2.0 would be reduced to 1.93.

The purity of the flow sort is integral to the accuracy of the method and has proved a limitation when other tissues have been assayed. The evaluation of replication timing in other cell lines involves the culture of adherent cells. In these cell types I found that the Hoechst staining, vital for the sorting into G1 and S phase, was not uniform. A suspension of single nuclei was also more difficult to obtain. The nuclei tended to clump as they passed through the cell sorter, decreasing the purity of the sort due to minor disruptions of the flow. This will lead to increased contamination and therefore distorted ratios. A comparison between a lymphoblastoid and a fibroblastoid (adherent cell) flow sorter profile can be seen in Figure 5.28.

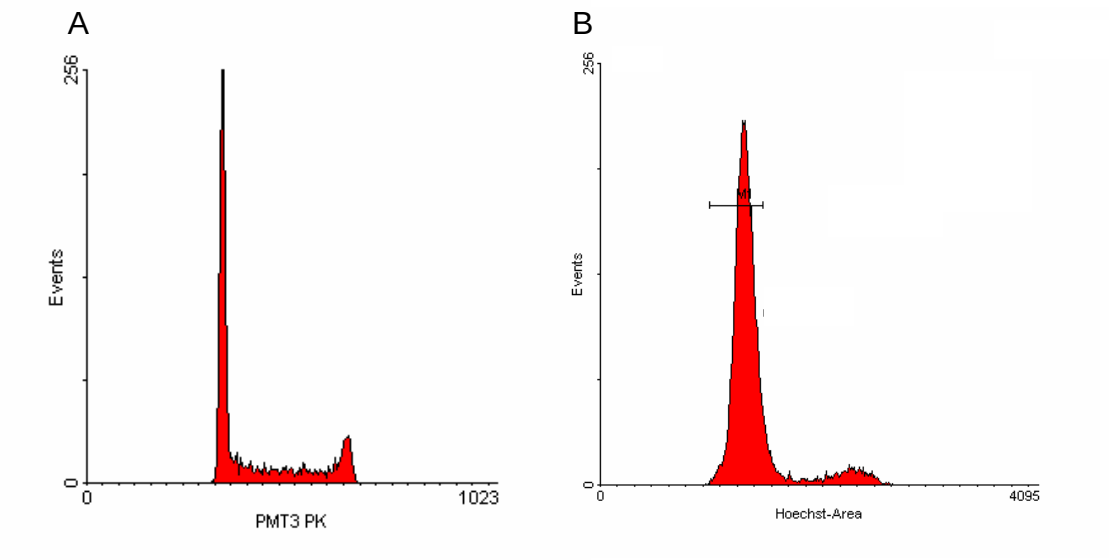


Figure 5.28: Comparison of the flow sort profiles of a lymphoblastoid cell line (A) and a fibroblastoid cell line (B).

Improving the discrimination between G1, S and G2 & M phase fractions obtained from other cell types will enable their replication timing to be assayed in the same way as achieved using the lymphoblastoid cell line. One way of achieving this would be to incorporate BrdU into nascent DNA. BrdU will therefore label cells in S phase only. Sorting nuclei using the combination of immunofluorescence against BrdU and propidium iodide staining for DNA content (Ormerod 2000) may be more effective at isolating nuclei in the S phase of the cell cycle for these cell types.

A further limitation of the method is the lack of heterochromatin and some other genome regions on the array. The clones on the array were selected from those used in the sequencing of the human genome (IHGSC 2001). Gaps in the draft sequence resulted in inevitable gaps in the replication timing profile of the chromosomes. Repeat regions of the genome are difficult to sequence leading to an under-representation of heterochromatic and centromeric DNA on the array. These sequences are late replicating (Gilbert 2002) and their lack of representation on the array will have an affect on the array normalisation. The array was normalised by the flow sort profile, which includes the representation of heterochromatin. An artefact of the normalisation process will therefore slightly bias measurements towards earlier replication.

The representation of repetitive heterochromatin sequences on the array in the appropriate amount would allow accurate normalisation of the array; however it would still be impossible to determine the exact replication timing of individual regions of highly repetitive heterochromatin using the current method. The effect of cross hybridisation of duplicated regions is detailed in 7.4.1. Repetitive heterochromatin represented on the array would cross hybridise with other repetitive sequences. The ratio reported by all repetitive regions would be an average replication timing of sequences that cross hybridise. The same limitations apply to regions of the genome with segmental duplications. Cross hybridisation results in the average replication timing of regions with similarity being reported. However it is unclear what effect the length and degree of homology will have on the replication timing ratio. As detailed in section 7.4.1, one way of circumnavigating this problem would be to use an array consisting of only unique sequence.

A related limitation of this method is the assay of regions where DNA replication is asynchronous, such as imprinted loci (Kawame, Gartler et al. 1995; Simon, Tenzen et al. 1999), the X chromosome in females (Avner and Heard 2001), immunoglobulin rearrangements and olfactory receptor genes (Goren and Cedar 2003). These are detailed in section 1.4.4. The flow sorting of the complete S phase ensures that both the early and late replicating alleles are hybridised to the same array. The ratio reported will be an average of the early and late replicating alleles and not the true replication time of either allele. To study the replication time of an imprinted region, cell lines with a uniparental disomy could possibly be used. S phase fractionation experiments such as those described in section 5.6 could also be used. However how regions involving immunoglobulin rearrangements and olfactory receptors would be assayed is unclear.

5.7.5: Verification of replication timing method

Replication timing has been previously assessed for a whole chromosome arm (Watanabe, Fujiyama et al. 2002), chromosome 11q, using flow sorting and real time PCR. This data was compared to the 1Mb resolution replication timing map of chromosome 11 obtained in section 5.5.1. The two replication timing profiles are very similar, with a correlation co-efficient of 0.69. This is despite the evaluation of

two different cell types and the total independence of the studies and methods used. This not only provides corroboration of our replication timing assay but also demonstrates the general similarity in the temporal programme of replication timing in these two different cell types.

Further verification was performed by selecting an early, two mid and a late replicating clone from the chromosome 22 replication timing profile described in section 5.3.1. and confirming the ratio of S:G1 by real time PCR. The two methods cannot be directly compared as the real time PCR can only assay a 150bp section of the genome. This is not comparable to the minimum 40Kb region sampled by the genomic arrays. A region of over 40Kb is likely to contain more than one replicon, whereas a region of just 150bp will not.

The standard deviation of the quadruplicates suggests the real time PCR is highly reproducible. The correlation between the replication time reported by arrays and that reported by real time PCR is strong ($r = 0.85$).

These two comparisons verify that genomic arrays can be used for the evaluation of replication timing. Unlike PCR based methods, the use of genomic arrays to assess replication ensures large regions of genome can be assessed at one time.

5.7.6: Assessment of replication timing using flow sorted S phase fractions.

Replication timing was assessed by detecting the increase in copy number within flow sorted fractions of S phase.

The majority of chromosome 22 loci were found to replicate in the first two fractions of S phase. Conversely most chromosome X loci replicated in the latter two fractions of S phase. This confirms what was found when the whole genome was analysed using complete S phase DNA hybridised against G1, with chromosome 22 replicating early and chromosome X replicating late.

In most cases the early replicating DNA increases copy number in the early S phase fractions. However, as shown in Figure 5.20 this did not reach the 2:1 ratio that would

signify that all replication was complete. 2:1 ratios were reached in the analysis of the G2/M fraction. This is not the case with late replicating fractions, which reached the 2:1 ratio in a linear fashion. It was observed that the late replicating clones exhibit a ratio of below 1:1 in the S1 and S2 fractions.

The scaling factors applied in Table 2.7 were calculated from the cell cycle profile. This means that the scaling factors applied are those that should be used for the whole genome. Application of an alternative normalisation factor, specific to chromosome 22 will increase the copy number change reported by the early replicating clones in the later S phase fractions to a ratio closer to 2:1. Contamination of the sorted S fractions by G1, G2 and other S DNA will also affect the ratios reported.

Loci reported by the single S:G1 phase hybridisation as early replicating display a copy number increase in the S1 and S2 fractions. This validates the single S:G1 hybridisation described in Chapter 5 for the assessment of replication timing.

Despite these limitations it is clear that genomic microarrays can be used to assess the replication timing of the majority of the genome. Verification by comparison with regions assayed by an alternative method shows that genomic arrays are highly effective at deducing replication timing over large regions of the genome.

5.7.7: Assessment of Replication Timing using High Resolution Arrays.

The construction of a high resolution array using 500bp PCR products is described in section 4.6. The data reported in section 3.4 from the test hybridisations displayed considerably more variation than that those reported for test hybridisations on other arrays. As described in section 4.6, the ratio reported by the PCR products from chromosome 22 reported more noise. It was therefore possible this noise in the ratios could be corrected. As a self:self hybridisation had been carried out, it was possible to divide the raw replication timing ratios by the self:self ratio reported for each sequence. The difference the inclusion of this analysis step makes is illustrated in Figure 5.29.

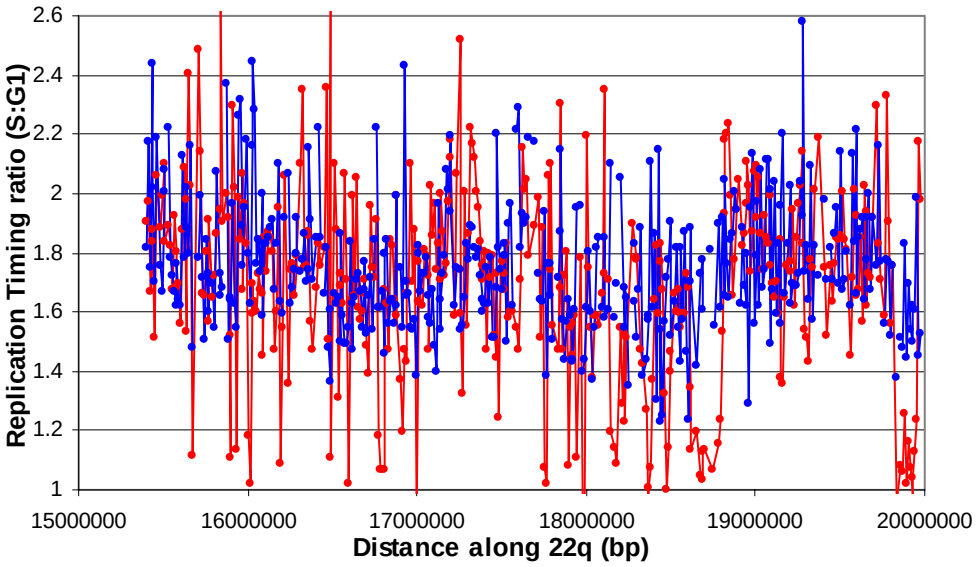


Figure 5.29: Correction against ratios reported for a self:self hybridisation on the high resolution array. Red: ratios reported prior to correction. Blue: ratios reported after correction as described.

Immediately it can be seen that the application of the additional correction factor tightens the ratios reported by the array from a spread of 0.84-2.75 and a standard deviation of 0.32, to a spread of 1.23-2.60 and a standard deviation of 0.22. However many points are still above the theoretical maximum of 2.0, reflecting inconsistencies with this method. Additional optimisation of the array, possibly by application of more accurate normalisation factors may increase the accuracy of the replication timing assay.

The analysis described in section 5.4.1 on the 10Kb resolution array shows that, in general the replication timing reported reflects that described by this region on the chromosome 22 tile path array. However this high resolution array does allow more detailed analysis. For example, a region 17.8-17.9Mb along chromosome 22 was reported as being early replicating by the 10Kb resolution array although the surrounding region was late replicating on the 22 tile path array. This 100Kb region is consistent with the size of one replicon, and so this result may identify a single early replicating replicon within a band of late replication. This indicates that although these arrays are currently not as accurate at reporting replication timing as the 22 tile

path array, they show the potential for detecting replication timing at a higher resolution.

The correlation between replication timing and GC content was also considered at high resolution. The GC content of each 500bp tile on the array was calculated and this was correlated with the replication time reported. The correlation at this level was weak with a correlation co-efficient of just 0.18, compared to a correlation coefficient of 0.27 when the same region is assayed using data from the tile path array. This shows a poorer correlation between replication timing and GC content when genomic DNA is assayed at this resolution.

The 500bp tile path resolution array showed increased variation compared to the tiling path arrays, although no ratios reported are above 2.2 or below 1. Additional peaks and troughs in the replication timing ratios were observed that were not apparent at the 10Kb resolution.

However, it can be argued that such marked peaks and troughs should not be seen at this high resolution. The array is made with 500bp overlapping PCR products. There are 275 PCR products on the array, covering 200Kb of sequence. As replicons are 40-100Kb in length the high resolution array is therefore thought to represent 2-5 replicons. The replication timing profile across the region does not reflect this. No groups of loci belonging to the same replicon, or replicon boundaries were identified. This could be due to the high coefficient of variation the consequence of which was that ratios within 0.4 of each other were not statistically different. This is very different to the 0.1 variation expected from the 22 tile path array, and would make it impossible to define replicons and replicon boundaries.

The high standard deviation of the method also makes it unfeasible to map replication origins using this method. DNA replicates at a rate of 50 base pairs per second. This means a replicon of 100Kb would take 33 minutes to replicate. A smaller 40Kb replicon would take 13 minutes to replicate. In the lymphoblastoid cell line assayed, S phase takes eight hours to complete. The tile path arrays have a standard error of 0.1 of S phase. This equates to 48 minutes. As this is greater than the time taken for a

replicon to replicate, variations within the replicon (such as the early replication of the origin) will not be identified.

Unfortunately time limitations did not allow full optimisation of the PCR product array. Despite these current limitations in assessing replication timing on the PCR product array it is clear from the work described in this Chapter that genomic clone arrays are highly effective at assessing replication timing and assaying large regions of the genome with a high accuracy.

5.7.8: Summary:

This Chapter has reported how microarrays have been used to assess human replication timing for the first time. Replication timing has been assayed on the whole genome at a 1Mb resolution, Chromosomes 1, 6, and 22 at a tile path resolution and a small region of chromosome 22 using 500bp PCR products.

Replication timing was then correlated with sequence features of the genome. Correlations with GC content, gene density and common repeat elements were observed.

The method was verified by comparison of the replication timing data produced for 11q with previously published data (Watanabe, Fujiyama et al. 2002). The replication timing of a selection of clones from chromosome 22 was assessed by real-time PCR. The replication timing of these clones was also found to correlate with the array method.

6. Results 4
Correlation between Replication Timing and Non-sequence Features of the Genome.

6.1 Introduction

Chapter five describes how microarrays can be used to assess replication timing. A correlation was observed between replication timing and sequence features of the genome. As described in section 1.3, replication timing has also been correlated with structural features of chromatin. This Chapter investigates the relationship between replication timing, transcriptional activity and histone modification of the genome. I also investigate how a change in chromatin, by a chromosomal translocation affects replication timing.

There is currently some controversy over whether there is a correlation between the time of replication and the transcriptional activity of regions of the genome. Experiments in yeast have shown no relationship between replication timing and transcriptional activity (Raghuraman, Winzeler et al. 2001), whilst analysis of the *Drosophila melanogaster* genome demonstrated no relationship with level of gene expression, but suggested a correlation with the probability of gene expression with genes that are transcribed being located within early replicating DNA (Schubeler, Scalzo et al. 2002). To investigate this link in the human genome, the expression of genes within a lymphoblastoid cell line was assayed and correlated with replication timing. This is described in section 6.2

Acetylation of histones within the nucleosomes has also been linked to transcriptionally active, early replicating regions of the chromosome (Grunstein 1997; Eberharter and Becker 2002; Vogelauer, Rubbi et al. 2002; Grewal and Moazed 2003). To investigate this correlation, DNA that had been immunoprecipitated with either of two antibodies, one anti-acetyl-Histone H3 (Upstate, USA) and a second anti-acetyl-Histone H4, ChIP grade for histone H4 acetylation (Upstate, USA), was applied to the array. Both antibodies are polyclonal and produced in rabbits. The Histone H3 recognises and is specific for acetylated human H3 of approx. 17kDa. The

Histone H4 antibody recognises acetylated histone proteins of approx 10kDa but is known to cross react with acetylated histone H2B and may cross react with other acetylated proteins.

The study was achieved in collaboration with the Microarrays, Transcriptional Regulation and Human Disease Group at the Sanger Institute. The chromatin immunoprecipitation was performed by Pawendeep Dharmi while I labelled and hybridised the DNA to the array. The results are reported in section 6.3.

Section 6.4 investigates chromosomal breakpoints and replication timing. Section 6.4.1 explores the effect a translocation between chromosomes 17 and 22 has on the replication timing of chromosome 22. The breakpoints had already been identified in our laboratory using an array painting technique (Fiegler, Gribble et al. 2003). This, and further resolution of the breakpoint by FISH showed that on chromosome 22 the breakpoint is within the clone bA46E17 (midpoint 11546117). Schleiermacher et *al* have shown that chromosomal breakpoints map to early replicating regions of the genome (Schleiermacher, Janoueix-Lerosey et al. 2003). This hypothesis is tested by mapping the position previously described breakpoints onto a normal replication timing profile (5.3.1) and is described in section 6.4.2.

6.2: Correlation between Replication Timing and Transcriptional Activity.

6.2.1 Correlation with Expression level on the 1 Mb Chip.

RNA was prepared from a cycling lymphoblastoid cell line (HRC 575) as described in section 2.6.1. The RNA was run on a 1% agarose gel to verify it was not degraded and is shown in Figure 6.1. The RNA was then labelled with Texas Red and applied to an Affymetrix U133A array in collaboration with Silvana Debenardi at the Molecular Oncology Unit, St Bartholomew's hospital.

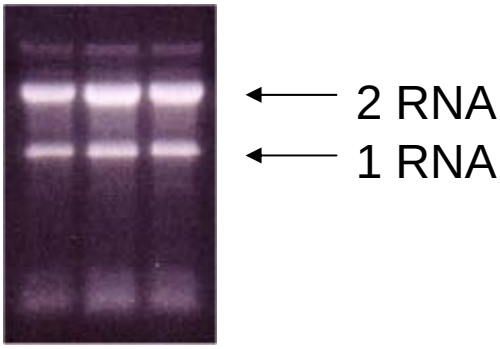


Figure 6.1: RNA prepared from a lymphoblastoid cell line.

The Affymetrix array contains oligonucleotides from approximately 13,000 human genes. Each gene loci is present as a pair of oligonucleotides, one contains the true gene sequence whilst the second oligonucleotide contains a mismatch. The Texas red ratio given by the sequence oligonucleotide is ratioed against the fluorescence obtained on the mismatched oligonucleotide. After hybridisation and sacanning the Affymetrix analysis program produces two values for each locus. Firstly, an expression level is given, this is obtained from the intensity of the fluorescence on the sequence oligonucleotide, secondly a ‘present’ or ‘absent’ call is given, a present call is obtained from the when the ratio of the intensities obtained from sequence: mismatch is above a set threshold.

The average expression level of each clone was correlated with the replication timing ratio for the 1Mb resolution array. Previous work by Bryan Young (Molecular Oncology, St Bartholemews hospital) had mapped Affymetrix data points within the clones present on the 1Mb array. Of the 3126 clones present on the 1Mb array only 1089 contained genes that were represented on the U133A Affymetrix chip. The replication timing ratio of each clone was plotted against the $\log_{10}$ of the average expression level of the clone (Figure 6.2).

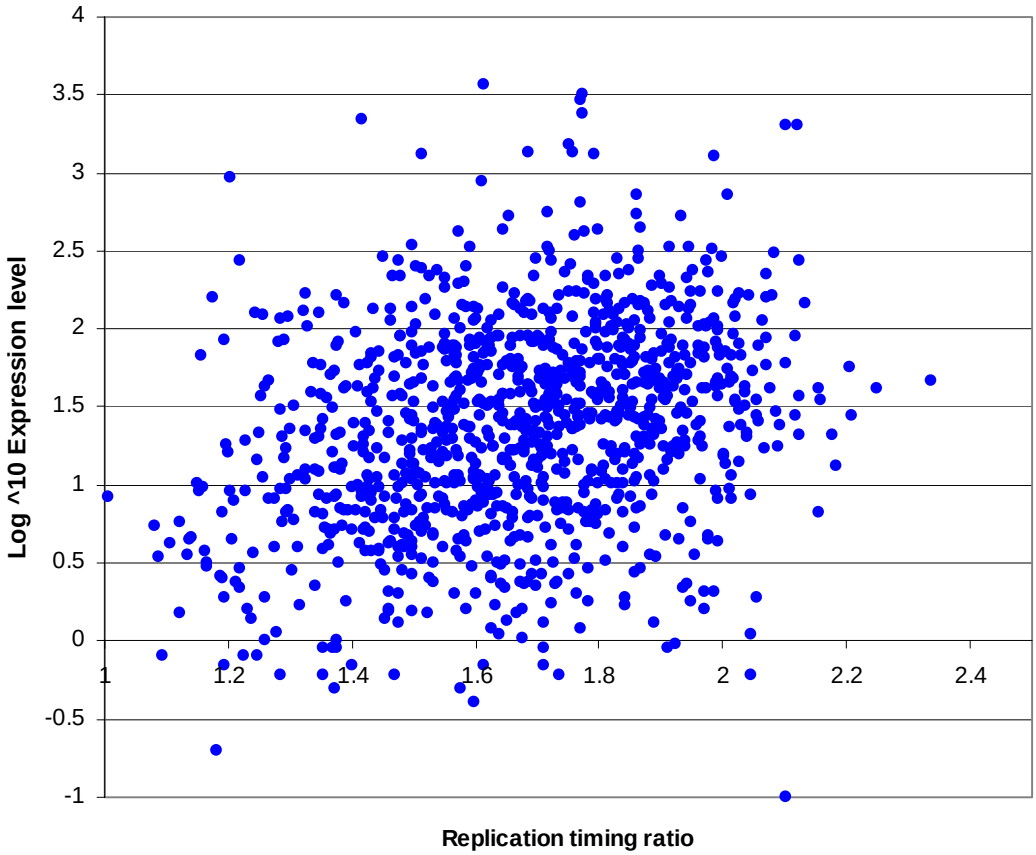


Figure 6.2: The correlation between replication timing ratio and $\log_{10}$ expression level of clones on the 1Mb array. $y = 0.057 + 0.86x$. $r = 0.30$.

These results show a weak correlation between the replication timing of a clone and the transcriptional activity of the genes within that clone. Hence early replicating clones are slightly more likely to be expressed at a higher level than those replicating later in S phase.

The expression level of the clone can also be plotted together with replication timing profiles reported in Chapter 5. An example of this is for chromosome 2 is shown in Figure 6.3. Other chromosome profiles are included in Appendix 8.

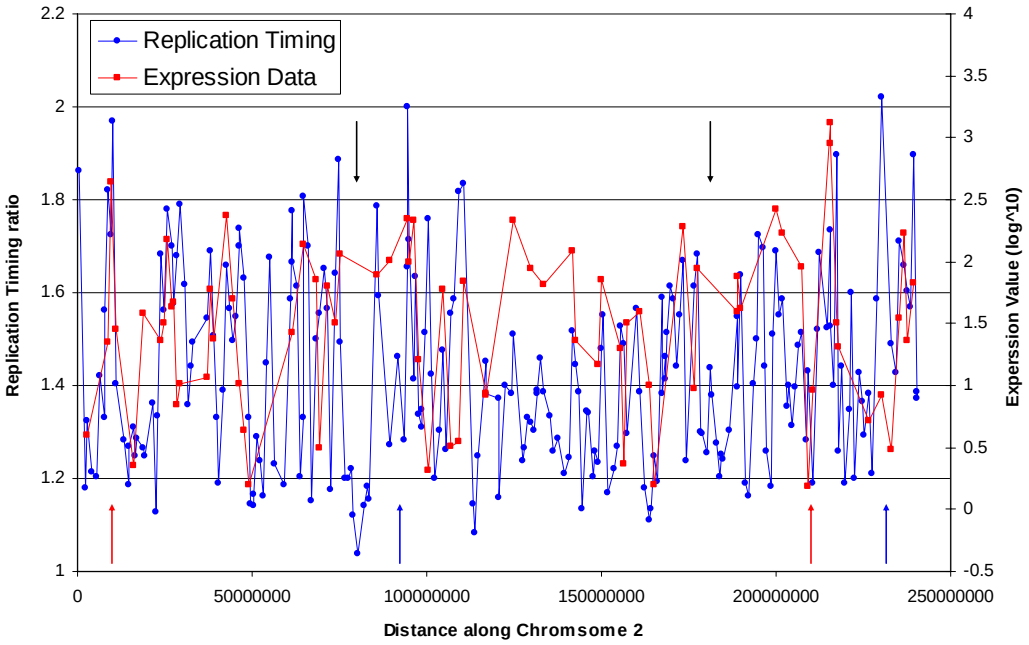


Figure 6.3: Replication timing and expression level profiles on Chromosome 2. Red arrows: regions where replication timing and transcriptional activity appear to correlate, Blue arrows: regions where replication timing and transcriptional activity do not correlate. Black arrow: regions that are late replicating, but are not represented on the U133A Affymetrix array.

The replication timing and expression level profiles of chromosome 2 show some regions where the two features appear to correlate (8.5-16Mb and 209-217Mb along chromosome 2), and regions where the two features are disparate (85-95Mb and 227-236Mb along chromosome 2).

6.2.2 Correlation with Expression level on the Tile path arrays.

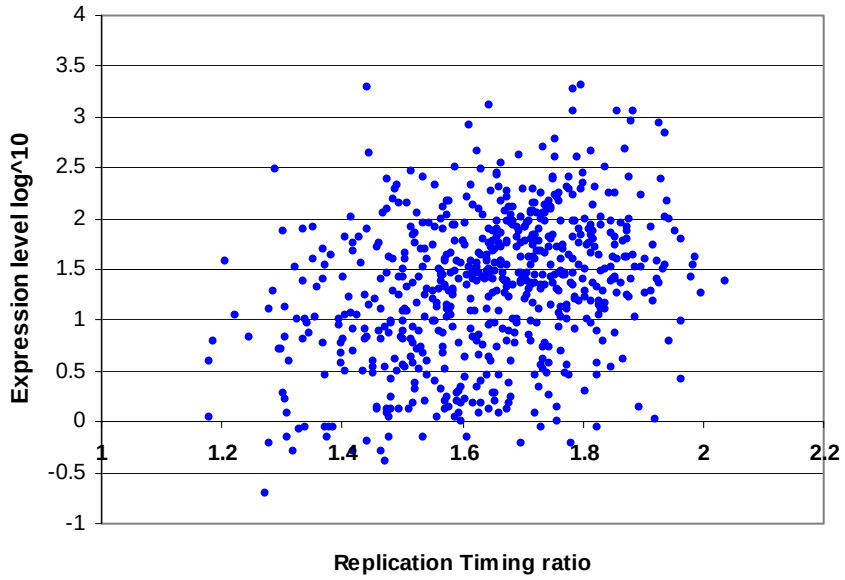
Analysis of replication timing and gene expression level was performed on tile path arrays for chromosomes 1, 6 and 22. This allowed the correlation between replication and transcription on a small, a medium and a large sized chromosome.

On chromosome 1, 647 of the 1961 tile path clones contained Affymetrix U133A data points. On chromosome 6, 430 of the 1651 tile path clones contained Affymetrix

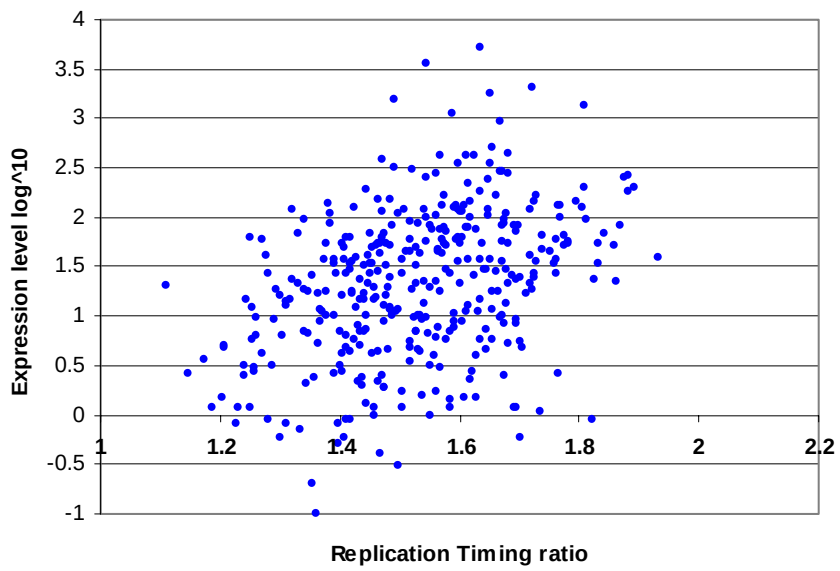
U133A data points and on chromosome 22, 147 of the 444 tile path clones contained Affymetrix U133A data points.

The correlation between replication timing and transcriptional activity was plotted for each chromosome as shown in Figure 6.4, The statistics for the linear regression line of each graph is shown in Table 6.1.

A



B



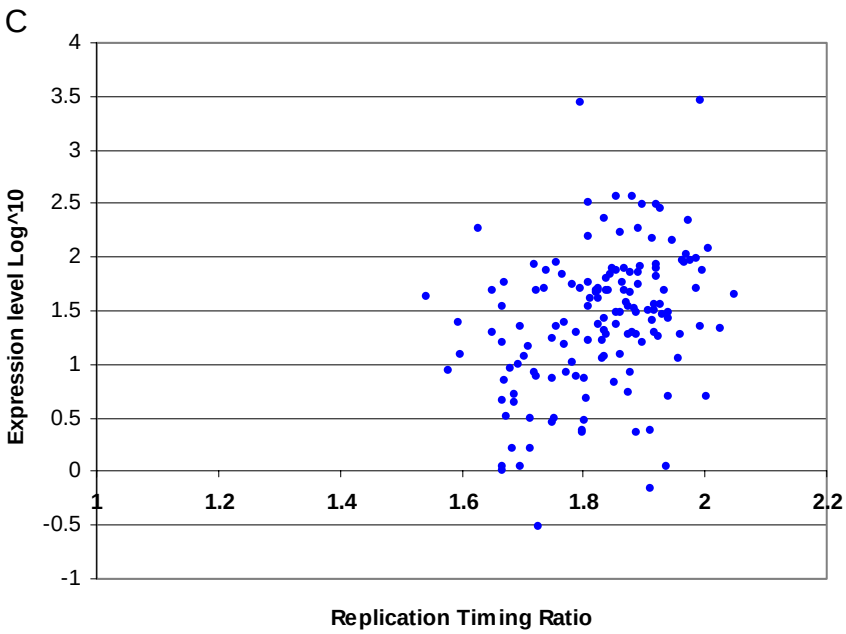


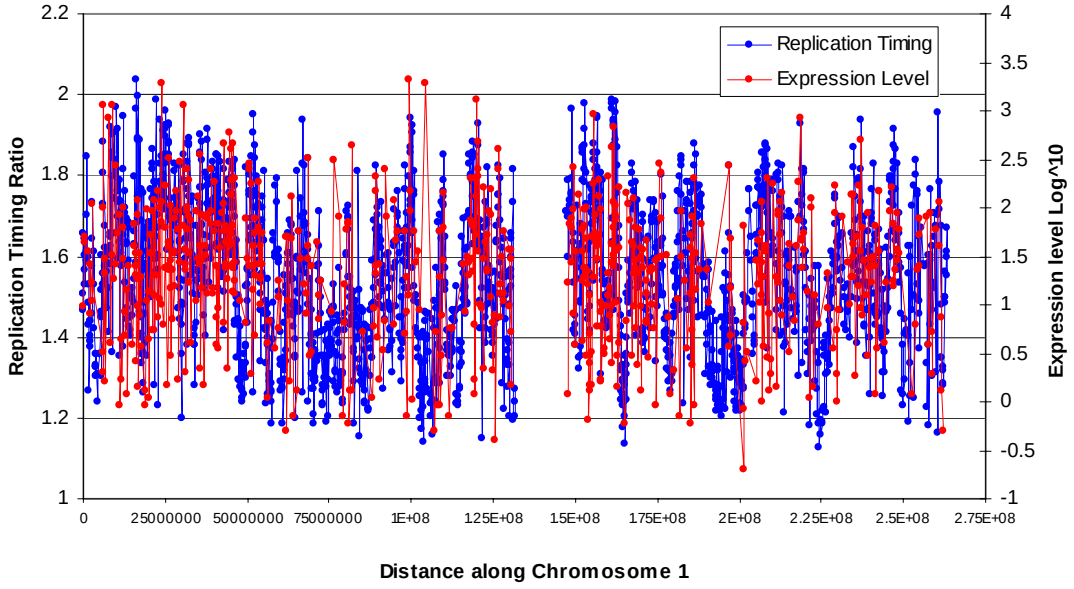
Figure 6.4: Replication timing ratio plotted against expression level (Log_{10}) for; A: Chromosome 1, B: Chromosome 6 and C: Chromosome 22.

Table 6.1: Regression features of Replication timing versus expression levels at tile path resolution.

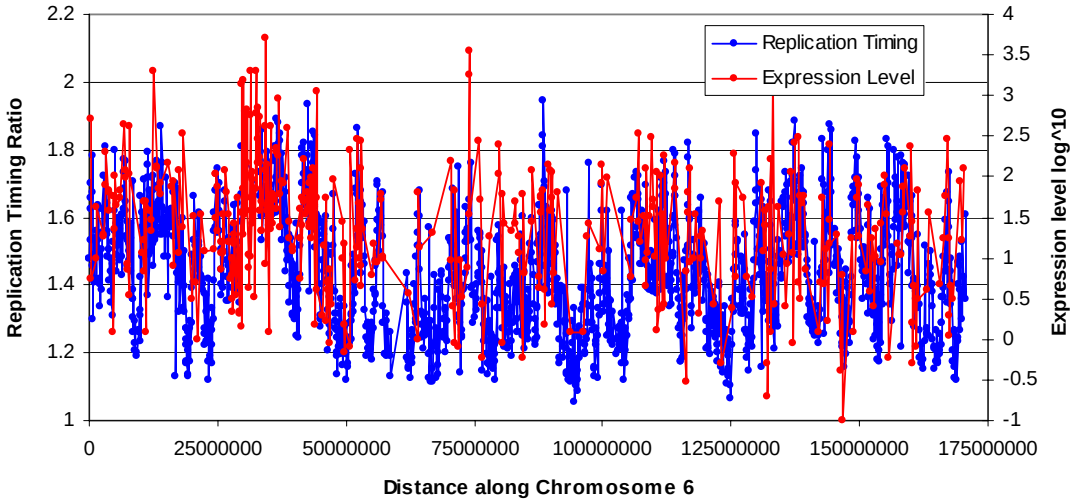
Chromosome	**Intercept**	**Regression coefficient**	**Correlation coefficient**
1	-1.02	1.43	0.33
6	-1.02	1.84	0.38
22	- 2.54	2.16	0.34

As before, the replication timing and expression level of the chromosome was plotted against chromosome position, for the three chromosomes examined at tile path resolution. This is shown in Figure 6.5.

A


Replication Timing
Expression Level
Replication Timing Ratio
Expression level Log^10
2.2
2
1.8
1.6
1.4
1.2
1
4
3.5
3
2.5
2
1.5
1
0.5
0
-0.5
-1
0
25000000
50000000
75000000
1E+08
125E+08
15E+08
175E+08
2E+08
2.25E+08
2.5E+08
2.75E+08
Distance along Chromosome 1

B


Replication Timing
Expression Level
Replication Timing Ratio
Expression level log^10
2.2
2
1.8
1.6
1.4
1.2
1
4
3.5
3
2.5
2
1.5
1
0.5
0
-0.5
-1
0
25000000
50000000
75000000
100000000
125000000
150000000
175000000
Distance along Chromosome 6

C

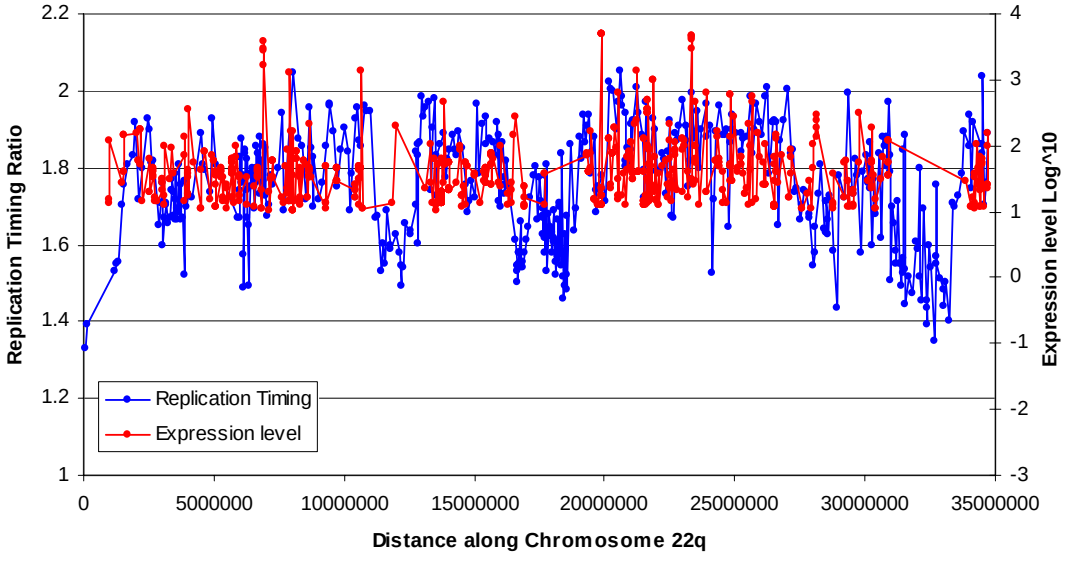

Replication Timing
Expression level
Replication Timing Ratio
Expression level Log^10
2.2
2
1.8
1.6
1.4
1.2
1
4
3
2
1
0
-1
-2
-3
0
5000000
10000000
15000000
20000000
25000000
30000000
35000000
Distance along Chromosome 22q

Figure 6.5: (Previous page) Replication timing ratio (blue) and Expression level (Log_{10} -red) plotted against chromosome position for; A: Chromosome 1, B: Chromosome 6 and C: Chromosome 22.

As with the same plots at a 1Mb resolution, there were regions where the replication timing and expression level appeared to correlate, and regions where they were disparate. Also, it can be seen that regions with a low density of Affymetrix points correlate with late replicating regions of the chromosomes studied. This coincides with the report in section 5.7.1, that gene poor regions are late replicating.

The correaltion coefficients for the three chromosomes studied at tile path resolution showed a weak positive correlation between early replication and transcriptional activity. The correlation coefficients of the three chromosomes studied and for the whole genome analysis on the 1Mb chip are comparable.

6.2.3: Correlation between Replication Timing and the Probability of Expression.

To ascertain whether there is a correlation between replication timing and probability of expression (an absolute measure of whether a clone contains genes that are expressed or not), an analysis was performed similar to that performed by Schubeler *et al* (Schubeler, Scalzo et al. 2002). Clones were ranked by replication timing and assembled into groups of 50 (1Mb array) or 25 (tile path array). Clones were then given a value of 1 if the clone contained a gene that was expressed or 0 if no genes within the clone were expressed. The average probability of expression was then calculated for each 50 or 25 clone window.

Firstly, of the approx 13,000 genes represented on the U133A array 2,063 genes were also represented within genomic clones on the 1Mb array. Of these, 1,013 were scored as “present” (expressed) while the remaining 1,050 genes were scored as “absent” (not expressed). The grouping of clones into windows of 50 and the analysis of the probability of expression showed a strong positive correlation between early replication and a high probability of expression. Logistic regression was performed on the data, (logistic regression is appropriate where the first variable can be any value, but the second variable is a choice between two discrete values, in this case either

presence or absence of expression),the correlation co-efficient was found to be 0.62 (see Figure 6.6).

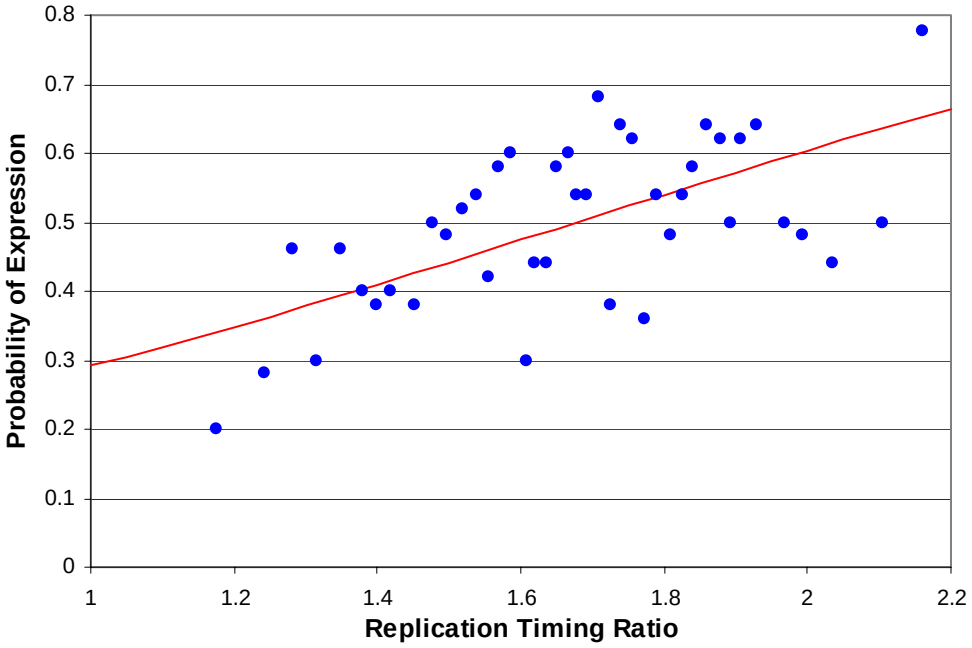
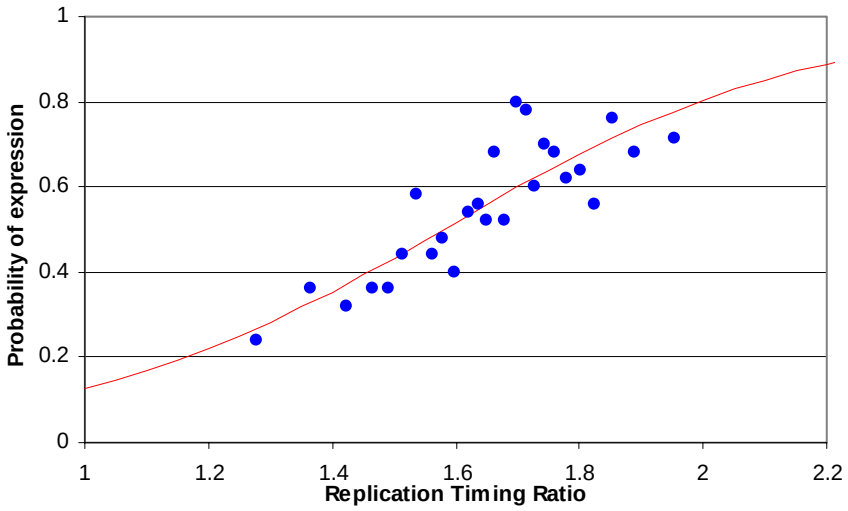


Figure 6.6: Correlation between replication timing and probability of expression for windows of 50 clones. The red line shows the logistic regression given by the equation $y = e^{-2.13 + 1.31x}/(1+ e^{-2.13 + 1.31x})$.

A similar analysis was performed on the individual tile path arrays, and stronger positive correlations were seen when logistic regression analysis was performed (Figure 6.7 and Table 6.2)

A



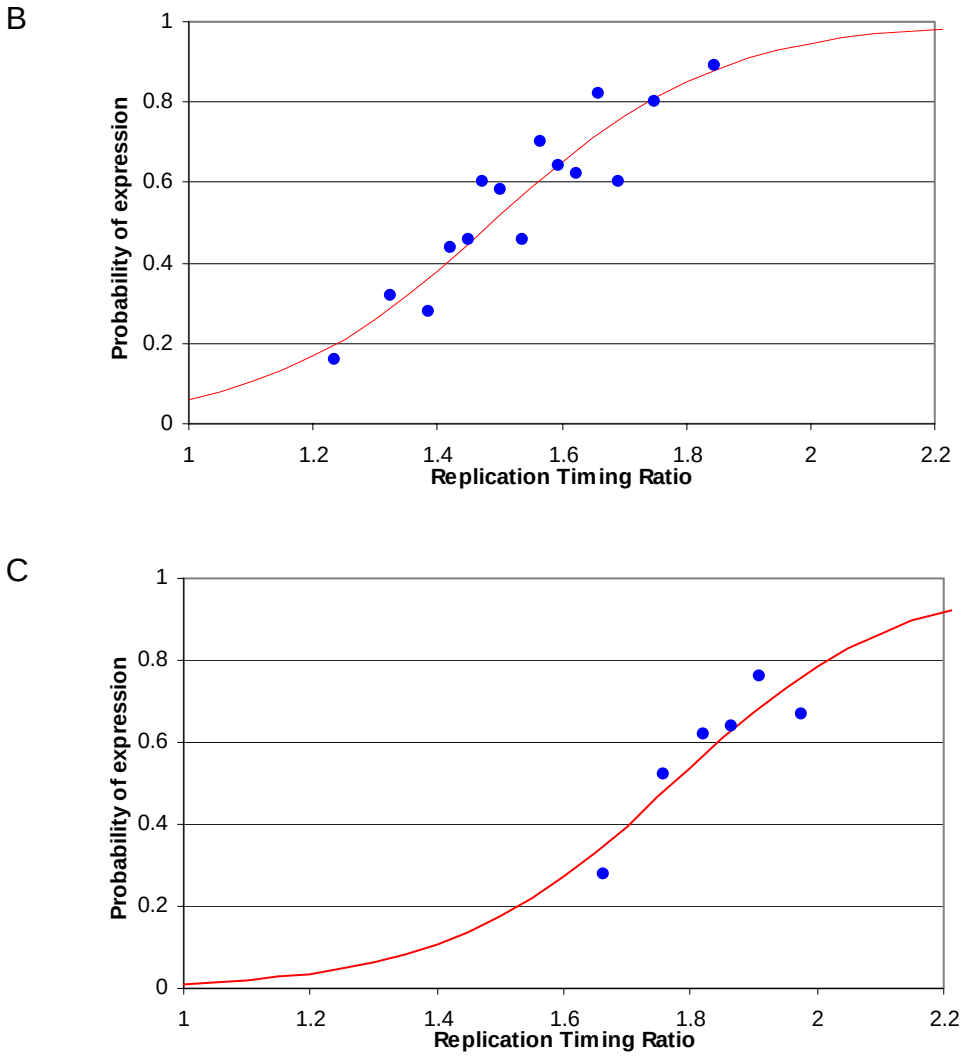


Figure 6.7: Correlation between replication and probability of expression at a tile path resolution. The red line shows the line of logistic regression and the equation is indicated in Table 5.3. A: Chromosome 1, B: Chromosome 6, C: Chromosome 22.

Table 6.2: features of the logistic regression performed, correlating replication timing ratio with probability of gene transcription. The equation for the regression line is $y = e^{a + bx}/(1+ e^{a + bx})$.

Array	Intercept (a)	Regression co-efficient (b)	Correlation co-efficient
1Mb Genome wide	-2.13	1.31	0.62
1	-5.29	3.35	0.83
6	-8.28	5.56	0.93
22	-10.13	5.71	0.89

This analysis shows that there is a strong correlation between replication timing and the probability of expression. This correlation with the probability of expression is stronger than the correlation with level of expression. The correlation also improves when an individual chromosome is studied, opposed to the entire genome.

6.3. Correlation between Histone Acetylation, Replication Timing and Sequence Features on the Chromosome 22 Tile Path Array.

To correlate replication timing with histone acetylation, separate chromatin immunoprecipitations were performed using an antibody binding to acetylated Histone H3 and a second antibody binding to acetylated Histone H4. The immunoprecipitated DNA was labelled in one colour and the same DNA that was used as the input to the immunoprecipitation was labelled in a second colour. The DNA was co-hybridised to the chromosome 22 tile path array. The ratio of chromatin immunoprecipitated DNA: input DNA was then plotted against chromosome position (Figure 6.8).

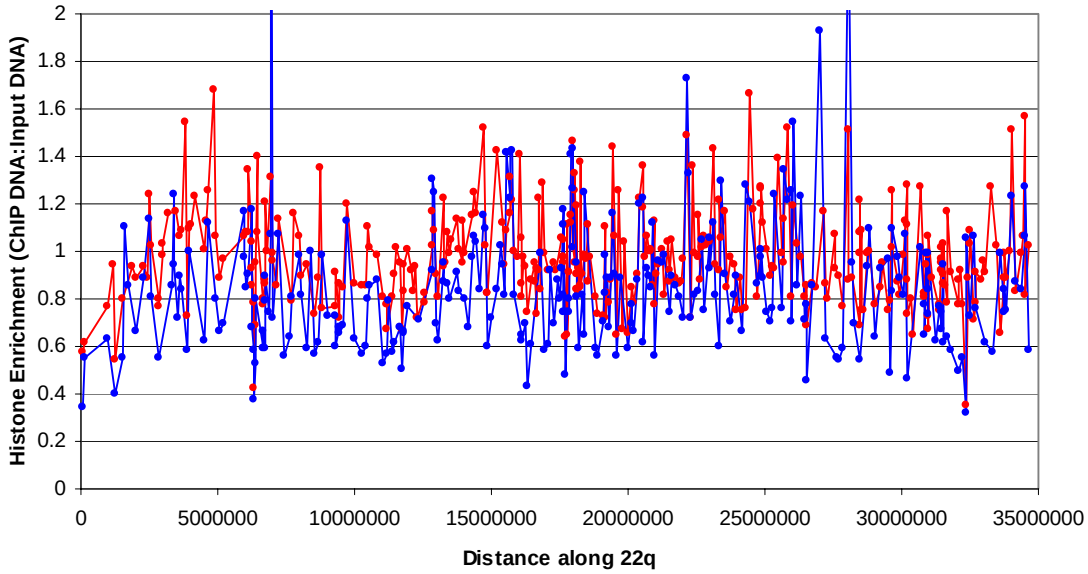


Figure 6.8: The ratio of Chromatin immunoprecipitated DNA:Input DNA plotted against position on chromosome 22. Red: H3 enrichment, Blue: H4 enrichment.

The relationship between histone H3 and histone H4 enrichment with acetylation was then plotted on Figure 6.9.

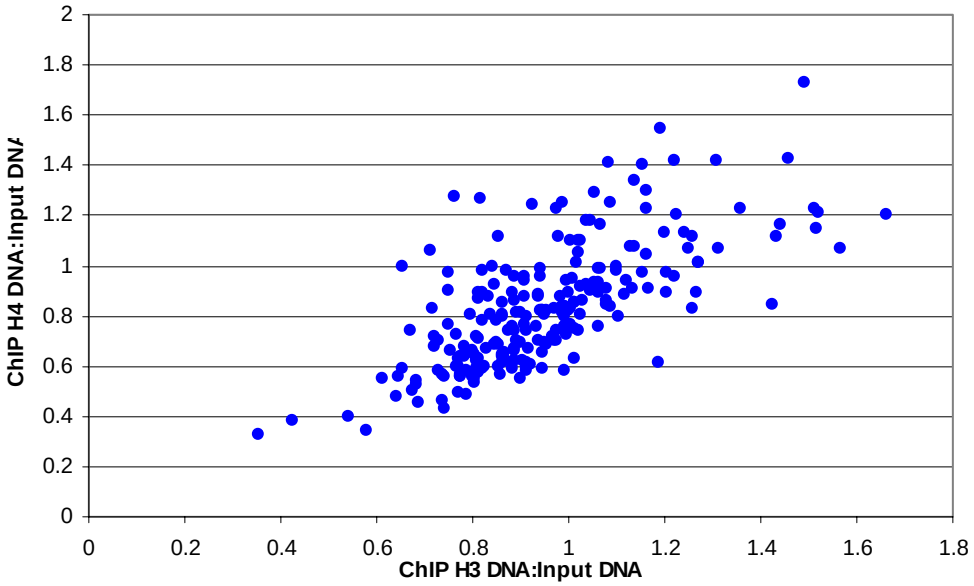
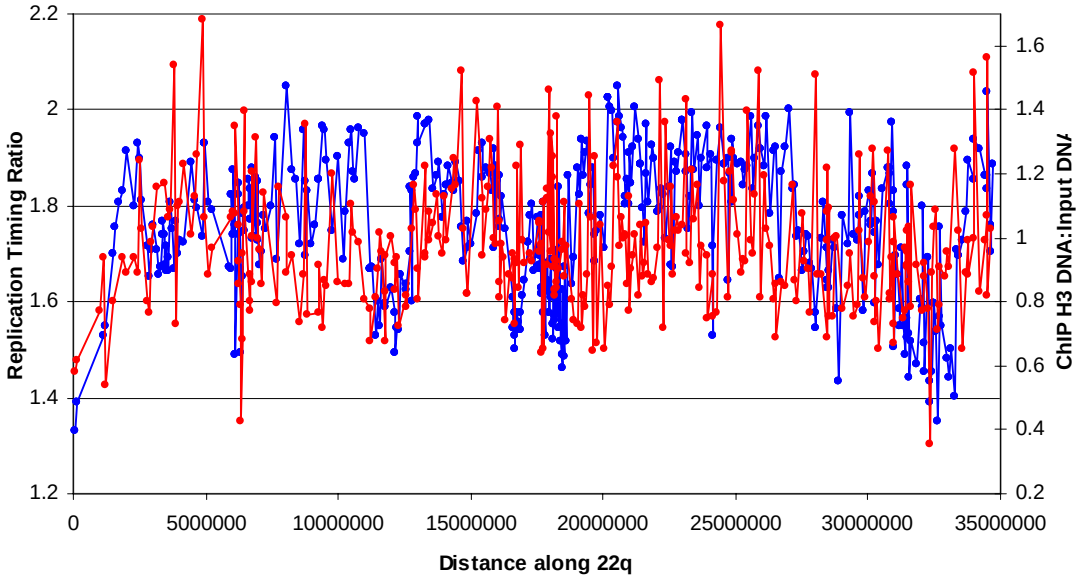


Figure 6.9: Relationship between H3 enriched DNA and H4 enriched DNA. Linear regression was performed. y = 0.50 + 0.54x and r=0.70.

There was a strong positive correlation between the distribution of acetylated histone H3 and acetylated histone H4.

The histone acetylation status was compared with the replication timing ratio of each clone (Figure 6.10).

A



B

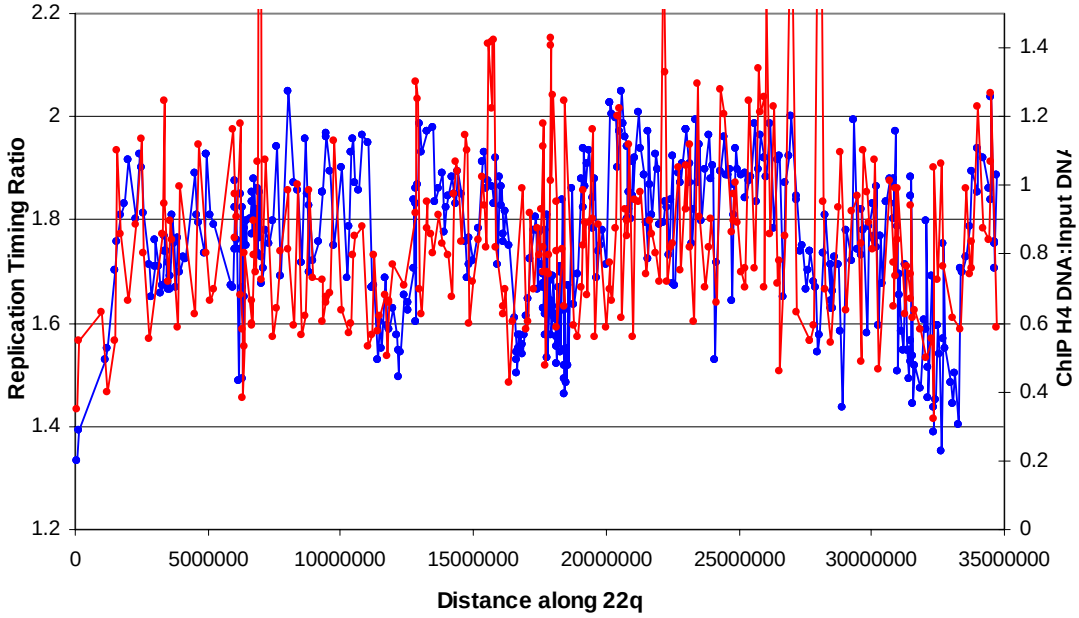


Figure 6.10: The ratio of immunoprecipitated DNA: Input DNA plotted with replication timing ratio, against position on chromosome 22. Red: Histone enrichment, Blue: Replication timing. A: Histone H3 enrichment, B: Histone H4 enrichment.

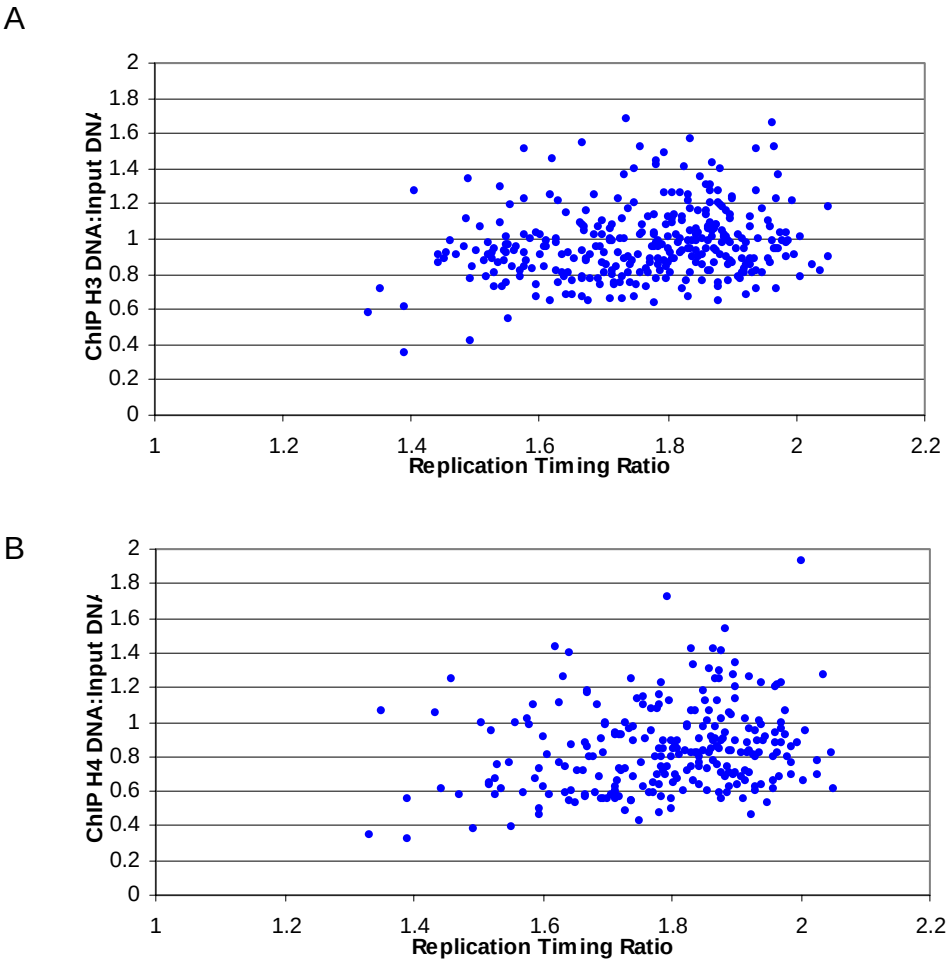


Figure 6.11: Correlation between replication timing and histone acetylation. A: Correlation with H3 acetylation, B: Correlation with H4 acetylation.

Linear regression was performed and the correlation between histone acetylation and replication timing is shown in Table 6.3.

Table 6.3: Linear regression statistics performed when replication timing was plotted versus Histone acetylation enrichment levels.

Histone Acetylation	**Intercept**	**Regression coefficient**	**Correlation co-efficient**
H3	0.48	0.28	0.21
H4	0.25	0.34	0.16

Linear regression was also performed to correlate histone acetylation with other genome parameters.

Table 6.4: Linear regression statistics performed when genome features were plotted versus histone enrichment levels. For the probability of expression statistics clones were clustered in groups of 20 and analysed as described in 6.2.3.

Histone	Feature	No. of observations	Intercept	Regression coefficient	Correlation coefficient
H3	**GC content (%)**	327	0.755	0.454	0.117
	Gene density	223	0.930	0.001	0.131
	***Alu* content**	223	0.878	0.004	0.246
	LINE content	223	1.022	-0.004	0.166
	Expression level	115	1.026	$4.53x10^{-5}$	0.074
	Prob. of Expression	-	0.872	0.308	0.663
H4	**GC content (%)**	256	0.488	0.776	0.125
	Gene density	166	0.793	0.002	0.204
	***Alu* content**	166	0.683	0.010	0.354
	LINE content	166	1.017	-0.011	0.256
	Expression level	103	0.879	0.0002	0.067
	Prob. of Expression	-	0.791	0.277	0.544

These statistics illustrate very weak correlations between levels of Histone H3 and H4 acetylation and other genome features on chromosome 22. However a stronger correlation can be seen between histone actylation and probability of transcription.

6.4: Study of the Replication Timing of Chromosomal Breakpoints using the Genomic Arrays

6.4.1: Assessment of the replication timing of a t(17q21.1:22q12.2) translocation on the chromosome 22 array.

A replication timing profile was generated for a lymphoblastoid cell line with a translocation between chromosomes 17 and 22 to investigate if the translocation affected the replication timing profile. The location of the chromosomal breakpoints had already been mapped by other members of the laboratory (Fiegler, Gribble et al. 2003), and the breakpoint on chromosome 22 had been located at approximately 11.5Mb along the q arm within clone bA46E17. The replication timing profile of

chromosome 22 was assessed on the tile path array. The average standard error of the points on the three replicates was 5.9%. The replication timing profile is shown in Figure 6.12.

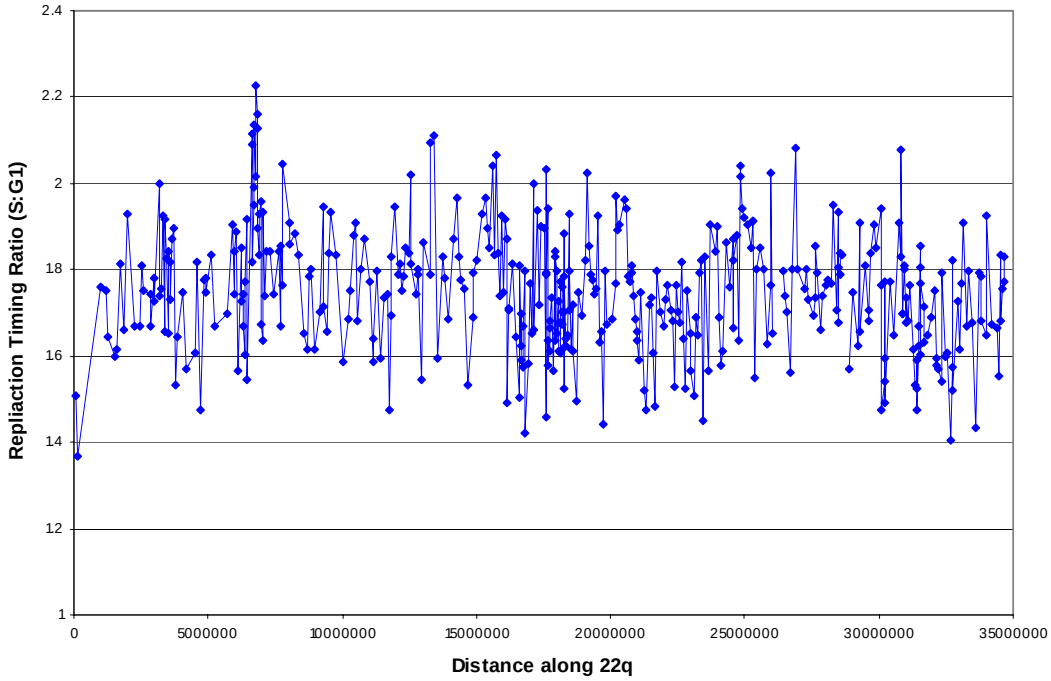


Figure 6.12: Replication timing profile of the 22 clones on the translocated cell line

The chromosome 22 replication profile for a normal lymphoblastoid cell line is described in section 5.3.1.

Comparison of the two replication timing profiles, illustrated in Figure 6.13 for the translocation and a normal cell line showed three differences. Firstly, six regions can be seen as having replication timings clearly different from the normal replication profile. The first of these is close to the breakpoint, about 430Kb (5 clones) downstream. Secondly, the data obtained from the translocated cell line displayed more local variation, with more points closer to 1:1 and 2:1, than that obtained from the normal cell line and thirdly, there are more data points over the hypothetical maximum ratio of 2:1, this is particularly noticeable at the VJ recombination region.

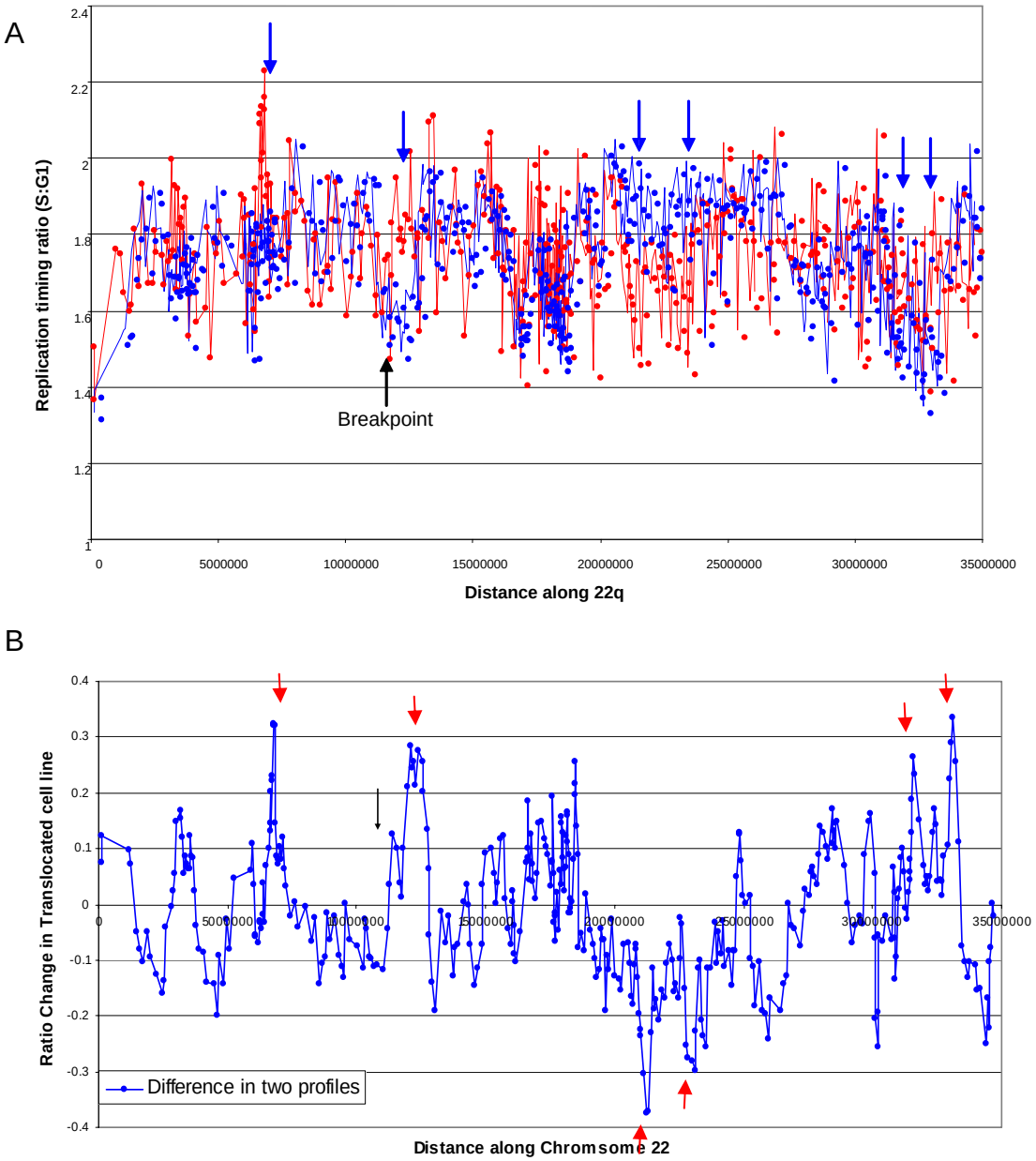


Figure 6.13: A: Comparison of the replication timing profile from a normal lymphoblastoid cell line and a lymphoblastoid cell line with a translocation between chromosomes 17 and 22. Blue: Replication profile from a normal cell line. Red: Replication profile from a translocated cell line. The breakpoint on chromosome 22 is indicated by the black arrow. Regions where there are significant changes in replication timing between the two regions are highlighted with blue arrows. B: The difference in replication timing between the translocated cell line and the normal cell line. The breakpoint is shown by the black arrow. Regions with 3 or more clones showing a greater than 0.2 difference (5 x standard deviation of a self:self

hybridisation on the 22 array) are shown by arrows in red. Positive values indicate that the time of replication was earlier in the translocated cell line. Negative values indicate replication occurred earlier in the normal cell line.

Linear regression between the replication timing profile of the normal and the translocated cell line gave a correlation coefficient of 0.32. This is much less than the correlation coefficient between two normal cell lines of 0.77 and reflects the regions of difference between the normal replication timing profile and that reported by the translocated cell line.

6.4.2 Assessment of the replication timing of constitutional breakpoints using the 1Mb array.

It has been suggested that regions of DNA that undergo chromosomal translocations are early replicating prior to translocation (Schleiermacher, Janoueix-Lerosey et al. 2003). The position of constitutional breakpoints that had already been identified at high resolution by other members of the laboratory were mapped onto a normal replication timing profile (Appendix 12). In total, nine chromosomal translocations were examined in this way. These are summarised in Table 6.5

Table 6.5: Replication timings of chromosomal breakpoints on a normal cell line.

Translocation	Derivative 1 Clone(s)	Av Replication Time of Clone(s)	Derivative 2 Clone(s)	Av Replication Time of Clone(s)
t(17:22) (q21.1-12.2)	dJ112G21 bA94L15	1.81	bK445C9 dJ353E16	1.65
t(2:7)a (q37.2-36.3)	bA263G22 (Spans)	1.49	bA269M19 dJ982E9	1.52
t(3:11) (q21-q12)	bA221E20 bA129J11	1.61	bA147G6 bA227P3	1.57
t(2:5) (q31.1-q23.2)	bA551O2 bA218M2	1.27	bA48C14 (Spans)	1.51
t(7:13) (q31.3-q21.3)	bA384A20 bA106F1	1.20	bA184L18 bA309H15	1.24
t(2:7)b	bA288C18 (Spans)	1.54	bA502P9 dJ855F16	1.18
t(2:7)c	bA84G18 bA260J21	1.57	dJ1143H19 bA175P8	1.48
t(1:6) (p22.1-q16.2)	dJ1043L3 bA427B20	1.45	bA117M4 dJ167P23	1.24

It can be seen that for five of the eight translocations the two breakpoints show a replication timing ratio within 0.1 of each other in a normal cell line. A further three have a replication timing ratio within 0.25 of each other. This correlation is seen in Figure 6.14.

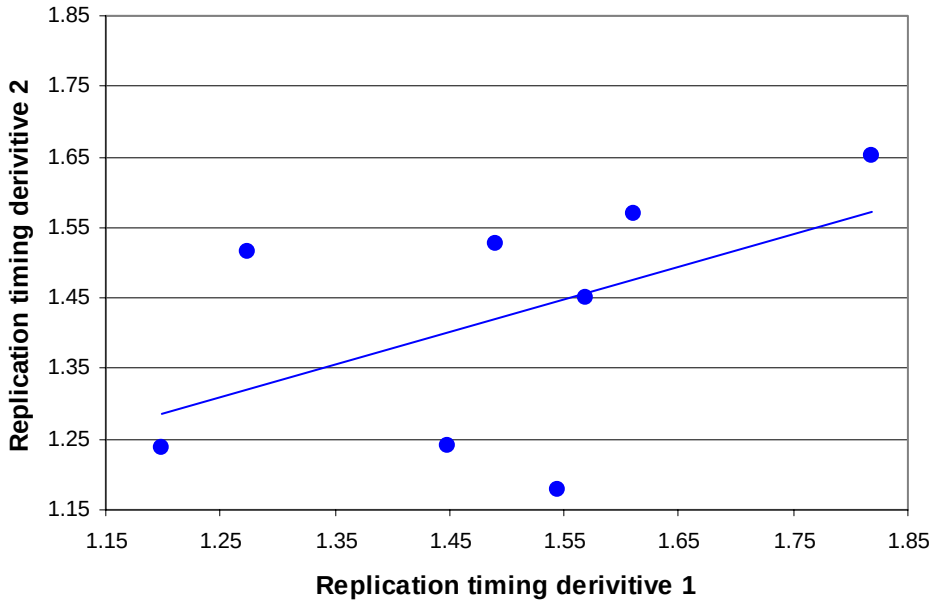


Figure 6.14: The correlation between the replication timing of the first and second breakpoint on a congenital translocation.

A positive correlation is seen between the replication timing of the two loci involved in the translocation with a correlation coefficient of 0.51. This suggests that replication timing of the two regions involved in a translocation would have to be similar for a translocation to occur.

6.5.: Discussion

6.5.1: Correlation between replication timing and gene expression

Experiments using genomic arrays to assess replication timing and Affymetrix arrays to assess transcriptional activity showed a weak correlation. However a strong correlation could be found between replication timing and the probability that a gene would be expressed. Thus expressed genes are more likely to be located within early replicating regions of the genome which reflects what was seen when *Drosophila* was studied on a genome wide level (Schubeler, Scalzo et al. 2002).

6.5.1.1: Late replicating regions are under represented on the Affymetrix chip

One of the drawbacks of trying to find a correlation between replication timing and transcription is that two different types of arrays were used to assess the features. Correlations could therefore only be performed on regions which are represented on both arrays.

The genome profiles of replication timing and expression shown in Appendix 8 highlight regions of several megabases which are under represented on the Affymetrix chip but are represented within the 1Mb clone set. As an example, analysis in Ensembl of the first region of this type identified (75-86Mb along chromosome 2) showed that this region is sparse in genes, with few genes overlapping the 1Mb clone set (Figure 6.15)

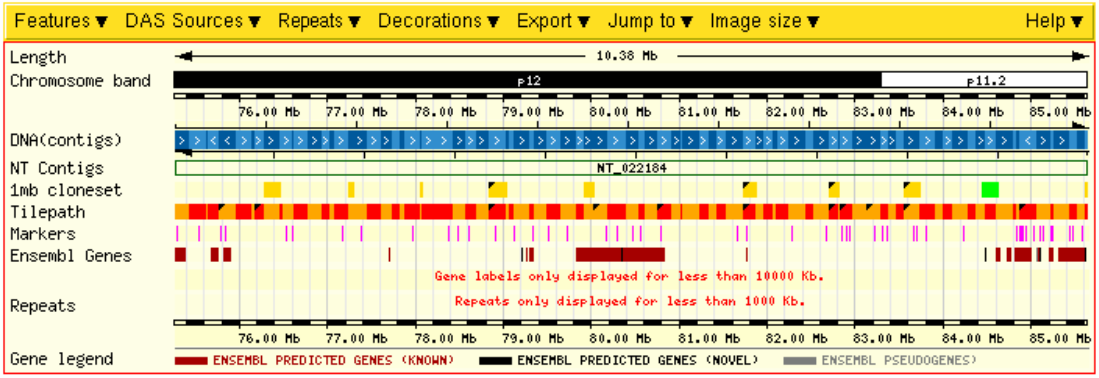
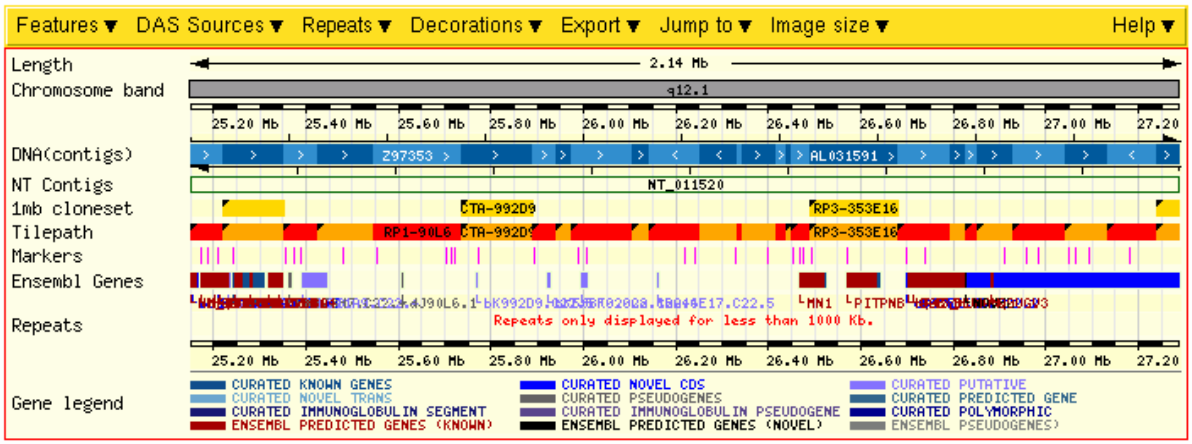


Figure 6.15: Ensembl view of a region of chromosome 2 (75-86Mb) to illustrate the position of the 1Mb clones in relation to genes in this region. The tracks of Ensembl show the cytogenetic location of the area studied and the representation of sequenced DNA contigs. The '1Mb cloneset' track shows the position of the 1Mb clones and the 'tile path' track shows the location of sequencing clones. The 'Ensembl gene' track shows the location of genes. Genes coloured in red have been confirmed, those in black are predicted.

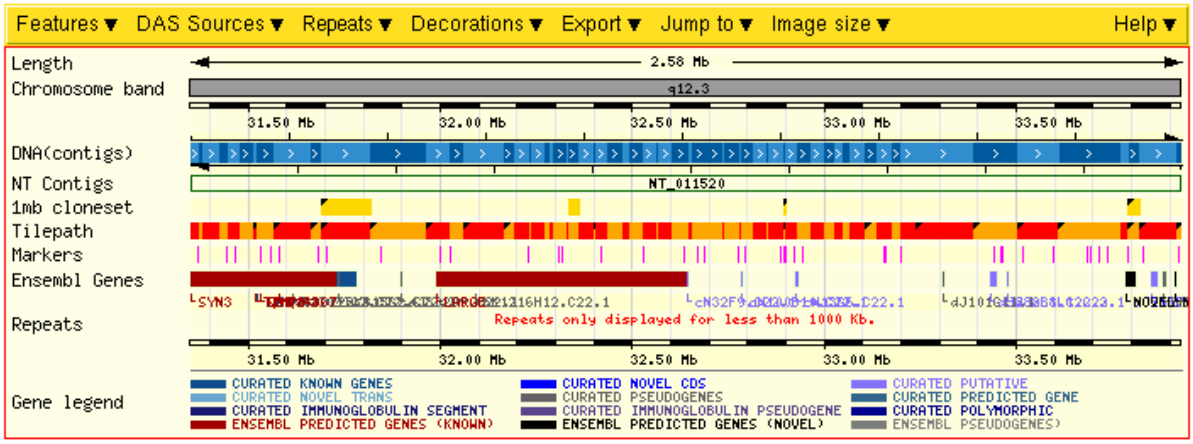
The chromosome 22 tile path array provides complete coverage of the whole q arm of chromosome 22 and ensures that the transcriptional activity of all chromosome 22 genes can be correlated with replication timing.

The study of chromosome 22 at a greater depth shows that the three major regions of late replication correspond to regions that are under represented on the Affymetrix array. When these regions were investigated using Ensembl it can be seen that these late replicating regions are lacking in verified genes (Figure 6.16).

A



B



C

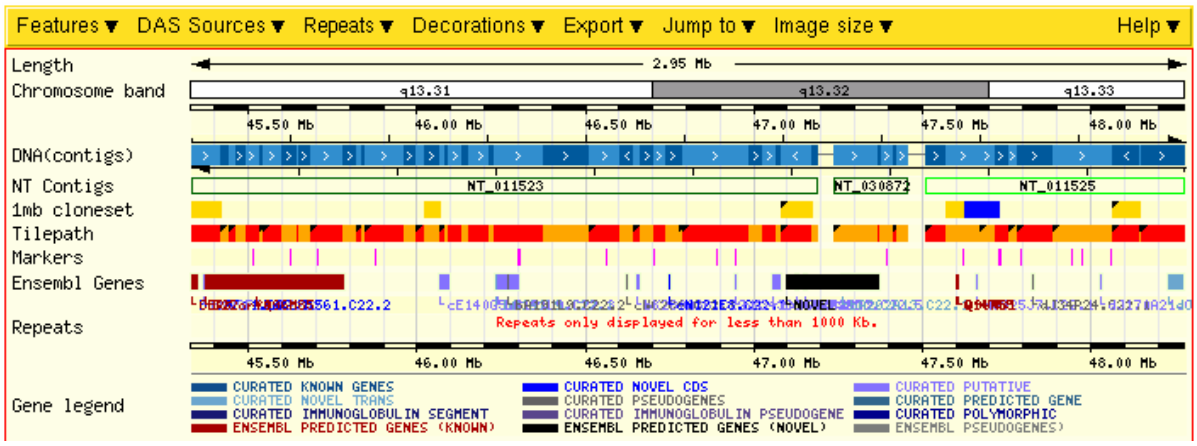


Figure 6.16: Ensembl pages illustrating those regions that are late replicating and under represented on the Affymetrix array are gene poor. A: Region 1, 25150000-27290000bp, B: Region 2, 31350000-33930000bp, C: Region 3 45330000-48280000bp. The tracks are as described in the legend to Figure 6.15.

The correlations as detailed in section 6.2 can only be performed if a locus is present on the Affymetrix chip and the genomic array. Analysis on the chromosome 22 tile path array showed that regions under represented on the Affymetrix array are late replicating. This follows the statistical analysis reported, regions with a low probability of expression (in this case, because there are no genes present) are late replicating.

6.5.1.2: Statistical analysis of the link between replication timing and transcriptional activity

The statistical analysis performed revealed a strong correlation between replication timing and probability of gene expression. The correlation coefficient of the analysis on the 1Mb chip was 0.62. This is stronger than most of the correlations with sequence features of the genome (except GC content). Analysis on the tile path arrays showed a higher correlation than that seen on the 1Mb array and better than any sequence feature investigated. The complete coverage afforded by the use of tile path arrays enables all genes that are represented on the Affymetrix array to be correlated with replication timing. This could explain why a better correlation between

probability of transcription and replication timing was found using the tilling path arrays compared to the less representative 1 Mb resolution arrays.

The probability of gene expression was calculated by grouping clones into 50 (1Mb array) or 25 (tile path arrays) and calculating the proportion of clones within each group containing an expressed gene. Groups of 50 were chosen for the 1Mb data to make the analysis comparable with that performed on the *Drosophila* genome (Schubeler, Scalzo et al. 2002). Smaller groups were chosen for the analysis of the tile path arrays to ensure sufficient data points for statistical analysis as fewer genes are represented.

The results obtained confirmed what was found in another high eukaryote, *Drosophila* (Schubeler, Scalzo et al. 2002). Schubeler *et al* also reported that early replicating genes are more likely to be expressed and yet I saw only a weak correlation is seen with level of expression. This is contrary to what has been reported in the yeast genome, where no correlation between replication and transcription was found (Raghuraman, Winzeler et al. 2001). However when the yeast genome was studied, the only correlation performed was between replication time and expression level. It is possible that further analysis of the yeast data could show a correlation between replication timing and the probability of expression. Unfortunately, the data that would allow this analysis has not been made available.

Our global analysis of replication timing confirms what has been observed in the interphase nuclei of HeLa cells, with early replicating DNA co-localising to transcriptionally active regions of the genome (Hassan, Errington et al. 1994). This is consistent with models linking replication timing and transcriptional activity (Section 1.4, Figure 1.14). Considering model A first, transcriptional activators are recruited into early replicating DNA, whilst repressors are recruited into late replicating DNA. Transcriptional activators will have to be present if a gene is expressed. The second model considers the condensation of the chromatin and theorises that replication initiators are recruited into transcriptionally active, open chromatin first, at the beginning of S phase. Only at the end of S phase is the transcriptionally inert, condensed chromatin replicated. Our results linking transcription and replication can support either of these models.

The correlation between replication timing and probability of transcription is stronger on the tile path arrays than any of the correlations with sequence features described in section 4.4. This suggests that the transcriptional status of the genome may be more important in determining the replication timing than sequence features of the genome or *vice-versa*. However sequence features also interrelate with transcriptional activity. The suggested dominance of transcriptional activity is unsurprising as during the 'timing decision point' in G1 phase of the cell cycle, it is the transcriptional activity of the genome that leads to repositioning in the interphase nucleus (described in section 1.2.3). After nuclear repositioning, the replication timing program of the genome is determined (Dimitrova and Gilbert 1999; Gilbert 2002).

However there are clones, within the groups of 50 used for analysis, which show very early replication, but contain no expressed genes. The distribution of the replicons in relation to the position of the clones may explain the appearance of some early replicating clones, with no transcribed genes. Replicons containing an expressed gene will replicate early. If the replicon spans more than one clone the early replication timing resulting from gene expression will also affect adjacent clones which may have no genes expressed. Another possibility is that the replication timing of several adjacent replicons is usually similar, leading to regions of approximately 1-2Mb with comparable replication timing. It may be that a clone within a gene rich and early replicating region, is itself gene poor but replicates early with the rest of the region. This may also explain why some clones with no transcriptional activity are early replicating.

6.5.2: Assessment of Histone modifications using the tile path array.

Modification of the Histone 3 and Histone 4 (H3 & H4) subunits of the nucleosome by acetylation was assessed on the 22q tile path array. The DNA was immunoprecipitated with antibodies to acetylated H3 or H4 and was hybridised to the 22 tile path array to identify regions that were enriched in either H3 or H4 acetylation and correlate these with replication timing.

Regions such as those 9.7-12.8, 16.1-17.5 and 31.0-33.5Mb along 22q show less acetylation on both H3 and H4 than the rest of chromosome 22, whilst regions such as those 12.7-13.7, 25.3- 26.3 and 34-34.5Mb (at the telomere) are hyperacetylated in comparison to the rest of chromosome 22 (Figure 6.10). The pattern of acetylation enrichment was very similar for H3 and H4 with a correlation coefficient of 0.7.

Histone acetylation was then correlated with other features of the genome, including replication timing. The correlation with replication timing was very weak. However when the replication profile was compared to histone acetylation enrichment (Figure 6.11), it was found that regions of late replication were not enriched in DNA associated with acetylated histones. The weak correlation observed with replication may be due to the sampling resolution of the 22q tile path array. The average sampling resolution of the 22q tile path array is 78Kb. This may include regions of acetylation enrichment and regions of hypoacetylation. Averaging of acetylation status over a 78Kb region may give an inaccurate report of the correlation between histone acetylation and replication timing. An array with a higher sampling resolution may show a better correlation. Smaller probes on the array would mean more accurate acetylation maps could be produced. It would also allow greater accuracy in determine acetylation status at regions such a gene promoters and replication origins.

The best correlation with acetylation was that with the probability of expression, although the correlations were not as strong as those between replicating timing and probability of expression. This is consistent with the open form of chromatin produced by acetylation of histones within the nucleosome aiding the likelihood that a gene will be transcribed. Weaker correlations with other genome features may be secondary to the correlation observed between histone acetylation and replication timing.

Because of the correlation between transcriptionally active chromatin and histone acetylation the lack of a correlation between histone acetylation and replication timing is surprising. This could be explained in a variety of ways. Firstly, the antibodies used detect acetylation on H3 or H4 are non-specific to the individual lysine residues within the histone tails. There are several lysine molecules on the amino tail of histone H3 and H4 that are potential sites for acetylation. An antibody that binds to

acetylation anywhere on H3 and H4 will not detect the subtleties of individual lysine acetylation. It is possible that the degree of acetylation on the different lysine molecules is related to gene transcription and replication timing. A second reason for the lack of correlation may be the resolution at which the histone acetylation was sampled. Replicons are known to be 40-200Kb in length. Therefore an array sampling 22q at a 78Kb resolution will be suitable for assessing replication timing. However this array may sample the genome at too low a resolution to be appropriate to sample histone acetylation. Future studies performed at a higher resolution may be more appropriate for assessing the correlation between histone acetylation, transcriptional activity and replication timing.

6.5.3: The Change of Replication Timing in a Translocated Cell Line.

The replication timing of a lymphoblastoid cell line with a translocation between chromosomes 17 and 22 was assessed using the chromosome 22 tile path array. This was compared to the replication timing of a normal lymphoblastoid cell line to investigate whether the translocation had any affect on the replication time.

The location of the breakpoint on chromosome 22 is within the clone bA46E17, which has a midpoint located 11546117bp along chromosome 22. In a normal lymphoblastoid cell line this clone reports a replication timing ratio of 1.55. This is much later then the chromosome 22 average of 1.75, and the clone is located within one of the late replicating bands of the chromosome.

In the translocated cell line the replication time of this clone is 1.56, and is not significantly different to the replication time reported in the normal cell line. However there is a significant change in replication timing just 430Kb (5 clones) down stream of the breakpoint. The replication timing of the translocated cell line becomes earlier replicating than the normal cell line. There is an additional shift towards early replication at the telomeric end of the chromosome. Conversely there is a movement from early to late replication approximately 21-24Mb along 22q (Figure 6.13). This places DNA that was previously in an early replicating band in a late replicating region. There is also a shift towards early replication at the VJ recombination region,

but this is more likely to be due to the epigenetic changes associated with IgL recombination, rather than being driven by the translocation. Assessment of DNA copy number from the translocated cell line as described in section 7.3 would verify this hypothesis.

Chromosomes 17 and 22 are both small, gene dense chromosomes located towards the centre of the nucleus (Cremer, von Hase et al. 2001). A greater number of chromosomal translocations between chromosomes 17 and 22 occur than would otherwise be expected for chromosomes of their size (Bickmore and Teague 2002). Translocation of the distal region of 22q onto the q arm of chromosome 17 has an affect on the replication timing of the translocated region of chromosome 22. Movement towards earlier replication suggests the chromatin is repackaged into a more open form. This may be due to relocation of the chromatin within the interphase nuclei. Both chromosomes 17 and 22 are located towards the centre of the interphase nuclei, however the movement of DNA within the specific areas that undergo a change in replication timing may result from relocation in relation to the nuclear matrix or the position in the interphase nuclei.

The time at which DNA undergoes replication is determined during the timing decision point (TDP, see section 1.2.1) during the GI phase of the cell cycle (Gilbert 2002). During the TDP, transcriptionally active DNA sequences are repositioned in the interphase nuclei in an environment favourable to gene expression. The repositioned transcriptionally active sequences are then programmed to replicate early.

Translocation between chromosomes 17:22 may result in sequences being repositioned in different places within the interphase nuclei and therefore may become programmed to replicate at the different time observed. To test this theory an expression array could be performed on the t(17:22) cell line. If the replication time of the cell line has changed because it has been moved to a more transcriptionally active or inactive region of the interphase nuclei, the mRNA expression of these areas would also change. Regions that replicated earlier in the t(17:22) cell line should be more transcriptionally active, and *vice-versa*. This hypothetical altered pattern of gene expression may lead to the phenotype observed within the patient.

The study of this one translocation has shown that replication timing changes when a translocation occurs. This event may happen in isolation, but is much more likely to be due to a change in position of the translocated portions of the chromosome. The change in replication timing may therefore indicate changes in nuclear position, transcriptional activity and other epigenetic factors. The construction of tile path arrays for all human chromosomes will allow other translocations to be studied in this way. The understanding of the replication timing and other epigenetic changes that occur in a translocated chromosome may help explain the molecular events involved in and which are the consequence of translocation and help us understand the link between a particular translocation and phenotype.

6.5.4: Replication Time of Constitutional Breakpoints in a Normal Cell Line.

A set of constitutional translocations were mapped onto the normal replication timing profiles produced on the 1Mb array. A correlation was observed between the replication timing of the two sites where the chromosomal breakpoints occurred. Unlike the work performed by Schleirmacher *et al*, it was not found that chromosomal breakpoints localised to regions of early replication (Schleiermacher, Janoueix-Lerosey et al. 2003). However those conclusions had been drawn by assessing translocations in neuroblastomas, not the constitutional translocations that were analysed in this study. In general, in this study it was found that that regions of early replication underwent translocation with other early replicating regions, whilst late replicating regions translocate with other late replicating regions.

A study analysing a large number of constitutional translocations (>10,000) provides evidence that the frequency of translocation is influenced by chromosomal size, gene density and nuclear position (Bickmore and Teague 2002). In general, gene dense regions of chromosomes are deficient in translocations. As gene dense regions of the chromosomes are also earlier replicating than gene sparse regions of the chromosome it may be expected that most breakpoints map to late replicating regions of the genome. However in the small number of constitutional breakpoints analysed on the normal lymphoblastoid replication profile this does not seem to be the case. Most of

the translocations studied map to mid replicating regions, which may indicate that other features, such as repeat content of sequence and nuclear location may be important in addition to gene density.

Nuclear position is also related to frequency of translocation. Translocations involving chromosomes 17, 19 and 22 are more frequent than would otherwise be predicted for the size and gene density of these chromosomes (Bickmore and Teague 2002). This was proposed to be due to nuclear position. Studies on non-constitutional breakpoints suggest that most translocations occur between sequences less than 1μm apart in the interphase nuclei (Savage 1996).

In interphase nuclei it is known that early replicating DNA is found in the internal nuclear environment, whilst DNA that replicates late localises adjacent to the nuclear periphery and nucleolus ((Cremer and Cremer 2001), see section 1.3.3). As sequences are positioned next to those with a similar replication timing within the interphase nuclei it is inevitable that regions with similar replication timing are more likely to undergo translocation. This is supported by the correlation shown in Figure 6.15.

6.5.5: Summary

In summary, the transcriptional activity of a lymphoblastoid cell line has been assayed on an Affymetrix U133A chip. This was then correlated with the replication timing data reported in Chapter 5. My experiments genome wide and tile path experiments on the human genome showed a weak correlation between replication timing and gene expression level. However, a strong correlation was seen between replication timing and probability of expression. This supports data previously reported in *Drosophila melanogaster* (Schubeler, Scalzo et al. 2002).

Histone acetylation was assayed on the 22q tile path array using ChIP on CHIP. The correlation between histone acetylation status and replication timing was very weak. This could be due to the sampling resolution of the array used.

The chromosome 22q tile path array was also used to assess the replication timing in a cell line that had undergone a translocation between chromosomes 17 and chromosome 22. The chromosome 22 replication timing profile of this cell line showed regions of replication timing clearly different to the replication timing profile of a normal lymphoblastoid cell line

7. Results 5
Assessment of Chromosomal Aberrations Using Genomic Arrays.

7.1: Introduction

7.1.1: Microdeletion Syndromes

7.1.1.1. Low Copy repeats at sites of Microdeletions.

Unlike the genomes of lower organisms the human genome consists of over 50% repetitive DNA (IHGSC 2001). These repeats fall into two different categories; common repeat elements, such as short and long interspersed nuclear elements (SINES and LINES), and segmental duplications. Segmental duplications (a subclass of low copy repeats) are regions of the genome, typically 50-500Kb long with high sequence similarity (98.5-99%). Segmental duplications can occur as intrachromosomal duplications, where the duplicated regions are on the same chromosome or as interchromosomal where the two or more duplicated regions are on different chromosomes.

The completion and publication of the whole draft genome sequence in 2001 (IHGSC 2001) allowed the comparison of regions of the genome, and the identification of segmental duplications. Within the finished sequence an estimated 3.3% of the genome was involved in segmental duplications. Intrachromosomal duplications account for about 2.64% of the total genome and interchromosomal duplications for 1.44% (Cheung, Estivill et al. 2003). Gene rich chromosomes show the highest incidence of segmental duplications.

Computational analysis of the human genome sequence by Bailey et al (Bailey, Gu et al. 2002) identified 169 large regions of the genome that had an over-representation of human shotgun sequence used to sequence the human genome by Celera Genomics. These sequences were found to be rich in these segmental duplications. Of the 169 regions identified, 24 are currently associated with disease; these include Gauchers disease on chromosome 1, Prader Willi and Angelman's syndrome on chromosome

15 and the DiGeorge region on Chromosome 22. The combined incidence of childhood disease involving segmental duplications is 1:750 (Eichler 2001). The reason for the association of duplicated regions with disease is due to the misalignment of chromosomes during meiosis where recombination occurs between duplicated regions rather than allelic loci. Recombination between homologous regions may result in deletion, amplification or inversion events. This can result in the disruption of a gene resulting in disease associated phenotypes. The mechanisms for this are shown in Figure 7.1.

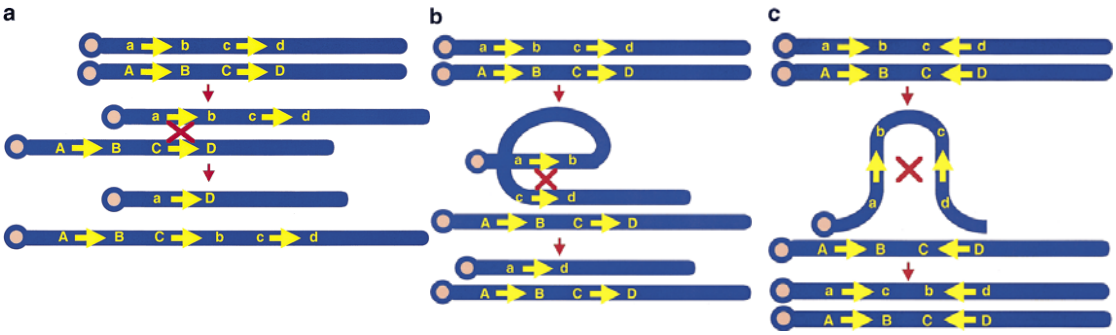


Figure 7.1: Mechanisms for segmental duplications a: recombination between repeats on two separate chromosomes leads to a deletion on one chromosome and an amplification on the other chromosome. b: recombination between repeats on the same chromosome leads to a deletion. c: recombination between repeats in an opposite orientation leads to an inversion (Adapted from (Ji, Eichler et al. 2000)).

One region of the genome particularly rich in segmental duplications is the subcentromeric region of the q arm of chromosome 22 (Dunham, Shimizu et al. 1999). In the 1.5 Mb of DNA adjacent to the centromere, which represents just 5% of the chromosome 22 sequence, 90% of sequence is duplicated on other chromosomes. 52% of the interchromosomal duplications on chromosome 22 were also located in this region (IHGSC 2001). Low copy repeats are also common within the next 7.5Mb of chromosome 22, with most of the sequence clones in the first 9Mb of the q arm containing some form of segmental duplication. The duplicated regions of chromosome 22 are represented in Figure 7.2. The highly duplicated region at the beginning of 22q includes the DiGeorge critical region (DGCR) that is deleted in patients suffering from DiGeorge syndrome. The same region involved in segmental

duplication is involved in velocardiofacial syndrome (VCFS) and CATCH22. The diversity in names is due to the variability of phenotypes observed.

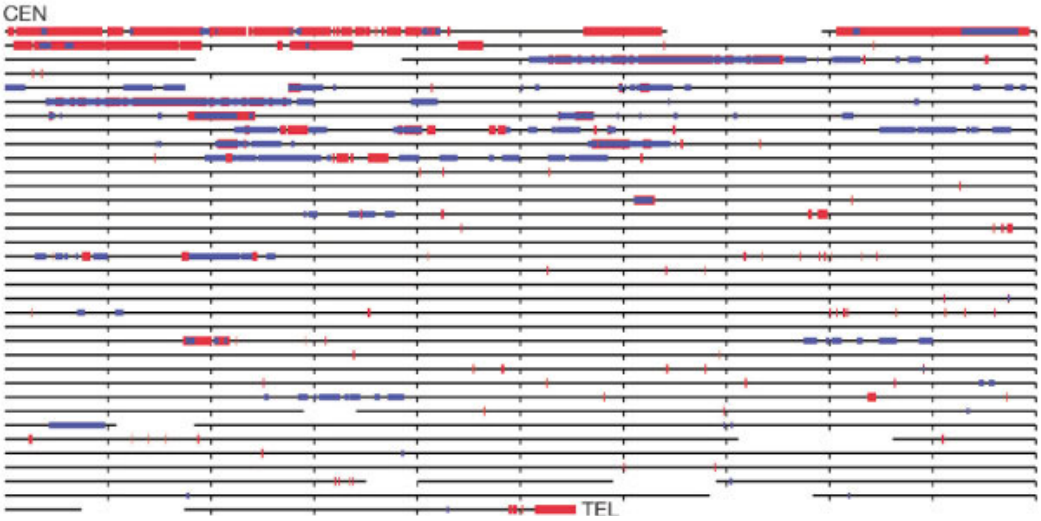


Figure 7.2: Segmental duplications on the sequenced q arm of chromosome 22. Each line represents 1Mb, with each intersection at 100Kb. Intrachromosomal repeats are shown in blue and interchromosomal repeats in red. Duplication alignments with > 90% nucleotide identity and > 1 kb long are shown (IHGSC 2001), (Bailey, Gu et al. 2002).

7.1.1.2. DiGeorge Syndrome & Conventional Diagnosis

DiGeorge is the most common microdeletion syndrome and is present in 1:4000 live births (Devriendt, Fryns et al. 1998). The syndrome is characterised by a variety of clinical features. These include a variety of heart defects, mainly affecting the aortic arch, immunodeficiencies due to a hypoplasic/absent thymus, hypocalcaemia – owing to hypoplasia of the parathyroids, and distinct facial features including low set ears and a cleft pallet. Cases presenting later in childhood tend to have a milder phenotype encompassing heart defects (OMIM entry 188400). The 3Mb deletion and the 1.5Mb deletion have indistinguishable phenotypes (Maynard, Haskell et al. 2002).

The DiGeorge critical region has been localised to chromosome 22 approx. 3966000 – 7888000bp along the q arm. The region is flanked by accession numbers AC008079-D86996 and the boundaries were defined by screening using high density genetic markers. Over 150 patients were screened with their unaffected parents used as

controls. A detectable deletion was found in 83% of the patients examined (Carlson, Sirotkin et al. 1997). Most deletions (approx. 90%) encompass a 3Mb deletion between two duplicated regions (Lindsay 2001). A further 8% of deletions have the same proximal boundary, but are smaller, encompassing 1.5Mb of DNA, between another low copy repeat (Figure 7.3). It has also been observed that rearrangement within the DiGeorge region may be associated with balanced translocations with 11q23 (Spiteri, Babcock et al. 2003). Non-22q11 deletions resulting in the DiGeorge phenotype may be due to deletions on other chromosomes. In these patients deletions have been detected on 10p13, 18q21.33 and 4q21.3-q25 (Greenberg, Elder et al. 1988 McDermid and Morrow 2002).

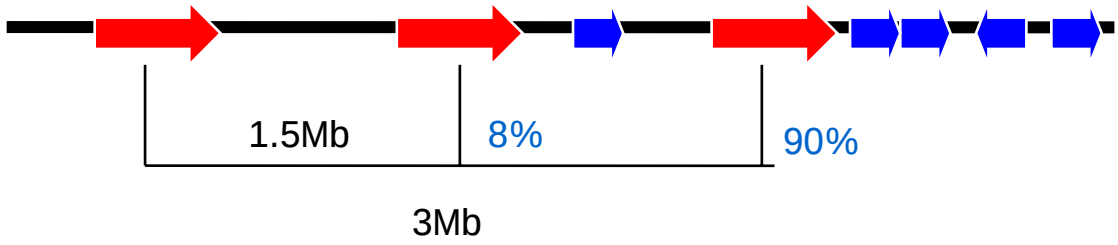


Figure 7.3: Patterns of deletion in DiGeorge patients. Red arrows: Duplications involved in DiGeorge deletions. Blue arrows: Other segmental duplications in the vicinity of the DiGeorge critical region.

The 3Mb deletion that is responsible for most patient phenotypes incorporates 30 genes. The smaller nested deletion encompasses 24 genes, with a variety of functions. However, no single gene has been identified as being solely responsible for DiGeorge syndrome. A low copy repeat (LCR 22) ranging in size from 40-350Kb and of 97-98% identity has been found at all three 22q11 breakpoint regions, as well as six adjacent locations over a 6.5Mb region. The mechanism of the chromosomal rearrangement at 22q11 is shown in Figure 7.4 (Maynard, Haskell et al. 2002).

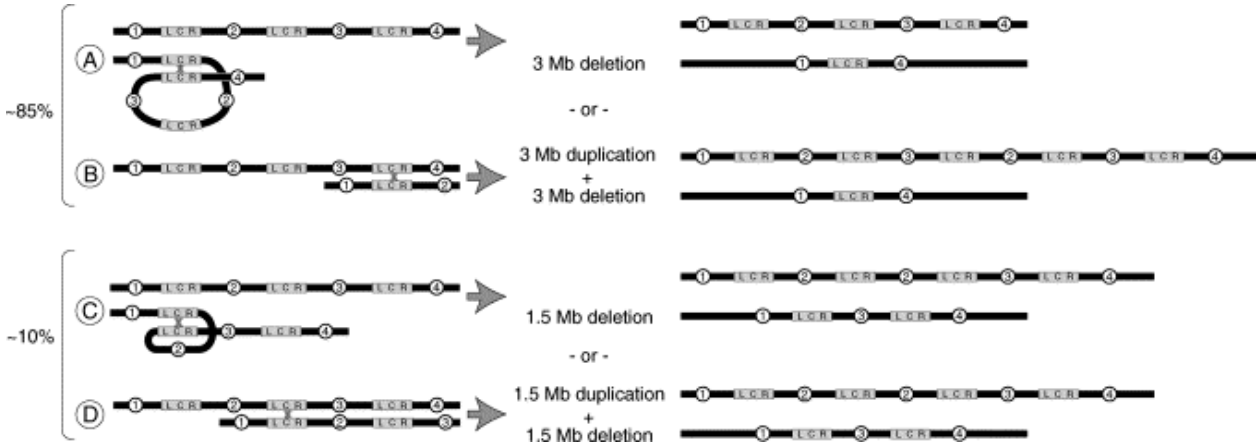
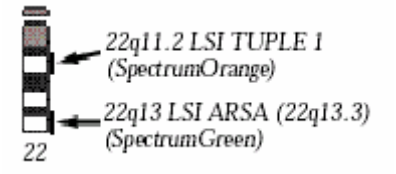


Figure 7.4: Mechanisms for deletion in DiGeorge patients (Maynard, Haskell et al. 2002).

The DiGeorge deletion in patients is usually clinically confirmed by FISH analysis of patient chromosomes using commercially available probes.

A



B

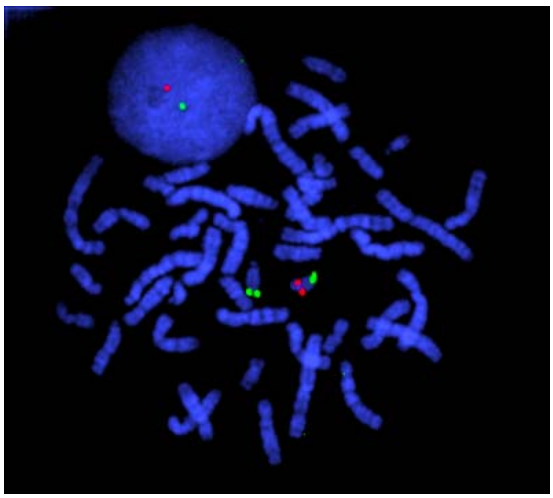


Figure 7.5: Detection of the DiGeorge deletion on patient metaphases using the commercially available probe set from Vysis. A: This is a two colour probe mix containing a spectrum orange probe mapping to the non-coding region of TUPLE 1 in the DiGeorge region and spectrum green labelled control probe hybridised to a region on 22q13.3. B: The Red signal can be seen as present on the normal chromosome 22 but absent on the copy of chromosome 22 containing the DiGeorge deletion. Reproduced from (Gribble 2003).

This commercially available Vysis probe illustrated in Figure 7.5 contains the markers D22S553, D22S609, D22S942, within the accession numbers AC000085, AC000092, AC000079. The probe is located within the first 1.5Mb of the DiGeorge critical region so will detect both the 1.5Mb and the 3Mb deletions. These commercial probes are used within clinical labs to detect the presence and absence of a deletion in the DiGeorge region of patients displaying a DiGeorge phenotype. They do not give any information about the size of the deletion. As the candidate gene(s) for DiGeorge deletion has not yet been identified, information about the deleted genes and their effect on phenotype is still critical in the understanding of DiGeorge syndrome.

The extent of the deletion in DiGeorge patients is conventionally sized using FISH probes covering contiguous regions along 22q11 (Lindsay 2001). Recently the production of a 22q tile path array has allowed the sizing of the deletion in a DiGeorge patient in a single hybridisation experiment (Buckley, Mantripragada et al. 2002). The hybridisation of DNA derived from a transformed DiGeorge lymphoblastoid cell line to the chromosome 22 tile path array showed a deletion spanning the DiGeorge critical region in the one patient examined. Some of the loci within the DiGeorge region gave ratios that are difficult to interpret on the CGH array. It was concluded that the reason for this is the high amount of common repeat elements within the cosmids located within this region. This preliminary study showed that a DiGeorge deletion could be detected on a DNA array. Further investigation of DiGeorge patients may confirm which genes are involved in the phenotype.

7.1.2: Immunoglobulin Rearrangements

7.1.2.1: Genome wide Immunoglobulin rearrangements

The human immune system relies on antibodies and T cell receptors to fight the large range of infectious agents that invade the body. As there is such a plethora of antigens the body has to deal with, the immune system has to produce a wide assortment of antibodies and T cell receptor (TCR) genes.

An individual produces more different types of antibody than all other proteins in the body put together. There are many more types of antibody in the body than there are genes in the genome, and therefore the conventional idea of one gene encoding one mRNA and one protein will not produce enough antibody diversity. A unique way of producing diversity has been observed in regions of the genome encoding antibodies.

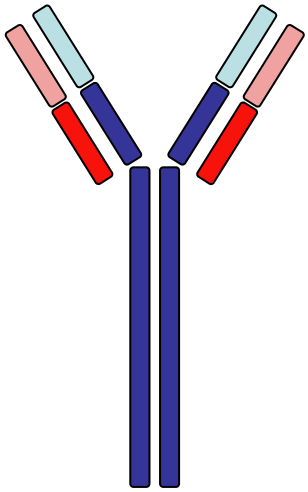


Figure 7.6: Basic four chain structure of an immunoglobulin protein. Blue: Two heavy chains, comprising of a constant region (C_H - dark blue) and a variable region (V_H – light blue). Red: Two light chains comprising of a constant region (C_L - red) and a variable region (V_L – pink).

Immunoglobulin gamma proteins have the same basic four chain structure (Figure 7.6). Antibiotic diversity is generated at the variable regions on the heavy (IgH) and light chains (IgL). The light chains are of two different types, either kappa (κ) or lambda (λ). The immunoglobulin heavy chain is encoded by a cluster of genes on chromosome 14q32.33 whilst the κ chain is encoded by genes on chromosome 2p11.2 and the λ chain is encoded on 22q11.22. The use of two different types of chain (heavy and light) immediately increases the antibody variability as any light chain can combine with any heavy chain.

In addition to the diversity generated by heavy and light chain association, immunoglobulin diversity is increased due to the rearrangement of germ-line variable

(V), diversity (D) and joining (J) gene segments. The IgH variable region exons are assembled from V, D and J gene segments. The IgL variable regions are assembled from just V and J gene segments. During B cell development the IgH variable region undergoes rearrangement first. Only 1 in 3 IgH rearrangements are in-frame and therefore successful, provided the IgH rearrangement is successful the IgL chain undergoes rearrangement. This V(D)J recombination is vital to produce diverse antibodies and only occurs in developing lymphocytes between immunoglobulin or TCR genes.

7.1.2.2: The Mechanism of λ Chain Rearrangement in Chromosome 22.

The lambda gene locus, encoded on chromosome 22q11 contains a set of variable genes and seven constant gene regions (Figure 7.7). The region was sequenced in 1997 (Kawasaki, Minoshima et al. 1997). In cells not producing antibodies, the variable genes and constant regions are found far apart. In cells that form antibodies the constant and variable genes are brought closer together, but still remain approximately 1500bp apart. The variable and constant regions are separated by a joining segment, which also contributes to diversity.

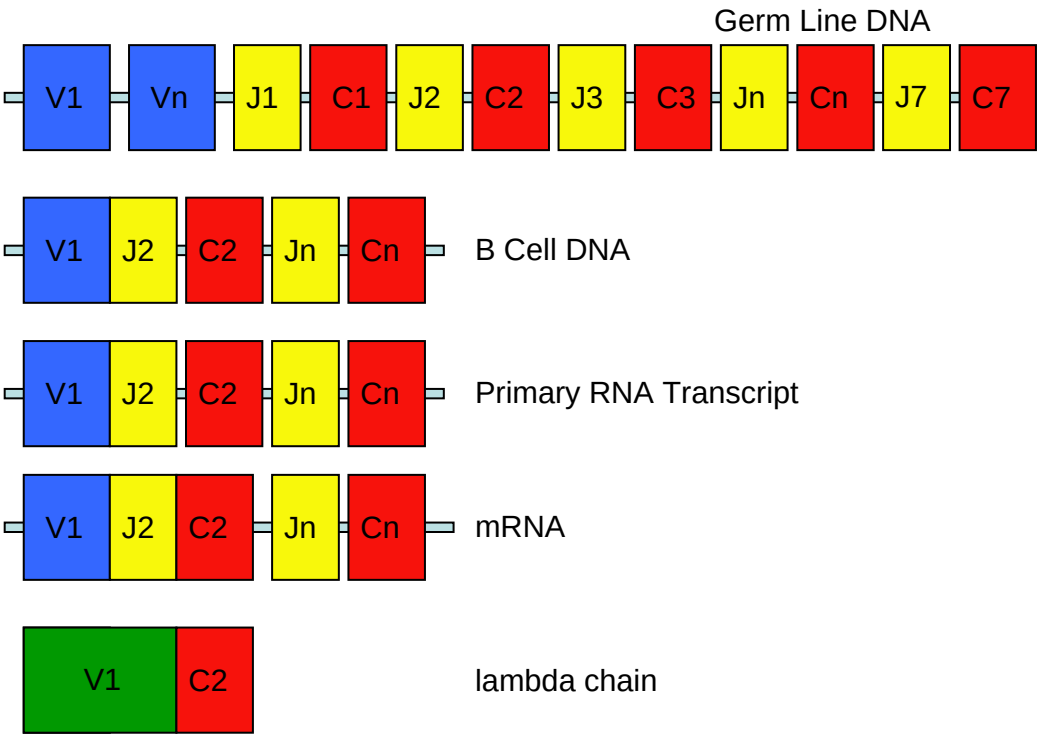


Figure 7.7: Recombination of the lambda chain of the immunoglobulin light chain. For details see text. Figure adapted from (Turner 2001).

In the germ-line of the λ chain loci there is a variety of V segments and seven different constant regions. Each constant gene region is accompanied by just one joining gene (unlike the κ chain). During B cell development the gene groups are rearranged and a region of the DNA is excised to bring variable and joining regions together. This rearrangement is mediated by a conserved recombination signal sequence which flanks the recombining regions. The recombination signal is an AT rich nonomer (ACAAAAACC) which is separated by a non-conserved 12 or 23bp spacer sequence and found upstream of the J segment. The VJ recombination occurs at sites of double stranded breaks, the recombination signal then provides complementary sequence so the ends can be precisely joined. The joining of the different segments can result in an inversion or deletion of the intervening sequences, resulting in copy number changes in these regions.

One allele will initially undergo rearrangement. If this is unsuccessful the second allele will undergo rearrangement. The IgLκ and IgLλ loci expression is under negative feedback control. The production of a functional IgL protein feeds back onto the IgLκ and IgLλ loci and prevents unnecessary rearrangement.

In this way different B cells contain different rearrangements of the constant, joining and variable genes leading to antibody diversity. The newly formed variable, joining and constant arrangement is transcribed into primary RNA. The introns are then removed and the spliced mRNA is translated into the lambda chain protein.

Expression of the successfully rearranged IgL chain is enhanced by epigenetic modification of the chromatin associated with active chromatin. The CpG islands are demethylated, histones are acetylated and transcriptional activators are also recruited to the chromatin (Blackwood and Kadonaga 1998). The active allele is switched to become early replicating whilst the inactive allele is late replicating (Goren and Cedar 2003).

Deletions in this region due to excision of DNA during VJ recombination can be detected, both by FISH and by using a CGH microarray (Buckley, Mantripragada et al. 2002). Assessment of the VJ recombination on CGH arrays will allow the sizing of rearrangements. The arrays may be able to detect incomplete VJ recombination, which leads to an imbalance in B and T cells in the immune system by reporting copy number change. They may also be used to detect aberrant VJ recombination which can lead to chromosomal translocations (Bassing, Swat et al. 2002).

7.1.3: Assessment of DiGeorge and IgL λ copy number change on genomic arrays.

The two previous chapters have described how genomic clone microarrays have been used to assess DNA replication timing in a human cell line. However the microarray sampling the genome at a 1Mb resolution and the 22q tile path array described in Chapter 4 also have other uses, such as detecting DNA copy number changes.

Section 7.2.1 describes how the chromosome 22q tiling path array can be used to detect a deletion in the DiGeorge region of 22q11. Patient DNA was obtained by collaboration with Charles Shaw-Smith, from the Department of Medical Genetics at Addenbrookes Hospital, Cambridge. Patients exhibiting features of the DiGeorge phenotype, but showing no 22q11 deletion by conventional FISH analysis were applied to the chromosome 22q array, and to the array sampling the genome at a 1Mb resolution. This is described in section 7.2.2. The DNA was obtained from collaboration with Katrina Prescott, from the Institute of Child Health, University College London. The arrays detected a deletion in one patient. Follow-up work, including additional FISH analysis and microsatallite analysis was performed at the Institute of Child Health.

The genome of B cells can undergo a rearrangement during development of the immunoglobulin light chain λ locus located in 22q11. Section 7.3 describes deletions observed in lymphoblastoid cell lines due to this rearrangement.

7.2: Array analysis of DiGeorge syndrome patients

7.2.1: Assessment of DiGeorge Patient DNA samples on the Chromosome 22q Tile path array.

DiGeorge Syndrome is a congenital defect caused by a deletion on chromosome 22q11. Five individuals who displayed a DiGeorge phenotype and who had demonstrated a deletion at 22q11 using a commercial diagnostic fluorescence *in situ* hybridisation (FISH) probe set were selected for analysis. The size of the 22q11 deletion was assessed by the hybridisation of DNA from these patients onto the 22q tile path array.

Initially, DNA from five different patients was hybridised to individual arrays using DNA from a male lymphoblastoid cell line (HRC 575) as a reference. A typical result is shown in Figure 7.8.

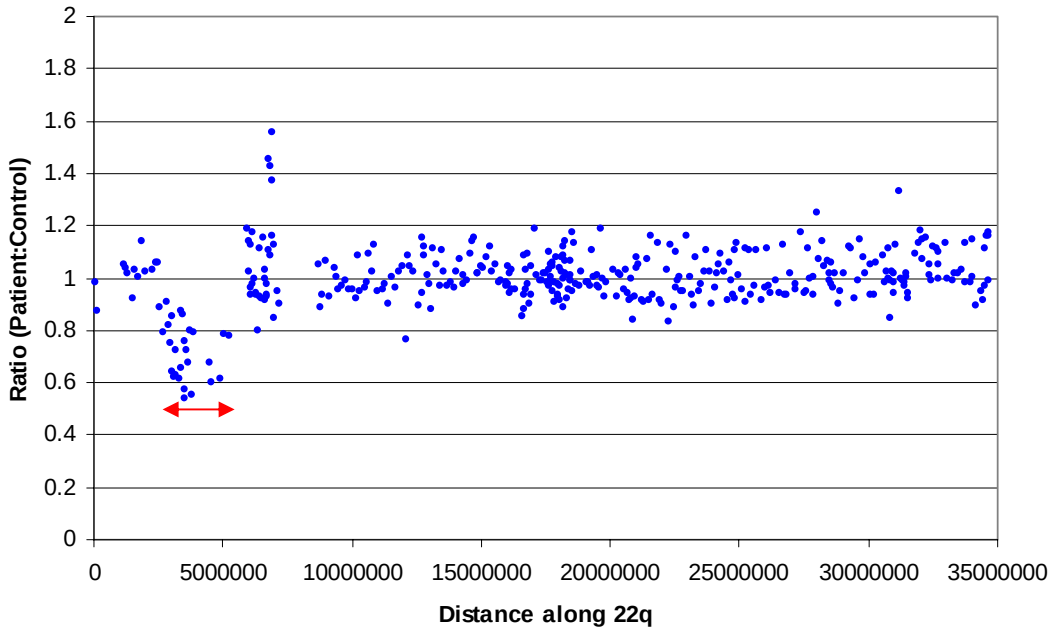


Figure 7.8: Hybridisation of DNA from a DiGeorge patient onto the 22q tile path array. DNA from a male lymphoblastoid cell line was used as a reference. Red arrow shows the region of the DiGeorge deletion.

In Figure 7.8, a deletion was can be seen between 2577096-5227316bp along chromosome 22q (between clones bac 519d21 – pac 52f6; international clone names

and accession numbers can be found in Appendix 9). A copy number change was defined as 5 times the standard deviation of the self:self hybridisation performed in 4.3.2 (0.2) to be highly statistically significant. A deletion was defined as clones that report a ratio below 0.8. A copy number gain was defined as clones with a ratio above 1.2. An amplification can also be seen in four of the clones between 6789448-6935464bp along chromosome 22q (between clones cN24A12 and cN9G6 – Table 7.1). However comparison with other experiments (See 4.3.3 and 7.3) showed that the amplified ratios seen were actually due to a deletion in the reference cell line at the immunoglobulin light chain λ locus. Consequently all the DiGeorge experiments were repeated using a pool of DNA extracted from units of donated blood from twenty anonymous individuals.

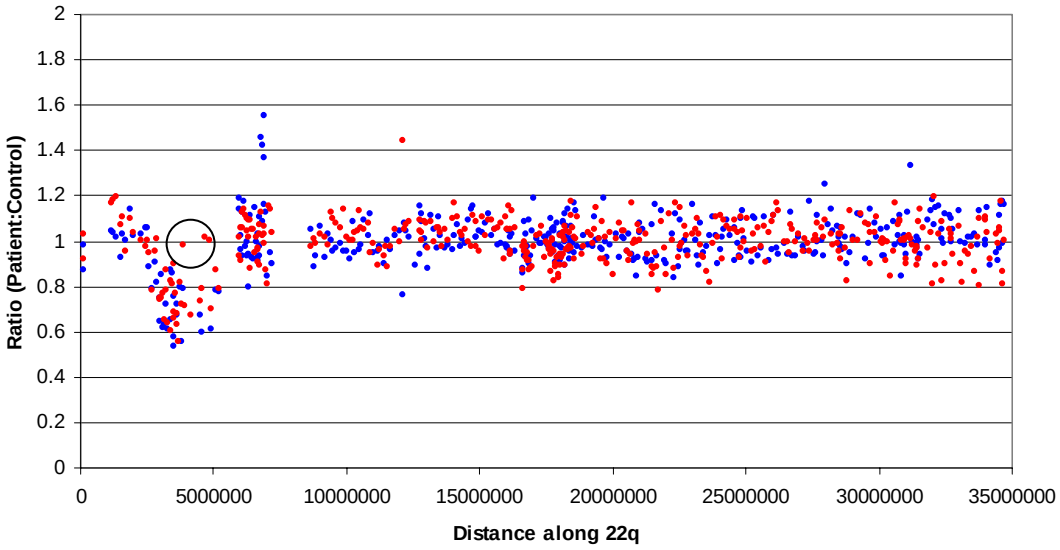


Figure 7.9: DNA from the same DiGeorge patient hybridised to two different arrays using two different control samples. Blue data points: Patient DNA hybridised against a single lymphoblastoid reference cell line (HRC 575). Red data points: Patient DNA hybridised against a pool of blood DNA from twenty different individuals. The three red points in the DiGeorge deletion region reporting normal ratios (circled) are discussed below.

When the patient DNA was hybridised against pooled DNA from twenty different individuals the four clones previously identified as having an elevated patient:control

ratio when hybridised against the lymphoblastoid cell line showed normal 1:1 ratios (Table 7.1).

Table 7.1: The region of 22q which exhibited a gain when the cell line HRC 575 DNA was used, and yet showed normal ratios when hybridised against a pool DNA control. Clones marked with an * exhibited the same copy number as the control DNA when hybridised against both control DNA's.

Clone	Midpoint of clone	Ratio V. cell line DNA	Ratio V. pool DNA
cN24A12	6789448.5	1.45	1.13
cN86D6*	6823353	1.09	1.02
cN92H4	6859162	1.42	Not Available
cN84E4	6894369	1.37	0.99
cN9C5*	6915218.5	1.16	1.03
cN9G6	6935464.5	1.56	1.06

DNA samples from five different DiGeorge patients were hybridised each to separate arrays using the pool DNA as a control. The results are shown in Figure 7.10 and Table 7.2.

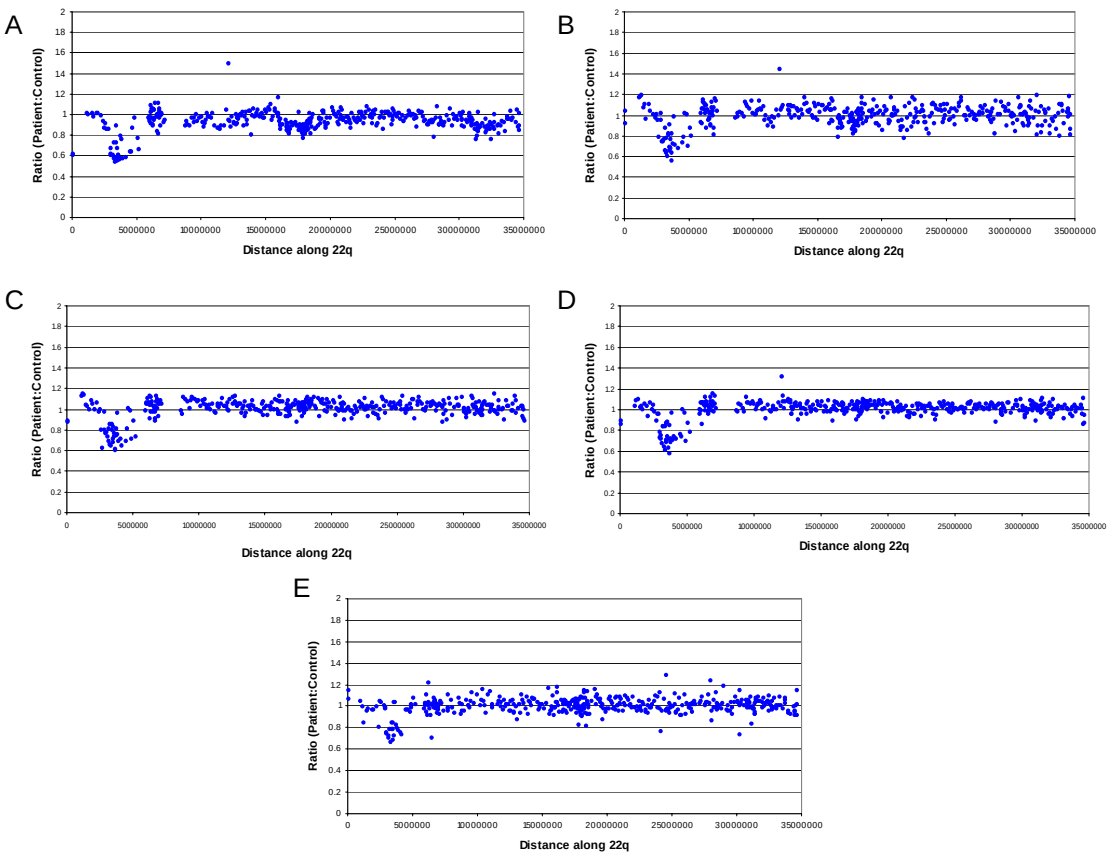


Figure 7.10: (Previous page) Patient:Control ratio profiles for five separate DiGeorge Patient DNA samples. A: Patient 1, B: Patient 2, C: Patient 3, D: Patient 4, E: Patient 5.

Table 7.2: Patient:Control ratios of clones in the DiGeorge region of chromosome 22. Shaded ratios indicated deletions in the loci represented on the array. Ratios shaded with pale grey only show a slightly reduced but not significant ratio, but are within the deleted region. NA: Clone ratio is not available. (Loci that did not pass the analysis criteria described in 2.5.2 and were rejected.) NP: clone was not present on early arrays.

Clone	Location	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5
b677f7	2523529	0.867599	1.00634	0.997294	1.041255	1.043958
bac519d21	2577097	0.910715	0.95083	0.799467	0.976757	1.036773
pac995o6	2710128	0.856811	0.782641	0.626426	0.893028	1.007301
p423	2881004	0.857811	1.012522	0.969525	0.958052	0.98568
fos41c7	2950903	NP	NP	NP	NP	0.979491
cN119F4	2979519	0.667419	0.746817	0.76231	0.723682	0.758473
18c3	3013238	0.616361	0.743239	0.772776	0.777589	0.747465
111f11	3052391	0.609105	0.750534	NA	0.753778	0.719907
b72f8	3192997	0.607635	0.873848	0.804798	0.841273	0.704972
fF39E1	3236756	0.727743	0.78112	0.852192	0.834134	0.785443
Cos98c4	3306156	0.568455	0.637356	0.656379	0.638873	NA
Cos49c12	3369955	0.534524	0.603945	0.647299	0.603859	0.664103
Cos 83c5	3406693	0.721168	0.825051	0.801886	0.836507	0.784862
Cos83e8	3443825	0.888218	0.813	0.847148	0.869298	0.84075
Cos 59f	3467897	0.578088	NA	NA	NA	0.779052
Cos105a	3493862	0.546522	0.658493	0.678754	0.679448	0.685424
Cos81h	3532078	0.67372	0.90295	0.828353	0.882227	0.84065
Cos31e	3569626	0.609019	0.768846	0.765578	0.737782	0.724199
Cos100h	3612315	0.600973	0.679369	0.729034	0.71584	1.036948
Cos91c	3652472	0.557729	0.63308	0.60015	0.63035	1.03047
Cos 89h	3731131	0.79355	0.818929	0.81366	0.842539	0.827075
c2h	3804446	0.571592	0.721842	0.75199	0.683078	0.794157
c56c	3878158	0.752982	0.981838	0.957891	0.969274	NA
p888c9	3925566	0.568191	0.712153	0.687371	0.705867	0.774482
p158I19	4056427	NA	NA	0.70935	0.740995	0.74914
b444p24	4165628	0.581054	0.676969	0.648217	0.719954	0.73486
b562F10	4491054	0.634745	0.734261	0.686802	0.760658	0.967303
p_m11	4591569	0.639606	0.787911	0.811635	0.740377	0.965098
bac32	4686897	0.86681	1.01556	1.005373	0.998268	0.967807
pac408	4854070	0.967018	1.00157	0.984653	0.969228	0.949261
b135h6	4935029	NA	0.69915	0.709847	0.697637	1.012156
p_n5	5082378	0.763858	0.872079	0.886715	0.864976	1.075634
p52f6	5227317	0.662279	0.793727	0.737265	0.778774	1.019238
cN109G12	5976695	0.964023	1.014324	0.981454	1.003975	1.02227

DiGeorge deletions at 22q11 can be seen in all five patients tested. An additional gain in clone dJ477H23 can also be seen in patients 1, 2 and 4, 12.1Mb along chromosome 22q.

Detailed plots within the DiGeorge region, for all five patients are shown superimposed in Figure 7.11.

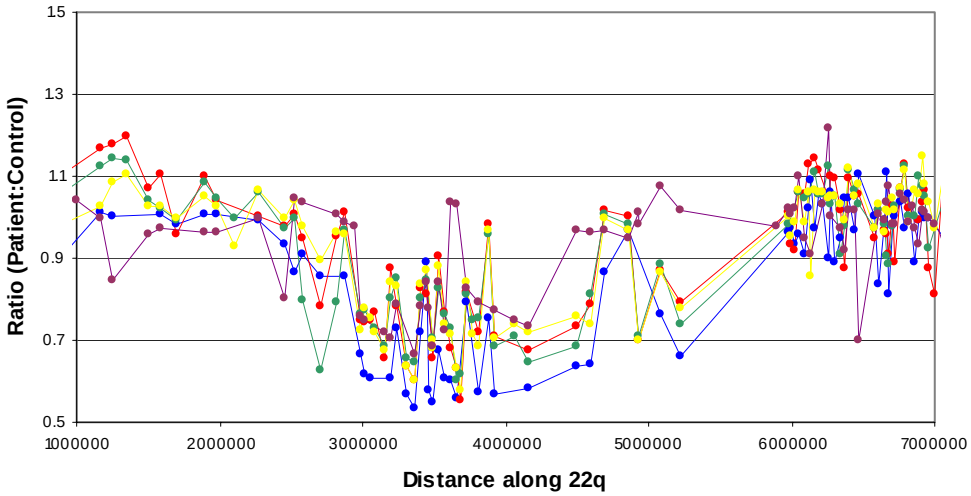


Figure 7.11: Patient:Control ratios obtained when five different patients are plotted on the same axis. Blue: Patient 1. Red: Patient 2. Green: Patient 3. Yellow: Patient 4. Purple: Patient 5.

A full single copy loss (ratios approximating 0.5:1) can only be seen in a few of the clones in the deleted region. The clones showing a full single copy loss are interspersed with clones reporting an intermediate ratio. There are also clones in the centre of the deleted region that report ratios that are modal as would be expected for non-deleted regions. The four clones that have this characteristic on multiple arrays are; Pac423 (average ratio = 0.95), Bac32 (0.91), Cos56c (0.97) and Pac408 (0.97).

These intermediate ratios make defining the boundaries of the deletion uncertain. At the edges of the deletion it is difficult to distinguish deleted ratios from the background variation. Because of this the shaded deleted regions in Table 7.2 do not have their boundaries defined exactly. It is however possible to determine that the deletion in patient 5 is smaller than the other deletions. This was confirmed by FISH

analysis of the patient chromosomes using clones bK562F10, p_n5 and p52F6 as probes (Figure 7.12).

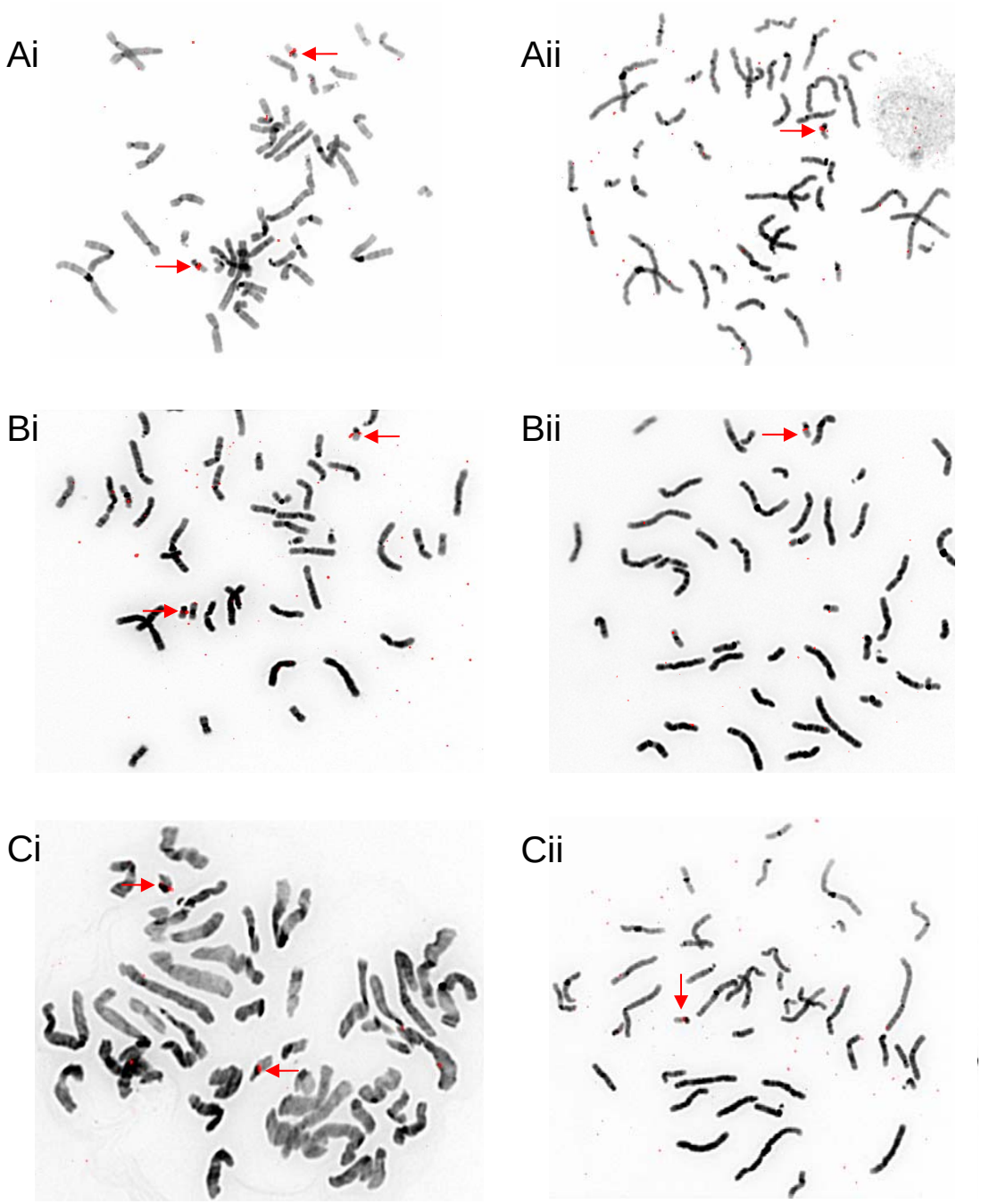
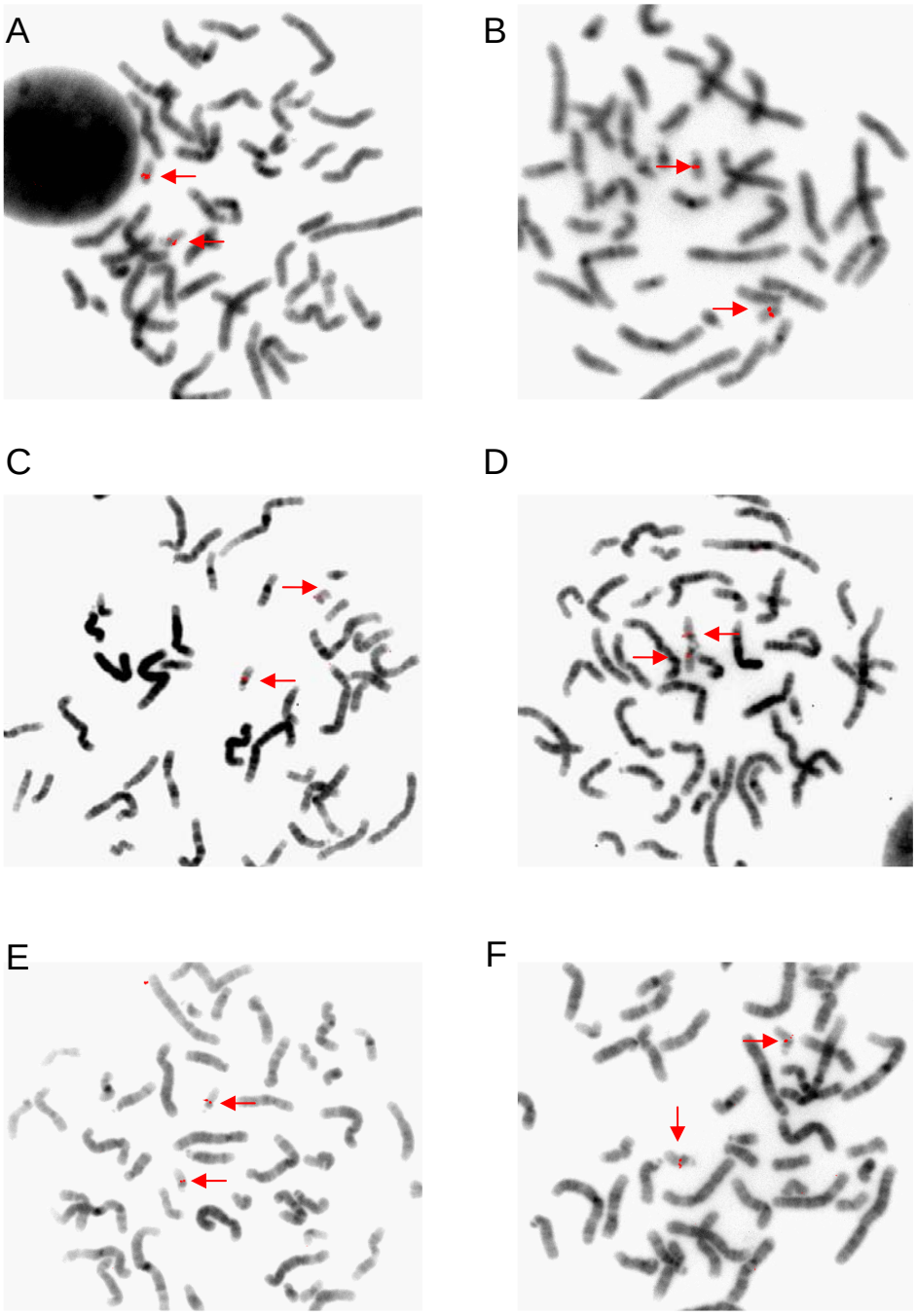


Figure 7.12: FISH analysis of the region that the array indicated is not deleted on patient 5. Two categories of metaphase spread from the patient are observed; (i) those that show a signal on both copies of the chromosome 22s, and (ii) those that show a signal on just one copy of chromosome 22. The probes are A: bK562F10, B: p_n5 and C: p52F6.

One hypothesis that would explain intermediate ratios observed in the DiGeorge region is the abundance of segmental duplications in 22q11. Duplications would mask the single copy deletion that is characteristic of DiGeorge syndrome, and would explain the intermediate ratios exhibited. A clone with a duplication at just one other loci within the genome would report a 3:4 ratio opposed to a 1:2 ratio. Because of this, selected clones from the DiGeorge region were mapped by fluorescence *in situ* hybridisation to normal chromosomes, and chromosomes isolated from two of the DiGeorge patients (Table 7.13 and Figure 7.13-7.15). Due to the limitation on the number of patient metaphases, experiments that were unsuccessful could not be repeated.

Table 7.3: Clones chosen for FISH analysis, and results on the patient metaphases. (A & B: clones on the edge of the DiGeorge deletion not reporting any copy number loss. C & D clones in the middle of the DiGeorge deletion not reporting any copy number loss. E-G: Clones reporting a single copy number loss on the arrays. H & I: clones reporting intermediate ratios on the array)

	Clone	**Accession no**	**FISH – Normal cell line**	**FISH – Patient 1**	**Array Ratio - 1**	**FISH – Patient 4**	**Array Ratio - 4**
A	519d21	AC008079	2x22	2x22	0.91	2x22	0.98
B	995o6	AC008132	2x22	2x22	0.85	2x22	0.89
C	Cos56c	Ac000080	2x22	NA	0.75	1x22	0.97
D	Bac32	Ac007050	2x22	1x22	0.87	2x22	1.00
E	49c12	Ac000079	2x22	1x22	0.53	1x22	0.60
F	98c4	Ac000092	2x22	NA	0.56	NA	0.64
G	52f6	Ac005500	2x22	NA	0.66	1x22	0.78
H	Pn_5	Ac002472	2x22	1x22	0.76	1x22	0.86
I	83c5	Ac000087	2x22	1x22	0.72	1x22	0.84


A
B
C
D
E
F

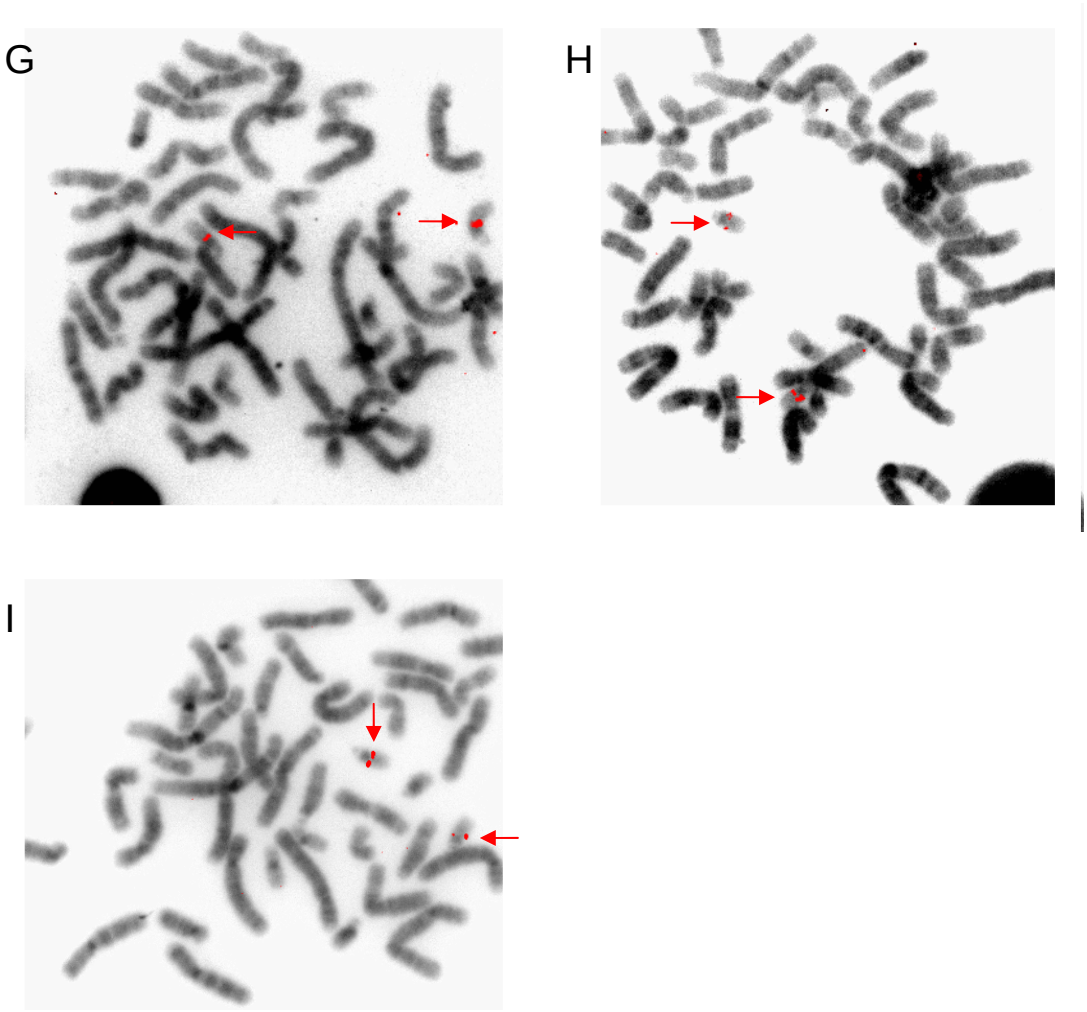


Figure 7.13: Probes hybridised to chromosomes prepared from a normal (46, XY) lymphoblastoid cell line. Lettered images relate to the probes described in Table 7.3.

The hybridisation of the probes to normal metaphase chromosomes showed no secondary signals that may indicate segmental duplications, although all the clones that were examined by FISH analysis have previously been shown to contain segmental duplications elsewhere on 22q (Buckley, Mantripragada et al. 2002). However the resolution of metaphase FISH would not enable intrachromosomal repeats elsewhere on chromosome 22q11 closer than 2 Mb from the FISHed clone to be resolved.

The results in the DiGeorge region indicate segmental duplications may affect the ratio reported by the arrays. However either the stringency of the FISH or the inability to resolve intrachromosomal repeats may produce disparate array and FISH results.

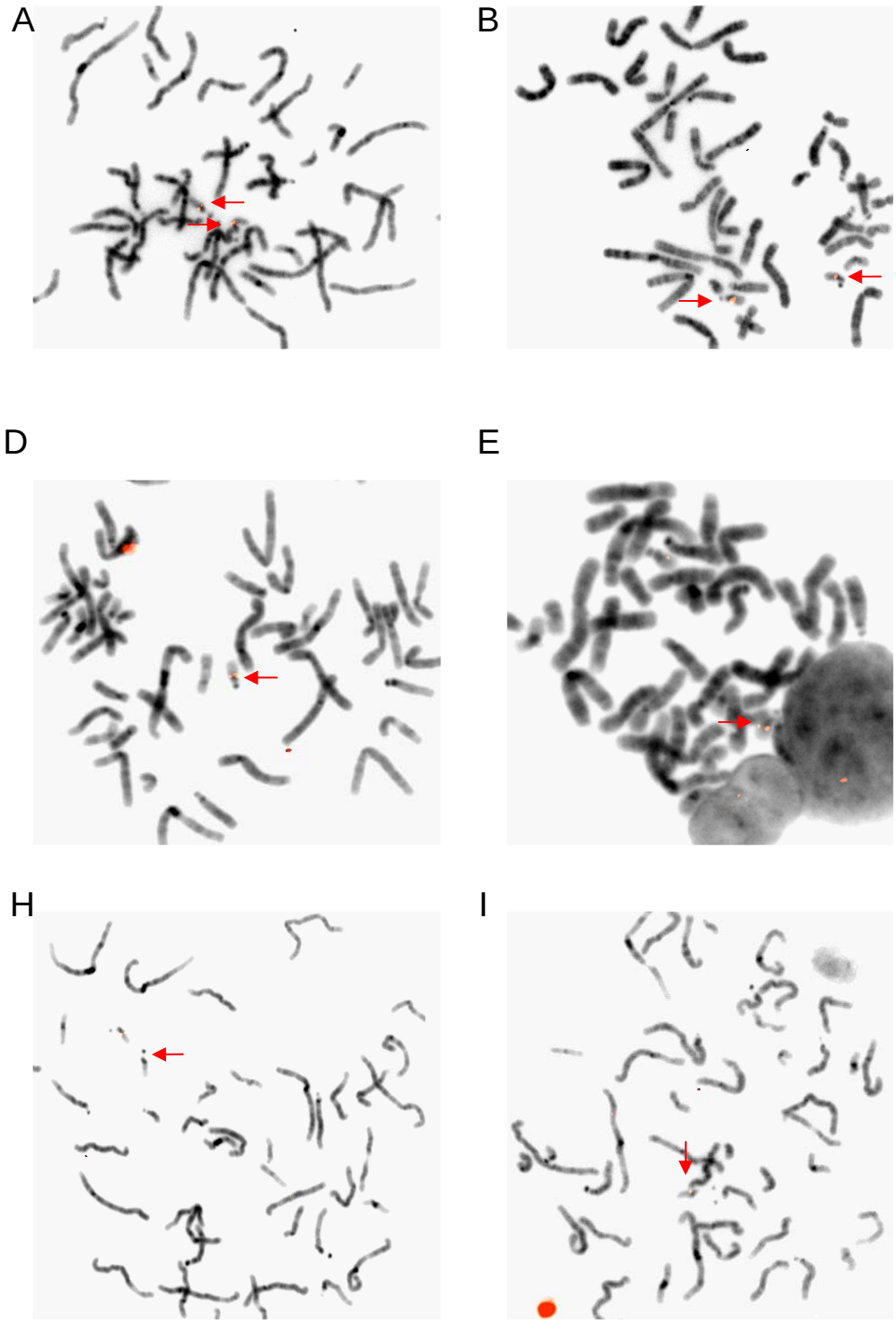
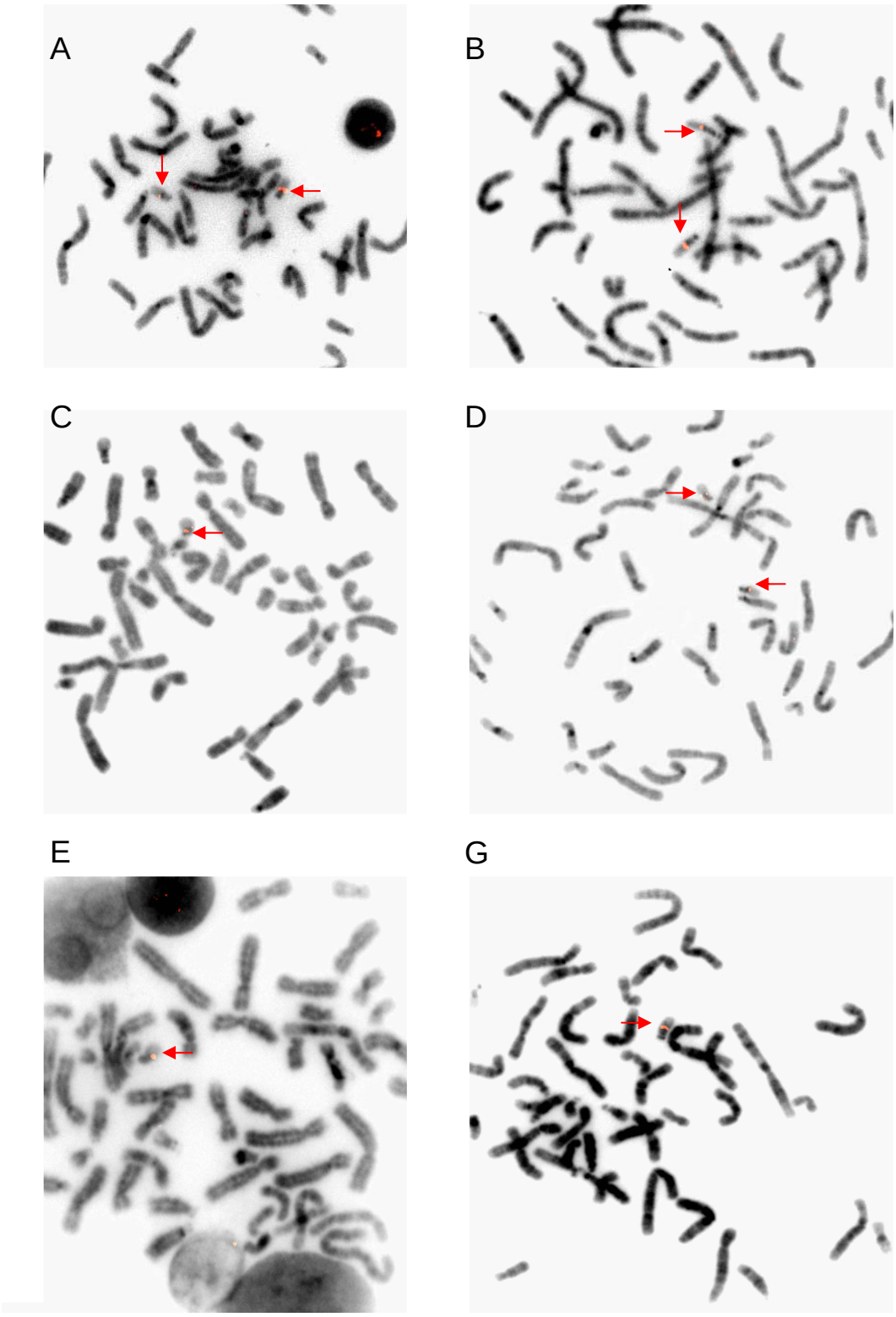


Figure 7.14: DiGeorge region probes hybridised to chromosomes isolated from patient 1. For clones used as the probe see Table 6.3 (letters correspond to clones used for hybridisation experiment.


A
B
C
D
E
G

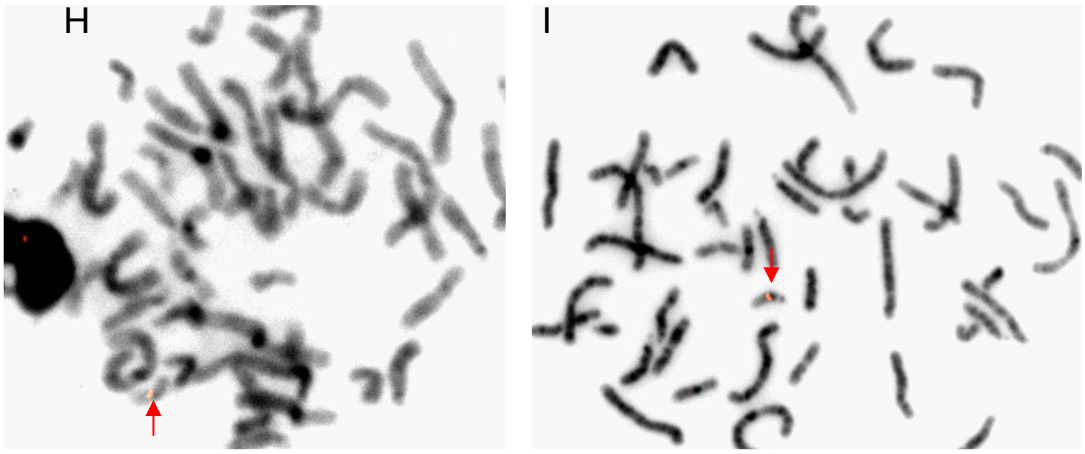


Figure 7.15: DiGeorge Region Probes hybridised to chromosomes isolated from patient 4

For clones used as the probe, see Table 7.3 (letters correspond to clones used for hybridisation experiment.

7.2.2: Assessment of patients with the DiGeorge phenotype that do not show a deletion in 22q by FISH analysis.

DNA from six patients that have aspects of the DiGeorge phenotype, but that had no deletion detected by conventional FISH, were applied to microarrays to characterise the patient DNA. Initially, patient DNA was hybridised to the 22q array to see if a deletion could be detected in the DiGeorge critical region that could not be detected by FISH. The DNA from the patient's blood was then hybridised to the 1Mb array for genome wide analysis to detect copy number changes elsewhere in the genome. The patient phenotypes are described in Table 7.4. The patient DNA samples were hybridised against pool DNA.

Table 7.4: The phenotype characteristics of patients showing some characteristics of DiGeorge syndrome, but with no 22q11 deletion when analysed by FISH.

Patient	Phenotype
1	Absent Thymus, bilateral cleft lip and pallet, tetralogy of fallot (heart defect), malformed ears, tracheoesophageal fistula, anomalous right subclavian artery, small testes, abnormal renal arteries, Arrinencphaly (absent optic tracts)
2	Facial dysmorphism, Coloboma (defect of the iris), Interrupted aortic arch, ventricular septal defect, atrial septal defect.
3	Ventricular septal defect, pulmonary atresia (obstruction of the pulmonary artery), cleft lip and palate, micropenis, undescended testes, hypoplastic scrotum, facial defects, thymus hypoplasia, deafness,
4	Nasal speech, nasal regurgitation, tetralogy of fallot, facial dysmorphism
5	Hypocalcaemia, aortic coarctation (heart defect), facial dysmorphism
6	Hypocalcaemia, interrupted aortic arch, low set ears, small mouth, interrupted aortic arch type B (heart defect)

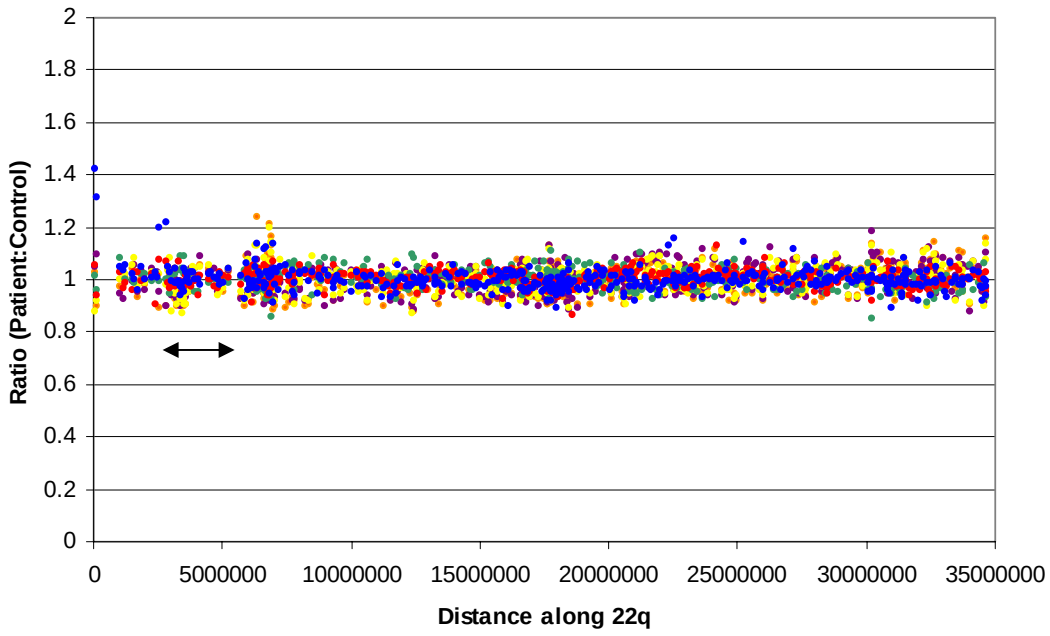


Figure 7.16: Patient:Control ratios obtained when six different patients are plotted against position on chromosome 22. Blue: Patient 1. Red: Patient 6. Green: Patient 3. Yellow: Patient 2. Purple: Patient 4, Orange: Patient 5. The DiGeorge region is indicated with a black arrow.

Patient 1 shows a copy number gain in the 2 clones adjacent to the centromere. However these two clones often show abnormal ratios (also reported in sections 4.3.3 and 7.3) so their elevation was not of note. The data was seen as being noisier at

around 6.5Mb along the q arm of chromosome 22. The standard deviation at this locus is 0.065 oppose to 0.043 along the rest of the chromosome arm. This coincides with the VJ recombination region of the immunoglobulin light chain λ region (Section 7.3). Clones bac519d21 and pac699j1 (located within the DiGeorge region) are also slightly elevated in one patient with ratios of 1.195 and 1.217 respectively.

CGH profiles of the six patients on the 1Mb array are shown in Appendix 10. Most loci on the graph that had been identified as containing gains or losses had already been identified as clones consistently reporting atypical ratios, as reported in Appendix 11. However some clones that did not consistently report atypical ratios were elevated or deleted. These are indicated in Table 7.5. No deletions were seen in the chromosome 22q clones on the 1Mb array.

Table 7.5: Clones showing amplification or deletion on the DiGeorge phenotype patients when analysed on the 1Mb array.

Patient	**Clones with a ratio >1.2**	**Clones with Ratio <0.8**
1	RP11-537N4 (Chr 19)	None
2	None	None
3	None	None
4	RP11-537N4 (Chr 19) RP11-383B4 (Chr 10)	CTD-2022G9 (Chr 5) RP11-412L4 (Chr 5) RP11-506H20 (Chr 5)
5	RP3-432E18 (Chr 12)	RP1-24K19 (Chr 21)
6	RP4-679K16 (Chr 1)	None

The chromosome 19 clone RP11-537N4 shows ratios elevated above 1.2 in two of the patients when hybridized against a female pool control. However this clone was also identified to contain segmental duplications, with interchromsomal duplications on chromosomes 11, 6 and 2 (Bailey, Gu et al. 2002), making the results obtained for this clone difficult to interpret.

Patient 4 shows a deletion across 3-4Mb of chromosome 5 (Figure 7.17). Three clones clearly showed deleted ratios (Table 7.6). One clone, RP11-92M7, proximal to the three deleted clones also shows a slightly reduced. This deletion was investigated further by our collaborators as discussed in section 7.4

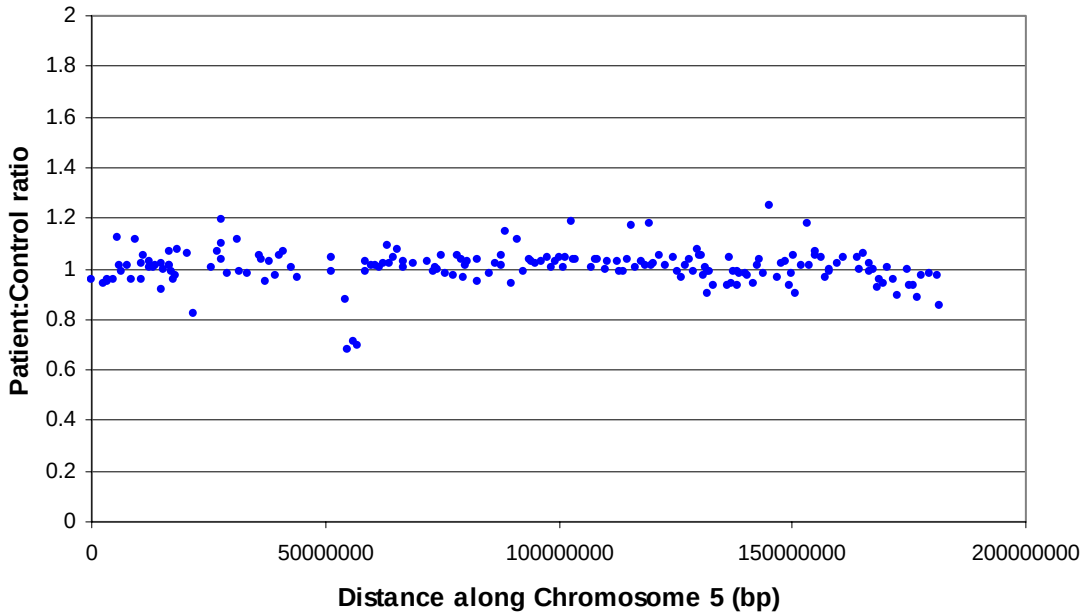


Figure 7.17: Deletion detected in chromosome 5 of patient 4 on the 1Mb array.

Table 7.6: The chromosome 5 clones deleted in patient 4

Clone	Chromosome	Position	Patient: Control
CTD-2022G9	5	54753069	0.68
RP11-506H20	5	56074899	0.71
RP11-412L4	5	57066879	0.70

The other gains and losses are also being investigated by our collaborators. Two of the six patients showed no gains or losses or deletions. This could be due to the fact that any deletions are not detected using an array of a 1Mb resolution or that the phenotype is not due to a DNA copy number change. Epigenetic changes in the genome may lead to the phenotypic effects observed. These would not be detected by the arrays.

7.3: Assessment of VJ recombination of the Immunoglobulin light chain λ using the 22q tile path array

The immunoglobulin light chain λ genes are located at approximately 6.5Mb along the q arm of chromosome 22. As lymphoblastoid cell lines are derived from differentiated B cells and are generally clonal, the immunoglobulin light chain λ genes will have undergone VJ recombination in these cells. As this leads to the

excision of DNA between the variable and the joining regions, clones on the arrays in this region will report these losses. Such changes were detected on the male: female control experiments reported in section 4.3.3. The VJ recombination in five different lymphoblastoid cell lines was assessed by hybridising DNA from the cell line against the pooled DNA from twenty anonymous blood donor samples (Fig 7.18).

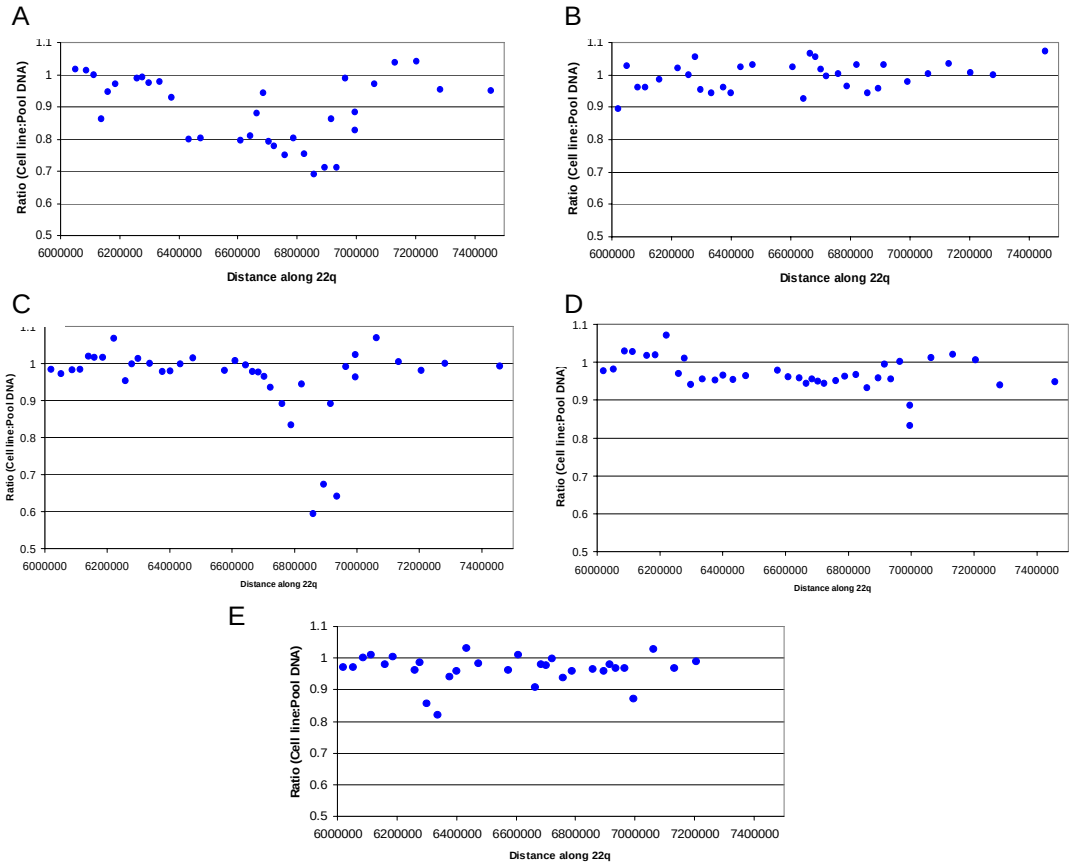


Figure 7.18: DNA from five lymphoblastoid cell lines with a normal karyotype were hybridised against DNA from a pool of 20 individuals A: Cell line HRC 575, B: Cell line HRC 146, C: Cell line HRC 159, D: Cell line HRC 160, E: Cell line HRC 196.

A deletion, defined using the <0.8 criteria described in 6.2.1, was seen in two out of five normal cell lines. The boundary of the deletion cannot be accurately defined. This is due to the large amount of segmental duplication within this region. The deletion in the HRC 575 cell line included clones cN22A12 – cN75C12 (midpoints 6433944-6995343) and represents a deletion of approximately 561Kb.

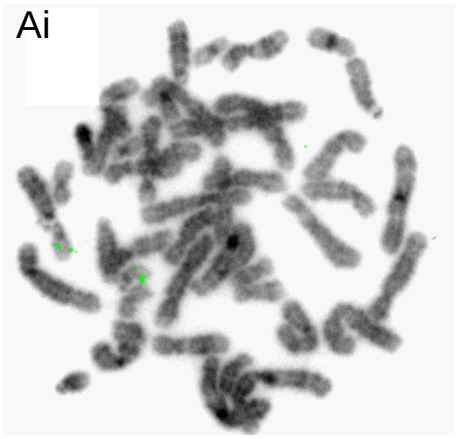
Comparison with the published map of the immunoglobulin light chain λ region allows determination of which constant and variable regions may be involved in the

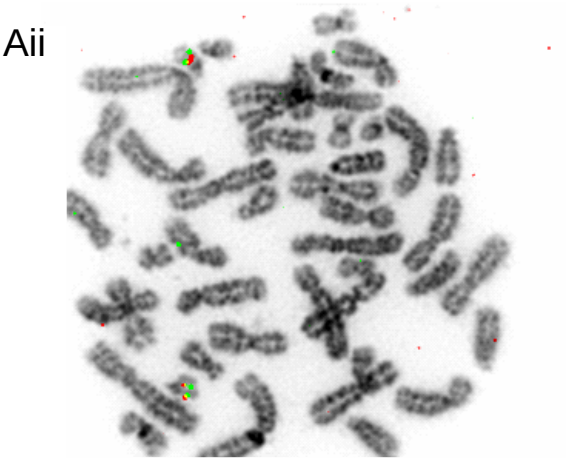
rearrangement (Kawasaki, Minoshima et al. 1997). The clone cN75C12 contains half the genes that encode the constant region of the IgL λ locus. It can therefore be determined that the constant region used is IgLC 3-7 (although IgLC 4,5, and 6 are known to be pseudogenes). The clone cN22A12 contains the IgLV genes 7-46, 5-45, 1-44 and 7-43. All IgLV genes more telomeric than this have been deleted. There was also a deletion in the cell line HRC 159. This deletion was smaller and covered approximately 76Kb between clones cN92H4 and cN9G6 (midpoints 6859162-6935464). The deletion mapped to between IgLC1-3 and IgLV 3-7. Again these clones contain a significant amount of segmental duplications so defining the exact size of the deletion was difficult. The other three cell lines show no deletion at this region, although there is slightly more background variation at the immunoglobulin light chain λ locus.

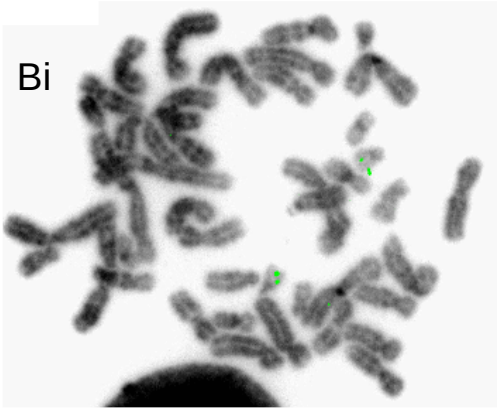
The deletion status at the immunoglobulin light chain λ locus was confirmed for two cell lines by FISH. HRC 575 showed a deletion, whereas the cell line HRC 160 did not. Biotin labelled FISH probes were made from the DNA from the same clones that were spotted onto the array and are shown in Table 7.7. A digoxygenin labelled control probe (bK57G9) from the non-deleted region of chromosome 22 was used to aid identification of chromosome 22. Selected images can be seen in Figure 7.19.

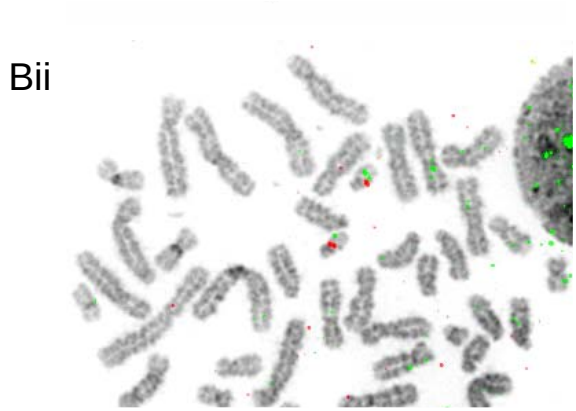
Table 7.7: Clones from the immunoglobulin light chain λ locus hybridised to metaphases from two different lymphoblastoid cell lines.

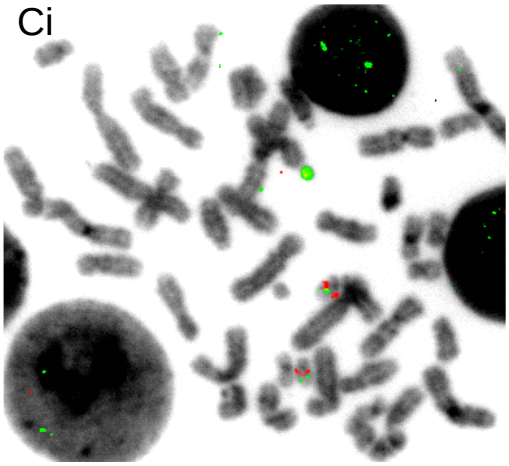
Clone	Accession no.	Signal on HRC 575	Signal on HRC 160
cN22A12	D86999	No	Yes
cN35B9	D87010	No	Yes
cN50D10	D87011	No	Yes
cN63E9	D87013	-	Yes
cN61E11	D87014	-	Yes
cN31F3	D87002	No	Yes
cN52F2	D87006	No	Yes
cN102D1	D86994	No	Yes
cN48A11	D87007	No	Yes
cN24A12	D86998	No	Yes
cN68D6	D87015	No	Yes
cN92H4	D87024	No	No
cN84E4	D87021	No	Yes
cN9C5	D87023	No	Yes
cN9G6	D87020	No	Yes
cN75C12	D87017	No	Yes
cN81C12	AP000360	Yes	Yes

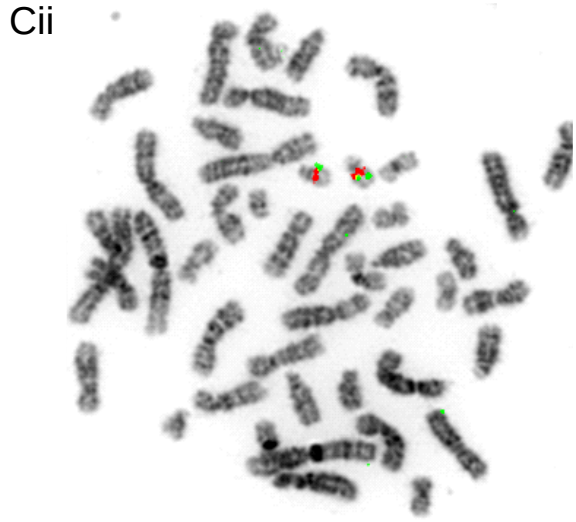

Ai


Aii


Bi


Bii


Ci


Cii

Di Dii

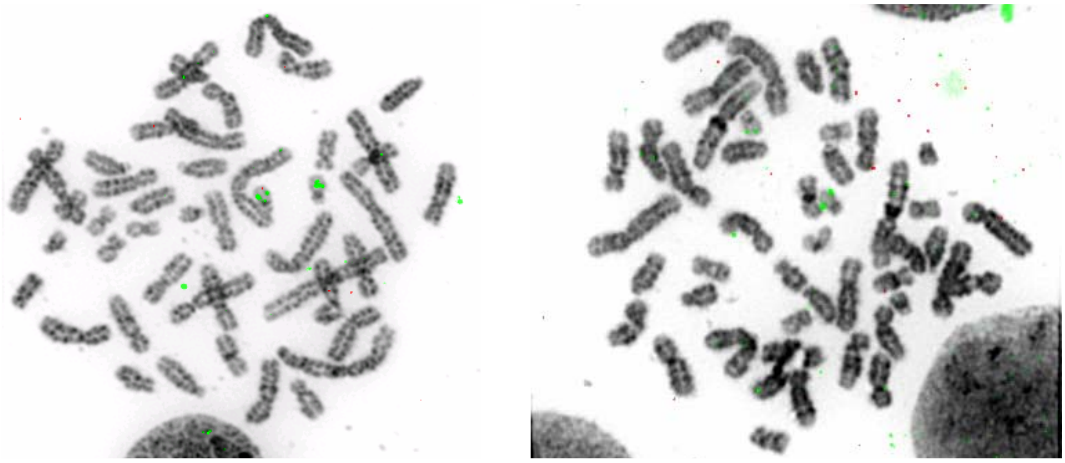


Figure 7.19: Fluorescence *in situ* hybridisation of clones from the immunoglobulin light chain λ locus (red) and a control probe (bK57G9 – green) to metaphases from the cell lines HRC 575 (i) and HRC 160 (ii). Letters relate to the clone used as the probe as reported in Table 7.8.

Table 7.8: Results from FISH experiments performed with clones from the VJ recombination region hybridised to metaphases from two different lymphoblastoid cell lines.

Hybridisation	**Clone**	**Result on HRC 575**	**Result on HRC 160**
A	cN68D6	deleted	present
B	cN75C12	deleted	present
C	cN81C12	present	present
D	cN92H4	deleted	deleted

The FISH analysis of HRC 575 cell line confirmed a deletion between cN68D6, and cN75C12, while no rearrangement was found in the HRC 160 cell line. cN81C12, was identified by arrays as being retained and distal to the HRC 575 deletion and was shown by FISH to be present in both cell lines. cN92H4 was found to be absent in both cell lines, despite showing a (non-deleted, although still reduced) ratio of 0.93 on the HRC 160 array.

7.4: Discussion

7.4.1: Segmental Duplications and the DiGeorge region.

A copy number loss was found in all the patients that had their DiGeorge status confirmed by FISH. However the reduced ratio observed for many clones rarely reached the 0.5:1 ratio that would indicate a full single copy number loss and deletion of one allele. This could be due to one of two explanations; either the clone is not fully deleted, or the clone DNA is cross hybridising with another region of the genome that is not deleted.

Examining the first possibility, the arrays have been shown to be quantitative (Fiegler, Gribble et al. 2003) and a deletion of only half a clone would report an intermediate ratio on the array. However it is unlikely that this is the reason for all the intermediate ratios seen as clones in the middle of the deleted region are affected. FISH analysis of some of these clones using metaphases from patients has shown that they are fully deleted on one copy of chromosome 22.

The second hypothesis is that non-deleted DNA from other regions of the genome are cross hybridising to the DNA on the microarray from the DiGeorge region. There is an abundance of segmental duplications in the 22q11 region of chromosome 22 (Figure 7.2) (Dunham, Shimizu et al. 1999; Bailey, Yavor et al. 2001; Bailey, Yavor et al. 2002). These repeated regions of the genome include intrachromosomal duplications, which exhibit homology to regions elsewhere on chromosome 22 and interchromosomal duplications which show homology to DNA sequences on other chromosomes. Duplications at the DiGeorge region can be seen in Figure 7.20.

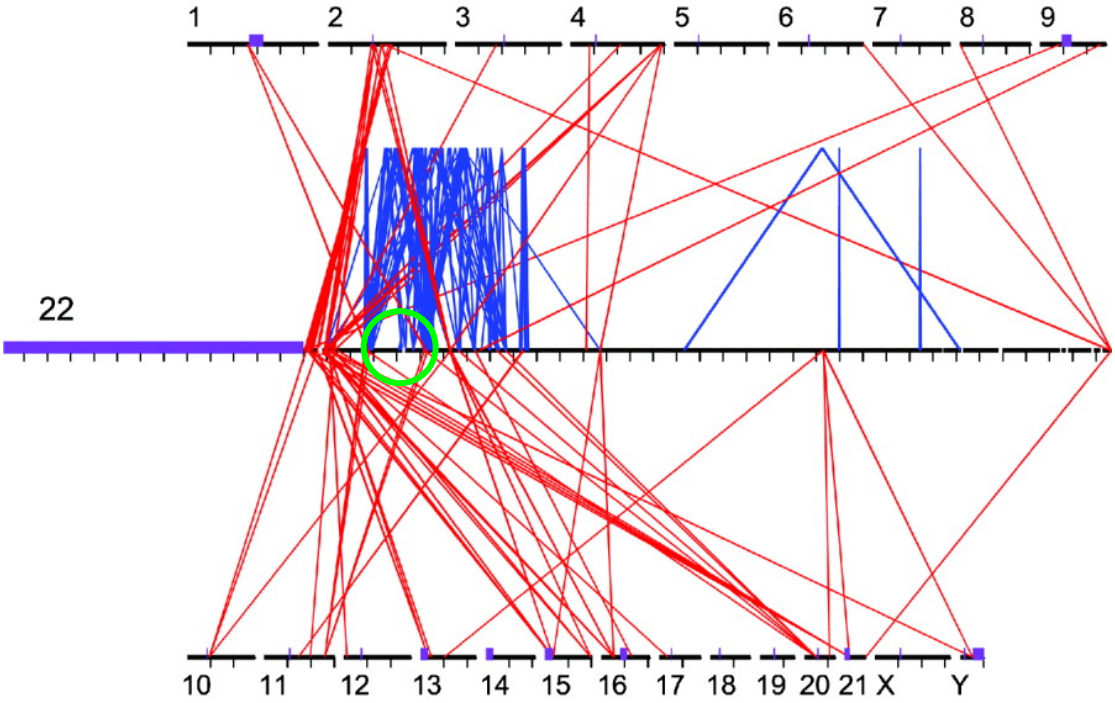


Figure 7.20: Segmental duplications on chromosome 22. Blue: Intrachromosomal deletions. Red: Interchromosomal deletions. DiGeorge region is indicated in green. Figure from (Bailey, Yavor et al. 2002).

In the experiments performed, whole genomic DNA from the DiGeorge patients is hybridised to the array. DNA that is not part of the deleted region, but has a high homology to clones within the deleted region will hybridise to these loci on the array. This would mask the single copy deletion that is characteristic of DiGeorge syndrome, and would explain the intermediate ratios exhibited.

Bioinformatic analysis of the clones in the DiGeorge region showed that many of these clones contained segmental duplications (Bailey, Yavor et al. 2002). Clones in the DiGeorge region such as pac699j1, pac995o6, bac519d21 and bac32 contain duplications in many other locations on 22q11. However, other clones in the DiGeorge region such as the cosmids 18c3, 111f11 and 119F4 contained no duplications and yet did not show a full single copy deletion on the array. However the bioinformatics approach to the detection of duplications may not be sufficient to find all regions of homology and is likely to underestimate the true amount of duplication (Eichler 2001). Most clones within the DiGeorge region contain a segmental duplication.

The chromosome 22 add-in experiment described in section 4.6 shows that the DNA from chromosome 22 clones spotted onto the array does not always show a full copy number change when an extra copy of chromosome 22 is added into the hybridisation mix. Cross hybridisation with other regions of the genome would mask the copy number change that occurs when an extra copy of chromosome 22 is added. If this is the case, then there should be a correlation between the slopes obtained from the add-in experiments and the ratios reported in the DiGeorge experiment. To test if this was so, the DiGeorge ratios reported were plotted against the slope obtained in the chromosome 22 add-in experiments, for clones in the DiGeorge region (Figure 7.21)

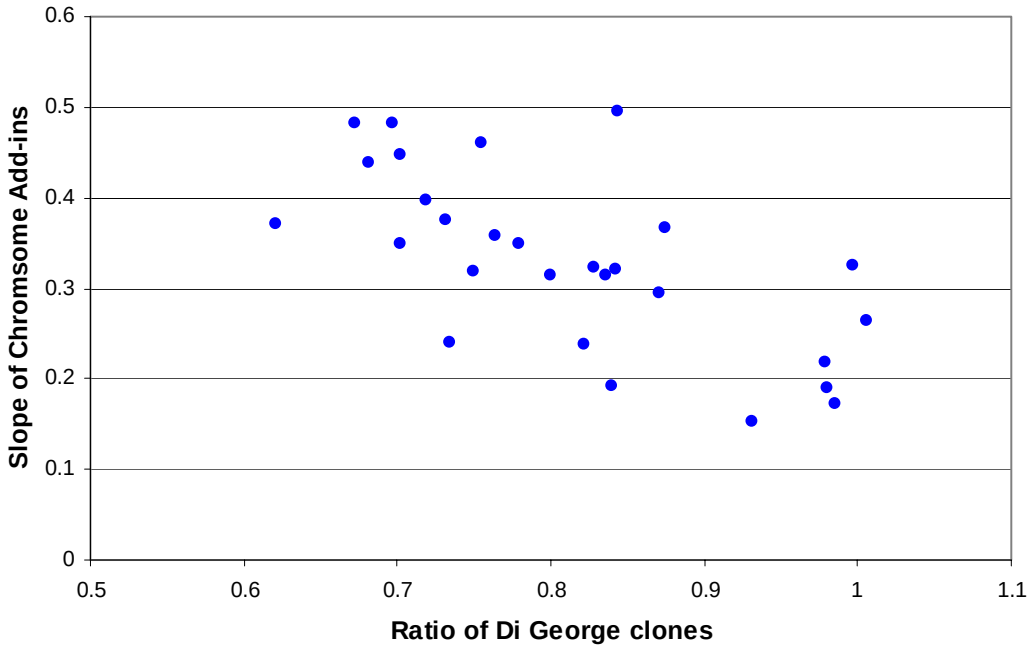


Figure 7.21: Correlation between DiGeorge ratios reported and the slope obtained from the chromosome 22 add-in experiments, for the clones in the DiGeorge region.

There is a negative correlation (regression coefficient 0.66) between the slope and the DiGeorge ratio. Clones that reported the largest slope (i.e. responded best to the add-in experiment) also reported the lowest DiGeorge ratio (those most consistent with a single copy deletion). Conversely, those clones that show a suppressed response to the chromosome 22 add-in experiments are those that show an incomplete single copy deletion ratio when hybridised with DiGeorge DNA. However the chromosome 22add-in experiment will not report the effect of intrachromosomal deletions. This analysis shows that the reason for the suppressed DiGeorge ratio is due to a

characteristic of the clone, and not due to differences in copy number within the deleted DiGeorge region.

It might be expected that the FISH experiments utilising clones with segmental duplications would show the regions with homology as secondary FISH signals. However, this was not seen, but it is unclear how the hybridisation kinetics of the arrays relate to the hybridisation kinetics of FISH. The DNA involved in the hybridisation is of different complexities, there are different relative amounts of Cot 1 present and different washing stringencies are used. Duplications would also have to be at least 2Mb apart to be resolved by metaphase FISH. As seen in Figure 7.20, most of the intrachromosomal duplications at the DiGeorge locus are not more than 2Mb apart. It is therefore unsurprising that segmental duplications have different consequences for arrays and metaphase FISH experiments.

Many of the chromosomal micro-deletion syndromes occur within regions of segmental duplication. It is therefore likely that underestimation of a full single copy loss on arrays will not be a characteristic unique to the DiGeorge region. By understanding how the duplications affect the arrays, their effects could be subtracted from the ratios obtained. This can be achieved in one of two ways;

Firstly, the utilisation of degenerate oligonucleotide primers to amplify the clone DNA ensures that full coverage of the clone DNA is represented on the array, including any segmental duplication contained within the clone DNA. To avoid the effect of segmental duplication on the ratios obtained on the array, the whole segmental duplication can be removed from the array by using an alternative strategy to amplify the clone DNA. Specifically designed PCR primers can be used to amplify all unique sequences within a clone, without amplifying segmental duplications or common repeat elements (Buckley, Mantripragada et al. 2002). In this way, segmental duplications would not be present on the array, and therefore would have no influence on the ratios reported. In addition, the removal of repeat elements would also improve the quantitation of DNA ratios reported by arrays. A secondary advantage may be a reduction in the amount of Cot 1 needed in the hybridisation mix, therefore reducing hybridisation costs. The drawback of this approach is that the design and production of individual PCR primers used for the amplification of the clone DNA is much more

expensive than using a universal primer to amplify all clone DNA. The removal of segmental duplications from the array also means the whole genome will not be covered. As many microdeletions and chromosomal breakpoints occur within segmental duplication regions, the removal of duplicate regions from arrays may limit their use in investigating these phenomena.

A second way of removing the effect of segmental duplications on the array ratio is to address the problem using a bioinformatic approach. Since the publication of the draft human genome sequence, work has been underway to map segmentally duplicated regions in the human genome. This has been achieved by identifying sequences that are over-represented in the Celera shotgun sequence and mapping them back against the draft sequence, or by repeat-masking regions of the sequence (to remove common repeat elements) and performing a global BLAST comparison with the rest of the genome (Bailey, Yavor et al. 2001; Bailey, Gu et al. 2002). By correctly identifying the amount of segmental duplication present in each clone, the number of duplications present elsewhere in the genome, and the degree of homology required for cross hybridisation, it should be possible to predict what ratio a clone containing a segmental duplication will produce on the array if a deletion or amplification is present. However the arrays will not be as sensitive when detecting single copy number changes. Currently, a problem with this analysis is that much of the genome sequence is still present in a draft form. Misalignment of the genome at duplicated regions in the draft sequence (as reviewed by Eichler (Eichler 2001)) will underestimate the amount of segmental duplications in the genome, and so make any correlation between reported array ratio and duplication inaccurate. Once all chromosomes have been sequenced to a ‘finished’ status these misalignments will be minimal and correlations between segmental duplication and the ratio reported by arrays should be possible. Work is currently underway in our group to correlate repeat content with ratios reported by the arrays. In summary the chromosome 22q tile path array can be used to detect DiGeorge deletions. However the presence of segmental duplications can make interpretation difficult and their presence should be taken into account when analysing these arrays.

7.4.2: Analysis of Patients showing the DiGeorge phenotype with no 22q11 deletion.

DNA from patients showing a DiGeorge phenotype but with no deletion in the DiGeorge critical region were analysed on the chromosome 22q tile path and the 1Mb arrays. Results from the 22q tile path arrays show that there was no deletion seen in any of these patients at the DiGeorge critical region. The DNA from all six patients was applied to the 1Mb array. One patient (patient 4) showed a 3-4Mb single copy deletion at 5q11.2. This patient's karyotype had previously been examined cytogenetically and no deletion had been detected on chromosome 5. This demonstrates that the arrays are more sensitive at detection of deletion than conventional cytogenetic methods.

Further studies were performed at the Institute of Child Health to confirm the results obtained by the array. Seven microsatalite markers analysed across the region were found to be homozygous, supporting the observation that this region has a single copy deletion. FISH analysis using selected clones confirmed the deletion (Table 7.9).

Table 7.9: Results from FISH experiments performed on metaphase chromosomes from patient 4.

Clone	**Chromosome**	**Position**	**Array ratio**	**FISH Results**	**FISH Comments**
RP11-497H16	5	21747447	0.82	present	Cross hybridises elsewhere on 5p and 5q
RP11-269M20	5	51407665	0.99	present	Secondary on chr 1
RP11-92M7	5	54387267	0.87	deleted	-
RP11-506H20	5	56074899	0.71	deleted	-
RP4-572A3	5	58515147	1.02	present	-

7.4.3: Analysis of the Immunoglobulin light chain λ recombination region.

During B cell development the loci encoding the immunoglobulin light chain (IgL) undergo rearrangement to produce antibody diversity. The IgL has 2 different classes;

κ and λ. The κ chain is encoded at 2p11.2 and the λ chain is encoded at 22q11. The rearrangement can occur on either allele at the κ of λ loci. Production of a functional protein initiates a feedback mechanism once a successful rearrangement has occurred; the other alleles are epigenetically silenced and not rearranged (Gorman and Alt 1998).

In Figure 7.11, a single copy deletion at 22q11 can be seen in two of the five cell lines analysed (HRC 575 and HRC 160) reflecting that the IgL λ has undergone rearrangement in these cell lines. The two rearranged cell lines studied show deletions of different sizes. HRC 575 has a large deletion of approx 560Kb. HRC 159 has a smaller 76Kb deletion. The clones involved in the deletion contain segmental duplications and so show incomplete reduction in ratios from those expected for a single copy loss. This makes defining the boundaries of the deletion difficult. However it can be seen that the two deleted cell lines do not share proximal or distal breakpoints and therefore it is likely that different V and J segments have been fused during recombination in these cell lines. Comparison of the deleted region with the sequence map (Kawasaki, Minoshima et al. 1997) indicates which constant and variable regions are involved in IgL λ rearrangements.

The other three cell lines do not show any rearrangement in the immunoglobulin light chain λ locus. This could be due to several different reasons. Firstly, the rearrangement and associated deletion may be too small to detect on the tile path array. Secondly, during B cell development it is the immunoglobulin heavy chain (IgH) on chromosome 14 that undergoes rearrangement first. Only one in three of these rearrangements are successful (Bassing, Swat et al. 2002). If IgH rearrangement is unsuccessful on both alleles the IgL will be prevented from rearrangement and no rearrangement will be seen at either the IgL κ or λ loci. Lastly, in humans IgL κ rearrangement occurs before IgL λ rearrangement (Nemazee and Weigert 2000). If either rearrangement at the IgL κ loci is successful, negative feedback by the transcribed IgL will prevent rearrangement at the IgL λ locus.

FISH was performed on two of the cell lines studied; one (HRC 575) showed a large deletion due to recombination of disparate V and J segments at one allele, the other

(HRC 160) showed no deletion. The FISH results, in the main, confirm the results obtained by the array analysis of the immunoglobulin light chain λ locus. The only clone that showed disparate results was cN92H4, which was absent in both cell lines when analysed by FISH. The ratio reported for this clone in the hybridisation using HRC 160 DNA (0.92) is not outside that expected due to experimental variation, and therefore it is not classed as deleted, however it is lower then the other clones in the immediate vicinity. This clone does include regions of segmental duplication. The deletion that would be reported on the array at this location may have been masked by the cross hybridisation of other regions of the genome, as described in 7.4.1.

These experiments show how the arrays can detect physiological rearrangements of the genome. Higher resolution arrays may allow the exact constant and variable genes rearranged in the B cells to be determined. Physiological rearrangements can be specific to one cell type. For example, IgL rearrangement is specific to B cells, whilst rearrangement of the T cell receptor family is specific to T cells (Turner 2001) and therefore should be taken into account when the type of DNA being used as control is decided. Knowledge of these regions means false deletions and amplifications are not detected as described in section 7.2.1.

7.4.4. Summary

This Chapter has shown how DNA microarrays can assess microdeletions at a much higher resolution than conventional cytogenetic techniques. They also give much more information about the size of the deletion than assessment of specific syndromes by commercial probes. The analysis of patient samples on a genome wide array enables detection of copy number gains and losses that might be missed if just one region of the genome was being screened.

8: Conclusions

8.1: Construction and Validation of the Chromosome 22 arrays.

In Chapter 4 I have described the construction of a 22q array from overlapping tile path clones. This was performed using the method described by Fiegler et al (Fiegler, Carr et al. 2003). Briefly, clone DNA was amplified using 3 different Degenerate Oligonucleotide Primers. This ensured complete representation of clone DNA on the array and made certain a minimal quantity of *E. Coli* DNA was amplified and represented.

Array verification experiments were performed to ensure that each locus on the array reported the correct copy number change. Self:self and male:female hybridisations on the 22q array revealed little variation in the ratios reported on the 22q clones. The standard deviations reported by the clones on the array were very small (0.04 and 0.09 respectively).

Further verification was performed by spiking a G1:G1 hybridisation with different copy numbers of chromosome 22. These experiments reported that 96% of the clones on the 22q tiling path array reported the expected response to the additional copies of chromosome 22.

The chromosome 22 tile path arrays were also used to detect microdeletions on chromosome 22. DNA from patients with confirmed DiGeorge syndrome was hybridised to the array. Clones on the array reported a reduced patient:control ratio; however ratios indicative of a full single copy deletion were not detected by most clones within the DiGeorge region. This is due to the large regions of segmental duplication within 22q11 which results in cross hybridisation with other regions of the genome. This masks the full single copy deletion.

Copy number change was also seen at the Immunoglobulin light chain λ locus which is also located in 22q11. Deletion at this locus due to VJ recombination can be

detected on the 22q array in lymphoblastoid cell lines. A deletion was detected in two of the five lymphoblastoid cell lines examined, however identification of a full single copy deletion is masked by the presence of segmental duplications at the immunoglobulin light chain λ locus

Chromosome X clones present on the array also show a reduced copy number in response to a male:female hybridisation. This verification indicated thatthe chromosome 22q tile path arrays are suitable for detecting the small copy number changes required for assaying replication timing.

An array was also constructed using 500bp PCR products. A 4.5Mb region located 15.5–20Mb along 22q was represented at a 10Kb resolution. A further 200Kb region 16.495-16.695Mb along chromosome 22q was represented with overlapping 500bp product arrays. This array was designed so that no duplicated regions were represented on the array. Products from chromosome X were also spotted to allow verification of copy number change. The standard deviation of results from a self:self hybridisation for the chromosome 22 500bp PCR products on the array was 2.5 times the standard deviation of the chromosome 22 clones represented on the tile path array. The chromosome X clones showed a single copy deletion, however the standard deviation reported by the X loci was also high. This excessive noise may be due to the considerably reduced Cy3 and Cy5 intensities obtained from these high resolution arrays. As a result these arrays will be inaccurate when reporting copy number change, especially when considering the accuracy needed to assay replication timing on arrays. Time limitations did not allow optimisation of the array to reduce the noise.

Further development of the high resolution PCR product array may improve the poor signal:background ratio that has been caused by the low intensities obtained from scanning these arrays. Accurate reporting of copy number change at this resolution will allow a greater in-depth analysis of replication timing and its correlation with other genomic features.

8.2: The use of Genomic Microarrays to assess Replication Timing.

This work has described for the first time replication timing assayed at a high resolution over a whole mammalian genome. Replication timing has been assayed using an array covering the whole euchromatic human genome at a 1Mb resolution. Furthermore this approach has been extended to examine replication timing of three chromosomes (chromosomes 1, 6 and 22) at a tile path resolution giving an unprecedented view of the detailed patterns of replication timing.

The method used to assess replication timing was developed on an array spanning 4.5Mb of chromosome 22 using the clone DNA detailed in Chapter 3. This verified that the method could detect the subtle copy number changes required for the assay of replication timing. The initial replication timing pilot studies described in Chapter 3 reported ratios within the expected 1:1 and 2:1 boundaries. In summary, late replicating regions were located within a G dark chromosome band, and were GC poor; conversely early replicating DNA was located within a G light band and was GC rich. This confirms previous reports (Tenzen, Yamagata et al. 1997; Watanabe, Fujiyama et al. 2002).

Expansion of these studies onto the 1Mb resolution genome wide array and the individual chromosome tile path arrays also reported ratios between the theoretical maximum and minimum of 2:1 and 1:1. Replicate experiments on the arrays showed this method to be highly reproducible. The average coefficient of variation between four replicate experiments on the 1Mb array was 5.5%. The tile path arrays also reported highly reproducible data.

Replication timing data obtained from the arrays was compared with previously published replication timing data. The 1Mb resolution data produced for chromosome 11q was compared to published data assaying 11q using PCR on flow sorted S phase fractions (Watanabe, Fujiyama et al. 2002). The correlation between the two replication timing profiles produced was strong; slight differences ($r = 0.69$) were likely to be due to the different cell types used. A change in replication timing between the MHC class II region and MHC class III region on chromosome 6

previously published (Tenzen, Yamagata et al. 1997) was also confirmed when assaying replication timing on the chromosome 6 tile path array.

Microarray technology has previously been used to assess replication timing on two other organisms. cDNA arrays have been used to assay replication timing of *Saccharomyces cerevisiae* (Raghuraman, Winzeler et al. 2001) and *Drosophila melanogaster* (Schubeler, Scalzo et al. 2002). However, due to the type of arrays used, these studies only allow the assay of replication timing of coding regions of the genome. Large amounts of the human genome are non-coding and therefore if replication timing was assayed on a human cDNA array, large regions of the genome such as the gene deserts on chromosomes 13 and 14, would remain unanalysed (IHGSC 2001). The use of genomic DNA from large insert clones ensures replication timing is assayed for both coding and non-coding DNA. The tile path arrays contain DNA from overlapping sequencing clones so that all sequenced coding and non-coding DNA for each chromosome is represented. The unbiased representation of coding and non-coding DNA on the arrays means that, for the first time, correlations between replication timing and sequence features of the genome can be calculated.

8.3: Large scale analysis of the correlation between replication timing and other features of the genome.

Published data has reported links between replication timing and other features of the human genome sequence as reported in Section 1.4. Due to the limitations of conventional methods of assaying replication timing, these associations have only been observed over small regions of the genome. Large scale analysis of the whole human genome at a 1Mb resolution reveals positive correlations between replication timing and GC content, gene density, *Alu* repeat density and probability of gene expression. A negative correlation is observed with LINE repeat density. These correlations were also seen on the individual chromosome tile path arrays, although the correlations of genome features with replication timing are weaker.

One problem with looking at the correlations in this way is that the sequence features of the genome such as GC content, sequence repeat density and transcriptional

activity correlate with each other as well as replication timing. Because of this it is difficult to establish which features have an influence on replication timing, and which features show a correlation as a secondary effect.

In summary, strong correlations were observed between replication timing and sequence features of the genome, especially when replication timing was averaged over the whole chromosome. The genome features that have been shown to associate with early replication, such as a high GC content, abundance of *Alu* repeat elements and transcriptional activity, are those that are associated with active euchromatin. Conversely features that correlate with late replication are those associated with inactive chromatin, such as AT rich DNA and a lack *Alu* repeats. Multiple regression analysis suggests several of these features have a combined affect on replication timing.

The correlation between replication timing and transcriptional activity is controversial. No correlation between replication timing and gene expression level was seen in a genome wide study of yeast produced by Raghuraman *et al* (Raghuraman, Winzeler et al. 2001). My observations on the human genome showed a correlation between replication timing and transcriptional activity. This was determined by experimentation on genomic and Affymetrix arrays. Analysis on the human genome is comparable to what was observed in *Drosophila* (Schubeler, Scalzo et al. 2002), where a correlation was seen with probability of transcription, but not expression level. The correlation with probability of gene expression also supports previously determined models linking early replication to gene transcription (Cook 1999; Gilbert 2002).

8.4: Future Work.

The work described in this thesis has shown how the replication timing of lymphoblastoid cell lines can be assayed on arrays constructed from genomic clone DNA. However this is only the beginning of the potential of the arrays to assay epigenetic features of the genome.

8.4.1: Optimisation of the high resolution PCR product array.

The high resolution array constructed from 500bp PCR products (described in Sections 4.5 and 5.4) reported high standard deviations when control experiments were performed. S:G1 hybridisations ratio reported by the arrays described several loci outside the theoretical boundaries of the experiment (1:1 – 2:1). This is likely to be due to the low Cy3 and Cy5 intensities detected from these arrays. The intensities reported were up to 100 times less than those reported by clone DNA arrays, therefore the signal:background ratio of each spot on the array is lower. Further development of the array to increase the spot signal intensity and increase the signal: background ratio would make the reporting of replication timing at this high resolution more accurate. Ways of achieving this may be to increase the concentration of the DNA spotted onto the array, or to decrease the amount of unlabelled herring sperm and Cot 1 DNA and yeast tRNA applied to the array during the pre-hybridisation and the hybridisation steps.

The accurate reporting of replication timing by arrays at a high resolution would be an important next step. Once the arrays have been optimised they can be used for a variety of applications;

- Fine mapping of replication timing, possibly allowing the mapping of replicon boundaries.
- Location of replication origins. This could be achieved by hybridisation of short nascent DNAs to the array to map regions containing replication origins.
- Mapping of epigenetic features (as described in Section 8.4.4). Correlations between replication timing and other epigenetic features at a high resolution would allow a greater understanding of the links between individual features of the genome.

8.4.2: The assay of replication timing within other tissues and cell lines.

To date, the replication timing of only human lymphoblastoid cells has been assayed on the genomic arrays. Tissue specific genes are early replicating in the tissues in

which they are expressed, but generally late replicating in other tissues (Hatton, Dhar et al. 1988). The assay of replication timing in other tissue types would allow this process to be investigated.

A fibroblast cell line was grown and flow sorted in preparation for assay of replication timing. However the purity of the sort made the DNA unsuitable for application to the array. Accurate separation of S and G1 phase nuclei in other cell lines and cell populations would allow the assay to be successful. The method described in this thesis separated G1 and S phase nuclei based on the Hoechst staining and therefore their DNA content. By labelling nascent DNA with BrdU, cells can be sorted by propidium iodide and BrdU intensity (Ormerod 2000). This would allow the more accurate sorting into G1 and S phase. Once the nuclei separation has been optimised this method can be used to assay replication timing in many different dividing cell types and during many different stages of their development.

8.4.3. Investigation of gene expression at regions which undergo changes in replication timing.

The assay of replication timing in a lymphoblastoid cell line with a translocation between chromosome 17 and 22 revealed several regions where the replication timing deviates from that seen in a normal cell line (Section 6.4). As transcriptional activity has been linked to replication timing it may be possible that the change in replication timing results in a change in transcriptional activity. This could be assayed by the application of RNA, extracted from the t(17;22) lymphoblastoid cell line, to the Affymetrix U133a array. As no genes at the translocation breakpoints were disrupted (S. Gribble, personal communication) it is possible that the phenotype exhibited by the patient is due to a change in transcriptional activity of the genome, resulting from the translocation previously identified.

8.4.4. Investigation of other epigenetic features on the arrays.

Studies reported in this thesis have shown how the genomic arrays can be used to assay copy number change (Chapter 7). Preliminary studies in this report also show how genomic arrays can be used to assess epigenetic features of the genome (Section 6.3). Histone acetylation was assayed on the chromosome 22 tile path array, and regions of clear difference in the acetylation status of the genome were identified.

DNA-protein interactions can be assayed by applying immunoprecipitated chromatin material to a genomic array (ChIP on CHIP). This technique involves the *in vivo* cross-linking of protein-DNA complexes and shearing of the DNA to produce small fragments. Specific protein-DNA interactions can then be purified using an antibody against the protein of interest, the cross-links are then reversed, and the protein removed from the DNA sample. The DNA can be labelled and co-hybridised to the array using differentially labelled input DNA as a control. Histone modifications, such as acetylation, methylation, phosphorylation and ubiquitination could be assayed in this way using both clone DNA arrays and high resolution PCR product arrays. Binding patterns of DNA-associate proteins, such as those involved in the origin recognition complex could also be investigated (van Steensel and Henikoff 2003). The methylation of CpG dinucleotides at CpG islands can also be investigated on a genome wide basis (Yan, Chen et al. 2001) utilising microarrays in this way.

These proposals illustrate how the genomic microarrays constructed for this thesis can be used for the investigation of many more features of the genome. Further large scale analysis of replication timing in other tissues in various developmental stages, will allow understanding of transcriptional changes within tissues. This method could also be applied to other model organisms, such as the mouse. This will allow the study of replication timing within tissues that cannot be obtained from humans, such as those in the developing foetus. Studies in animals at different developmental stages may reveal changes in replication timing, providing insights into the control of transcriptional activity during the progression of an organism to maturity.

8.5: Conclusions

In this thesis, I have described how replication timing has been assessed on a genome wide basis. Microarrays allow rapid analysis of replication timing for large regions of the genome. The microarrays produced as described in this thesis are also a valuable tool for the study of other epigenetic features of the genome and DNA copy number changes associated with cancer and microdeletion syndromes.

Replication timing was correlated with several other features of the genome. This may explain what determines whether a piece of DNA will be early or late replicating. The follow-up experiments described in Section 8.4 will allow further investigation of the interplay between replication timing and other sequence or epigenetic features of the genome.

References

Abdurashidova, G., M. Deganuto, et al. (2000). "Start sites of bidirectional DNA synthesis at the human lamin B2 origin." Science **287**(5460): 2023-6.

Albertson, D. G. (2003). "Profiling breast cancer by array CGH." Breast Cancer Res Treat **78**(3): 289-98.

Albertson, D. G., B. Ylstra, et al. (2000). "Quantitative mapping of amplicon structure by array CGH identifies CYP24 as a candidate oncogene." Nat Genet **25**(2): 144-6.

Amiel, A., T. Litmanovitch, et al. (1998). "Temporal differences in replication timing of homologous loci in malignant cells derived from CML and lymphoma patients." Genes Chromosomes Cancer **22**(3): 225-31.

Avner, P. and E. Heard (2001). "X-chromosome inactivation: counting, choice and initiation." Nat Rev Genet **2**(1): 59-67.

Azuara, V., K. E. Brown, et al. (2003). "Heritable gene silencing in lymphocytes delays chromatid resolution without affecting the timing of DNA replication." Nat Cell Biol **5**(7): 668-74.

Bailey, J. A., Z. Gu, et al. (2002). "Recent segmental duplications in the human genome." Science **297**(5583): 1003-7.

Bailey, J. A., A. M. Yavor, et al. (2001). "Segmental duplications: organization and impact within the current human genome project assembly." Genome Res **11**(6): 1005-17.

Bailey, J. A., A. M. Yavor, et al. (2002). "Human-specific duplication and mosaic transcripts: the recent paralogous structure of chromosome 22." Am J Hum Genet **70**(1): 83-100.

Bannister, A. J., P. Zegerman, et al. (2001). "Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain." Nature **410**(6824): 120-4.

Bassing, C. H., W. Swat, et al. (2002). "The mechanism and regulation of chromosomal V(D)J recombination." Cell **109 Suppl**: S45-55.

Bickmore, W. A. and P. Teague (2002). "Influences of chromosome size, gene density and nuclear position on the frequency of constitutional translocations in the human population." Chromosome Res **10**(8): 707-15.

Blackwood, E. M. and J. T. Kadonaga (1998). "Going the distance: a current view of enhancer action." Science **281**(5373): 61-3.

Boggs, B. A. and A. C. Chinault (1997). "Analysis of DNA replication by fluorescence in situ hybridization." Methods **13**(3): 259-70.

Buckley, P. G., K. K. Mantripragada, et al. (2002). "A full-coverage, high-resolution human chromosome 22 genomic microarray for clinical and research applications." Hum Mol Genet **11**(25): 3221-9.

Carlson, C., H. Sirotkin, et al. (1997). "Molecular definition of 22q11 deletions in 151 velo-cardio-facial syndrome patients." Am J Hum Genet **61**(3): 620-9.

Chess, A., I. Simon, et al. (1994). "Allelic inactivation regulates olfactory receptor gene expression." Cell **78**(5): 823-34.

Cheung, J., X. Estivill, et al. (2003). "Genome-wide detection of segmental duplications and potential assembly errors in the human genome sequence." Genome Biol **4**(4): R25.

Cimbora, D. M., D. Schubeler, et al. (2000). "Long-distance control of origin choice and replication timing in the human beta-globin locus are independent of the locus control region." Mol Cell Biol **20**(15): 5581-91.

Cohen, S. M., E. R. Cobb, et al. (1998). "Identification of chromosomal bands replicating early in the S phase of normal human fibroblasts." Exp Cell Res **245**(2): 321-9.

Cook, P. R. (1994). "RNA polymerase: structural determinant of the chromatin loop and the chromosome." Bioessays **16**(6): 425-30.

Cook, P. R. (1995). "A chromomeric model for nuclear and chromosome structure." J Cell Sci **108**(Pt 9): 2927-35.

Cook, P. R. (1999). "The organization of replication and transcription." Science **284**(5421): 1790-5.

Craig, J. M., S. Boyle, et al. (1997). "Scaffold attachments within the human genome." J Cell Sci **110**(Pt 21): 2673-82.

Cremer, M., J. von Hase, et al. (2001). "Non-random radial higher-order chromatin arrangements in nuclei of diploid human cells." Chromosome Res **9**(7): 541-67.

Cremer, T. and C. Cremer (2001). "Chromosome territories, nuclear architecture and gene regulation in mammalian cells." Nat Rev Genet **2**(4): 292-301.

Cross, S. H., V. H. Clark, et al. (2000). "CpG island libraries from human chromosomes 18 and 22: landmarks for novel genes." Mamm Genome **11**(5): 373-83.

References

Abdurashidova, G., M. Deganuto, et al. (2000). "Start sites of bidirectional DNA synthesis at the human lamin B2 origin." Science **287**(5460): 2023-6.

Albertson, D. G. (2003). "Profiling breast cancer by array CGH." Breast Cancer Res Treat **78**(3): 289-98.

Albertson, D. G., B. Ylstra, et al. (2000). "Quantitative mapping of amplicon structure by array CGH identifies CYP24 as a candidate oncogene." Nat Genet **25**(2): 144-6.

Amiel, A., T. Litmanovitch, et al. (1998). "Temporal differences in replication timing of homologous loci in malignant cells derived from CML and lymphoma patients." Genes Chromosomes Cancer **22**(3): 225-31.

Avner, P. and E. Heard (2001). "X-chromosome inactivation: counting, choice and initiation." Nat Rev Genet **2**(1): 59-67.

Azuara, V., K. E. Brown, et al. (2003). "Heritable gene silencing in lymphocytes delays chromatid resolution without affecting the timing of DNA replication." Nat Cell Biol **5**(7): 668-74.

Bailey, J. A., Z. Gu, et al. (2002). "Recent segmental duplications in the human genome." Science **297**(5583): 1003-7.

Bailey, J. A., A. M. Yavor, et al. (2001). "Segmental duplications: organization and impact within the current human genome project assembly." Genome Res **11**(6): 1005-17.

Bailey, J. A., A. M. Yavor, et al. (2002). "Human-specific duplication and mosaic transcripts: the recent paralogous structure of chromosome 22." Am J Hum Genet **70**(1): 83-100.

Bannister, A. J., P. Zegerman, et al. (2001). "Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain." Nature **410**(6824): 120-4.

Bassing, C. H., W. Swat, et al. (2002). "The mechanism and regulation of chromosomal V(D)J recombination." Cell **109 Suppl**: S45-55.

Bickmore, W. A. and P. Teague (2002). "Influences of chromosome size, gene density and nuclear position on the frequency of constitutional translocations in the human population." Chromosome Res **10**(8): 707-15.

Blackwood, E. M. and J. T. Kadonaga (1998). "Going the distance: a current view of enhancer action." Science **281**(5373): 61-3.

Boggs, B. A. and A. C. Chinault (1997). "Analysis of DNA replication by fluorescence in situ hybridization." Methods **13**(3): 259-70.

Buckley, P. G., K. K. Mantripragada, et al. (2002). "A full-coverage, high-resolution human chromosome 22 genomic microarray for clinical and research applications." Hum Mol Genet **11**(25): 3221-9.

Carlson, C., H. Sirotkin, et al. (1997). "Molecular definition of 22q11 deletions in 151 velo-cardio-facial syndrome patients." Am J Hum Genet **61**(3): 620-9.

Chess, A., I. Simon, et al. (1994). "Allelic inactivation regulates olfactory receptor gene expression." Cell **78**(5): 823-34.

Cheung, J., X. Estivill, et al. (2003). "Genome-wide detection of segmental duplications and potential assembly errors in the human genome sequence." Genome Biol **4**(4): R25.

Cimbora, D. M., D. Schubeler, et al. (2000). "Long-distance control of origin choice and replication timing in the human beta-globin locus are independent of the locus control region." Mol Cell Biol **20**(15): 5581-91.

Cohen, S. M., E. R. Cobb, et al. (1998). "Identification of chromosomal bands replicating early in the S phase of normal human fibroblasts." Exp Cell Res **245**(2): 321-9.

Cook, P. R. (1994). "RNA polymerase: structural determinant of the chromatin loop and the chromosome." Bioessays **16**(6): 425-30.

Cook, P. R. (1995). "A chromomeric model for nuclear and chromosome structure." J Cell Sci **108**(Pt 9): 2927-35.

Cook, P. R. (1999). "The organization of replication and transcription." Science **284**(5421): 1790-5.

Craig, J. M., S. Boyle, et al. (1997). "Scaffold attachments within the human genome." J Cell Sci **110**(Pt 21): 2673-82.

Cremer, M., J. von Hase, et al. (2001). "Non-random radial higher-order chromatin arrangements in nuclei of diploid human cells." Chromosome Res **9**(7): 541-67.

Cremer, T. and C. Cremer (2001). "Chromosome territories, nuclear architecture and gene regulation in mammalian cells." Nat Rev Genet **2**(4): 292-301.

Cross, S. H., V. H. Clark, et al. (2000). "CpG island libraries from human chromosomes 18 and 22: landmarks for novel genes." Mamm Genome **11**(5): 373-83.

Delgado, S., M. Gomez, et al. (1998). "Initiation of DNA replication at CpG islands in mammalian chromosomes." Embo J **17**(8): 2426-35.

DePamphilis, M. L. (2003). "The 'ORC cycle': a novel pathway for regulating eukaryotic DNA replication." Gene **310**: 1-15.

Devriendt, K., J. P. Fryns, et al. (1998). "The annual incidence of DiGeorge/velocardiofacial syndrome." J Med Genet **35**(9): 789-90.

Diffley, J. F. and K. Labib (2002). "The chromosome replication cycle." J Cell Sci **115**(Pt 5): 869-72.

Dijkwel, P. A. and J. L. Hamlin (1995). "The Chinese hamster dihydrofolate reductase origin consists of multiple potential nascent-strand start sites." Mol Cell Biol **15**(6): 3023-31.

Dijkwel, P. A., L. D. Mesner, et al. (2000). "Dispersive initiation of replication in the Chinese hamster rhodopsin locus." Exp Cell Res **256**(1): 150-7.

Dijkwel, P. A., S. Wang, et al. (2002). "Initiation sites are distributed at frequent intervals in the Chinese hamster dihydrofolate reductase origin of replication but are used with very different efficiencies." Mol Cell Biol **22**(9): 3053-65.

Dimitrova, D. S. and D. M. Gilbert (1999). "The spatial position and replication timing of chromosomal domains are both established in early G1 phase." Mol Cell **4**(6): 983-93.

Dimitrova, D. S. and D. M. Gilbert (2000). "Temporally coordinated assembly and disassembly of replication factories in the absence of DNA synthesis." Nat Cell Biol **2**(10): 686-94.

Djeliova, V., G. Russev, et al. (2001). "Dynamics of association of origins of DNA replication with the nuclear matrix during the cell cycle." Nucleic Acids Res **29**(15): 3181-7.

Drouin, R., N. Lemieux, et al. (1990). "Analysis of DNA replication during S-phase by means of dynamic chromosome banding at high resolution." Chromosoma **99**(4): 273-80.

Dunham, I., N. Shimizu, et al. (1999). "The DNA sequence of human chromosome 22." Nature **402**(6761): 489-95.

Dutrillaux, B., J. Couturier, et al. (1976). "Sequence of DNA replication in 277 R- and Q-bands of human chromosomes using a BrdU treatment." Chromosoma **58**(1): 51-61.

Eberharter, A. and P. B. Becker (2002). "Histone acetylation: a switch between repressive and permissive chromatin: Second in review series on chromatin dynamics." EMBO Rep **3**(3): 224-9.

Eichler, E. E. (2001). "Segmental duplications: what's missing, misassigned, and misassembled-- and should we care?" Genome Res **11**(5): 653-6.

Falck, J., J. H. Petrini, et al. (2002). "The DNA damage-dependent intra-S phase checkpoint is regulated by parallel pathways." Nat Genet **30**(3): 290-4.

Ferreira, J., G. Paolella, et al. (1997). "Spatial organization of large-scale chromatin domains in the nucleus: a magnified view of single chromosome territories." J Cell Biol **139**(7): 1597-610.

Fiegler, H., P. Carr, et al. (2003). "DNA microarrays for comparative genomic hybridization based on DOP-PCR amplification of BAC and PAC clones." Genes Chromosomes Cancer **36**(4): 361-74.

Fiegler, H., S. M. Gribble, et al. (2003). "Array painting: a method for the rapid analysis of aberrant chromosomes using DNA microarrays." J Med Genet **40**(9): 664-70.

Ford, H. L. and A. B. Pardee (1998). "The S phase: beginning, middle, and end: a perspective." J Cell Biochem Suppl **31**: 1-7.

Fowler, J. C., L and Jarvis, P. (1998). Practical Statistics for Field Biology, John Wiley and Sons.

Francke, U. (1994). "Digitized and differentially shaded human chromosome ideograms for genomic applications." Cytogenet Cell Genet **65**(3): 206-18.

Furstenthal, L., C. Swanson, et al. (2001). "Triggering ubiquitination of a CDK inhibitor at origins of DNA replication." Nat Cell Biol **3**(8): 715-22.

Ganner, E. and H. J. Evans (1971). "The relationship between patterns of DNA replication and of quinacrine fluorescence in the human chromosome complement." Chromosoma **35**(3): 326-41.

Gatti, M. and B. S. Baker (1989). "Genes controlling essential cell-cycle functions in Drosophila melanogaster." Genes Dev **3**(4): 438-53.

Giacca, M., L. Zentilin, et al. (1994). "Fine mapping of a replication origin of human DNA." Proc Natl Acad Sci U S A **91**(15): 7119-23.

Gilbert, D. M. (1986). "Temporal order of replication of Xenopus laevis 5S ribosomal RNA genes in somatic cells." Proc Natl Acad Sci U S A **83**(9): 2924-8.

Gilbert, D. M. (2001). "Making sense of eukaryotic DNA replication origins." Science **294**(5540): 96-100.

Gilbert, D. M. (2001). "Nuclear position leaves its mark on replication timing." J Cell Biol **152**(2): F11-5.

Gilbert, D. M. (2002). "Replication timing and metazoan evolution." Nat Genet **32**(3): 336-7.

Gilbert, D. M. (2002). "Replication timing and transcriptional control: beyond cause and effect." Curr Opin Cell Biol **14**(3): 377-83.

Goren, A. and H. Cedar (2003). "Replicating by the clock." Nat Rev Mol Cell Biol **4**(1): 25-32.

Gorman, J. R. and F. W. Alt (1998). "Regulation of immunoglobulin light chain isotype expression." Adv Immunol **69**: 113-81.

Greenberg, F., F. F. Elder, et al. (1988). "Cytogenetic findings in a prospective series of patients with DiGeorge anomaly." Am J Hum Genet **43**(5): 605-11.

Grewal, S. I. and D. Moazed (2003). "Heterochromatin and epigenetic control of gene expression." Science **301**(5634): 798-802.

Gribble, S. (2003). Fluorescence In Situ Hybridisation, Flow sorting and related technologies. Genome Mapping and Sequencing. I. Dunham. Wymondham, UK, Horizon Scientific Press**:** 129-166.

Grunstein, M. (1997). "Histone acetylation in chromatin structure and transcription." Nature **389**(6649): 349-52.

Hassan, A. B. and P. R. Cook (1993). "Visualization of replication sites in unfixed human cells." J Cell Sci **105 (Pt 2)**: 541-50.

Hassan, A. B. and P. R. Cook (1994). "Does transcription by RNA polymerase play a direct role in the initiation of replication?" J Cell Sci **107 (Pt 6)**: 1381-7.

Hassan, A. B., R. J. Errington, et al. (1994). "Replication and transcription sites are colocalized in human cells." J Cell Sci **107 (Pt 2)**: 425-34.

Hatton, K. S., V. Dhar, et al. (1988). "Replication program of active and inactive multigene families in mammalian cells." Mol Cell Biol **8**(5): 2149-58.

Heitz, E. (1928). "Das Heterochromatin der Moose." 1 Jahrbuecher Wiss Bot. **69**: 762-818.

Heitz, E. (1930). "Der Bau der samatischen Kerne von *Drosophila melanogaster*." Z Indukt Abstammungs-Vererbungslehre **54**: 248-249.

Holmquist, G., M. Gray, et al. (1982). "Characterization of Giemsa dark- and light-band DNA." Cell **31**(1): 121-9.

Holmquist, G. P. (1987). "Role of replication time in the control of tissue-specific gene expression." Am J Hum Genet **40**(2): 151-73.

Hozak, P., A. B. Hassan, et al. (1993). "Visualization of replication factories attached to nucleoskeleton." Cell **73**(2): 361-73.

Hozak, P., D. A. Jackson, et al. (1994). "Replication factories and nuclear bodies: the ultrastructural characterization of replication sites during the cell cycle." J Cell Sci **107 (Pt 8)**: 2191-202.

Hultdin, M., E. Gronlund, et al. (2001). "Replication timing of human telomeric DNA and other repetitive sequences analyzed by fluorescence in situ hybridization and flow cytometry." Exp Cell Res **271**(2): 223-9.

International Human Genome Sequencing Consortium (2001). "Initial sequencing and analysis of the human genome." Nature **409**(6822): 860-921.

Ina, S., T. Sasaki, et al. (2001). "A broad replication origin of Drosophila melanogaster, oriDalpha, consists of AT-rich multiple discrete initiation sites." Chromosoma **109**(8): 551-64.

Izumikawa, Y., K. Naritomi, et al. (1991). "Replication asynchrony between homologs 15q11.2: cytogenetic evidence for genomic imprinting." Hum Genet **87**(1): 1-5.

Jackson, J. P., A. M. Lindroth, et al. (2002). "Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase." Nature **416**(6880): 556-60.

Ji, Y., E. E. Eichler, et al. (2000). "Structure of chromosomal duplicons and their role in mediating human genomic disorders." Genome Res **10**(5): 597-610.

Kalejta, R. F., X. Li, et al. (1998). "Distal sequences, but not ori-beta/OBR-1, are essential for initiation of DNA replication in the Chinese hamster DHFR origin." Mol Cell **2**(6): 797-806.

Kallioniemi, A., O. P. Kallioniemi, et al. (1992). "Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors." Science **258**(5083): 818-21.

Kawame, H., S. M. Gartler, et al. (1995). "Allele-specific replication timing in imprinted domains: absence of asynchrony at several loci." Hum Mol Genet **4**(12): 2287-93.

Kawasaki, K., S. Minoshima, et al. (1997). "One-megabase sequence analysis of the human immunoglobulin lambda gene locus." Genome Res **7**(3): 250-61.

Kelly, T. J. and G. W. Brown (2000). "Regulation of chromosome replication." Annu Rev Biochem **69**: 829-80.

Kerem, B. S., R. Goitein, et al. (1984). "Mapping of DNAase I sensitive regions on mitotic chromosomes." Cell **38**(2): 493-9.

Kimura, K. and T. Hirano (1997). "ATP-dependent positive supercoiling of DNA by 13S condensin: a biochemical implication for chromosome condensation." Cell **90**(4): 625-34.

Kimura, K., V. V. Rybenkov, et al. (1999). "13S condensin actively reconFigures DNA by introducing global positive writhe: implications for chromosome condensation." Cell **98**(2): 239-48.

Kitsberg, D., S. Selig, et al. (1993). "Replication structure of the human beta-globin gene domain." Nature **366**(6455): 588-90.

Klein, C. A., O. Schmidt-Kittler, et al. (1999). "Comparative genomic hybridization, loss of heterozygosity, and DNA sequence analysis of single cells." Proc Natl Acad Sci U S A **96**(8): 4494-9.

Lachner, M., D. O'Carroll, et al. (2001). "Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins." Nature **410**(6824): 116-20.

Latt, S. A. (1973). "Microfluorometric detection of deoxyribonucleic acid replication in human metaphase chromosomes." Proc Natl Acad Sci U S A **70**(12): 3395-9.

Li, E. (2002). "Chromatin modification and epigenetic reprogramming in mammalian development." Nat Rev Genet **3**(9): 662-73.

Li, F., J. Chen, et al. (2001). "The replication timing program of the Chinese hamster beta-globin locus is established coincident with its repositioning near peripheral heterochromatin in early G1 phase." J Cell Biol **154**(2): 283-92.

Lindsay, E. A. (2001). "Chromosomal microdeletions: dissecting del22q11 syndrome." Nat Rev Genet **2**(11): 858-68.

Little, R. D., T. H. Platt, et al. (1993). "Initiation and termination of DNA replication in human rRNA genes." Mol Cell Biol **13**(10): 6600-13.

Maynard, T. M., G. T. Haskell, et al. (2002). "22q11 DS: genomic mechanisms and gene function in DiGeorge/velocardiofacial syndrome." Int J Dev Neurosci **20**(3-5): 407-19.

McCune, H. J. and A. D. Donaldson (2003). "DNA replication: telling time with microarrays." Genome Biol **4**(2): 204.

McDermid, H. E. and B. E. Morrow (2002). "Genomic disorders on 22q11." Am J Hum Genet **70**(5): 1077-88.

Miller, O. L., Jr. (1965). "Fine structure of lampbrush chromosomes." Natl Cancer Inst Monogr **18**: 79-99.

Mostoslavsky, R., N. Singh, et al. (2001). "Asynchronous replication and allelic exclusion in the immune system." Nature **414**(6860): 221-5.

Munkel, C., R. Eils, et al. (1999). "Compartmentalization of interphase chromosomes observed in simulation and experiment." J Mol Biol **285**(3): 1053-65.

Nakamura, H. M., T. and Sato, C (1986). "Structural Organisation of replicon domains during DNA synthetic phase in the mammalian nucleus." Exp Cell Res **165**: 291-297.

Natale, D. A., C. J. Li, et al. (2000). "Selective instability of Orc1 protein accounts for the absence of functional origin recognition complexes during the M-G(1) transition in mammals." Embo J **19**(11): 2728-38.

Nemazee, D. and M. Weigert (2000). "Revising B cell receptors." J Exp Med **191**(11): 1813-7.

Nishitani, H. and Z. Lygerou (2002). "Control of DNA replication licensing in a cell cycle." Genes Cells **7**(6): 523-34.

Nothias, J. Y., M. Miranda, et al. (1996). "Uncoupling of transcription and translation during zygotic gene activation in the mouse." Embo J **15**(20): 5715-25.

Ofir, R., A. C. Wong, et al. (1999). "Position effect of human telomeric repeats on replication timing." Proc Natl Acad Sci U S A **96**(20): 11434-9.

Ohno, S. (1985). "Dispensable Genes." Trends in Genetics **1**: 160-164.

Ormerod, M. G. (2000). Flow Cytometry, Oxford University Press.

Ormerod, M. G., A. W. Payne, et al. (1987). "Improved program for the analysis of DNA histograms." Cytometry **8**(6): 637-41.

Paulson, J. R. and U. K. Laemmli (1977). "The structure of histone-depleted metaphase chromosomes." Cell **12**(3): 817-28.

Pflumm, M. F. (2002). "The role of DNA replication in chromosome condensation." Bioessays **24**(5): 411-8.

Pflumm, M. F. and M. R. Botchan (2001). "Orc mutants arrest in metaphase with abnormally condensed chromosomes." Development **128**(9): 1697-707.

Pinkel, D., R. Segraves, et al. (1998). "High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays." Nat Genet **20**(2): 207-11.

Pollard, T. a. E., W C (2002). Cell Biology, Elsevier Science.

Raghuraman, M. K., E. A. Winzeler, et al. (2001). "Replication dynamics of the yeast genome." Science **294**(5540): 115-21.

Redi, C. A., S. Garagna, et al. (2001). "The other chromatin." Chromosoma **110**(3): 136-47.

Rountree, M. R., K. E. Bachman, et al. (2000). "DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci." Nat Genet **25**(3): 269-77.

Savage, J. R. (1996). "Insight into sites." Mutat Res **366**(2): 81-95.

Schermelleh, L., I. Solovei, et al. (2001). "Two-color fluorescence labeling of early and mid-to-late replicating chromatin in living cells." Chromosome Res **9**(1): 77-80.

Schleiermacher, G., I. Janoueix-Lerosey, et al. (2003). "Combined 24-color karyotyping and comparative genomic hybridization analysis indicates predominant rearrangements of early replicating chromosome regions in neuroblastoma." Cancer Genet Cytogenet **141**(1): 32-42.

Schubeler, D., D. Scalzo, et al. (2002). "Genome-wide DNA replication profile for Drosophila melanogaster: a link between transcription and replication timing." Nat Genet **32**(3): 438-42.

Selig, S., K. Okumura, et al. (1992). "Delineation of DNA replication time zones by fluorescence in situ hybridization." Embo J **11**(3): 1217-25.

Simon, I., T. Tenzen, et al. (1999). "Asynchronous replication of imprinted genes is established in the gametes and maintained during development." Nature **401**(6756): 929-32.

Singh, N., F. A. Ebrahimi, et al. (2003). "Coordination of the random asynchronous replication of autosomal loci." Nat Genet **33**(3): 339-41.

Sinnett, D., A. Flint, et al. (1993). "Determination of DNA replication kinetics in synchronized human cells using a PCR-based assay." Nucleic Acids Res **21**(14): 3227-32.

Smith, Z. E. and D. R. Higgs (1999). "The pattern of replication at a human telomeric region (16p13.3): its relationship to chromosome structure and gene expression." Hum Mol Genet **8**(8): 1373-86.

Snijders, A. M., N. Nowak, et al. (2001). "Assembly of microarrays for genome-wide measurement of DNA copy number." Nat Genet **29**(3): 263-4.

Solinas-Toldo, S., S. Lampel, et al. (1997). "Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances." Genes Chromosomes Cancer **20**(4): 399-407.

Spiteri, E., M. Babcock, et al. (2003). "Frequent translocations occur between low copy repeats on chromosome 22q11.2 (LCR22s) and telomeric bands of partner chromosomes." Hum Mol Genet **12**(15): 1823-37.

Stevenson, J. B. and D. E. Gottschling (1999). "Telomeric chromatin modulates replication timing near chromosome ends." Genes Dev **13**(2): 146-51.

Strachan, T. a. R. A. P. (2001). Human Molecular Genetics 2, BIOS Scientific publishers.

Strehl, S., J. M. LaSalle, et al. (1997). "High-resolution analysis of DNA replication domain organization across an R/G-band boundary." Mol Cell Biol **17**(10): 6157-66.

Ten Hagen, K. G., D. M. Gilbert, et al. (1990). "Replication timing of DNA sequences associated with human centromeres and telomeres." Mol Cell Biol **10**(12): 6348-55.

Tenzen, T., T. Yamagata, et al. (1997). "Precise switching of DNA replication timing in the GC content transition area in the human major histocompatibility complex." Mol Cell Biol **17**(7): 4043-50.

Turner, M. (2001). Antibodies. Immunology. I. B. Roitt, J and Male, D. London, Harcourt Publishers Ltd.

van Steensel, B. and S. Henikoff (2003). "Epigenomic profiling using microarrays." Biotechniques **35**(2): 346-50, 352-4, 356-7.

Vashee, S., C. Cvetic, et al. (2003). "Sequence-independent DNA binding and replication initiation by the human origin recognition complex." Genes Dev **17**(15): 1894-908.

Vogelauer, M., L. Rubbi, et al. (2002). "Histone acetylation regulates the time of replication origin firing." Mol Cell **10**(5): 1223-33.

Wang, Z., I. B. Castano, et al. (2000). "Pol kappa: A DNA polymerase required for sister chromatid cohesion." Science **289**(5480): 774-9.

Watanabe, Y., A. Fujiyama, et al. (2002). "Chromosome-wide assessment of replication timing for human chromosomes 11q and 21q: disease-related genes in timing-switch regions." Hum Mol Genet **11**(1): 13-21.

Watson, J. V., S. H. Chambers, et al. (1987). "A pragmatic approach to the analysis of DNA histograms with a definable G1 peak." Cytometry **8**(1): 1-8.

Yan, P. S., C. M. Chen, et al. (2001). "Dissecting complex epigenetic alterations in breast cancer using CpG island microarrays." Cancer Res **61**(23): 8375-80.

Yunis, J. J. (1981). "Mid-prophase human chromosomes. The attainment of 2000 bands." Hum Genet **56**(3): 293-8.

Zhang, J., F. Xu, et al. (2002). "Establishment of transcriptional competence in early and late S phase." Nature **420**(6912): 198-202.

Zhou, J., O. V. Ermakova, et al. (2002). "Replication and subnuclear location dynamics of the immunoglobulin heavy-chain locus in B-lineage cells." Mol Cell Biol **22**(13): 4876-89.

Zink, D., H. Bornfleth, et al. (1999). "Organization of early and late replicating DNA in human chromosome territories." Exp Cell Res **247**(1): 176-88.

Appendices

Appendix 1: Reagents and buffers used.

Appendix 2: PCR primers for the High Resolution Array

Appendix 3: Primers for quantitative PCR

Appendix 4: Male:male hybridisation on 1Mb array

Appendix 5: Male:female hybridisation on 1Mb array

Appendix 6: Replication timing profiles for all 24 chromosomes.

Appendix 7: Perl program to identify regions of co-ordinated replication

Appendix 8: Replication timing and Expression level profiles for all 24 chromosomes.

Appendix 9: Chromosome 22 sequencing-clone information

Appendix 10: 1Mb profiles of patients with DiGeorge phenotype and no 22q11 deletion.

Appendix 11: Clones known to report an incorrect copy number change on the 1Mb resolution array

Appendix 12: Position of Chromosomal Breakpoints on the Replication Timing Profiles

Appendix 13: The significance of a correlation co-efficient.

Appendix 14: Publications arising from this work.

‘The Replication Timing of the Human Genome’ Woodfine *et al*.

Appendix 1: Reagents and buffers used.

Amino linking Buffer (10x)

500mM KCl,
25mM MgCl2,
50mM Tris/HCl pH 8.5
Made with autoclaved distilled water.

HindIII Digestion mix (for a 96 well plate)

Hind III (Boehringer 40U/ml),	55µl
Buffer B (Boehringer)	99µl
Sterilised water	286µl

Hybridisation Buffer

50% deionised formamide
2xSSC
10% dextran sulphate
0.1% SDS
10mM Tris pH 7.4
0.1% Tween 20

LB Agar

Tryptone	10g
Yeast Extract	5g
NaCl	10g
Agar	15g

Make up to 1 litre with autoclaved distilled water.

LB Broth

Tryptone	10g
Yeast Extract	5g
NaCl	10g

pH to 7.5 (using 1M NaOH)
Make up to 1 litre with autoclaved distilled water.
Autoclave at 121^{o}C for 15 minutes.

Orange G (10mls)

Orange G	0.1g
Ficoll	1.2g

Make up to 10ml with sterilised distilled water

Polyamine isolation buffer (PAB)

80mM KCl
20mM NaCl
2mM EDTA
0.5mM EGTA
15mM Tris
3mM dithiothreitol

0.25% (vol:vol) Triton X-100
pH adjusted to 7.2

Sheath Buffer

10mM Tris-HCl pH 8.0
1mM EDTA
100mM NaCl
0.5mM Sodium Azide

SSC (1x)

0.15M NaCl
0.015M Sodium Ctrate
pH 7.0

TAPS 2 Buffer (10x)

250mM TAPS pH 9.3,
166mM $(NH_4)_2SO_4$,
25m $MgCl_2$,
0.165% w/v Bovine serum albumin (Sigma),
0.7% v/v 2-mercaptoethanol
Made with autoclaved distilled water.

TBE Buffer (10x)

Tris Base	121g
Boric Acid	61.83g
EDTA	18.612g

pH 8.0
Make up to 1 litre with autoclaved distilled water.

TY Media (2x)

Bacto-tryptone	16g
Bacto-yeast Extract	10g
NaCl	5g

Make up to 1 litre with autoclaved distilled water.
Autoclave at 121^{o}C for 15 minutes.

Vista Green Stain (for 500ml – 1 gel)

1Ml Tris HCL	5ml
0.5M EDTA pH 7.4	0.5ml
Vistra Green	0.05ml

Make up to 500ml with sterilised distilled water.

Appendix 2: PCR primers for the High Resolution Array

2a: Primer sequence for PCR products in the high resolution array

STS Primer	**Forward primer**	**Reverse primer**
500bp overlapping PCR product array		
stSG494879	TGACCATGGACGGGAGAGAAAACATCCA	GAAAATGTGTGGCAGGTTCA
stSG494880	TGACCATGTGTCTCCCTTGGTGACATGA	CTCCCCACATGAGACCAGAT
stSG494881	TGACCATGAAGGCTAATGGGAAAGAGGC	TTCTGTCCCCTTTTGATTGC
stSG494882	TGACCATGCTAGGAAGAGGTTCCAGGGG	CTGAGCCTTCCTGTGTGGAT
stSG494883	TGACCATGGGAAACCATGCACCTCAGTT	GACCAGAAGGAAATGTTGGC
stSG494884	TGACCATGAAGACGGCTCTCAACCTTCA	GAAGACTCCAGCTGTGTCCC
stSG494885	TGACCATGCTCTTTGCTCGCAGTCATCA	CACAAGAGAAACACAGGCTCTC
stSG494886	No Unique Sequence	No Unique Sequence
stSG494887	TGACCATGGTCCCAACACCTCCATTTTG	CTGAACTTGGCCCATAAAACT
stSG494888	TGACCATGCGGACTCAAAAGAACAAGGC	CCTCTGAAACCGGCAGAATA
stSG494889	TGACCATGATCATTGAAGGTGCCAAGGA	TGTGCTTCAGCAAAACATCC
stSG494890	TGACCATGTACTCTTCAGTGGCCCGAAC	TATTGGCGGCCATCTACTTT
stSG494891	TGACCATGGTGCTAATTTCCACCACAGTCA	TGAAGGAAATGGAAAAGGGA
stSG494892	TGACCATGCCACTGCCTGCCAGTTAGAT	GTGCCGATCGAGACTCTTCT
stSG494893	TGACCATGGGCAAAATTCAAATCCTCCA	CTGATCTGCCTCCATCCATT
stSG494894	TGACCATGCCAGGTCACTGCCCTAAAAA	CCCAGGTCAGTTGTTTGTGA
stSG494895	TGACCATGTGAGGACTCCTGGGTTCAAG	TTCCAAACAGAGGCCTTCAT
stSG494896	TGACCATGGGTTTTCTGGACAGTTGACACA	GGAAAATGGACAAGCAGTTGA
stSG494897	TGACCATGGTGTCTTGGAGACTCCCTGG	TCCATAATTTCCGGGTTTCTA
stSG494898	TGACCATGCCTGTGGAAATCCCTCATGT	AGGACACAGGTTTGCTTTCA
stSG494899	TGACCATGGTGGCCTCTAACTCTGGCAT	CCCATACCTTTCTGAATCTGC
stSG494900	TGACCATGAATGACACCATCACCAGCAA	AGTTTCAATCACCGTGCCAT
stSG494901	TGACCATGCCCATCCTATGCCCTGTATG	GCAGCTGCAGTCAACTAACAGA
stSG494902	TGACCATGCATCTCCCAAGCTTTGCCTA	TGCACATGGTGAAATGAACA
stSG494903	TGACCATGTCCTCATGCCTCATGTCATC	TTTGGGAATACAGACAGGGG
stSG494904	TGACCATGTGGGGACAGAGGTAATCTGG	TTGCATGTGATCTGCACGTA
stSG494905	TGACCATGAAAGGTCACCCATTGCTTTT	TGGGATAAGTGAGGGTCTGC
stSG494906	TGACCATGGGAGGCTTTGGTTGTGTTTC	GTTGTTGGGGGAAGGAAAGT
stSG494907	TGACCATGAGGGTGTGACCCTGAGAGG	GCCACTGGCTGTTCAGATTA
stSG494908	TGACCATGGTGAAGGCTTGGCTGATACC	TGAAACATCTTCTGCCTCCA
stSG494909	TGACCATGGCAACTCTCCAAGTTCTGCC	GGATGGAGAAGGAAGTGCAG
stSG494910	TGACCATGTAATCTGGAAGGGCAGGAGA	CTCCCCTGAAGTGAGAGCTG
stSG494911	TGACCATGATGCCCTGACTCCAAAACTG	CCGCTGGAATTGTATCCTGT
stSG494912	TGACCATGACTCTGGAAGCCAAAAAGCA	CCAAACCGAAACAAAAAGGA
stSG494913	TGACCATGTTTTCCTTGGAACCCTTTATGA	GGTGTTTGTAAGGCAAGGAAA
stSG494914	TGACCATGCAAGTATGGCGCATCTCTCA	GGAAGTTCACGAGGGACAAA
stSG494915	TGACCATGACCCCATTCAGCTCACAAAA	ATCTGGCAGGATTTCTTGGA
stSG494916	TGACCATGAGGGGCTTGTGAAGACACAC	GGCTGGAATTCCGTCTCATA
stSG494917	No Unique Sequence	No Unique Sequence
stSG494918	TGACCATGGGAGCTCACCTTTTGGGTC	GCAGGAATAGAAGTGGGAGC
stSG494919	TGACCATGGGCCCTCCTAAGCTATTTGG	TGGGGTGTGATCACTGAGAA
stSG494920	TGACCATGGGTTCAATCTGTTGCCGTTT	GTGTTTGCATGGTTGAGCAC
stSG494921	No Unique Sequence	No Unique Sequence
stSG494922	TGACCATGGTGTACAGGGGAAGAGCGAG	GGGAAAGGAAAACTGAACCA

stSG494923	TGACCATGCTTCGTCTCTATGGTCCCCC	TAACCAACTGGAGGCAGAGG
stSG494924	TGACCATGTGTCCATTTCCTTTAGTGCG	CCGTGAACAGTAACTCCCTAGC
stSG494925	TGACCATGACAGGGTGCAGTGTAGTCCC	GGCTCCCCACAACAAGTTT
stSG494926	TGACCATGACTTCTCCCATGTGTTGTTCC	AGGCAGGGGAGCCTATCTAA
stSG494927	TGACCATGATGGGTGCTGTTCTTGTTCC	TTGGAAAACTGCAAATCAGC
stSG494928	TGACCATGTGATACCCTTCTCCTGCTCC	TGAGCACCTGGGTACAGACA
stSG494929	TGACCATGCTCACTGGGCTGGCTCTATC	TGCTTTCTTACACAAGACCCA
stSG494930	TGACCATGATCAGGTGGGAATGATGCTC	AGAGGTTGCCCAAAACACAC
stSG494931	TGACCATGAAATGAGCAAACTTGGCAGC	TCACCTGGCCAAAACAATTT
stSG494932	TGACCATGTGACTGTCTCAGAGCTGAATGA	CCAAGCCAAGATTCCTTTGA
stSG494933	TGACCATGGGTGGGTGAGACTTGAGGAA	AATTCCATGTCCCCACCATA
stSG494934	TGACCATGTGCTTCCTCCTCCTGTGACT	GTGAGCTACACCTTTGGCCT
stSG494935	TGACCATGTACCCATCAAGCCTACCTGG	TTCTCCCTTTCCTCAGTCCC
stSG494936	TGACCATGAGGCCTTTGAATAGCAAGCA	GCTGGACTATTGGCTTCTGC
stSG494937	TGACCATGTGAGAAAAACCCACTCAGGG	GGTTTCAACCCAGGAAGACA
stSG494938	TGACCATGGAGGTTAGGCTCAAGGGGAC	ACCTGTGTTGGGCTCTTGAC
stSG494939	TGACCATGTGAGTGCTTCCTGTGTCCTG	GTTTGTTAGCGTATGGGCGT
stSG494940	TGACCATGGCTTTGCTTGCTACTTGGCT	CCATCTCGTTTCCAGGACTC
stSG494941	TGACCATGTTTGTTGAGCACTGTCTGGC	TCTCTTCCACATGGACCCTC
stSG494942	TGACCATGCTCAAATCACACCACACACG	TGGTGTCAGCTGAGAAGAGC
stSG494943	TGACCATGTGGCAGCATATTCGAGTGAG	GCTCACAGCCTCTCTGCTTT
stSG494944	TGACCATGTGATCCCCAACTAGAGAAAAGG	AGCAAATGTTATTTCCCCTCC
stSG494945	TGACCATGTGTCAGCCGATCAGTCAGTC	TGGGCTCCACATATTTCCTC
stSG494946	TGACCATGGGTGAATTCTCCACCAGTCC	CTCCCTAGCTGTGCCAGAAC
stSG494947	TGACCATGTTCCTGCCTGGCTAACTGAT	GCATAGAGAAGGGACTAGAGGG
stSG494948	TGACCATGACCCGTCAAATCCTCAGATG	CGGGTACTGGGACTTTACCA
stSG494949	TGACCATGTGAGCAACGGCATAGAGATG	CCGGCCCACAATTTTAATAGA
stSG494950	TGACCATGGACACACCAGGCATCAGAGA	TGCCATGGATGGTGAGACTA
stSG494951	TGACCATGTCTGGCTTCCAGTCCTTGTT	GAGGCAAGCAGATTTTGGAG
stSG494952	TGACCATGACTTTTGGAACTTGGCATGG	CCTTTGCACTCAATGCTTCA
stSG494953	TGACCATGTACATAGGGATCTGGGCTGG	AAATCCTGTGGCTCCTTGTG
stSG494954	TGACCATGCCTGCCAGCTTCTGACTTCT	AACAGATTTCCTCCCATTGC
stSG494955	TGACCATGGGCTGACCTACTGGAGCAAA	TCAAGAGGAATTGACCTGAACA
stSG494956	TGACCATGCTAAGTTTCTCCCCGCTCCT	GCCTAAGGCCAGATTGATGA
stSG494957	TGACCATGGTCTCTGGCTCTTTGTGGCT	CCATTCTACCCAGGCATCTG
stSG494958	TGACCATGTTGACAGTAGCTGCAGGTGG	TTGGTGAGGAGGGAGATGAC
stSG494959	TGACCATGTTGGGTAGGCTGATCAGAGG	TTCTGAAGACCCTGGAATGG
stSG494960	TGACCATGAAACCCACCTTCCAAAGTCC	GTCTCCAAGAGAGGAGCGGT
stSG494961	TGACCATGGAGAGGCTAACGGACATGCT	GGCCACAGTCCTGTTCATTT
stSG494962	TGACCATGGAAACTGAGGTGTTGCGGAT	AGGGGCATCAGTTCAACATC
stSG494963	TGACCATGGCACATCTTCAGTGGGACCT	CAGGAAATACCTGAGGCCAA
stSG494964	TGACCATGGTGATTGGGGATGTGTGTGA	AGGAAAGCCATTATTTGGGG
stSG494965	TGACCATGTAGGACATGGAAGACCGGAG	GACAAAGCGGATGAAACCAT
stSG494966	TGACCATGGACGTCATCACGAAGATCTGA	CTTTCAGCATGAACCAAGCC
stSG494967	TGACCATGGTGTTTGCTTCATTGGCCTT	TTGGAACCCTCTTCCTCCTT
stSG494968	TGACCATGCAGGTCCCACATCAGGACTT	TCCAGGGAGAGGAAGACAGA
stSG494969	TGACCATGTTAAGGACCACACCCTGGAG	AGGGGACAAGTGACATCCTG
stSG494970	TGACCATGGGCCTCCTACCACACTCACT	GCAATTTCCTTAGAATGACCCA
stSG494971	TGACCATGGAGCTCCGGAGACTGACAAC	TGTGCACCTCCTTTATGGAA
stSG494972	TGACCATGCAACCTGCCACAAGACCTG	GAATTGCCTCGCCCATACT
stSG494973	TGACCATGCCCAGAAAAACCTGGGATATG	GAGTCGCCACCGTAACATTT
stSG494974	TGACCATGAATATGCACAGGGGAGAACG	AAATTGGACTAGTGGCCCAG

stSG494975	TGACCATGCAACATCAGCTTCCGTGAGA	CCCAGCAGACTAGGGAGATG
stSG494976	TGACCATGGCTGTGGAAGGAAAGACCAA	GATGGAGAAACAATGGGTGC
stSG494977	TGACCATGTCTGATGTCAAAGCAACCTGA	CACTCATCAGCCTAGATGCAA
stSG494978	No Unique Sequence	No Unique Sequence
stSG494979	TGACCATGGGCAGGAGTGGAGGTGATTA	CACAGGGCAGGTACCAAGTT
stSG494980	TGACCATGCATGCTCTGCTTTCCTTTCC	AGCAGCTCATGCTAATGCAG
stSG494981	TGACCATGGGCGCCGCATAATCTAAATA	TGGGAGATTTTCCAAGATGG
stSG494982	TGACCATGTTCAACAGAGCCGTGAACAG	GCCATTTGTGTAGCATTAGCC
stSG494983	TGACCATGAAAACAATAACGGACCGATCA	CACAAGCAATGGCCTTAACA
stSG494984	TGACCATGCAGGCACTACTGATGCTCCA	AATGCAGGAACACACATCCA
stSG494985	TGACCATGAATTTCAAACTGAGCAGGGG	ATGGCCAGCCGTTTACATAC
stSG494986	TGACCATGTTCTTAGGAAGGTTTTGAGCCT	ACAATAACCCCTGCAGTCCA
stSG494987	TGACCATGGGGCATTGAGTTTTTGGATG	GCCTCACCCAAACTGGTTAT
stSG494988	TGACCATGAATGGGCTGTACCTCATGCT	CTGCCTCCCTTGTCCATAAA
stSG494989	TGACCATGCGAGAGATACAGAGCCCAGG	TACGAAATTGGGGTTTCCAA
stSG494990	TGACCATGAGAGCACTTGCTGTAGGTCCA	GTAGGGCTCTAGACCTGGGC
stSG494991	TGACCATGACCAGGCCAACACTGGTACT	GGATGGGAGGTAAGCACTCA
stSG494992	TGACCATGGCAAAGTGAGAGAGAGATGTCC	CCATCCTTTCATTCCTTTAACC
stSG494993	TGACCATGCCCCATCTCCACCTACACAT	GGGAAAGTTTCTGGCTAATGC
stSG494994	TGACCATGCACTAAGGGAAGCACAGGGA	CTGCTTTTCAGTTTGGCCTC
stSG494995	TGACCATGTGTCCCTATCCCTCCCTCTT	GGCAGGCTCAGATCTGTAATG
stSG494996	TGACCATGGGATGACTTAGTAGGGGCCA	GGTGAGCACCACACCTCTCT
stSG494997	TGACCATGTCTTCCATGAGGGAATTTGG	CAAAATGGCATGGAGATACAAA
stSG494998	TGACCATGTGAGTGCCAAAGAATGGTGA	ATTGAGATGAATTGGCAGGC
stSG494999	TGACCATGAGAGTAAGGGTGGTGGGCTT	CCTTCAAGCTGGCTTTTGAC
stSG495000	TGACCATGAAGTGAGGAGTAGGGCTGGA	CGAATCAGGGGAAACTGAAG
stSG495001	TGACCATGGAATCCCCACGGTAGAGACA	TTAGCCATTCAGAGGGTTGG
stSG495002	TGACCATGCACCTTTCTTGCTCTGGAGG	AGAGCACTTGTTCCTGGCAT
stSG495003	TGACCATGGCAATGAAGGAATGAACCAAA	TGCCAATTACTGATCAGGCT
stSG495004	TGACCATGGCTTGCCTATGGGTGTGTCT	GATGTGGAGGAATGTGGCTT
stSG495005	TGACCATGCTCCCACAACCCTTGACCTA	AGCCTACCTTCCCCTTGAGA
stSG495006	TGACCATGGGCACTAAACTGGCTCCCTA	GCCATCCTGCAAGAGAAGTC
stSG495007	TGACCATGTTTTCCTCCCAACCACTTGT	TGCTGGCCTATCCCAAATTA
stSG495008	TGACCATGTGACTTGTGGGAAACAGCAA	TGTGGACCAATGCAAACACT
stSG495009	TGACCATGTTGGGAAGAAGGAGGGTTTT	AGAGATCTTGCTCACCCCAA
stSG495010	TGACCATGCACTGAAAGACTGGGGCTCT	ATCTGTCACCATCCTCAGCC
stSG495011	TGACCATGGGCTGAAGTCTGCAAATCCT	TATTTGTTCCCTGCCTTTGG
stSG495012	TGACCATGAATCCCTGGGAAGCTAAACG	ATGAGGTCCCCCAATTTCTC
stSG495013	TGACCATGATGCCAGCATTGATGTGTGT	TTGGTTGCAGCATCAGTAGC
stSG495014	TGACCATGTTTTCATCATTGGCTCCACA	CCGGGTAAAACAGACTCCCT
stSG495015	TGACCATGCACCTAAAAACACACCCTCCC	CCCTGGAAAGTTCCCAATTT
stSG495016	TGACCATGTTGCAGCTGCTGACTCAATC	TCTCCTCCCTCACTTCACCA
stSG495017	TGACCATGGATGAGGGTGAAGACTGGGA	TCACTTTTTCCTGCAAGGCT
stSG495018	TGACCATGTATGGCCAGTGCTTCTAGGC	GTGGGAGGGCAGTTTCTGTA
stSG495019	TGACCATGCACATGCTCCAGTGCTGAGT	AATCAGATTGGTTGGCAAGG
stSG495020	TGACCATGGAAAGGGGAGGAAACAGTCC	CTGGGGGTTTTATTGAAGACA
stSG495021	TGACCATGCAGTGATGTCAAGGCCAGTG	CCCAAACAGAGGTTCACCAT
stSG495022	TGACCATGTGGGAGATGCAGAGTTGACA	GAGGAAAGGCACAGATTGGT
stSG495023	TGACCATGGCAGTTTCTGGTGGTGACCT	TCAAGTTCAATGCCTAGGGG
stSG495024	TGACCATGGGTGTGAGATCCCAAAAGGA	CAATCTCCGGGTGCAGTTAT
stSG495025	TGACCATGGGCTGGTGGAAACAACACTT	ACAGCCTAGTGCAGCCTCAT
stSG495026	TGACCATGATCCTCCCTCTCACCCTCAT	TAAGGCAGTCCTGGAGGAGA

stSG495027	TGACCATGTTAATGGCTCCTCACCCTTG	AAGGGATGGAAGAAAGGAGG
stSG495028	TGACCATGGATGAGGTAATGCGGCTCTG	GCTTACCGATGTCGGAGTTG
stSG495029	TGACCATGGAGGTCGCAAAATGGGTAGA	ATCCTTGACAAGTGTTGGGC
stSG495030	TGACCATGGAGAGATGCCAGCAGTGACA	TGCAGGAAGTATCCCTCCAC
stSG495031	TGACCATGCCACTTGATATGTGGGGGTC	TCAGCTCCTTGCCTCAATTT
stSG495032	TGACCATGCACTTCCAGCTGCTCTCCTT	TGGGAAGCTACGTGTGATTTC
stSG495033	TGACCATGTGGTCAGCAGAGAGCTGAGA	GGATAGAAGGGCACTGACCA
stSG495034	TGACCATGTGAGTGTGAGGCACCTGAAG	AAATCATGGCTTCCCAACTG
stSG495035	TGACCATGGAGACATCCACTCCTCCCTG	TCACATGTGGGATCTTGAGG
stSG495036	TGACCATGGCTGGTTCCCTAGGTTCTCC	AAGACCATCAGGCGTTTCAC
stSG495037	TGACCATGGCCCTGATGGATTTTTCTGA	TAAGACAGGAAAAGGGGGCT
stSG495038	TGACCATGGCTCACAGTCATCCTGCTTG	AGAGTTGGGGGTCTTCTGGT
stSG495039	TGACCATGAATAAAAGATGGCTGCACGG	ATATTGACCTCCATCGTGCC
stSG495040	TGACCATGTGGTGGAGTGGACAAAGATTC	GGTGGGGAACAAGGAAAAGT
stSG495041	TGACCATGCAAAGGAGTTAGGTCGCAGG	TCTGGCATGTTCTGGTCTTG
stSG495042	TGACCATGCCTCAAAGGCTCTGTCCTGA	CTAGCAACAAATGCGCAAAA
stSG495043	TGACCATGTTCCTCAGGCCATTCAGAGT	CAACGGGAGTCACCTCAAAT
stSG495044	TGACCATGGCTCTAAGGAGCATGGTTGG	CTGTTACCTGGGGGACTTCA
stSG495045	TGACCATGCAGAACTCACGGGTCACGTA	GTTCCAAAAGCATTGCAGGT
stSG495046	TGACCATGGACACATCCTCAGCCATCCT	ACTTGAGCCTCCAATCTTATCC
stSG495047	TGACCATGTGCCTGTGCTTTTTCCTACC	CTTGGGCAAAGTCTGAGGAG
stSG495048	TGACCATGTCCTAATCCAGACTGCCCTG	TTAGTGGTTGATGTCTGCCG
stSG495049	TGACCATGCATTTTCCAGCCACTCTGTG	TGGGCAGAAGTTACCTGAGAA
stSG495050	TGACCATGTCCAGTGATTGAACTTCCTGTG	GCCGTGTTGTGTTTACATGG
stSG495051	TGACCATGTCCCTCTGGAAAGCAGAGAA	GACCTGAGAAGGGCATGG
stSG495052	TGACCATGGTACTCCCTCTCTCCCCCTG	CCCCCACACTTTTATTTCCA
stSG495053	TGACCATGTGAGGCACACATGCCTACAT	GCTCACCAGGAGCTACAAGG
stSG495054	TGACCATGAAGGCCTCAGTGCTGTCAGT	GCCACCTTTTGTCAGCTCTC
stSG495055	TGACCATGAGAGGAGCCACAGGCTATGA	TCCCAATTTCTGATCCTTGC
stSG495056	TGACCATGTGCATGTGAAGACGTAGGGA	TAACTGGCAGAATTCCCAGC
stSG495057	TGACCATGGGGGACACAGGATGTAACCA	TGGGATGTCTCTGATCTGGTC
stSG495058	TGACCATGTGCAAGCCTCCTTTTCTCAT	CATCCTTTGGGACATGCTTT
stSG495059	TGACCATGTGGAAAGCAGAAACCCACTC	CAGGCCTTCCACTGTCTGTT
stSG495060	No Unique Sequence	No Unique Sequence
stSG495061	TGACCATGAGGTGAGAAAGCAACCATCG	CACAGAATCACAGTGGCACA
stSG495062	TGACCATGCTGAGGTTGTTCCAAGCCAT	AAAGACCAGAAGGAGCAGCA
stSG495063	No Unique Sequence	No Unique Sequence
stSG495064	TGACCATGGCTGGCTTTCTATTTCCCGT	TCCAATGTCAGACAGAGAAAGG
stSG495065	TGACCATGAGTCCACAAAGAAGGGAGCC	AGGATTTCCCTGGTGTCTCA
stSG495066	TGACCATGGGCAGGTTCAAAGGGTTTTT	TCACTCAAGTGTGAAGGGGA
stSG495067	TGACCATGTCCTAGTCCGTGGTTTCACC	GTCACTGCACTTGCCTTTCA
stSG495068	TGACCATGGGTCATCATCTGGGAGAGGT	TCATGTCAAAGCAGACCTAAGC
stSG495069	TGACCATGAATGGCAAGAGAACGACACC	GGCAATGACTCACCCACATT
stSG495070	TGACCATGGGGACCACCTGCTGAGTAAA	GGCTGGTTCCTATTTGGTGA
stSG495071	TGACCATGGAAGATTTGGAGGGGACCAT	TTTGCAGGCTGAGAGAAACA
stSG495072	TGACCATGTCTTAGCAGGTGGGAACCCT	CCCTCAAGACCCTGTGAGAA
stSG495073	TGACCATGTGCATCAGCCAGTGACTTTC	GCACTTCTTTACGAGCCAGG
stSG495074	TGACCATGCCAGAAGTTGAGGAGGGTGA	AGAAATCCTGCCCGTCTTTT
stSG495075	TGACCATGTTTTGCAATTACCTCTGCCC	CAGCCACCTTGCTTTCACTT
stSG495076	TGACCATGAGGTGCTCAGCCATCAGACT	CCAAGACTCAAACTCCAGGC
stSG495077	TGACCATGTGGCCATTTAGAAGTTCCCTT	ATCGTGACCATTGTGGGACT
stSG495078	TGACCATGTGTCTGCTTATTTGGGGCTT	TGCAGAGTCACTTGAGGTGG

stSG495079	TGACCATGAGCCAGCAGAGAGGTGAGAA	GGCTCCCAGAATGATACCAG
stSG495080	TGACCATGAATGAAGGGCTTCCAGTTCA	TTCCTGGAGCCATTTACCTG
stSG495081	TGACCATGGGGGATACTGGGGATTGTTT	AAGAAAAGAAAGAGCTGCACAA
stSG495082	TGACCATGACTGAGGGAAGGAACTGGGT	AGGAGAAAGGAGGGCAGAAC
stSG495083	TGACCATGAACACAGGGATGGGTTCTTG	GCCACCATATCATGCCTTCT
stSG495084	TGACCATGGCACTTACTCCCTGCTCTGG	GAAAGGGAAAGCAGAGGGTC
stSG495085	TGACCATGTGCTGCATGTGATTTTCAGG	TAGCACGGGAAGTTTCTTGG
stSG495086	TGACCATGGCCCTCTGTGAGGAAGAATG	CGAGCAGTGCTACAGAGACG
stSG495087	TGACCATGGTAATGACCCATGTCCCCCT	CTTTTTCCTTCCCCTTCTGG
stSG495088	TGACCATGCATCTCCATGGTTCAAGCCT	CCTCTTCATGGAGGGTGAAG
stSG495089	TGACCATGGCCTCTTCCCATCACAAATG	TGCATGTGATTTGCTTGTTTG
stSG495090	TGACCATGTCCCAGGTTGACCAATAAGG	GATGCAAAGCTGTGCTGTGT
stSG495091	TGACCATGGAGGAAGAGGCTGCCCTAGT	CCACGTCCACTTGAGGTCTT
stSG495092	TGACCATGTGTCTGAGTGCAGGATGTCTG	AGCAGCAGCTGAGTTTGAGA
stSG495093	TGACCATGCCATCACACACACAAGAGCC	GCAGCTGAGCGTTCTTTTCT
stSG495094	TGACCATGGGAGCCTGCCTATCCCTACT	CTTCTGGGGTTGAATTCATTG
stSG495095	TGACCATGCCACCCAAAATAATGCCAAT	TTGGATGGTCTTCCCATTGT
stSG495096	TGACCATGGAAACACCACCATTCACGG	TTCCAAGACTCCTGCTTTTGA
stSG495097	TGACCATGAAAAGACAATGCTCGACGCT	AGCCATAAGGCCACATCAAG
stSG495098	TGACCATGAGATCGCCTCTGTGTTTGCT	AGCTAACGTCCATGTCACCC
stSG495099	TGACCATGAAGGGTGATATTTCCCTGGC	GGAATCAAAGGAGGAAAGGC
stSG495100	TGACCATGAGTCTGCTCTGCCTGACTCC	GCCAGTGGGACATCTCATTT
stSG495101	TGACCATGAATATGTTGCACCGATGCTG	TTTGGTCCTCTTCATCCCTG
stSG495102	TGACCATGGCAAGCCGACTTTTTGACTT	CGTACAGTAGGGGCTCAACC
stSG495103	TGACCATGTGGACATGAACCTGTGCAAT	CACAGCTATTGTGGATGCGT
stSG495104	TGACCATGCCCTGGCCAATAATGGTATG	GCCAGGTCATGGAATAGGAA
stSG495105	TGACCATGGAGCATCTATGAGGCGGTGT	AGGACTGGGGGACTGAAAGT
stSG495106	TGACCATGCTGGCTTGTTTTCCATGCTT	TGACTGTGAAGGTGATGGGA
stSG495107	TGACCATGTCCACTTCCCTTCTGCTTTC	AGCTTTGGCCACAGAAAAGA
stSG495108	TGACCATGACGCTAGGGTGTGATGTGGT	TCCTGTCCTCTAACCCGATG
stSG495109	TGACCATGTGGGTGGATCATACAACAACA	AGGTCTTGGTTGTCACCTGG
stSG495110	TGACCATGTTGTCAGCGTTGGCATTAGT	CTGACCACCTTGGACCAAAT
stSG495111	TGACCATGCACCATCACTGCAGGCTAAT	TTCAGGTGTAGACAGGAAGGC
stSG495112	TGACCATGAAGAGGGCAAAGGGACTGAT	GGTCCTGTTCAGAACCTCCA
stSG495113	TGACCATGGAGCTCTTCCTATGCAAACTCC	TGCCTCAGTTTTTATTGCAGG
stSG495114	TGACCATGTGCCAGACAAATCAGCAAAG	GAACACTCTCTGGACCAGGC
stSG495115	TGACCATGTCCCTAACACAACATTTGGCT	AAACCCAGGGGTGTACATGA
stSG495116	TGACCATGACATTTGCAGGGGATGATGT	GACCCTGAATGTGCCTCTGT
stSG495117	TGACCATGAGAATAAGCACGGCAGAGGA	GCCTTTTGGCACAAAACATT
stSG495118	TGACCATGTGCTGGAAACAGAACAGTGC	TTACTTCATTGTGCCCTCCC
stSG495119	TGACCATGTGGCCAAATGATATGAGCAG	TGCAGCACACGTTAGCTCTT
stSG495120	TGACCATGGGCAAAAATTCCTGTTTCCA	TCAGGATGCACAGTCCAGAG
stSG495121	TGACCATGTGGGTAATTTGAAGAGCGTG	CTTGTTGTTTCCGTAAGCCC
stSG495122	TGACCATGTTGCAAACATTTCTGGTGGA	TCAGCCAAGGAGCAGTTCTT
stSG495123	TGACCATGGAAACGATTGCTACAGTTTCCA	TGCAAATCTTGATGGTAGCAG
stSG495124	TGACCATGTGTTCCCTTTCTTCCCTCCT	GAAACACAGCACGTGGTTCA
stSG495125	TGACCATGGCCAGGAGAAACTGTTCCC	GGGCAGTTTCTTGGTGTGAT
stSG495126	TGACCATGTCACACTGACGTGTTCCAGA	CTCCTCCCCAAGCTCTCTTT
stSG495127	TGACCATGTTGCTGCCTAAAGGGAAAAG	GGCCATAGTGCGTTCTGTTT
stSG495128	TGACCATGGCAAGAGTGACTGAAAACGGA	AAAAACACACAGGGAGGTAGG
stSG495129	TGACCATGCGTGGCTGAAGAGAATTTCC	GTCAGCCCATTTCCTGTGTT
stSG495130	TGACCATGTAGCCCCAATCAATGACTCC	CTCCCAAGGGCACACATAGT

stSG495131	TGACCATGGCACTCTCAAGCCACTCACA	ATTATGGGAGCCCAGGAAAG
stSG495132	TGACCATGGTTTTTGAGGGAGCTTTCCA	AGGGAGACCCACACTCACAC
stSG495133	TGACCATGGGCCTTACACTTTTCCAGCA	GGCATAGTCGCTTGGTGAAT
stSG495134	TGACCATGCGGCTAGCTGTTCCTCACTC	ACCTTCCCTGCCCTTTTCTA
stSG495135	TGACCATGCTCCTTCCCCTGCACATAAG	GTGCTGTGTGTGGAGTGACC
stSG495136	TGACCATGGTGTGGGAAGGCTGGTCTAA	GAGGGCTTTGCAGTGTTAGC
stSG495137	TGACCATGAGACAGGTGCAGGAAGGAGA	TGGCTTTTGAGAAGGCATTT
stSG495138	TGACCATGTCTTCCATAAAGACAATCCCCT	TTCTGCCTGTGACAAACCTG
stSG495139	TGACCATGCATGCTGGACAACAACCATC	CCTTTCTTCAGGAGTGGTGC
stSG495140	TGACCATGGAAGTAGGAAGTTTCCCCGC	GCCTCCTCTGGGTCTTCTCT
stSG495141	TGACCATGTGGGATACTAGCCGCAGACT	GCGAACAACCTTCAGAAAGC
stSG495142	TGACCATGGAAGGGTTATGGACCTCAGTG	CTCACCAAGGTCCTTCCAAA
stSG495143	TGACCATGTGTTTGGTAAATAGAGGCCAGC	GCTGAGTCCATGGATTTGCT
stSG495144	TGACCATGCTTCCTCTCTTCTGTCCCCA	TCTGGCTCCTGATTGAAGGT
stSG495145	TGACCATGAGCCCAAAGTGAGCATCAAC	CTCCCCCAATGCATAAAATAA
stSG495146	TGACCATGTCCCAACTTAATGTGCCGAT	TTCCATTTCACATGGCCTAA
stSG495147	TGACCATGTTCCTCCTATCAGCACCACC	TTATCCAGGTTGCAAAAGCC
stSG495148	TGACCATGCATGCTTAGGAGGGGTTGAA	GGGTAATTTCTGACAGGGCA
stSG495149	TGACCATGTCTCCTCTTTTCAATGGGACAT	CTGCCAGCATGATGACAAAT
stSG495150	TGACCATGAAGCTCTCTCTGCAACAGCC	GCAGATGACTTGGCAATGAA
stSG495151	No Unique Sequence	No Unique Sequence
stSG495152	TGACCATGCAAGACTATGTGATAGGCTGGC	GATGGTCAGACACACCAATCAG
stSG495153	TGACCATGGGGGAGGAGACAGGAGAATC	CACAGCCTGCCTTATCCATT
stSG495154	TGACCATGCAGCATTTCCAAAATCAAGC	GAGCACTGTAGGCCTTGTGG
stSG495155	TGACCATGTTGCCACCTGTTCTGTTTCA	CAGATATGGGAAGGTGGGTG
stSG495156	TGACCATGGGGGAAAGAAGCAAAGGAAA	CCAGCTTTATGAAGGTCAGCTC
stSG495157	TGACCATGTCAGATTTATGATGCGGACAA	GGATCACTGACTTTGGCTCTG
stSG495158	TGACCATGCAGAGTTCCATGTGGGTGAA	GCAGGAATTAATGAATGGTGG
stSG495159	TGACCATGAGAATGTGATGCTGGAGGCT	TCACCCCACATGGAACACTA
stSG495160	TGACCATGGGGCTGTGTCTTAGCCATGT	CACAACAACCAATGCCTCAC
stSG495161	TGACCATGTGAGGGCAGAAGCTATGACA	TATACACCAGGCAGGATGGG
stSG495162	TGACCATGGCAAAACCTAAGTGACTCCCTC	GAAGGAATGTTCCCTCTCTTTG
stSG495163	TGACCATGTCTGCAGAACTGTGCTCTTTG	ATCTTTCCACATCCTTCCCC
stSG495164	TGACCATGATTGACCCTGGCTCCTCTTC	TTCCCTTTGCTGTTTTTGCT
stSG495165	TGACCATGCCAATGCATGTACGAGGAAG	CCATGATGACATTCATCCCA
stSG495166	TGACCATGACCCATAAGTTTGGCATTGG	TTGGGAAAGCCATCTGGTAA
stSG495167	TGACCATGATGACCAACTGGGTCTGCTC	ATCACCTTCCAATCAGTGCC
stSG495168	TGACCATGCCAGCTTATCAACCCGATTC	CTGGTGTTTTGCCTTCCATT
stSG495169	TGACCATGGATGATTGTCCCAGGCCTTA	CCACTGTCTAAGGGCGTTCT
stSG495170	TGACCATGGGAAGGACAGGGGAAAAATC	CAATATCCCCTCCTGGATCA
stSG495171	TGACCATGTGAGGGGGAGTCATCAAAAT	ACAAAGGTGGCCACAGATTC
stSG495172	TGACCATGGGATATTGGGCCAGAAAAGC	TCAGATCATGCAGTGTCGGT
stSG495173	No Unique Sequence	No Unique Sequence
stSG495174	TGACCATGCTGAAAAGCATCCTTTGGGA	TCAACTGCAGGAGCACATTC
stSG495175	TGACCATGCAGGCAAAGACTTAGGACACAA	TGACATACAGAAGCATTCGCT
stSG495176	TGACCATGCAAGTGACCCCCAACTCTGT	CCAGCCCAAACAAGAAGTGT
stSG495177	TGACCATGTCAAGTCTTCCCCCTCATTG	CAATTCTGTCCTAAGGCCCA
stSG495178	TGACCATGGCAGAAGGCTTTGTCCTCAG	TGGGTCCCAGATAAGTGGTAA
stSG495179	TGACCATGCATGAATTGGTCAGTCGGAA	GCCACAAAAGGAAACCAAAA
stSG495180	TGACCATGTGTCTCTCTCTCTCGTTGCATT	TGGACATGTTGAATCAGGAAA
stSG495181	TGACCATGAGGTGGTGGCAAATAGTTCG	CAAGACTCAGGCACACATGG
stSG495182	TGACCATGTGTGCTGCTACAGATGTGCTT	TGTCATGGTCAGTCTCCAGC

stSG495183	TGACCATGTGACCAGGCAAAGAGGAAAG	CACAGTCAAATGAGGCAGGA
stSG495184	TGACCATGAGTCTGGGGAGTGAGTGCAT	CCTCACCACCCTCAAGTGAT
stSG495185	TGACCATGCCTGAGATTCTGGGAAGCAG	AATGCTGAAGACCCCTCTGA
stSG495186	TGACCATGTTCCTAAGGGAGCAAAACACA	TGCCTTCTCACTATTGCTGC
stSG495187	TGACCATGGCCTACAGGCTGAGTCCAAG	CTGCCCAACTTGACAAGAGA
stSG495188	TGACCATGAGAGCTCGTTGGGAAATCAA	GATGACTGGGTGGAAAAAGC
stSG495189	TGACCATGTGCAGGGGATCTAAGGACAC	TCACCCCATTTTCCAAATGTA
stSG495190	No Unique Sequence	No Unique Sequence
stSG495191	TGACCATGTTCAAGCAGGGAAGCTGTTT	AGGAGTTTTGGAGGCAAGGT
stSG495192	TGACCATGAGGCCACAATCTTTGCAACT	TGGAGTGTGGAGAACTGCTG
stSG495193	TGACCATGAAGAGGGCATAGGTCACTGG	AAGGAAGGAAGAGGAGGCAG
stSG495194	TGACCATGGCCTAATTTATTCCTTCCCCC	TTCGGTACTGTGTTTGCAATTT
stSG495195	TGACCATGCAGCAGCTTCAGTGCTATGC	CAAGATGCAAATCCTGCTCA
stSG495196	TGACCATGGCATTCCAGCGTGGATATTT	CAACAGGGGTCCCATCTCTA
stSG495197	TGACCATGTCTCTTTCCTCTGACAGCCC	CCTGTCCCTGTTCTGCTGAT
stSG495198	TGACCATGATGGCATAAGAACAGGTGCC	GAATGCTAAAGCAATGGGGA
stSG495199	TGACCATGCTTGTCCCCACAGGGAATTA	TTCCATAACTCCAGGTTGCC
stSG495200	TGACCATGACCGCCTGAGTGATGAAAAA	CCATGTGGCTTACTGGTTGA
stSG495201	TGACCATGGTGTGCATGTGTGTGTTTGG	TGGAAGCCAGAATTGTGTGA
stSG495202	No Unique Sequence	No Unique Sequence
stSG495203	No Unique Sequence	No Unique Sequence
stSG495204	TGACCATGGCCCTGATGTGCTTAGGGT	CTGGAAAACGCAAGAGAAGG
stSG495205	No Unique Sequence	No Unique Sequence
stSG495206	TGACCATGGGAGGCCTGAATGCTCTTTA	CGTGAAAACAAAAACTTCCTGA
stSG495207	TGACCATGCTCAGCCAGCACTCACTCTG	TCTGTGGAATCGGACATTCA
stSG495208	TGACCATGGTGATTGCACAGGTGGATTG	ACATTCATGTGCAGGTGAGG
stSG495209	TGACCATGCAACTCTAGGGGACTGCCTG	CCCAGGGCCTCAGTTTTATT
stSG495210	TGACCATGGCCATTTGCTCACAACAAAG	TTTTTAAACAACCCCCTCCA
stSG495211	No Unique Sequence	No Unique Sequence
stSG495212	TGACCATGCCAGCATCCTGTTTTGCTTT	TGAAAGGACCAGTAGTGCCC
stSG495213	TGACCATGCCCAACAAAGGCTACGGTAA	ATGGCAGCTATGTTGGACCT
stSG495214	No Unique Sequence	No Unique Sequence
stSG495215	TGACCATGCCTCCAGACCACAATGACAA	TTATTCTCCCCTGACCTCCA
stSG495216	No Unique Sequence	No Unique Sequence
stSG495217	TGACCATGGTAATGGCGGCTGATCTTTC	GCCAACATTTCAGGTTGCTT
stSG495218	TGACCATGTCACAGTCTACCGGAGGAGG	GCTTCATTGATTGGCTGGAT
stSG495219	TGACCATGAGCTTCAGTCACCTTCCCG	CACACCTGCTTGGGAATTTT
stSG495220	TGACCATGCCCAAAGAAGCCAAATTACTG	GGGGCTTCCTAAAGCAAACT
stSG495221	TGACCATGCACCACCCAAAGGCATAAAC	CCCAGAGCATCTCAGAGGAG
stSG495222	TGACCATGAACCGAGCATACCATTCCTG	CAAGGGGAAGATGCAGTGAT
stSG495223	TGACCATGTGTCCTTGGTTCTAGGCCAC	GGTTCTGTTCTGGCTCATCC
stSG495224	TGACCATGCACAGCCACCAATAAGTCCA	CAGCGGCACTGTTCTTGTAG
stSG495225	TGACCATGGTGGAAAAGAGCACAGCACA	ATAGCAAATGCACCTCGGTC
stSG495226	TGACCATGACAATATTGGGCAGCTGAGG	GGCCAGCTCTCTTACCCTTT
stSG495227	No Unique Sequence	No Unique Sequence
stSG495228	TGACCATGGGTTTGGCAGAGAGTCAAGG	AGTGTCGGAACAGAGGATGG
stSG495229	TGACCATGTAAACTCCCAGCCTCAGTCC	TTTGATGCCTTGCTTATATTCC
stSG495230	TGACCATGTGTCTTGGTTGTTATGCAGTGA	CCTCCTTGGATGTCGAGAAG
stSG495231	TGACCATGCAAACCGCAATAGAGAGCCT	TAAAGAGGGGCAGCTTCTGA
stSG495232	TGACCATGAGCAATGCAGAGTGATTCCC	GATCAAGATTCGCATCCCAT
stSG495233	TGACCATGTCATCAGTGAAAAGGACAAAGC	TGACCTCCTGATTGTGTGTCA
stSG495234	TGACCATGTTCTGCCACCTCCAGTCTCT	ACCACAAGGCCACACTTAGG

stSG495235	TGACCATGTGAGCAGCCCGTAACTTAGC	GTGCAAGGTCTTTTGCCCTA
stSG495236	TGACCATGCGATCAGCTCTGTTTCCCTC	CACCAACATTCTGTGCCATC
stSG495237	TGACCATGCCTCCCCTGGAGAAAGGTAG	CAGGCTTTCAGCACTCTGTG
stSG495238	TGACCATGTGATGGCTACAAAGGAGCTG	GCCACAACCTCATTGGAGAT
stSG495239	TGACCATGGACAATGTGCCAGTTTGTGG	AGAAGCCGTGCTCATCAGTT
stSG495240	TGACCATGGGACAGCTGAAGGATTAAGGTC	GCCATGATTCCAGCTTGC
stSG495241	TGACCATGGTTGAGTCTGGGAGCTCAGG	TTTCCATTTTGCCACGTGTA
stSG495242	TGACCATGGATCAGCCACCAACAGTGAC	TTATTCACCTGGTTGCGACA
stSG495243	TGACCATGATTGACGTACCACCTCCTGC	TGAGTCTGGGTGAGCTTCCT
stSG495244	TGACCATGCAAAAGGTGGTTGGTTGAGC	TCTCCAGCTCCCTCTTGAAA
stSG495245	TGACCATGCCAAGAGATCCTTCTGCCAA	CTCCATGAGAGCCTTCTGCT
stSG495246	TGACCATGCCCATTCCAAATCTCTACGG	GCTGGCCAAATTGTCAGAGT
stSG495247	TGACCATGAATTCAAGCACAACCTTGACC	CTCCCGTTATTCTGAGCAGC
stSG495248	TGACCATGGGAGTTAAGGGGAGGCAAAG	CCTTTTCTCCCAGAGGGAAT
stSG495249	TGACCATGGGGCTGTTACGGTTACTGGA	AGCTTCTCAGCAAGGAGGTG
stSG495250	TGACCATGCGTCCTCTGGATAAAGGCTG	CACTCTCTCTCTCCCGCTTC
stSG495251	TGACCATGGGTGCTATTGAGGTGGGGTA	CCTTACCTGGGCTTTCACAG
stSG495252	TGACCATGGGAAAACTTACGAAACCCTCAA	TCTTGCTTCAGTTTTTCCTGG
stSG495253	TGACCATGGGACACTATTTCTGCCGACC	GATGGATCCAAATGGCAGAT
stSG495254	TGACCATGCACCTTTCCCACATGCTCTT	CCATCGCCTCTCTCTTTTTG
stSG495255	TGACCATGGCTTTGGTGTCTTTGCCATT	CCAGGGTACCTCTCCCTTTC
stSG495256	TGACCATGGGTCAACTGTTTCCCTGCTG	AAAGGCTGAGCTCGATAGCA
stSG495257	TGACCATGAGATCCACCCCCACACACTA	TCCCTCCTGTTATTTGCTGG
stSG495258	TGACCATGACCAGATTGCTATCAGCCCT	GGGCCATTCGTCTTTAAGGT
stSG495259	TGACCATGAACAGCACGAACAAAAACCC	CCCCAGGACATATCAACCAC
stSG495260	TGACCATGTTGGCTGATTTCTCCCACTC	TGGTTCTCCTTCCAAGATGG
stSG495261	TGACCATGCTTGAGCTCTGGTCAGGTCC	TGCGACCACTTTGCAGTAAG
stSG495262	TGACCATGCAGCATGTGCTTGGCAGTAG	GAAATTTCCGTTCCCTGTGA
stSG495263	TGACCATGGATTCATCGACCCCTAGTCCT	TTTCTCTGCTCATGTGGTCAA
stSG495264	TGACCATGGTGGGAGAGTTAAGGGGAGC	AGCATGGCATGTTGTGGTTA
stSG495265	TGACCATGCCCAAGATGTCTGCTGTTCC	TCTGAGACAGATGCCAGGTG
stSG495266	TGACCATGAAGCCCTGCAGATACAAGGA	TTCAGCGAGTAGGTTACTTCCA
stSG495267	TGACCATGGCTCTAAAACCAATTCACCACA	CAGGGTATCTGACAGCCCAT
stSG495268	TGACCATGCGCCTTAGGCACTTTCATTT	CATCCCTGCACAACTCTCAA
stSG495269	TGACCATGAGCTCCATGCTGTGAGGTTT	GGACTGGTTTATGCTGGGAA
stSG495270	TGACCATGAGGCTTTGTATGGGGGAGAT	GCACAAGACTGGCATCTCAA
stSG495271	TGACCATGAAGACGCTCAATGTCCCAGT	CTAGGGTTAGGAAGGGCAGG
stSG495272	No Unique Sequence	No Unique Sequence
stSG495273	TGACCATGCCATTGCCAATACAATGCCT	TTTTCCCACACACCTGGTAA
stSG495274	TGACCATGGTCTCCCTATTTTCACCTAGCC	AGTGATCCAAGGGCTGACTG
stSG495275	TGACCATGGCTGGTTTGGGAGAAGAAGA	TTCCATTGCCACTGAGATGA
stSG495276	TGACCATGGGCCAAACAAGCTGTGAAAT	ATGAAAACCAAACCAGCCAC
stSG495277	TGACCATGTGGGGAACTACAGGGGTTTT	CCCCTATACTTTCTGAGGGCA
stSG495278	TGACCATGGCTCTGTGTTGTAATCGCCA	TCCCCAAGGTAAGGCTCTTT
stSG495279	TGACCATGAGGTCCATGCCGTAGTTGTC	TGCAGAGGTTTGTGACTTGC
stSG495280	TGACCATGCTAGGCTTCTTCAGCCCTCC	GCCAGCTGCTCTATCTGTCC
stSG495281	TGACCATGCCTGCCCACCTTGTATCATT	CACTTATGCCACAAGCCTCA
stSG495282	TGACCATGGGATTCATCGGAACCAACAC	GGGAATCATGTTCTCACGGT
stSG495283	TGACCATGTGCTGTGGGACTACACAAGG	GAGAGGGTGGGAGGCACT
stSG495284	TGACCATGTCATGATGGTCAAGAGCCAA	ACCTCACTCTGCCCATTCAC
stSG495285	TGACCATGGATTGGCAGAGGCATTGTTT	AGCAGCTGGAACTCTGAGGA
stSG495286	TGACCATGGGCATCTGTCCCTCACTAGC	GCAAATCTCATGACCCCCTA

stSG495287	TGACCATGCCTTTTTAGGCCTTTGGTCC	GTCTCCCTACCCCACCAAAT
stSG495288	TGACCATGCTGGCTCCTGCTTCTAGGTG	TGAAAGCCACTTGCTGTGAG
stSG495289	TGACCATGGAGGTCAAATCACTGGGGAA	GGGCTGAAACAAACCTGAAA
stSG495290	TGACCATGCTCTCCTCCTTTCTCCCCAC	AGATTTTCCCCAGCGTAGGT
stSG495291	TGACCATGTCTTTGCCAGCAGTCTCTGA	AGAGGCACAGAGGGAACTGA
stSG495292	TGACCATGCCACCCACCTGTTTACTTGC	AATAGATGTGTGGCGGGAAG
stSG495293	TGACCATGGTGGATTCTCAGGGATGCAG	GAGGGAGAAAGAGGTGGGAC
stSG495294	TGACCATGGCAGGTAAATGGTGGCTGTT	CTGCCAGCTAGCAACTGATG
stSG495295	TGACCATGACCGAGCTTAGCGGTATGAC	AGTCTTCATGGCCACACTCC
stSG495296	TGACCATGCTGCATGCTTCACAGGTAGC	CCTTGAGAATGCACCACAGA
stSG495297	TGACCATGTCTCCGATTCTCTGCGTCTT	CTCTCCTCAGCCTCAACCAC
stSG495298	TGACCATGTACCGAGGTGTCTGTGTCCA	AGACCTTCTCCCTCAGCCTC
stSG495299	TGACCATGCCCTTAGCTCGGTACCAACA	AACTGTGGAAATGGCTTTGG
stSG495300	TGACCATGTTGTTACTGCCTTTGGGGTC	ACTGCTTTATTCTGGTGGGG
stSG495301	TGACCATGACAAGGAGGAGGGTATGGCT	AGATCTGTCTGGCTGCAGAG
stSG495302	TGACCATGCAGCCCATAGTTGTCCCATT	AGTGACACTCGTGCTCATGC
stSG495303	TGACCATGTATTCTTCCCCCAGGATTGC	GGAATGTGAGGCATGGATTT
stSG495304	TGACCATGAAAAGAACCAGAGCGTGGGT	AGGCAGGACCAAAAATTCCT
stSG495305	TGACCATGTTAGTGCACACCAGGGACAC	TAATTTGGCAAGCCATCCAT
stSG495306	TGACCATGTGCCTGGACGTGCTTATGTA	GTCACTGTGGTCAAGAGGCA
stSG495307	TGACCATGTGCCAACTAACCTTGCCTCT	GAGAGAGCGAGAGAGGGGAT
stSG495308	TGACCATGCAAAATCCCTCCACACCTGA	AAATGGCTTCCTTCCTTGGT
stSG495309	No Unique Sequence	No Unique Sequence
stSG495310	TGACCATGCTGCATGCACAGTCAGTCCT	GCAAGAAAAGCACAGGAAGG
stSG495311	TGACCATGTAAAACATCACAGCCACCCA	CAAGAGAGGCTGGGAATGAG
stSG495312	No Unique Sequence	No Unique Sequence
stSG495313	TGACCATGACAAGACCAAGGGCTTCTCA	GCCCCAAAGCAATTTTCATA
stSG495314	TGACCATGGGGTTCCTTAGCTGTGGTGA	ACTGTTTGCCATTGTCCCTC
stSG495315	TGACCATGCACTGAATGACAGTGAGCCC	TGTATGGAGCAGATGCTTTGA
stSG495316	TGACCATGTTGAGCCTGAGCAAGTGTTG	TCTCACAGAAACACCATCGC
stSG495317	TGACCATGCTCGGTATAAACTCGCCTGC	ATGTGGCACCATCTCAAACA
stSG495318	TGACCATGGCAGAACAAACTCAAGGCAA	TTGTTAAGTCAACCAACCCAA
stSG495319	TGACCATGGTGGGAGGCAGTGGTATCAT	TGAGCAATTAGGGTTCTGCAT
stSG495320	TGACCATGCAAGAATACAGGGGGCTGAA	TGGAAAGTGGTATTCCTGCC
stSG495321	TGACCATGATCCAGGCCAATCAGCTATG	ATGGGGATAGTGACCACCAA
stSG495322	TGACCATGCCCACATGCTTAGCTACTGAAA	TGAGGTGATATTCTGTGTGCAT
stSG495323	TGACCATGAGAGTCACAAAAGGGCCAGA	TTTATTTTTCTGTGATGGGCA
stSG495324	TGACCATGAGGCTCATTATGATGGGTGG	TTCACATTGTACACCAAAAGGC
stSG495325	TGACCATGTGCCATCTACAAACACACCC	CCCTAGTTCTGTCTTCCGCA
stSG495326	TGACCATGTCTCTCTGTTTCCAGGCTGC	TTCTGAAATTTCCATCCCTGA
stSG495327	TGACCATGCAAAAGGGTCCAAACCTCAA	AAGTGCGCTCTTGGAAGAAA
stSG495328	TGACCATGTGGCTGGTTTTCGCTTACTA	TTAGGTAAAGCAAATGTGGTCG
stSG495329	TGACCATGTTTTTCCCCATGACCTGAAA	TTTTGACCTTTGCCCTGAAC
stSG495330	TGACCATGTTGGAAGGGATAGCACAAGG	GTTGTTTTCTGCCTCCATCC
stSG495331	TGACCATGTTCCTTCCCCGTTGTAAGAT	GCAAGGTGGGACTGAAGGT
stSG495332	TGACCATGTTGGGGGAGCAGTTTCTCTA	GTTGTGAGCTGGTTGCTTCA
stSG495333	TGACCATGGGCACTCAGGCAACTGTCTA	CCTAGAGGGCCTTTTTGCTT
stSG495334	TGACCATGGCAGGACTTCTAGGGCCTTT	TGGCAGACATTGCAGTTAGC
stSG495335	TGACCATGTGCCGAGACGTGTGTTACTC	TATGGACAAAGTGGGGGAAA
stSG495336	TGACCATGTGGGAGCACAGTTTATGCAA	GATGGCTCTTAGGGGTTTCC
10Kb Resolution Array		
stSG495337	TGACCATGTGCAAGTGCACATACACACAC	CTGTGGTTATGGGGTGCTTT

stSG495338	TGACCATGATAACCACAGCAGAAGGCGT	TGGACCATAGCCTTTGTGAA
stSG495339	TGACCATGTGCACATGGAGAGAAGTGCT	CGATCTATTGGGCTTCCTGA
stSG495340	TGACCATGGCCCAATAGATCGCTGAAAA	ATTCTTCCTGGGTCAGGGTC
stSG495341	TGACCATGGACCCTGACCCAGGAAGAAT	TTCTGAATGCATTTGGCTAGAA
stSG495342	TGACCATGGGTGCACTGATCCAGGAATAA	ACAATGACCTTTGGACAGCC
stSG495343	TGACCATGGGCTGTCCAAAGGTCATTGT	TCTGGAATTTGAAGCTGGGT
stSG495344	TGACCATGACCCAGCTTCAAATTCCAGA	TGGGAGCAGCTCGTAGTCTT
stSG495345	TGACCATGGCACTGGACAGAATGCACTG	CCTTCTGTGTGCTGCCTCTA
stSG495346	TGACCATGAAGATGACTTCATCACCATTGC	TATGGCACTTGCTTCATTGC
stSG495347	TGACCATGGCAATGAAGCAAGTGCCATA	TGACTCAGCCTTTGATGCAC
stSG495348	TGACCATGCTGAGTCAGAACACCTGCCA	CACTGAGGGAAGAAAAAGCG
stSG495349	TGACCATGCGCTTTTTCTTCCCTCAGTG	GCAGGGCCTGTTTGATTTTA
stSG495350	TGACCATGAATTCCAGTCCCAGCAAATG	CTGCTGGAGGATCAAGAAGG
stSG495351	TGACCATGGAAGCCCTTCTTGATCCTCC	GGGTGTAAACTCCAAGCCAA
stSG495352	TGACCATGGAGGGGAAAAAGATAACGCC	GGGGAATGGAAAGAGGAGAG
stSG495353	TGACCATGCAACAACTGCAACTTTGGGA	GTTGCACAGACATCAAGGCT
stSG495354	TGACCATGAAGCCTTGATGTCTGTGCAA	TTCTGACCAGGTGGGTAAGC
stSG495355	TGACCATGGCTTACCCACCTGGTCAGAA	TCCTTTCCAGGGTTTCCTTT
stSG495356	TGACCATGCCATACGTGGAAACTTGGCT	CACAAAGTAGGGCCAGGGTA
stSG495357	TGACCATGTACCCTGGCCCTACTTTGTG	TGTTTGGGGAGGTGGAAATA
stSG495358	TGACCATGGAAAACGATTCGATCCCTGA	TCCTCTGTCCAATCCACACA
stSG495359	TGACCATGTGTGTGGATTGGACAGAGGA	CGGGGAGAATGAAAAGATCA
stSG495360	TGACCATGATCTTTTCATTCTCCCCGGT	ATGCTGTGGTGGCACATTTA
stSG495361	TGACCATGTAAATGTGCCACCACAGCAT	CTTTCGAAGCACCCATGATT
stSG495362	TGACCATGGTACCATGCCTGGTTTGGAT	AGGTCCCAACTGCAAAAGTG
stSG495363	TGACCATGTATTGCCATCCCCAGAACAC	TGTGTGTGTGACAGGCTGAG
stSG495364	TGACCATGCTCAGCCTGTCACACACACA	TCCCAGCCTTGTGACTTTCT
stSG495365	TGACCATGGAAAGTCACAAGGCTGGGAA	AAGCACTGGAAGAACGAGGA
stSG495366	TGACCATGCTACCTCAATGTTGCTGCGA	CTGGTTACAAATCCAAGCCC
stSG495367	TGACCATGTTTGTAACCAGGGTGGGTTC	GATGTGGCTTTCTGATGGGT
stSG495368	TGACCATGTCGTTGCAGATAAGGGAAGC	GGAGGACTCCCTTCTACTTGG
stSG495369	TGACCATGAGAGCACTTGCTGTAGGTCCA	GTAGGGCTCTAGACCTGGGC
stSG495370	TGACCATGCAACAGGTTACCCAGGTGCT	AGCCTACGCTGGTTTCTTCA
stSG495371	TGACCATGTCTCTAGCATGTCCTTGGCA	AGAATGCCACAGTGGGGTAG
stSG495372	TGACCATGCTACCCCACTGTGGCATTCT	CTCTTTGGAGGCAACCTCAG
stSG495373	TGACCATGCTGAGGTTGCCTCCAAAGAG	CAAGGGGAGAAAATGGTGAA
stSG495374	TGACCATGAATAGAGGTGTTGGGCTTGG	TGGTTGCCATTCAATAGCTC
stSG495375	TGACCATGTGAATGGTCCTGGAGGTAGG	AAGTGGGTGGAGATCAGTGG
stSG495376	TGACCATGCCTCTGCAAAGTCTTGCCTC	TCTGATGCCTTAGCTGGGTC
stSG495377	TGACCATGGCAACAGACAACAAGCCAGA	AGGAAAGGAAAGGCTGGAAA
stSG495378	TGACCATGTTTCCAGCCTTTCCTTTCCT	GGCTGATGAAACTCTGAACCA
stSG495379	TGACCATGTGGTTCAGAGTTTCATCAGCC	GCATTTACCAAGCCCTTCAA
stSG495380	TGACCATGGGGCTTGGTAAATGCGAGTA	TGCATCTGGACAAGGAAGTG
stSG495381	TGACCATGATCTGGCCCATTACAGGAGA	ACACTAAGGGGGAATGTTGG
stSG495382	TGACCATGACCAGGCCAACACTGGTACT	GGATGGGAGGTAAGCACTCA
stSG495383	TGACCATGTGAGTGCTTACCTCCCATCC	TGGACTTCGTAAGTCTGGGG
stSG495384	TGACCATGCAATCCCTAGCCCGTAGGTT	CTCTCCCGGGTAAGTGCTC
stSG495385	TGACCATGTAAGATAGCCACTGCTGCCA	GGGGAGTTGTCAAGTTTCCA
stSG495386	TGACCATGGGAAACTTGACAACTCCCCA	CACCGAAGAGGTGAGGAGAG
stSG495387	TGACCATGAGGCTTTCTGACTCCTTGACC	TAGCTGGTACACGTTGGCAC
stSG495388	TGACCATGCTTCTCCTGAAGAGGGCAGA	CATCAGAATCCTCAGCCTGG
stSG495389	TGACCATGCCACTAAATTGGAGAGGCCA	AATGGTGGTCAGGAAAGCAG

stSG495390	TGACCATGAGATGGAGATGAAAAGCAAGAA	TGGATGCAGATGGATGAAAG
stSG495391	TGACCATGGCAAAGTGAGAGAGAGATGTCC	CCATCCTTTCATTCCTTTAACC
stSG495392	TGACCATGAAACAGTTTGATTGCTGGGG	TTGCAAGTGAAGTCAAAGCG
stSG495393	TGACCATGGTCACTGAATGCCAGCTGAA	GTGCTCTCCAACTAGCCCAG
stSG495394	TGACCATGTGTGCTTGGCATAGTGGGTA	CTCAAGGACCCCAAACTTGA
stSG495395	TGACCATGATGCACACTCATTGGCACTC	ATAGGGCCTGGAAGTGTCCT
stSG495396	TGACCATGTTCTTGTTGGTTGTTTGCCC	GGCAACCACAGTAAAATGCC
stSG495397	TGACCATGTAGGCGATGCCTGAGTATCC	AAAGTGTTGCCAGTCTTGCC
stSG495398	TGACCATGGGCAAGACTGGCAACACTTT	CTTAAGCCTGAGCCCAACTG
stSG495399	TGACCATGGGGAAGTGTGTTGAATGGCT	TCACAATGGCCTCAATGGTA
stSG495400	TGACCATGTACCATTGAGGCCATTGTGA	TCGTGAAACACTGGAACTGC
stSG495401	TGACCATGAGTGTTTCACGAAAGGGGTG	GTTACCCTTTTGACAGCCCA
stSG495402	TGACCATGGATGGCCACATGCATTATAAGA	CCAAGTCAAGCACTCTTTGC
stSG495403	TGACCATGACCTGTTTCTCCCCATCTCC	GGGAAAGTTTCTGGCTAATGC
stSG495404	TGACCATGGCATTAGCCAGAAACTTTCCC	CCTGCTACTCCAACCCTTTG
stSG495405	TGACCATGCCTTATACTGCCTTTGGGGC	CCCAGATTTAAGCGGTTCTG
stSG495406	TGACCATGGCTCAAGAATTTTTGGCCTG	TGCCTCTGGCTCTATCTTCA
stSG495407	TGACCATGAACCTTGGAGGGGTCTGTTC	TCAATGATTTATGCCTCCACA
stSG495408	TGACCATGATCCAGAGCTTGGAGGAACA	TGATGTGCTGATGTGCAAAG
stSG495409	TGACCATGTTTGCACATCAGCACATCAG	GAGGGAATGGACTGCAAAAA
stSG495410	TGACCATGTCCATTCCCTCTCCACAATC	AACATGGGCTGTCTCCTGAC
stSG495411	TGACCATGACCTAGTGGGCCTGAGGTCT	GTGATGAGCCCAGGGTAAAA
stSG495412	TGACCATGGTGCCTTATGGGTTCTCCAG	CAGCAGTGATAGCTCACCCA
stSG495413	TGACCATGAGATAGAGGGAGGGGTCAGG	TTTGGCCTCCAGCAAAATAC
stSG495414	TGACCATGGGTCTTCACCTGTCTGTGGG	ATCTGTTTCGATGGATGCCT
stSG495415	TGACCATGACAGGAGGGGCCAGAGTAGT	CAGCACAGCAGAAGATCAGC
stSG495416	TGACCATGGTCCTGGAAAGGGGGATAAG	CTTTCCCATTGCATTCCATT
stSG495417	TGACCATGATGGAATGCAATGGGAAAGA	AAGGAAAATGGTGTCAAGCAA
stSG495418	TGACCATGACGTTCAAGCATCACAGTGC	ATTCCCCAGAAGCAACAAAA
stSG495419	TGACCATGTTAAAAACCCCACCAGCAAG	GCCAGAGCAAAACTAGCCAC
stSG495420	TGACCATGGTGGCTAGTTTTGCTCTGGC	GCTTTGGACACTTCCTCTGC
stSG495421	TGACCATGGCAGAGGAAGTGTCCAAAGC	ATCCACCACAAGTCCCAAAG
stSG495422	TGACCATGTCTGAAAGCATGAGGCTTGA	CAAATCCCTGGCCATCTAAA
stSG495423	TGACCATGTTAGATGGCCAGGGATTTGA	ACTCCCCTGAGAAGCTGTGA
stSG495424	TGACCATGGACGCTCTTCATGGGATCAT	TAACTTCTGCCCCAACTCCA
stSG495425	TGACCATGTGGAGTTGGGGCAGAAGTTA	GCTGCTCTGGAGAGGAGAGA
stSG495426	TGACCATGGGTCCGAGGTCTCCTCTTTC	TTCTGGGGTTTTCACCAAAG
stSG495427	TGACCATGTTGGTGAAAACCCCAGAATC	TTTTTCCCATAAATCAACCCA
stSG495428	TGACCATGTGGGTTGATTTATGGGAAAAA	CTGCTGGCTTTTGTTGTTGA
stSG495429	TGACCATGTTGTCTTTGCAGATCAGGGA	TAATCCCTCTTCCCTCCCAC
stSG495430	TGACCATGGTGGGAGGGAAGAGGGATTA	GCCTAGCAACAGCATCCTTC
stSG495431	TGACCATGGAAGGATGCTGTTGCTAGGC	CCATCGATTCCCAGATGTTT
stSG495432	TGACCATGGTGTGAGCCCAATCCAAGTAG	CCGCAAGGGCTAAACAGAAT
stSG495433	TGACCATGTTTGTCCCTATCCCTCCCTC	GGCAGGCTCAGATCTGTAATG
stSG495434	TGACCATGCTGTCGGACATAGATTACTGGC	ACAATACAGTCGCCACTCCC
stSG495435	TGACCATGGTCTGTCTGCAGCATCCCTA	AGCGTACCTGTACCCCCAG
stSG495436	TGACCATGTACCCCTCAGCTCCGGTAGT	CAGTTGTGGAAATTGTGCGT
stSG495437	TGACCATGGGTGAGTCATGGCTTCTGGT	AGGGAAACAAAGCAGCTTCA
stSG495438	TGACCATGTTGGGGACAATAATTTGGGA	TCCAGGTAACCAGCCTCAAC
stSG495439	TGACCATGGGCTGGTTACCTGGACAGAA	ATGGGTCATTTGTCACAGCA
stSG495440	TGACCATGTGCTGTGACAAATGACCCAT	ACATAGCGCAAACCCAAAAG
stSG495441	TGACCATGGGTTTGCGCTATGTTCCACT	ATTCAGACTCCAGCGCTCTC

stSG495442	TGACCATGGAGAGCGCTGGAGTCTGAAT	GGGGGACGGGATACAGTAAT
stSG495443	TGACCATGATTACTGTATCCCGTCCCCC	TCCCATATGAAGCCAAGGTC
stSG495444	TGACCATGCCACCAGCTCCAATCAGACT	CAGCCCCATATTGGTACAGG
stSG495445	TGACCATGCTGTACCAATATGGGGCTGG	TGAAAGACAGCAAGGGAACA
stSG495446	TGACCATGGCACTGCTGGTTTTCTGTGA	AACAGGTTCCAGTCTGGGTG
stSG495447	TGACCATGCTAAGCCCTTGCCAACCATA	AAAGGTGATCATTTGCACCA
stSG495448	TGACCATGTTATCTGTGTGTTTCTGCCTCA	CCTGCTACCCGTTTCTCATT
stSG495449	TGACCATGGAGGTGTGGTGCTCACCTTT	TGCAAAGCATACGTTTCGTC
stSG495450	TGACCATGTTTGCATTAAGAGTTGGGCA	AACGGGCTTGATTACTGCAC
stSG495451	TGACCATGGTGCAGTAATCAAGCCCGTT	CATTTCCCTCAATTGTGCCT
stSG495452	TGACCATGGCACAATTGAGGGAAATGCT	GGGTTCAGAAATGGGAGACA
stSG495453	TGACCATGGTCTCCCATTTCTGAACCCA	GGCAACTTGAGAGCAGGAAC
stSG495454	TGACCATGGTTCCTGCTCTCAAGTTGCC	TTTGCCATTTCCTTTCATCC
stSG495455	TGACCATGGGATGAAAGGAAATGGCAAA	CGCTCAACTTCCACTTCTCC
stSG495456	TGACCATGGAGAAGTGGAAGTTGAGCGG	AGCCATCCACAGGCATAAAG
stSG495457	TGACCATGTGCCTGTGGATGGCTTTATT	AGCCCTCCCAATCTTACCAC
stSG495458	TGACCATGCCCCCAACAAATGTCACTCT	GGGGAATGAGAACTATCCAACA
stSG495459	TGACCATGTGTTGGATAGTTCTCATTCCCC	GGCAGAAACTGTTGACACTCTG
stSG495460	TGACCATGTGTCAACAGTTTCTGCCTTCA	TGACATCACCAGAGGGTTCA
stSG495461	TGACCATGATCACTGGGCTCTTTCATGG	GCAGCTGCAATCTTTTCACA
stSG495462	TGACCATGAGGAAATGCAAGCCCATACT	CAGATCCCCTCCATATTGGTT
stSG495463	TGACCATGGGAGGGGATCTGTGTTTCAT	AGGCCTTCAAAGCAACAATG
stSG495464	TGACCATGAACTTGAAATTCCTTGGCTACG	ATTTGGCTCAAGGGCTTTTT
stSG495465	TGACCATGCCAGGTTGCTGAAGGAAAAC	ACTAGCAAACTGCAGCCGAG
stSG495466	TGACCATGCTGCAGTTTGCTAGTGCGTC	AGCTGTGCCCAACCTCTCTA
stSG495467	TGACCATGGCTAGAGAGGTTGGGCACAG	AGCAGAAAAGAGGGCAGTCA
stSG495468	TGACCATGTGACTGCCCTCTTTTCTGCT	CTAAGTGCCTGCAAAGAGCC
stSG495469	TGACCATGGGCTCTTTGCAGGCACTTAG	GCCACCATTCAACTTGACAC
stSG495470	TGACCATGCAAGTTGAATGGTGGCATGT	AGCCAATGGTCTCTTCTGTCTC
stSG495471	TGACCATGGAGACAGAAGAGACCATTGGC	CTCCACAGGAGTGGGTCATT
stSG495472	TGACCATGGGATGTTCGTTCCGTCTTGT	TGGCCTTGTAGCAGGAAATC
stSG495473	TGACCATGTACAAGGCCAAAAGCCTGAT	TTCTTGCTGCCAATTGTGAC
stSG495474	TGACCATGTCCTGAGGTCCTCTTTCTGC	CTCTGCTTTTTCTCGGTGCT
stSG495475	TGACCATGAGCACCGAGAAAAAGCAGAG	TACACCAAACTGGGCAACAA
stSG495476	TGACCATGATTCAGGACATCGTTGGGAG	CGTCCTGCATTAAACAGTGG
stSG495477	TGACCATGGAGCAGCAATGATCACCCTT	GCTGATGACTACTCCAGCACA
stSG495478	TGACCATGCAGCTGTACAGGAAGAGGCA	ACTATTCCCAAGGCCAACCT
stSG495479	TGACCATGGGTTGGCCTTGGGAATAGTT	ACCAAATGGCCTTTCAACAG
stSG495480	TGACCATGCTGTTGAAAGGCCATTTGGT	GCACATAACATTCCAAGCCA
stSG495481	TGACCATGATTGCCAATGTCTTCCTGCT	CACACATCCCCTGCATAGTG
stSG495482	TGACCATGGCAGGGGATGTGTGTATGTG	CTTGCTTGCTTCCATGACAA
stSG495483	TGACCATGTTGTCATGGAAGCAAGCAAG	CCCAGCCACATAAAACCTGT
stSG495484	TGACCATGGTGCCATATGCATGAGCAGT	GGTCACCATTCTTTGGCACT
stSG495485	TGACCATGATGGTGACCTTGCTTCTGCT	GAAGGCTGGGCATCAAGTAA
stSG495486	TGACCATGCAGCCTTCCAATTTGTCTCC	GCAGAGTTCCAAACAGCACA
stSG495487	TGACCATGCCCAAGGAGAGGTCTCATGTT	TCCGTCCCTGCTGAATTAAC
stSG495488	TGACCATGTAATTCAGCAGGGACGGAAT	GTCTCATGGCGACCCTAAAA
stSG495489	TGACCATGCTTTTTGCCCTTTCCCATTT	GCCAGGCATCCTGATTTTTA
stSG495490	TGACCATGCTGATTTGGAGCTTGGAAGG	GCAGGGTGTAACCATGAGGT
stSG495491	TGACCATGTGCCCCAGATCCTTCTAATG	GTCAGGTGATGGCAAGGAAT
stSG495492	TGACCATGTTCCAAGGGAGTGGTGAAAG	AAGCCCACCACCCTTACTCT
stSG495493	TGACCATGAGAGTAAGGGTGGTGGGCTT	CCTTCAAGCTGGCTTTTGAC

stSG495494	TGACCATGAGAGGGCAATGTGAAGAGGA	TGGAAACATTGTAGGTGCCA
stSG495495	TGACCATGTGGCACCTACAATGTTTCCA	AGGAATGCCGTTTCCTTTTT
stSG495496	TGACCATGGACCCTTTCCTTGGGAAGTC	CCTCCAGGTTCCTCAAAACA
stSG495497	TGACCATGCCTGTTTTGAGGAACCTGGA	CCCAAGACCCATTTCTTTGA
stSG495498	TGACCATGGGGCTACCCCAATCATCATA	AAAGAATTCCAAAAGCGGGT
stSG495499	TGACCATGACCCGCTTTTGGAATTCTTT	GACAGTCCCTGCGTTGAAGT
stSG495500	TGACCATGGTATACACGGAGGGTCACGG	CAAGCTCAGTCTCCTCAGCC
stSG495501	TGACCATGTCTCACGGGTATTTTCCACA	TGGCAAGAATAACCCCACTC
stSG495502	TGACCATGTGAAAACTACACCACGCAGG	TGATGCTGCAATTTAATCCAA
stSG495503	TGACCATGTGGAAGTGAGGAGTAGGGCT	CGAATCAGGGGAAACTGAAG
stSG495504	TGACCATGCTTCAGTTTCCCCTGATTCG	AATGCCCAGTGAATTAACGC
stSG495505	TGACCATGGCCTAAGCACAGACATGAAGC	TAACTAATGCAGTGCCCCGT
stSG495506	TGACCATGTCTCCTGCTTTTCCAGAAGG	TGTGCACCAAGAAACCAAAG
stSG495507	TGACCATGTTTGGTTTCTTGGTGCACAG	TGCGAGGTAAAAGTTGAGGC
stSG495508	TGACCATGGCCTCAACTTTTACCTCGCA	AGAAAGCATGCAGTGAGGGT
stSG495509	TGACCATGCTCACTGCATGCTTTCTTGC	CCCACCATGGATTACCAGAC
stSG495510	TGACCATGGTCTGGTAATCCATGGTGGG	GGGTAAGACCCTCACGATCA
stSG495511	TGACCATGCAGGATGGTGAAGAAGGGAA	GCCGAATTGAACTACCTCCA
stSG495512	TGACCATGGTCCTCCATGCAAATCACCT	CTTTGAGAACAGCCCAGCTC
stSG495513	TGACCATGGAGCTGGGCTGTTCTCAAAG	GTGGATAAGCTGTCCCGTGT
stSG495514	TGACCATGCCGTTCTCACCTGGTTTCAC	CTTGGTGGGAATTAGCCTGA
stSG495515	TGACCATGCAAGCACTGGAACAGCACAC	GGAGCCTGAGGGATCCTAGT
stSG495516	TGACCATGGGGGAAACTAGGATCCCTCA	GGGGATTCCAAAATGAACCT
stSG495517	TGACCATGGAATCCCCACGGTAGAGACA	TTAGCCATTCAGAGGGTTGG
stSG495518	TGACCATGCCAACCCTCTGAATGGCTAA	CCCACTCTGGAGAACAGCTC
stSG495519	TGACCATGTGTTCATCCTGGACTCCCTC	CCTCCATGTCTTCCCAGTGT
stSG495520	TGACCATGACACTGGGAAGACATGGAGG	ACAGGCCTAAGGGAAGGAAA
stSG495521	TGACCATGTTTCCTTCCCTTAGGCCTGT	CTTCTCTCCCTCTACCCGCT
stSG495522	TGACCATGAGCGGGTAGAGGGAGAGAAG	ACATCAAGTGGCTGGAAAGG
stSG495523	TGACCATGCCTTTCCAGCCACTTGATGT	TCTCACATGCTCCGTGCTAC
stSG495524	TGACCATGGTAGCACGGAGCATGTGAGA	CTGATCAGAGAGCCCAGAGG
stSG495525	TGACCATGCTCTCTGATCAGGGTCCTCG	CTATCCCCACAGGAGCAAAA
stSG495526	TGACCATGTTTTGCTCCTGTGGGGATAG	GCTGCACCTAATCCAGAACC
stSG495527	TGACCATGGCTGGTTCTGGATTAGGTGC	CTTAAGGCTCCTCCTCTGCC
stSG495528	TGACCATGGCTTTTTGAGTTCACAGCCC	TCTCAAGCGTCCTTCCATCT
stSG495529	TGACCATGATGGAAGGACGCTTGAGAGA	AGCAGATCAGTGACGAGGGT
stSG495530	TGACCATGTCCAGTTCCCAGAGATGGAG	GGCCTTCCTAATCTTCACCA
stSG495531	TGACCATGGGAGAATGAGGGCAGTGTGT	CTGGATTCTCCCCCAGTGTA
stSG495532	TGACCATGAGGGTGAACTGGTGAGAGGA	TTACCGAGTTTCCTGGACCTT
stSG495533	TGACCATGAAAACTGGGACAAGGTGTCG	TCTGTGTGGGTAGCTTGTGC
stSG495534	TGACCATGAGTCAGTGCCCCATAAATGC	TTCATGGCATCCCTACTGGT
stSG495535	TGACCATGCCTCTATTTCCACTGGGCAA	TTTGGGGACAAATCAAGGAG
stSG495536	TGACCATGCCAAAACCCTCAGCAAGGTA	TCATCCTCCCACACAGATCA
stSG495537	TGACCATGCACCCTATGCCAGGAACAAG	TACACACCATGCACACATGC
stSG495538	TGACCATGGCATGTGTGCATGGTGTGTA	CCTCTCTGTGTTCCTGGCTC
stSG495539	TGACCATGCAGAACAGAGGCTGACTCCC	CCCTGAGATGGTTCAAGGAA
stSG495540	TGACCATGGGTCTTTGTTAAAGCAGCCAA	TTTTGGCAATTCCGATTCTC
stSG495541	TGACCATGTTCTTTGGCACCTTGGTTTC	TGCTTTCTCCCTTTGCTCTC
stSG495542	TGACCATGAGCAAAGGGAGAAAGCACAG	GCCTCTCCTGAAGCTTTGAA
stSG495543	TGACCATGTTCAAAGCTTCAGGAGAGGC	CCTTCTAGTTTCTTGCCCCC
stSG495544	TGACCATGACGGATTCTACCCCTGGAAC	GGCTTCCTGTTTTCAGCTTG
stSG495545	TGACCATGGCCCTCAATGAGCTGTGATT	TGCAAGAGGGAAACAGATGG

stSG495546	TGACCATGAATGCCATCTGTTTCCCTCTT	AAACCCATTCAGAAGATTTGGA
stSG495547	TGACCATGTCCAGAGGTGTTTGAGAGGAA	CAGCCAATCATCAAAGAGCA
stSG495548	TGACCATGTGCTCTTTGATGATTGGCTG	TTGCATTTATTGGCCATCTG
stSG495549	TGACCATGCAGTTTGCAGATGGCCAATA	AAGGCCAGAGTAGGCTGACA
stSG495550	TGACCATGAATCTTACATGGGGGAGCAG	CATGCTGGTAAATTGCCTCC
stSG495551	TGACCATGCCTCTTACGAAAGCTGAAGGC	TACCCCTTTGGAATGAGCTG
stSG495552	TGACCATGGCAATGGGGACTTGCAAAA	CCAAAAGTCATCACATTAGGGC
stSG495553	TGACCATGTACCCAATGACCCAATGACC	GAAGACTTCTGCACCCATCC
stSG495554	TGACCATGGACGTATCCAGACAAGCCCT	GGGGCCAATCTAATCCTTCT
stSG495555	TGACCATGTTAGATTGGCCCCTCTCCTT	GATTCCAGTGGGGGATACCT
stSG495556	TGACCATGAGGTATCCCCCACTGGAATC	TTATCTTCCCACCCAACCCT
stSG495557	TGACCATGCCACCCACAAATGGGAAAG	AAAGGTCCTCTGCTGCTGAA
stSG495558	TGACCATGGCTTGCCTATGGGTGTGTCT	GATGTGGAGGAATGTGGCTT
stSG495559	TGACCATGAAGTCCTGAGGAGCCCATTT	ATGCAATGAAGGTGGGAAAG
stSG495560	TGACCATGCTTTCCCACCTTCATTGCAT	GCTTGGCTTGGTCTGTTTTC
stSG495561	TGACCATGAGGACACAGGATCAACCAGG	CAGTTGACATGACCCTCCCT
stSG495562	TGACCATGAAGTTGATGGATCAGGGTGG	AGGTCAGCTCTGCACCACTT
stSG495563	TGACCATGCGCAATCCTTAGGCAGTGAT	GTGTACAGTCCGGGAGCATT
stSG495564	TGACCATGTCCCGGACTGTACACAAACA	AAGCAGTTGTGGTCCAGGAG
stSG495565	TGACCATGGGCAATGGTTTTCTGCAAAT	CCTTCTGAAACTGGGGATCA
stSG495566	TGACCATGCTGGACTTCCACAGGGCTT	CCTAGGACACTCTCCGGTTG
stSG495567	TGACCATGCCACAAATGGAAGGTATGGC	CCTCCCTAGAAGGCAGTGTG
stSG495568	TGACCATGGCCACAATGGCTGGACTTAT	TGGGAGAGAAACATGCACAG
stSG495569	TGACCATGATGCCGCATTTAGCAACTCT	TTCCTCAGACTGCCTCCTGT
stSG495570	TGACCATGCAGGAGGCAGTCTGAGGAAG	TAGGTCAAGGGTTGTGGGAG
stSG495571	TGACCATGCTCCCACAACCCTTGACCTA	AGCCTACCTTCCCCTTGAGA
stSG495572	TGACCATGTTGGGAGAGCTTGGCTTAAA	AGTCCTGGGGCTGGTGTATT
stSG495573	TGACCATGTCCTCTGTTCCCCATCTCAC	TTACCGGCTTTCTCTGCAAT
stSG495574	TGACCATGTGCACAAATGGCTTGATTGT	CCTTCCTTCCCCTGTGAGTT
stSG495575	TGACCATGGGCCCAGTTCACTCATGTTT	TGGTGGTTTTATTTCCTGCC
stSG495576	TGACCATGCTCTCAGGGCCCTTTCCTT	TGAAACACTAGCAAGCGTGG
stSG495577	TGACCATGCCACGCTTGCTAGTGTTTCA	GTTTGAAAACCACCCGCTTA
stSG495578	TGACCATGGTCAAAAGAGCAAAGCCAGG	CTACCGTGCCCAGAGTCATT
stSG495579	TGACCATGCCTCCACTCACCAAGAGAGC	TGCTTCATTTTATTTCCGGC
stSG495580	TGACCATGCATTCTGAGCAGCTTGCTTG	CTGTGATCAAGGCAGAATGAA
stSG495581	TGACCATGCAATCAGGTGGCAAGACAAA	GTGCCAAGCTGTTTGGAGTT
stSG495582	TGACCATGTGGCACCAAATCCATCAGTA	CCTGTTGTTCCCATCACCTT
stSG495583	TGACCATGGAGCTCAAAGGTGTCCTTGC	TGTAAGCTCTGTGGACGCAC
stSG495584	TGACCATGAGTCAGGCGCTAGAGGAAGC	CACTGAATTTGGCCTTACCC
stSG495585	TGACCATGAGGCACTAAACTGGCTCCCT	GCCATCCTGCAAGAGAAGTC
stSG495586	TGACCATGCTCATGGTAATGCCTGGTCC	CAGACGGTCCTGAGCTCTTC
stSG495587	TGACCATGGCTCAGGACCGTCTGACTTC	GGAAGTGAAACCAGCCACAT
stSG495588	TGACCATGAGGAGCTTTTGGTGATTGGA	TACAAGGCAAGGAGCCAACT
stSG495589	TGACCATGTTGGCTCCTTGCCTTGTACT	AGAGTATGGGCTTTGGGCTT
stSG495590	TGACCATGAAGCCCAAAGCCCATACTCT	AATTTGCTTCCTGCCTTTGA
stSG495591	TGACCATGTGTTGCATTTGTGGAGAGGA	ACCTACCTGCCACTCCCTTT
stSG495592	TGACCATGGGAGTGGCAGGTAGGTGAGA	CAGACACCCCTGTCTGTTCC
stSG495593	TGACCATGGTCTGCAGAGGTTTCCCAAC	GAGGCTGCAGTCACAAATGA
stSG495594	TGACCATGGCCCTGAGAGCCTGAATCTA	ACCTCAGCGTTTCCATCGTA
stSG495595	TGACCATGTACGATGGAAACGCTGAGGT	CCTGACCAGCCCAATTAAGA
stSG495596	TGACCATGTTTCCTCCCAACCACTTGTC	TGCTGGCCTATCCCAAATTA
stSG495597	TGACCATGATTTGGGATAGGCCAGCAAT	CCCCAACAGGACATAAAAAGG

stSG495598	TGACCATGGCCTGAAGGGAATGGAGTTT	CTAAGCTCACCATCCCCAAA
stSG495599	TGACCATGGATGCACATGGTTTGACTGG	AGGGCTGCTGACACCTAGAA
stSG495600	TGACCATGGCTTACAACATGGCTGTGGA	CTATGGAACAAGCAGCACCC
stSG495601	TGACCATGAGTGGAAGGGCTGTTTCTCA	AAGACAGGAGTATGCCAGGAA
stSG495602	TGACCATGTTCCTGGCATACTCCTGTCTT	CTAAGGGAGGTGACGCAGAG
stSG495603	TGACCATGGTCACCTCCCTTAGGAAGCC	AGGACAGACCAGGCAAGAGA
stSG495604	TGACCATGCTGTGCATCACAAAGCCATT	TCACCATAGACACCAGGGTATG
stSG495605	TGACCATGCATACCCTGGTGTCTATGGTGA	CAGGGGCTTCAGCTGTCTAA
stSG495606	TGACCATGTTAGACAGCTGAAGCCCCTG	TGTCTCAACCTTTGGTGTGC
stSG495607	TGACCATGGGTCATGTGCAAGTCTCCAG	CCTGGTCAGAGCCTCATTTC
stSG495608	TGACCATGCCAGAGGAAATGAGGCTCTG	TCTGTTCAGCAATCACCTGC
stSG495609	TGACCATGGCAGGTGATTGCTGAACAGA	TGCTGTTTCCCACAAGTCAA
stSG495610	TGACCATGAGTTCAGGTTGCTTGGATGG	CACACTGGGGAGGTGAGATT
stSG495611	TGACCATGTGTATACACCCCTCCTCCCA	AAAATCTCCTGGACTGGCCT

*TCACCATG – Amino linking adaptor added to the 5’ end of all forward primers

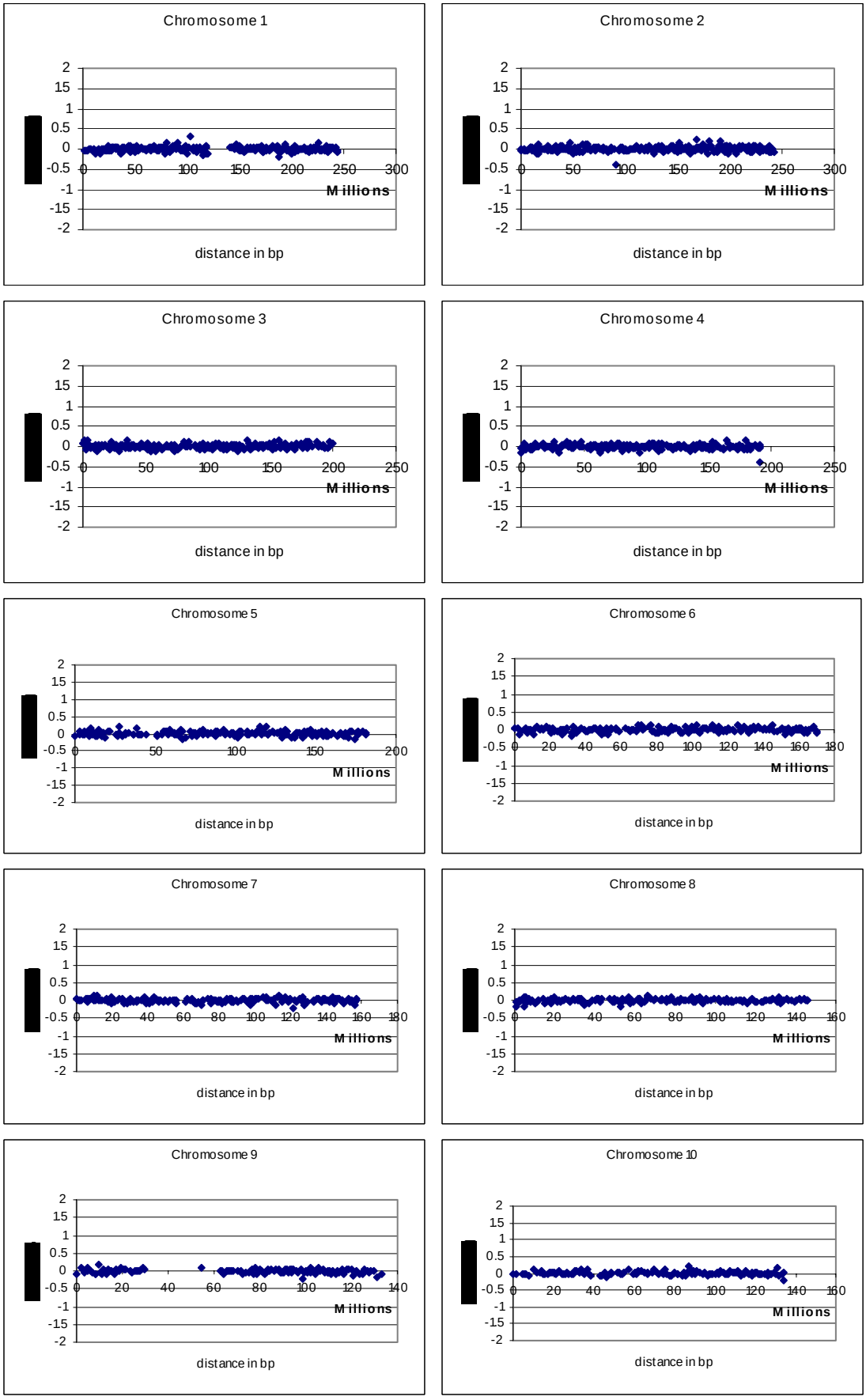
2b: The 96 well format of primers STSG 495474-495569.

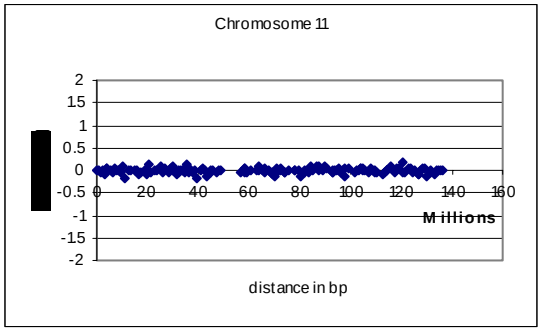
	1	2	3	4	5	6	7	8	9	10	11	12
A	stSG495474	stSG495475	stSG495476	stSG495477	stSG495478	stSG495479	stSG495480	stSG495481	stSG495482	stSG495483	stSG495484	stSG495485
B	stSG495486	stSG495487	stSG495488	stSG495489	stSG495490	stSG495491	stSG495492	stSG495493	stSG495494	stSG495495	stSG495496	stSG495497
C	stSG495498	stSG495499	stSG495500	stSG495501	stSG495502	stSG495503	stSG495504	stSG495505	stSG495506	stSG495507	stSG495508	stSG495509
D	stSG495510	stSG495511	stSG495512	stSG495513	stSG495514	stSG495515	stSG495516	stSG495517	stSG495518	stSG495519	stSG495520	stSG495521
E	stSG495522	stSG495523	stSG495524	stSG495525	stSG495526	stSG495527	stSG495528	stSG495529	stSG495530	stSG495531	stSG495532	stSG495533
F	stSG495534	stSG495535	stSG495536	stSG495537	stSG495538	stSG495539	stSG495540	stSG495541	stSG495542	stSG495543	stSG495544	stSG495545
G	stSG495546	stSG495547	stSG495548	stSG495549	stSG495550	stSG495551	stSG495552	stSG495553	stSG495554	stSG495555	stSG495556	stSG495557
H	stSG495558	stSG495559	stSG495560	stSG495561	stSG495562	stSG495563	stSG495564	stSG495565	stSG495566	stSG495567	stSG495568	stSG495569

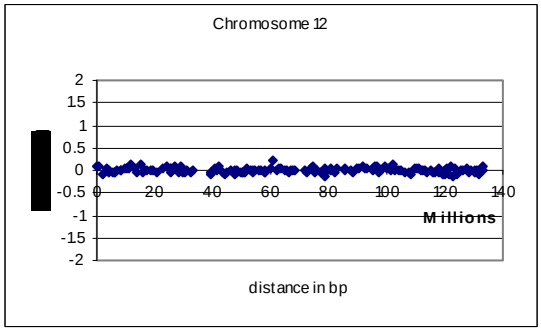
Appendix 3: Primers for quantitative PCR

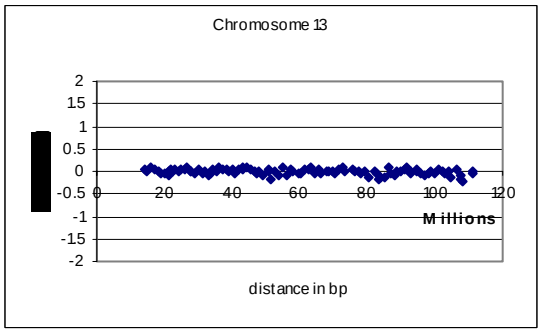
Clone	Forward Primer	Reverse Primer
cE140F8-1	TGTTCTATGAGTATGCGACTTTCCA	TTCAAACTGTGGGATGGTGAGA
cE140F8-2	TCTGCATCTTAAAGTGAGAGCTATGTTAC	TGAAGCTCTGATCTCCAGAAAGAG
cE140F8-3	GCTGATTTCCTCGTTCCCTCTATT	GTGTTAGGCAGTGGAATCATGTTC
cE140F8-4	GGATTCTGTCTTGTCTGGCCTTT	CTCCCGCGGTGCCTTT
cE140F8-5	GGCACCGCGGGAGAAG	GGCTGCATTGTTACAAATCTTTTTT
cN69F4-1	GGTTGAGGTCTGAAGCCCTTT	GGTCACTGCCCAGGCTCTT
cN69F4-2	CCTTGTCATCCCAAATACACCAT	AGACAGCTCCTGGGTCTTCCA
cN69F4-3	CAGAAACTGGCTTTGGAGAGATC	GAGACGTGGCTGAGCACAGA
cN69F4-4	GCACAAAATGTTCGAGACTGATACA	TTTACAACAAAGGCCAAATGCA
bK57G9-1	GGTGAGCCACATTTGTTATATTTGAA	GACTCACCCTTCCCCCTCTAAG
bK57G9-3	CTGTGCTGTGAATAGATCCATGTG	TGGCCGGGTGAACTCTTC
bK57G9-4	ACAATGGGTGCCAAGTTGGTA	CCCACAACCTGCTGCAGACT
bK57G9-5	TGGGCAGAGTCCCTGATTCT	AACTGGAAGGTGAACCCCAAA
bK57G9-6	GACTTCCAGGCCCTATGTCAGA	AAGTGGGAAGTTGCTGCTATGC
bK57G9-7	GATGCATGGGTGGGTGATG	TCCTGAGCCTCATTTGTTCTCA
bK57G9-8	CGGGCTTTGTCACAGCATCT	CAAAACTGGGAACAGCCTAAACA
cB13C9-1	TCAACAAGATATGTGCAAGCTTCTC	AAACTCCACCGGGCTCAAT
cB13C9-2	TTGCTGAGATTATGAATGGGTTTC	CTAGAGCTATTTTCTGTTTCCGACATACT
cB13C9-3	GCTGCACAAGCCATCCATTT	GGCCAGTGTGATTGATAAACTGAGT
cB13C9-4	GGGAGAATCCCAGCAAGTCA	CACCTCCCTGGTTGGTCATC
cB13C9-5	CTGCACCCCTCTTGTCTGTAACT	CGTCCTGAAACTTGGCATCTG

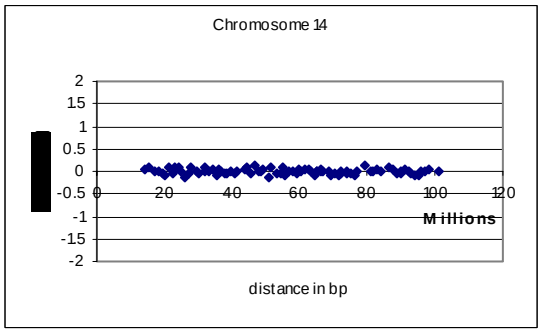
Appendix 4: Male:male hybridisation on 1Mb array

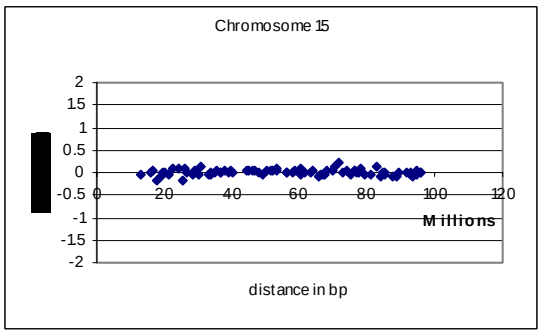


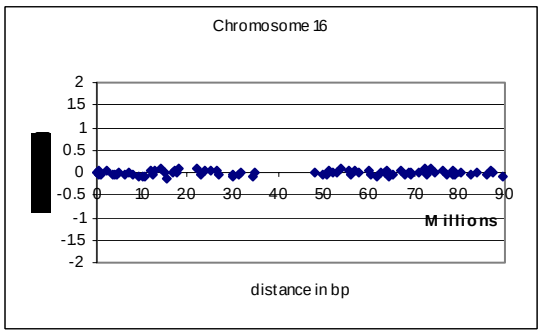

Chromosome 11
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
Millions
distance in bp

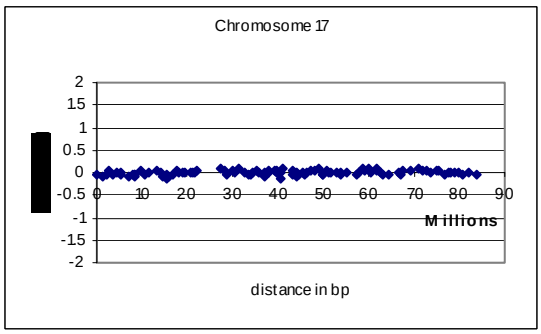

Chromosome 12
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
Millions
distance in bp

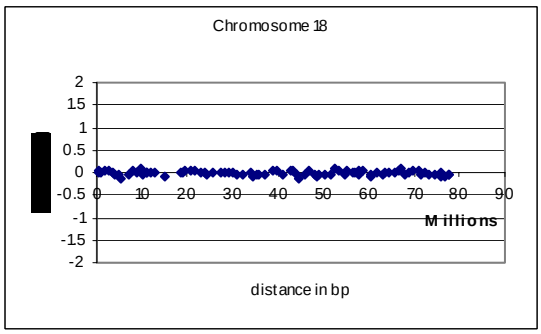

Chromosome 13
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

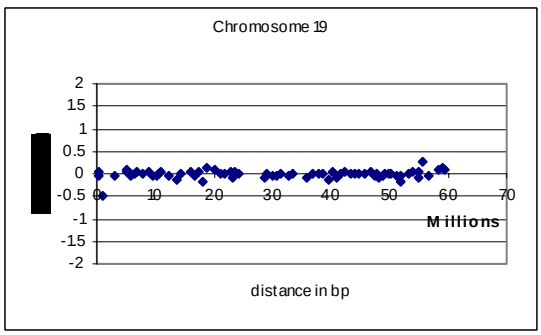

Chromosome 14
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

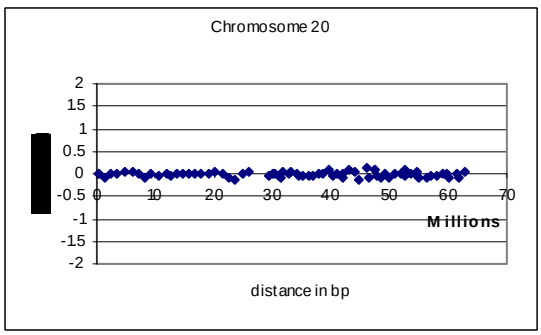

Chromosome 15
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

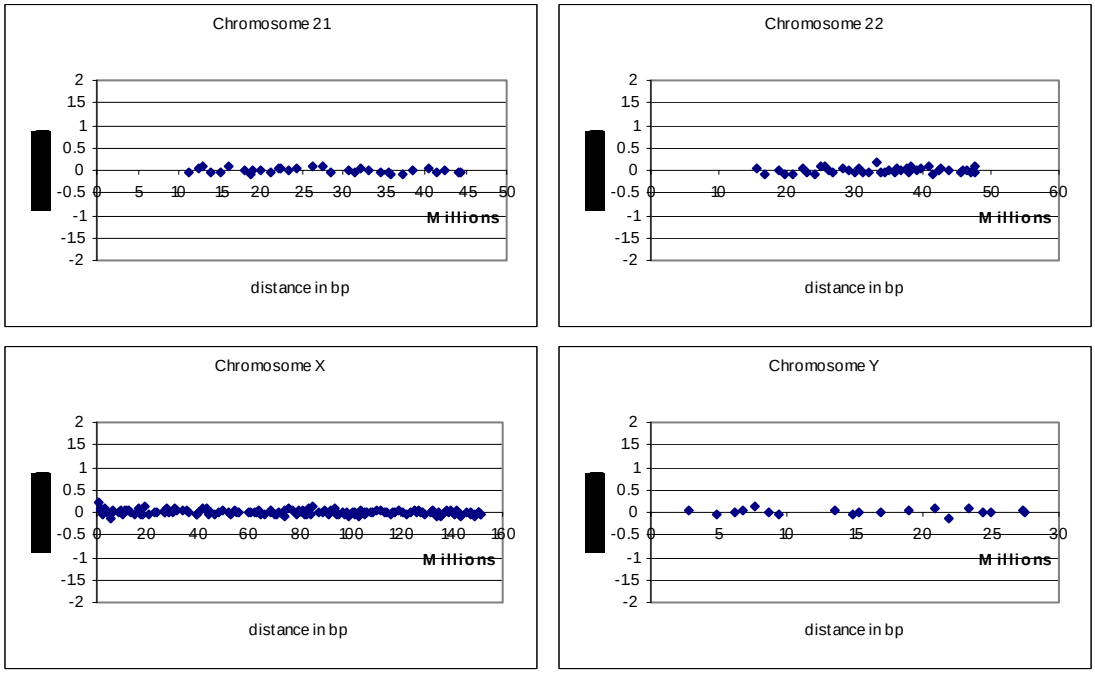

Chromosome 16
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp


Chromosome 17
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp


Chromosome 18
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp

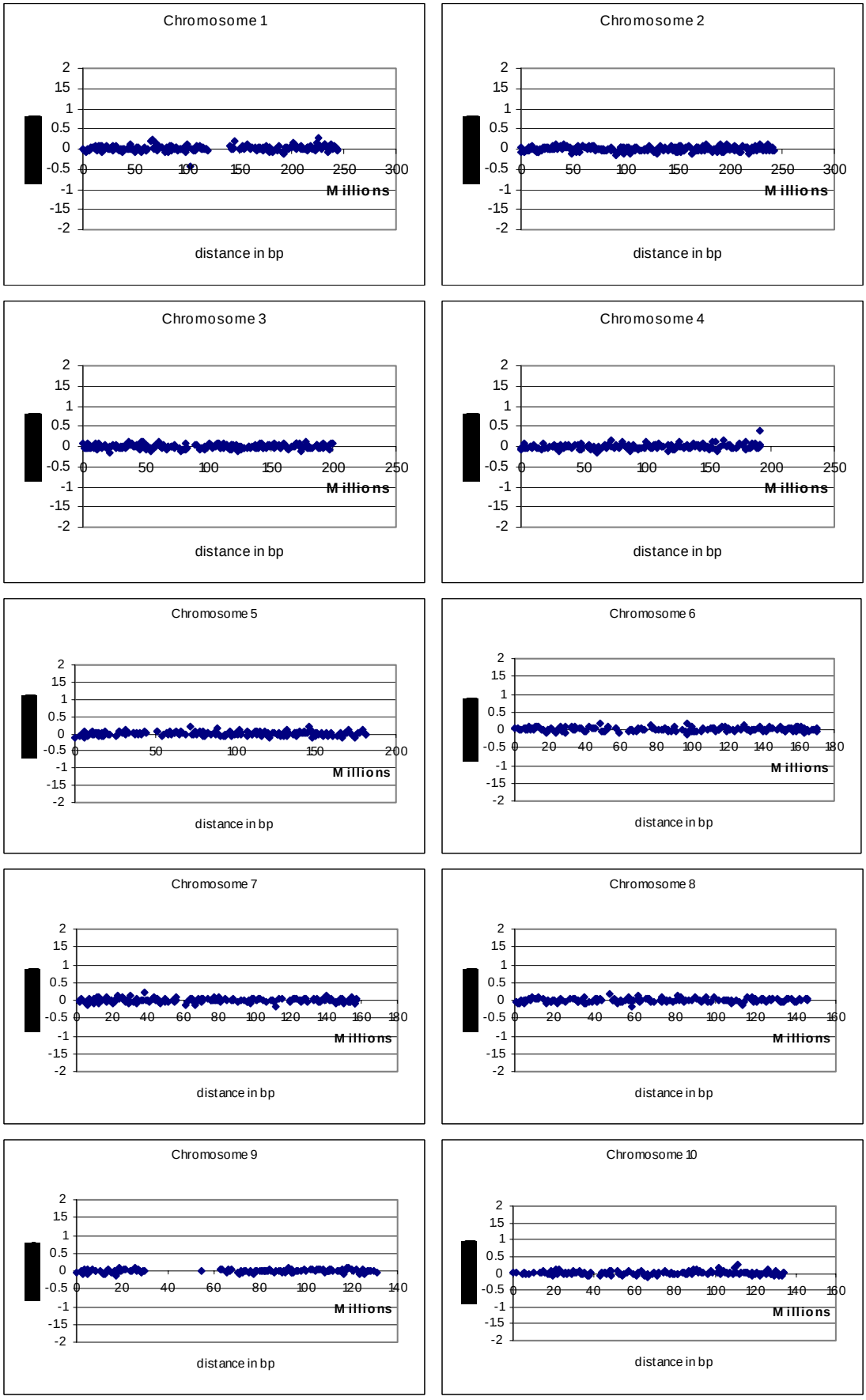

Chromosome 19
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
Millions
distance in bp

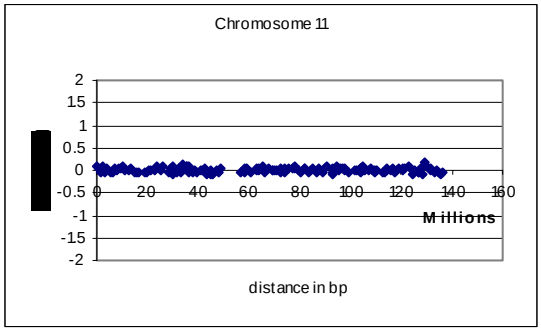

Chromosome 20
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
Millions
distance in bp

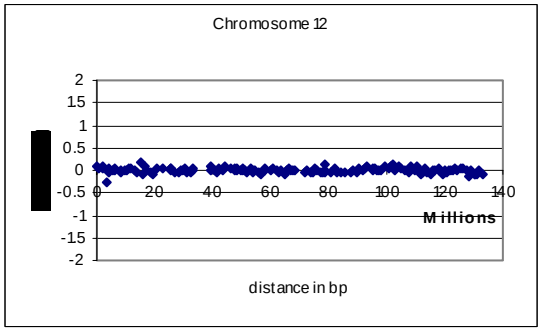


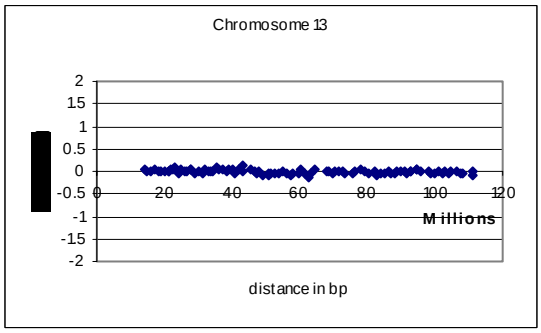
Log^2 ratios are given; therefore a 1:1 ratio will report a log^2 ratio of 0.

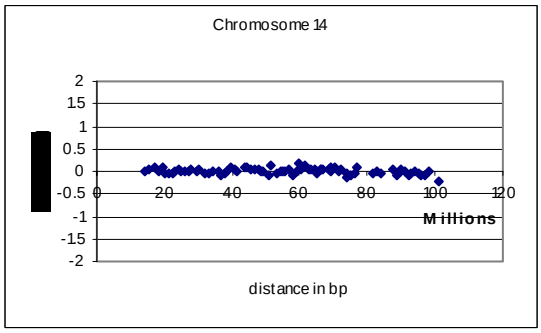
Appendix 5: Male:female hybridisation on 1Mb array

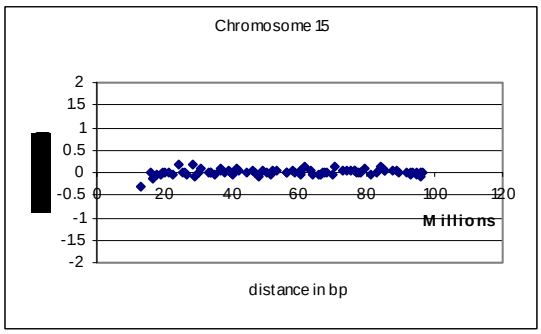


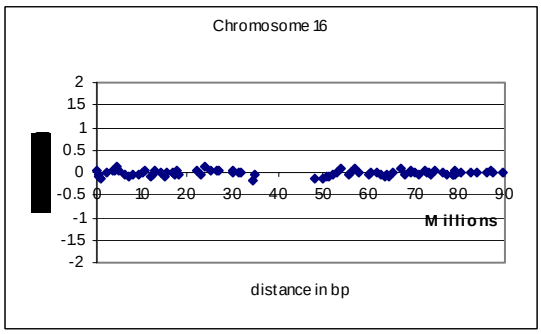

Chromosome 11
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
Millions
distance in bp

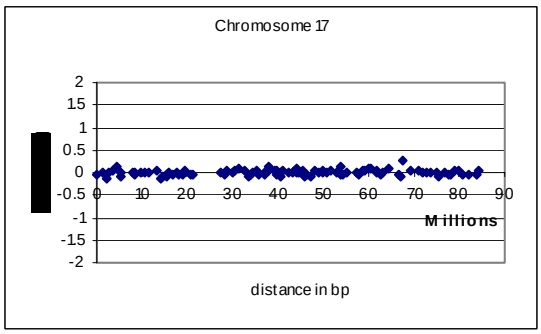

Chromosome 12
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
Millions
distance in bp

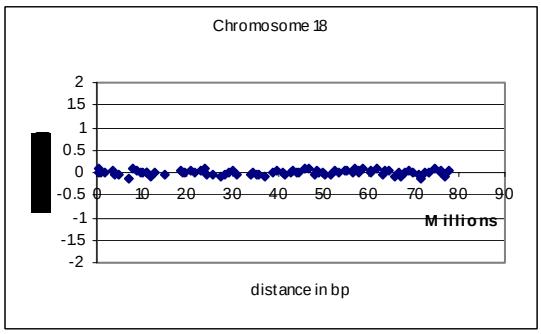

Chromosome 13
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

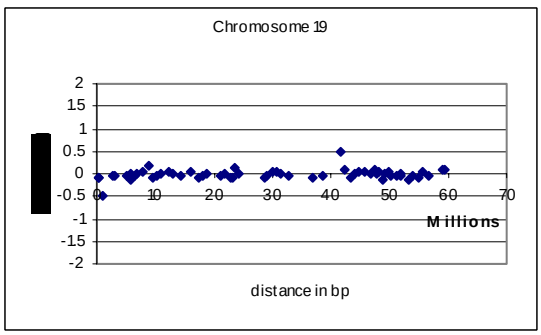

Chromosome 14
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

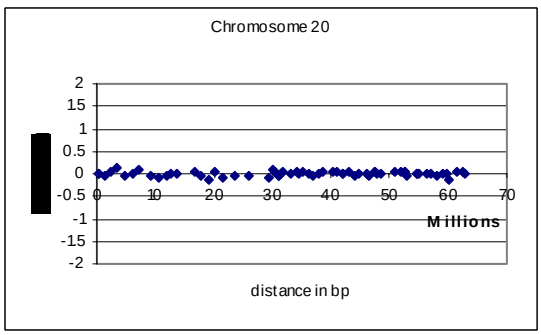

Chromosome 15
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

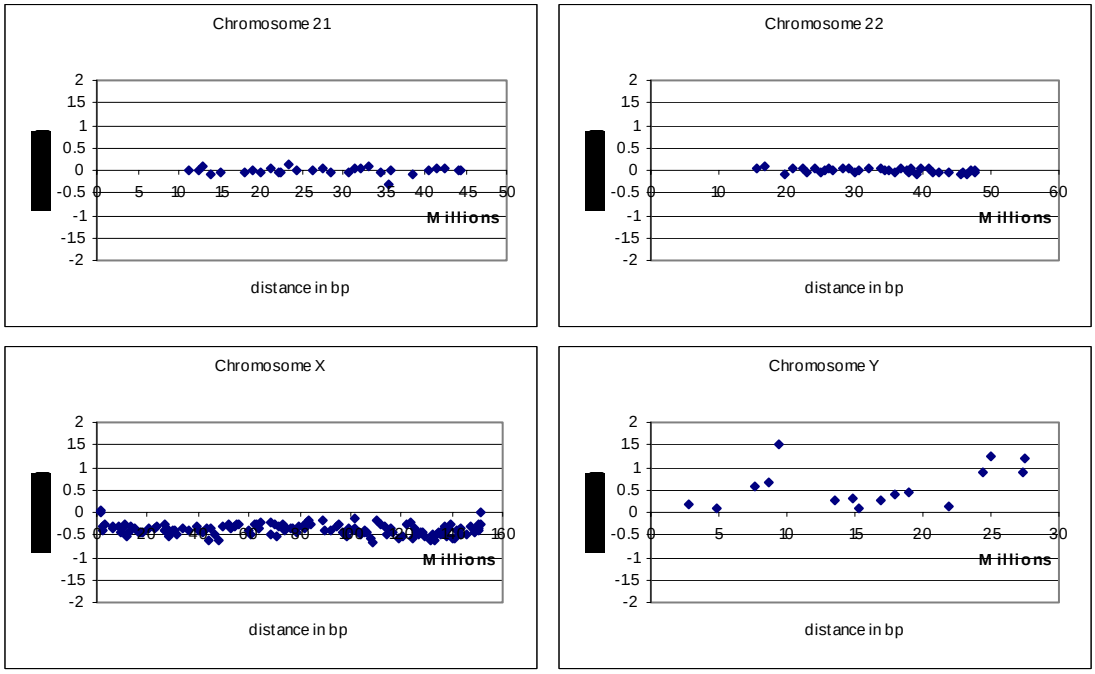

Chromosome 16
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp


Chromosome 17
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp


Chromosome 18
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp

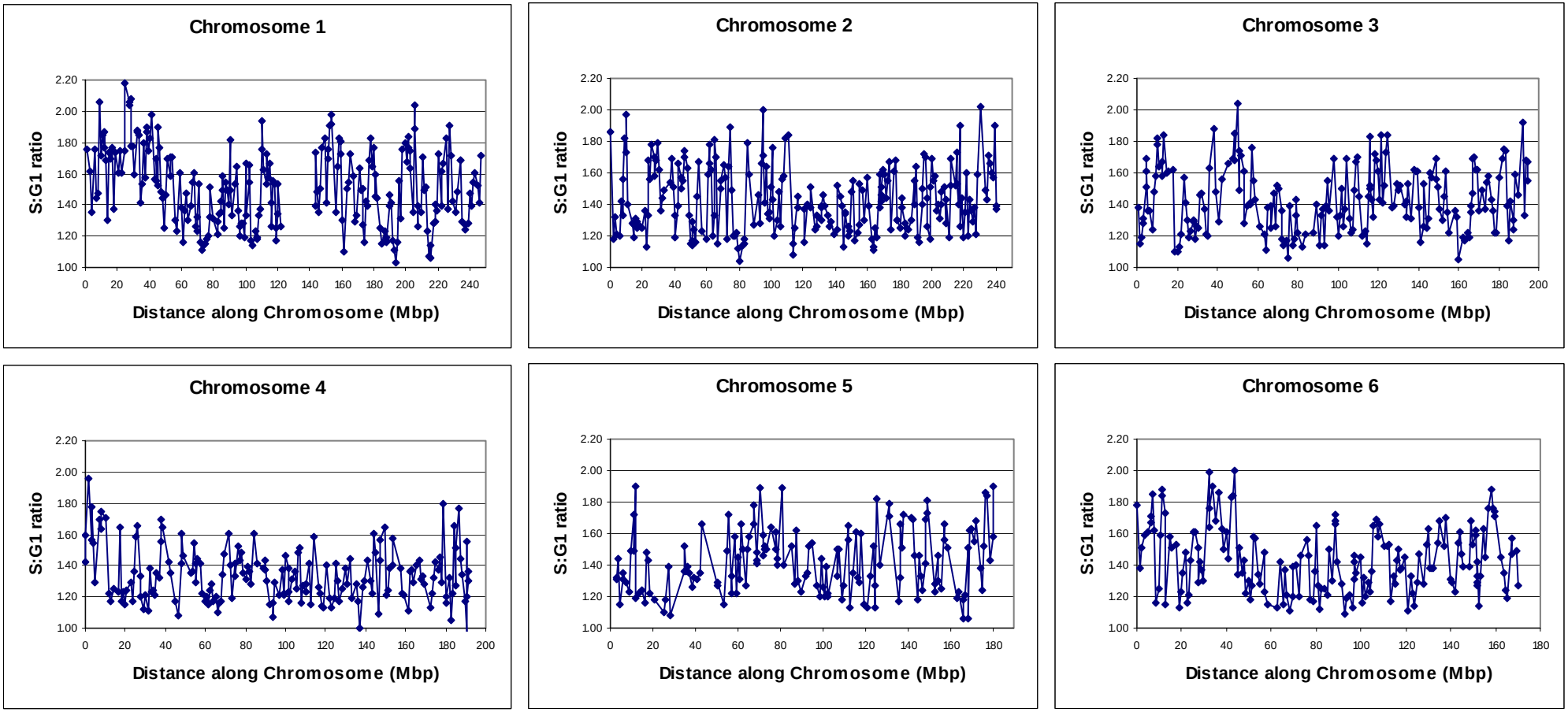

Chromosome 19
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
Millions
distance in bp

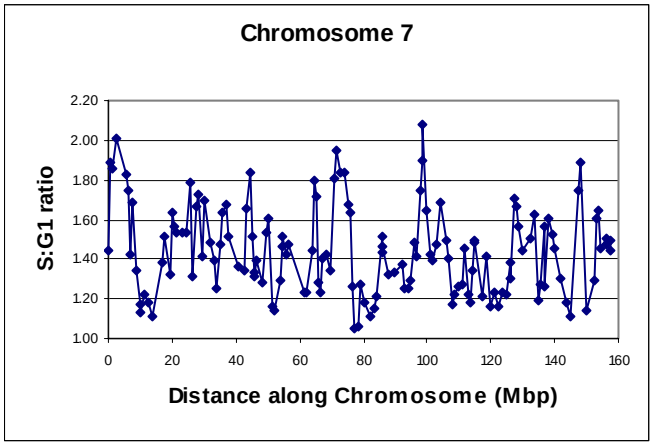

Chromosome 20
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
Millions
distance in bp

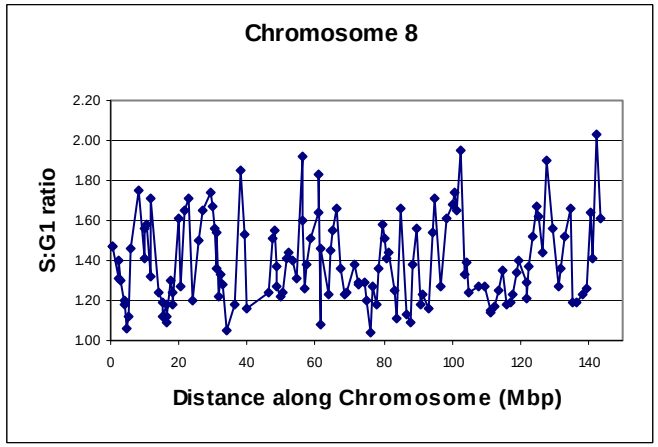


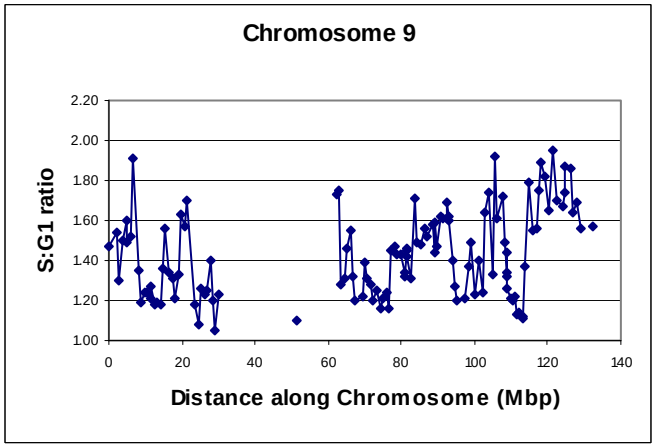
Log^2 ratios are given; therefore a 1:1 ratio will report a log^2 ratio of 0, this is seen on the autosomes. A ratio of 0.5:1 representing a single copy loss on the X clones will report a ratio of 0.5. Clones on the Y chromosome report a variety of ratios. This is because there is only Cy 3 labelled Y chromosome DNA within the hybridisation mix, so there is no Cy 5 DNA to hybridise against.

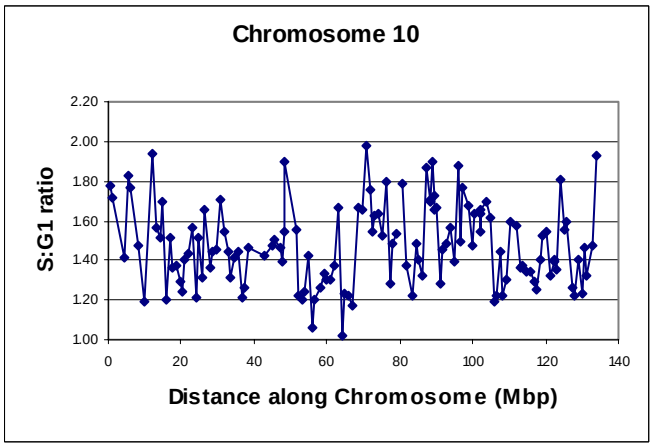
Appendix 6: Replication timing profiles for all 24 chromosomes.

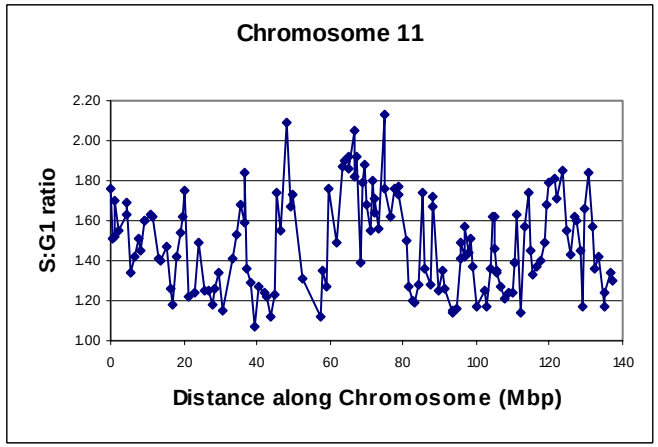


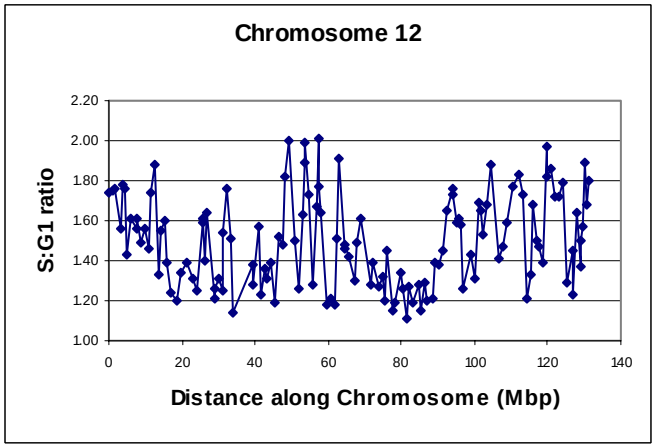

Chromosome 7
S:G1 ratio
2.20
2.00
1.80
1.60
1.40
1.20
1.00
0
20
40
60
80
100
120
140
160
Distance along Chromosome (Mbp)

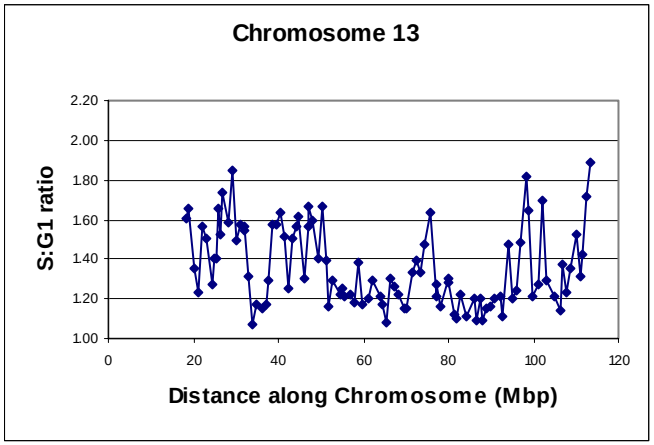

Chromosome 8
S:G1 ratio
2.20
2.00
1.80
1.60
1.40
1.20
1.00
0
20
40
60
80
100
120
140
Distance along Chromosome (Mbp)

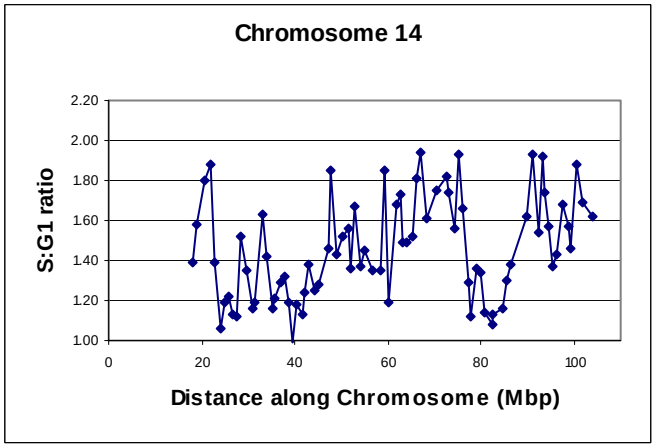

Chromosome 9
S:G1 ratio
2.20
2.00
1.80
1.60
1.40
1.20
1.00
0
20
40
60
80
100
120
140
Distance along Chromosome (Mbp)

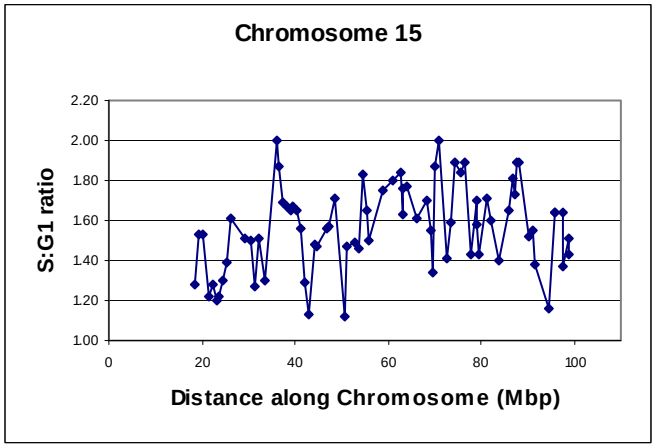

Chromosome 10
S:G1 ratio
2.20
2.00
1.80
1.60
1.40
1.20
1.00
0
20
40
60
80
100
120
140
Distance along Chromosome (Mbp)

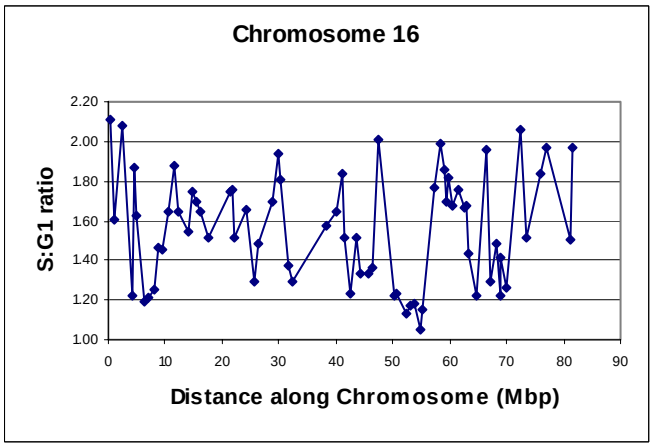

Chromosome 11
S:G1 ratio
2.20
2.00
1.80
1.60
1.40
1.20
1.00
0
20
40
60
80
100
120
140
Distance along Chromosome (Mbp)

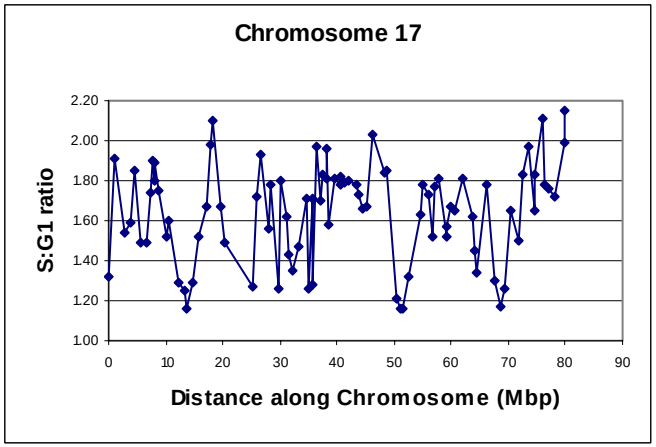

Chromosome 12
S:G1 ratio
2.20
2.00
1.80
1.60
1.40
1.20
1.00
0
20
40
60
80
100
120
140
Distance along Chromosome (Mbp)

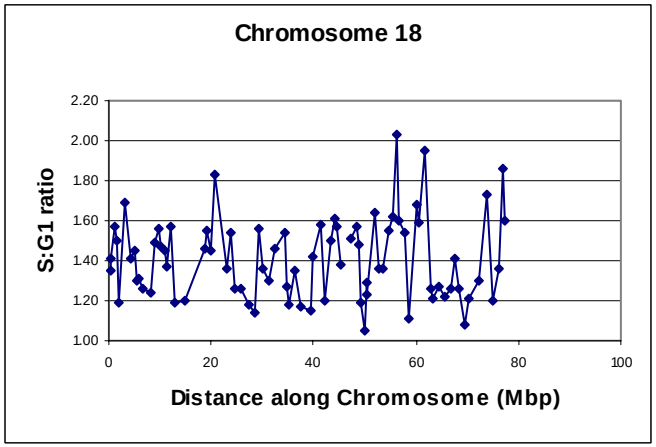

Chromosome 13
S:G1 ratio
2.20
2.00
1.80
1.60
1.40
1.20
1.00
0
20
40
60
80
100
120
Distance along Chromosome (Mbp)

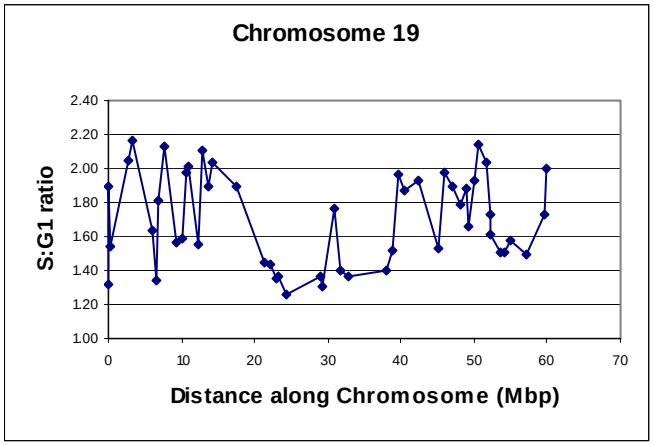

Chromosome 14
S:G1 ratio
2.20
2.00
1.80
1.60
1.40
1.20
1.00
0
20
40
60
80
100
Distance along Chromosome (Mbp)

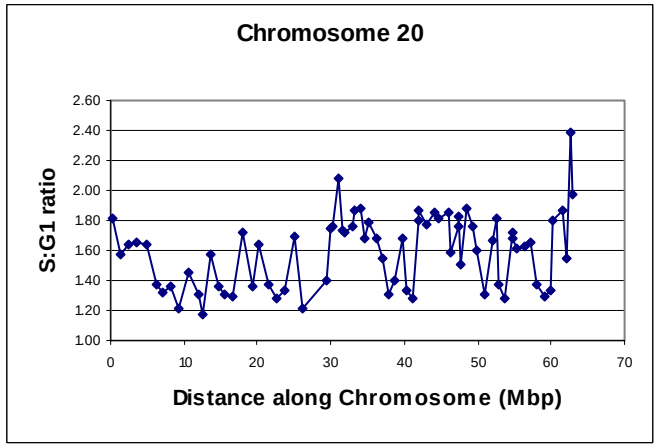

Chromosome 15
S:G1 ratio
2.20
2.00
1.80
1.60
1.40
1.20
1.00
0
20
40
60
80
100
Distance along Chromosome (Mbp)

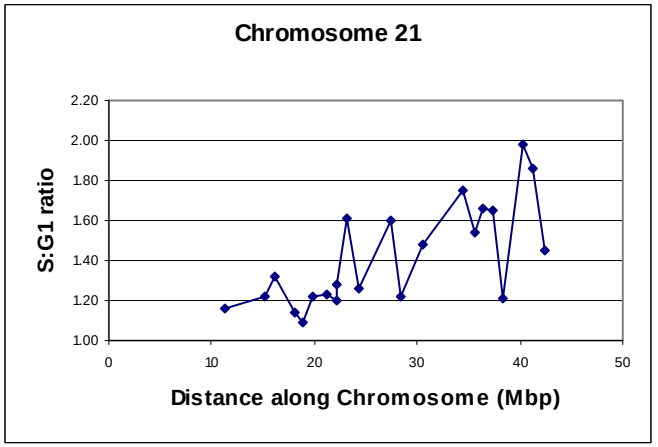

Chromosome 16
S:G1 ratio
2.20
2.00
1.80
1.60
1.40
1.20
1.00
0
10
20
30
40
50
60
70
80
90
Distance along Chromosome (Mbp)

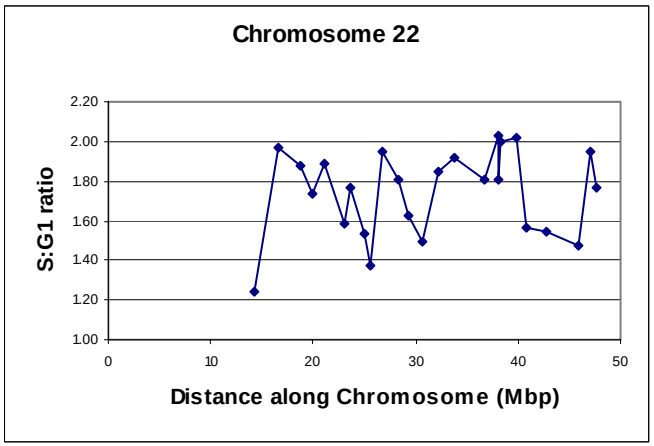

Chromosome 17
S:G1 ratio
2.20
2.00
1.80
1.60
1.40
1.20
1.00
0
10
20
30
40
50
60
70
80
90
Distance along Chromosome (Mbp)

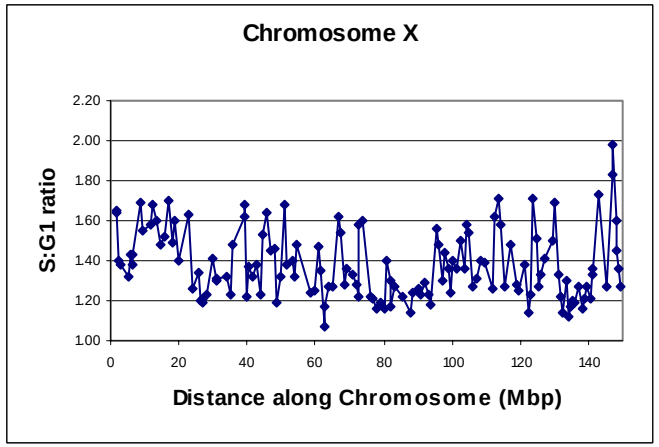

Chromosome 18
S:G1 ratio
2.20
2.00
1.80
1.60
1.40
1.20
1.00
0
20
40
60
80
100
Distance along Chromosome (Mbp)

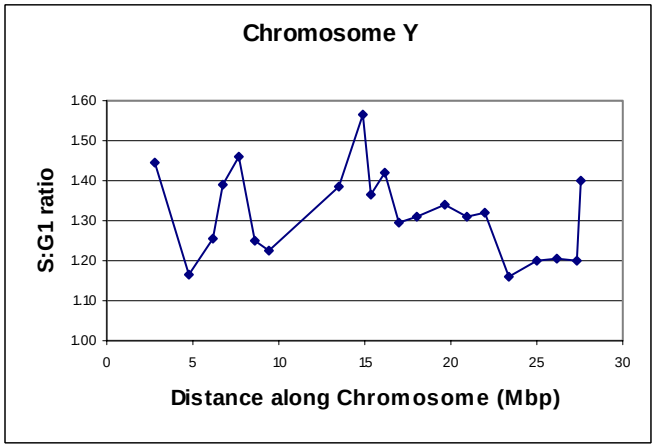

Chromosome 19
S:G1 ratio
2.40
2.20
2.00
1.80
1.60
1.40
1.20
1.00
0
10
20
30
40
50
60
70
Distance along Chromosome (Mbp)


Chromosome 20
S:G1 ratio
2.60
2.40
2.20
2.00
1.80
1.60
1.40
1.20
1.00
0
10
20
30
40
50
60
70
Distance along Chromosome (Mbp)


Chromosome 21
S:G1 ratio
2.20
2.00
1.80
1.60
1.40
1.20
1.00
0
10
20
30
40
50
Distance along Chromosome (Mbp)


Chromosome 22
S:G1 ratio
2.20
2.00
1.80
1.60
1.40
1.20
1.00
0
10
20
30
40
50
Distance along Chromosome (Mbp)

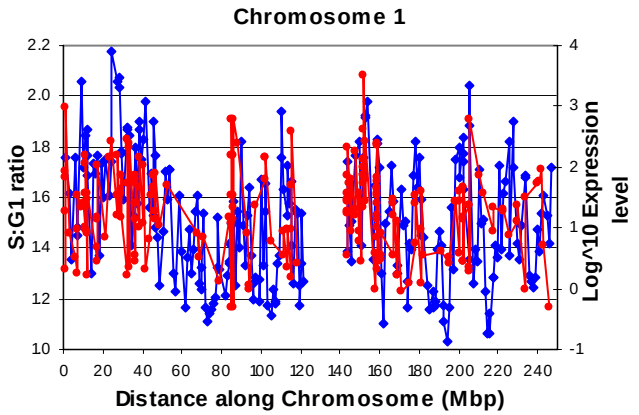

Chromosome X
S:G1 ratio
2.20
2.00
1.80
1.60
1.40
1.20
1.00
0
20
40
60
80
100
120
140
Distance along Chromosome (Mbp)

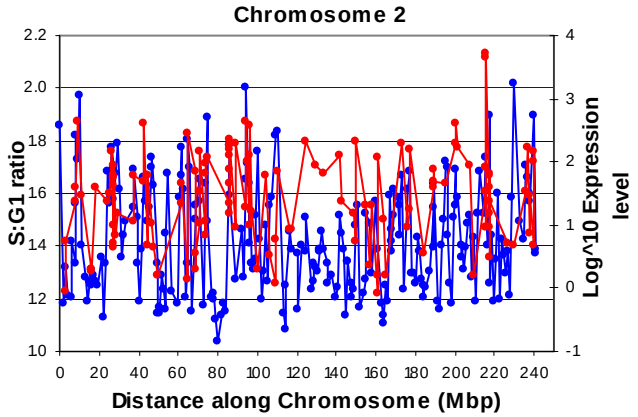

Chromosome Y
S:G1 ratio
1.60
1.50
1.40
1.30
1.20
1.10
1.00
0
5
10
15
20
25
30
Distance along Chromosome (Mbp)

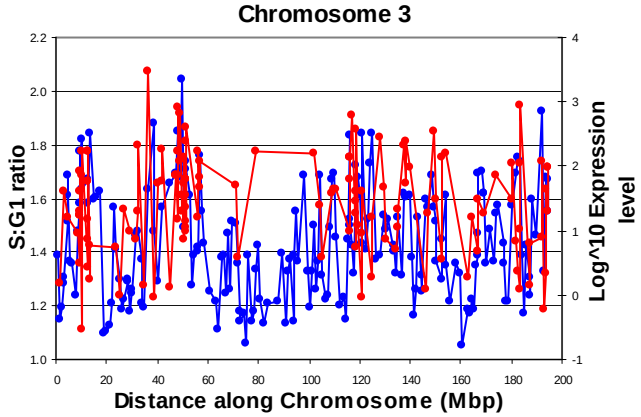
Appendix 7: Perl program to identify regions of co-ordinated replication

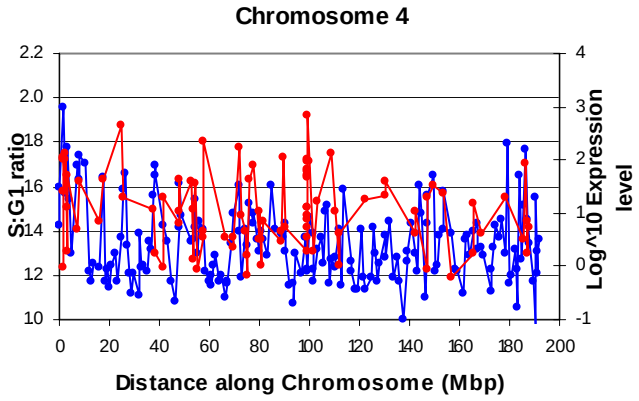
A purpose-written perl program was used to find the optimal segmentation of the replication timing (RT) data. Suppose a chromosome contains n RT signals arranged in genome order. Within each segment, starting at coordinate i and ending at coordinate j, we define the score S_{ij} equal to the sum of squared deviations of the RT values from the mean RT signal μ_{ij} for the segment. The optimal segmentation pattern (ie the number of segments and coordinates of segment boundaries) is chosen which minimises a function, W_n, based on the sum of segment scores plus a penalty score B for each segment transition. Let W_k be the score of the optimal segmentation for coordinates 1 through k. Then $W_0 = 0$ and $W_k = \min_{i<k} \{ W_{i-1}+B+S_{ik} \}$ for all $k>0$. The degree of segmentation is controlled by the value of B. The optimal segmentation is found by backtracking from the terminal value W_n. The statistical significance of W was determined by re-running the program on 1000 permuted data sets in which the order of observed RT signals was shuffled. The P-value for the test of the null hypothesis that the observed segmentation score could have arisen by chance is estimated as the proportion of times the permuted W score exceeded the observed score.

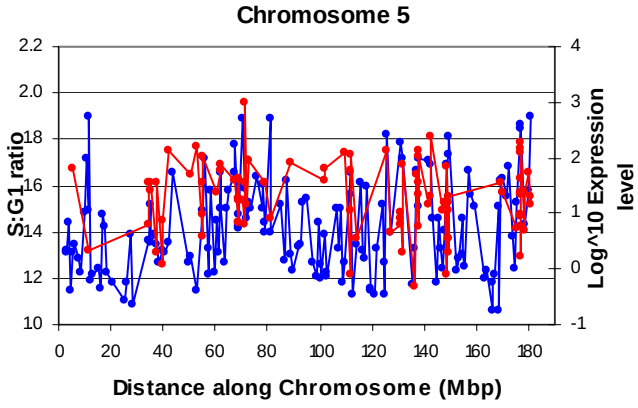
Appendix 8: Replication timing and Expression level profiles for all 24 chromosomes.

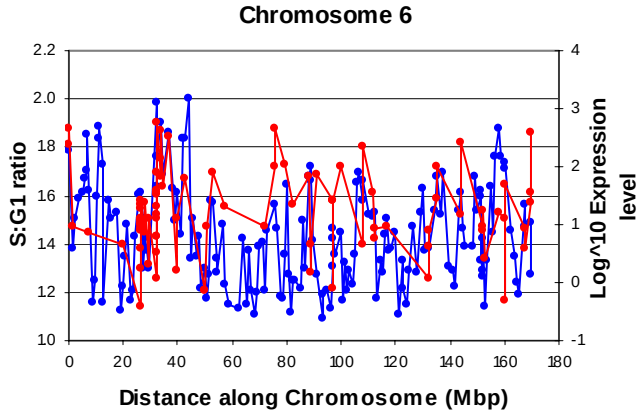


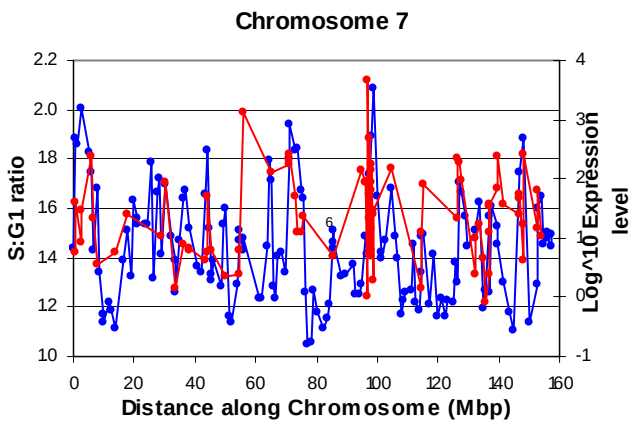


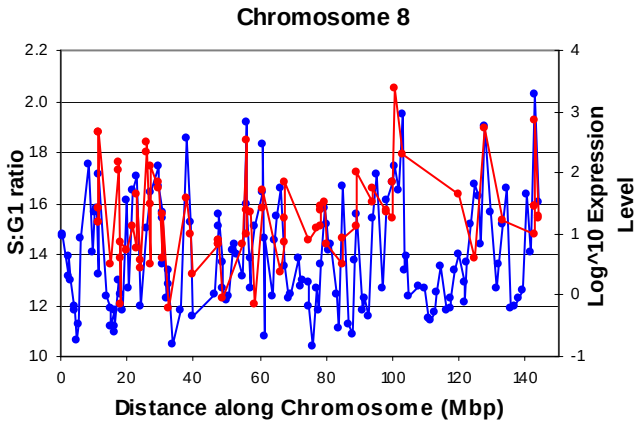


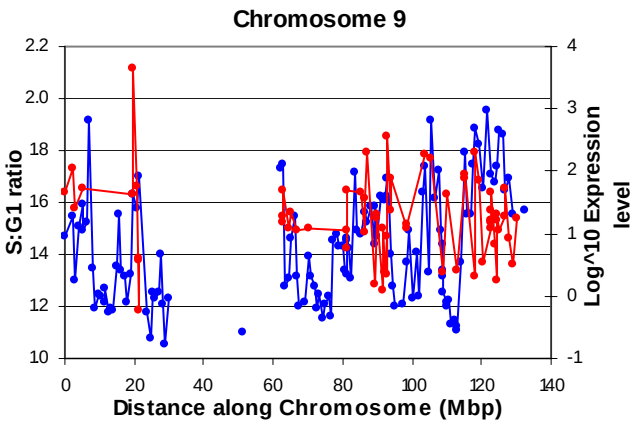


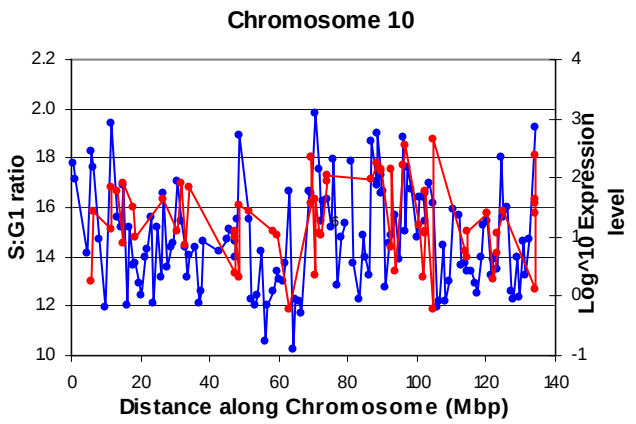


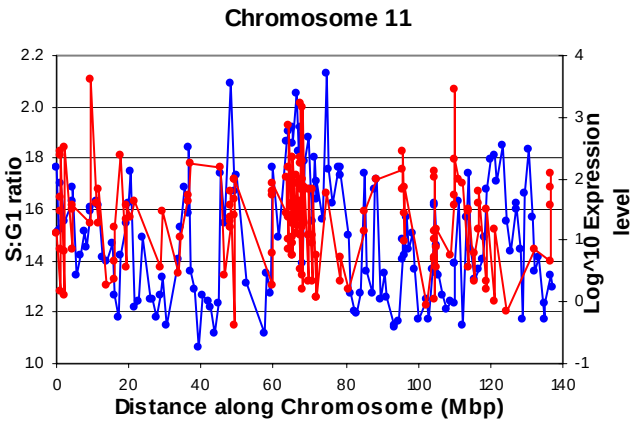


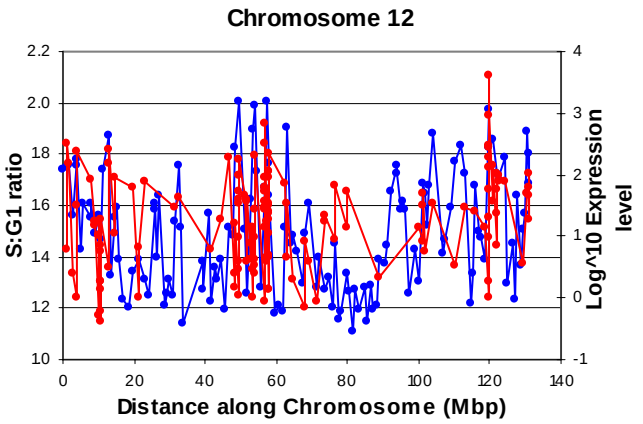

Chromosome 7
S:G1 ratio
Log^10 Expression level
Distance along Chromosome (Mbp)

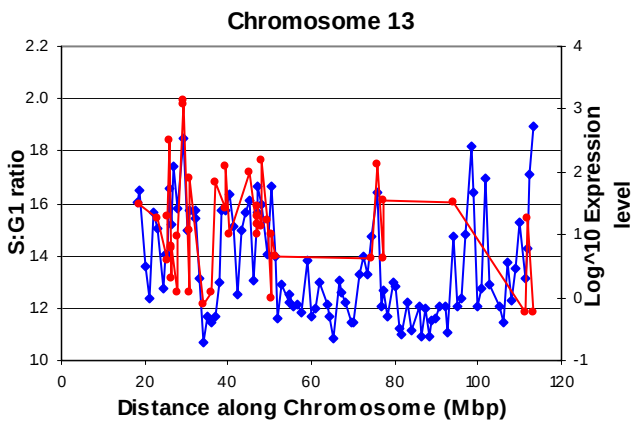

Chromosome 8
S:G1 ratio
Log^10 Expression Level
Distance along Chromosome (Mbp)

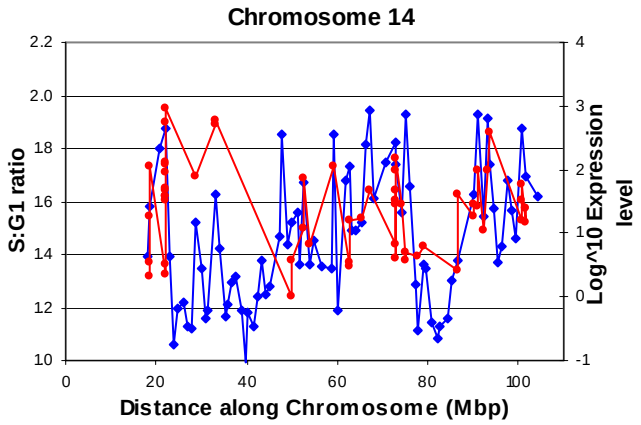

Chromosome 9
S:G1 ratio
Log^10 Expression level
Distance along Chromosome (Mbp)

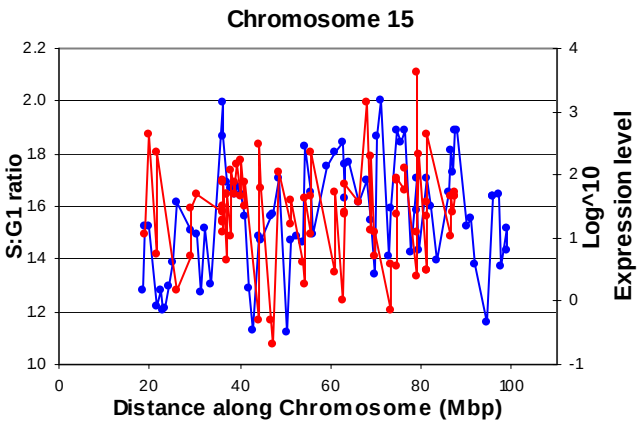

Chromosome 10
S:G1 ratio
Log^10 Expression level
Distance along Chromosome (Mbp)

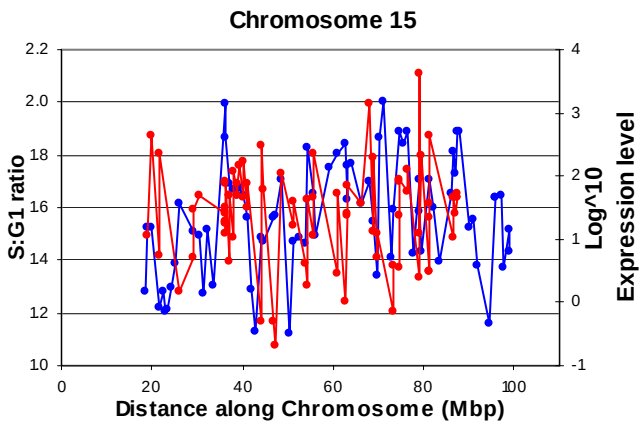

Chromosome 11
S:G1 ratio
Log^10 Expression level
Distance along Chromosome (Mbp)

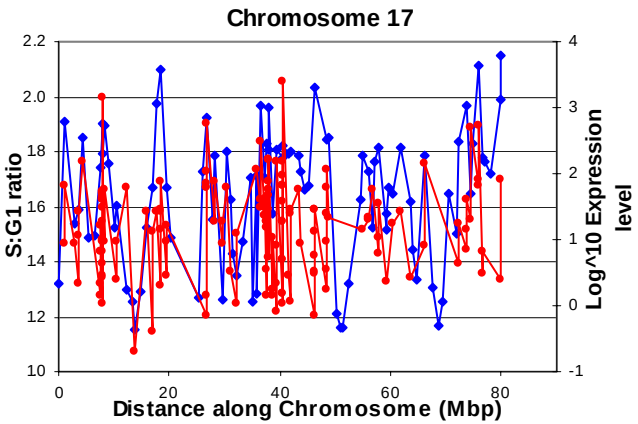

Chromosome 12
S:G1 ratio
Log^10 Expression level
Distance along Chromosome (Mbp)

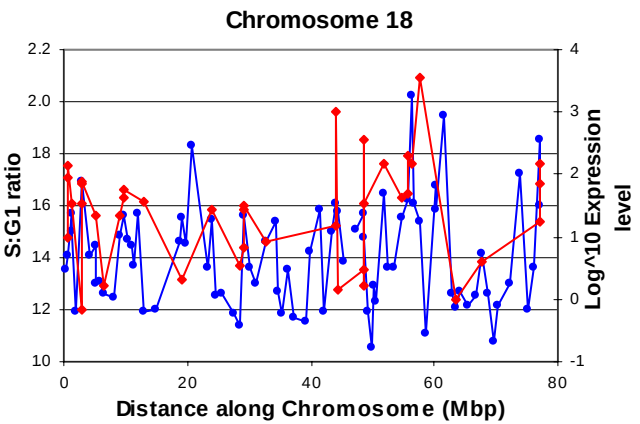

Chromosome 13
S:G1 ratio
Log^10 Expression level
Distance along Chromosome (Mbp)

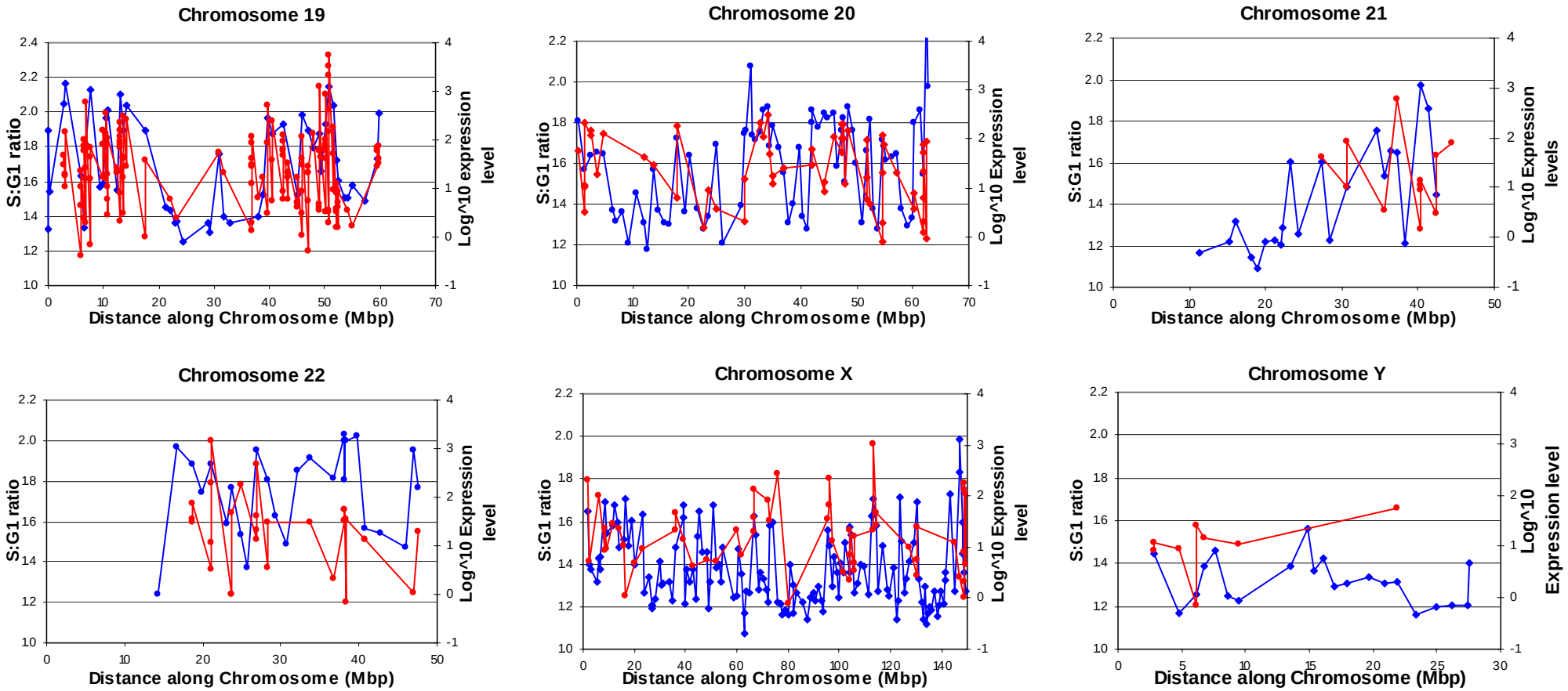

Chromosome 14
S:G1 ratio
Log^10 Expression level
Distance along Chromosome (Mbp)


Chromosome 15
S:G1 ratio
Log^10 Expression level
Distance along Chromosome (Mbp)


Chromosome 15
S:G1 ratio
Log^10 Expression level
Distance along Chromosome (Mbp)


Chromosome 17
S:G1 ratio
Log^10 Expression level
Distance along Chromosome (Mbp)


Chromosome 18
S:G1 ratio
Log^10 Expression level
Distance along Chromosome (Mbp)



Blue: Replication timing ratio. Red: Expression level of clones in the 1Mb set

Appendix 9: Chromosome 22 sequencing-clone information

9a: International Names for chromosome 22 clones

Accession No.	International Clone name	Sanger Clone name
AP000522	AP000522	cN4G1
AP000523	AP000523	c60H5
AP000524	AP000524	c70D1
AP000525	AP000525	cN14H11
AP000526	AP000526	cN64E9
AP000527	AP000527	cNN83F12
AP000528	AP000528	cN91G6
AP000529	AP000529	cN3G11
AP000530	AP000530	cN65E1
AP000531	AP000531	cN59E1
AP000532	AP000532	cN2F2
AP000533	AP000533	cN60G3
AP000534	AP000534	cN23H5
AP000535	AP000535	cN58F10
AP000536	AP000536	cN64C8
AP000537	AP000537	cN54B2
AP000538	AP000538	cN65B12
AP000539	AP000539	cN72E11
AP000540	AP000540	cN53D1
AP000541	AP000541	cN13E4
AP000542	AP000542	cN60D12
AP000543	AP000543	cN20H12
AP000544	AP000544	cN17H1
AP000545	AP000545	cN68B10
AP000546	AP000546	cN18E3
AP000547	KB-67B5	KB67B5
AP000365	KB-7G2	KB7G2
AC005301	AC005301	p15j16
AC007064	AC007064	p87o8
AC006548	AC006548	p20k14
AC006946	AC006946	p109l3
AC005300	AC005300	p143i13
AC005399	AC005399	p238m15
AC004019	AC004019	357f7
AC007666	AC007666	p273a17
AC006285	AC006285	p1087l10
AC016026	AC016026	b461k10
AC008101	XXbac-677f7	b677f7
AC008079	AC008079	bac519d21
AC008132	AC008132	pac995o6
AC008103	AC008103	pac699j1
AC007326	AC007326	p423
AC000095	AC000095	fF41C7
AC004461	AC004461	cN119F4

AC004462	AC004462	18c3
AC004471	AC004471	111f11
AC004463	AC004463	79h12
AC000081	AC000081	59c10
AC000094	AC000094	fF39E1
AC000085	AC000085	72f8
AC000092	AC000092	98c4
AC000079	AC000079	49c12
AC000068	AC000068	102g9
AC000087	AC000087	83c5
AC000088	AC000088	83e8
AC000082	AC000082	59f
AC000070	AC000070	105a
AC000086	AC000086	81h
AC000077	AC000077	31e
AC000067	AC000067	100h
AC000093	AC000093	carlaa
AC000091	AC000091	91c
AC000089	AC000089	89h
AC000076	AC000076	2h
AC000078	AC000078	33e
AC000090	AC000090	8c
AC000080	AC000080	56c
AC005663	AC005663	p888c9
AC006547	AC006547	p158l19
AC007663	AC007663	b444p24
AC007731	AC007731	b562f10
AC005500	AC005500	p52f6
AC004033	AC004033	p_M11
AC007050	AC007050	bac32
AC007308	AC007308	pac408
AC002470	AC002470	bK135H6
AC002472	AC002472	P_N5
AP000550	KB-1592A4	KB1592A4
AP000551	KB-876E2	KB876E2
AP000552	KB-1183D5	KB1183D5
AP000556	KB-1172D5	KB1172D5
AP000557	KB-1323B2	KB1323B2
AP000558	KB-1802C5	KB1802C5
AP000553	KB-1440D3	KB1440D3
AP000554	KB-666H9	KB666H9
AP000555	KB-1027C11	KB1027C11
D86995	D86995	cN109G12
D87019	D87019	cN86G7
D87012	D87012	cN61D6
D88268	D88268	cN47H9
D86993	D86993	cN23C6
D87004	D87004	cN4E7
D87022	D87022	cN88E1
D88271	D88271	cN114H4
D88269	D88269	cN33B6
D87000	D87000	cN30E12

D86996	D86996	cN23F1
D86989	D86989	cN110H3
D88270	D88270	cN123E1
D87003	D87003	cN2H8
D87018	D87018	cN80A10
D87016	D87016	cN75A1
D86999	D86999	cN22A12
D87010	D87010	cN35B9
D87009	D87009	KB288A10
D87011	D87011	cN50D10
D87013	D87013	cN63E9
D87014	D87014	cN61E11
D86991	D86991	cN29D3
D87002	D87002	cN31F3
D87006	D87006	cN52F2
D86994	D86994	cN102D1
D87007	D87007	cN48A11
D87015	D87015	cN68D6
D86998	D86998	cN24A12
D87021	D87021	cN84E4
D87024	D87024	cN92H4
D87020	D87020	cN9G6
D87023	D87023	cN9C5
D87017	D87017	cN75C12
AP000360	AP000360	cN81C12
AP000361	AP000361	cN8E4
AP000362	AP000362	cN75A12
AC000029	AC000029	bK865E9
AC000102	AC000102	bK60B5
AP000343	KB-282B12	kB282B12
AP000344	KB-1269D1	kB1269D1
AP000345	KB-208E9	KB208E9
AP000346	KB-1572G7	kB1572G7
AP000347	KB-113H7	KB113H7
AP000348	AP000348	cN27C7
AP000349	KB-1839H6	kB1839H6
AP000350	KB-1125A3	kB1125A3
AP000351	KB-226F1	KB226F1
AP000352	KB-1561E1	kB1561E1
AP000353	KB-318B8	kB318B8
AP000354	KB-1674E1	kB1674E1
AP000355	KB-1896H10	kB1896H10
AP000356	KB-1995A5	kB1995A5
AP000357	AP000357	cN95F10
AP000358	AP000358	cN110F4
AP000359	KB-63E7	KB63E7
AL049759	RP5-930L11	dJ930L11
AL050312	RP11-9F11	bA9F11
AL022323	CTA-243E7	bK243E7
Z99916	CTA-221G9	bK221G9
AL022332	RP3-462D8	dJ462D8
AL022324	CTA-246H3	bK246H3

AL008721	CTA-390C10	bK390C10
AL022329	CTA-407F11	bK407F11
AL080245	RP11-89B2	bA89B2
Z98949	CTA-125H2	bK125H2
AL079300	CTA-109D1	bK109D1
AL022337	CTA-796E4	bK796E4
AL080273	RP11-259P1	bA259P1
AL023513	RP1-268D13	dJ268D13
AL078460	RP3-341O5	dJ341O5
AL035044	RP1-40G4	dJ40G4p
Z99714	CTB-48E9	bK1048E9
Z95115	CTA-445C9	bK445C9
Z99774	CTA-373H7	bK373H7
Z95889	CTA-211A9	bK211A9
Z97353	RP1-90L6	dJ90L6
AL008638	CTA-992D9	bK992D9
AL021153	CTA-503F6	bK503F6
AL034386	RP5-1172A22	dJ1172A22
AL020994	CTA-929C8	bK929C8
AL049536	RP1-205F14	dJ205F14p
AL050402	RP11-46E17	bA46E17
AL133456	AL133456	dJ231P7p
AL390209	AL390209	bK437G10
AL121885	RP11-375H17	bA375H17
AL031591	RP3-353E16	dJ353E16
AL033538	RP3-477H23	dJ477H23
AL035453	SC22CB-42E1	cB42E1
AL050313	CTA-754D9	bK754D9
AL035397	RP6-45P1	dA45P1
AL023281	CTA-544A11	bK544A11
AL008722	CTA-732E4	bK732E4
AL080241	RP11-541J16	bA541J16
AL118497	RP11-329J7	bA329J7
AL121825	RP11-436C9	bA436C9
AL117330	RP11-444G7	bA444G7
AL023494	RP3-366L4	dJ366L4
Z93930	CTA-292E10	bK292E10
AL031596	RP4-745C22	dJ745C22
Z95113	CTA-175E3	bK175E3
AL021393	CTA-747E2	bK747E2
Z95116	CTA-57G9	bK57G9
AL031186	CTA-984G1	bK984G1
AC000026	AC000026	bK58B8
AC000041	AC000041	cE42H1
AC000035	AC000035	cN47G11
AC005529	AC005529	bK256D12
AC004882	RP1-76B20	dJ76B20
Z82171	SC22CB-11B7	cB11B7
AC004819	RP1-15I23	dJ15I23
AC003681	RP3-394A18	dJ394A18
AC003071	CTA-85E5	bK85E5
AC002378	RP3-438O4	dJ438O4

AC004264	RP1-102K2	dJ102K2
AC004997	RP1-130H16	dJ130H16
AC004832	RP4-539M6	dJ539M6
AC005006	RP1-56J10	dJ56J10
AC003072	CTA-963H5	bK963H5
AL079299	RP11-492A7	bA492A7
AL022336	RP1-78F24	dJ78F24
AC004542	RP3-430N8	dJ430N8
AC005233	RP5-1198O21	dJ1198O21
AC005005	RP3-412A9	dJ412A9
AC002073	RP3-515N1	dJ515N1
AC005003	RP3-400N23	dJ400N23
AL096702	RP11-254F5	bA254F5
AL096701	RP11-247I13	bA247I13
AL109802	RP11-163M1	bA163M1
AL096768	RP5-858B16	dJ858B16
AL031255	RP4-694E4	dJ694E4
AC005004	RP3-403E2	dJ403E2
AL022331	CTA-440B3	bK440B3
Z82190	RP1-180M12	dJ180M12
Z83856	LL22NC03-113A11	cN113A11
Z82248	LL22NC03-44A4	cN44A4
AL008719	CTA-342B11	bK342B11
Z74021	SC22CB-1E7	cB1E7
Z80998	SC22CB-36G12	cB36G12
Z83849	CITF22-65B7	fF65B7
AL022321	RP1-20O8	dJ20O8
Z83839	RP1-127L4	dJ127L4
AL008723	RP1-90G24	dJ90G24
AL021937	RP1-149A16	dJ149A16
AL035068	RP1-116G19	dJ116G19
Z71183	LL22NC03-28H9	cN28H9
Z82181	LL22NC01-86D10	cE86D10
Z80902	LL22NC03-80H12	cN80H12
AL021452	CTA-285F3	bK285F3
Z82246	LL22NC03-104C7	cN104C7
Z75744	LL22NC03-117B5	cN117B5
Z81309	LL22NC01-92H8	cE92H8
Z69714	LL22NC03-37F10	cN37F10
Z72521	LL22NC03-29F4	cN29F4
Z72520	LL22NC03-19H5	cN19H5
Z73495	LL22NC01-116C6	cE116C6
AL023282	CTA-766E1	bK766E1
Z73979	SC22CB-10B1	cB10B1
Z98256	RP1-309I22	dJ309I22
AL031592	CTA-366B10	bK366B10
Z83846	CTA-415G2	bK415G2
Z82198	RP1-302D9	dJ302D9
AL008630	CTA-282F2	bK282F2
Z82173	SC22CB-1D7	cB1D7
AL008715	CTC-216H12	bK1216H12
AL023577	CTA-566G5	bK566G5

Z82179	LL22NC01-140F8	cE140F8
Z73421	LL22NC03-37D7	cN37D7
AL133451	LL22NC03-120B6	cN120B6
Z69943	LL22NC03-4F11	cN4F11
AL008640	SC22CB-33D11	cB33D11
Z70288	LL22NC01-78G1	cE78G1
Z97354	LL22NC03-117F11	cN117F11
Z69042	LL22NC01-95B1	cE95B1
Z68287	LL22NC03-38E12	cN38E12
Z76736	RP1-75E8	dJ75E8
Z69713	LL22NC03-20A6	cN20A6
Z68324	LL22NC03-7A10	cN7A10
AL096754	LL22NC03-2E9	cN2E9
Z49866	LL22NC03-73A10	cN73A10
Z54073	LL22NC03-13E1	cN13E1
Z77853	LL22NC03-53F3	cN53F3
Z69715	LL22NC03-74G7	cN74G7
Z73429	LL22NC03-32F9	cN32F9
Z69925	LL22NC03-116A5	cN116A5
Z68223	LL22NC01-110C7	cE110C7
Z83852	SC22CB-49C12	cB49C12
Z82182	LL22NC01-90C2	cE90C2
Z99704	LL22NC01-75B8	cE75B8
Z74581	LL22NC01-127C11	cE127C11
AL008641	LL22NC03-100B10	cN100B10
Z69707	LL22NC01-95B9	cE95B9
Z68285	LL22NC03-11D4	cN11D4
AL022338	LL22NC01-82F7	cE82F7
Z82250	LL22NC03-86D4	cN86D4
AL008717	CTA-221H1	bK221H1
Z68323	LL22NC03-13G6	cN13G6
Z68288	LL22NC03-5E4	cN5E4
Z68325	LL22NC03-98E6	cN98E6
Z69712	LL22NC03-12G10	cN12G10
Z68754	LL22NC01-78H10	cE78H10
Z68758	LL22NC03-85E10	cN85E10
Z50860	LL22NC03-76A1	cN76A1
AL020992	CTA-363A12	bK363A12
Z68224	LL22NC01-129H9	cE129H9
AL049750	LL22NC01-141E2	cE141E2
Z69907	LL22NC03-22D1	cN22D1
AL031001	RP1-281O16	dJ281O16
AL021877	RP1-101G11	dJ101G11
Z82196	RP1-288L1	dJ288L1
AL024495	RP3-404L14	dJ404L14
Z82194	RP1-272J12	dJ272J12
Z83853	SC22CB-109E1	cB109E1
AL024494	RP1-215F16	dJ215F16
Z99755	CTA-714B7	bK714B7
AL031300	RP3-323A16	dJ323A16
AL008635	RP3-510H16	dJ510H16
Z82244	CTA-286B10	bK286B10

AL009049	RP5-824I19	dJ824I19
AL022334	RP4-569D19	dJ569D19
AL049747	CITF22-62D4	fF62D4
Z79996	SC22CB-33F2	cB33F2
AL049748	RP1-41P2	dJ41P2
AL079295	RP1-106I20	dJ106I20
Z82217	RP1-78B3	dJ78B3
Z95114	CTA-212A2	bK212A2
AL031426	CTA-191D12	bK191D12
Z82215	RP1-68O2	dJ68O2
AL022302	RP4-633O19	dJ633O19
AL022313	RP5-1119A7	dJ1119A7
AL031845	CTA-566H6	bK566H6
Z70289	CITF22-4G12	fF4G12
AL049749	RP1-293L6	dJ293L6
Z80897	LL22NC01-132D12	cE132D12
Z82184	CITF22-126G10	fF126G10
Z82185	CITF22-24E5	fF24E5
AL008637	CTA-833B7	bK833B7
AL133392	CITF22-45C1	fF45C1
Z82180	LL22NC01-81G9	cE81G9
Z73420	LL22NC01-146D10	cE146D10
AL022314	RP5-1170K4	dJ1170K4
Z82188	RP1-151B14	dJ151B14
Z94160	RP1-63G5	dJ63G5
AL049850	RP5-889J22	dJ889J22
Z93096	CTA-390B3	bK390B3
AL022315	RP5-1177I5	dJ1177I5
AL109980	RP4-697G8	dJ697G8
AL035496	RP3-437O22	dJ437O22
Z83844	RP1-37E16	dJ37E16
Z97630	RP3-466N1	dJ466N1
AL022311	RP5-1014D13	dJ1014D13
AL031587	RP5-1039K5	dJ1039K5
AL022322	CTA-228A9	bK228A9
AL021977	CTA-447C4	bK447C4
AL020993	RP1-5O6	dJ5O6
Z98749	RP3-449O17	dJ449O17
Z97056	RP3-434P1	dJ434P1
AL022320	RP1-199H16	dJ199H16
AL035495	RP1-319F24	dJ319F24
AL021707	RP3-508I15	dJ508I15
AL021806	RP4-779B17	dJ779B17
AL008583	RP3-327J16	dJ327J16
AL022318	CTA-150C2	bK150C2
AL031846	RP4-742C19	dJ742C19
Z81010	LL22NC03-10C3	cN10C3
AL031590	CTA-232D4	bK232D4
AL022326	RP3-333H23	dJ333H23
Z83845	RP3-407F17	dJ407F17
AL022312	RP5-1104E15	dJ1104E15
AL008716	CTA-206C7	bK206C7

AL022319	RP1-172B20	dJ172B20
AL022353	CTA-352E11	bK352E11
Z82206	RP3-370M22	dJ370M22
Z83847	RP3-496C20	dJ496C20
AL033547	RP3-340K22	dJ340K22
AL031589	RP6-11O7	dA11O7
Z93783	RP3-377F16	dJ377F16
AL022238	RP5-1042K10	dJ1042K10
AL031594	RP4-591N18	dJ591N18
Z86090	CTA-229A8	bK229A8
AL096703	RP4-735G18	dJ735G18
AL049764	RP3-362J20	dJ362J20
Z98048	RP3-408N23	dJ408N23
AL035450	RP5-1057D18	dJ1057D18
AL080242	RP11-554C12	bA554C12
AL080243	RP11-12M9	bA12M9
AL096765	RP11-422A16	bA422A16
AL035658	RP1-85F18	dJ85F18
AL035681	RP4-756G23	dJ756G23
AL035659	RP5-979N1	dJ979N1
AL008582	CTA-223H9	bK223H9
AL023553	RP3-347H13	dJ347H13
Z83840	CTA-216E10	bK216E10
AL023879	CTA-109G6	bK109G6
AL021453	RP5-821D11	dJ821D11
Z99716	CTA-250D10	bK250D10
Z82192	RP1-186O1	dJ186O1
AL021878	RP1-257I20	dJ257I20
AL031346	RP4-597B2	dJ597B2
Z83851	CTA-989H11	bK989H11
AL022316	CTA-126B4	bK126B4
AL035418	RP1-141I3	dJ141I3
Z93241	RP1-222E13	dJ222E13
Z82176	SC22CB-33B7	cB33B7
AL049757	RP1-47A17	dJ47A17
AL049758	RP3-437M21	dJ437M21
AL022476	RP3-323M22	dJ323M22
AL022237	CTB-191B2	bK1191B2
Z82214	RP3-526I14	dJ526I14
Z99756	RP1-100N22	dJ100N22
AL096761	RP4-754E20	dJ754E20
Z82172	SC22CB-13C9	cB13C9
AL118498	RP1-185D5	dJ185D5
AL031843	RP1-246D7	dJ246D7
Z82201	RP3-345P10	dJ345P10
AL023801	RP4-786D3	dJ786D3
Z97055	RP3-388M5	dJ388M5
AL023654	RP4-549K18	dJ549K18
AL035398	RP4-796I17	dJ796I17
Z82178	SC22CB-79B4	cB79B4
Z82174	SC22CB-20F6	cB20F6
AL033543	CTA-414D7	bK414D7

AL031595	RP4-671O14	dJ671O14
AL022339	LL22NC03-75B3	cN75B3
AL591914	CITF22-4H11	fF4H11
AL671760	CITF22-11D1	fF11D1
AL929500	CITF22-57B10	fF57B10
Z85994	RP1-32I10	dJ32I10
Z81308	CTA-397C4	bK397C4
AL023973	RP5-1033E15	dJ1033E15
Z75407	LL22NC03-128A12	cN128A12
AL022317	RP1-140L1	dJ140L1
AL022333	RP3-474I12	dJ474I12
Z98743	RP1-181C9	dJ181C9
Z93244	CTA-116F5	bK116F5
Z83838	RP1-127B20	dJ127B20
AL079301	RP4-753M9	dJ753M9
Z82243	CTA-217C2	bK217C2
AL008718	CTA-268H5	bK268H5
AL021391	RP1-102D24	dJ102D24
Z98047	RP1-162H14	dJ162H14
Z95331	CTA-941F9	bK941F9
Z93784	RP3-398C22	dJ398C22
Z84478	RP1-37M3	dJ37M3
AL049811	RP11-140I15	bA140I15
AL929387	RP11-398F12	bA398F12
Z94161	LL22NC03-102C10	cN102C10
AL049856	RP4-695O20	dJ695O20
AL078611	SC22cB-5E3	cB5E3
AL031034	LL22NC03-98G1	cN98G1
Z93024	LL22NC03-5H6	cN5H6
AL031844	RP3-361H15	dJ361H15
AL031588	RP5-1163J1	dJ1163J1
AL031597	RP5-996D20	dJ996D20
AL021392	RP3-439F8	dJ439F8
AL096766	RP6-59H18	dA59H18
AL118516	CTA-29F11	bK29F11
AL078642	RP5-917C11	dJ917C11
Z97351	RP1-116M15	dJ116M15
Z82187	CITF22-64F4	fF64F4
U51561	LLcos-79E2	cN79E2
AL023576	CTA-358H9	bK358H9
Z80896	RP1-67C13	dJ67C13
Z79999	CITF22-111A3	fF111A3
AL023733	SC22cB-58F5	cB58F5
Z81000	CITF22-67D6	fF67D6
AL096755	RP3-477J10	dJ477J10
U51559	LLcos-65D1	cN65D1
AL096756	RP3-477J10	dJ494G10
Z82183	LL22NC01-98F6	cE98F6
Z83836	RP1-111J24	dJ111J24
Z82186	CITF22-49D8	fF49D8
AL035069	RP3-404P13	dJ404P13
Z84496	LL22NC03-75H12	cN75H12

AL021306	CTB-109B5	bK1109B5
Z80999	LL22NC01-140G5	cE140G5
Z73416	LL22NC03-114B2	cN114B2
AL033544	LL22NC01-107C5	cE107C5
AL049568	SC22CB-23F1	cB23F1
Z80901	LL22NC03-119A7	cN119A7
AC000034	AC000034	cN119A4
Z83854	SC22CB-44H3	cB44H3
AL117329	RP11-191L9	bA191L9
AL121580	RP13-455A7	bB455A7
AL110122	CTA-280A3	bK280A3
Z73963	LL22NC03-62C4	cN62C4
AL110121	CTA-280A3	bK280A3
Z70688	LL22NC03-27C5	cN27C5
Z81002	LL22NC03-53A9	cN53A9
Z80772	CITF22-37F6	fF37F6
AL118553	CITF22-91B7	fF91B7
Z78421	LL22NC03-121E8	cN121E8
AL118554	CITF22-91B7	fF91B7
AL050314	RP1-100G10	dJ100G10
AL008720	CTA-343C1	bK343C1
Z72006	LL22NC03-69F4	cN69F4
Z82249	LL22NC03-4D6	cN4D6
AL078640	RP11-536P6	bA536P6
Z84468	CTA-299D3	bK299D3
AL096853	CITF22-96H12	fF96H12
AL096843	RP11-262A13	bA262A13
Z83837	CITF22-113D11	fF113D11
Z83855	LL22NC03-104C4	cN104C4
AL078607	RP4-619N21	dJ619N21
AL954742	CITF22-114E11	fF114E11
AL078613	RP11-354I12	bA354I12
AL078622	RP5-925J7	dJ925J7
AL078632	RP11-255N20	bA255N20
AL049773	RP5-1061O18	dJ1061O18
Z94162	LL22NC03-21F1	cN21F1
Z82202	RP1-34P24	dJ34P24
Z82189	RP1-170A21	dJ170A21
AL008636	CTA-722E9	bK722E9
AL031593	RP4-566L20	dJ566L20
Z97192	RP1-29C18	dJ29C18
AL023802	RP5-983L19	dJ983L19
Z98885	RP3-522J7	dJ522J7
AL117328	RP11-494O16	bA494O16
AL080240	RP11-232E17	bA232E17
AL022327	RP3-355C18	dJ355C18
AL034546	RP5-898I4	dJ898I4
AL022328	RP3-402G11	dJ402G11
AL954743	CITF22-49G11	fF49G11
AL671545	CITF22-53E7	fF53Et
AL096767	RP4-579N16	dJ579N16
Z94802	CTA-999D10	bK999D10

U62317	U62317	bK384D8
Z82168	RP1-104C13	dJ104C13
Z82245	CTA-799F10	bK799F10
AC000050	AC000050	cN66C4
AC000036	AC000036	cN85A3
AC002056	AC002056	cN94H12
AC002055	AC002055	cN1G3

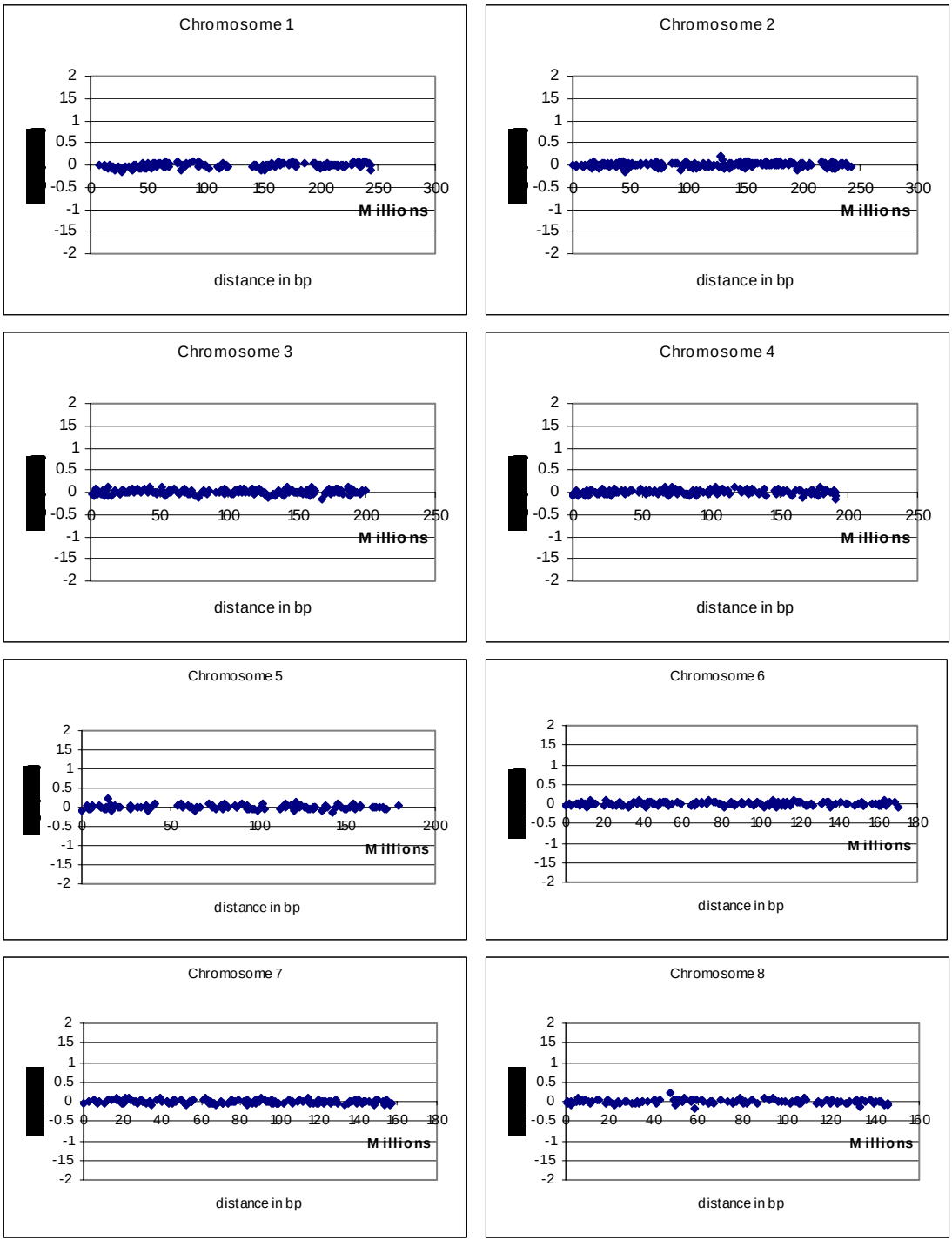
9b: Clone Libraries

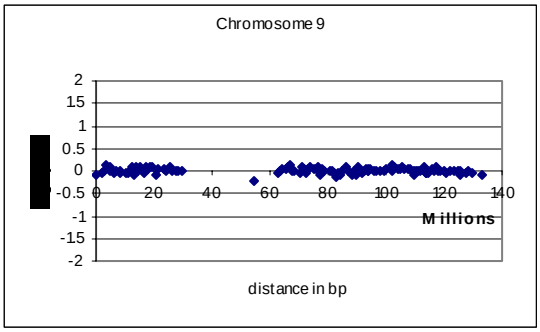
Table adapted from, http://www.sanger.ac.uk/HGP/methods/mapping/info/lib-details.shtml

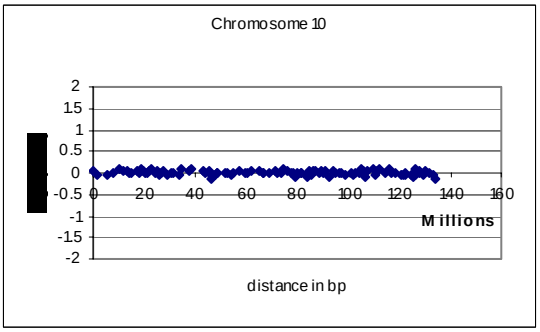
Library Type	**Library**	**Library code**	**Antibiotic**
Cosmid	Sc22cB, Sanger flow sorted chromosome 22	cB	Kanamycin 30μg/ml
Cosmid	LL22NC01 “E”, Paris flow sorted chromosome 22	cE	Kanamycin 30μg/ml
Cosmid	LL22NC03 “N”, Lawerence Livermore flow sorted chromosome 22	cN	Kanamycin 30μg/ml
Fosmid	CITF22, Caltech flow sorted chromosome 22 fosmid library	fF	Chloramphenicol 25μg/ml
RPCI Human PAC	RPCI-1-5, de Jong whole genome male PAC library	dJ	Kanamycin 25μg/ml
RPCI Human PAC	RPCI-6, de Jong whole genome female PAC library	dA	Kanamycin 25μg/ml
BAC	CIT978SK (CTA, CTB and CTC) Caltech whole genome BAC library	bK	Chloramphenicol 12.5μg/ml
RPCI Human BAC	RPCI-11, Whole genome male BAC library	bA	Chloramphenicol 25μg/ml
RPCI Human BAC	RPCI-13, Whole genome female BAC library	bB	Chloramphenicol 25μg/ml

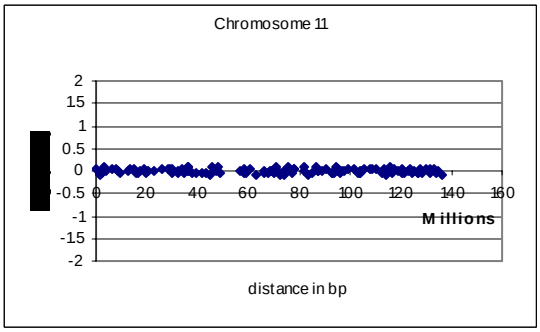
Appendix 10: 1Mb profiles of patients with DiGeorge phenotype and no 22q11 deletion.

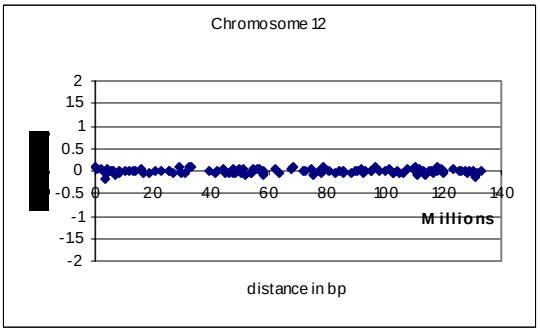
Patient 1:

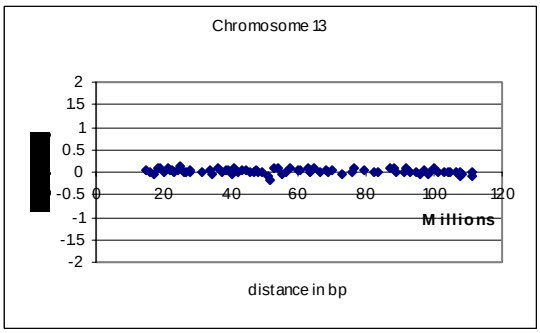


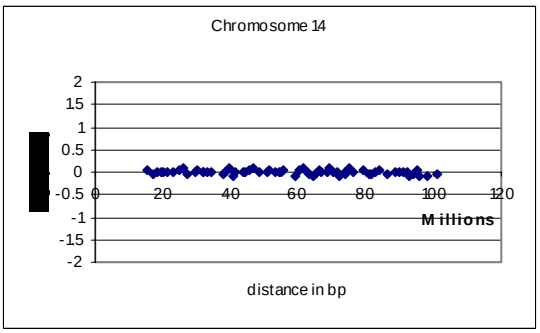

Chromosome 9
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
Millions
distance in bp

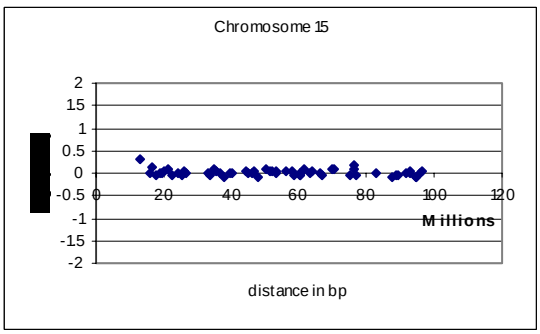

Chromosome 10
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
Millions
distance in bp

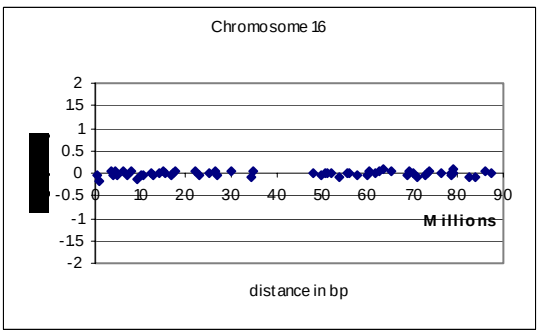

Chromosome 11
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
Millions
distance in bp

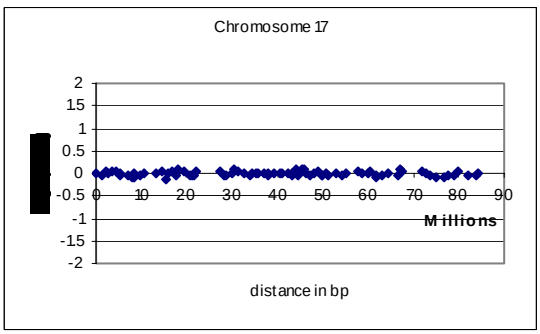

Chromosome 12
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
Millions
distance in bp

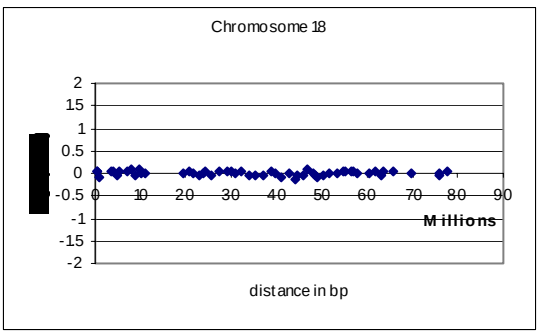

Chromosome 13
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

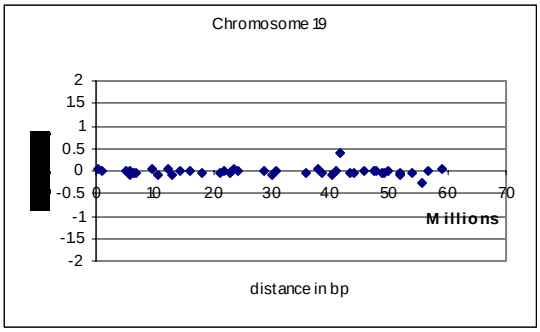

Chromosome 14
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

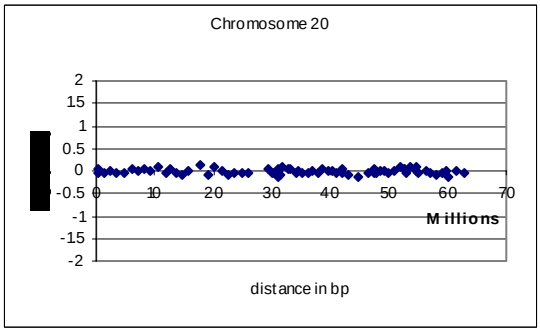

Chromosome 15
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

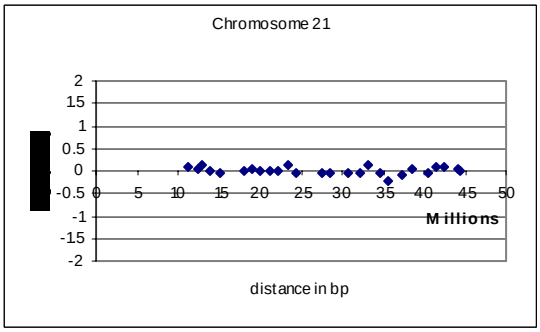

Chromosome 16
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp

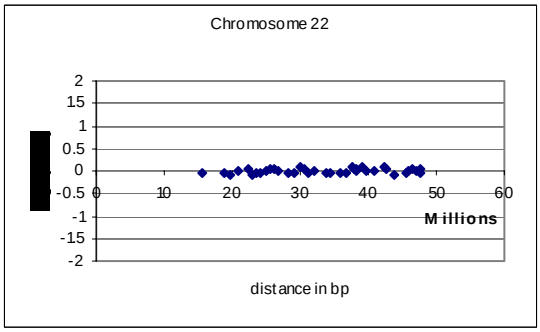

Chromosome 17
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp

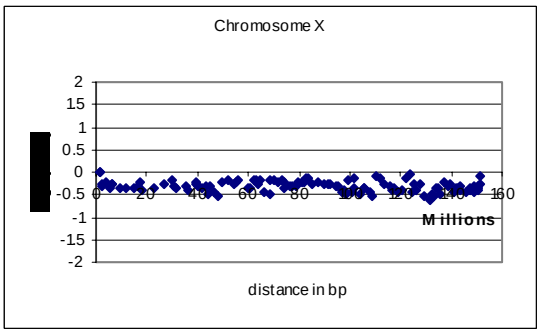

Chromosome 18
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp

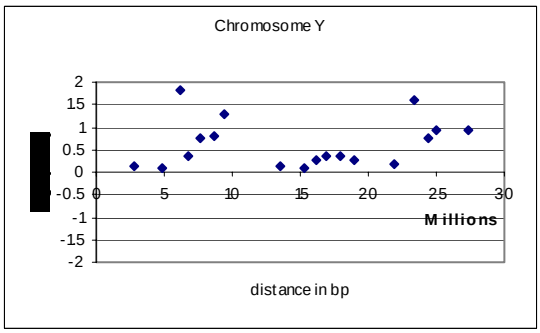

Chromosome 19
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
Millions
distance in bp


Chromosome 20
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
Millions
distance in bp

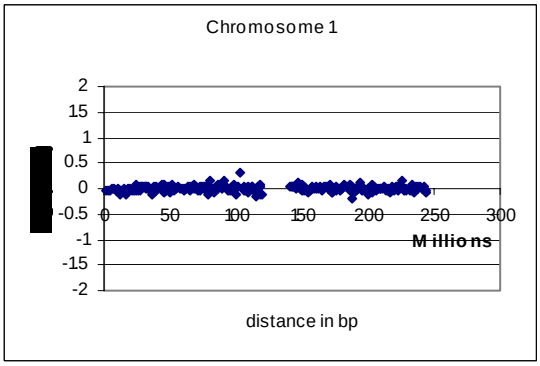

Chromosome 21
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
5
10
15
20
25
30
35
40
45
50
Millions
distance in bp

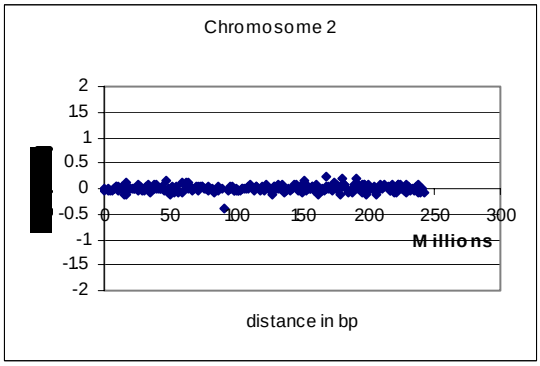

Chromosome 22
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
Millions
distance in bp

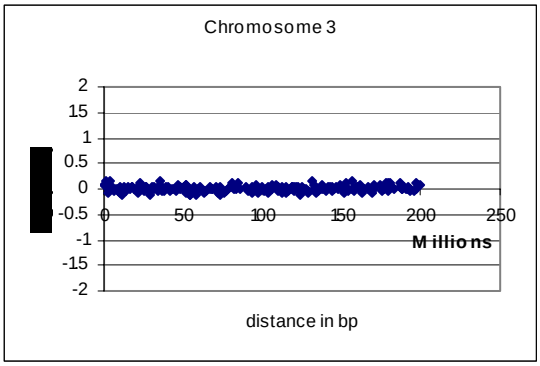

Chromosome X
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
Millions
distance in bp

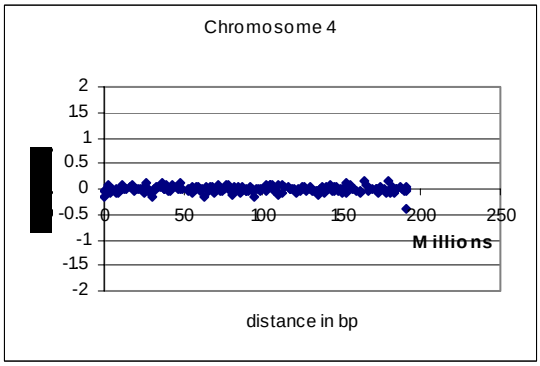

Chromosome Y
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
5
10
15
20
25
30
Millions
distance in bp

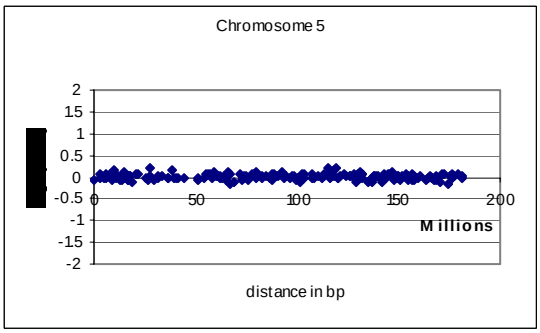
Patient 2:

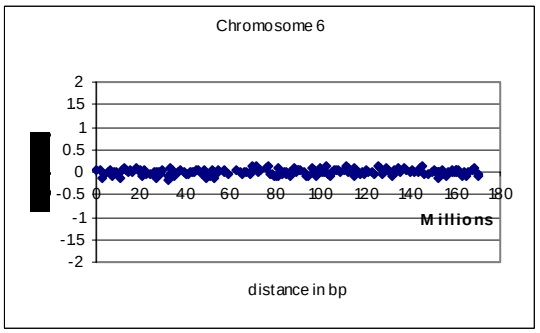

Chromosome 1
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
50
100
150
200
250
300
Millions
distance in bp

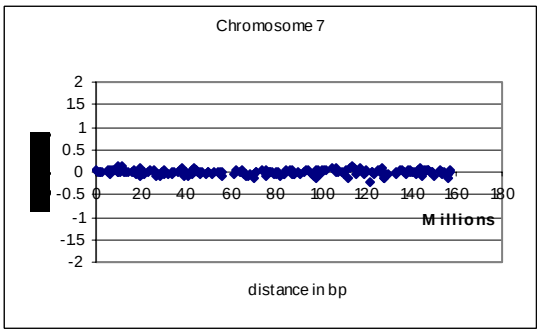

Chromosome 2
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
50
100
150
200
250
300
Millions
distance in bp

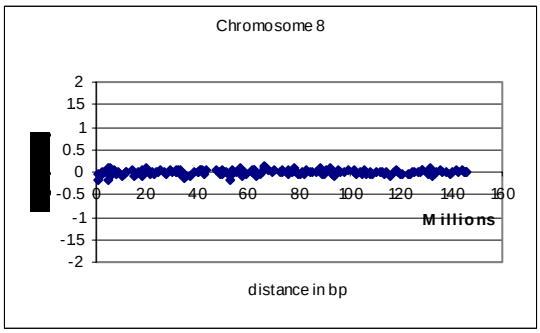

Chromosome 3
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
50
100
150
200
250
Millions
distance in bp

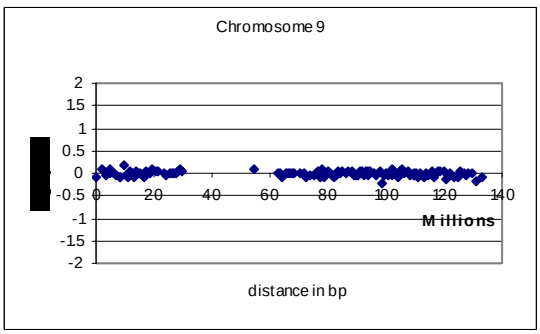

Chromosome 4
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
50
100
150
200
250
Millions
distance in bp

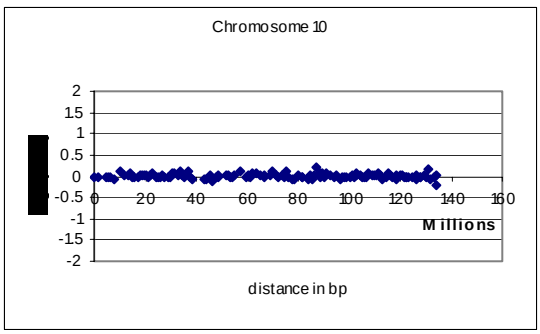

Chromosome 5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
50
100
150
200
Millions
distance in bp

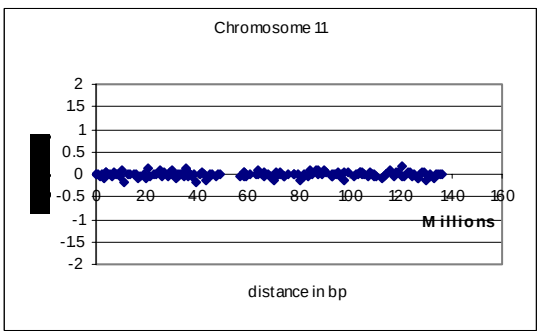

Chromosome 6
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
180
Millions
distance in bp

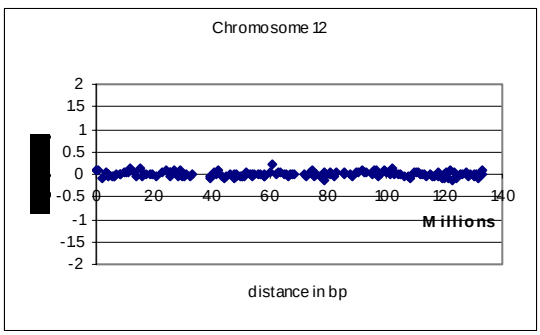

Chromosome 7
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
180
Millions
distance in bp

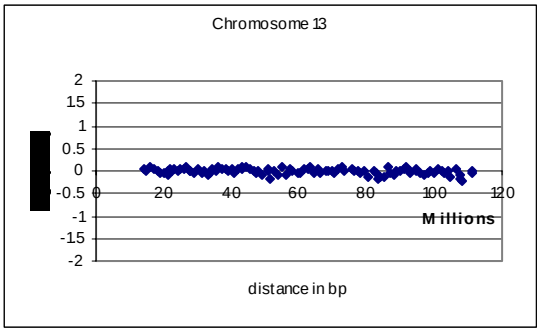

Chromosome 8
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
Millions
distance in bp

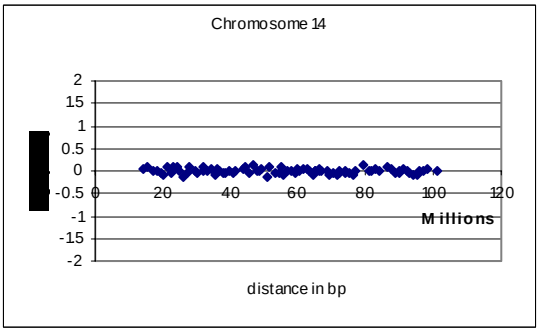

Chromosome 9
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
Millions
distance in bp

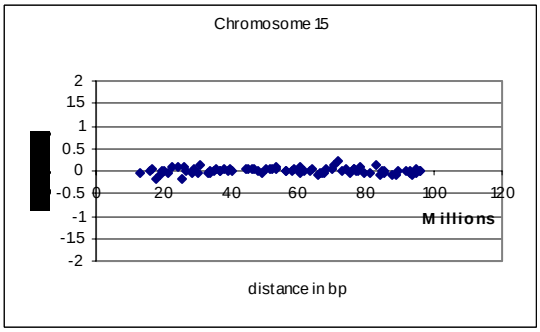

Chromosome 10
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
Millions
distance in bp

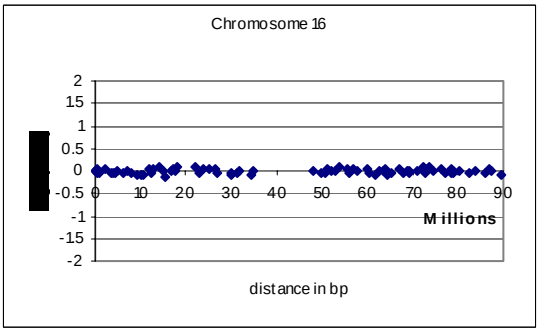

Chromosome 11
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
Millions
distance in bp

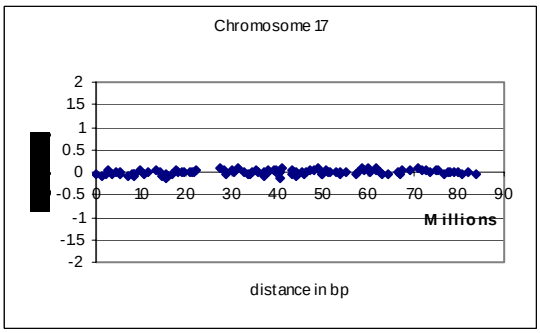

Chromosome 12
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
Millions
distance in bp

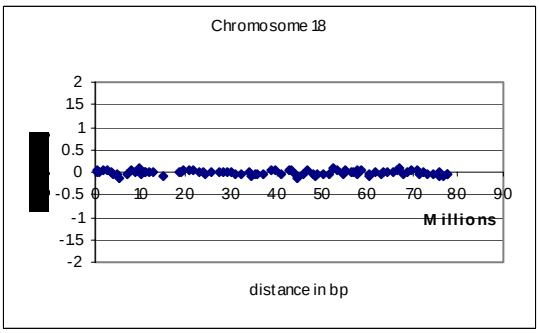

Chromosome 13
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

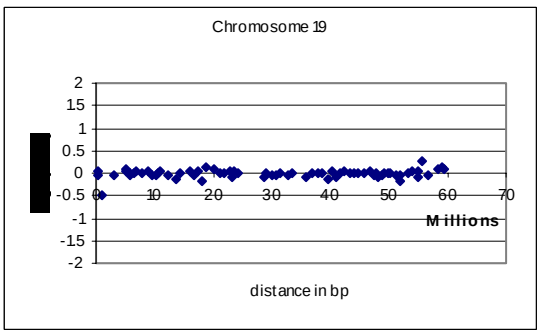

Chromosome 14
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

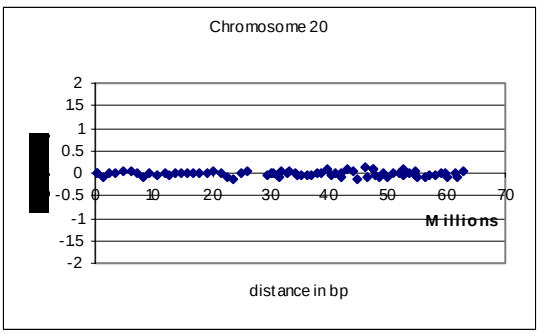

Chromosome 15
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

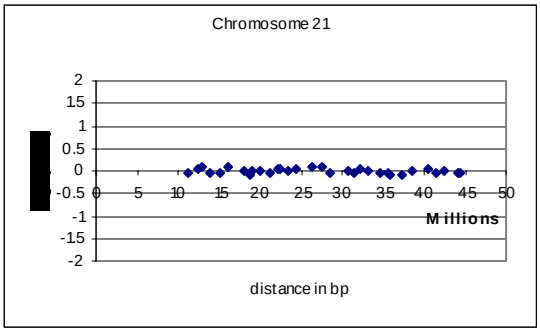

Chromosome 16
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp

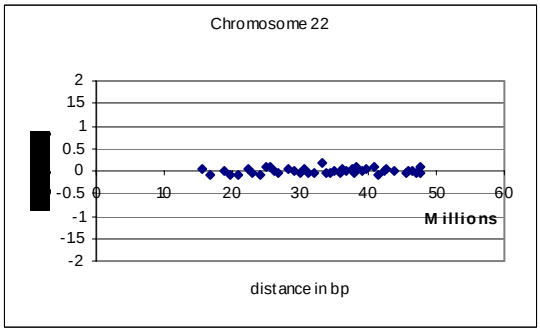

Chromosome 17
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp

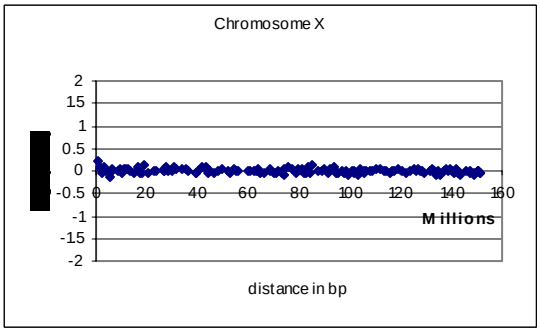

Chromosome 18
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp

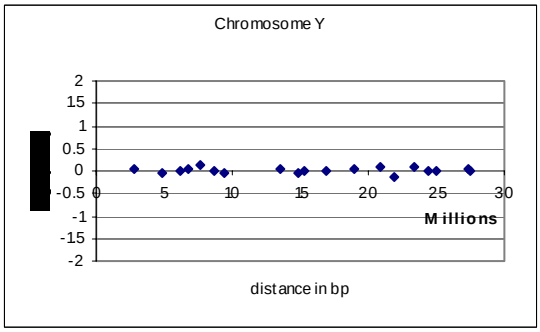

Chromosome 19
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
Millions
distance in bp


Chromosome 20
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
Millions
distance in bp

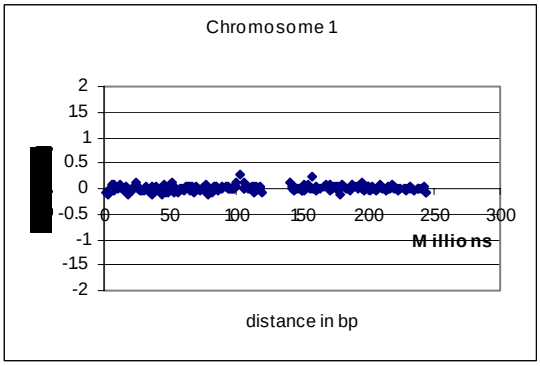

Chromosome 21
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0 5 10 15 20 25 30 35 40 45 50
Millions
distance in bp

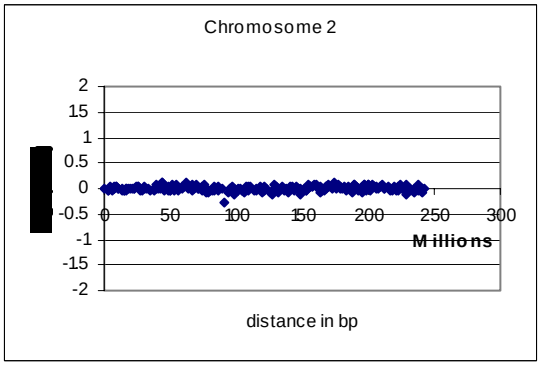

Chromosome 22
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0 10 20 30 40 50 60
Millions
distance in bp

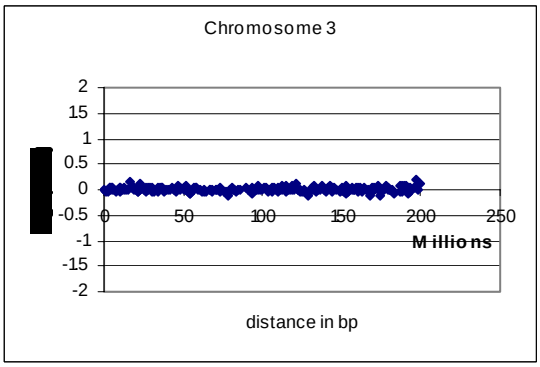

Chromosome X
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0 20 40 60 80 100 120 140 160
Millions
distance in bp

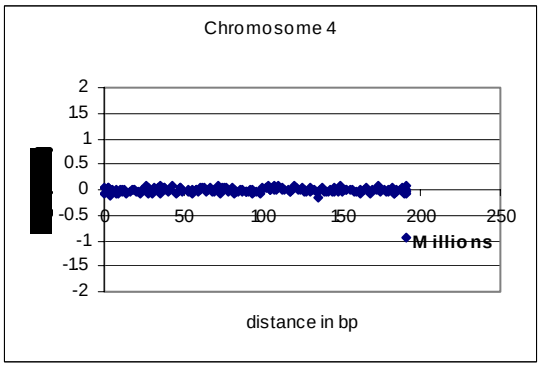

Chromosome Y
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0 5 10 15 20 25 30
Millions
distance in bp

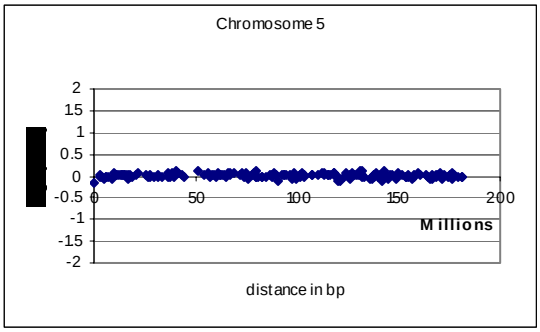
Patient 3:

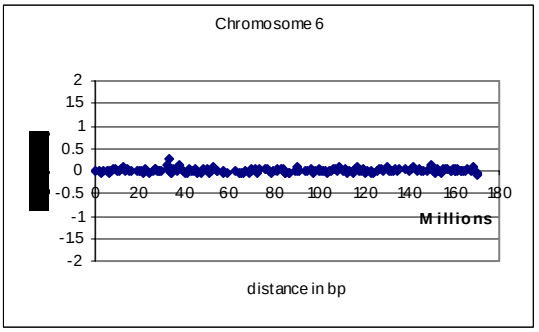

Chromosome 1
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0 50 100 150 200 250 300
Millions
distance in bp

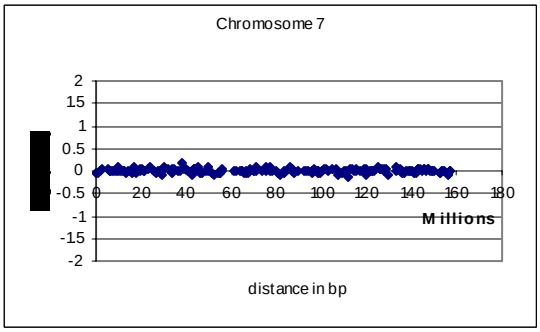

Chromosome 2
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0 50 100 150 200 250 300
Millions
distance in bp

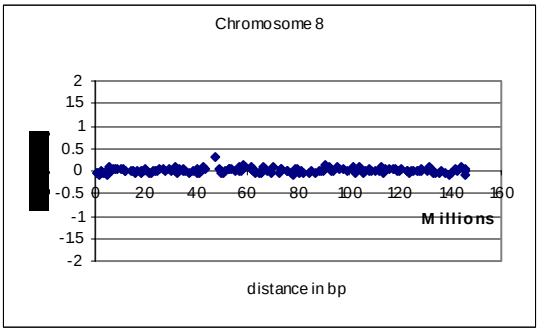

Chromosome 3
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0 50 100 150 200 250
Millions
distance in bp

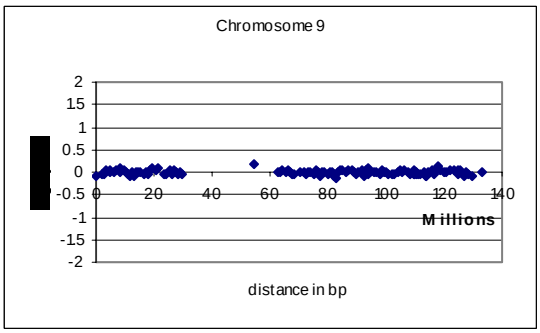

Chromosome 4
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0 50 100 150 200 250
Millions
distance in bp

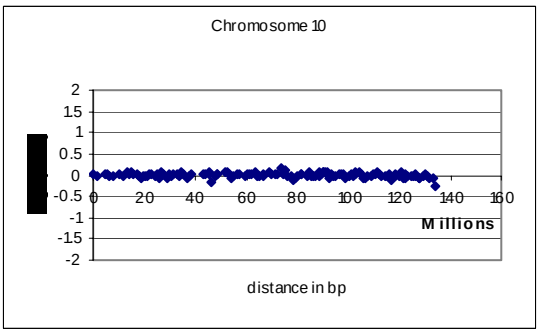

Chromosome 5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
50
100
150
200
Millions
distance in bp

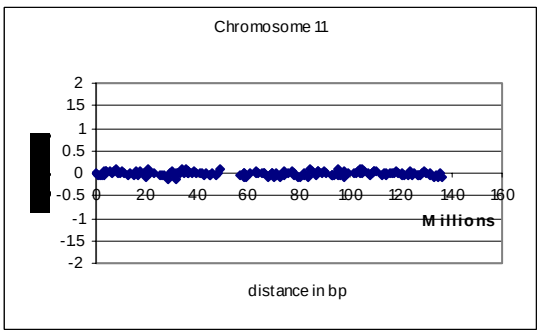

Chromosome 6
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
180
Millions
distance in bp

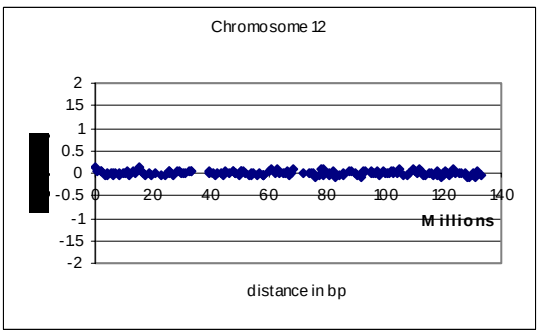

Chromosome 7
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
180
Millions
distance in bp

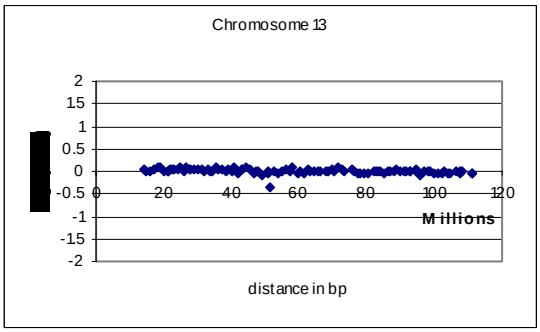

Chromosome 8
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
Millions
distance in bp

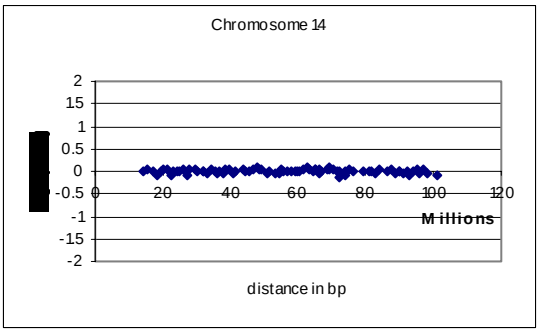

Chromosome 9
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
Millions
distance in bp

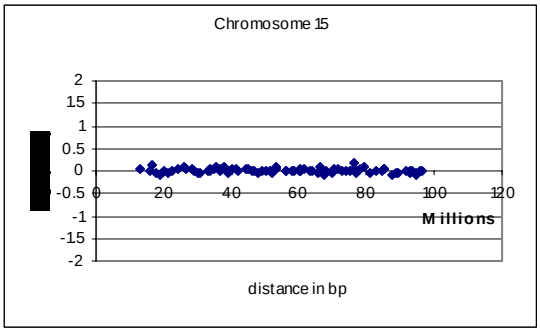

Chromosome 10
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
Millions
distance in bp

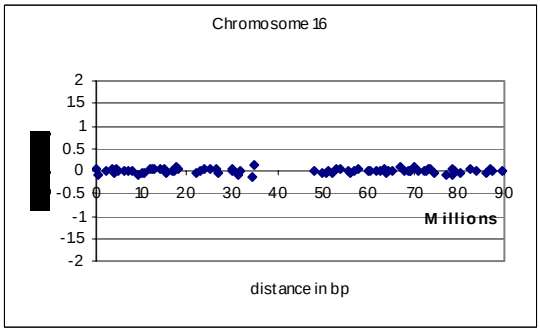

Chromosome 11
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
Millions
distance in bp

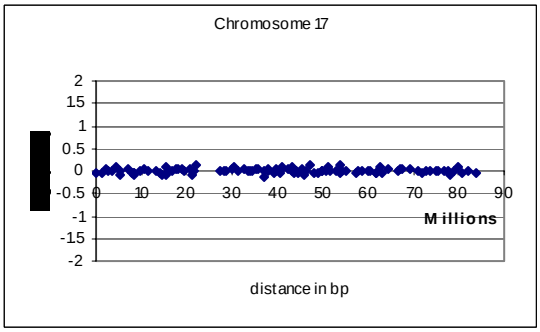

Chromosome 12
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
Millions
distance in bp

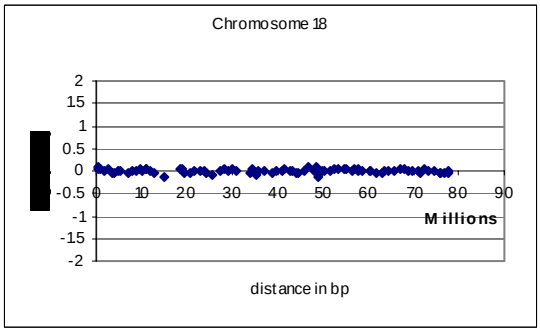

Chromosome 13
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

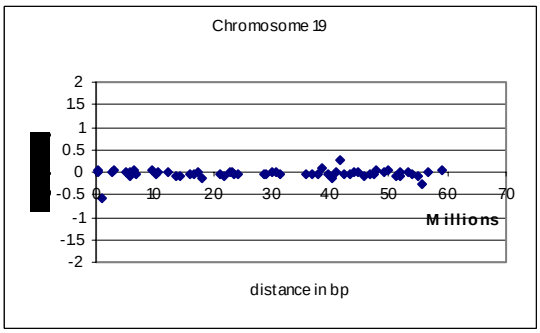

Chromosome 14
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

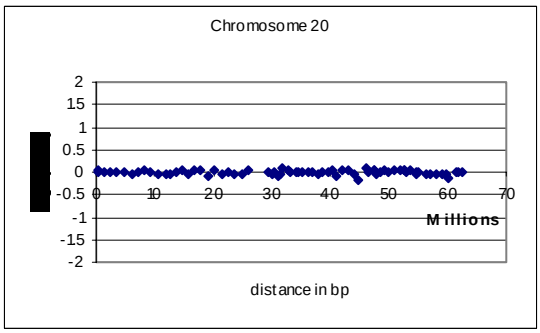

Chromosome 15
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

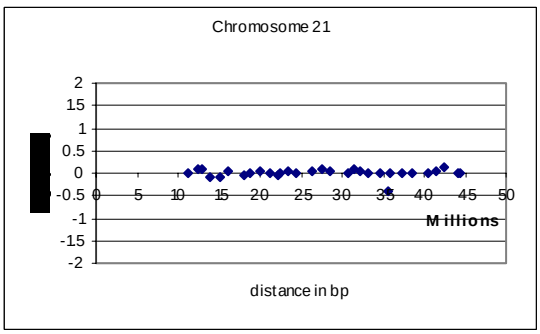

Chromosome 16
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp

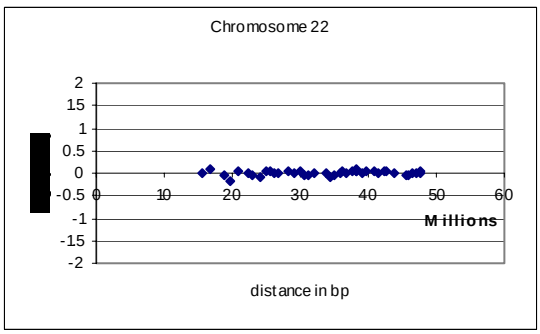

Chromosome 17
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp

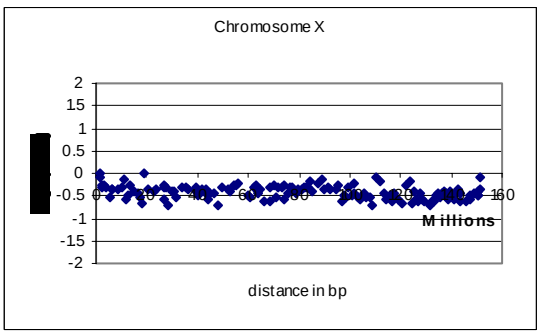

Chromosome 18
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp

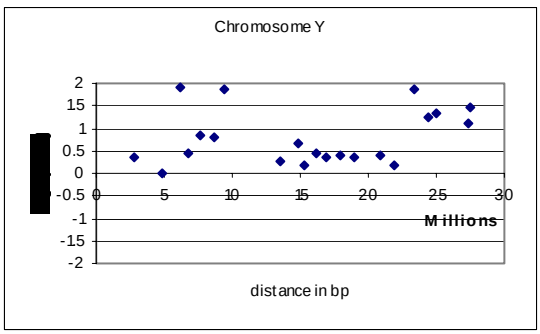

Chromosome 19
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
Millions
distance in bp


Chromosome 20
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
Millions
distance in bp

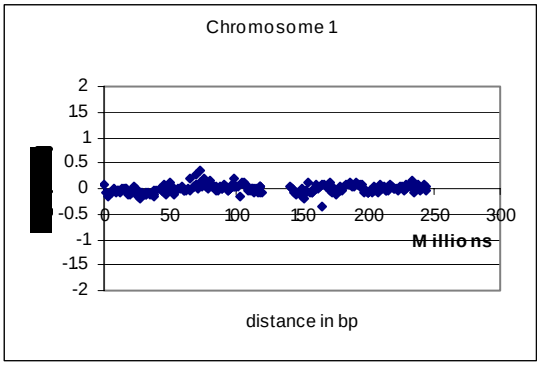

Chromosome 21
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
5
10
15
20
25
30
35
40
45
50
Millions
distance in bp

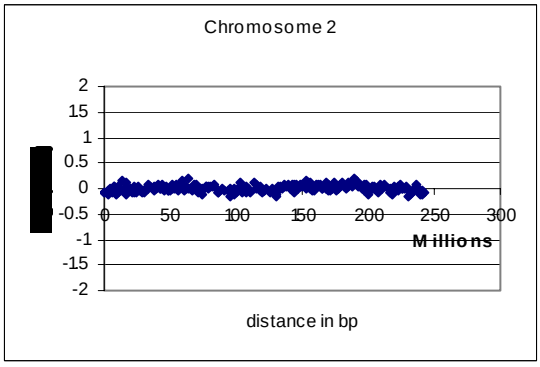

Chromosome 22
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
Millions
distance in bp

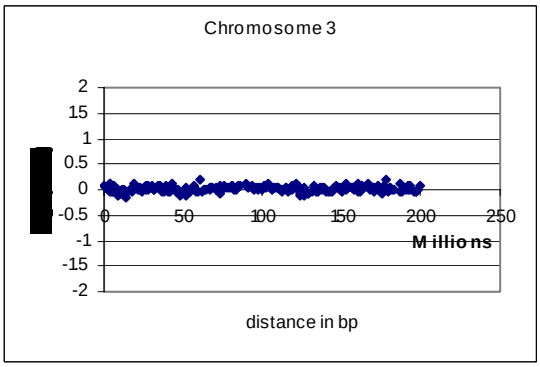

Chromosome X
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
Millions
distance in bp

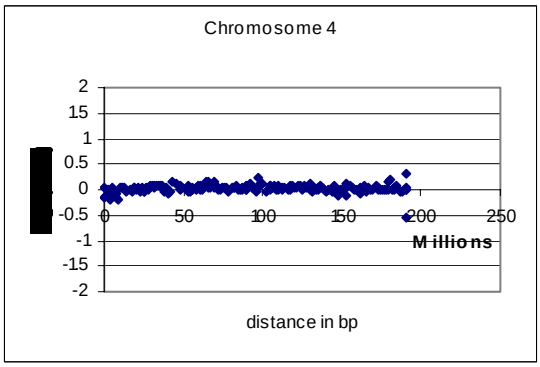

Chromosome Y
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
5
10
15
20
25
30
Millions
distance in bp

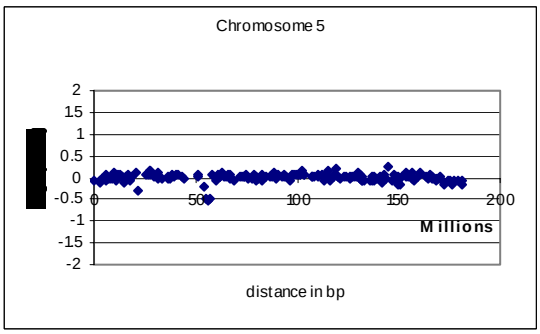
Patient 4:

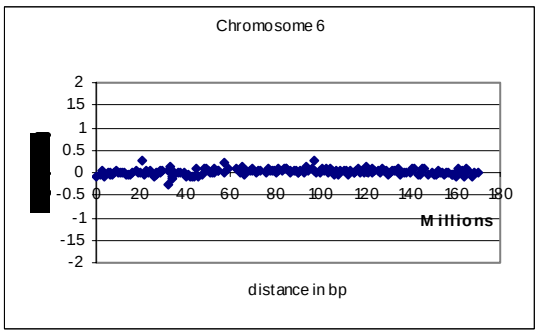

Chromosome 1
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
50
100
150
200
250
300
Millions
distance in bp

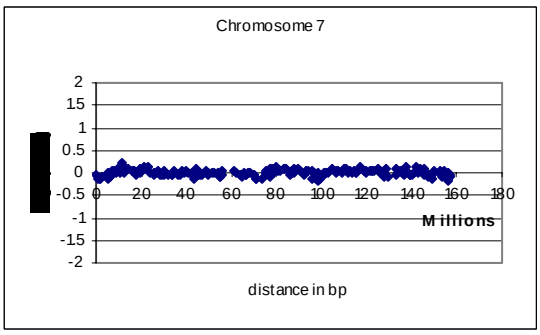

Chromosome 2
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
50
100
150
200
250
300
Millions
distance in bp

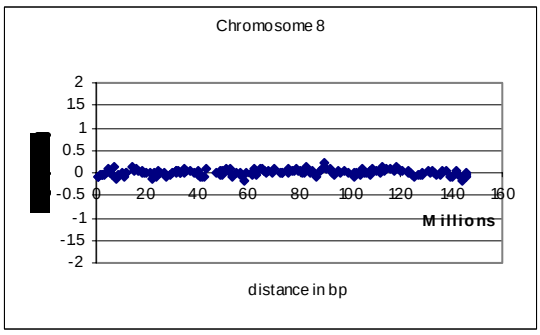

Chromosome 3
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
50
100
150
200
250
Millions
distance in bp

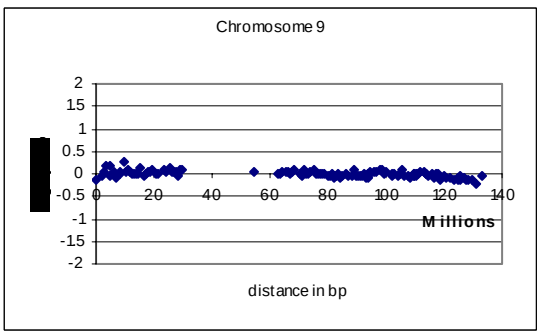

Chromosome 4
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
50
100
150
200
250
Millions
distance in bp

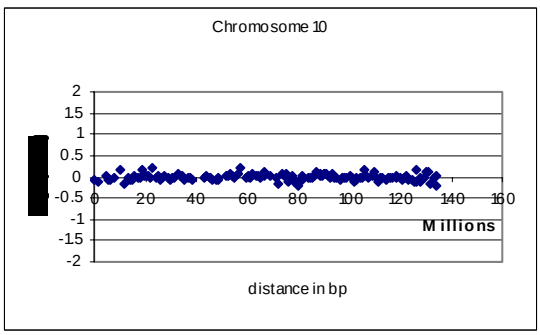

Chromosome 5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
50
100
150
200
Millions
distance in bp

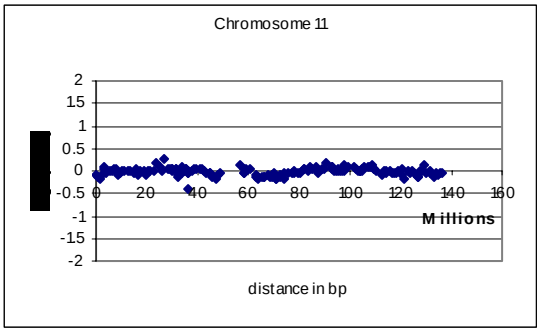

Chromosome 6
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
180
Millions
distance in bp

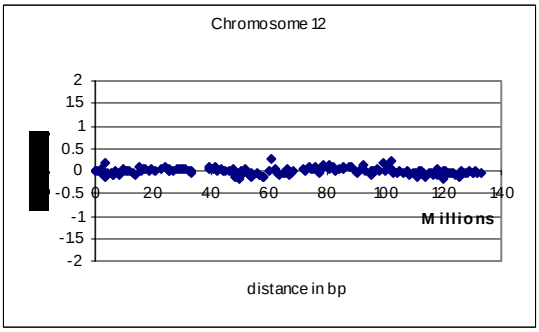

Chromosome 7
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
180
Millions
distance in bp

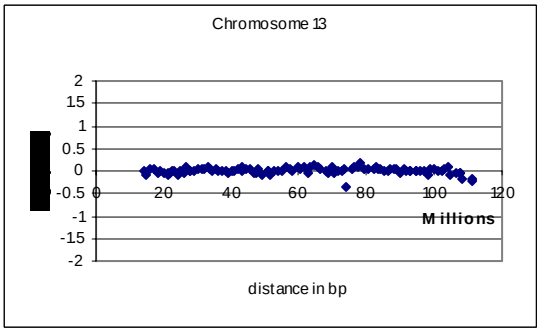

Chromosome 8
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
Millions
distance in bp

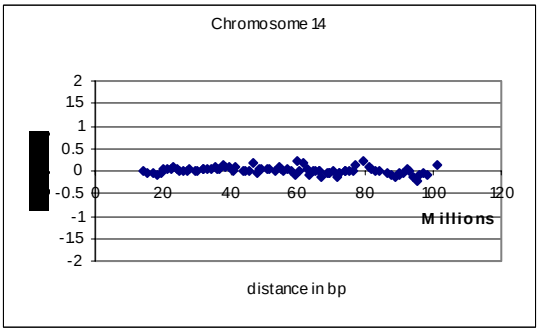

Chromosome 9
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
Millions
distance in bp

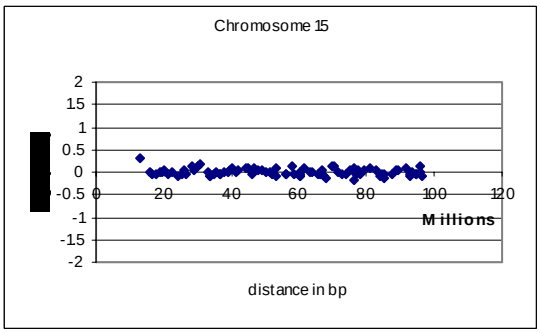

Chromosome 10
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
Millions
distance in bp

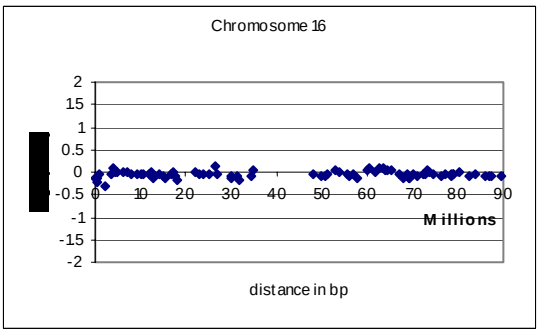

Chromosome 11
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
Millions
distance in bp

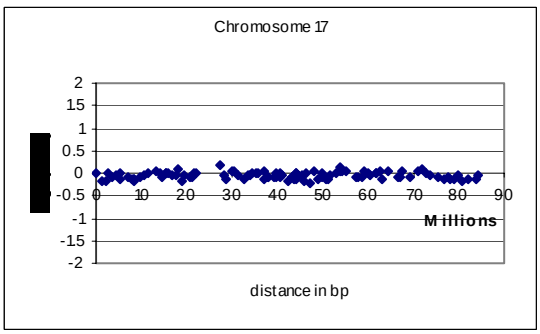

Chromosome 12
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
Millions
distance in bp

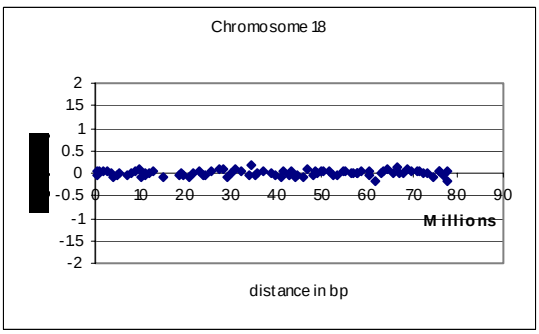

Chromosome 13
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

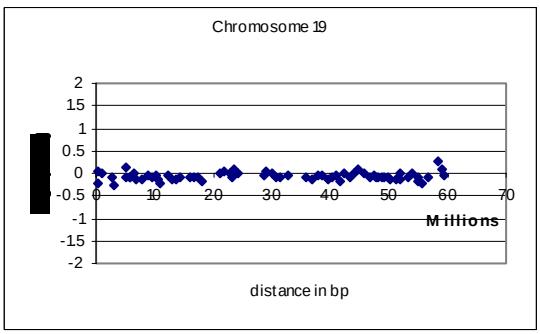

Chromosome 14
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

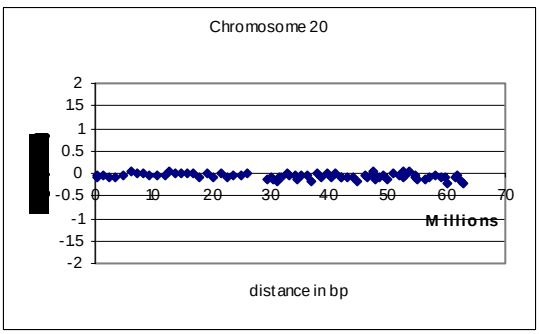

Chromosome 15
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

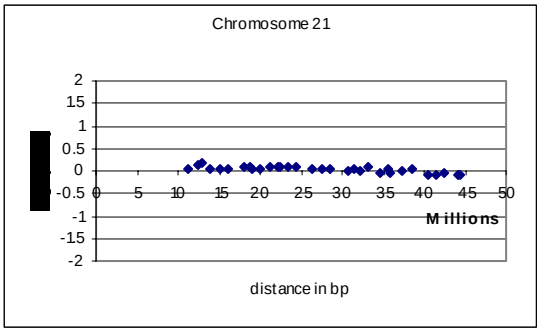

Chromosome 16
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp

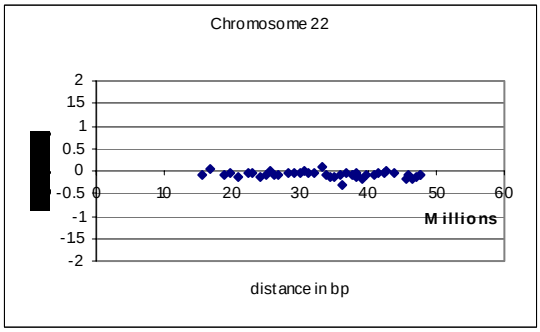

Chromosome 17
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp

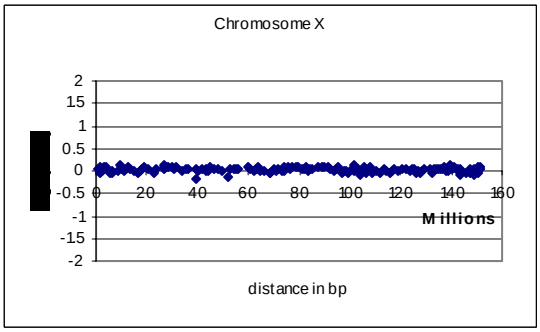

Chromosome 18
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp

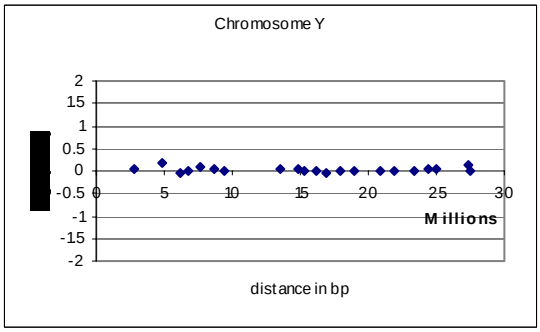

Chromosome 19
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
Millions
distance in bp


Chromosome 20
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
Millions
distance in bp

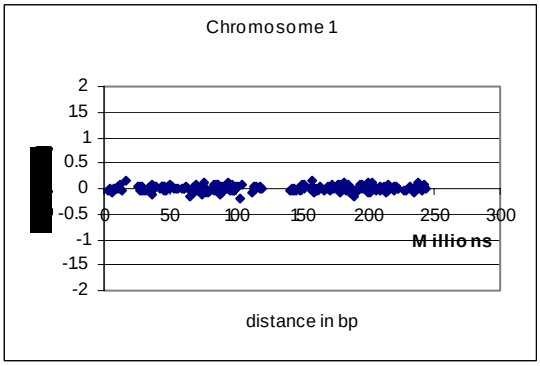

Chromosome 21
2
15
1
0.5
0
-0.5
-1
-15
-2
0 5 10 15 20 25 30 35 40 45 50
Millions
distance in bp

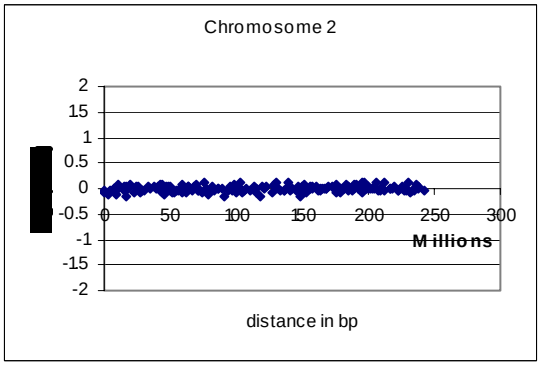

Chromosome 22
2
15
1
0.5
0
-0.5
-1
-15
-2
0 10 20 30 40 50 60
Millions
distance in bp

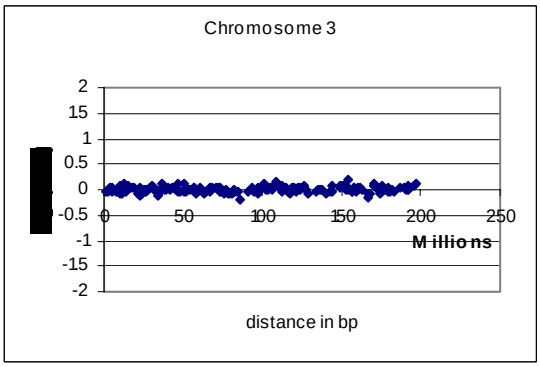

Chromosome X
2
15
1
0.5
0
-0.5
-1
-15
-2
0 20 40 60 80 100 120 140 160
Millions
distance in bp

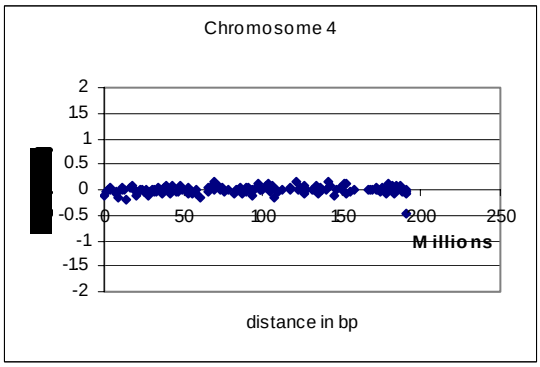

Chromosome Y
2
15
1
0.5
0
-0.5
-1
-15
-2
0 5 10 15 20 25 30
Millions
distance in bp

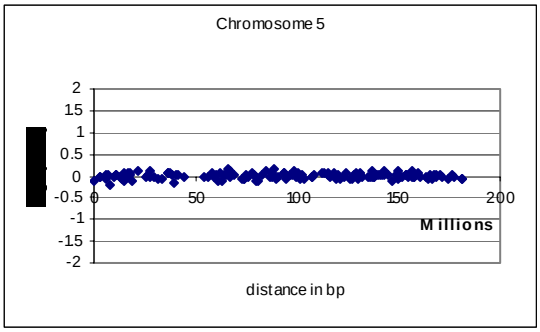
Patient 5:

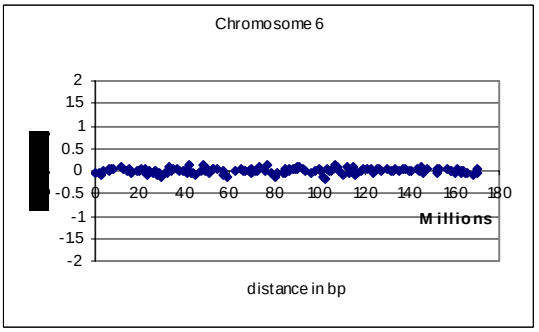

Chromosome 1
2
15
1
0.5
0
-0.5
-1
-15
-2
0 50 100 150 200 250 300
Millions
distance in bp

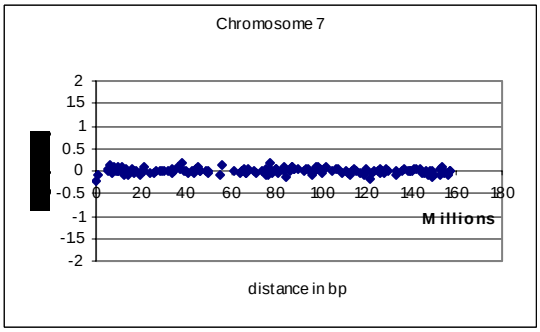

Chromosome 2
2
15
1
0.5
0
-0.5
-1
-15
-2
0 50 100 150 200 250 300
Millions
distance in bp

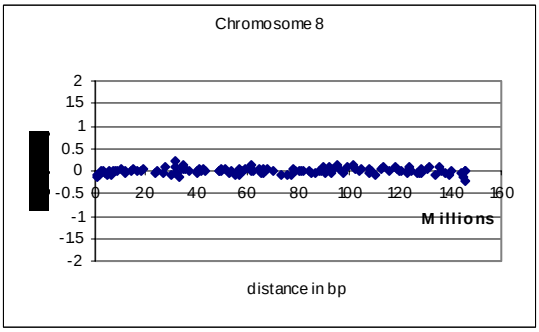

Chromosome 3
2
15
1
0.5
0
-0.5
-1
-15
-2
0 50 100 150 200 250
Millions
distance in bp

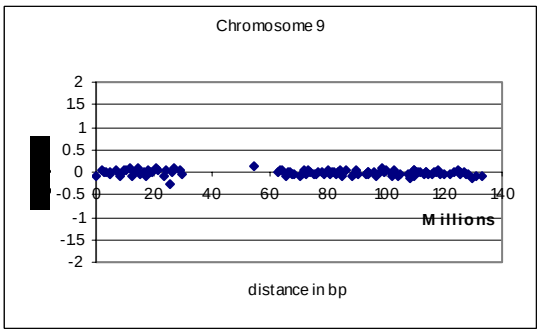

Chromosome 4
2
15
1
0.5
0
-0.5
-1
-15
-2
0 50 100 150 200 250
Millions
distance in bp

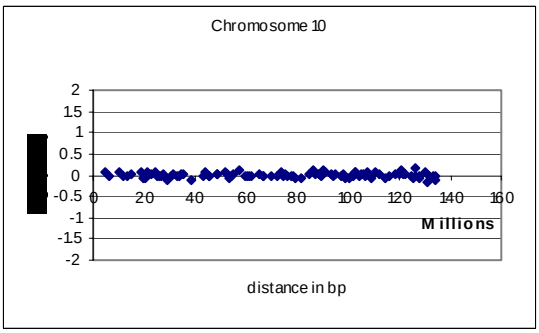

Chromosome 5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
50
100
150
200
Millions
distance in bp

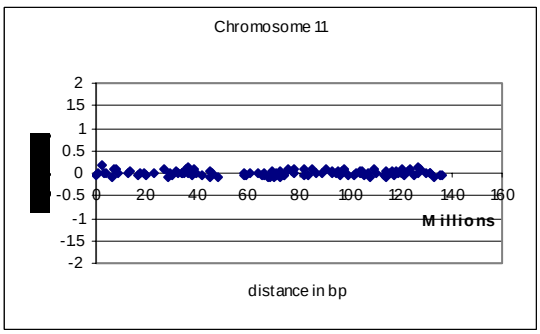

Chromosome 6
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
180
Millions
distance in bp

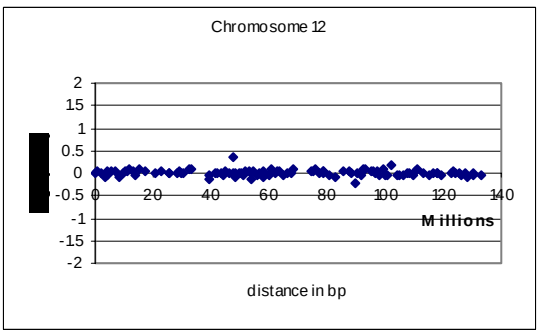

Chromosome 7
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
180
Millions
distance in bp

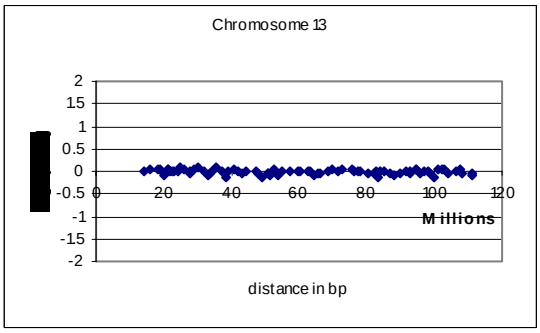

Chromosome 8
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
Millions
distance in bp

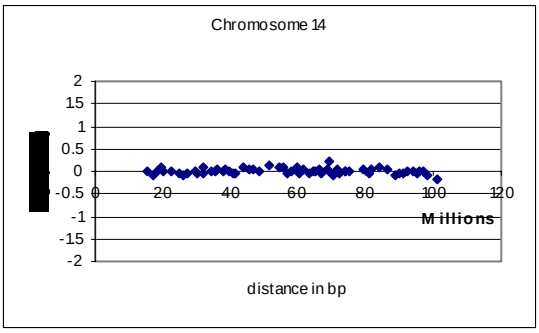

Chromosome 9
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
Millions
distance in bp

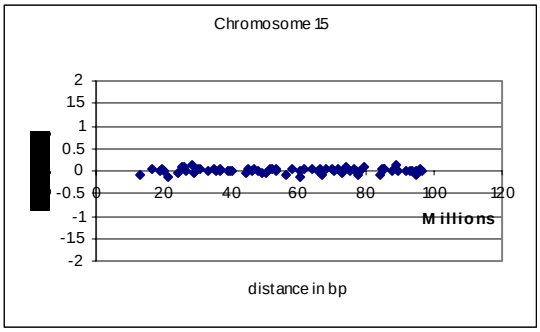

Chromosome 10
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
Millions
distance in bp

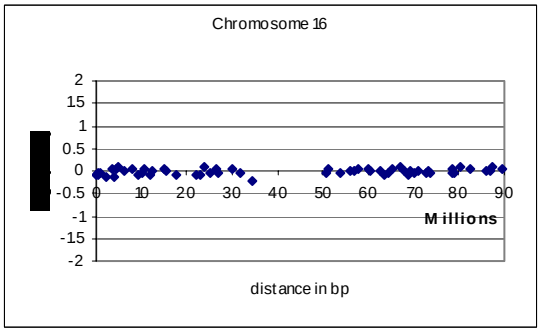

Chromosome 11
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
Millions
distance in bp

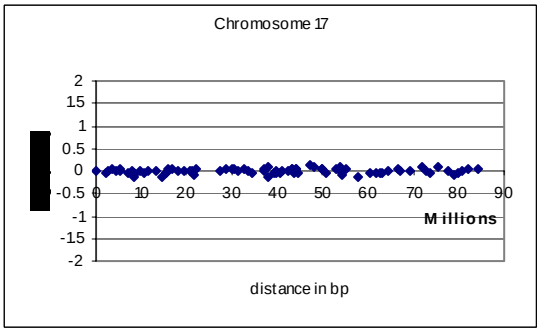

Chromosome 12
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
Millions
distance in bp

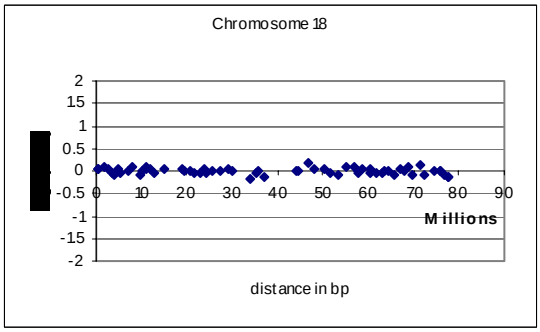

Chromosome 13
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

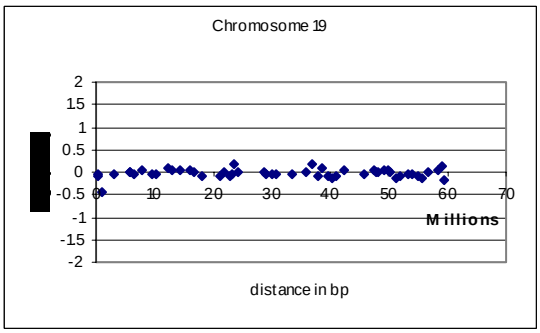

Chromosome 14
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

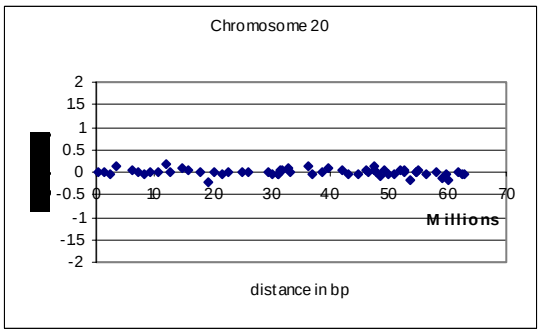

Chromosome 15
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

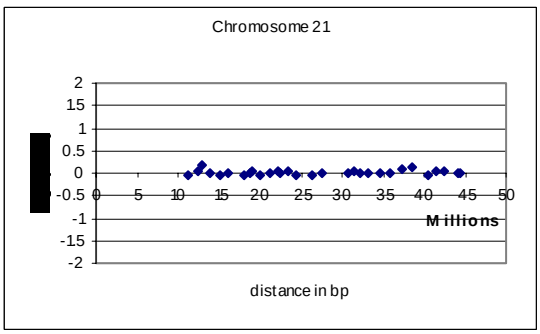

Chromosome 16
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp

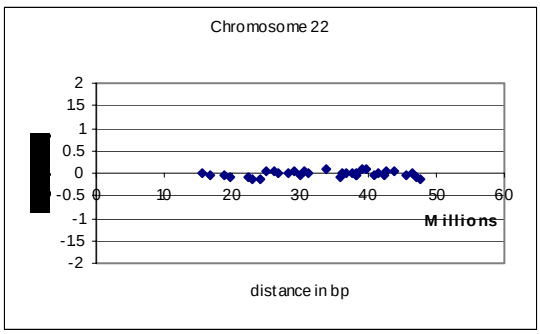

Chromosome 17
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp

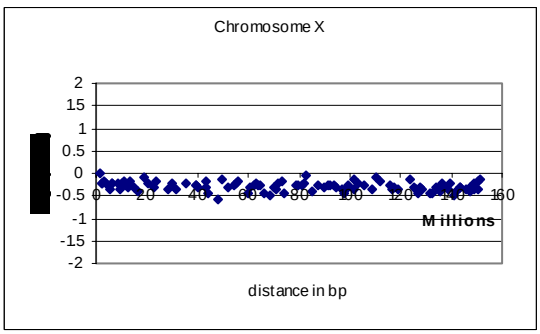

Chromosome 18
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp

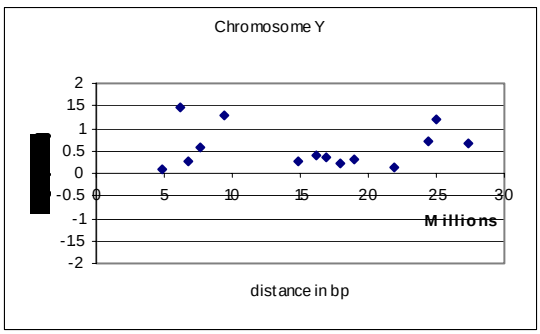

Chromosome 19
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
Millions
distance in bp


Chromosome 20
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
Millions
distance in bp

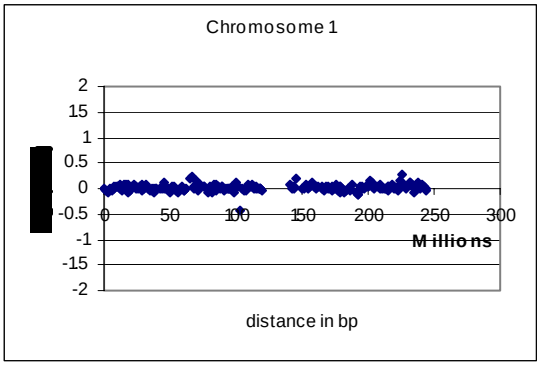

Chromosome 21
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
5
10
15
20
25
30
35
40
45
50
Millions
distance in bp

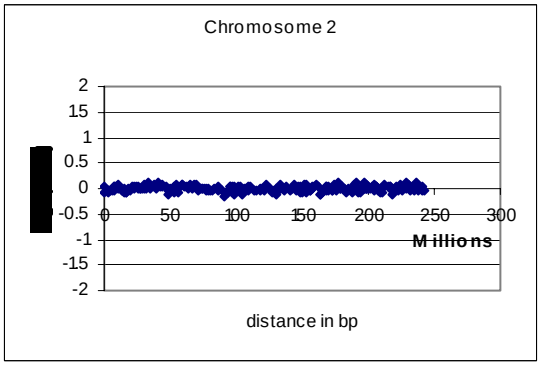

Chromosome 22
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
Millions
distance in bp

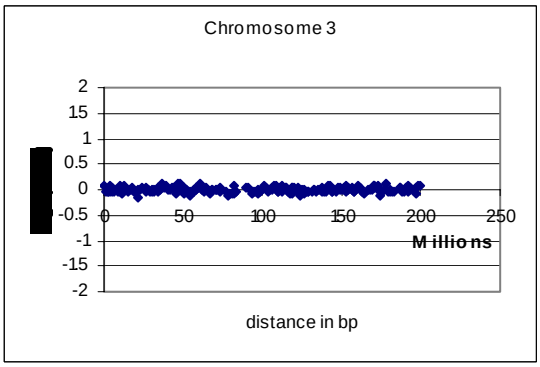

Chromosome X
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
Millions
distance in bp

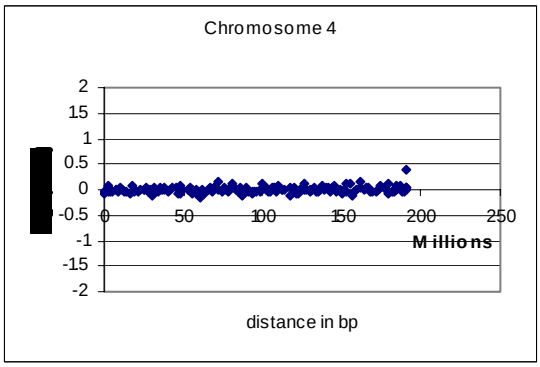

Chromosome Y
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
5
10
15
20
25
30
Millions
distance in bp

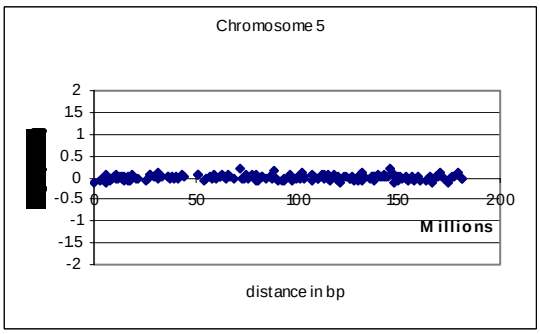
Patient 6:

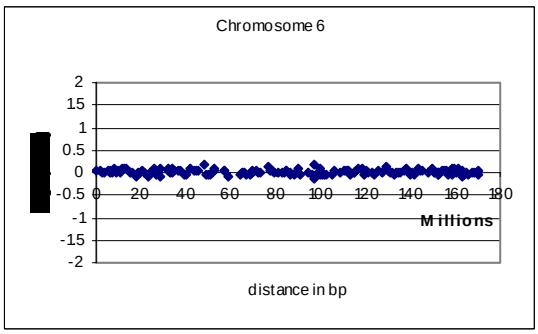

Chromosome 1
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
50
100
150
200
250
300
Millions
distance in bp

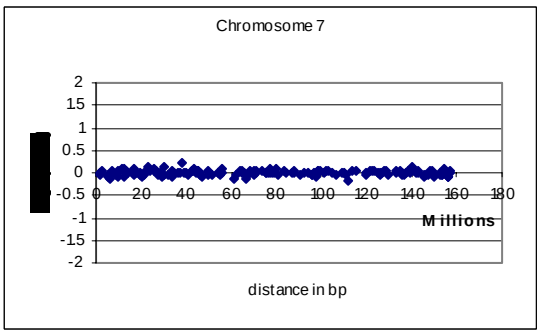

Chromosome 2
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
50
100
150
200
250
300
Millions
distance in bp

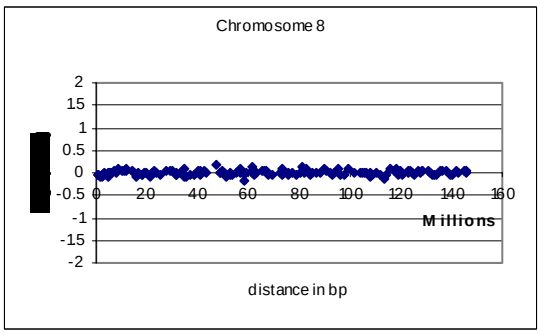

Chromosome 3
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
50
100
150
200
250
Millions
distance in bp

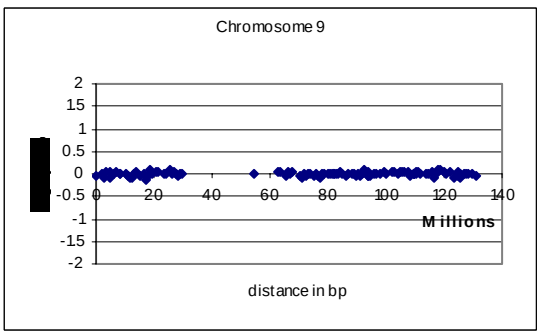

Chromosome 4
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
50
100
150
200
250
Millions
distance in bp

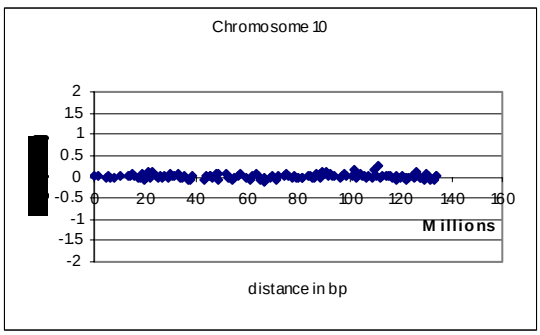

Chromosome 5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
50
100
150
200
Millions
distance in bp

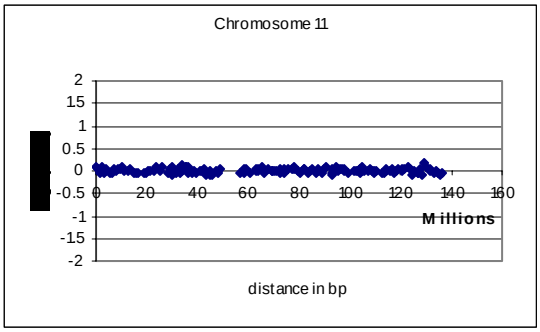

Chromosome 6
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
180
Millions
distance in bp

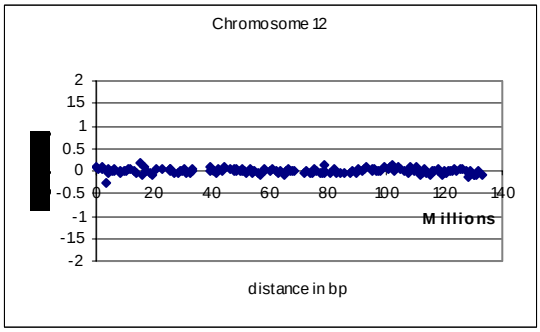

Chromosome 7
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
180
Millions
distance in bp

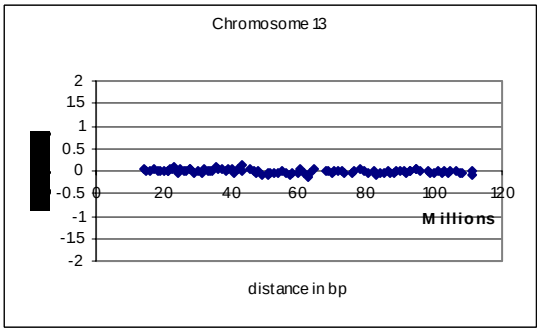

Chromosome 8
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
Millions
distance in bp

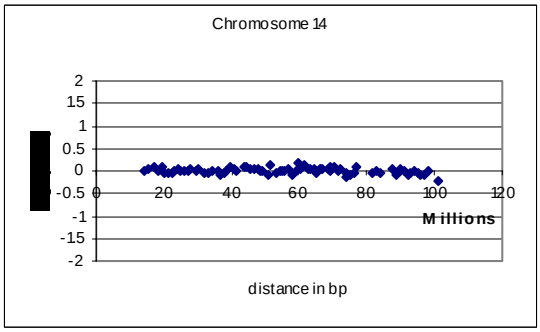

Chromosome 9
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
Millions
distance in bp

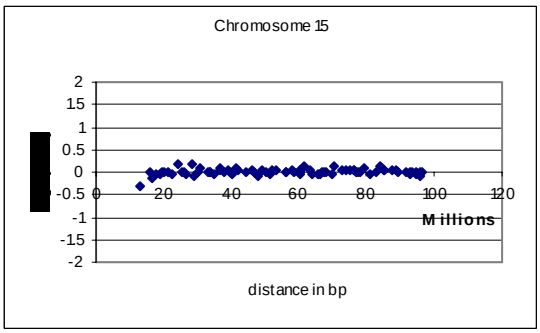

Chromosome 10
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
Millions
distance in bp

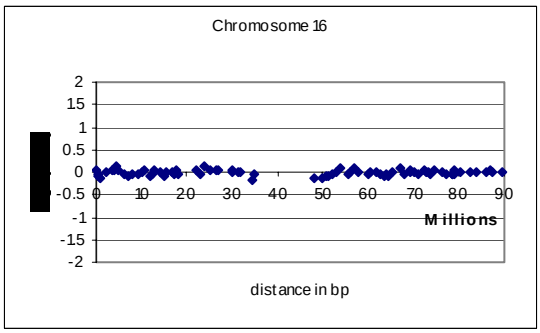

Chromosome 11
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
Millions
distance in bp

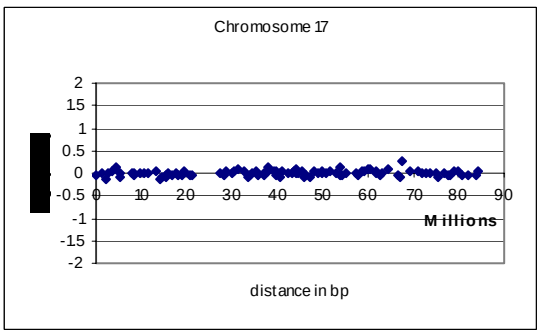

Chromosome 12
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
Millions
distance in bp

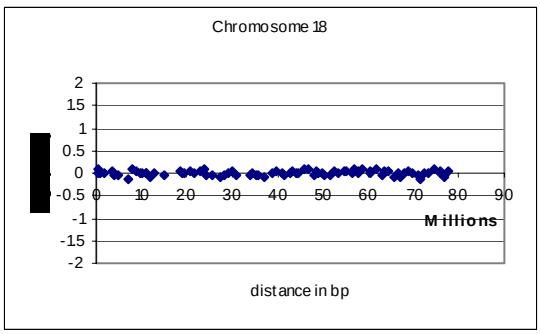

Chromosome 13
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

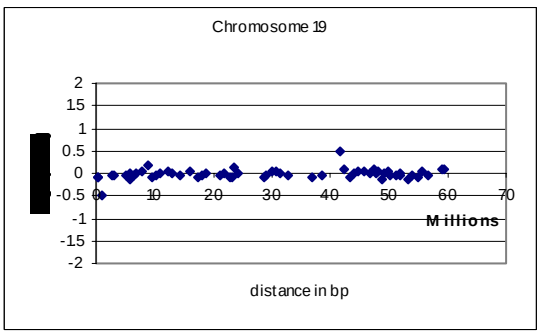

Chromosome 14
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

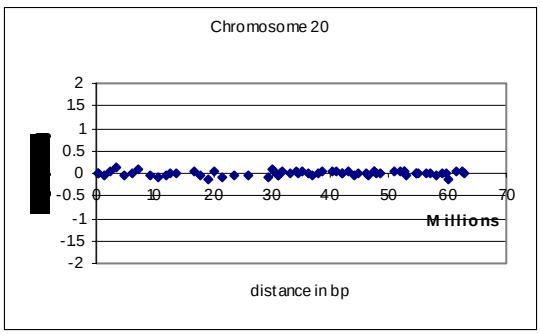

Chromosome 15
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

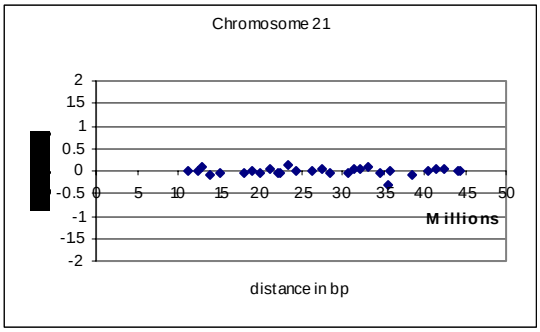

Chromosome 16
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp

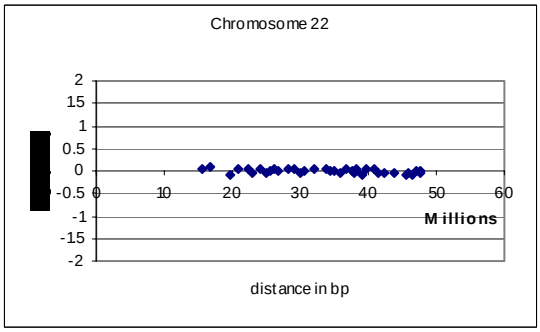

Chromosome 17
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp

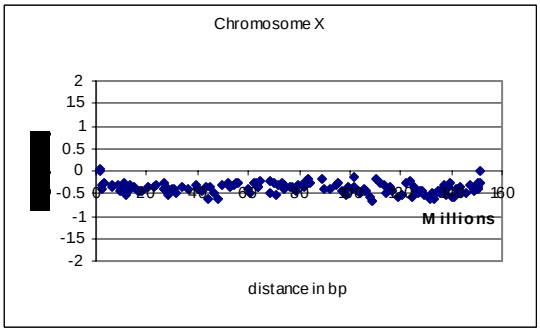

Chromosome 18
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp

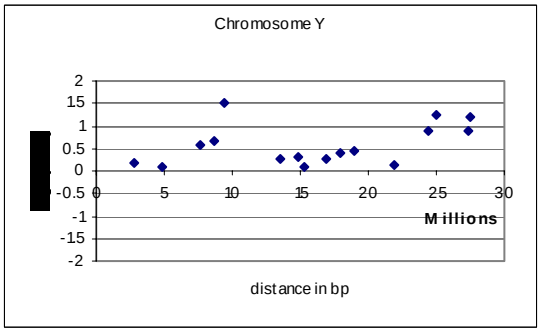

Chromosome 19
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
Millions
distance in bp


Chromosome 20
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
Millions
distance in bp


Chromosome 21
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
5
10
15
20
25
30
35
40
45
50
Millions
distance in bp


Chromosome 22
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
Millions
distance in bp


Chromosome X
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
Millions
distance in bp


Chromosome Y
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
5
10
15
20
25
30
Millions
distance in bp

Appendix 11: Clones known to report an incorrect copy number change on the 1Mb resolution array

Autosomes reporting an unexpected linear ratio

Clone	Chr	Position (bp)
bA326G21	1	143533568
bA5K23	1	159201779.5
dJ1108M17	1	104815713
dJ97P20	1	167462445
bA32C20	2	128082133
bA400O18	2	184873490
963K6	4	191632378
bA94E2	5	18186914.5
dJ159G19	6	80412976
dJ93N13	6	32596730.5
bA17M8	8	136548513.5
bA350F16	8	46413667
bB445N5	10	38414815
bA13E1	10	48233085
221K18	12	131110572
bA25J23	13	78143302
bA279F15	13	55656472
820M16	14	104124908
bA2F9	15	18507931
bA161M6	16	1085464
dJ843B9	17	43741286
bA220N20	17	44390565.5
bA416K7	17	45289524.5
bA294I20	19	86263
bA50L23	22	19845427.5

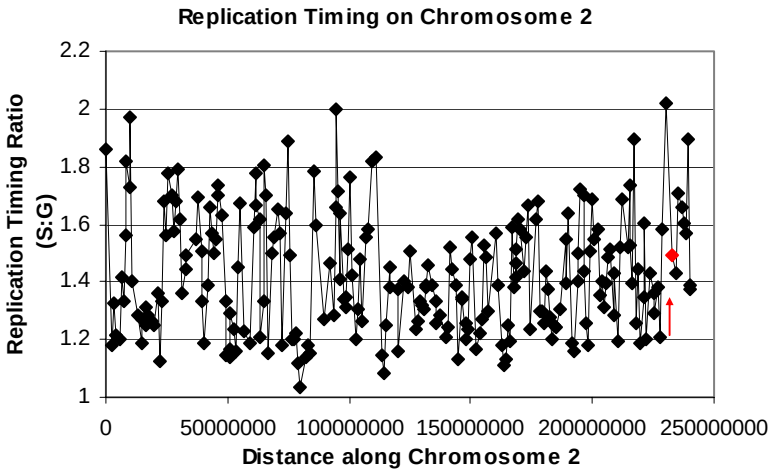
X Clones reporting an unexpected linear ratio

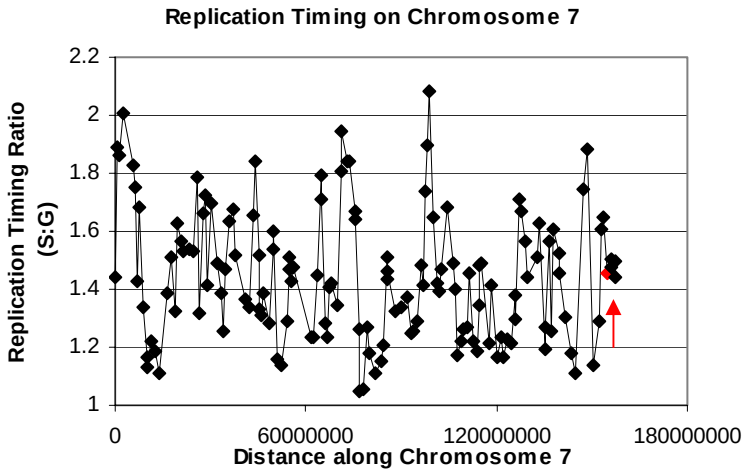
Clone	Chr	Position (bp)
98C4	X	490000
bA155F12	X	1834539.5
bA457M7	X	2000055.5
bA418N20	X	2326367.5
bA483M24	X	5959670
bA323F16	X	6434519
bA431J24	X	15946832
bA2J15	X	16774468
bA268G12	X	25943688.5
bA163L4	X	39575156.5
bA56H2	X	49574586.5

bB188A5	X	53052560
bA445O16	X	53863026
dJ966K21	X	54681885
dJ323B6	X	60874153.5
bB130F17	X	61742036.5
dJ583H20	X	66927306.5
bB260P4	X	68630583
bA236O12	X	70675479
dJ411B6	X	71885166.5
dJ875J14	X	72373814
dJ93L7	X	80445116.5
dJ225D2	X	80937973.5
bB166C10	X	85629051
bA122L9	X	86374786.5
bA156J23	X	87618715.5
dJ421I20	X	98011402
dJ312P4	X	100360142.5
dJ290B4	X	108330281
dJ93I3	X	109797538
bA434C1	X	111084399
dJ428A13	X	121210464
bA218L14	X	148668646
225F6	X	149149818

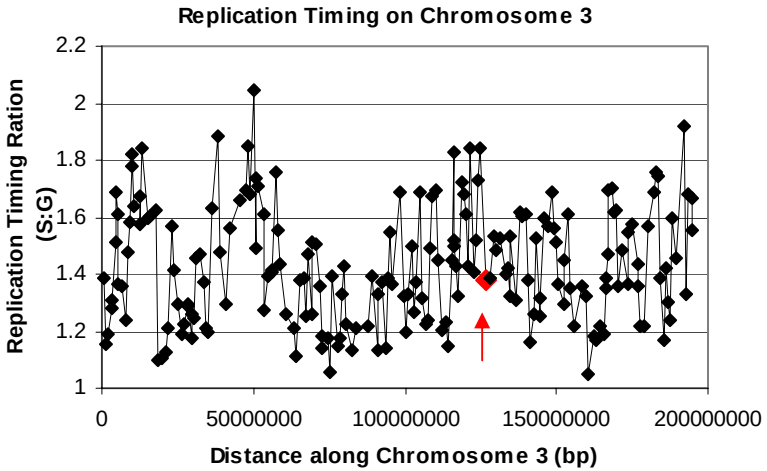
Appendix 12: Position of Chromosomal Breakpoints on the Replication Timing Profiles Location of breakpoints are indicated by red clones and red arrows.

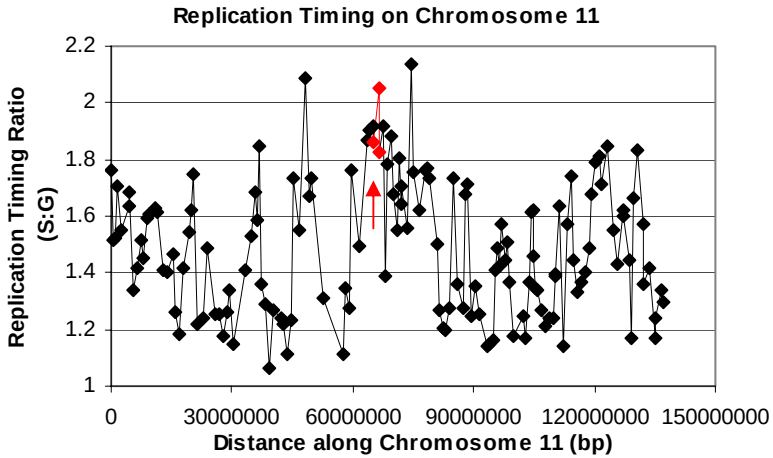
t(2:7)a



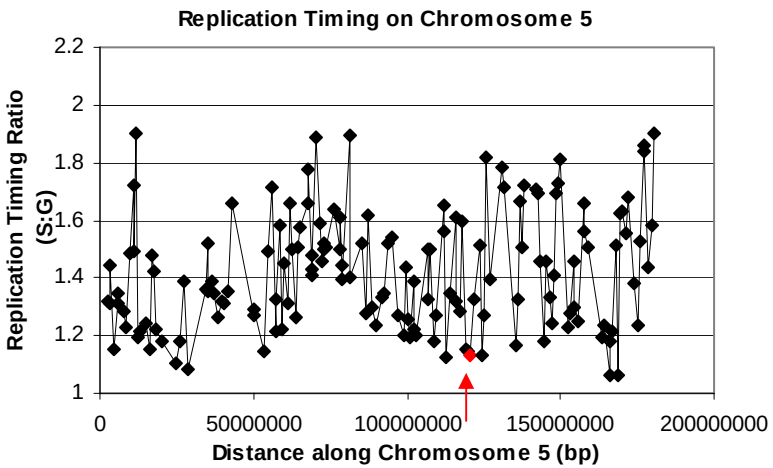


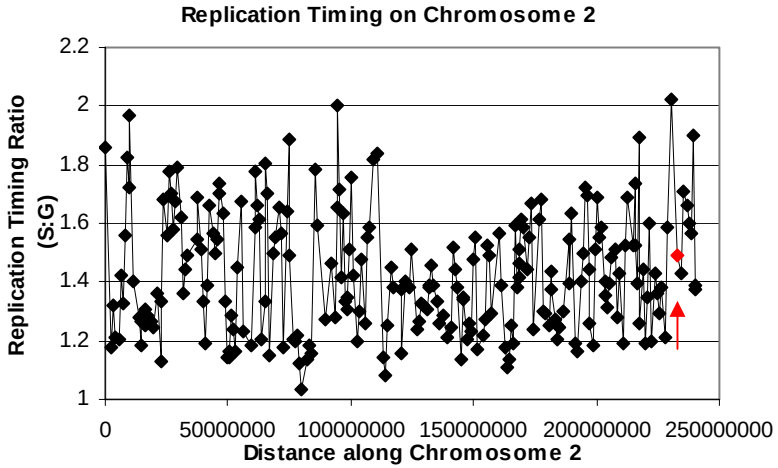
(3:11)



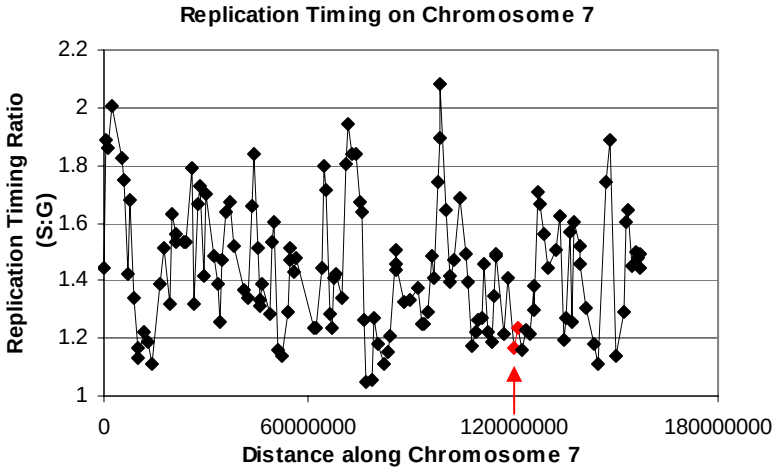


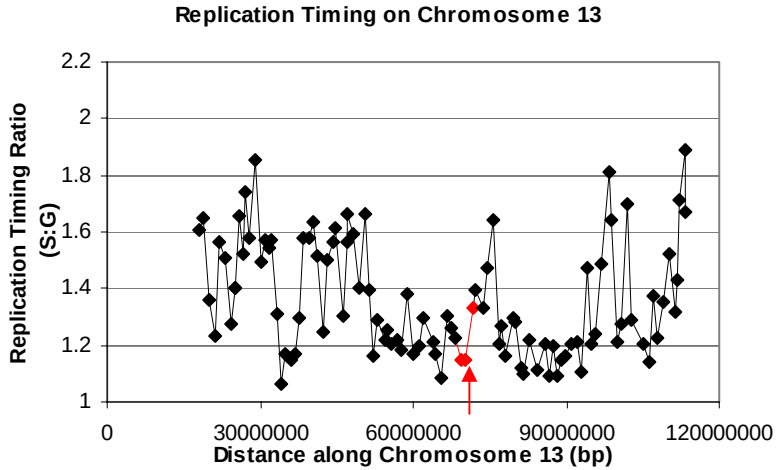
t(2:5)


Replication Timing on Chromosome 5
Replication Timing Ratio (S:G)
2.2
2
1.8
1.6
1.4
1.2
1
0
50000000
100000000
150000000
200000000
Distance along Chromosome 5 (bp)

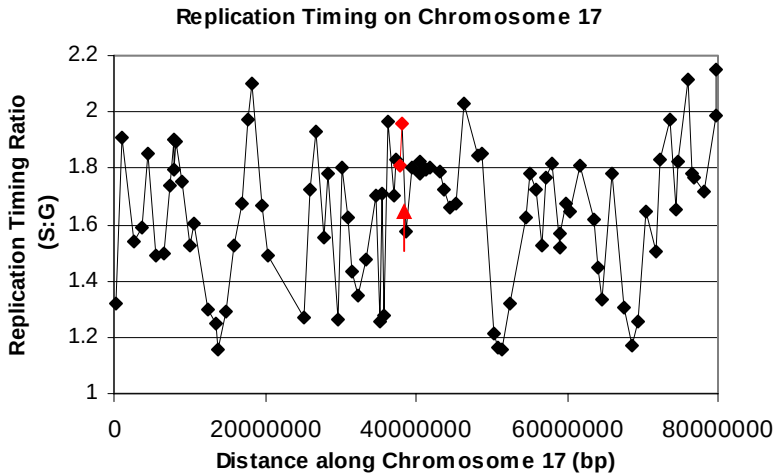

Replication Timing on Chromosome 2
Replication Timing Ratio (S:G)
2.2
2
1.8
1.6
1.4
1.2
1
0
50000000
100000000
150000000
200000000
250000000
Distance along Chromosome 2

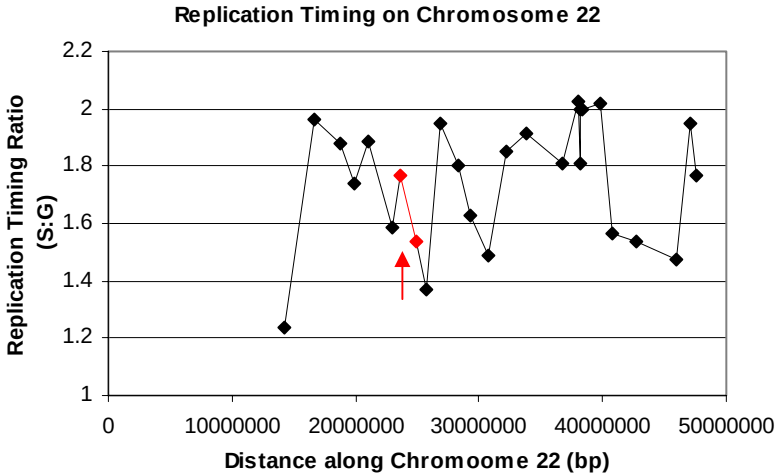
t(7:13)


Replication Timing on Chromosome 7
Replication Timing Ratio (S:G)
2.2
2
1.8
1.6
1.4
1.2
1
0
60000000
120000000
180000000
Distance along Chromosome 7

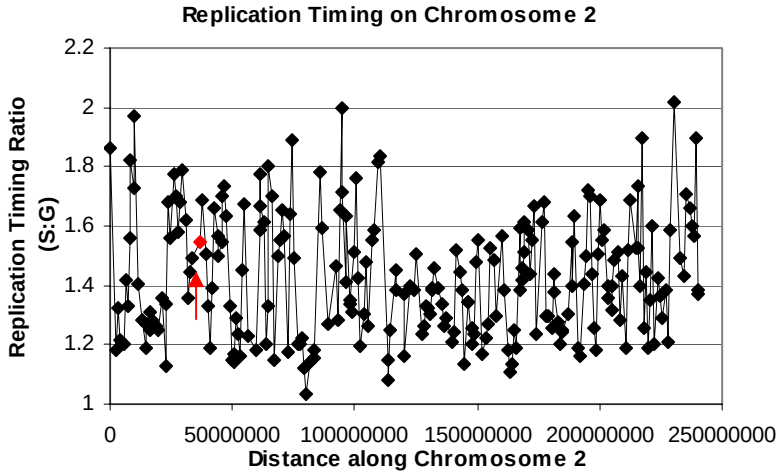

Replication Timing on Chromosome 13
Replication Timing Ratio (S:G)
2.2
2
1.8
1.6
1.4
1.2
1
0
30000000
60000000
90000000
120000000
Distance along Chromosome 13 (bp)

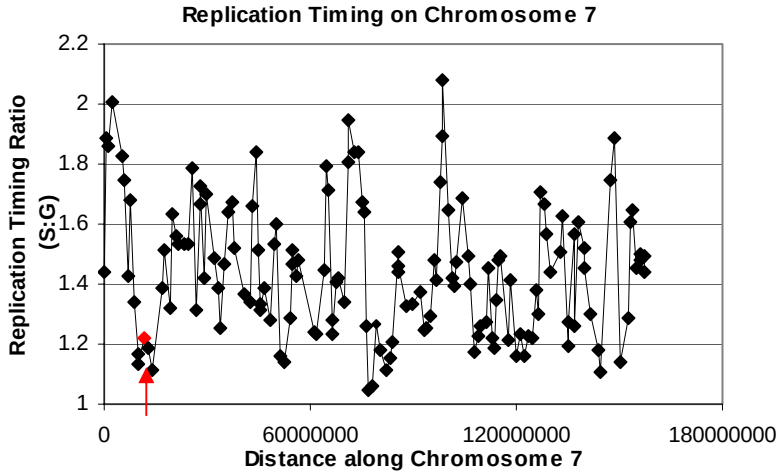
t(17:22)


Replication Timing on Chromosome 17
Replication Timing Ratio (S:G)
2.2
2
1.8
1.6
1.4
1.2
1
0
20000000
40000000
60000000
80000000
Distance along Chromosome 17 (bp)

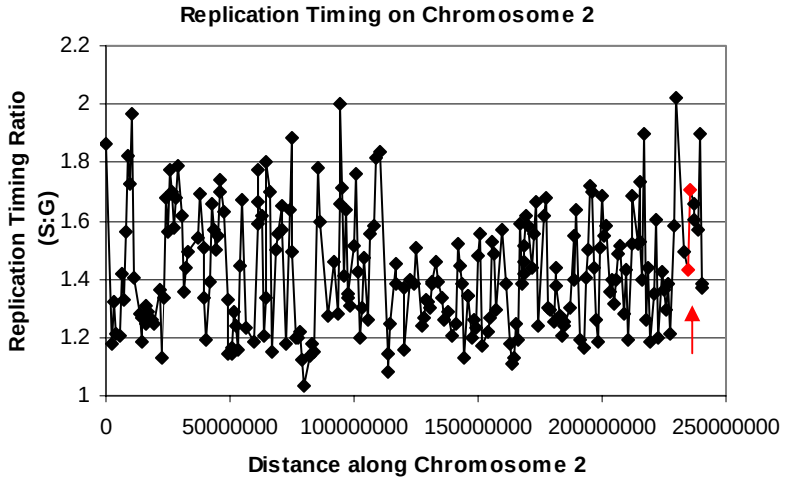

Replication Timing on Chromosome 22
Replication Timing Ratio (S:G)
2.2
2
1.8
1.6
1.4
1.2
1
0
10000000
20000000
30000000
40000000
50000000
Distance along Chromoome 22 (bp)

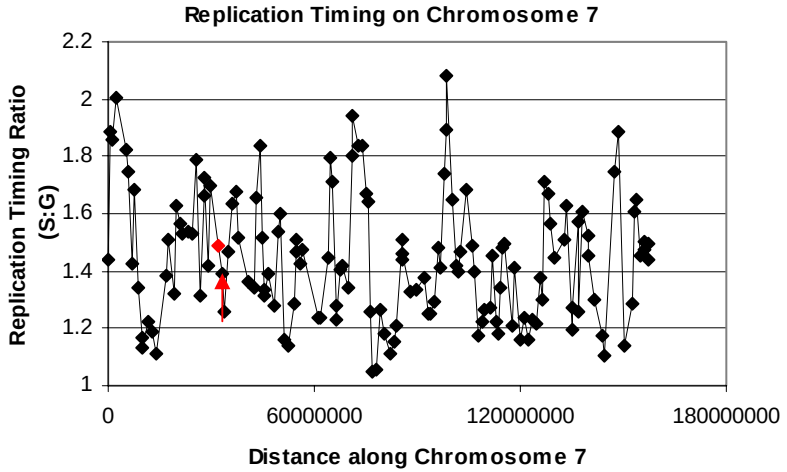
t(2:7)b


Replication Timing on Chromosome 2
Replication Timing Ratio (S:G)
2.2
2
1.8
1.6
1.4
1.2
1
0
50000000
100000000
150000000
200000000
250000000
Distance along Chromosome 2

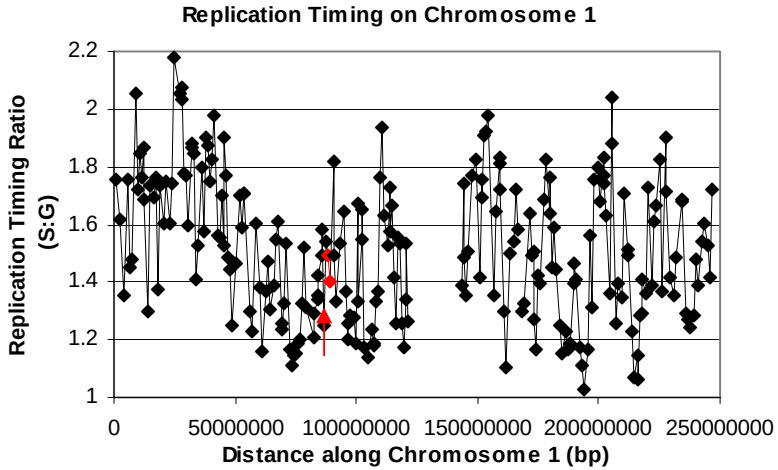

Replication Timing on Chromosome 7
Replication Timing Ratio (S:G)
2.2
2
1.8
1.6
1.4
1.2
1
0
60000000
120000000
180000000
Distance along Chromosome 7

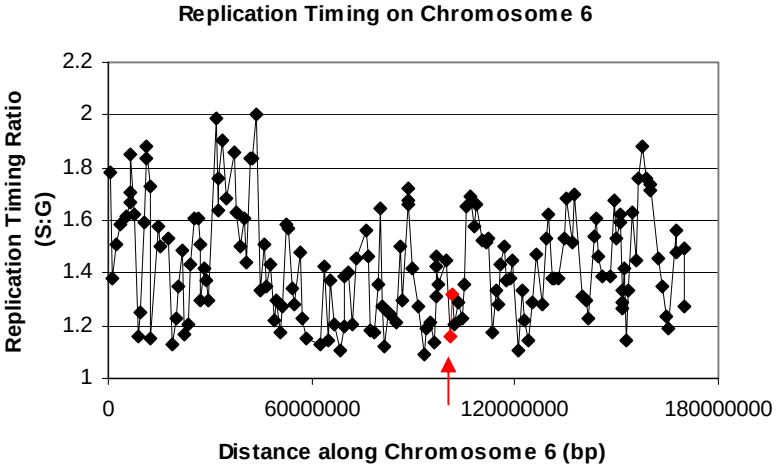
t(2:7)c


Replication Timing on Chromosome 2
Replication Timing Ratio (S:G)
2.2
2
1.8
1.6
1.4
1.2
1
0
50000000
100000000
150000000
200000000
250000000
Distance along Chromosome 2


Replication Timing on Chromosome 7
Replication Timing Ratio (S:G)
2.2
2
1.8
1.6
1.4
1.2
1
0
60000000
120000000
180000000
Distance along Chromosome 7

t(1:6)


Replication Timing on Chromosome 1
Replication Timing Ratio (S:G)
2.2
2
1.8
1.6
1.4
1.2
1
0
50000000
100000000
150000000
200000000
250000000
Distance along Chromosome 1 (bp)


Replication Timing on Chromosome 6
Replication Timing Ratio (S:G)
2.2
2
1.8
1.6
1.4
1.2
1
0
60000000
120000000
180000000
Distance along Chromosome 6 (bp)

Appendix 13: The significance of a correlation co-efficient.

Value of Coefficient (r)	**Meaning**
0.00-0.19	A very weak correlation
0.20-0.39	A weak correlation
0.40-0.69	A modest correlation
0.70-0.89	A strong correlation
0.90-1.00	A very strong correlation

Table from ‘Practical Statistics for Field Biologists’ (Fowler 1998).

Appendix 14: Publications arising from this work.

‘The Replication timing of the human genome.’ Woodfine et al. Human Molecular Genetics

Appendix 1: Reagents and buffers used.

Amino linking Buffer (10x)

500mM KCl,
25mM MgCl2,
50mM Tris/HCl pH 8.5
Made with autoclaved distilled water.

HindIII Digestion mix (for a 96 well plate)

Hind III (Boehringer 40U/ml),	55µl
Buffer B (Boehringer)	99µl
Sterilised water	286µl

Hybridisation Buffer

50% deionised formamide
2xSSC
10% dextran sulphate
0.1% SDS
10mM Tris pH 7.4
0.1% Tween 20

LB Agar

Tryptone	10g
Yeast Extract	5g
NaCl	10g
Agar	15g

Make up to 1 litre with autoclaved distilled water.

LB Broth

Tryptone	10g
Yeast Extract	5g
NaCl	10g

pH to 7.5 (using 1M NaOH)
Make up to 1 litre with autoclaved distilled water.
Autoclave at 121^{o}C for 15 minutes.

Orange G (10mls)

Orange G	0.1g
Ficoll	1.2g

Make up to 10ml with sterilised distilled water

Polyamine isolation buffer (PAB)

80mM KCl
20mM NaCl
2mM EDTA
0.5mM EGTA
15mM Tris
3mM dithiothreitol

0.25% (vol:vol) Triton X-100
pH adjusted to 7.2

Sheath Buffer

10mM Tris-HCl pH 8.0
1mM EDTA
100mM NaCl
0.5mM Sodium Azide

SSC (1x)

0.15M NaCl
0.015M Sodium Ctrate
pH 7.0

TAPS 2 Buffer (10x)

250mM TAPS pH 9.3,
166mM $(NH_4)_2SO_4$,
25m $MgCl_2$,
0.165% w/v Bovine serum albumin (Sigma),
0.7% v/v 2-mercaptoethanol
Made with autoclaved distilled water.

TBE Buffer (10x)

Tris Base	121g
Boric Acid	61.83g
EDTA	18.612g

pH 8.0
Make up to 1 litre with autoclaved distilled water.

TY Media (2x)

Bacto-tryptone	16g
Bacto-yeast Extract	10g
NaCl	5g

Make up to 1 litre with autoclaved distilled water.
Autoclave at 121^{o}C for 15 minutes.

Vista Green Stain (for 500ml – 1 gel)

1Ml Tris HCL	5ml
0.5M EDTA pH 7.4	0.5ml
Vistra Green	0.05ml

Make up to 500ml with sterilised distilled water.

Appendix 2: PCR primers for the High Resolution Array

2a: Primer sequence for PCR products in the high resolution array

STS Primer	**Forward primer**	**Reverse primer**
500bp overlapping PCR product array		
stSG494879	TGACCATGGACGGGAGAGAAAACATCCA	GAAAATGTGTGGCAGGTTCA
stSG494880	TGACCATGTGTCTCCCTTGGTGACATGA	CTCCCCACATGAGACCAGAT
stSG494881	TGACCATGAAGGCTAATGGGAAAGAGGC	TTCTGTCCCCTTTTGATTGC
stSG494882	TGACCATGCTAGGAAGAGGTTCCAGGGG	CTGAGCCTTCCTGTGTGGAT
stSG494883	TGACCATGGGAAACCATGCACCTCAGTT	GACCAGAAGGAAATGTTGGC
stSG494884	TGACCATGAAGACGGCTCTCAACCTTCA	GAAGACTCCAGCTGTGTCCC
stSG494885	TGACCATGCTCTTTGCTCGCAGTCATCA	CACAAGAGAAACACAGGCTCTC
stSG494886	No Unique Sequence	No Unique Sequence
stSG494887	TGACCATGGTCCCAACACCTCCATTTTG	CTGAACTTGGCCCATAAAACT
stSG494888	TGACCATGCGGACTCAAAAGAACAAGGC	CCTCTGAAACCGGCAGAATA
stSG494889	TGACCATGATCATTGAAGGTGCCAAGGA	TGTGCTTCAGCAAAACATCC
stSG494890	TGACCATGTACTCTTCAGTGGCCCGAAC	TATTGGCGGCCATCTACTTT
stSG494891	TGACCATGGTGCTAATTTCCACCACAGTCA	TGAAGGAAATGGAAAAGGGA
stSG494892	TGACCATGCCACTGCCTGCCAGTTAGAT	GTGCCGATCGAGACTCTTCT
stSG494893	TGACCATGGGCAAAATTCAAATCCTCCA	CTGATCTGCCTCCATCCATT
stSG494894	TGACCATGCCAGGTCACTGCCCTAAAAA	CCCAGGTCAGTTGTTTGTGA
stSG494895	TGACCATGTGAGGACTCCTGGGTTCAAG	TTCCAAACAGAGGCCTTCAT
stSG494896	TGACCATGGGTTTTCTGGACAGTTGACACA	GGAAAATGGACAAGCAGTTGA
stSG494897	TGACCATGGTGTCTTGGAGACTCCCTGG	TCCATAATTTCCGGGTTTCTA
stSG494898	TGACCATGCCTGTGGAAATCCCTCATGT	AGGACACAGGTTTGCTTTCA
stSG494899	TGACCATGGTGGCCTCTAACTCTGGCAT	CCCATACCTTTCTGAATCTGC
stSG494900	TGACCATGAATGACACCATCACCAGCAA	AGTTTCAATCACCGTGCCAT
stSG494901	TGACCATGCCCATCCTATGCCCTGTATG	GCAGCTGCAGTCAACTAACAGA
stSG494902	TGACCATGCATCTCCCAAGCTTTGCCTA	TGCACATGGTGAAATGAACA
stSG494903	TGACCATGTCCTCATGCCTCATGTCATC	TTTGGGAATACAGACAGGGG
stSG494904	TGACCATGTGGGGACAGAGGTAATCTGG	TTGCATGTGATCTGCACGTA
stSG494905	TGACCATGAAAGGTCACCCATTGCTTTT	TGGGATAAGTGAGGGTCTGC
stSG494906	TGACCATGGGAGGCTTTGGTTGTGTTTC	GTTGTTGGGGGAAGGAAAGT
stSG494907	TGACCATGAGGGTGTGACCCTGAGAGG	GCCACTGGCTGTTCAGATTA
stSG494908	TGACCATGGTGAAGGCTTGGCTGATACC	TGAAACATCTTCTGCCTCCA
stSG494909	TGACCATGGCAACTCTCCAAGTTCTGCC	GGATGGAGAAGGAAGTGCAG
stSG494910	TGACCATGTAATCTGGAAGGGCAGGAGA	CTCCCCTGAAGTGAGAGCTG
stSG494911	TGACCATGATGCCCTGACTCCAAAACTG	CCGCTGGAATTGTATCCTGT
stSG494912	TGACCATGACTCTGGAAGCCAAAAAGCA	CCAAACCGAAACAAAAAGGA
stSG494913	TGACCATGTTTTCCTTGGAACCCTTTATGA	GGTGTTTGTAAGGCAAGGAAA
stSG494914	TGACCATGCAAGTATGGCGCATCTCTCA	GGAAGTTCACGAGGGACAAA
stSG494915	TGACCATGACCCCATTCAGCTCACAAAA	ATCTGGCAGGATTTCTTGGA
stSG494916	TGACCATGAGGGGCTTGTGAAGACACAC	GGCTGGAATTCCGTCTCATA
stSG494917	No Unique Sequence	No Unique Sequence
stSG494918	TGACCATGGGAGCTCACCTTTTGGGTC	GCAGGAATAGAAGTGGGAGC
stSG494919	TGACCATGGGCCCTCCTAAGCTATTTGG	TGGGGTGTGATCACTGAGAA
stSG494920	TGACCATGGGTTCAATCTGTTGCCGTTT	GTGTTTGCATGGTTGAGCAC
stSG494921	No Unique Sequence	No Unique Sequence
stSG494922	TGACCATGGTGTACAGGGGAAGAGCGAG	GGGAAAGGAAAACTGAACCA

stSG494923	TGACCATGCTTCGTCTCTATGGTCCCCC	TAACCAACTGGAGGCAGAGG
stSG494924	TGACCATGTGTCCATTTCCTTTAGTGCG	CCGTGAACAGTAACTCCCTAGC
stSG494925	TGACCATGACAGGGTGCAGTGTAGTCCC	GGCTCCCCACAACAAGTTT
stSG494926	TGACCATGACTTCTCCCATGTGTTGTTCC	AGGCAGGGGAGCCTATCTAA
stSG494927	TGACCATGATGGGTGCTGTTCTTGTTCC	TTGGAAAACTGCAAATCAGC
stSG494928	TGACCATGTGATACCCTTCTCCTGCTCC	TGAGCACCTGGGTACAGACA
stSG494929	TGACCATGCTCACTGGGCTGGCTCTATC	TGCTTTCTTACACAAGACCCA
stSG494930	TGACCATGATCAGGTGGGAATGATGCTC	AGAGGTTGCCCAAAACACAC
stSG494931	TGACCATGAAATGAGCAAACTTGGCAGC	TCACCTGGCCAAAACAATTT
stSG494932	TGACCATGTGACTGTCTCAGAGCTGAATGA	CCAAGCCAAGATTCCTTTGA
stSG494933	TGACCATGGGTGGGTGAGACTTGAGGAA	AATTCCATGTCCCCACCATA
stSG494934	TGACCATGTGCTTCCTCCTCCTGTGACT	GTGAGCTACACCTTTGGCCT
stSG494935	TGACCATGTACCCATCAAGCCTACCTGG	TTCTCCCTTTCCTCAGTCCC
stSG494936	TGACCATGAGGCCTTTGAATAGCAAGCA	GCTGGACTATTGGCTTCTGC
stSG494937	TGACCATGTGAGAAAAACCCACTCAGGG	GGTTTCAACCCAGGAAGACA
stSG494938	TGACCATGGAGGTTAGGCTCAAGGGGAC	ACCTGTGTTGGGCTCTTGAC
stSG494939	TGACCATGTGAGTGCTTCCTGTGTCCTG	GTTTGTTAGCGTATGGGCGT
stSG494940	TGACCATGGCTTTGCTTGCTACTTGGCT	CCATCTCGTTTCCAGGACTC
stSG494941	TGACCATGTTTGTTGAGCACTGTCTGGC	TCTCTTCCACATGGACCCTC
stSG494942	TGACCATGCTCAAATCACACCACACACG	TGGTGTCAGCTGAGAAGAGC
stSG494943	TGACCATGTGGCAGCATATTCGAGTGAG	GCTCACAGCCTCTCTGCTTT
stSG494944	TGACCATGTGATCCCCAACTAGAGAAAAGG	AGCAAATGTTATTTCCCCTCC
stSG494945	TGACCATGTGTCAGCCGATCAGTCAGTC	TGGGCTCCACATATTTCCTC
stSG494946	TGACCATGGGTGAATTCTCCACCAGTCC	CTCCCTAGCTGTGCCAGAAC
stSG494947	TGACCATGTTCCTGCCTGGCTAACTGAT	GCATAGAGAAGGGACTAGAGGG
stSG494948	TGACCATGACCCGTCAAATCCTCAGATG	CGGGTACTGGGACTTTACCA
stSG494949	TGACCATGTGAGCAACGGCATAGAGATG	CCGGCCCACAATTTTAATAGA
stSG494950	TGACCATGGACACACCAGGCATCAGAGA	TGCCATGGATGGTGAGACTA
stSG494951	TGACCATGTCTGGCTTCCAGTCCTTGTT	GAGGCAAGCAGATTTTGGAG
stSG494952	TGACCATGACTTTTGGAACTTGGCATGG	CCTTTGCACTCAATGCTTCA
stSG494953	TGACCATGTACATAGGGATCTGGGCTGG	AAATCCTGTGGCTCCTTGTG
stSG494954	TGACCATGCCTGCCAGCTTCTGACTTCT	AACAGATTTCCTCCCATTGC
stSG494955	TGACCATGGGCTGACCTACTGGAGCAAA	TCAAGAGGAATTGACCTGAACA
stSG494956	TGACCATGCTAAGTTTCTCCCCGCTCCT	GCCTAAGGCCAGATTGATGA
stSG494957	TGACCATGGTCTCTGGCTCTTTGTGGCT	CCATTCTACCCAGGCATCTG
stSG494958	TGACCATGTTGACAGTAGCTGCAGGTGG	TTGGTGAGGAGGGAGATGAC
stSG494959	TGACCATGTTGGGTAGGCTGATCAGAGG	TTCTGAAGACCCTGGAATGG
stSG494960	TGACCATGAAACCCACCTTCCAAAGTCC	GTCTCCAAGAGAGGAGCGGT
stSG494961	TGACCATGGAGAGGCTAACGGACATGCT	GGCCACAGTCCTGTTCATTT
stSG494962	TGACCATGGAAACTGAGGTGTTGCGGAT	AGGGGCATCAGTTCAACATC
stSG494963	TGACCATGGCACATCTTCAGTGGGACCT	CAGGAAATACCTGAGGCCAA
stSG494964	TGACCATGGTGATTGGGGATGTGTGTGA	AGGAAAGCCATTATTTGGGG
stSG494965	TGACCATGTAGGACATGGAAGACCGGAG	GACAAAGCGGATGAAACCAT
stSG494966	TGACCATGGACGTCATCACGAAGATCTGA	CTTTCAGCATGAACCAAGCC
stSG494967	TGACCATGGTGTTTGCTTCATTGGCCTT	TTGGAACCCTCTTCCTCCTT
stSG494968	TGACCATGCAGGTCCCACATCAGGACTT	TCCAGGGAGAGGAAGACAGA
stSG494969	TGACCATGTTAAGGACCACACCCTGGAG	AGGGGACAAGTGACATCCTG
stSG494970	TGACCATGGGCCTCCTACCACACTCACT	GCAATTTCCTTAGAATGACCCA
stSG494971	TGACCATGGAGCTCCGGAGACTGACAAC	TGTGCACCTCCTTTATGGAA
stSG494972	TGACCATGCAACCTGCCACAAGACCTG	GAATTGCCTCGCCCATACT
stSG494973	TGACCATGCCCAGAAAAACCTGGGATATG	GAGTCGCCACCGTAACATTT
stSG494974	TGACCATGAATATGCACAGGGGAGAACG	AAATTGGACTAGTGGCCCAG

stSG494975	TGACCATGCAACATCAGCTTCCGTGAGA	CCCAGCAGACTAGGGAGATG
stSG494976	TGACCATGGCTGTGGAAGGAAAGACCAA	GATGGAGAAACAATGGGTGC
stSG494977	TGACCATGTCTGATGTCAAAGCAACCTGA	CACTCATCAGCCTAGATGCAA
stSG494978	No Unique Sequence	No Unique Sequence
stSG494979	TGACCATGGGCAGGAGTGGAGGTGATTA	CACAGGGCAGGTACCAAGTT
stSG494980	TGACCATGCATGCTCTGCTTTCCTTTCC	AGCAGCTCATGCTAATGCAG
stSG494981	TGACCATGGGCGCCGCATAATCTAAATA	TGGGAGATTTTCCAAGATGG
stSG494982	TGACCATGTTCAACAGAGCCGTGAACAG	GCCATTTGTGTAGCATTAGCC
stSG494983	TGACCATGAAAACAATAACGGACCGATCA	CACAAGCAATGGCCTTAACA
stSG494984	TGACCATGCAGGCACTACTGATGCTCCA	AATGCAGGAACACACATCCA
stSG494985	TGACCATGAATTTCAAACTGAGCAGGGG	ATGGCCAGCCGTTTACATAC
stSG494986	TGACCATGTTCTTAGGAAGGTTTTGAGCCT	ACAATAACCCCTGCAGTCCA
stSG494987	TGACCATGGGGCATTGAGTTTTTGGATG	GCCTCACCCAAACTGGTTAT
stSG494988	TGACCATGAATGGGCTGTACCTCATGCT	CTGCCTCCCTTGTCCATAAA
stSG494989	TGACCATGCGAGAGATACAGAGCCCAGG	TACGAAATTGGGGTTTCCAA
stSG494990	TGACCATGAGAGCACTTGCTGTAGGTCCA	GTAGGGCTCTAGACCTGGGC
stSG494991	TGACCATGACCAGGCCAACACTGGTACT	GGATGGGAGGTAAGCACTCA
stSG494992	TGACCATGGCAAAGTGAGAGAGAGATGTCC	CCATCCTTTCATTCCTTTAACC
stSG494993	TGACCATGCCCCATCTCCACCTACACAT	GGGAAAGTTTCTGGCTAATGC
stSG494994	TGACCATGCACTAAGGGAAGCACAGGGA	CTGCTTTTCAGTTTGGCCTC
stSG494995	TGACCATGTGTCCCTATCCCTCCCTCTT	GGCAGGCTCAGATCTGTAATG
stSG494996	TGACCATGGGATGACTTAGTAGGGGCCA	GGTGAGCACCACACCTCTCT
stSG494997	TGACCATGTCTTCCATGAGGGAATTTGG	CAAAATGGCATGGAGATACAAA
stSG494998	TGACCATGTGAGTGCCAAAGAATGGTGA	ATTGAGATGAATTGGCAGGC
stSG494999	TGACCATGAGAGTAAGGGTGGTGGGCTT	CCTTCAAGCTGGCTTTTGAC
stSG495000	TGACCATGAAGTGAGGAGTAGGGCTGGA	CGAATCAGGGGAAACTGAAG
stSG495001	TGACCATGGAATCCCCACGGTAGAGACA	TTAGCCATTCAGAGGGTTGG
stSG495002	TGACCATGCACCTTTCTTGCTCTGGAGG	AGAGCACTTGTTCCTGGCAT
stSG495003	TGACCATGGCAATGAAGGAATGAACCAAA	TGCCAATTACTGATCAGGCT
stSG495004	TGACCATGGCTTGCCTATGGGTGTGTCT	GATGTGGAGGAATGTGGCTT
stSG495005	TGACCATGCTCCCACAACCCTTGACCTA	AGCCTACCTTCCCCTTGAGA
stSG495006	TGACCATGGGCACTAAACTGGCTCCCTA	GCCATCCTGCAAGAGAAGTC
stSG495007	TGACCATGTTTTCCTCCCAACCACTTGT	TGCTGGCCTATCCCAAATTA
stSG495008	TGACCATGTGACTTGTGGGAAACAGCAA	TGTGGACCAATGCAAACACT
stSG495009	TGACCATGTTGGGAAGAAGGAGGGTTTT	AGAGATCTTGCTCACCCCAA
stSG495010	TGACCATGCACTGAAAGACTGGGGCTCT	ATCTGTCACCATCCTCAGCC
stSG495011	TGACCATGGGCTGAAGTCTGCAAATCCT	TATTTGTTCCCTGCCTTTGG
stSG495012	TGACCATGAATCCCTGGGAAGCTAAACG	ATGAGGTCCCCCAATTTCTC
stSG495013	TGACCATGATGCCAGCATTGATGTGTGT	TTGGTTGCAGCATCAGTAGC
stSG495014	TGACCATGTTTTCATCATTGGCTCCACA	CCGGGTAAAACAGACTCCCT
stSG495015	TGACCATGCACCTAAAAACACACCCTCCC	CCCTGGAAAGTTCCCAATTT
stSG495016	TGACCATGTTGCAGCTGCTGACTCAATC	TCTCCTCCCTCACTTCACCA
stSG495017	TGACCATGGATGAGGGTGAAGACTGGGA	TCACTTTTTCCTGCAAGGCT
stSG495018	TGACCATGTATGGCCAGTGCTTCTAGGC	GTGGGAGGGCAGTTTCTGTA
stSG495019	TGACCATGCACATGCTCCAGTGCTGAGT	AATCAGATTGGTTGGCAAGG
stSG495020	TGACCATGGAAAGGGGAGGAAACAGTCC	CTGGGGGTTTTATTGAAGACA
stSG495021	TGACCATGCAGTGATGTCAAGGCCAGTG	CCCAAACAGAGGTTCACCAT
stSG495022	TGACCATGTGGGAGATGCAGAGTTGACA	GAGGAAAGGCACAGATTGGT
stSG495023	TGACCATGGCAGTTTCTGGTGGTGACCT	TCAAGTTCAATGCCTAGGGG
stSG495024	TGACCATGGGTGTGAGATCCCAAAAGGA	CAATCTCCGGGTGCAGTTAT
stSG495025	TGACCATGGGCTGGTGGAAACAACACTT	ACAGCCTAGTGCAGCCTCAT
stSG495026	TGACCATGATCCTCCCTCTCACCCTCAT	TAAGGCAGTCCTGGAGGAGA

stSG495027	TGACCATGTTAATGGCTCCTCACCCTTG	AAGGGATGGAAGAAAGGAGG
stSG495028	TGACCATGGATGAGGTAATGCGGCTCTG	GCTTACCGATGTCGGAGTTG
stSG495029	TGACCATGGAGGTCGCAAAATGGGTAGA	ATCCTTGACAAGTGTTGGGC
stSG495030	TGACCATGGAGAGATGCCAGCAGTGACA	TGCAGGAAGTATCCCTCCAC
stSG495031	TGACCATGCCACTTGATATGTGGGGGTC	TCAGCTCCTTGCCTCAATTT
stSG495032	TGACCATGCACTTCCAGCTGCTCTCCTT	TGGGAAGCTACGTGTGATTTC
stSG495033	TGACCATGTGGTCAGCAGAGAGCTGAGA	GGATAGAAGGGCACTGACCA
stSG495034	TGACCATGTGAGTGTGAGGCACCTGAAG	AAATCATGGCTTCCCAACTG
stSG495035	TGACCATGGAGACATCCACTCCTCCCTG	TCACATGTGGGATCTTGAGG
stSG495036	TGACCATGGCTGGTTCCCTAGGTTCTCC	AAGACCATCAGGCGTTTCAC
stSG495037	TGACCATGGCCCTGATGGATTTTTCTGA	TAAGACAGGAAAAGGGGGCT
stSG495038	TGACCATGGCTCACAGTCATCCTGCTTG	AGAGTTGGGGGTCTTCTGGT
stSG495039	TGACCATGAATAAAAGATGGCTGCACGG	ATATTGACCTCCATCGTGCC
stSG495040	TGACCATGTGGTGGAGTGGACAAAGATTC	GGTGGGGAACAAGGAAAAGT
stSG495041	TGACCATGCAAAGGAGTTAGGTCGCAGG	TCTGGCATGTTCTGGTCTTG
stSG495042	TGACCATGCCTCAAAGGCTCTGTCCTGA	CTAGCAACAAATGCGCAAAA
stSG495043	TGACCATGTTCCTCAGGCCATTCAGAGT	CAACGGGAGTCACCTCAAAT
stSG495044	TGACCATGGCTCTAAGGAGCATGGTTGG	CTGTTACCTGGGGGACTTCA
stSG495045	TGACCATGCAGAACTCACGGGTCACGTA	GTTCCAAAAGCATTGCAGGT
stSG495046	TGACCATGGACACATCCTCAGCCATCCT	ACTTGAGCCTCCAATCTTATCC
stSG495047	TGACCATGTGCCTGTGCTTTTTCCTACC	CTTGGGCAAAGTCTGAGGAG
stSG495048	TGACCATGTCCTAATCCAGACTGCCCTG	TTAGTGGTTGATGTCTGCCG
stSG495049	TGACCATGCATTTTCCAGCCACTCTGTG	TGGGCAGAAGTTACCTGAGAA
stSG495050	TGACCATGTCCAGTGATTGAACTTCCTGTG	GCCGTGTTGTGTTTACATGG
stSG495051	TGACCATGTCCCTCTGGAAAGCAGAGAA	GACCTGAGAAGGGCATGG
stSG495052	TGACCATGGTACTCCCTCTCTCCCCCTG	CCCCCACACTTTTATTTCCA
stSG495053	TGACCATGTGAGGCACACATGCCTACAT	GCTCACCAGGAGCTACAAGG
stSG495054	TGACCATGAAGGCCTCAGTGCTGTCAGT	GCCACCTTTTGTCAGCTCTC
stSG495055	TGACCATGAGAGGAGCCACAGGCTATGA	TCCCAATTTCTGATCCTTGC
stSG495056	TGACCATGTGCATGTGAAGACGTAGGGA	TAACTGGCAGAATTCCCAGC
stSG495057	TGACCATGGGGGACACAGGATGTAACCA	TGGGATGTCTCTGATCTGGTC
stSG495058	TGACCATGTGCAAGCCTCCTTTTCTCAT	CATCCTTTGGGACATGCTTT
stSG495059	TGACCATGTGGAAAGCAGAAACCCACTC	CAGGCCTTCCACTGTCTGTT
stSG495060	No Unique Sequence	No Unique Sequence
stSG495061	TGACCATGAGGTGAGAAAGCAACCATCG	CACAGAATCACAGTGGCACA
stSG495062	TGACCATGCTGAGGTTGTTCCAAGCCAT	AAAGACCAGAAGGAGCAGCA
stSG495063	No Unique Sequence	No Unique Sequence
stSG495064	TGACCATGGCTGGCTTTCTATTTCCCGT	TCCAATGTCAGACAGAGAAAGG
stSG495065	TGACCATGAGTCCACAAAGAAGGGAGCC	AGGATTTCCCTGGTGTCTCA
stSG495066	TGACCATGGGCAGGTTCAAAGGGTTTTT	TCACTCAAGTGTGAAGGGGA
stSG495067	TGACCATGTCCTAGTCCGTGGTTTCACC	GTCACTGCACTTGCCTTTCA
stSG495068	TGACCATGGGTCATCATCTGGGAGAGGT	TCATGTCAAAGCAGACCTAAGC
stSG495069	TGACCATGAATGGCAAGAGAACGACACC	GGCAATGACTCACCCACATT
stSG495070	TGACCATGGGGACCACCTGCTGAGTAAA	GGCTGGTTCCTATTTGGTGA
stSG495071	TGACCATGGAAGATTTGGAGGGGACCAT	TTTGCAGGCTGAGAGAAACA
stSG495072	TGACCATGTCTTAGCAGGTGGGAACCCT	CCCTCAAGACCCTGTGAGAA
stSG495073	TGACCATGTGCATCAGCCAGTGACTTTC	GCACTTCTTTACGAGCCAGG
stSG495074	TGACCATGCCAGAAGTTGAGGAGGGTGA	AGAAATCCTGCCCGTCTTTT
stSG495075	TGACCATGTTTTGCAATTACCTCTGCCC	CAGCCACCTTGCTTTCACTT
stSG495076	TGACCATGAGGTGCTCAGCCATCAGACT	CCAAGACTCAAACTCCAGGC
stSG495077	TGACCATGTGGCCATTTAGAAGTTCCCTT	ATCGTGACCATTGTGGGACT
stSG495078	TGACCATGTGTCTGCTTATTTGGGGCTT	TGCAGAGTCACTTGAGGTGG

stSG495079	TGACCATGAGCCAGCAGAGAGGTGAGAA	GGCTCCCAGAATGATACCAG
stSG495080	TGACCATGAATGAAGGGCTTCCAGTTCA	TTCCTGGAGCCATTTACCTG
stSG495081	TGACCATGGGGGATACTGGGGATTGTTT	AAGAAAAGAAAGAGCTGCACAA
stSG495082	TGACCATGACTGAGGGAAGGAACTGGGT	AGGAGAAAGGAGGGCAGAAC
stSG495083	TGACCATGAACACAGGGATGGGTTCTTG	GCCACCATATCATGCCTTCT
stSG495084	TGACCATGGCACTTACTCCCTGCTCTGG	GAAAGGGAAAGCAGAGGGTC
stSG495085	TGACCATGTGCTGCATGTGATTTTCAGG	TAGCACGGGAAGTTTCTTGG
stSG495086	TGACCATGGCCCTCTGTGAGGAAGAATG	CGAGCAGTGCTACAGAGACG
stSG495087	TGACCATGGTAATGACCCATGTCCCCCT	CTTTTTCCTTCCCCTTCTGG
stSG495088	TGACCATGCATCTCCATGGTTCAAGCCT	CCTCTTCATGGAGGGTGAAG
stSG495089	TGACCATGGCCTCTTCCCATCACAAATG	TGCATGTGATTTGCTTGTTTG
stSG495090	TGACCATGTCCCAGGTTGACCAATAAGG	GATGCAAAGCTGTGCTGTGT
stSG495091	TGACCATGGAGGAAGAGGCTGCCCTAGT	CCACGTCCACTTGAGGTCTT
stSG495092	TGACCATGTGTCTGAGTGCAGGATGTCTG	AGCAGCAGCTGAGTTTGAGA
stSG495093	TGACCATGCCATCACACACACAAGAGCC	GCAGCTGAGCGTTCTTTTCT
stSG495094	TGACCATGGGAGCCTGCCTATCCCTACT	CTTCTGGGGTTGAATTCATTG
stSG495095	TGACCATGCCACCCAAAATAATGCCAAT	TTGGATGGTCTTCCCATTGT
stSG495096	TGACCATGGAAACACCACCATTCACGG	TTCCAAGACTCCTGCTTTTGA
stSG495097	TGACCATGAAAAGACAATGCTCGACGCT	AGCCATAAGGCCACATCAAG
stSG495098	TGACCATGAGATCGCCTCTGTGTTTGCT	AGCTAACGTCCATGTCACCC
stSG495099	TGACCATGAAGGGTGATATTTCCCTGGC	GGAATCAAAGGAGGAAAGGC
stSG495100	TGACCATGAGTCTGCTCTGCCTGACTCC	GCCAGTGGGACATCTCATTT
stSG495101	TGACCATGAATATGTTGCACCGATGCTG	TTTGGTCCTCTTCATCCCTG
stSG495102	TGACCATGGCAAGCCGACTTTTTGACTT	CGTACAGTAGGGGCTCAACC
stSG495103	TGACCATGTGGACATGAACCTGTGCAAT	CACAGCTATTGTGGATGCGT
stSG495104	TGACCATGCCCTGGCCAATAATGGTATG	GCCAGGTCATGGAATAGGAA
stSG495105	TGACCATGGAGCATCTATGAGGCGGTGT	AGGACTGGGGGACTGAAAGT
stSG495106	TGACCATGCTGGCTTGTTTTCCATGCTT	TGACTGTGAAGGTGATGGGA
stSG495107	TGACCATGTCCACTTCCCTTCTGCTTTC	AGCTTTGGCCACAGAAAAGA
stSG495108	TGACCATGACGCTAGGGTGTGATGTGGT	TCCTGTCCTCTAACCCGATG
stSG495109	TGACCATGTGGGTGGATCATACAACAACA	AGGTCTTGGTTGTCACCTGG
stSG495110	TGACCATGTTGTCAGCGTTGGCATTAGT	CTGACCACCTTGGACCAAAT
stSG495111	TGACCATGCACCATCACTGCAGGCTAAT	TTCAGGTGTAGACAGGAAGGC
stSG495112	TGACCATGAAGAGGGCAAAGGGACTGAT	GGTCCTGTTCAGAACCTCCA
stSG495113	TGACCATGGAGCTCTTCCTATGCAAACTCC	TGCCTCAGTTTTTATTGCAGG
stSG495114	TGACCATGTGCCAGACAAATCAGCAAAG	GAACACTCTCTGGACCAGGC
stSG495115	TGACCATGTCCCTAACACAACATTTGGCT	AAACCCAGGGGTGTACATGA
stSG495116	TGACCATGACATTTGCAGGGGATGATGT	GACCCTGAATGTGCCTCTGT
stSG495117	TGACCATGAGAATAAGCACGGCAGAGGA	GCCTTTTGGCACAAAACATT
stSG495118	TGACCATGTGCTGGAAACAGAACAGTGC	TTACTTCATTGTGCCCTCCC
stSG495119	TGACCATGTGGCCAAATGATATGAGCAG	TGCAGCACACGTTAGCTCTT
stSG495120	TGACCATGGGCAAAAATTCCTGTTTCCA	TCAGGATGCACAGTCCAGAG
stSG495121	TGACCATGTGGGTAATTTGAAGAGCGTG	CTTGTTGTTTCCGTAAGCCC
stSG495122	TGACCATGTTGCAAACATTTCTGGTGGA	TCAGCCAAGGAGCAGTTCTT
stSG495123	TGACCATGGAAACGATTGCTACAGTTTCCA	TGCAAATCTTGATGGTAGCAG
stSG495124	TGACCATGTGTTCCCTTTCTTCCCTCCT	GAAACACAGCACGTGGTTCA
stSG495125	TGACCATGGCCAGGAGAAACTGTTCCC	GGGCAGTTTCTTGGTGTGAT
stSG495126	TGACCATGTCACACTGACGTGTTCCAGA	CTCCTCCCCAAGCTCTCTTT
stSG495127	TGACCATGTTGCTGCCTAAAGGGAAAAG	GGCCATAGTGCGTTCTGTTT
stSG495128	TGACCATGGCAAGAGTGACTGAAAACGGA	AAAAACACACAGGGAGGTAGG
stSG495129	TGACCATGCGTGGCTGAAGAGAATTTCC	GTCAGCCCATTTCCTGTGTT
stSG495130	TGACCATGTAGCCCCAATCAATGACTCC	CTCCCAAGGGCACACATAGT

stSG495131	TGACCATGGCACTCTCAAGCCACTCACA	ATTATGGGAGCCCAGGAAAG
stSG495132	TGACCATGGTTTTTGAGGGAGCTTTCCA	AGGGAGACCCACACTCACAC
stSG495133	TGACCATGGGCCTTACACTTTTCCAGCA	GGCATAGTCGCTTGGTGAAT
stSG495134	TGACCATGCGGCTAGCTGTTCCTCACTC	ACCTTCCCTGCCCTTTTCTA
stSG495135	TGACCATGCTCCTTCCCCTGCACATAAG	GTGCTGTGTGTGGAGTGACC
stSG495136	TGACCATGGTGTGGGAAGGCTGGTCTAA	GAGGGCTTTGCAGTGTTAGC
stSG495137	TGACCATGAGACAGGTGCAGGAAGGAGA	TGGCTTTTGAGAAGGCATTT
stSG495138	TGACCATGTCTTCCATAAAGACAATCCCCT	TTCTGCCTGTGACAAACCTG
stSG495139	TGACCATGCATGCTGGACAACAACCATC	CCTTTCTTCAGGAGTGGTGC
stSG495140	TGACCATGGAAGTAGGAAGTTTCCCCGC	GCCTCCTCTGGGTCTTCTCT
stSG495141	TGACCATGTGGGATACTAGCCGCAGACT	GCGAACAACCTTCAGAAAGC
stSG495142	TGACCATGGAAGGGTTATGGACCTCAGTG	CTCACCAAGGTCCTTCCAAA
stSG495143	TGACCATGTGTTTGGTAAATAGAGGCCAGC	GCTGAGTCCATGGATTTGCT
stSG495144	TGACCATGCTTCCTCTCTTCTGTCCCCA	TCTGGCTCCTGATTGAAGGT
stSG495145	TGACCATGAGCCCAAAGTGAGCATCAAC	CTCCCCCAATGCATAAAATAA
stSG495146	TGACCATGTCCCAACTTAATGTGCCGAT	TTCCATTTCACATGGCCTAA
stSG495147	TGACCATGTTCCTCCTATCAGCACCACC	TTATCCAGGTTGCAAAAGCC
stSG495148	TGACCATGCATGCTTAGGAGGGGTTGAA	GGGTAATTTCTGACAGGGCA
stSG495149	TGACCATGTCTCCTCTTTTCAATGGGACAT	CTGCCAGCATGATGACAAAT
stSG495150	TGACCATGAAGCTCTCTCTGCAACAGCC	GCAGATGACTTGGCAATGAA
stSG495151	No Unique Sequence	No Unique Sequence
stSG495152	TGACCATGCAAGACTATGTGATAGGCTGGC	GATGGTCAGACACACCAATCAG
stSG495153	TGACCATGGGGGAGGAGACAGGAGAATC	CACAGCCTGCCTTATCCATT
stSG495154	TGACCATGCAGCATTTCCAAAATCAAGC	GAGCACTGTAGGCCTTGTGG
stSG495155	TGACCATGTTGCCACCTGTTCTGTTTCA	CAGATATGGGAAGGTGGGTG
stSG495156	TGACCATGGGGGAAAGAAGCAAAGGAAA	CCAGCTTTATGAAGGTCAGCTC
stSG495157	TGACCATGTCAGATTTATGATGCGGACAA	GGATCACTGACTTTGGCTCTG
stSG495158	TGACCATGCAGAGTTCCATGTGGGTGAA	GCAGGAATTAATGAATGGTGG
stSG495159	TGACCATGAGAATGTGATGCTGGAGGCT	TCACCCCACATGGAACACTA
stSG495160	TGACCATGGGGCTGTGTCTTAGCCATGT	CACAACAACCAATGCCTCAC
stSG495161	TGACCATGTGAGGGCAGAAGCTATGACA	TATACACCAGGCAGGATGGG
stSG495162	TGACCATGGCAAAACCTAAGTGACTCCCTC	GAAGGAATGTTCCCTCTCTTTG
stSG495163	TGACCATGTCTGCAGAACTGTGCTCTTTG	ATCTTTCCACATCCTTCCCC
stSG495164	TGACCATGATTGACCCTGGCTCCTCTTC	TTCCCTTTGCTGTTTTTGCT
stSG495165	TGACCATGCCAATGCATGTACGAGGAAG	CCATGATGACATTCATCCCA
stSG495166	TGACCATGACCCATAAGTTTGGCATTGG	TTGGGAAAGCCATCTGGTAA
stSG495167	TGACCATGATGACCAACTGGGTCTGCTC	ATCACCTTCCAATCAGTGCC
stSG495168	TGACCATGCCAGCTTATCAACCCGATTC	CTGGTGTTTTGCCTTCCATT
stSG495169	TGACCATGGATGATTGTCCCAGGCCTTA	CCACTGTCTAAGGGCGTTCT
stSG495170	TGACCATGGGAAGGACAGGGGAAAAATC	CAATATCCCCTCCTGGATCA
stSG495171	TGACCATGTGAGGGGGAGTCATCAAAAT	ACAAAGGTGGCCACAGATTC
stSG495172	TGACCATGGGATATTGGGCCAGAAAAGC	TCAGATCATGCAGTGTCGGT
stSG495173	No Unique Sequence	No Unique Sequence
stSG495174	TGACCATGCTGAAAAGCATCCTTTGGGA	TCAACTGCAGGAGCACATTC
stSG495175	TGACCATGCAGGCAAAGACTTAGGACACAA	TGACATACAGAAGCATTCGCT
stSG495176	TGACCATGCAAGTGACCCCCAACTCTGT	CCAGCCCAAACAAGAAGTGT
stSG495177	TGACCATGTCAAGTCTTCCCCCTCATTG	CAATTCTGTCCTAAGGCCCA
stSG495178	TGACCATGGCAGAAGGCTTTGTCCTCAG	TGGGTCCCAGATAAGTGGTAA
stSG495179	TGACCATGCATGAATTGGTCAGTCGGAA	GCCACAAAAGGAAACCAAAA
stSG495180	TGACCATGTGTCTCTCTCTCTCGTTGCATT	TGGACATGTTGAATCAGGAAA
stSG495181	TGACCATGAGGTGGTGGCAAATAGTTCG	CAAGACTCAGGCACACATGG
stSG495182	TGACCATGTGTGCTGCTACAGATGTGCTT	TGTCATGGTCAGTCTCCAGC

stSG495183	TGACCATGTGACCAGGCAAAGAGGAAAG	CACAGTCAAATGAGGCAGGA
stSG495184	TGACCATGAGTCTGGGGAGTGAGTGCAT	CCTCACCACCCTCAAGTGAT
stSG495185	TGACCATGCCTGAGATTCTGGGAAGCAG	AATGCTGAAGACCCCTCTGA
stSG495186	TGACCATGTTCCTAAGGGAGCAAAACACA	TGCCTTCTCACTATTGCTGC
stSG495187	TGACCATGGCCTACAGGCTGAGTCCAAG	CTGCCCAACTTGACAAGAGA
stSG495188	TGACCATGAGAGCTCGTTGGGAAATCAA	GATGACTGGGTGGAAAAAGC
stSG495189	TGACCATGTGCAGGGGATCTAAGGACAC	TCACCCCATTTTCCAAATGTA
stSG495190	No Unique Sequence	No Unique Sequence
stSG495191	TGACCATGTTCAAGCAGGGAAGCTGTTT	AGGAGTTTTGGAGGCAAGGT
stSG495192	TGACCATGAGGCCACAATCTTTGCAACT	TGGAGTGTGGAGAACTGCTG
stSG495193	TGACCATGAAGAGGGCATAGGTCACTGG	AAGGAAGGAAGAGGAGGCAG
stSG495194	TGACCATGGCCTAATTTATTCCTTCCCCC	TTCGGTACTGTGTTTGCAATTT
stSG495195	TGACCATGCAGCAGCTTCAGTGCTATGC	CAAGATGCAAATCCTGCTCA
stSG495196	TGACCATGGCATTCCAGCGTGGATATTT	CAACAGGGGTCCCATCTCTA
stSG495197	TGACCATGTCTCTTTCCTCTGACAGCCC	CCTGTCCCTGTTCTGCTGAT
stSG495198	TGACCATGATGGCATAAGAACAGGTGCC	GAATGCTAAAGCAATGGGGA
stSG495199	TGACCATGCTTGTCCCCACAGGGAATTA	TTCCATAACTCCAGGTTGCC
stSG495200	TGACCATGACCGCCTGAGTGATGAAAAA	CCATGTGGCTTACTGGTTGA
stSG495201	TGACCATGGTGTGCATGTGTGTGTTTGG	TGGAAGCCAGAATTGTGTGA
stSG495202	No Unique Sequence	No Unique Sequence
stSG495203	No Unique Sequence	No Unique Sequence
stSG495204	TGACCATGGCCCTGATGTGCTTAGGGT	CTGGAAAACGCAAGAGAAGG
stSG495205	No Unique Sequence	No Unique Sequence
stSG495206	TGACCATGGGAGGCCTGAATGCTCTTTA	CGTGAAAACAAAAACTTCCTGA
stSG495207	TGACCATGCTCAGCCAGCACTCACTCTG	TCTGTGGAATCGGACATTCA
stSG495208	TGACCATGGTGATTGCACAGGTGGATTG	ACATTCATGTGCAGGTGAGG
stSG495209	TGACCATGCAACTCTAGGGGACTGCCTG	CCCAGGGCCTCAGTTTTATT
stSG495210	TGACCATGGCCATTTGCTCACAACAAAG	TTTTTAAACAACCCCCTCCA
stSG495211	No Unique Sequence	No Unique Sequence
stSG495212	TGACCATGCCAGCATCCTGTTTTGCTTT	TGAAAGGACCAGTAGTGCCC
stSG495213	TGACCATGCCCAACAAAGGCTACGGTAA	ATGGCAGCTATGTTGGACCT
stSG495214	No Unique Sequence	No Unique Sequence
stSG495215	TGACCATGCCTCCAGACCACAATGACAA	TTATTCTCCCCTGACCTCCA
stSG495216	No Unique Sequence	No Unique Sequence
stSG495217	TGACCATGGTAATGGCGGCTGATCTTTC	GCCAACATTTCAGGTTGCTT
stSG495218	TGACCATGTCACAGTCTACCGGAGGAGG	GCTTCATTGATTGGCTGGAT
stSG495219	TGACCATGAGCTTCAGTCACCTTCCCG	CACACCTGCTTGGGAATTTT
stSG495220	TGACCATGCCCAAAGAAGCCAAATTACTG	GGGGCTTCCTAAAGCAAACT
stSG495221	TGACCATGCACCACCCAAAGGCATAAAC	CCCAGAGCATCTCAGAGGAG
stSG495222	TGACCATGAACCGAGCATACCATTCCTG	CAAGGGGAAGATGCAGTGAT
stSG495223	TGACCATGTGTCCTTGGTTCTAGGCCAC	GGTTCTGTTCTGGCTCATCC
stSG495224	TGACCATGCACAGCCACCAATAAGTCCA	CAGCGGCACTGTTCTTGTAG
stSG495225	TGACCATGGTGGAAAAGAGCACAGCACA	ATAGCAAATGCACCTCGGTC
stSG495226	TGACCATGACAATATTGGGCAGCTGAGG	GGCCAGCTCTCTTACCCTTT
stSG495227	No Unique Sequence	No Unique Sequence
stSG495228	TGACCATGGGTTTGGCAGAGAGTCAAGG	AGTGTCGGAACAGAGGATGG
stSG495229	TGACCATGTAAACTCCCAGCCTCAGTCC	TTTGATGCCTTGCTTATATTCC
stSG495230	TGACCATGTGTCTTGGTTGTTATGCAGTGA	CCTCCTTGGATGTCGAGAAG
stSG495231	TGACCATGCAAACCGCAATAGAGAGCCT	TAAAGAGGGGCAGCTTCTGA
stSG495232	TGACCATGAGCAATGCAGAGTGATTCCC	GATCAAGATTCGCATCCCAT
stSG495233	TGACCATGTCATCAGTGAAAAGGACAAAGC	TGACCTCCTGATTGTGTGTCA
stSG495234	TGACCATGTTCTGCCACCTCCAGTCTCT	ACCACAAGGCCACACTTAGG

stSG495235	TGACCATGTGAGCAGCCCGTAACTTAGC	GTGCAAGGTCTTTTGCCCTA
stSG495236	TGACCATGCGATCAGCTCTGTTTCCCTC	CACCAACATTCTGTGCCATC
stSG495237	TGACCATGCCTCCCCTGGAGAAAGGTAG	CAGGCTTTCAGCACTCTGTG
stSG495238	TGACCATGTGATGGCTACAAAGGAGCTG	GCCACAACCTCATTGGAGAT
stSG495239	TGACCATGGACAATGTGCCAGTTTGTGG	AGAAGCCGTGCTCATCAGTT
stSG495240	TGACCATGGGACAGCTGAAGGATTAAGGTC	GCCATGATTCCAGCTTGC
stSG495241	TGACCATGGTTGAGTCTGGGAGCTCAGG	TTTCCATTTTGCCACGTGTA
stSG495242	TGACCATGGATCAGCCACCAACAGTGAC	TTATTCACCTGGTTGCGACA
stSG495243	TGACCATGATTGACGTACCACCTCCTGC	TGAGTCTGGGTGAGCTTCCT
stSG495244	TGACCATGCAAAAGGTGGTTGGTTGAGC	TCTCCAGCTCCCTCTTGAAA
stSG495245	TGACCATGCCAAGAGATCCTTCTGCCAA	CTCCATGAGAGCCTTCTGCT
stSG495246	TGACCATGCCCATTCCAAATCTCTACGG	GCTGGCCAAATTGTCAGAGT
stSG495247	TGACCATGAATTCAAGCACAACCTTGACC	CTCCCGTTATTCTGAGCAGC
stSG495248	TGACCATGGGAGTTAAGGGGAGGCAAAG	CCTTTTCTCCCAGAGGGAAT
stSG495249	TGACCATGGGGCTGTTACGGTTACTGGA	AGCTTCTCAGCAAGGAGGTG
stSG495250	TGACCATGCGTCCTCTGGATAAAGGCTG	CACTCTCTCTCTCCCGCTTC
stSG495251	TGACCATGGGTGCTATTGAGGTGGGGTA	CCTTACCTGGGCTTTCACAG
stSG495252	TGACCATGGGAAAACTTACGAAACCCTCAA	TCTTGCTTCAGTTTTTCCTGG
stSG495253	TGACCATGGGACACTATTTCTGCCGACC	GATGGATCCAAATGGCAGAT
stSG495254	TGACCATGCACCTTTCCCACATGCTCTT	CCATCGCCTCTCTCTTTTTG
stSG495255	TGACCATGGCTTTGGTGTCTTTGCCATT	CCAGGGTACCTCTCCCTTTC
stSG495256	TGACCATGGGTCAACTGTTTCCCTGCTG	AAAGGCTGAGCTCGATAGCA
stSG495257	TGACCATGAGATCCACCCCCACACACTA	TCCCTCCTGTTATTTGCTGG
stSG495258	TGACCATGACCAGATTGCTATCAGCCCT	GGGCCATTCGTCTTTAAGGT
stSG495259	TGACCATGAACAGCACGAACAAAAACCC	CCCCAGGACATATCAACCAC
stSG495260	TGACCATGTTGGCTGATTTCTCCCACTC	TGGTTCTCCTTCCAAGATGG
stSG495261	TGACCATGCTTGAGCTCTGGTCAGGTCC	TGCGACCACTTTGCAGTAAG
stSG495262	TGACCATGCAGCATGTGCTTGGCAGTAG	GAAATTTCCGTTCCCTGTGA
stSG495263	TGACCATGGATTCATCGACCCCTAGTCCT	TTTCTCTGCTCATGTGGTCAA
stSG495264	TGACCATGGTGGGAGAGTTAAGGGGAGC	AGCATGGCATGTTGTGGTTA
stSG495265	TGACCATGCCCAAGATGTCTGCTGTTCC	TCTGAGACAGATGCCAGGTG
stSG495266	TGACCATGAAGCCCTGCAGATACAAGGA	TTCAGCGAGTAGGTTACTTCCA
stSG495267	TGACCATGGCTCTAAAACCAATTCACCACA	CAGGGTATCTGACAGCCCAT
stSG495268	TGACCATGCGCCTTAGGCACTTTCATTT	CATCCCTGCACAACTCTCAA
stSG495269	TGACCATGAGCTCCATGCTGTGAGGTTT	GGACTGGTTTATGCTGGGAA
stSG495270	TGACCATGAGGCTTTGTATGGGGGAGAT	GCACAAGACTGGCATCTCAA
stSG495271	TGACCATGAAGACGCTCAATGTCCCAGT	CTAGGGTTAGGAAGGGCAGG
stSG495272	No Unique Sequence	No Unique Sequence
stSG495273	TGACCATGCCATTGCCAATACAATGCCT	TTTTCCCACACACCTGGTAA
stSG495274	TGACCATGGTCTCCCTATTTTCACCTAGCC	AGTGATCCAAGGGCTGACTG
stSG495275	TGACCATGGCTGGTTTGGGAGAAGAAGA	TTCCATTGCCACTGAGATGA
stSG495276	TGACCATGGGCCAAACAAGCTGTGAAAT	ATGAAAACCAAACCAGCCAC
stSG495277	TGACCATGTGGGGAACTACAGGGGTTTT	CCCCTATACTTTCTGAGGGCA
stSG495278	TGACCATGGCTCTGTGTTGTAATCGCCA	TCCCCAAGGTAAGGCTCTTT
stSG495279	TGACCATGAGGTCCATGCCGTAGTTGTC	TGCAGAGGTTTGTGACTTGC
stSG495280	TGACCATGCTAGGCTTCTTCAGCCCTCC	GCCAGCTGCTCTATCTGTCC
stSG495281	TGACCATGCCTGCCCACCTTGTATCATT	CACTTATGCCACAAGCCTCA
stSG495282	TGACCATGGGATTCATCGGAACCAACAC	GGGAATCATGTTCTCACGGT
stSG495283	TGACCATGTGCTGTGGGACTACACAAGG	GAGAGGGTGGGAGGCACT
stSG495284	TGACCATGTCATGATGGTCAAGAGCCAA	ACCTCACTCTGCCCATTCAC
stSG495285	TGACCATGGATTGGCAGAGGCATTGTTT	AGCAGCTGGAACTCTGAGGA
stSG495286	TGACCATGGGCATCTGTCCCTCACTAGC	GCAAATCTCATGACCCCCTA

stSG495287	TGACCATGCCTTTTTAGGCCTTTGGTCC	GTCTCCCTACCCCACCAAAT
stSG495288	TGACCATGCTGGCTCCTGCTTCTAGGTG	TGAAAGCCACTTGCTGTGAG
stSG495289	TGACCATGGAGGTCAAATCACTGGGGAA	GGGCTGAAACAAACCTGAAA
stSG495290	TGACCATGCTCTCCTCCTTTCTCCCCAC	AGATTTTCCCCAGCGTAGGT
stSG495291	TGACCATGTCTTTGCCAGCAGTCTCTGA	AGAGGCACAGAGGGAACTGA
stSG495292	TGACCATGCCACCCACCTGTTTACTTGC	AATAGATGTGTGGCGGGAAG
stSG495293	TGACCATGGTGGATTCTCAGGGATGCAG	GAGGGAGAAAGAGGTGGGAC
stSG495294	TGACCATGGCAGGTAAATGGTGGCTGTT	CTGCCAGCTAGCAACTGATG
stSG495295	TGACCATGACCGAGCTTAGCGGTATGAC	AGTCTTCATGGCCACACTCC
stSG495296	TGACCATGCTGCATGCTTCACAGGTAGC	CCTTGAGAATGCACCACAGA
stSG495297	TGACCATGTCTCCGATTCTCTGCGTCTT	CTCTCCTCAGCCTCAACCAC
stSG495298	TGACCATGTACCGAGGTGTCTGTGTCCA	AGACCTTCTCCCTCAGCCTC
stSG495299	TGACCATGCCCTTAGCTCGGTACCAACA	AACTGTGGAAATGGCTTTGG
stSG495300	TGACCATGTTGTTACTGCCTTTGGGGTC	ACTGCTTTATTCTGGTGGGG
stSG495301	TGACCATGACAAGGAGGAGGGTATGGCT	AGATCTGTCTGGCTGCAGAG
stSG495302	TGACCATGCAGCCCATAGTTGTCCCATT	AGTGACACTCGTGCTCATGC
stSG495303	TGACCATGTATTCTTCCCCCAGGATTGC	GGAATGTGAGGCATGGATTT
stSG495304	TGACCATGAAAAGAACCAGAGCGTGGGT	AGGCAGGACCAAAAATTCCT
stSG495305	TGACCATGTTAGTGCACACCAGGGACAC	TAATTTGGCAAGCCATCCAT
stSG495306	TGACCATGTGCCTGGACGTGCTTATGTA	GTCACTGTGGTCAAGAGGCA
stSG495307	TGACCATGTGCCAACTAACCTTGCCTCT	GAGAGAGCGAGAGAGGGGAT
stSG495308	TGACCATGCAAAATCCCTCCACACCTGA	AAATGGCTTCCTTCCTTGGT
stSG495309	No Unique Sequence	No Unique Sequence
stSG495310	TGACCATGCTGCATGCACAGTCAGTCCT	GCAAGAAAAGCACAGGAAGG
stSG495311	TGACCATGTAAAACATCACAGCCACCCA	CAAGAGAGGCTGGGAATGAG
stSG495312	No Unique Sequence	No Unique Sequence
stSG495313	TGACCATGACAAGACCAAGGGCTTCTCA	GCCCCAAAGCAATTTTCATA
stSG495314	TGACCATGGGGTTCCTTAGCTGTGGTGA	ACTGTTTGCCATTGTCCCTC
stSG495315	TGACCATGCACTGAATGACAGTGAGCCC	TGTATGGAGCAGATGCTTTGA
stSG495316	TGACCATGTTGAGCCTGAGCAAGTGTTG	TCTCACAGAAACACCATCGC
stSG495317	TGACCATGCTCGGTATAAACTCGCCTGC	ATGTGGCACCATCTCAAACA
stSG495318	TGACCATGGCAGAACAAACTCAAGGCAA	TTGTTAAGTCAACCAACCCAA
stSG495319	TGACCATGGTGGGAGGCAGTGGTATCAT	TGAGCAATTAGGGTTCTGCAT
stSG495320	TGACCATGCAAGAATACAGGGGGCTGAA	TGGAAAGTGGTATTCCTGCC
stSG495321	TGACCATGATCCAGGCCAATCAGCTATG	ATGGGGATAGTGACCACCAA
stSG495322	TGACCATGCCCACATGCTTAGCTACTGAAA	TGAGGTGATATTCTGTGTGCAT
stSG495323	TGACCATGAGAGTCACAAAAGGGCCAGA	TTTATTTTTCTGTGATGGGCA
stSG495324	TGACCATGAGGCTCATTATGATGGGTGG	TTCACATTGTACACCAAAAGGC
stSG495325	TGACCATGTGCCATCTACAAACACACCC	CCCTAGTTCTGTCTTCCGCA
stSG495326	TGACCATGTCTCTCTGTTTCCAGGCTGC	TTCTGAAATTTCCATCCCTGA
stSG495327	TGACCATGCAAAAGGGTCCAAACCTCAA	AAGTGCGCTCTTGGAAGAAA
stSG495328	TGACCATGTGGCTGGTTTTCGCTTACTA	TTAGGTAAAGCAAATGTGGTCG
stSG495329	TGACCATGTTTTTCCCCATGACCTGAAA	TTTTGACCTTTGCCCTGAAC
stSG495330	TGACCATGTTGGAAGGGATAGCACAAGG	GTTGTTTTCTGCCTCCATCC
stSG495331	TGACCATGTTCCTTCCCCGTTGTAAGAT	GCAAGGTGGGACTGAAGGT
stSG495332	TGACCATGTTGGGGGAGCAGTTTCTCTA	GTTGTGAGCTGGTTGCTTCA
stSG495333	TGACCATGGGCACTCAGGCAACTGTCTA	CCTAGAGGGCCTTTTTGCTT
stSG495334	TGACCATGGCAGGACTTCTAGGGCCTTT	TGGCAGACATTGCAGTTAGC
stSG495335	TGACCATGTGCCGAGACGTGTGTTACTC	TATGGACAAAGTGGGGGAAA
stSG495336	TGACCATGTGGGAGCACAGTTTATGCAA	GATGGCTCTTAGGGGTTTCC
10Kb Resolution Array		
stSG495337	TGACCATGTGCAAGTGCACATACACACAC	CTGTGGTTATGGGGTGCTTT

stSG495338	TGACCATGATAACCACAGCAGAAGGCGT	TGGACCATAGCCTTTGTGAA
stSG495339	TGACCATGTGCACATGGAGAGAAGTGCT	CGATCTATTGGGCTTCCTGA
stSG495340	TGACCATGGCCCAATAGATCGCTGAAAA	ATTCTTCCTGGGTCAGGGTC
stSG495341	TGACCATGGACCCTGACCCAGGAAGAAT	TTCTGAATGCATTTGGCTAGAA
stSG495342	TGACCATGGGTGCACTGATCCAGGAATAA	ACAATGACCTTTGGACAGCC
stSG495343	TGACCATGGGCTGTCCAAAGGTCATTGT	TCTGGAATTTGAAGCTGGGT
stSG495344	TGACCATGACCCAGCTTCAAATTCCAGA	TGGGAGCAGCTCGTAGTCTT
stSG495345	TGACCATGGCACTGGACAGAATGCACTG	CCTTCTGTGTGCTGCCTCTA
stSG495346	TGACCATGAAGATGACTTCATCACCATTGC	TATGGCACTTGCTTCATTGC
stSG495347	TGACCATGGCAATGAAGCAAGTGCCATA	TGACTCAGCCTTTGATGCAC
stSG495348	TGACCATGCTGAGTCAGAACACCTGCCA	CACTGAGGGAAGAAAAAGCG
stSG495349	TGACCATGCGCTTTTTCTTCCCTCAGTG	GCAGGGCCTGTTTGATTTTA
stSG495350	TGACCATGAATTCCAGTCCCAGCAAATG	CTGCTGGAGGATCAAGAAGG
stSG495351	TGACCATGGAAGCCCTTCTTGATCCTCC	GGGTGTAAACTCCAAGCCAA
stSG495352	TGACCATGGAGGGGAAAAAGATAACGCC	GGGGAATGGAAAGAGGAGAG
stSG495353	TGACCATGCAACAACTGCAACTTTGGGA	GTTGCACAGACATCAAGGCT
stSG495354	TGACCATGAAGCCTTGATGTCTGTGCAA	TTCTGACCAGGTGGGTAAGC
stSG495355	TGACCATGGCTTACCCACCTGGTCAGAA	TCCTTTCCAGGGTTTCCTTT
stSG495356	TGACCATGCCATACGTGGAAACTTGGCT	CACAAAGTAGGGCCAGGGTA
stSG495357	TGACCATGTACCCTGGCCCTACTTTGTG	TGTTTGGGGAGGTGGAAATA
stSG495358	TGACCATGGAAAACGATTCGATCCCTGA	TCCTCTGTCCAATCCACACA
stSG495359	TGACCATGTGTGTGGATTGGACAGAGGA	CGGGGAGAATGAAAAGATCA
stSG495360	TGACCATGATCTTTTCATTCTCCCCGGT	ATGCTGTGGTGGCACATTTA
stSG495361	TGACCATGTAAATGTGCCACCACAGCAT	CTTTCGAAGCACCCATGATT
stSG495362	TGACCATGGTACCATGCCTGGTTTGGAT	AGGTCCCAACTGCAAAAGTG
stSG495363	TGACCATGTATTGCCATCCCCAGAACAC	TGTGTGTGTGACAGGCTGAG
stSG495364	TGACCATGCTCAGCCTGTCACACACACA	TCCCAGCCTTGTGACTTTCT
stSG495365	TGACCATGGAAAGTCACAAGGCTGGGAA	AAGCACTGGAAGAACGAGGA
stSG495366	TGACCATGCTACCTCAATGTTGCTGCGA	CTGGTTACAAATCCAAGCCC
stSG495367	TGACCATGTTTGTAACCAGGGTGGGTTC	GATGTGGCTTTCTGATGGGT
stSG495368	TGACCATGTCGTTGCAGATAAGGGAAGC	GGAGGACTCCCTTCTACTTGG
stSG495369	TGACCATGAGAGCACTTGCTGTAGGTCCA	GTAGGGCTCTAGACCTGGGC
stSG495370	TGACCATGCAACAGGTTACCCAGGTGCT	AGCCTACGCTGGTTTCTTCA
stSG495371	TGACCATGTCTCTAGCATGTCCTTGGCA	AGAATGCCACAGTGGGGTAG
stSG495372	TGACCATGCTACCCCACTGTGGCATTCT	CTCTTTGGAGGCAACCTCAG
stSG495373	TGACCATGCTGAGGTTGCCTCCAAAGAG	CAAGGGGAGAAAATGGTGAA
stSG495374	TGACCATGAATAGAGGTGTTGGGCTTGG	TGGTTGCCATTCAATAGCTC
stSG495375	TGACCATGTGAATGGTCCTGGAGGTAGG	AAGTGGGTGGAGATCAGTGG
stSG495376	TGACCATGCCTCTGCAAAGTCTTGCCTC	TCTGATGCCTTAGCTGGGTC
stSG495377	TGACCATGGCAACAGACAACAAGCCAGA	AGGAAAGGAAAGGCTGGAAA
stSG495378	TGACCATGTTTCCAGCCTTTCCTTTCCT	GGCTGATGAAACTCTGAACCA
stSG495379	TGACCATGTGGTTCAGAGTTTCATCAGCC	GCATTTACCAAGCCCTTCAA
stSG495380	TGACCATGGGGCTTGGTAAATGCGAGTA	TGCATCTGGACAAGGAAGTG
stSG495381	TGACCATGATCTGGCCCATTACAGGAGA	ACACTAAGGGGGAATGTTGG
stSG495382	TGACCATGACCAGGCCAACACTGGTACT	GGATGGGAGGTAAGCACTCA
stSG495383	TGACCATGTGAGTGCTTACCTCCCATCC	TGGACTTCGTAAGTCTGGGG
stSG495384	TGACCATGCAATCCCTAGCCCGTAGGTT	CTCTCCCGGGTAAGTGCTC
stSG495385	TGACCATGTAAGATAGCCACTGCTGCCA	GGGGAGTTGTCAAGTTTCCA
stSG495386	TGACCATGGGAAACTTGACAACTCCCCA	CACCGAAGAGGTGAGGAGAG
stSG495387	TGACCATGAGGCTTTCTGACTCCTTGACC	TAGCTGGTACACGTTGGCAC
stSG495388	TGACCATGCTTCTCCTGAAGAGGGCAGA	CATCAGAATCCTCAGCCTGG
stSG495389	TGACCATGCCACTAAATTGGAGAGGCCA	AATGGTGGTCAGGAAAGCAG

stSG495390	TGACCATGAGATGGAGATGAAAAGCAAGAA	TGGATGCAGATGGATGAAAG
stSG495391	TGACCATGGCAAAGTGAGAGAGAGATGTCC	CCATCCTTTCATTCCTTTAACC
stSG495392	TGACCATGAAACAGTTTGATTGCTGGGG	TTGCAAGTGAAGTCAAAGCG
stSG495393	TGACCATGGTCACTGAATGCCAGCTGAA	GTGCTCTCCAACTAGCCCAG
stSG495394	TGACCATGTGTGCTTGGCATAGTGGGTA	CTCAAGGACCCCAAACTTGA
stSG495395	TGACCATGATGCACACTCATTGGCACTC	ATAGGGCCTGGAAGTGTCCT
stSG495396	TGACCATGTTCTTGTTGGTTGTTTGCCC	GGCAACCACAGTAAAATGCC
stSG495397	TGACCATGTAGGCGATGCCTGAGTATCC	AAAGTGTTGCCAGTCTTGCC
stSG495398	TGACCATGGGCAAGACTGGCAACACTTT	CTTAAGCCTGAGCCCAACTG
stSG495399	TGACCATGGGGAAGTGTGTTGAATGGCT	TCACAATGGCCTCAATGGTA
stSG495400	TGACCATGTACCATTGAGGCCATTGTGA	TCGTGAAACACTGGAACTGC
stSG495401	TGACCATGAGTGTTTCACGAAAGGGGTG	GTTACCCTTTTGACAGCCCA
stSG495402	TGACCATGGATGGCCACATGCATTATAAGA	CCAAGTCAAGCACTCTTTGC
stSG495403	TGACCATGACCTGTTTCTCCCCATCTCC	GGGAAAGTTTCTGGCTAATGC
stSG495404	TGACCATGGCATTAGCCAGAAACTTTCCC	CCTGCTACTCCAACCCTTTG
stSG495405	TGACCATGCCTTATACTGCCTTTGGGGC	CCCAGATTTAAGCGGTTCTG
stSG495406	TGACCATGGCTCAAGAATTTTTGGCCTG	TGCCTCTGGCTCTATCTTCA
stSG495407	TGACCATGAACCTTGGAGGGGTCTGTTC	TCAATGATTTATGCCTCCACA
stSG495408	TGACCATGATCCAGAGCTTGGAGGAACA	TGATGTGCTGATGTGCAAAG
stSG495409	TGACCATGTTTGCACATCAGCACATCAG	GAGGGAATGGACTGCAAAAA
stSG495410	TGACCATGTCCATTCCCTCTCCACAATC	AACATGGGCTGTCTCCTGAC
stSG495411	TGACCATGACCTAGTGGGCCTGAGGTCT	GTGATGAGCCCAGGGTAAAA
stSG495412	TGACCATGGTGCCTTATGGGTTCTCCAG	CAGCAGTGATAGCTCACCCA
stSG495413	TGACCATGAGATAGAGGGAGGGGTCAGG	TTTGGCCTCCAGCAAAATAC
stSG495414	TGACCATGGGTCTTCACCTGTCTGTGGG	ATCTGTTTCGATGGATGCCT
stSG495415	TGACCATGACAGGAGGGGCCAGAGTAGT	CAGCACAGCAGAAGATCAGC
stSG495416	TGACCATGGTCCTGGAAAGGGGGATAAG	CTTTCCCATTGCATTCCATT
stSG495417	TGACCATGATGGAATGCAATGGGAAAGA	AAGGAAAATGGTGTCAAGCAA
stSG495418	TGACCATGACGTTCAAGCATCACAGTGC	ATTCCCCAGAAGCAACAAAA
stSG495419	TGACCATGTTAAAAACCCCACCAGCAAG	GCCAGAGCAAAACTAGCCAC
stSG495420	TGACCATGGTGGCTAGTTTTGCTCTGGC	GCTTTGGACACTTCCTCTGC
stSG495421	TGACCATGGCAGAGGAAGTGTCCAAAGC	ATCCACCACAAGTCCCAAAG
stSG495422	TGACCATGTCTGAAAGCATGAGGCTTGA	CAAATCCCTGGCCATCTAAA
stSG495423	TGACCATGTTAGATGGCCAGGGATTTGA	ACTCCCCTGAGAAGCTGTGA
stSG495424	TGACCATGGACGCTCTTCATGGGATCAT	TAACTTCTGCCCCAACTCCA
stSG495425	TGACCATGTGGAGTTGGGGCAGAAGTTA	GCTGCTCTGGAGAGGAGAGA
stSG495426	TGACCATGGGTCCGAGGTCTCCTCTTTC	TTCTGGGGTTTTCACCAAAG
stSG495427	TGACCATGTTGGTGAAAACCCCAGAATC	TTTTTCCCATAAATCAACCCA
stSG495428	TGACCATGTGGGTTGATTTATGGGAAAAA	CTGCTGGCTTTTGTTGTTGA
stSG495429	TGACCATGTTGTCTTTGCAGATCAGGGA	TAATCCCTCTTCCCTCCCAC
stSG495430	TGACCATGGTGGGAGGGAAGAGGGATTA	GCCTAGCAACAGCATCCTTC
stSG495431	TGACCATGGAAGGATGCTGTTGCTAGGC	CCATCGATTCCCAGATGTTT
stSG495432	TGACCATGGTGTGAGCCCAATCCAAGTAG	CCGCAAGGGCTAAACAGAAT
stSG495433	TGACCATGTTTGTCCCTATCCCTCCCTC	GGCAGGCTCAGATCTGTAATG
stSG495434	TGACCATGCTGTCGGACATAGATTACTGGC	ACAATACAGTCGCCACTCCC
stSG495435	TGACCATGGTCTGTCTGCAGCATCCCTA	AGCGTACCTGTACCCCCAG
stSG495436	TGACCATGTACCCCTCAGCTCCGGTAGT	CAGTTGTGGAAATTGTGCGT
stSG495437	TGACCATGGGTGAGTCATGGCTTCTGGT	AGGGAAACAAAGCAGCTTCA
stSG495438	TGACCATGTTGGGGACAATAATTTGGGA	TCCAGGTAACCAGCCTCAAC
stSG495439	TGACCATGGGCTGGTTACCTGGACAGAA	ATGGGTCATTTGTCACAGCA
stSG495440	TGACCATGTGCTGTGACAAATGACCCAT	ACATAGCGCAAACCCAAAAG
stSG495441	TGACCATGGGTTTGCGCTATGTTCCACT	ATTCAGACTCCAGCGCTCTC

stSG495442	TGACCATGGAGAGCGCTGGAGTCTGAAT	GGGGGACGGGATACAGTAAT
stSG495443	TGACCATGATTACTGTATCCCGTCCCCC	TCCCATATGAAGCCAAGGTC
stSG495444	TGACCATGCCACCAGCTCCAATCAGACT	CAGCCCCATATTGGTACAGG
stSG495445	TGACCATGCTGTACCAATATGGGGCTGG	TGAAAGACAGCAAGGGAACA
stSG495446	TGACCATGGCACTGCTGGTTTTCTGTGA	AACAGGTTCCAGTCTGGGTG
stSG495447	TGACCATGCTAAGCCCTTGCCAACCATA	AAAGGTGATCATTTGCACCA
stSG495448	TGACCATGTTATCTGTGTGTTTCTGCCTCA	CCTGCTACCCGTTTCTCATT
stSG495449	TGACCATGGAGGTGTGGTGCTCACCTTT	TGCAAAGCATACGTTTCGTC
stSG495450	TGACCATGTTTGCATTAAGAGTTGGGCA	AACGGGCTTGATTACTGCAC
stSG495451	TGACCATGGTGCAGTAATCAAGCCCGTT	CATTTCCCTCAATTGTGCCT
stSG495452	TGACCATGGCACAATTGAGGGAAATGCT	GGGTTCAGAAATGGGAGACA
stSG495453	TGACCATGGTCTCCCATTTCTGAACCCA	GGCAACTTGAGAGCAGGAAC
stSG495454	TGACCATGGTTCCTGCTCTCAAGTTGCC	TTTGCCATTTCCTTTCATCC
stSG495455	TGACCATGGGATGAAAGGAAATGGCAAA	CGCTCAACTTCCACTTCTCC
stSG495456	TGACCATGGAGAAGTGGAAGTTGAGCGG	AGCCATCCACAGGCATAAAG
stSG495457	TGACCATGTGCCTGTGGATGGCTTTATT	AGCCCTCCCAATCTTACCAC
stSG495458	TGACCATGCCCCCAACAAATGTCACTCT	GGGGAATGAGAACTATCCAACA
stSG495459	TGACCATGTGTTGGATAGTTCTCATTCCCC	GGCAGAAACTGTTGACACTCTG
stSG495460	TGACCATGTGTCAACAGTTTCTGCCTTCA	TGACATCACCAGAGGGTTCA
stSG495461	TGACCATGATCACTGGGCTCTTTCATGG	GCAGCTGCAATCTTTTCACA
stSG495462	TGACCATGAGGAAATGCAAGCCCATACT	CAGATCCCCTCCATATTGGTT
stSG495463	TGACCATGGGAGGGGATCTGTGTTTCAT	AGGCCTTCAAAGCAACAATG
stSG495464	TGACCATGAACTTGAAATTCCTTGGCTACG	ATTTGGCTCAAGGGCTTTTT
stSG495465	TGACCATGCCAGGTTGCTGAAGGAAAAC	ACTAGCAAACTGCAGCCGAG
stSG495466	TGACCATGCTGCAGTTTGCTAGTGCGTC	AGCTGTGCCCAACCTCTCTA
stSG495467	TGACCATGGCTAGAGAGGTTGGGCACAG	AGCAGAAAAGAGGGCAGTCA
stSG495468	TGACCATGTGACTGCCCTCTTTTCTGCT	CTAAGTGCCTGCAAAGAGCC
stSG495469	TGACCATGGGCTCTTTGCAGGCACTTAG	GCCACCATTCAACTTGACAC
stSG495470	TGACCATGCAAGTTGAATGGTGGCATGT	AGCCAATGGTCTCTTCTGTCTC
stSG495471	TGACCATGGAGACAGAAGAGACCATTGGC	CTCCACAGGAGTGGGTCATT
stSG495472	TGACCATGGGATGTTCGTTCCGTCTTGT	TGGCCTTGTAGCAGGAAATC
stSG495473	TGACCATGTACAAGGCCAAAAGCCTGAT	TTCTTGCTGCCAATTGTGAC
stSG495474	TGACCATGTCCTGAGGTCCTCTTTCTGC	CTCTGCTTTTTCTCGGTGCT
stSG495475	TGACCATGAGCACCGAGAAAAAGCAGAG	TACACCAAACTGGGCAACAA
stSG495476	TGACCATGATTCAGGACATCGTTGGGAG	CGTCCTGCATTAAACAGTGG
stSG495477	TGACCATGGAGCAGCAATGATCACCCTT	GCTGATGACTACTCCAGCACA
stSG495478	TGACCATGCAGCTGTACAGGAAGAGGCA	ACTATTCCCAAGGCCAACCT
stSG495479	TGACCATGGGTTGGCCTTGGGAATAGTT	ACCAAATGGCCTTTCAACAG
stSG495480	TGACCATGCTGTTGAAAGGCCATTTGGT	GCACATAACATTCCAAGCCA
stSG495481	TGACCATGATTGCCAATGTCTTCCTGCT	CACACATCCCCTGCATAGTG
stSG495482	TGACCATGGCAGGGGATGTGTGTATGTG	CTTGCTTGCTTCCATGACAA
stSG495483	TGACCATGTTGTCATGGAAGCAAGCAAG	CCCAGCCACATAAAACCTGT
stSG495484	TGACCATGGTGCCATATGCATGAGCAGT	GGTCACCATTCTTTGGCACT
stSG495485	TGACCATGATGGTGACCTTGCTTCTGCT	GAAGGCTGGGCATCAAGTAA
stSG495486	TGACCATGCAGCCTTCCAATTTGTCTCC	GCAGAGTTCCAAACAGCACA
stSG495487	TGACCATGCCCAAGGAGAGGTCTCATGTT	TCCGTCCCTGCTGAATTAAC
stSG495488	TGACCATGTAATTCAGCAGGGACGGAAT	GTCTCATGGCGACCCTAAAA
stSG495489	TGACCATGCTTTTTGCCCTTTCCCATTT	GCCAGGCATCCTGATTTTTA
stSG495490	TGACCATGCTGATTTGGAGCTTGGAAGG	GCAGGGTGTAACCATGAGGT
stSG495491	TGACCATGTGCCCCAGATCCTTCTAATG	GTCAGGTGATGGCAAGGAAT
stSG495492	TGACCATGTTCCAAGGGAGTGGTGAAAG	AAGCCCACCACCCTTACTCT
stSG495493	TGACCATGAGAGTAAGGGTGGTGGGCTT	CCTTCAAGCTGGCTTTTGAC

stSG495494	TGACCATGAGAGGGCAATGTGAAGAGGA	TGGAAACATTGTAGGTGCCA
stSG495495	TGACCATGTGGCACCTACAATGTTTCCA	AGGAATGCCGTTTCCTTTTT
stSG495496	TGACCATGGACCCTTTCCTTGGGAAGTC	CCTCCAGGTTCCTCAAAACA
stSG495497	TGACCATGCCTGTTTTGAGGAACCTGGA	CCCAAGACCCATTTCTTTGA
stSG495498	TGACCATGGGGCTACCCCAATCATCATA	AAAGAATTCCAAAAGCGGGT
stSG495499	TGACCATGACCCGCTTTTGGAATTCTTT	GACAGTCCCTGCGTTGAAGT
stSG495500	TGACCATGGTATACACGGAGGGTCACGG	CAAGCTCAGTCTCCTCAGCC
stSG495501	TGACCATGTCTCACGGGTATTTTCCACA	TGGCAAGAATAACCCCACTC
stSG495502	TGACCATGTGAAAACTACACCACGCAGG	TGATGCTGCAATTTAATCCAA
stSG495503	TGACCATGTGGAAGTGAGGAGTAGGGCT	CGAATCAGGGGAAACTGAAG
stSG495504	TGACCATGCTTCAGTTTCCCCTGATTCG	AATGCCCAGTGAATTAACGC
stSG495505	TGACCATGGCCTAAGCACAGACATGAAGC	TAACTAATGCAGTGCCCCGT
stSG495506	TGACCATGTCTCCTGCTTTTCCAGAAGG	TGTGCACCAAGAAACCAAAG
stSG495507	TGACCATGTTTGGTTTCTTGGTGCACAG	TGCGAGGTAAAAGTTGAGGC
stSG495508	TGACCATGGCCTCAACTTTTACCTCGCA	AGAAAGCATGCAGTGAGGGT
stSG495509	TGACCATGCTCACTGCATGCTTTCTTGC	CCCACCATGGATTACCAGAC
stSG495510	TGACCATGGTCTGGTAATCCATGGTGGG	GGGTAAGACCCTCACGATCA
stSG495511	TGACCATGCAGGATGGTGAAGAAGGGAA	GCCGAATTGAACTACCTCCA
stSG495512	TGACCATGGTCCTCCATGCAAATCACCT	CTTTGAGAACAGCCCAGCTC
stSG495513	TGACCATGGAGCTGGGCTGTTCTCAAAG	GTGGATAAGCTGTCCCGTGT
stSG495514	TGACCATGCCGTTCTCACCTGGTTTCAC	CTTGGTGGGAATTAGCCTGA
stSG495515	TGACCATGCAAGCACTGGAACAGCACAC	GGAGCCTGAGGGATCCTAGT
stSG495516	TGACCATGGGGGAAACTAGGATCCCTCA	GGGGATTCCAAAATGAACCT
stSG495517	TGACCATGGAATCCCCACGGTAGAGACA	TTAGCCATTCAGAGGGTTGG
stSG495518	TGACCATGCCAACCCTCTGAATGGCTAA	CCCACTCTGGAGAACAGCTC
stSG495519	TGACCATGTGTTCATCCTGGACTCCCTC	CCTCCATGTCTTCCCAGTGT
stSG495520	TGACCATGACACTGGGAAGACATGGAGG	ACAGGCCTAAGGGAAGGAAA
stSG495521	TGACCATGTTTCCTTCCCTTAGGCCTGT	CTTCTCTCCCTCTACCCGCT
stSG495522	TGACCATGAGCGGGTAGAGGGAGAGAAG	ACATCAAGTGGCTGGAAAGG
stSG495523	TGACCATGCCTTTCCAGCCACTTGATGT	TCTCACATGCTCCGTGCTAC
stSG495524	TGACCATGGTAGCACGGAGCATGTGAGA	CTGATCAGAGAGCCCAGAGG
stSG495525	TGACCATGCTCTCTGATCAGGGTCCTCG	CTATCCCCACAGGAGCAAAA
stSG495526	TGACCATGTTTTGCTCCTGTGGGGATAG	GCTGCACCTAATCCAGAACC
stSG495527	TGACCATGGCTGGTTCTGGATTAGGTGC	CTTAAGGCTCCTCCTCTGCC
stSG495528	TGACCATGGCTTTTTGAGTTCACAGCCC	TCTCAAGCGTCCTTCCATCT
stSG495529	TGACCATGATGGAAGGACGCTTGAGAGA	AGCAGATCAGTGACGAGGGT
stSG495530	TGACCATGTCCAGTTCCCAGAGATGGAG	GGCCTTCCTAATCTTCACCA
stSG495531	TGACCATGGGAGAATGAGGGCAGTGTGT	CTGGATTCTCCCCCAGTGTA
stSG495532	TGACCATGAGGGTGAACTGGTGAGAGGA	TTACCGAGTTTCCTGGACCTT
stSG495533	TGACCATGAAAACTGGGACAAGGTGTCG	TCTGTGTGGGTAGCTTGTGC
stSG495534	TGACCATGAGTCAGTGCCCCATAAATGC	TTCATGGCATCCCTACTGGT
stSG495535	TGACCATGCCTCTATTTCCACTGGGCAA	TTTGGGGACAAATCAAGGAG
stSG495536	TGACCATGCCAAAACCCTCAGCAAGGTA	TCATCCTCCCACACAGATCA
stSG495537	TGACCATGCACCCTATGCCAGGAACAAG	TACACACCATGCACACATGC
stSG495538	TGACCATGGCATGTGTGCATGGTGTGTA	CCTCTCTGTGTTCCTGGCTC
stSG495539	TGACCATGCAGAACAGAGGCTGACTCCC	CCCTGAGATGGTTCAAGGAA
stSG495540	TGACCATGGGTCTTTGTTAAAGCAGCCAA	TTTTGGCAATTCCGATTCTC
stSG495541	TGACCATGTTCTTTGGCACCTTGGTTTC	TGCTTTCTCCCTTTGCTCTC
stSG495542	TGACCATGAGCAAAGGGAGAAAGCACAG	GCCTCTCCTGAAGCTTTGAA
stSG495543	TGACCATGTTCAAAGCTTCAGGAGAGGC	CCTTCTAGTTTCTTGCCCCC
stSG495544	TGACCATGACGGATTCTACCCCTGGAAC	GGCTTCCTGTTTTCAGCTTG
stSG495545	TGACCATGGCCCTCAATGAGCTGTGATT	TGCAAGAGGGAAACAGATGG

stSG495546	TGACCATGAATGCCATCTGTTTCCCTCTT	AAACCCATTCAGAAGATTTGGA
stSG495547	TGACCATGTCCAGAGGTGTTTGAGAGGAA	CAGCCAATCATCAAAGAGCA
stSG495548	TGACCATGTGCTCTTTGATGATTGGCTG	TTGCATTTATTGGCCATCTG
stSG495549	TGACCATGCAGTTTGCAGATGGCCAATA	AAGGCCAGAGTAGGCTGACA
stSG495550	TGACCATGAATCTTACATGGGGGAGCAG	CATGCTGGTAAATTGCCTCC
stSG495551	TGACCATGCCTCTTACGAAAGCTGAAGGC	TACCCCTTTGGAATGAGCTG
stSG495552	TGACCATGGCAATGGGGACTTGCAAAA	CCAAAAGTCATCACATTAGGGC
stSG495553	TGACCATGTACCCAATGACCCAATGACC	GAAGACTTCTGCACCCATCC
stSG495554	TGACCATGGACGTATCCAGACAAGCCCT	GGGGCCAATCTAATCCTTCT
stSG495555	TGACCATGTTAGATTGGCCCCTCTCCTT	GATTCCAGTGGGGGATACCT
stSG495556	TGACCATGAGGTATCCCCCACTGGAATC	TTATCTTCCCACCCAACCCT
stSG495557	TGACCATGCCACCCACAAATGGGAAAG	AAAGGTCCTCTGCTGCTGAA
stSG495558	TGACCATGGCTTGCCTATGGGTGTGTCT	GATGTGGAGGAATGTGGCTT
stSG495559	TGACCATGAAGTCCTGAGGAGCCCATTT	ATGCAATGAAGGTGGGAAAG
stSG495560	TGACCATGCTTTCCCACCTTCATTGCAT	GCTTGGCTTGGTCTGTTTTC
stSG495561	TGACCATGAGGACACAGGATCAACCAGG	CAGTTGACATGACCCTCCCT
stSG495562	TGACCATGAAGTTGATGGATCAGGGTGG	AGGTCAGCTCTGCACCACTT
stSG495563	TGACCATGCGCAATCCTTAGGCAGTGAT	GTGTACAGTCCGGGAGCATT
stSG495564	TGACCATGTCCCGGACTGTACACAAACA	AAGCAGTTGTGGTCCAGGAG
stSG495565	TGACCATGGGCAATGGTTTTCTGCAAAT	CCTTCTGAAACTGGGGATCA
stSG495566	TGACCATGCTGGACTTCCACAGGGCTT	CCTAGGACACTCTCCGGTTG
stSG495567	TGACCATGCCACAAATGGAAGGTATGGC	CCTCCCTAGAAGGCAGTGTG
stSG495568	TGACCATGGCCACAATGGCTGGACTTAT	TGGGAGAGAAACATGCACAG
stSG495569	TGACCATGATGCCGCATTTAGCAACTCT	TTCCTCAGACTGCCTCCTGT
stSG495570	TGACCATGCAGGAGGCAGTCTGAGGAAG	TAGGTCAAGGGTTGTGGGAG
stSG495571	TGACCATGCTCCCACAACCCTTGACCTA	AGCCTACCTTCCCCTTGAGA
stSG495572	TGACCATGTTGGGAGAGCTTGGCTTAAA	AGTCCTGGGGCTGGTGTATT
stSG495573	TGACCATGTCCTCTGTTCCCCATCTCAC	TTACCGGCTTTCTCTGCAAT
stSG495574	TGACCATGTGCACAAATGGCTTGATTGT	CCTTCCTTCCCCTGTGAGTT
stSG495575	TGACCATGGGCCCAGTTCACTCATGTTT	TGGTGGTTTTATTTCCTGCC
stSG495576	TGACCATGCTCTCAGGGCCCTTTCCTT	TGAAACACTAGCAAGCGTGG
stSG495577	TGACCATGCCACGCTTGCTAGTGTTTCA	GTTTGAAAACCACCCGCTTA
stSG495578	TGACCATGGTCAAAAGAGCAAAGCCAGG	CTACCGTGCCCAGAGTCATT
stSG495579	TGACCATGCCTCCACTCACCAAGAGAGC	TGCTTCATTTTATTTCCGGC
stSG495580	TGACCATGCATTCTGAGCAGCTTGCTTG	CTGTGATCAAGGCAGAATGAA
stSG495581	TGACCATGCAATCAGGTGGCAAGACAAA	GTGCCAAGCTGTTTGGAGTT
stSG495582	TGACCATGTGGCACCAAATCCATCAGTA	CCTGTTGTTCCCATCACCTT
stSG495583	TGACCATGGAGCTCAAAGGTGTCCTTGC	TGTAAGCTCTGTGGACGCAC
stSG495584	TGACCATGAGTCAGGCGCTAGAGGAAGC	CACTGAATTTGGCCTTACCC
stSG495585	TGACCATGAGGCACTAAACTGGCTCCCT	GCCATCCTGCAAGAGAAGTC
stSG495586	TGACCATGCTCATGGTAATGCCTGGTCC	CAGACGGTCCTGAGCTCTTC
stSG495587	TGACCATGGCTCAGGACCGTCTGACTTC	GGAAGTGAAACCAGCCACAT
stSG495588	TGACCATGAGGAGCTTTTGGTGATTGGA	TACAAGGCAAGGAGCCAACT
stSG495589	TGACCATGTTGGCTCCTTGCCTTGTACT	AGAGTATGGGCTTTGGGCTT
stSG495590	TGACCATGAAGCCCAAAGCCCATACTCT	AATTTGCTTCCTGCCTTTGA
stSG495591	TGACCATGTGTTGCATTTGTGGAGAGGA	ACCTACCTGCCACTCCCTTT
stSG495592	TGACCATGGGAGTGGCAGGTAGGTGAGA	CAGACACCCCTGTCTGTTCC
stSG495593	TGACCATGGTCTGCAGAGGTTTCCCAAC	GAGGCTGCAGTCACAAATGA
stSG495594	TGACCATGGCCCTGAGAGCCTGAATCTA	ACCTCAGCGTTTCCATCGTA
stSG495595	TGACCATGTACGATGGAAACGCTGAGGT	CCTGACCAGCCCAATTAAGA
stSG495596	TGACCATGTTTCCTCCCAACCACTTGTC	TGCTGGCCTATCCCAAATTA
stSG495597	TGACCATGATTTGGGATAGGCCAGCAAT	CCCCAACAGGACATAAAAAGG

stSG495598	TGACCATGGCCTGAAGGGAATGGAGTTT	CTAAGCTCACCATCCCCAAA
stSG495599	TGACCATGGATGCACATGGTTTGACTGG	AGGGCTGCTGACACCTAGAA
stSG495600	TGACCATGGCTTACAACATGGCTGTGGA	CTATGGAACAAGCAGCACCC
stSG495601	TGACCATGAGTGGAAGGGCTGTTTCTCA	AAGACAGGAGTATGCCAGGAA
stSG495602	TGACCATGTTCCTGGCATACTCCTGTCTT	CTAAGGGAGGTGACGCAGAG
stSG495603	TGACCATGGTCACCTCCCTTAGGAAGCC	AGGACAGACCAGGCAAGAGA
stSG495604	TGACCATGCTGTGCATCACAAAGCCATT	TCACCATAGACACCAGGGTATG
stSG495605	TGACCATGCATACCCTGGTGTCTATGGTGA	CAGGGGCTTCAGCTGTCTAA
stSG495606	TGACCATGTTAGACAGCTGAAGCCCCTG	TGTCTCAACCTTTGGTGTGC
stSG495607	TGACCATGGGTCATGTGCAAGTCTCCAG	CCTGGTCAGAGCCTCATTTC
stSG495608	TGACCATGCCAGAGGAAATGAGGCTCTG	TCTGTTCAGCAATCACCTGC
stSG495609	TGACCATGGCAGGTGATTGCTGAACAGA	TGCTGTTTCCCACAAGTCAA
stSG495610	TGACCATGAGTTCAGGTTGCTTGGATGG	CACACTGGGGAGGTGAGATT
stSG495611	TGACCATGTGTATACACCCCTCCTCCCA	AAAATCTCCTGGACTGGCCT

*TCACCATG – Amino linking adaptor added to the 5’ end of all forward primers

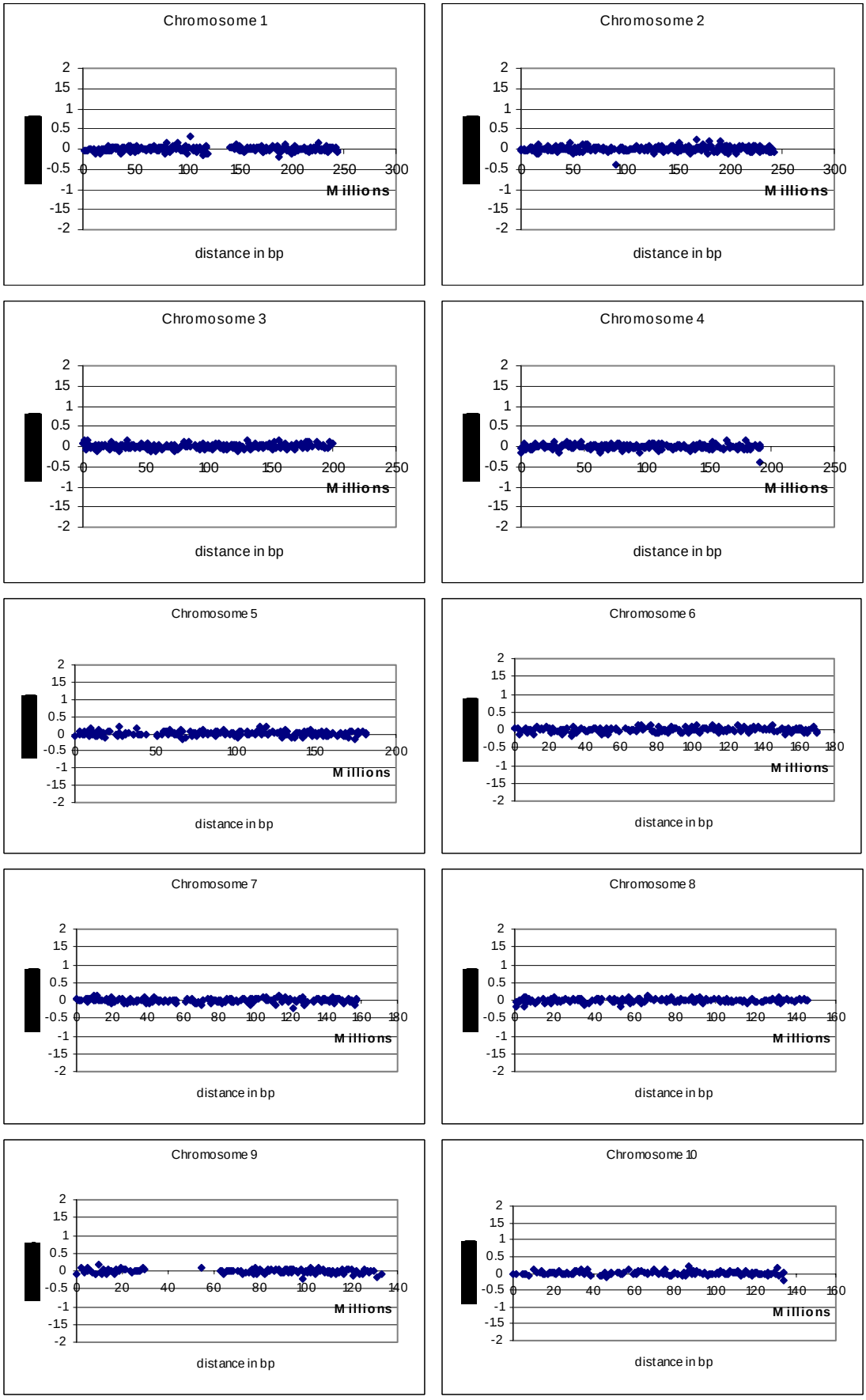
2b: The 96 well format of primers STSG 495474-495569.

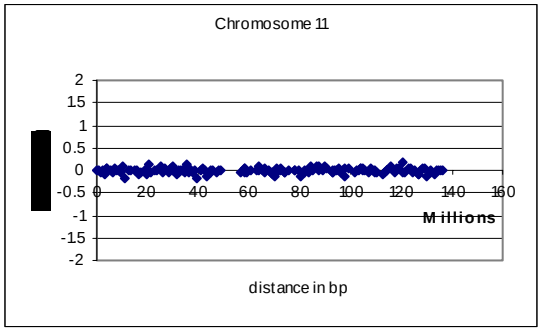
	1	2	3	4	5	6	7	8	9	10	11	12
A	stSG495474	stSG495475	stSG495476	stSG495477	stSG495478	stSG495479	stSG495480	stSG495481	stSG495482	stSG495483	stSG495484	stSG495485
B	stSG495486	stSG495487	stSG495488	stSG495489	stSG495490	stSG495491	stSG495492	stSG495493	stSG495494	stSG495495	stSG495496	stSG495497
C	stSG495498	stSG495499	stSG495500	stSG495501	stSG495502	stSG495503	stSG495504	stSG495505	stSG495506	stSG495507	stSG495508	stSG495509
D	stSG495510	stSG495511	stSG495512	stSG495513	stSG495514	stSG495515	stSG495516	stSG495517	stSG495518	stSG495519	stSG495520	stSG495521
E	stSG495522	stSG495523	stSG495524	stSG495525	stSG495526	stSG495527	stSG495528	stSG495529	stSG495530	stSG495531	stSG495532	stSG495533
F	stSG495534	stSG495535	stSG495536	stSG495537	stSG495538	stSG495539	stSG495540	stSG495541	stSG495542	stSG495543	stSG495544	stSG495545
G	stSG495546	stSG495547	stSG495548	stSG495549	stSG495550	stSG495551	stSG495552	stSG495553	stSG495554	stSG495555	stSG495556	stSG495557
H	stSG495558	stSG495559	stSG495560	stSG495561	stSG495562	stSG495563	stSG495564	stSG495565	stSG495566	stSG495567	stSG495568	stSG495569

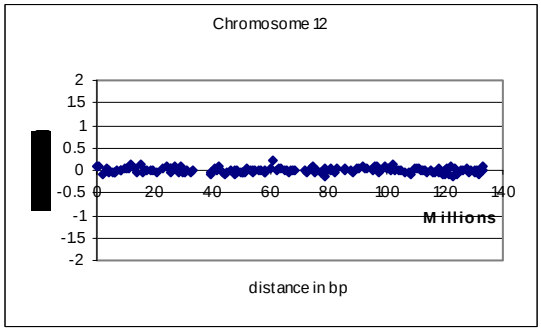
Appendix 3: Primers for quantitative PCR

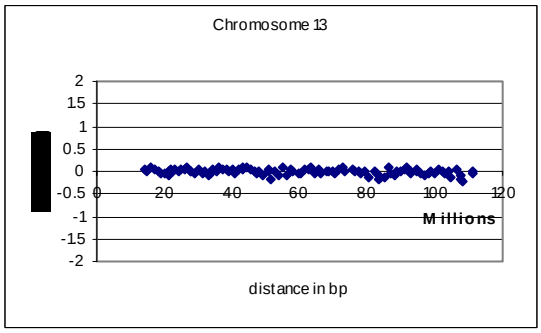
Clone	Forward Primer	Reverse Primer
cE140F8-1	TGTTCTATGAGTATGCGACTTTCCA	TTCAAACTGTGGGATGGTGAGA
cE140F8-2	TCTGCATCTTAAAGTGAGAGCTATGTTAC	TGAAGCTCTGATCTCCAGAAAGAG
cE140F8-3	GCTGATTTCCTCGTTCCCTCTATT	GTGTTAGGCAGTGGAATCATGTTC
cE140F8-4	GGATTCTGTCTTGTCTGGCCTTT	CTCCCGCGGTGCCTTT
cE140F8-5	GGCACCGCGGGAGAAG	GGCTGCATTGTTACAAATCTTTTTT
cN69F4-1	GGTTGAGGTCTGAAGCCCTTT	GGTCACTGCCCAGGCTCTT
cN69F4-2	CCTTGTCATCCCAAATACACCAT	AGACAGCTCCTGGGTCTTCCA
cN69F4-3	CAGAAACTGGCTTTGGAGAGATC	GAGACGTGGCTGAGCACAGA
cN69F4-4	GCACAAAATGTTCGAGACTGATACA	TTTACAACAAAGGCCAAATGCA
bK57G9-1	GGTGAGCCACATTTGTTATATTTGAA	GACTCACCCTTCCCCCTCTAAG
bK57G9-3	CTGTGCTGTGAATAGATCCATGTG	TGGCCGGGTGAACTCTTC
bK57G9-4	ACAATGGGTGCCAAGTTGGTA	CCCACAACCTGCTGCAGACT
bK57G9-5	TGGGCAGAGTCCCTGATTCT	AACTGGAAGGTGAACCCCAAA
bK57G9-6	GACTTCCAGGCCCTATGTCAGA	AAGTGGGAAGTTGCTGCTATGC
bK57G9-7	GATGCATGGGTGGGTGATG	TCCTGAGCCTCATTTGTTCTCA
bK57G9-8	CGGGCTTTGTCACAGCATCT	CAAAACTGGGAACAGCCTAAACA
cB13C9-1	TCAACAAGATATGTGCAAGCTTCTC	AAACTCCACCGGGCTCAAT
cB13C9-2	TTGCTGAGATTATGAATGGGTTTC	CTAGAGCTATTTTCTGTTTCCGACATACT
cB13C9-3	GCTGCACAAGCCATCCATTT	GGCCAGTGTGATTGATAAACTGAGT
cB13C9-4	GGGAGAATCCCAGCAAGTCA	CACCTCCCTGGTTGGTCATC
cB13C9-5	CTGCACCCCTCTTGTCTGTAACT	CGTCCTGAAACTTGGCATCTG

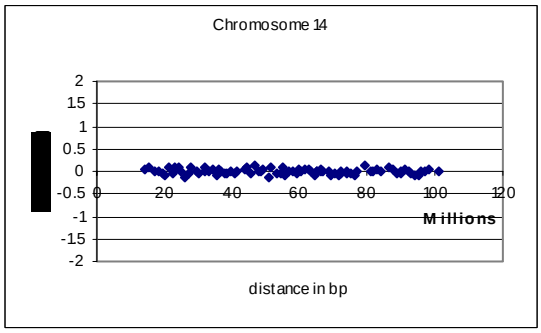
Appendix 4: Male:male hybridisation on 1Mb array

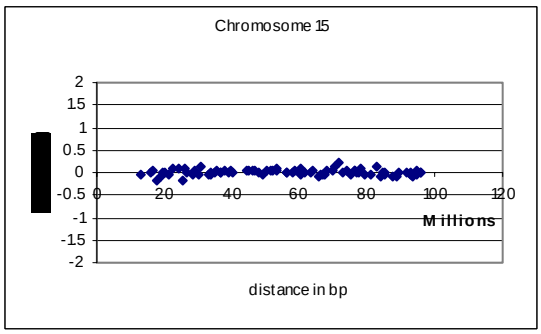


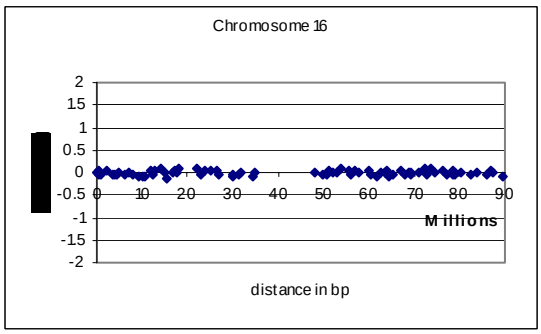

Chromosome 11
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
Millions
distance in bp

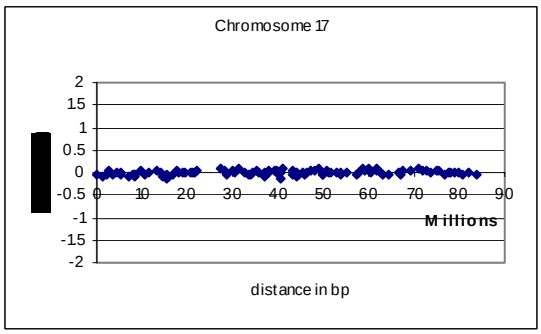

Chromosome 12
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
Millions
distance in bp

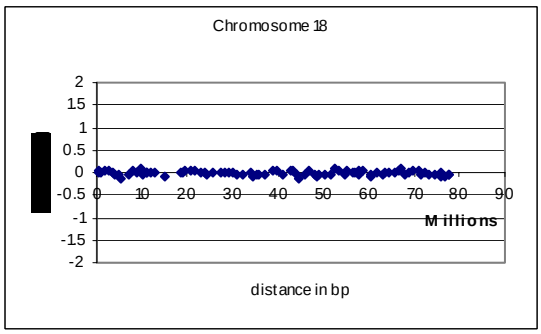

Chromosome 13
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

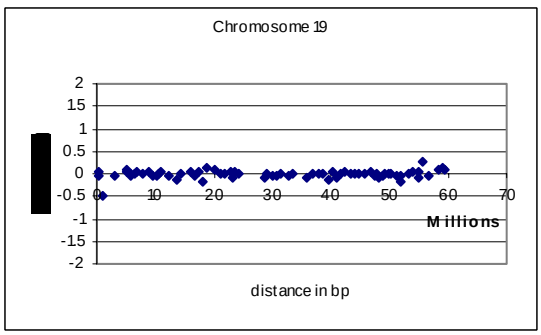

Chromosome 14
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

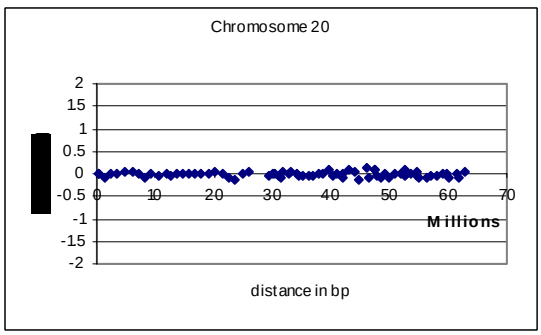

Chromosome 15
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

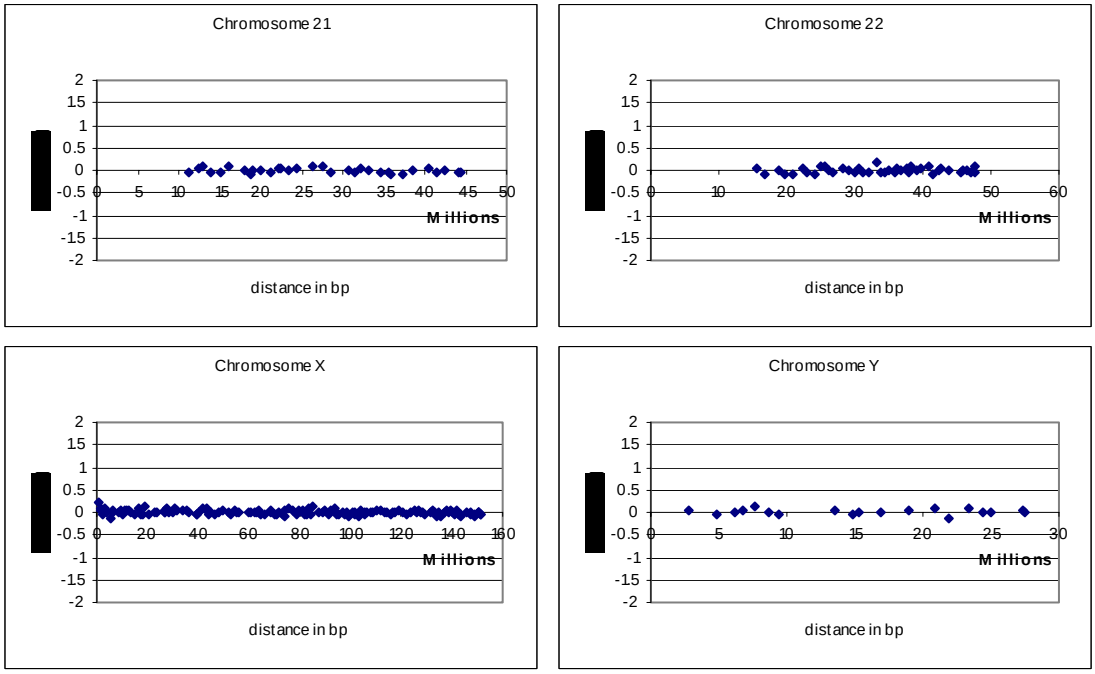

Chromosome 16
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp


Chromosome 17
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp


Chromosome 18
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp

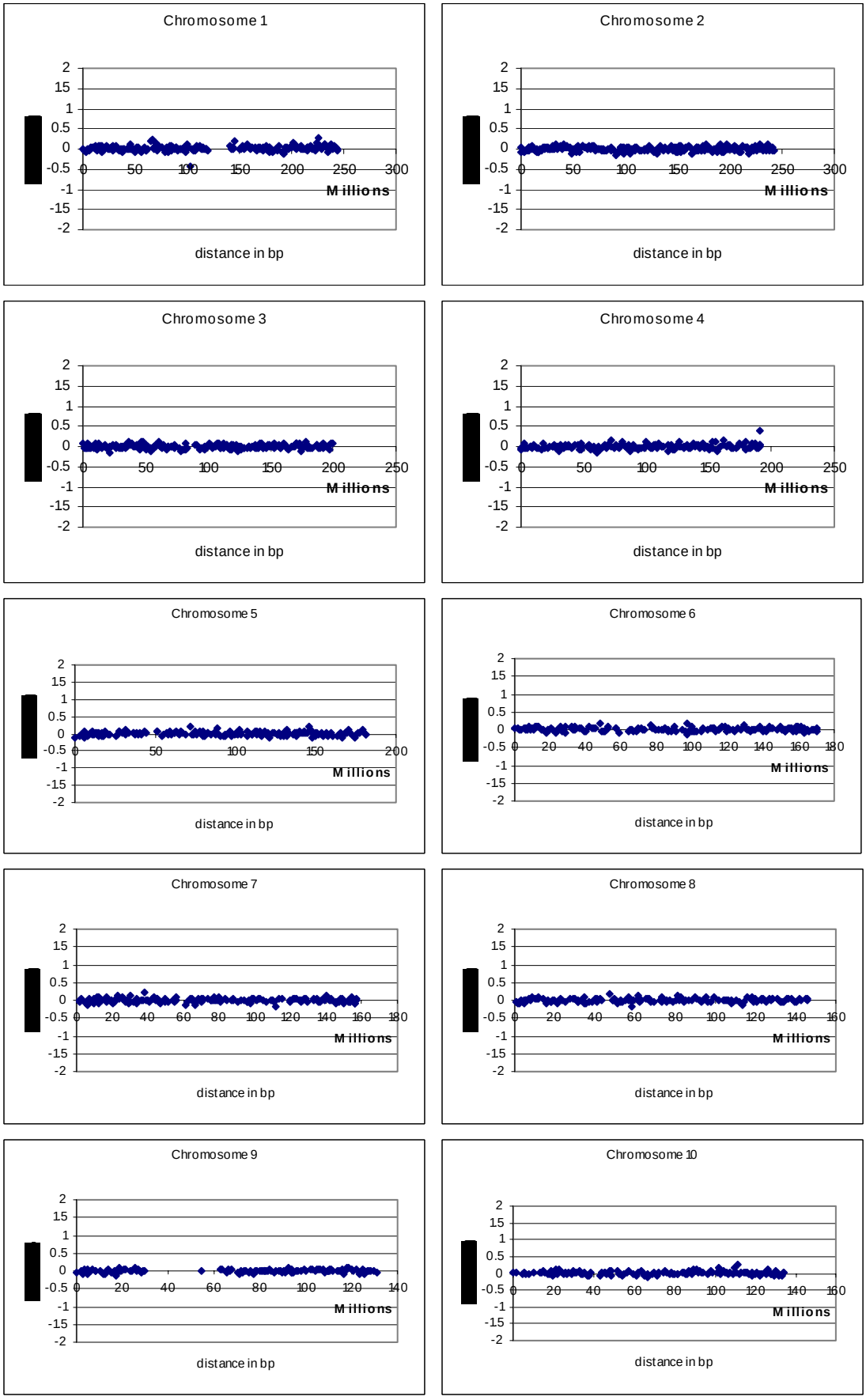

Chromosome 19
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
Millions
distance in bp

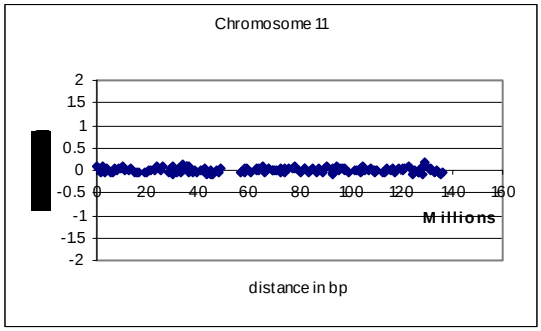

Chromosome 20
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
Millions
distance in bp

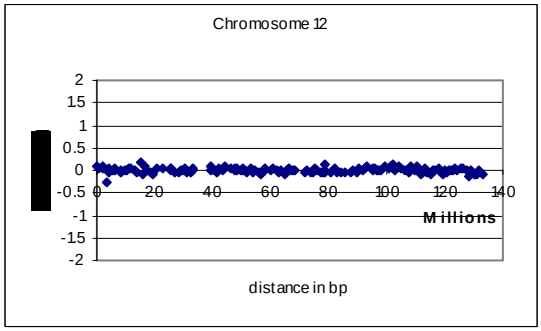


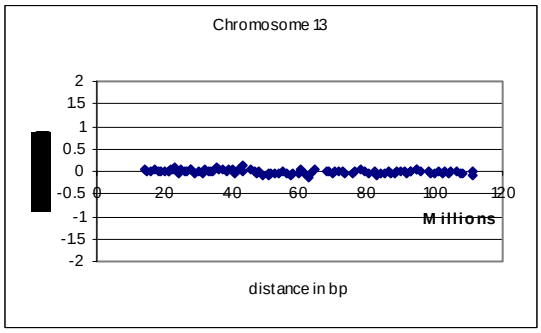
Log^2 ratios are given; therefore a 1:1 ratio will report a log^2 ratio of 0.

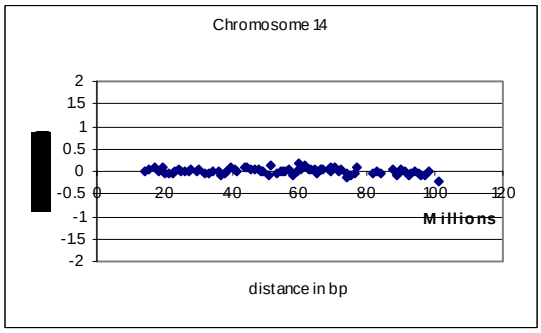
Appendix 5: Male:female hybridisation on 1Mb array

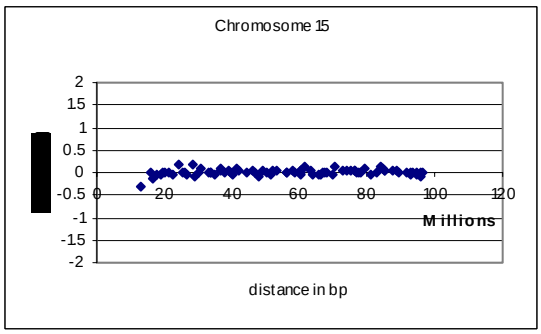


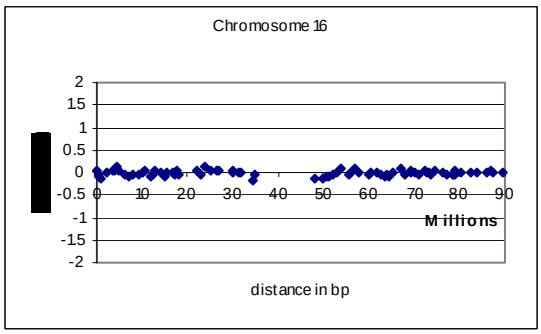

Chromosome 11
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
Millions
distance in bp

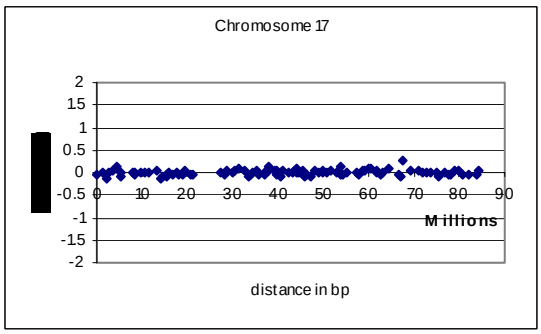

Chromosome 12
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
Millions
distance in bp

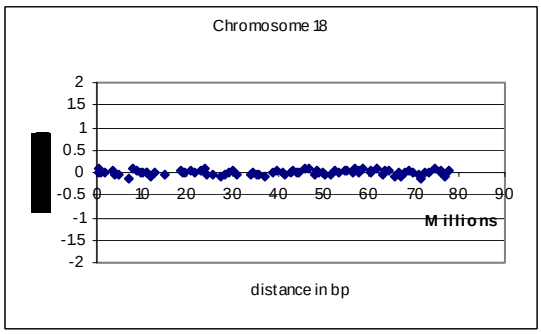

Chromosome 13
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

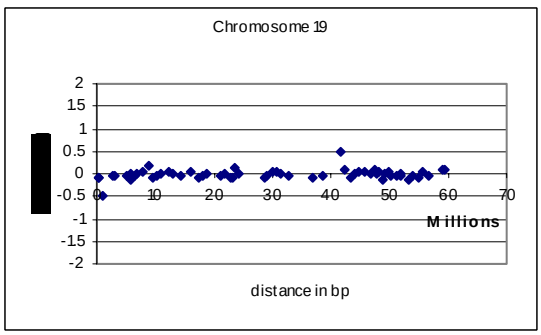

Chromosome 14
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

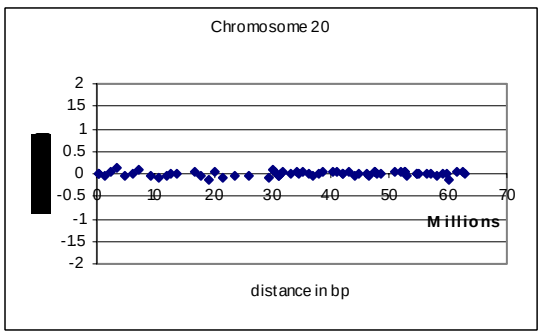

Chromosome 15
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

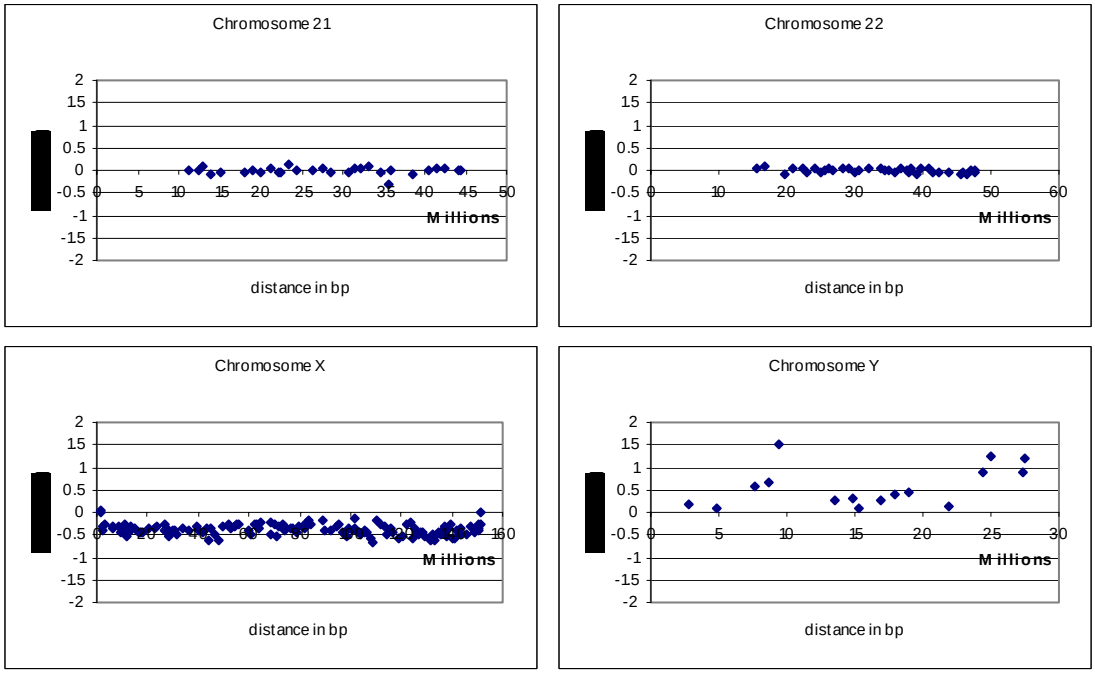

Chromosome 16
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp


Chromosome 17
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp


Chromosome 18
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp

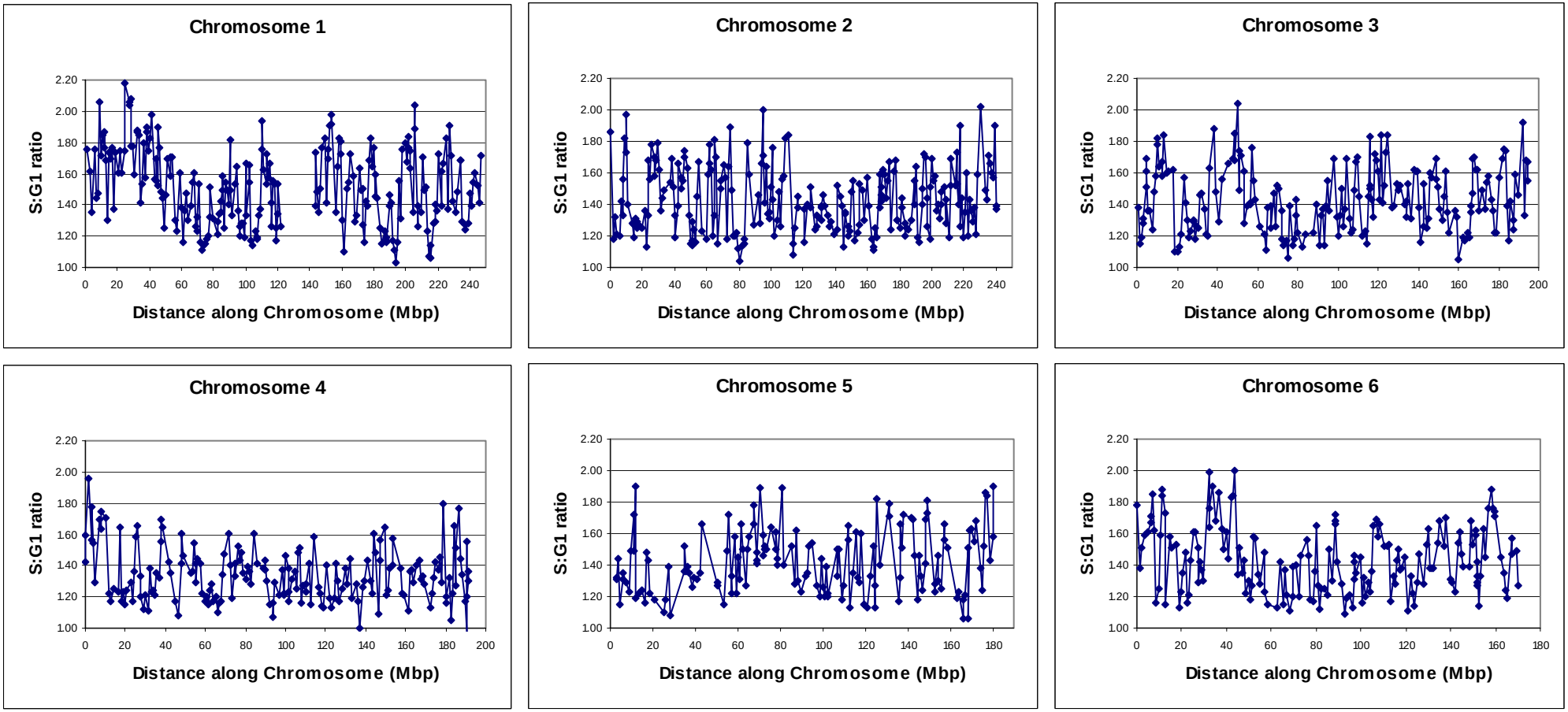

Chromosome 19
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
Millions
distance in bp

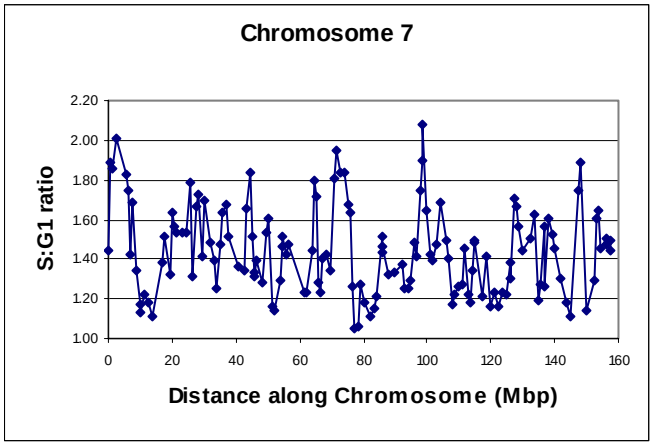

Chromosome 20
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
Millions
distance in bp

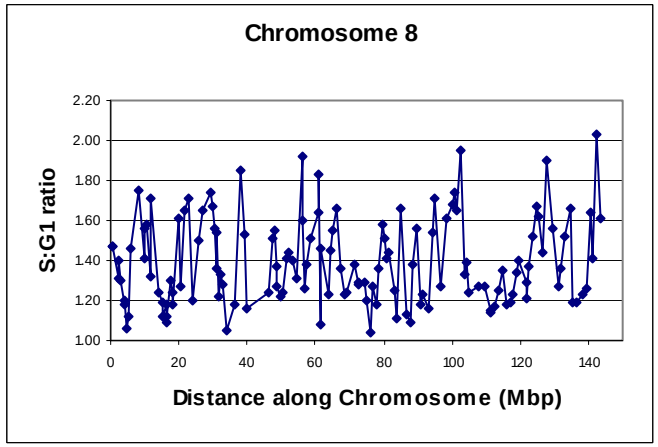


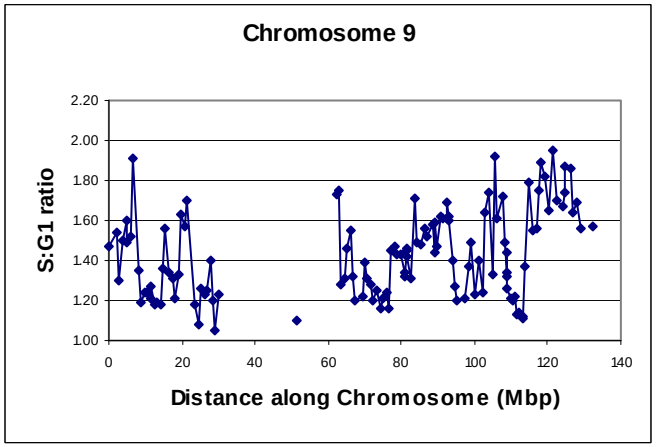
Log^2 ratios are given; therefore a 1:1 ratio will report a log^2 ratio of 0, this is seen on the autosomes. A ratio of 0.5:1 representing a single copy loss on the X clones will report a ratio of 0.5. Clones on the Y chromosome report a variety of ratios. This is because there is only Cy 3 labelled Y chromosome DNA within the hybridisation mix, so there is no Cy 5 DNA to hybridise against.

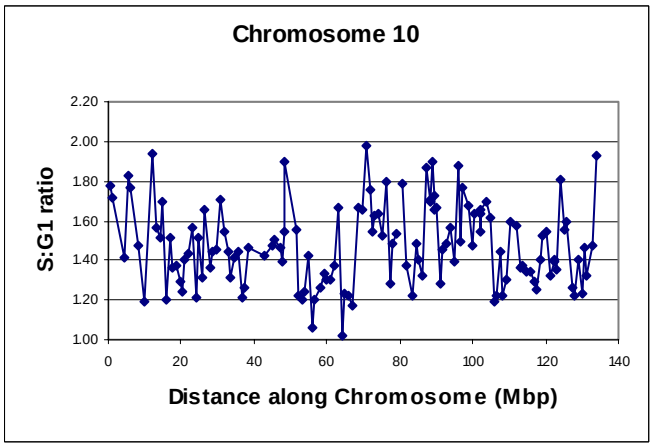
Appendix 6: Replication timing profiles for all 24 chromosomes.

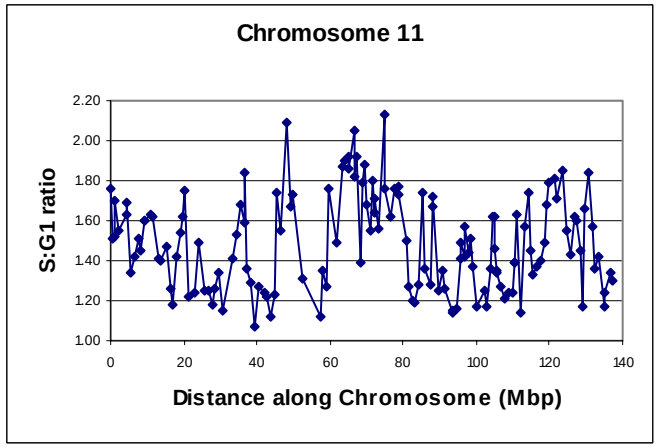


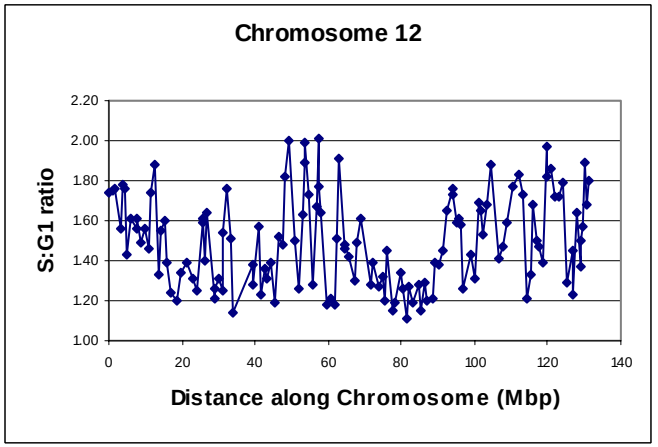

Chromosome 7
S:G1 ratio
2.20
2.00
1.80
1.60
1.40
1.20
1.00
0
20
40
60
80
100
120
140
160
Distance along Chromosome (Mbp)

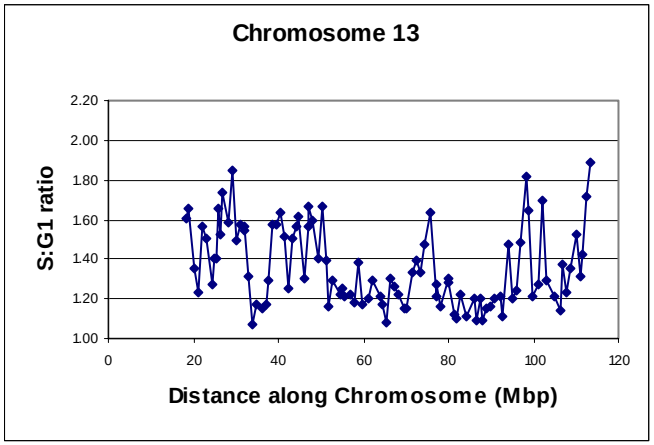

Chromosome 8
S:G1 ratio
2.20
2.00
1.80
1.60
1.40
1.20
1.00
0
20
40
60
80
100
120
140
Distance along Chromosome (Mbp)

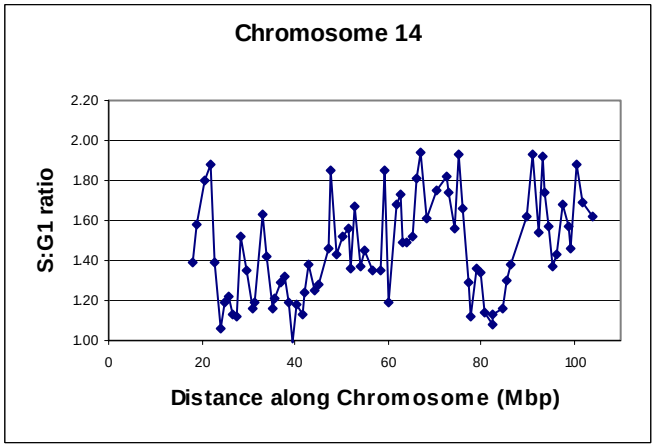

Chromosome 9
S:G1 ratio
2.20
2.00
1.80
1.60
1.40
1.20
1.00
0
20
40
60
80
100
120
140
Distance along Chromosome (Mbp)

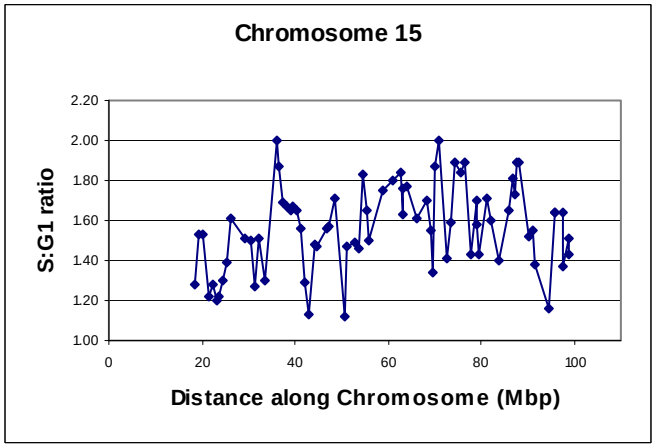

Chromosome 10
S:G1 ratio
2.20
2.00
1.80
1.60
1.40
1.20
1.00
0
20
40
60
80
100
120
140
Distance along Chromosome (Mbp)

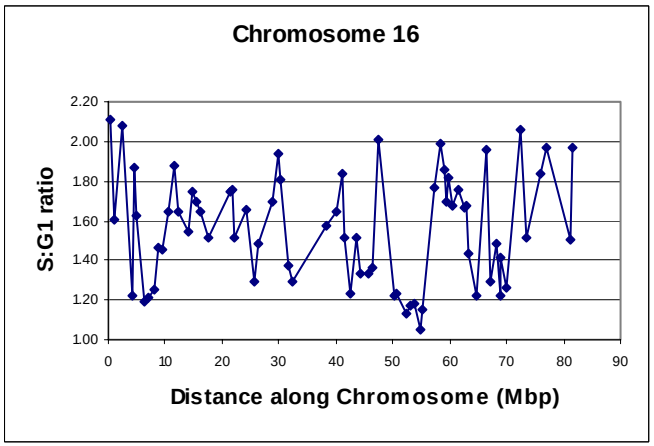

Chromosome 11
S:G1 ratio
2.20
2.00
1.80
1.60
1.40
1.20
1.00
0
20
40
60
80
100
120
140
Distance along Chromosome (Mbp)

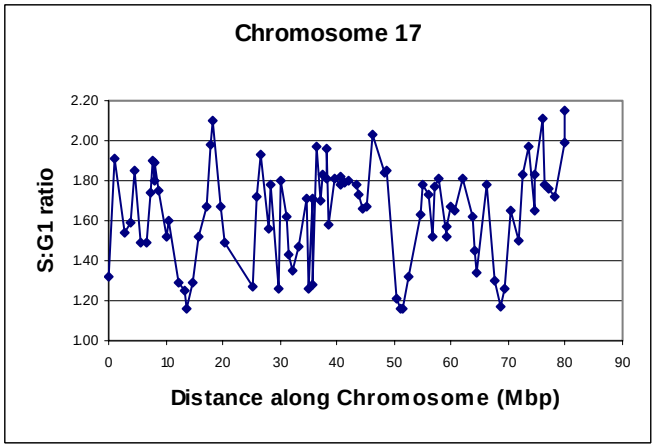

Chromosome 12
S:G1 ratio
2.20
2.00
1.80
1.60
1.40
1.20
1.00
0
20
40
60
80
100
120
140
Distance along Chromosome (Mbp)

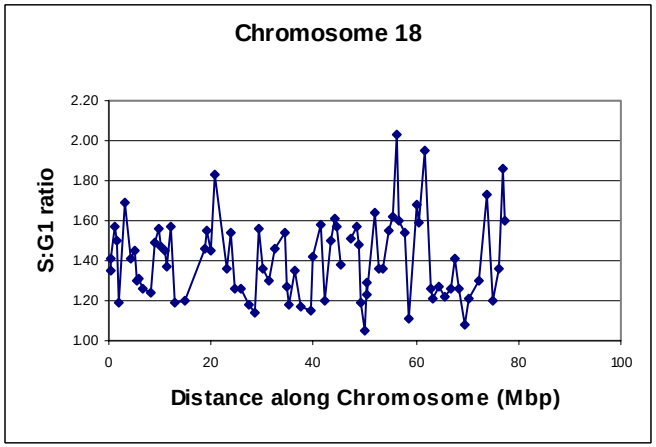

Chromosome 13
S:G1 ratio
2.20
2.00
1.80
1.60
1.40
1.20
1.00
0
20
40
60
80
100
120
Distance along Chromosome (Mbp)

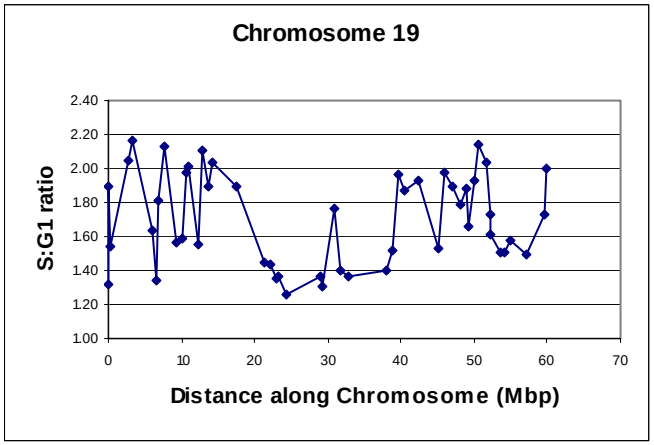

Chromosome 14
S:G1 ratio
2.20
2.00
1.80
1.60
1.40
1.20
1.00
0
20
40
60
80
100
Distance along Chromosome (Mbp)

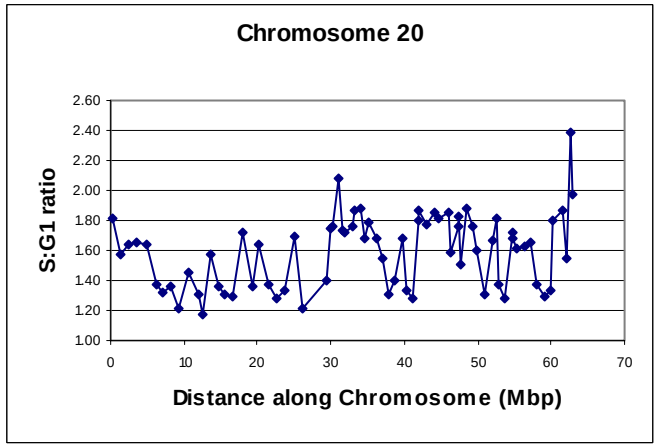

Chromosome 15
S:G1 ratio
2.20
2.00
1.80
1.60
1.40
1.20
1.00
0
20
40
60
80
100
Distance along Chromosome (Mbp)

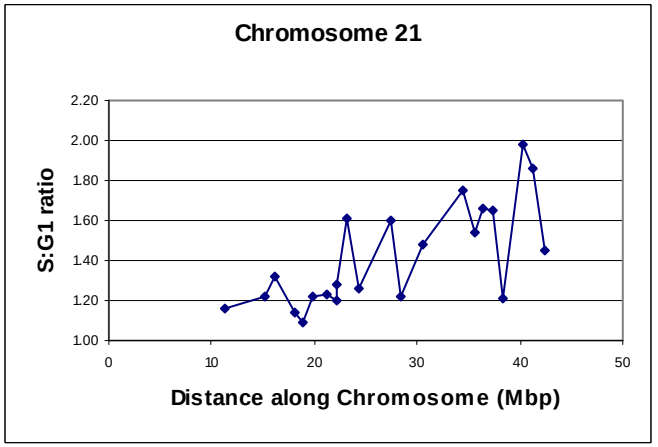

Chromosome 16
S:G1 ratio
2.20
2.00
1.80
1.60
1.40
1.20
1.00
0
10
20
30
40
50
60
70
80
90
Distance along Chromosome (Mbp)

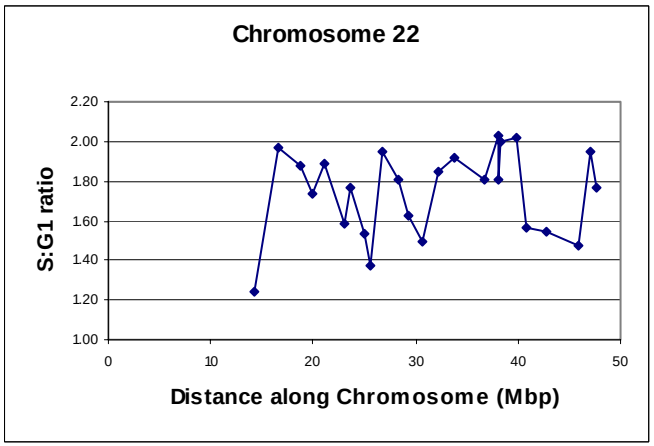

Chromosome 17
S:G1 ratio
2.20
2.00
1.80
1.60
1.40
1.20
1.00
0
10
20
30
40
50
60
70
80
90
Distance along Chromosome (Mbp)

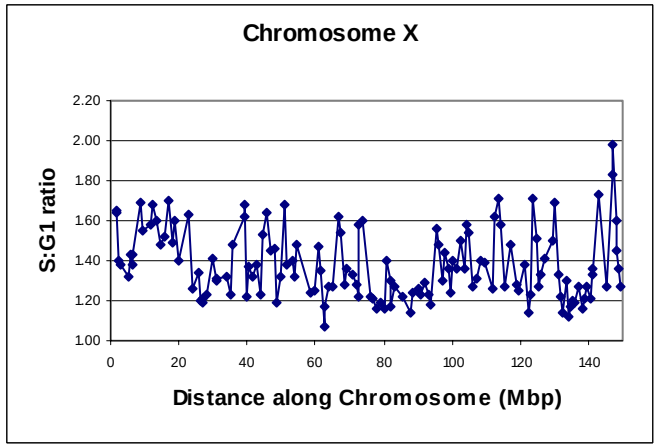

Chromosome 18
S:G1 ratio
2.20
2.00
1.80
1.60
1.40
1.20
1.00
0
20
40
60
80
100
Distance along Chromosome (Mbp)

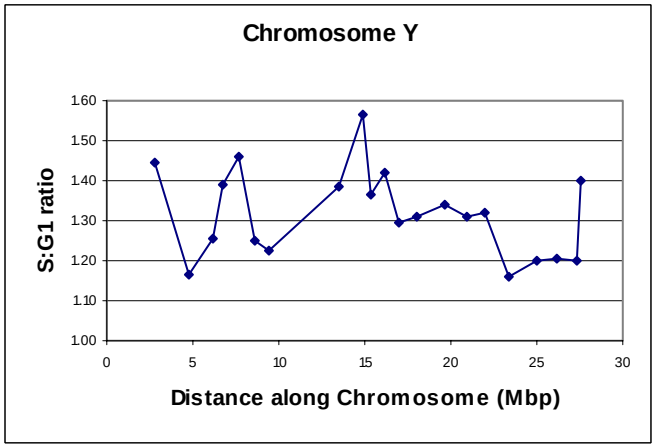

Chromosome 19
S:G1 ratio
2.40
2.20
2.00
1.80
1.60
1.40
1.20
1.00
0
10
20
30
40
50
60
70
Distance along Chromosome (Mbp)


Chromosome 20
S:G1 ratio
2.60
2.40
2.20
2.00
1.80
1.60
1.40
1.20
1.00
0
10
20
30
40
50
60
70
Distance along Chromosome (Mbp)


Chromosome 21
S:G1 ratio
2.20
2.00
1.80
1.60
1.40
1.20
1.00
0
10
20
30
40
50
Distance along Chromosome (Mbp)


Chromosome 22
S:G1 ratio
2.20
2.00
1.80
1.60
1.40
1.20
1.00
0
10
20
30
40
50
Distance along Chromosome (Mbp)

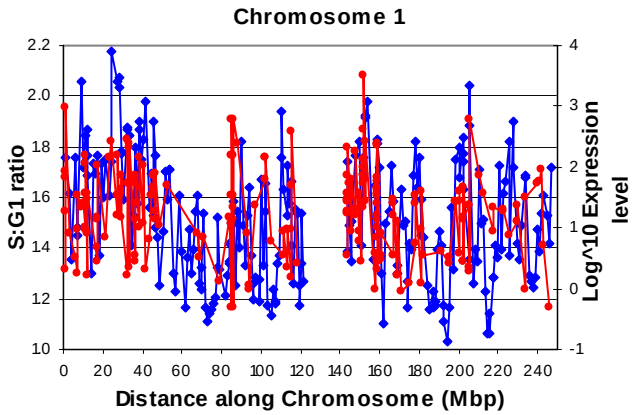

Chromosome X
S:G1 ratio
2.20
2.00
1.80
1.60
1.40
1.20
1.00
0
20
40
60
80
100
120
140
Distance along Chromosome (Mbp)

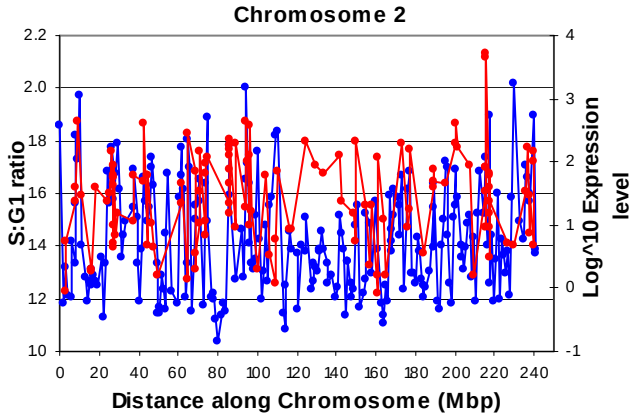

Chromosome Y
S:G1 ratio
1.60
1.50
1.40
1.30
1.20
1.10
1.00
0
5
10
15
20
25
30
Distance along Chromosome (Mbp)

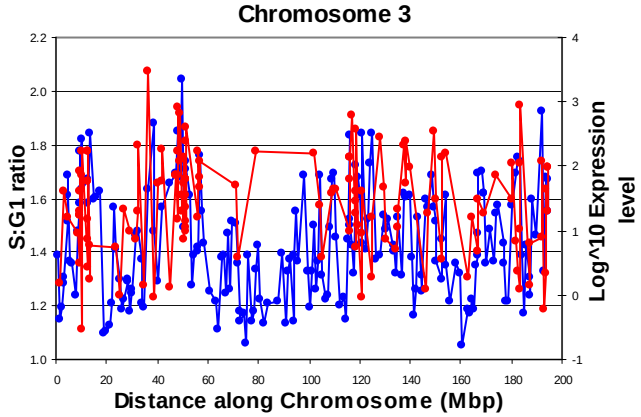
Appendix 7: Perl program to identify regions of co-ordinated replication

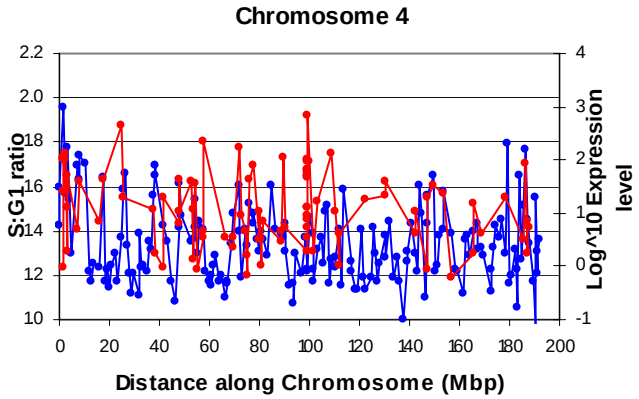
A purpose-written perl program was used to find the optimal segmentation of the replication timing (RT) data. Suppose a chromosome contains n RT signals arranged in genome order. Within each segment, starting at coordinate i and ending at coordinate j, we define the score S_{ij} equal to the sum of squared deviations of the RT values from the mean RT signal μ_{ij} for the segment. The optimal segmentation pattern (ie the number of segments and coordinates of segment boundaries) is chosen which minimises a function, W_n, based on the sum of segment scores plus a penalty score B for each segment transition. Let W_k be the score of the optimal segmentation for coordinates 1 through k. Then $W_0 = 0$ and $W_k = \min_{i<k} \{ W_{i-1}+B+S_{ik} \}$ for all $k>0$. The degree of segmentation is controlled by the value of B. The optimal segmentation is found by backtracking from the terminal value $W_{n.}$. The statistical significance of W was determined by re-running the program on 1000 permuted data sets in which the order of observed RT signals was shuffled. The P-value for the test of the null hypothesis that the observed segmentation score could have arisen by chance is estimated as the proportion of times the permuted W score exceeded the observed score.

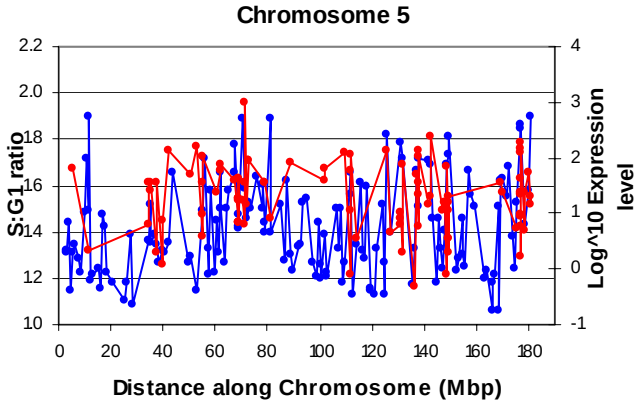
Appendix 8: Replication timing and Expression level profiles for all 24 chromosomes.

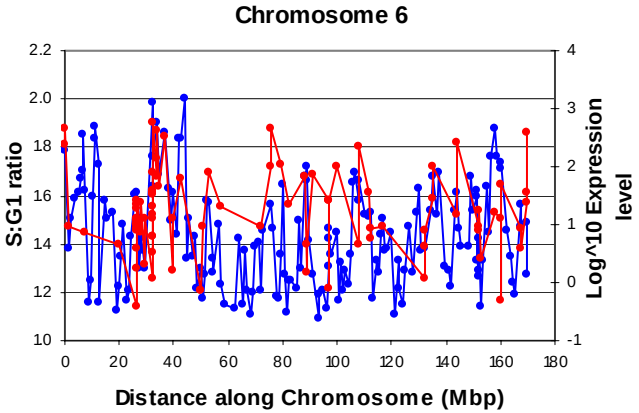


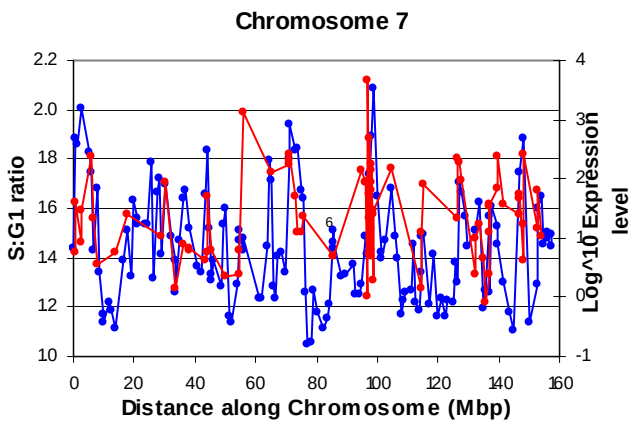


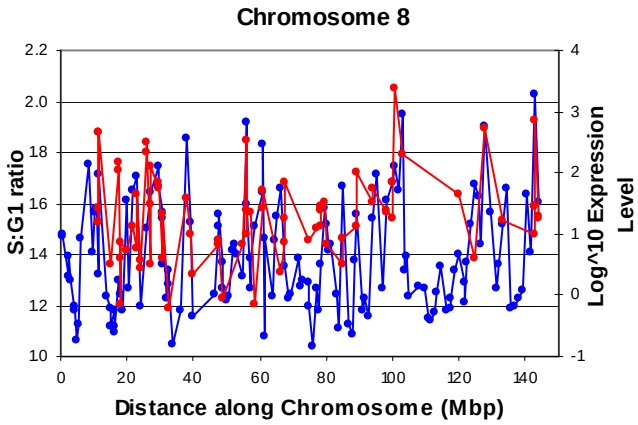


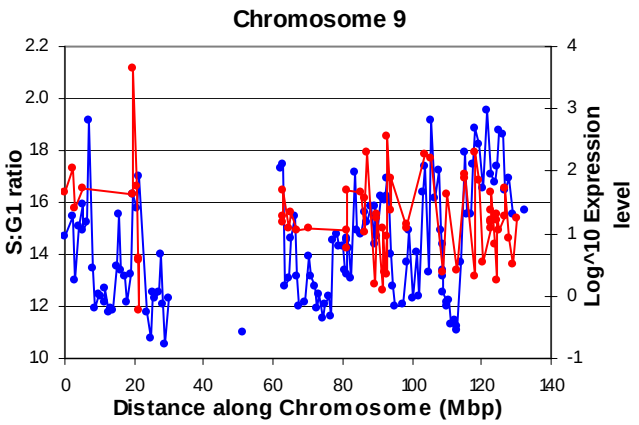


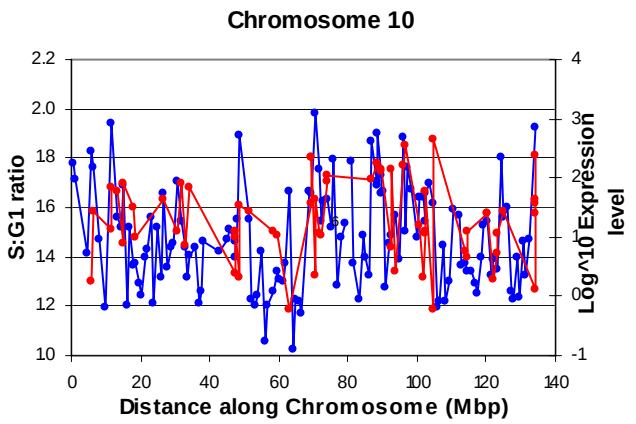


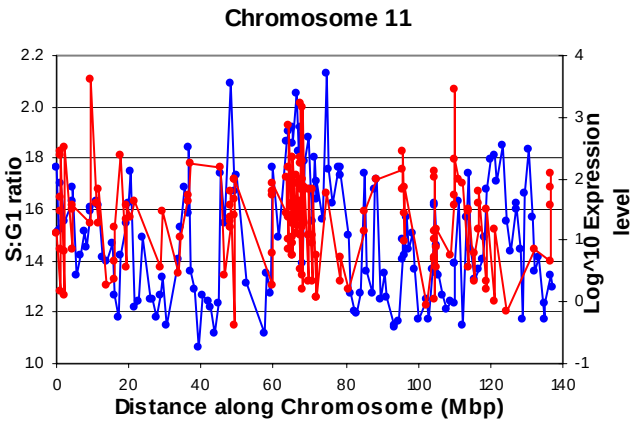


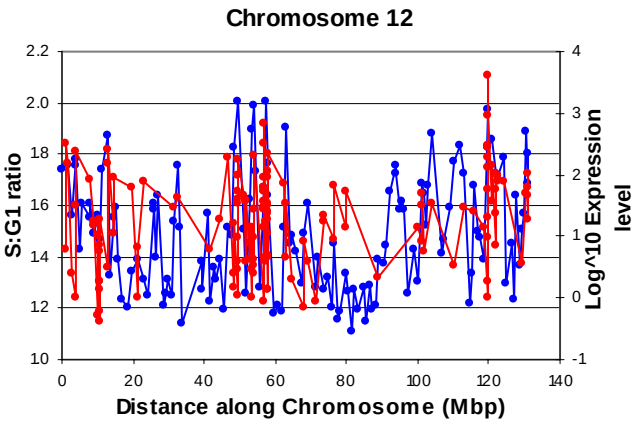

Chromosome 7
S:G1 ratio
Log^10 Expression level
Distance along Chromosome (Mbp)

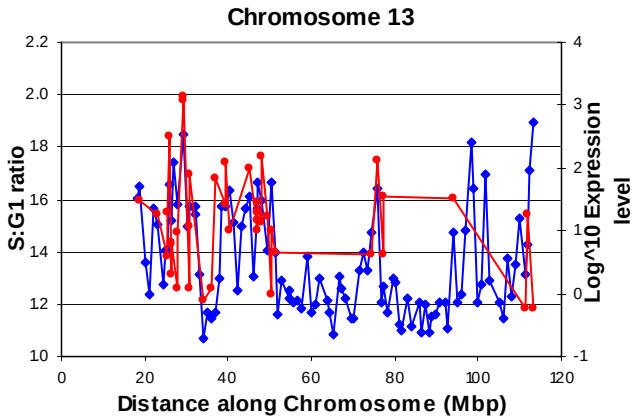

Chromosome 8
S:G1 ratio
Log^10 Expression Level
Distance along Chromosome (Mbp)

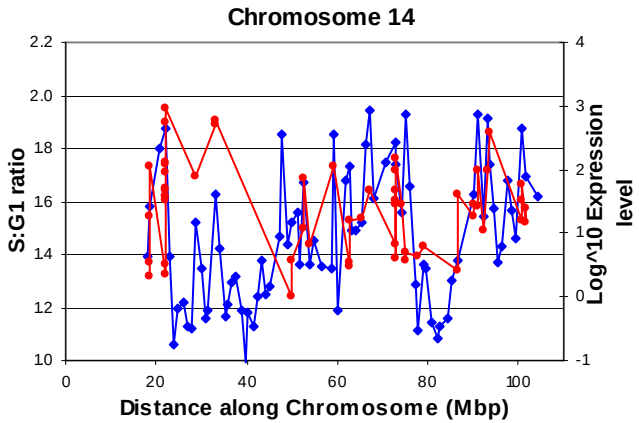

Chromosome 9
S:G1 ratio
Log^10 Expression level
Distance along Chromosome (Mbp)

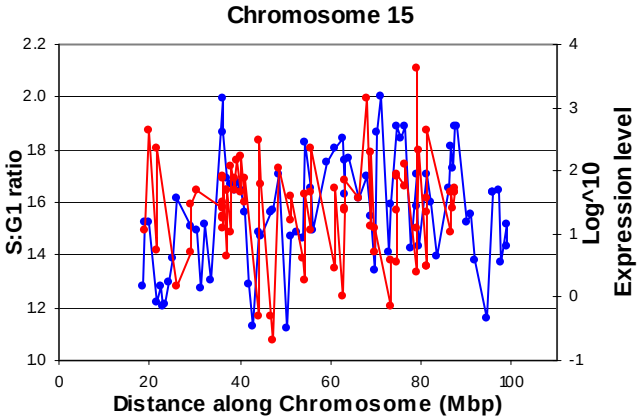

Chromosome 10
S:G1 ratio
Log^10 Expression level
Distance along Chromosome (Mbp)

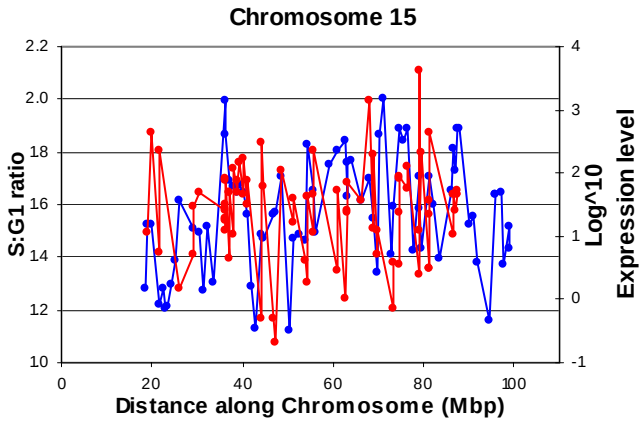

Chromosome 11
S:G1 ratio
Log^10 Expression level
Distance along Chromosome (Mbp)

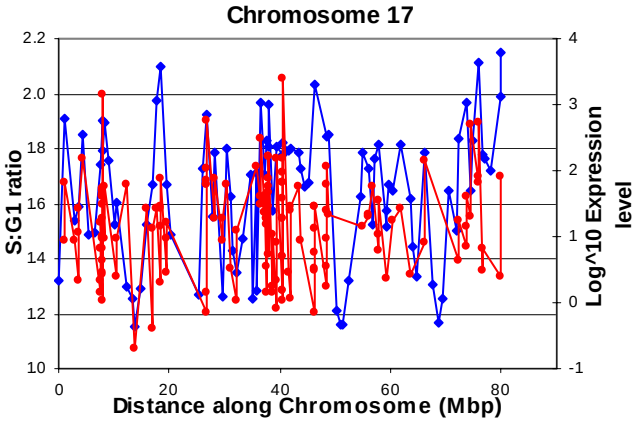

Chromosome 12
S:G1 ratio
Log^10 Expression level
Distance along Chromosome (Mbp)

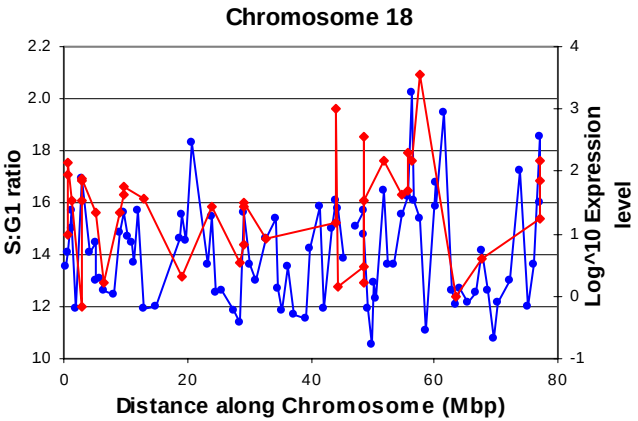

Chromosome 13
S:G1 ratio
Log^10 Expression level
Distance along Chromosome (Mbp)

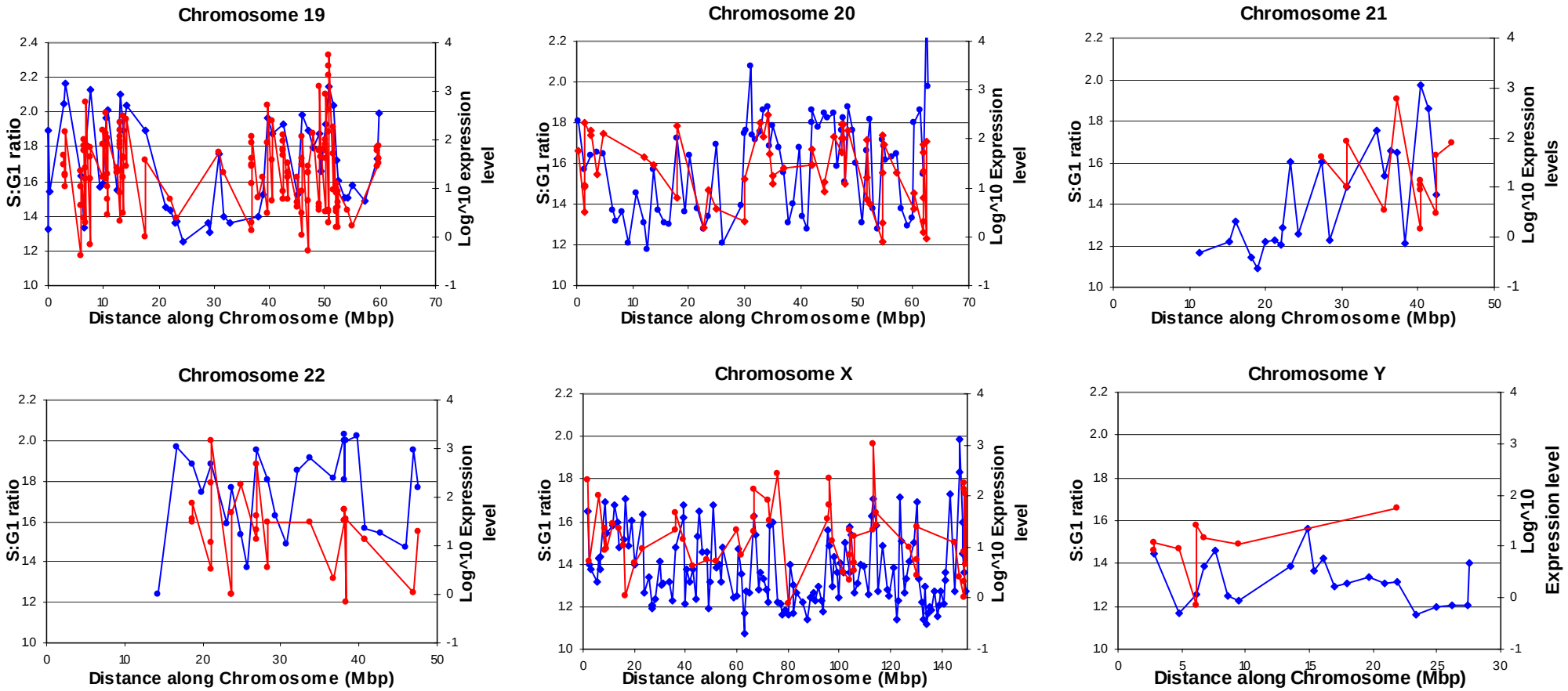

Chromosome 14
S:G1 ratio
Log^10 Expression level
Distance along Chromosome (Mbp)


Chromosome 15
S:G1 ratio
Log^10 Expression level
Distance along Chromosome (Mbp)


Chromosome 15
S:G1 ratio
Log^10 Expression level
Distance along Chromosome (Mbp)


Chromosome 17
S:G1 ratio
Log^10 Expression level
Distance along Chromosome (Mbp)


Chromosome 18
S:G1 ratio
Log^10 Expression level
Distance along Chromosome (Mbp)



Blue: Replication timing ratio. Red: Expression level of clones in the 1Mb set

Appendix 9: Chromosome 22 sequencing-clone information

9a: International Names for chromosome 22 clones

Accession No.	International Clone name	Sanger Clone name
AP000522	AP000522	cN4G1
AP000523	AP000523	c60H5
AP000524	AP000524	c70D1
AP000525	AP000525	cN14H11
AP000526	AP000526	cN64E9
AP000527	AP000527	cNN83F12
AP000528	AP000528	cN91G6
AP000529	AP000529	cN3G11
AP000530	AP000530	cN65E1
AP000531	AP000531	cN59E1
AP000532	AP000532	cN2F2
AP000533	AP000533	cN60G3
AP000534	AP000534	cN23H5
AP000535	AP000535	cN58F10
AP000536	AP000536	cN64C8
AP000537	AP000537	cN54B2
AP000538	AP000538	cN65B12
AP000539	AP000539	cN72E11
AP000540	AP000540	cN53D1
AP000541	AP000541	cN13E4
AP000542	AP000542	cN60D12
AP000543	AP000543	cN20H12
AP000544	AP000544	cN17H1
AP000545	AP000545	cN68B10
AP000546	AP000546	cN18E3
AP000547	KB-67B5	KB67B5
AP000365	KB-7G2	KB7G2
AC005301	AC005301	p15j16
AC007064	AC007064	p87o8
AC006548	AC006548	p20k14
AC006946	AC006946	p109l3
AC005300	AC005300	p143i13
AC005399	AC005399	p238m15
AC004019	AC004019	357f7
AC007666	AC007666	p273a17
AC006285	AC006285	p1087l10
AC016026	AC016026	b461k10
AC008101	XXbac-677f7	b677f7
AC008079	AC008079	bac519d21
AC008132	AC008132	pac995o6
AC008103	AC008103	pac699j1
AC007326	AC007326	p423
AC000095	AC000095	fF41C7
AC004461	AC004461	cN119F4

AC004462	AC004462	18c3
AC004471	AC004471	111f11
AC004463	AC004463	79h12
AC000081	AC000081	59c10
AC000094	AC000094	fF39E1
AC000085	AC000085	72f8
AC000092	AC000092	98c4
AC000079	AC000079	49c12
AC000068	AC000068	102g9
AC000087	AC000087	83c5
AC000088	AC000088	83e8
AC000082	AC000082	59f
AC000070	AC000070	105a
AC000086	AC000086	81h
AC000077	AC000077	31e
AC000067	AC000067	100h
AC000093	AC000093	carlaa
AC000091	AC000091	91c
AC000089	AC000089	89h
AC000076	AC000076	2h
AC000078	AC000078	33e
AC000090	AC000090	8c
AC000080	AC000080	56c
AC005663	AC005663	p888c9
AC006547	AC006547	p158l19
AC007663	AC007663	b444p24
AC007731	AC007731	b562f10
AC005500	AC005500	p52f6
AC004033	AC004033	p_M11
AC007050	AC007050	bac32
AC007308	AC007308	pac408
AC002470	AC002470	bK135H6
AC002472	AC002472	P_N5
AP000550	KB-1592A4	KB1592A4
AP000551	KB-876E2	KB876E2
AP000552	KB-1183D5	KB1183D5
AP000556	KB-1172D5	KB1172D5
AP000557	KB-1323B2	KB1323B2
AP000558	KB-1802C5	KB1802C5
AP000553	KB-1440D3	KB1440D3
AP000554	KB-666H9	KB666H9
AP000555	KB-1027C11	KB1027C11
D86995	D86995	cN109G12
D87019	D87019	cN86G7
D87012	D87012	cN61D6
D88268	D88268	cN47H9
D86993	D86993	cN23C6
D87004	D87004	cN4E7
D87022	D87022	cN88E1
D88271	D88271	cN114H4
D88269	D88269	cN33B6
D87000	D87000	cN30E12

D86996	D86996	cN23F1
D86989	D86989	cN110H3
D88270	D88270	cN123E1
D87003	D87003	cN2H8
D87018	D87018	cN80A10
D87016	D87016	cN75A1
D86999	D86999	cN22A12
D87010	D87010	cN35B9
D87009	D87009	KB288A10
D87011	D87011	cN50D10
D87013	D87013	cN63E9
D87014	D87014	cN61E11
D86991	D86991	cN29D3
D87002	D87002	cN31F3
D87006	D87006	cN52F2
D86994	D86994	cN102D1
D87007	D87007	cN48A11
D87015	D87015	cN68D6
D86998	D86998	cN24A12
D87021	D87021	cN84E4
D87024	D87024	cN92H4
D87020	D87020	cN9G6
D87023	D87023	cN9C5
D87017	D87017	cN75C12
AP000360	AP000360	cN81C12
AP000361	AP000361	cN8E4
AP000362	AP000362	cN75A12
AC000029	AC000029	bK865E9
AC000102	AC000102	bK60B5
AP000343	KB-282B12	kB282B12
AP000344	KB-1269D1	kB1269D1
AP000345	KB-208E9	KB208E9
AP000346	KB-1572G7	kB1572G7
AP000347	KB-113H7	KB113H7
AP000348	AP000348	cN27C7
AP000349	KB-1839H6	kB1839H6
AP000350	KB-1125A3	kB1125A3
AP000351	KB-226F1	KB226F1
AP000352	KB-1561E1	kB1561E1
AP000353	KB-318B8	kB318B8
AP000354	KB-1674E1	kB1674E1
AP000355	KB-1896H10	kB1896H10
AP000356	KB-1995A5	kB1995A5
AP000357	AP000357	cN95F10
AP000358	AP000358	cN110F4
AP000359	KB-63E7	KB63E7
AL049759	RP5-930L11	dJ930L11
AL050312	RP11-9F11	bA9F11
AL022323	CTA-243E7	bK243E7
Z99916	CTA-221G9	bK221G9
AL022332	RP3-462D8	dJ462D8
AL022324	CTA-246H3	bK246H3

AL008721	CTA-390C10	bK390C10
AL022329	CTA-407F11	bK407F11
AL080245	RP11-89B2	bA89B2
Z98949	CTA-125H2	bK125H2
AL079300	CTA-109D1	bK109D1
AL022337	CTA-796E4	bK796E4
AL080273	RP11-259P1	bA259P1
AL023513	RP1-268D13	dJ268D13
AL078460	RP3-341O5	dJ341O5
AL035044	RP1-40G4	dJ40G4p
Z99714	CTB-48E9	bK1048E9
Z95115	CTA-445C9	bK445C9
Z99774	CTA-373H7	bK373H7
Z95889	CTA-211A9	bK211A9
Z97353	RP1-90L6	dJ90L6
AL008638	CTA-992D9	bK992D9
AL021153	CTA-503F6	bK503F6
AL034386	RP5-1172A22	dJ1172A22
AL020994	CTA-929C8	bK929C8
AL049536	RP1-205F14	dJ205F14p
AL050402	RP11-46E17	bA46E17
AL133456	AL133456	dJ231P7p
AL390209	AL390209	bK437G10
AL121885	RP11-375H17	bA375H17
AL031591	RP3-353E16	dJ353E16
AL033538	RP3-477H23	dJ477H23
AL035453	SC22CB-42E1	cB42E1
AL050313	CTA-754D9	bK754D9
AL035397	RP6-45P1	dA45P1
AL023281	CTA-544A11	bK544A11
AL008722	CTA-732E4	bK732E4
AL080241	RP11-541J16	bA541J16
AL118497	RP11-329J7	bA329J7
AL121825	RP11-436C9	bA436C9
AL117330	RP11-444G7	bA444G7
AL023494	RP3-366L4	dJ366L4
Z93930	CTA-292E10	bK292E10
AL031596	RP4-745C22	dJ745C22
Z95113	CTA-175E3	bK175E3
AL021393	CTA-747E2	bK747E2
Z95116	CTA-57G9	bK57G9
AL031186	CTA-984G1	bK984G1
AC000026	AC000026	bK58B8
AC000041	AC000041	cE42H1
AC000035	AC000035	cN47G11
AC005529	AC005529	bK256D12
AC004882	RP1-76B20	dJ76B20
Z82171	SC22CB-11B7	cB11B7
AC004819	RP1-15I23	dJ15I23
AC003681	RP3-394A18	dJ394A18
AC003071	CTA-85E5	bK85E5
AC002378	RP3-438O4	dJ438O4

AC004264	RP1-102K2	dJ102K2
AC004997	RP1-130H16	dJ130H16
AC004832	RP4-539M6	dJ539M6
AC005006	RP1-56J10	dJ56J10
AC003072	CTA-963H5	bK963H5
AL079299	RP11-492A7	bA492A7
AL022336	RP1-78F24	dJ78F24
AC004542	RP3-430N8	dJ430N8
AC005233	RP5-1198O21	dJ1198O21
AC005005	RP3-412A9	dJ412A9
AC002073	RP3-515N1	dJ515N1
AC005003	RP3-400N23	dJ400N23
AL096702	RP11-254F5	bA254F5
AL096701	RP11-247I13	bA247I13
AL109802	RP11-163M1	bA163M1
AL096768	RP5-858B16	dJ858B16
AL031255	RP4-694E4	dJ694E4
AC005004	RP3-403E2	dJ403E2
AL022331	CTA-440B3	bK440B3
Z82190	RP1-180M12	dJ180M12
Z83856	LL22NC03-113A11	cN113A11
Z82248	LL22NC03-44A4	cN44A4
AL008719	CTA-342B11	bK342B11
Z74021	SC22CB-1E7	cB1E7
Z80998	SC22CB-36G12	cB36G12
Z83849	CITF22-65B7	fF65B7
AL022321	RP1-20O8	dJ20O8
Z83839	RP1-127L4	dJ127L4
AL008723	RP1-90G24	dJ90G24
AL021937	RP1-149A16	dJ149A16
AL035068	RP1-116G19	dJ116G19
Z71183	LL22NC03-28H9	cN28H9
Z82181	LL22NC01-86D10	cE86D10
Z80902	LL22NC03-80H12	cN80H12
AL021452	CTA-285F3	bK285F3
Z82246	LL22NC03-104C7	cN104C7
Z75744	LL22NC03-117B5	cN117B5
Z81309	LL22NC01-92H8	cE92H8
Z69714	LL22NC03-37F10	cN37F10
Z72521	LL22NC03-29F4	cN29F4
Z72520	LL22NC03-19H5	cN19H5
Z73495	LL22NC01-116C6	cE116C6
AL023282	CTA-766E1	bK766E1
Z73979	SC22CB-10B1	cB10B1
Z98256	RP1-309I22	dJ309I22
AL031592	CTA-366B10	bK366B10
Z83846	CTA-415G2	bK415G2
Z82198	RP1-302D9	dJ302D9
AL008630	CTA-282F2	bK282F2
Z82173	SC22CB-1D7	cB1D7
AL008715	CTC-216H12	bK1216H12
AL023577	CTA-566G5	bK566G5

Z82179	LL22NC01-140F8	cE140F8
Z73421	LL22NC03-37D7	cN37D7
AL133451	LL22NC03-120B6	cN120B6
Z69943	LL22NC03-4F11	cN4F11
AL008640	SC22CB-33D11	cB33D11
Z70288	LL22NC01-78G1	cE78G1
Z97354	LL22NC03-117F11	cN117F11
Z69042	LL22NC01-95B1	cE95B1
Z68287	LL22NC03-38E12	cN38E12
Z76736	RP1-75E8	dJ75E8
Z69713	LL22NC03-20A6	cN20A6
Z68324	LL22NC03-7A10	cN7A10
AL096754	LL22NC03-2E9	cN2E9
Z49866	LL22NC03-73A10	cN73A10
Z54073	LL22NC03-13E1	cN13E1
Z77853	LL22NC03-53F3	cN53F3
Z69715	LL22NC03-74G7	cN74G7
Z73429	LL22NC03-32F9	cN32F9
Z69925	LL22NC03-116A5	cN116A5
Z68223	LL22NC01-110C7	cE110C7
Z83852	SC22CB-49C12	cB49C12
Z82182	LL22NC01-90C2	cE90C2
Z99704	LL22NC01-75B8	cE75B8
Z74581	LL22NC01-127C11	cE127C11
AL008641	LL22NC03-100B10	cN100B10
Z69707	LL22NC01-95B9	cE95B9
Z68285	LL22NC03-11D4	cN11D4
AL022338	LL22NC01-82F7	cE82F7
Z82250	LL22NC03-86D4	cN86D4
AL008717	CTA-221H1	bK221H1
Z68323	LL22NC03-13G6	cN13G6
Z68288	LL22NC03-5E4	cN5E4
Z68325	LL22NC03-98E6	cN98E6
Z69712	LL22NC03-12G10	cN12G10
Z68754	LL22NC01-78H10	cE78H10
Z68758	LL22NC03-85E10	cN85E10
Z50860	LL22NC03-76A1	cN76A1
AL020992	CTA-363A12	bK363A12
Z68224	LL22NC01-129H9	cE129H9
AL049750	LL22NC01-141E2	cE141E2
Z69907	LL22NC03-22D1	cN22D1
AL031001	RP1-281O16	dJ281O16
AL021877	RP1-101G11	dJ101G11
Z82196	RP1-288L1	dJ288L1
AL024495	RP3-404L14	dJ404L14
Z82194	RP1-272J12	dJ272J12
Z83853	SC22CB-109E1	cB109E1
AL024494	RP1-215F16	dJ215F16
Z99755	CTA-714B7	bK714B7
AL031300	RP3-323A16	dJ323A16
AL008635	RP3-510H16	dJ510H16
Z82244	CTA-286B10	bK286B10

AL009049	RP5-824I19	dJ824I19
AL022334	RP4-569D19	dJ569D19
AL049747	CITF22-62D4	fF62D4
Z79996	SC22CB-33F2	cB33F2
AL049748	RP1-41P2	dJ41P2
AL079295	RP1-106I20	dJ106I20
Z82217	RP1-78B3	dJ78B3
Z95114	CTA-212A2	bK212A2
AL031426	CTA-191D12	bK191D12
Z82215	RP1-68O2	dJ68O2
AL022302	RP4-633O19	dJ633O19
AL022313	RP5-1119A7	dJ1119A7
AL031845	CTA-566H6	bK566H6
Z70289	CITF22-4G12	fF4G12
AL049749	RP1-293L6	dJ293L6
Z80897	LL22NC01-132D12	cE132D12
Z82184	CITF22-126G10	fF126G10
Z82185	CITF22-24E5	fF24E5
AL008637	CTA-833B7	bK833B7
AL133392	CITF22-45C1	fF45C1
Z82180	LL22NC01-81G9	cE81G9
Z73420	LL22NC01-146D10	cE146D10
AL022314	RP5-1170K4	dJ1170K4
Z82188	RP1-151B14	dJ151B14
Z94160	RP1-63G5	dJ63G5
AL049850	RP5-889J22	dJ889J22
Z93096	CTA-390B3	bK390B3
AL022315	RP5-1177I5	dJ1177I5
AL109980	RP4-697G8	dJ697G8
AL035496	RP3-437O22	dJ437O22
Z83844	RP1-37E16	dJ37E16
Z97630	RP3-466N1	dJ466N1
AL022311	RP5-1014D13	dJ1014D13
AL031587	RP5-1039K5	dJ1039K5
AL022322	CTA-228A9	bK228A9
AL021977	CTA-447C4	bK447C4
AL020993	RP1-5O6	dJ5O6
Z98749	RP3-449O17	dJ449O17
Z97056	RP3-434P1	dJ434P1
AL022320	RP1-199H16	dJ199H16
AL035495	RP1-319F24	dJ319F24
AL021707	RP3-508I15	dJ508I15
AL021806	RP4-779B17	dJ779B17
AL008583	RP3-327J16	dJ327J16
AL022318	CTA-150C2	bK150C2
AL031846	RP4-742C19	dJ742C19
Z81010	LL22NC03-10C3	cN10C3
AL031590	CTA-232D4	bK232D4
AL022326	RP3-333H23	dJ333H23
Z83845	RP3-407F17	dJ407F17
AL022312	RP5-1104E15	dJ1104E15
AL008716	CTA-206C7	bK206C7

AL022319	RP1-172B20	dJ172B20
AL022353	CTA-352E11	bK352E11
Z82206	RP3-370M22	dJ370M22
Z83847	RP3-496C20	dJ496C20
AL033547	RP3-340K22	dJ340K22
AL031589	RP6-11O7	dA11O7
Z93783	RP3-377F16	dJ377F16
AL022238	RP5-1042K10	dJ1042K10
AL031594	RP4-591N18	dJ591N18
Z86090	CTA-229A8	bK229A8
AL096703	RP4-735G18	dJ735G18
AL049764	RP3-362J20	dJ362J20
Z98048	RP3-408N23	dJ408N23
AL035450	RP5-1057D18	dJ1057D18
AL080242	RP11-554C12	bA554C12
AL080243	RP11-12M9	bA12M9
AL096765	RP11-422A16	bA422A16
AL035658	RP1-85F18	dJ85F18
AL035681	RP4-756G23	dJ756G23
AL035659	RP5-979N1	dJ979N1
AL008582	CTA-223H9	bK223H9
AL023553	RP3-347H13	dJ347H13
Z83840	CTA-216E10	bK216E10
AL023879	CTA-109G6	bK109G6
AL021453	RP5-821D11	dJ821D11
Z99716	CTA-250D10	bK250D10
Z82192	RP1-186O1	dJ186O1
AL021878	RP1-257I20	dJ257I20
AL031346	RP4-597B2	dJ597B2
Z83851	CTA-989H11	bK989H11
AL022316	CTA-126B4	bK126B4
AL035418	RP1-141I3	dJ141I3
Z93241	RP1-222E13	dJ222E13
Z82176	SC22CB-33B7	cB33B7
AL049757	RP1-47A17	dJ47A17
AL049758	RP3-437M21	dJ437M21
AL022476	RP3-323M22	dJ323M22
AL022237	CTB-191B2	bK1191B2
Z82214	RP3-526I14	dJ526I14
Z99756	RP1-100N22	dJ100N22
AL096761	RP4-754E20	dJ754E20
Z82172	SC22CB-13C9	cB13C9
AL118498	RP1-185D5	dJ185D5
AL031843	RP1-246D7	dJ246D7
Z82201	RP3-345P10	dJ345P10
AL023801	RP4-786D3	dJ786D3
Z97055	RP3-388M5	dJ388M5
AL023654	RP4-549K18	dJ549K18
AL035398	RP4-796I17	dJ796I17
Z82178	SC22CB-79B4	cB79B4
Z82174	SC22CB-20F6	cB20F6
AL033543	CTA-414D7	bK414D7

AL031595	RP4-671O14	dJ671O14
AL022339	LL22NC03-75B3	cN75B3
AL591914	CITF22-4H11	fF4H11
AL671760	CITF22-11D1	fF11D1
AL929500	CITF22-57B10	fF57B10
Z85994	RP1-32I10	dJ32I10
Z81308	CTA-397C4	bK397C4
AL023973	RP5-1033E15	dJ1033E15
Z75407	LL22NC03-128A12	cN128A12
AL022317	RP1-140L1	dJ140L1
AL022333	RP3-474I12	dJ474I12
Z98743	RP1-181C9	dJ181C9
Z93244	CTA-116F5	bK116F5
Z83838	RP1-127B20	dJ127B20
AL079301	RP4-753M9	dJ753M9
Z82243	CTA-217C2	bK217C2
AL008718	CTA-268H5	bK268H5
AL021391	RP1-102D24	dJ102D24
Z98047	RP1-162H14	dJ162H14
Z95331	CTA-941F9	bK941F9
Z93784	RP3-398C22	dJ398C22
Z84478	RP1-37M3	dJ37M3
AL049811	RP11-140I15	bA140I15
AL929387	RP11-398F12	bA398F12
Z94161	LL22NC03-102C10	cN102C10
AL049856	RP4-695O20	dJ695O20
AL078611	SC22cB-5E3	cB5E3
AL031034	LL22NC03-98G1	cN98G1
Z93024	LL22NC03-5H6	cN5H6
AL031844	RP3-361H15	dJ361H15
AL031588	RP5-1163J1	dJ1163J1
AL031597	RP5-996D20	dJ996D20
AL021392	RP3-439F8	dJ439F8
AL096766	RP6-59H18	dA59H18
AL118516	CTA-29F11	bK29F11
AL078642	RP5-917C11	dJ917C11
Z97351	RP1-116M15	dJ116M15
Z82187	CITF22-64F4	fF64F4
U51561	LLcos-79E2	cN79E2
AL023576	CTA-358H9	bK358H9
Z80896	RP1-67C13	dJ67C13
Z79999	CITF22-111A3	fF111A3
AL023733	SC22cB-58F5	cB58F5
Z81000	CITF22-67D6	fF67D6
AL096755	RP3-477J10	dJ477J10
U51559	LLcos-65D1	cN65D1
AL096756	RP3-477J10	dJ494G10
Z82183	LL22NC01-98F6	cE98F6
Z83836	RP1-111J24	dJ111J24
Z82186	CITF22-49D8	fF49D8
AL035069	RP3-404P13	dJ404P13
Z84496	LL22NC03-75H12	cN75H12

AL021306	CTB-109B5	bK1109B5
Z80999	LL22NC01-140G5	cE140G5
Z73416	LL22NC03-114B2	cN114B2
AL033544	LL22NC01-107C5	cE107C5
AL049568	SC22CB-23F1	cB23F1
Z80901	LL22NC03-119A7	cN119A7
AC000034	AC000034	cN119A4
Z83854	SC22CB-44H3	cB44H3
AL117329	RP11-191L9	bA191L9
AL121580	RP13-455A7	bB455A7
AL110122	CTA-280A3	bK280A3
Z73963	LL22NC03-62C4	cN62C4
AL110121	CTA-280A3	bK280A3
Z70688	LL22NC03-27C5	cN27C5
Z81002	LL22NC03-53A9	cN53A9
Z80772	CITF22-37F6	fF37F6
AL118553	CITF22-91B7	fF91B7
Z78421	LL22NC03-121E8	cN121E8
AL118554	CITF22-91B7	fF91B7
AL050314	RP1-100G10	dJ100G10
AL008720	CTA-343C1	bK343C1
Z72006	LL22NC03-69F4	cN69F4
Z82249	LL22NC03-4D6	cN4D6
AL078640	RP11-536P6	bA536P6
Z84468	CTA-299D3	bK299D3
AL096853	CITF22-96H12	fF96H12
AL096843	RP11-262A13	bA262A13
Z83837	CITF22-113D11	fF113D11
Z83855	LL22NC03-104C4	cN104C4
AL078607	RP4-619N21	dJ619N21
AL954742	CITF22-114E11	fF114E11
AL078613	RP11-354I12	bA354I12
AL078622	RP5-925J7	dJ925J7
AL078632	RP11-255N20	bA255N20
AL049773	RP5-1061O18	dJ1061O18
Z94162	LL22NC03-21F1	cN21F1
Z82202	RP1-34P24	dJ34P24
Z82189	RP1-170A21	dJ170A21
AL008636	CTA-722E9	bK722E9
AL031593	RP4-566L20	dJ566L20
Z97192	RP1-29C18	dJ29C18
AL023802	RP5-983L19	dJ983L19
Z98885	RP3-522J7	dJ522J7
AL117328	RP11-494O16	bA494O16
AL080240	RP11-232E17	bA232E17
AL022327	RP3-355C18	dJ355C18
AL034546	RP5-898I4	dJ898I4
AL022328	RP3-402G11	dJ402G11
AL954743	CITF22-49G11	fF49G11
AL671545	CITF22-53E7	fF53Et
AL096767	RP4-579N16	dJ579N16
Z94802	CTA-999D10	bK999D10

U62317	U62317	bK384D8
Z82168	RP1-104C13	dJ104C13
Z82245	CTA-799F10	bK799F10
AC000050	AC000050	cN66C4
AC000036	AC000036	cN85A3
AC002056	AC002056	cN94H12
AC002055	AC002055	cN1G3

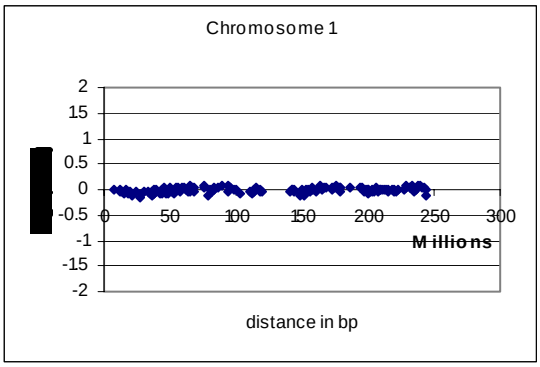
9b: Clone Libraries

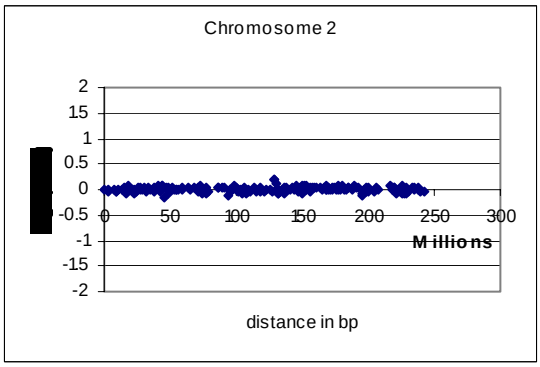
Table adapted from, http://www.sanger.ac.uk/HGP/methods/mapping/info/lib-details.shtml

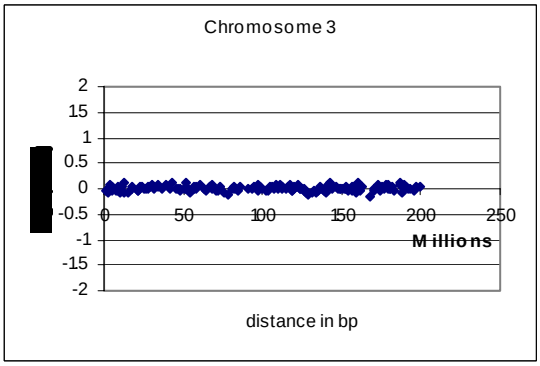
Library Type	**Library**	**Library code**	**Antibiotic**
Cosmid	Sc22cB, Sanger flow sorted chromosome 22	cB	Kanamycin 30μg/ml
Cosmid	LL22NC01 “E”, Paris flow sorted chromosome 22	cE	Kanamycin 30μg/ml
Cosmid	LL22NC03 “N”, Lawerence Livermore flow sorted chromosome 22	cN	Kanamycin 30μg/ml
Fosmid	CITF22, Caltech flow sorted chromosome 22 fosmid library	fF	Chloramphenicol 25μg/ml
RPCI Human PAC	RPCI-1-5, de Jong whole genome male PAC library	dJ	Kanamycin 25μg/ml
RPCI Human PAC	RPCI-6, de Jong whole genome female PAC library	dA	Kanamycin 25μg/ml
BAC	CIT978SK (CTA, CTB and CTC) Caltech whole genome BAC library	bK	Chloramphenicol 12.5μg/ml
RPCI Human BAC	RPCI-11, Whole genome male BAC library	bA	Chloramphenicol 25μg/ml
RPCI Human BAC	RPCI-13, Whole genome female BAC library	bB	Chloramphenicol 25μg/ml

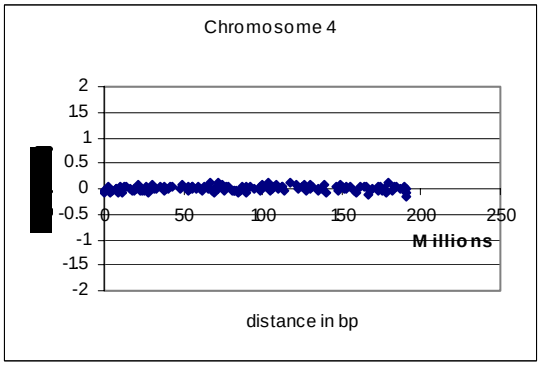
Appendix 10: 1Mb profiles of patients with DiGeorge phenotype and no 22q11 deletion.

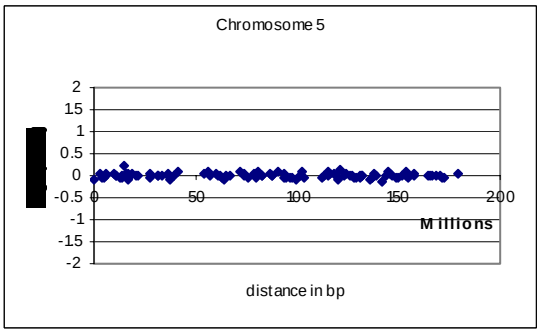
Patient 1:

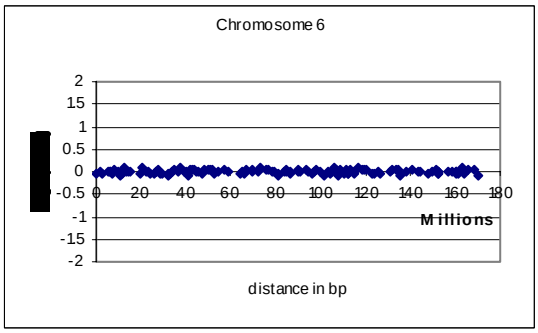


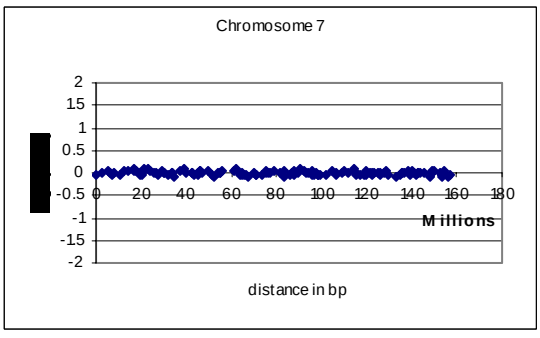


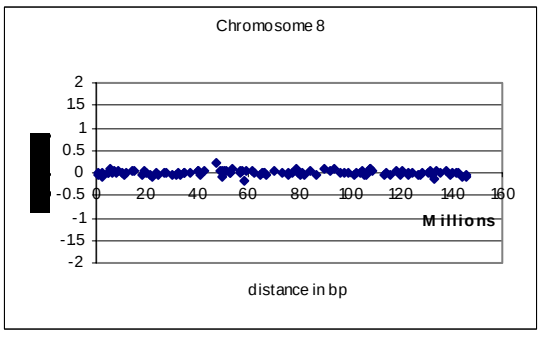


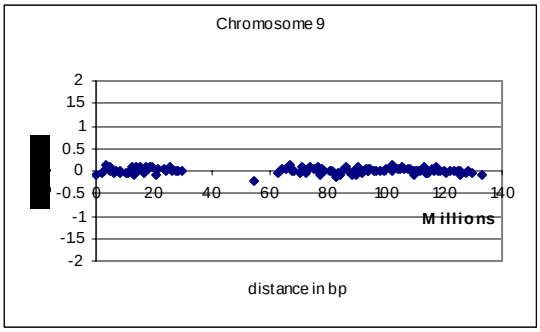


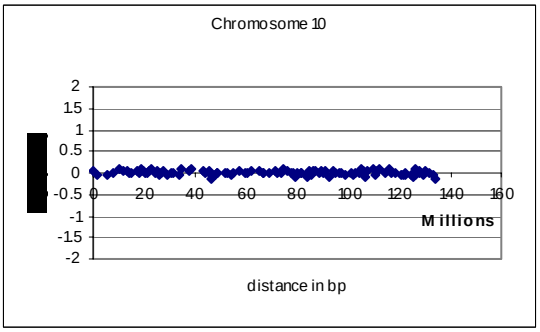


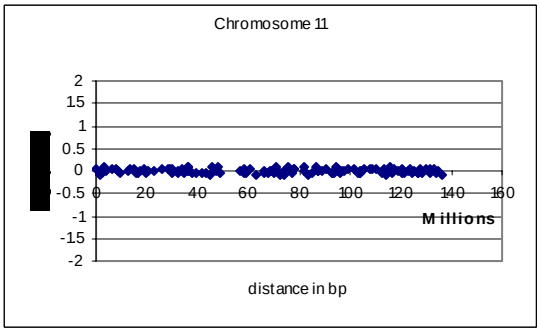


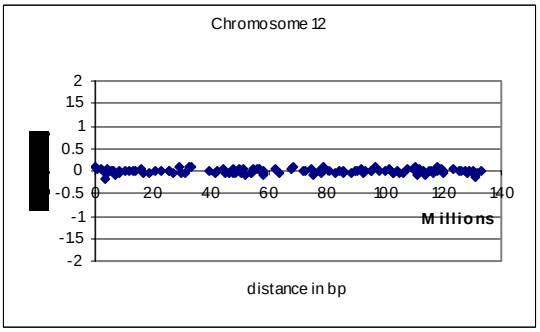


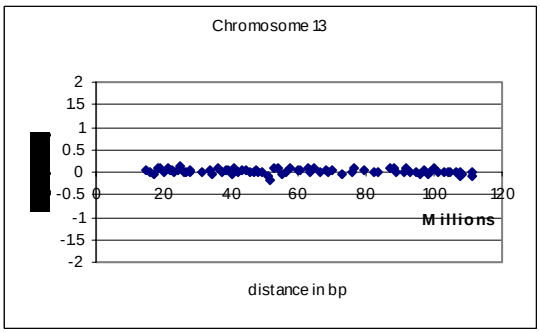


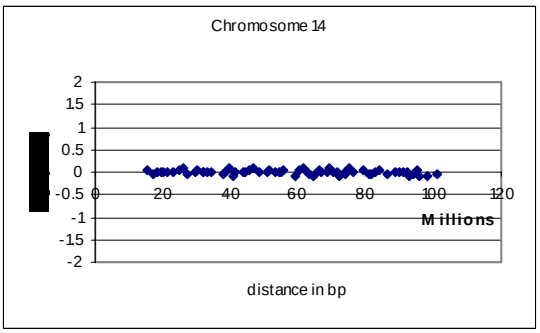

Chromosome 9
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0 20 40 60 80 100 120 140
Millions
distance in bp

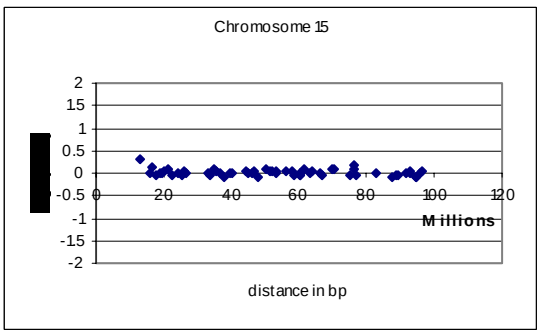

Chromosome 10
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0 20 40 60 80 100 120 140 160
Millions
distance in bp

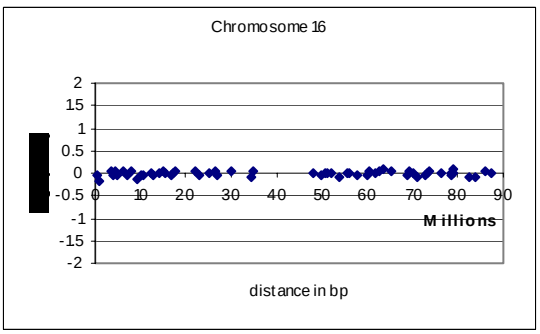

Chromosome 11
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0 20 40 60 80 100 120 140 160
Millions
distance in bp

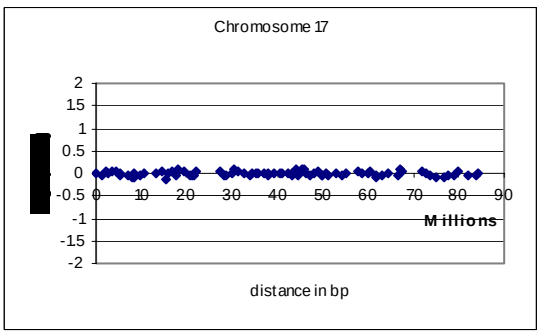

Chromosome 12
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0 20 40 60 80 100 120 140
Millions
distance in bp

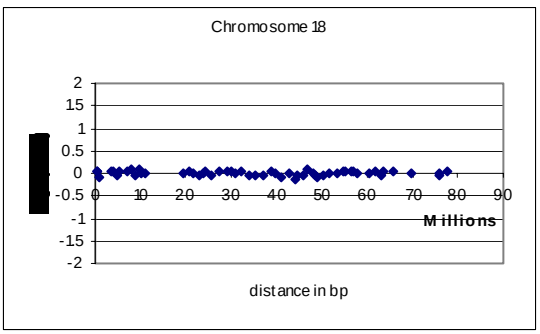

Chromosome 13
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0 20 40 60 80 100 120
Millions
distance in bp

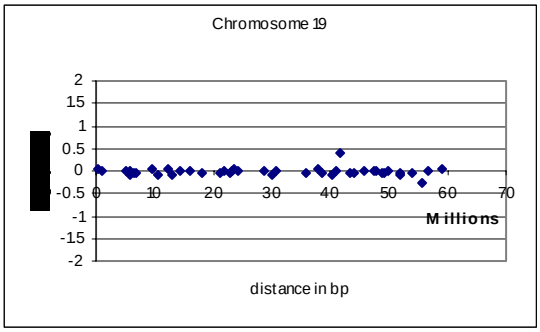

Chromosome 14
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0 20 40 60 80 100 120
Millions
distance in bp

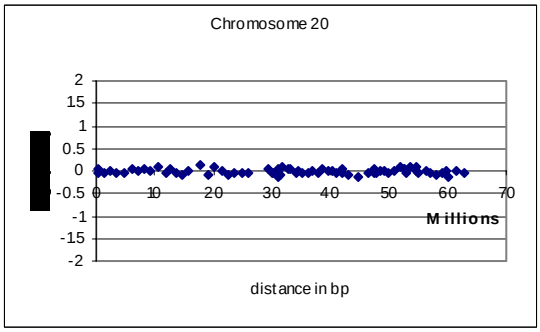

Chromosome 15
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0 20 40 60 80 100 120
Millions
distance in bp

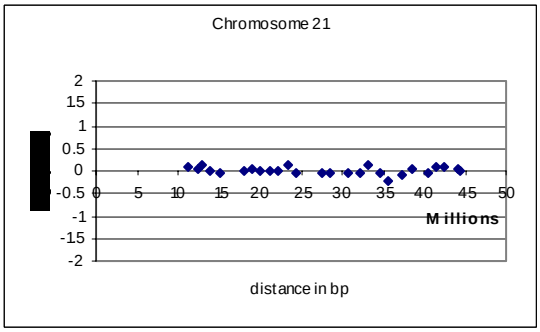

Chromosome 16
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0 10 20 30 40 50 60 70 80 90
Millions
distance in bp

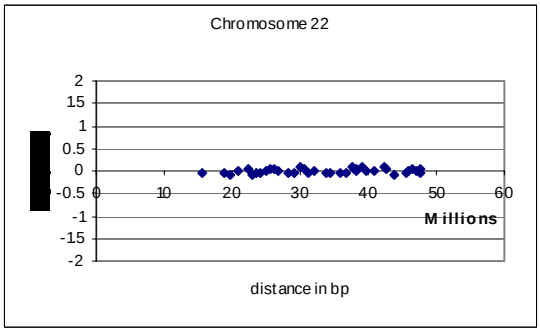

Chromosome 17
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0 10 20 30 40 50 60 70 80 90
Millions
distance in bp

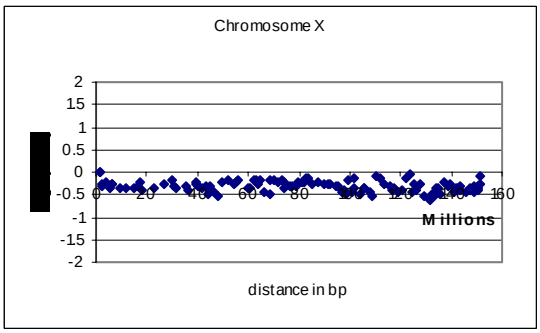

Chromosome 18
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0 10 20 30 40 50 60 70 80 90
Millions
distance in bp

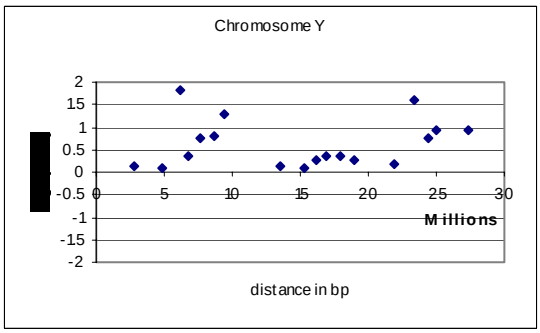

Chromosome 19
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
Millions
distance in bp


Chromosome 20
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
Millions
distance in bp

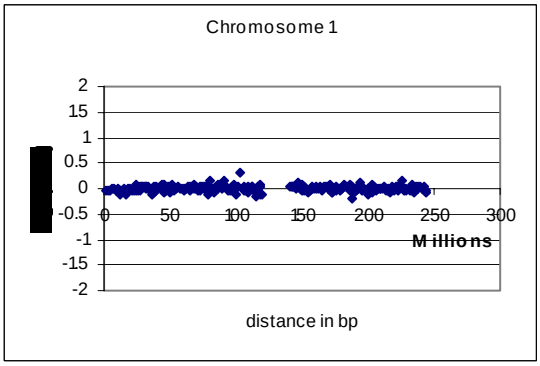

Chromosome 21
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
5
10
15
20
25
30
35
40
45
50
Millions
distance in bp

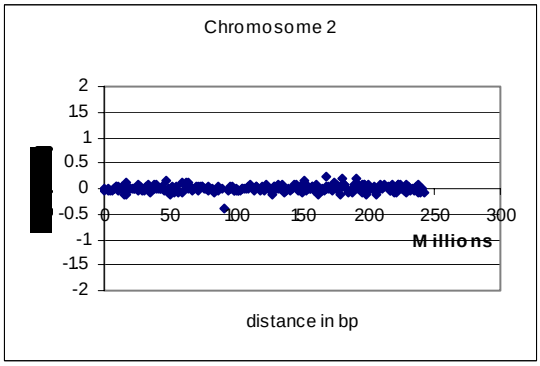

Chromosome 22
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
Millions
distance in bp

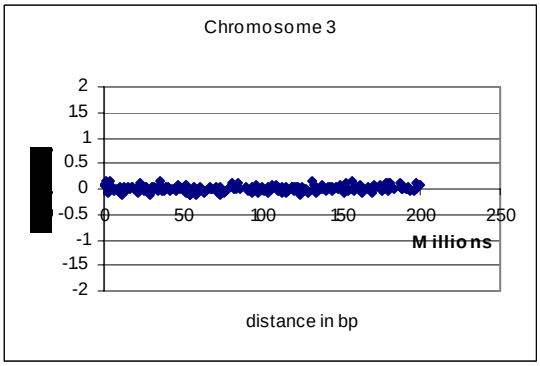

Chromosome X
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
Millions
distance in bp

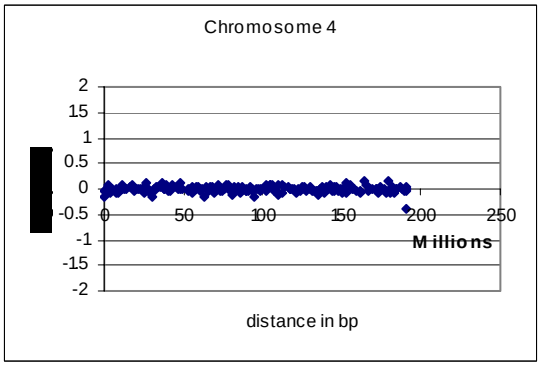

Chromosome Y
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
5
10
15
20
25
30
Millions
distance in bp

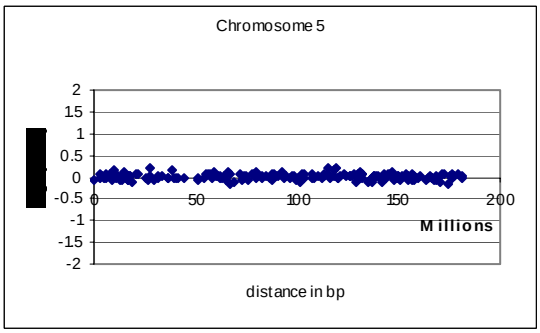
Patient 2:

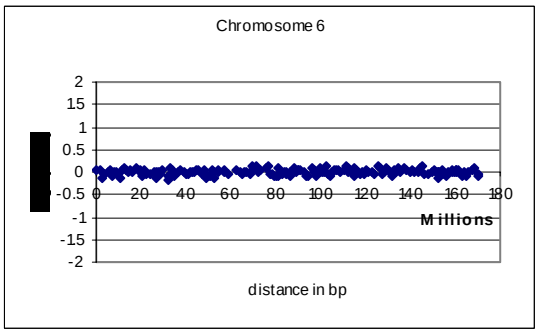

Chromosome 1
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
50
100
150
200
250
300
Millions
distance in bp

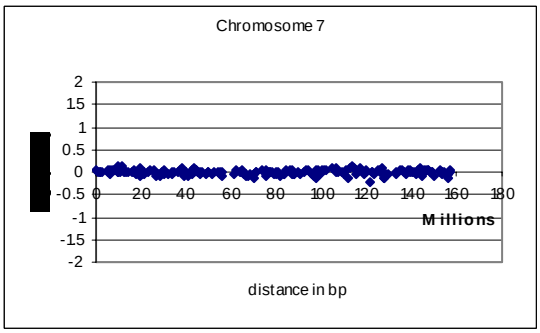

Chromosome 2
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
50
100
150
200
250
300
Millions
distance in bp

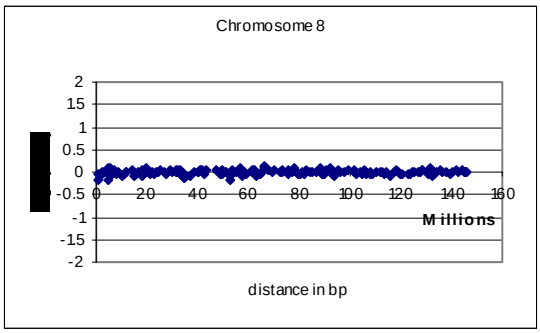

Chromosome 3
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
50
100
150
200
250
Millions
distance in bp

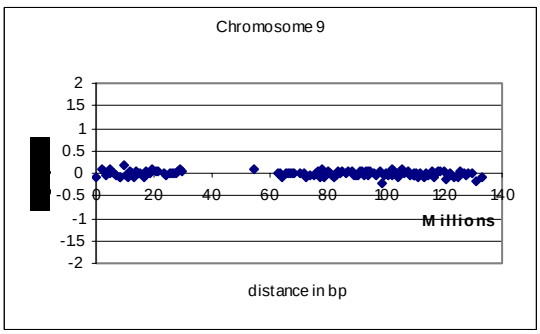

Chromosome 4
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
50
100
150
200
250
Millions
distance in bp

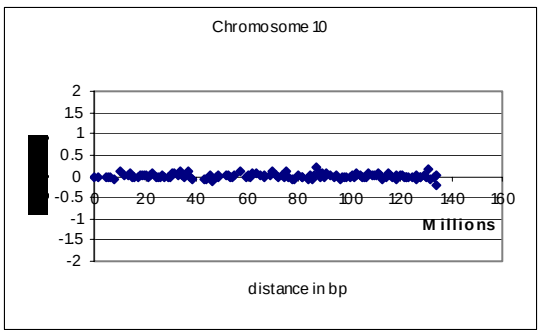

Chromosome 5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
50
100
150
200
Millions
distance in bp

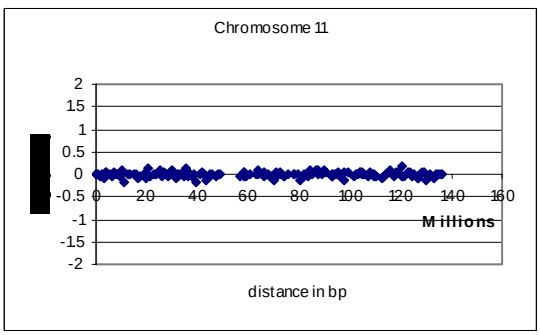

Chromosome 6
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
180
Millions
distance in bp

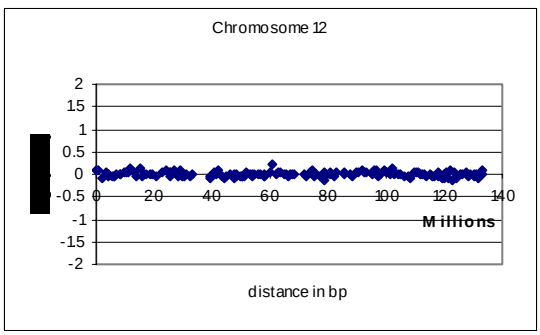

Chromosome 7
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
180
Millions
distance in bp

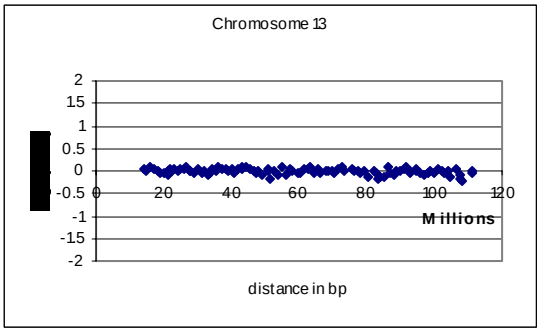

Chromosome 8
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
Millions
distance in bp

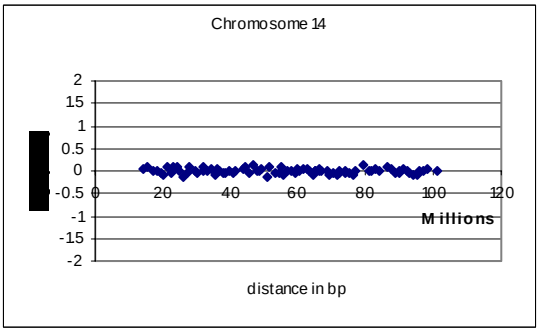

Chromosome 9
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
Millions
distance in bp

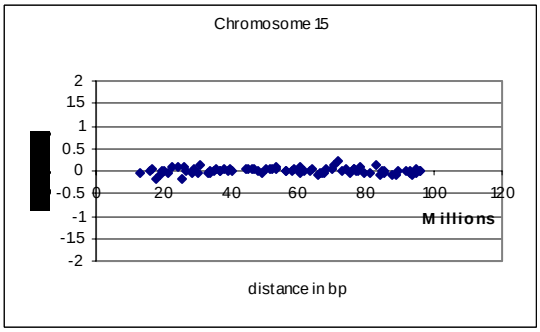

Chromosome 10
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
Millions
distance in bp

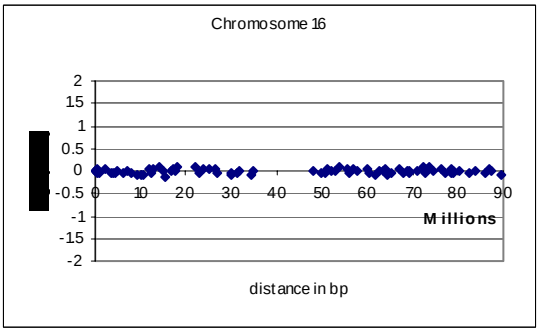

Chromosome 11
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
Millions
distance in bp

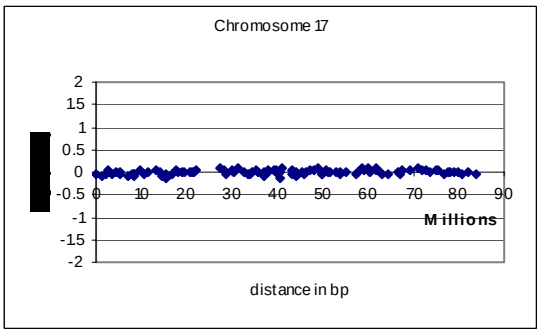

Chromosome 12
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
Millions
distance in bp

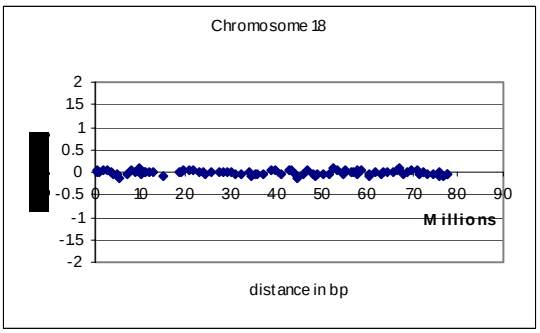

Chromosome 13
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

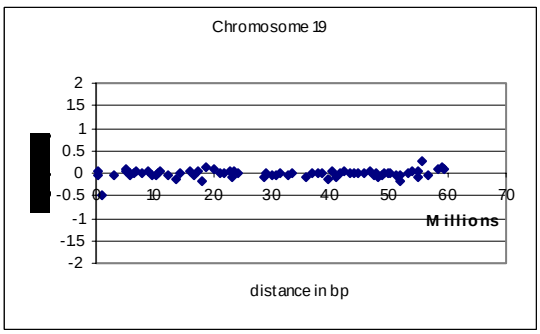

Chromosome 14
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

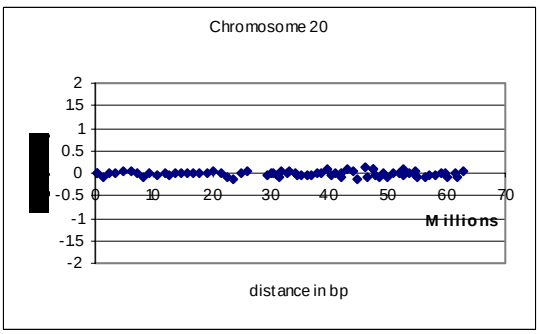

Chromosome 15
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

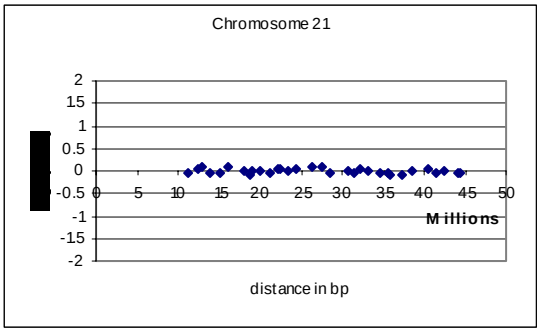

Chromosome 16
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp

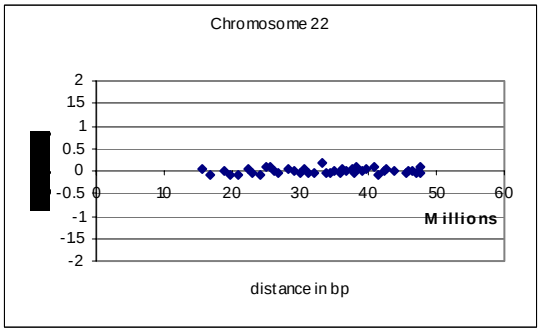

Chromosome 17
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp

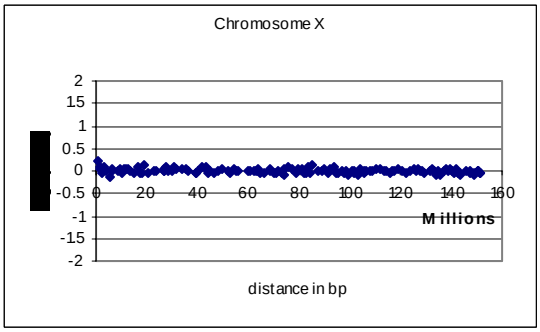

Chromosome 18
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp

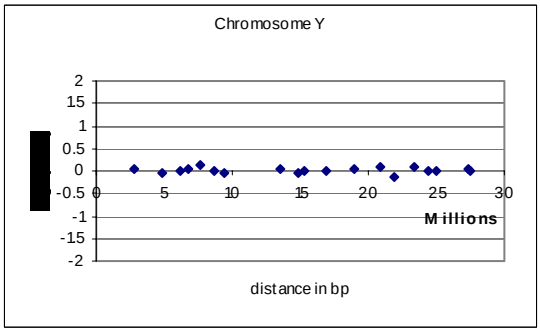

Chromosome 19
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
Millions
distance in bp


Chromosome 20
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
Millions
distance in bp

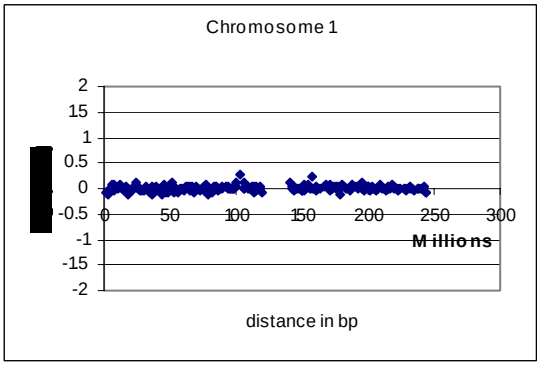

Chromosome 21
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0 5 10 15 20 25 30 35 40 45 50
Millions
distance in bp

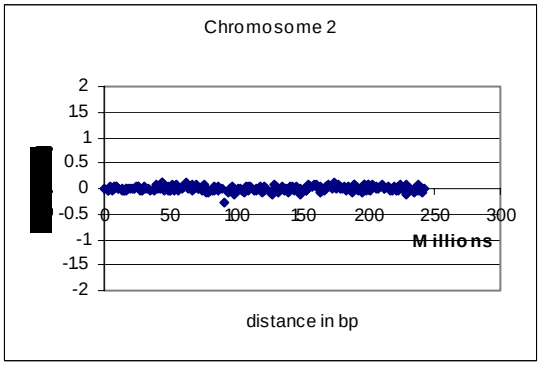

Chromosome 22
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0 10 20 30 40 50 60
Millions
distance in bp

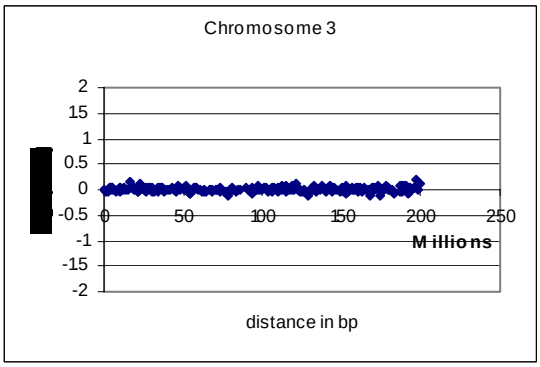

Chromosome X
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0 20 40 60 80 100 120 140 160
Millions
distance in bp

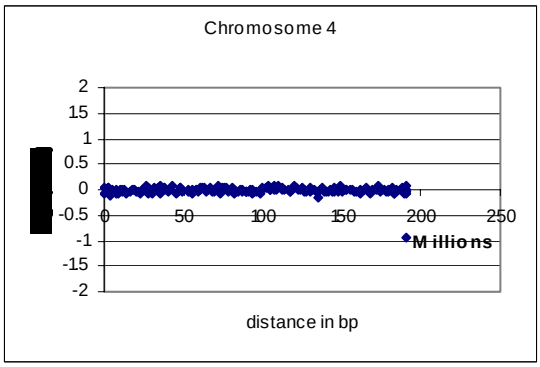

Chromosome Y
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0 5 10 15 20 25 30
Millions
distance in bp

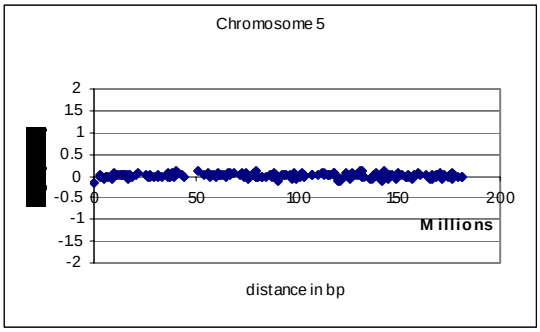
Patient 3:

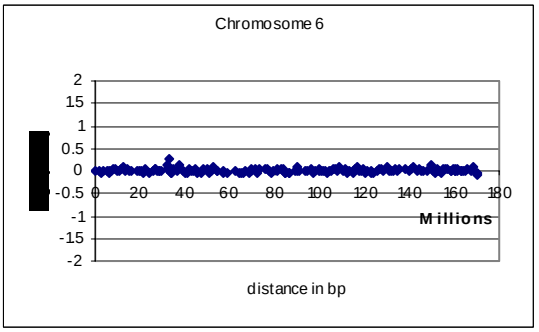

Chromosome 1
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0 50 100 150 200 250 300
Millions
distance in bp

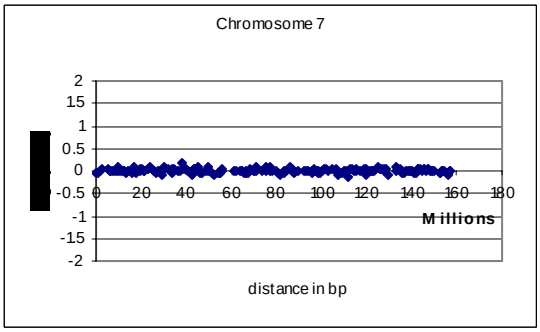

Chromosome 2
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0 50 100 150 200 250 300
Millions
distance in bp

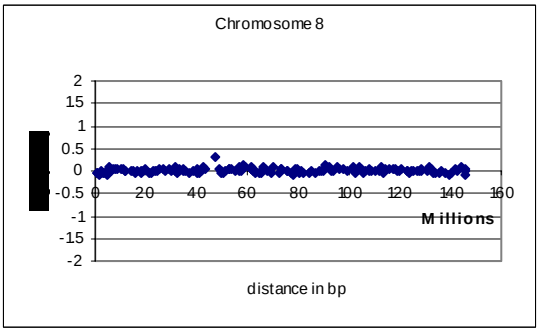

Chromosome 3
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0 50 100 150 200 250
Millions
distance in bp

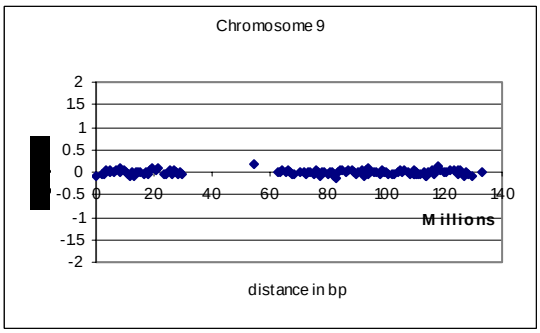

Chromosome 4
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0 50 100 150 200 250
Millions
distance in bp

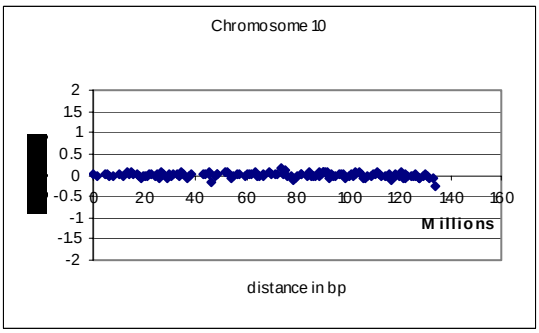

Chromosome 5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
50
100
150
200
Millions
distance in bp

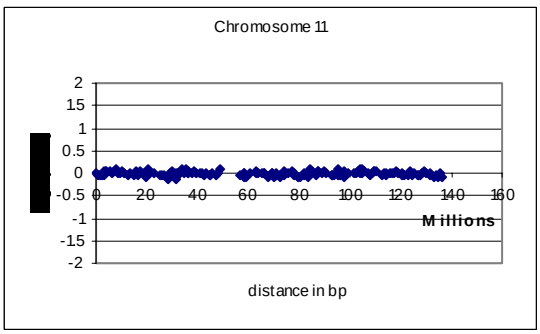

Chromosome 6
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
180
Millions
distance in bp

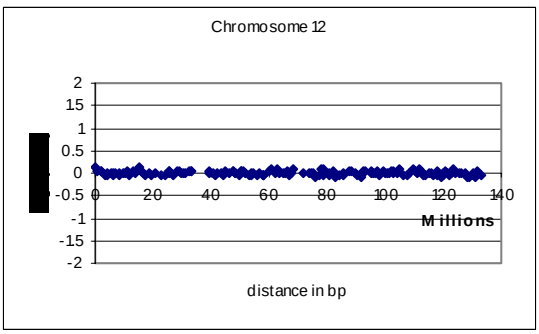

Chromosome 7
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
180
Millions
distance in bp

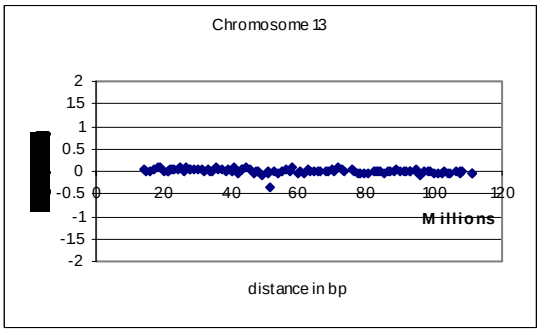

Chromosome 8
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
Millions
distance in bp

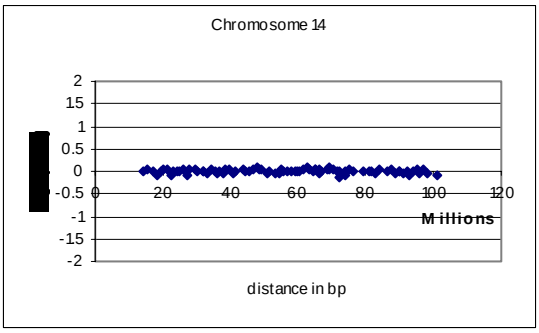

Chromosome 9
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
Millions
distance in bp

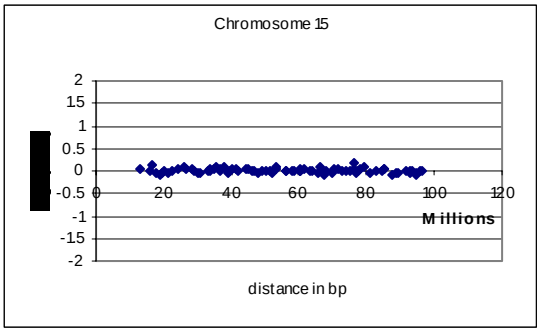

Chromosome 10
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
Millions
distance in bp

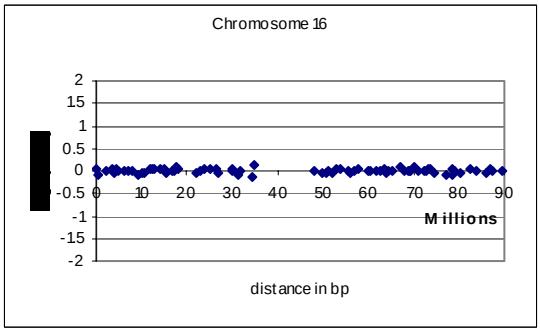

Chromosome 11
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
Millions
distance in bp

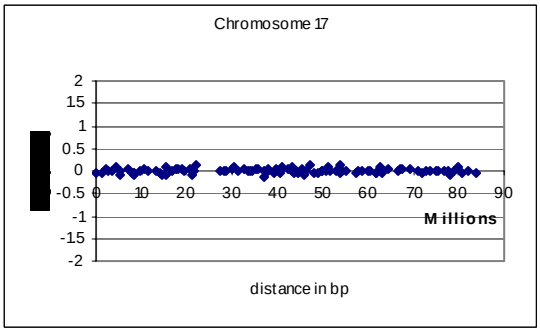

Chromosome 12
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
Millions
distance in bp

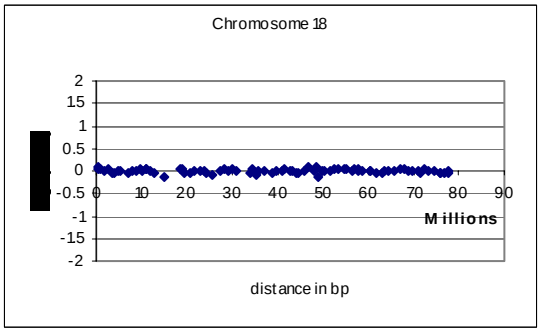

Chromosome 13
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

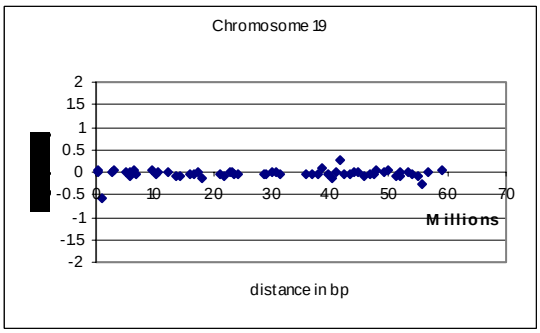

Chromosome 14
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

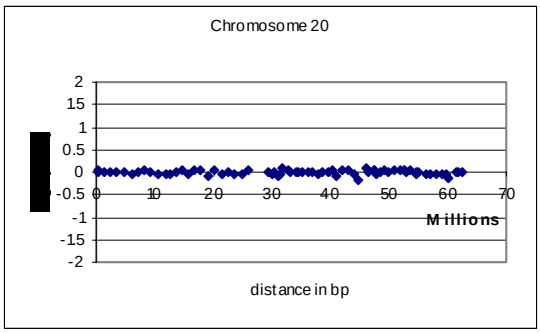

Chromosome 15
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

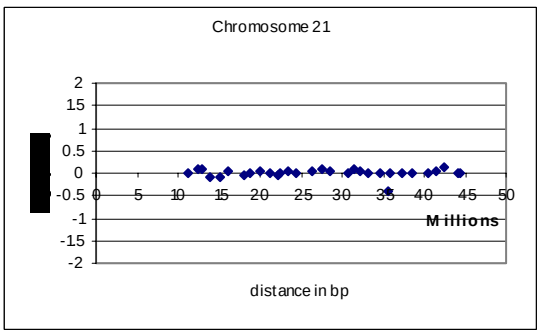

Chromosome 16
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp

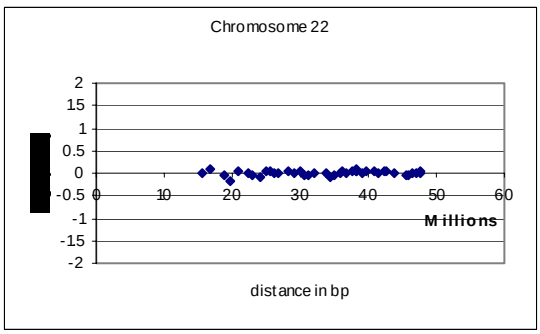

Chromosome 17
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp

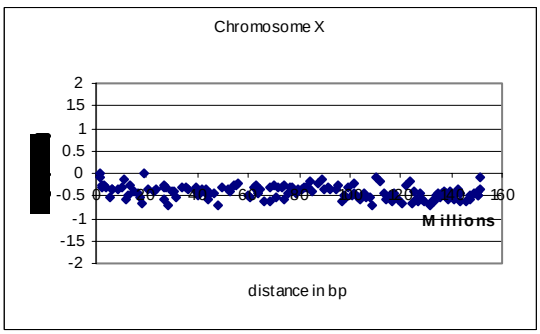

Chromosome 18
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp

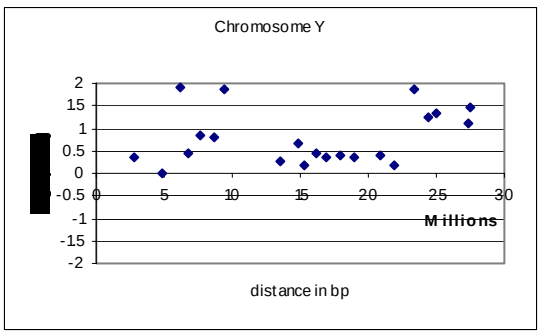

Chromosome 19
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
Millions
distance in bp


Chromosome 20
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
Millions
distance in bp

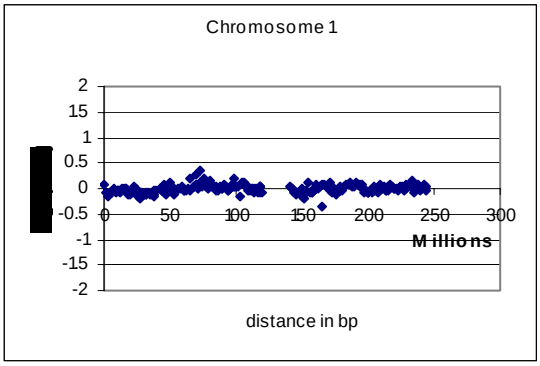

Chromosome 21
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
5
10
15
20
25
30
35
40
45
50
Millions
distance in bp

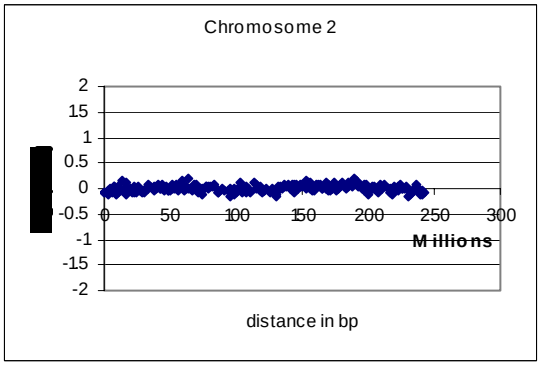

Chromosome 22
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
Millions
distance in bp

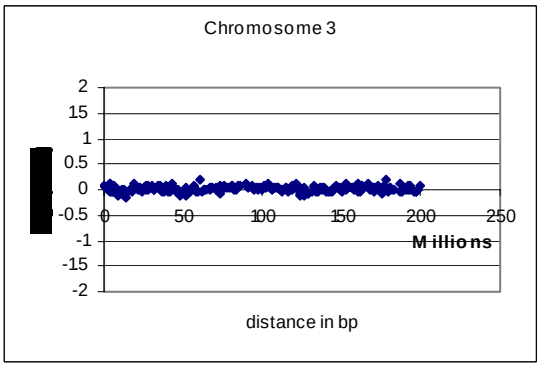

Chromosome X
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
Millions
distance in bp

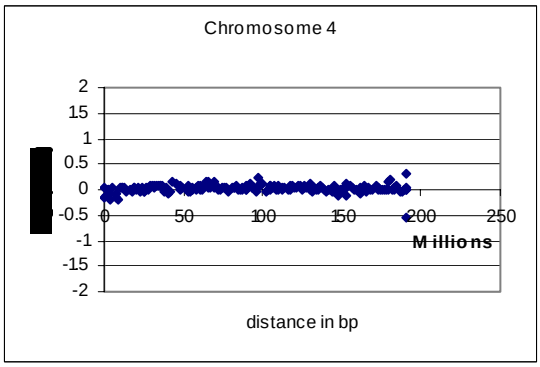

Chromosome Y
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
5
10
15
20
25
30
Millions
distance in bp

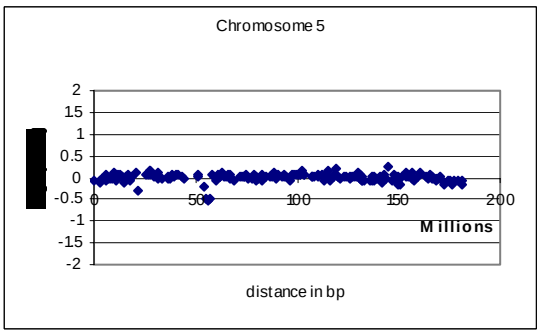
Patient 4:

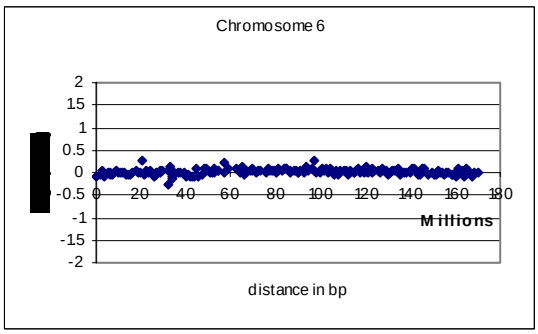

Chromosome 1
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
50
100
150
200
250
300
Millions
distance in bp

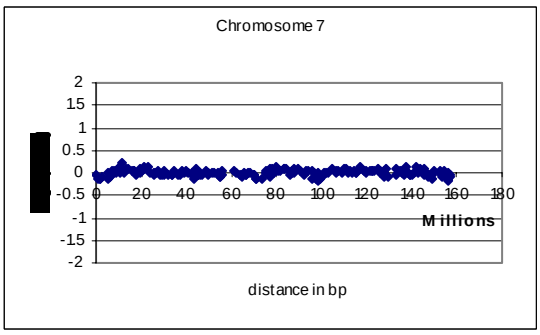

Chromosome 2
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
50
100
150
200
250
300
Millions
distance in bp

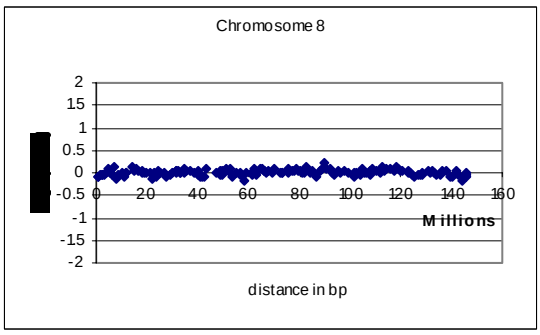

Chromosome 3
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
50
100
150
200
250
Millions
distance in bp

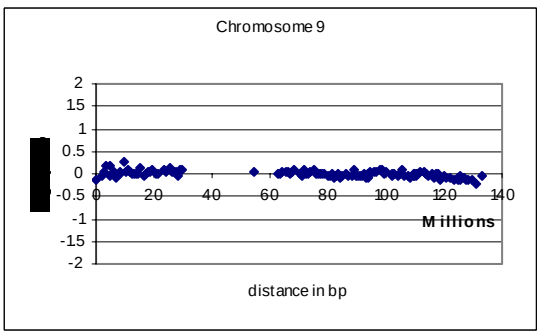

Chromosome 4
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
50
100
150
200
250
Millions
distance in bp

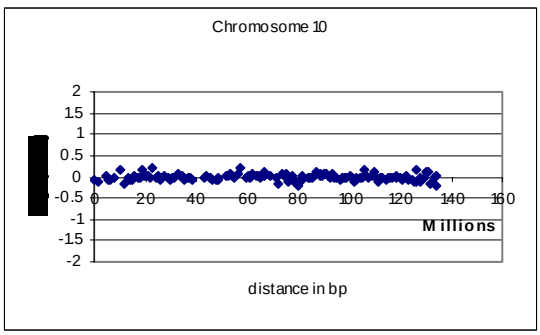

Chromosome 5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
50
100
150
200
Millions
distance in bp

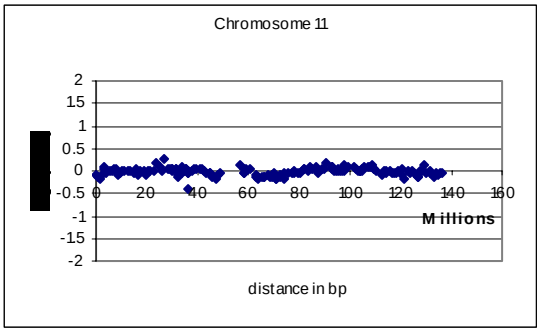

Chromosome 6
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
180
Millions
distance in bp

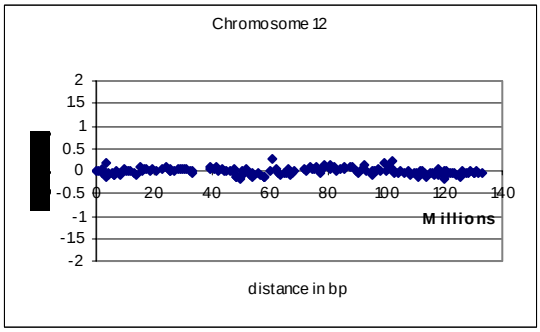

Chromosome 7
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
180
Millions
distance in bp

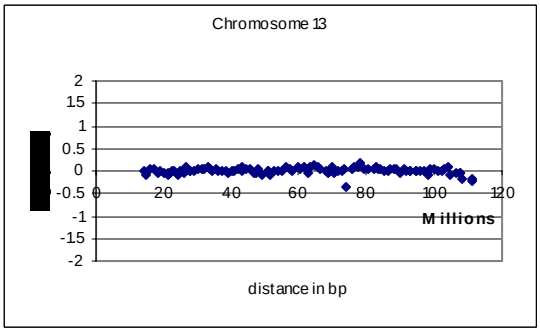

Chromosome 8
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
Millions
distance in bp

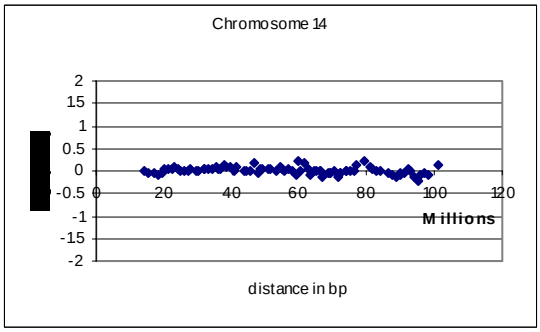

Chromosome 9
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
Millions
distance in bp

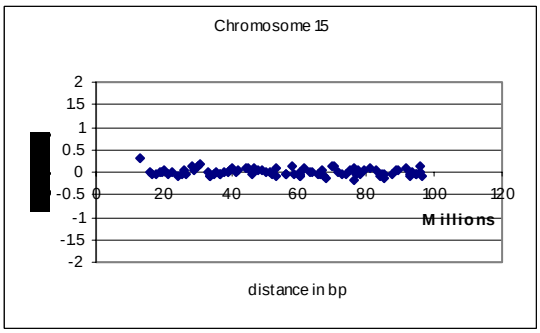

Chromosome 10
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
Millions
distance in bp

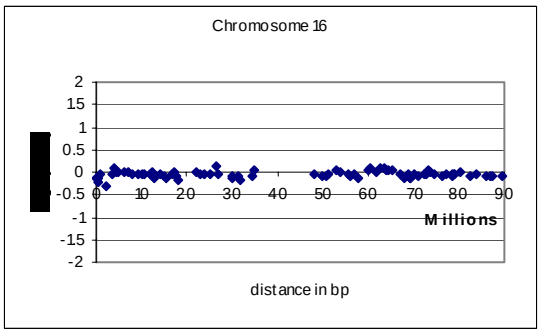

Chromosome 11
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
Millions
distance in bp

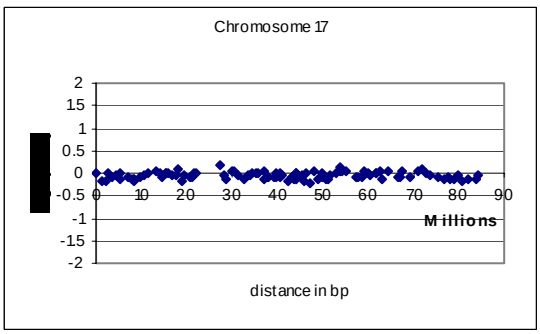

Chromosome 12
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
Millions
distance in bp

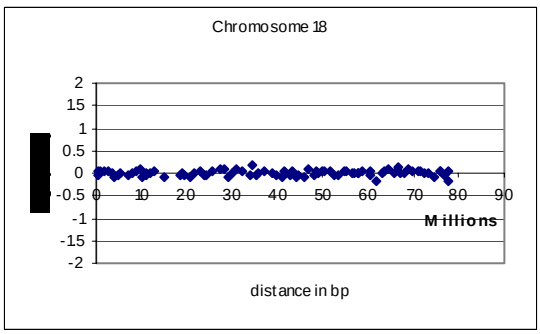

Chromosome 13
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

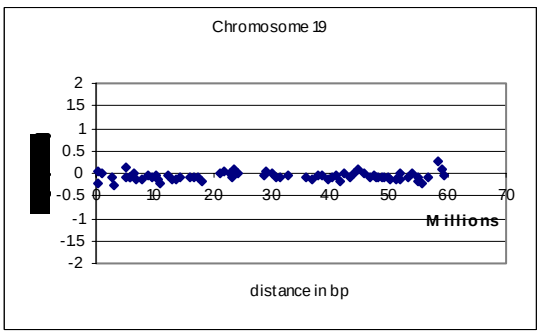

Chromosome 14
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

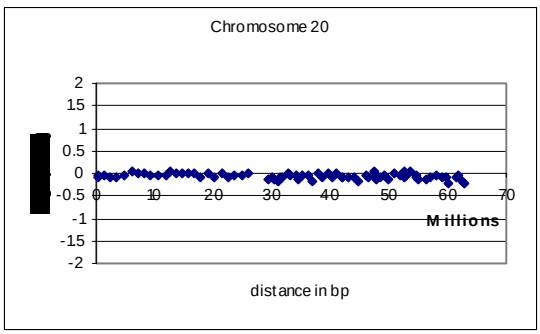

Chromosome 15
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

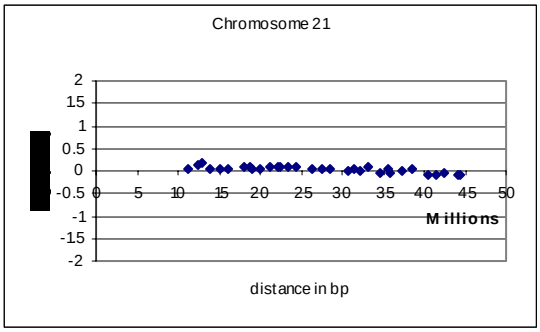

Chromosome 16
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp

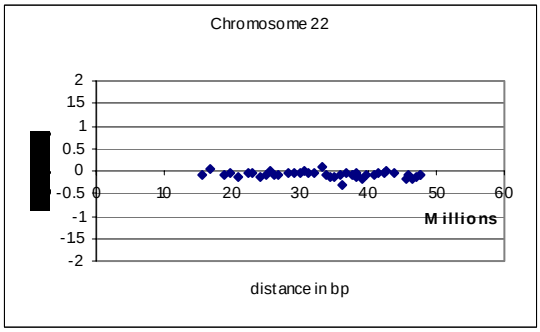

Chromosome 17
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp

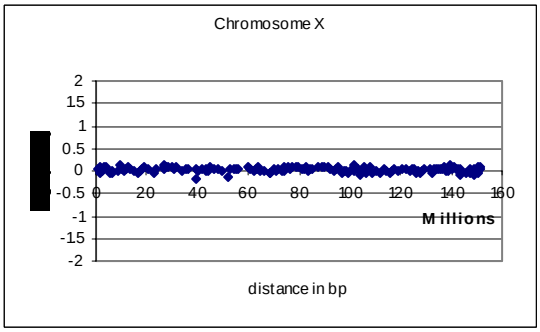

Chromosome 18
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp

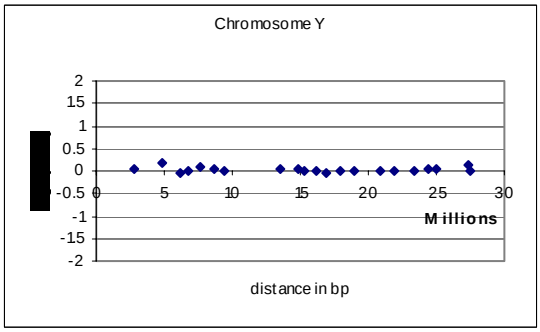

Chromosome 19
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
Millions
distance in bp


Chromosome 20
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
Millions
distance in bp

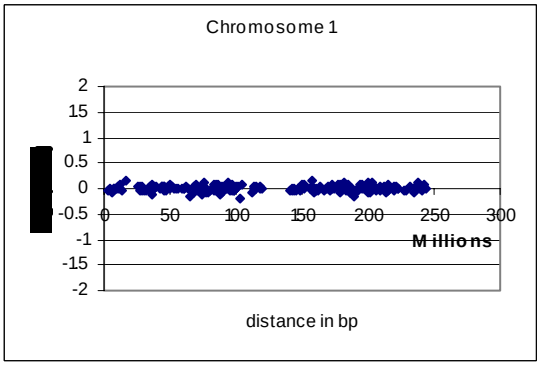

Chromosome 21
2
15
1
0.5
0
-0.5
-1
-15
-2
0 5 10 15 20 25 30 35 40 45 50
Millions
distance in bp

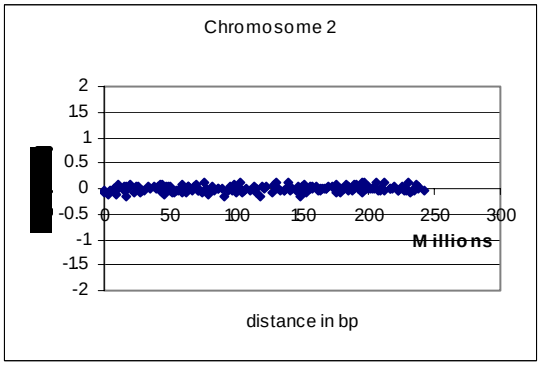

Chromosome 22
2
15
1
0.5
0
-0.5
-1
-15
-2
0 10 20 30 40 50 60
Millions
distance in bp

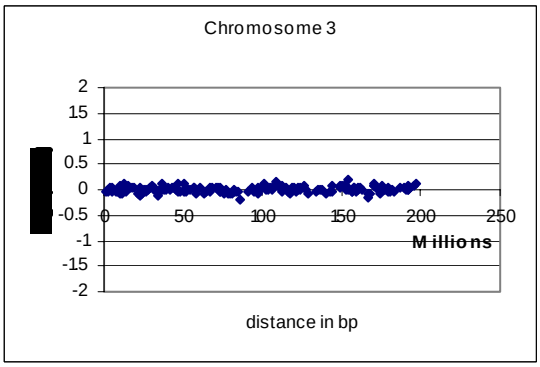

Chromosome X
2
15
1
0.5
0
-0.5
-1
-15
-2
0 20 40 60 80 100 120 140 160
Millions
distance in bp

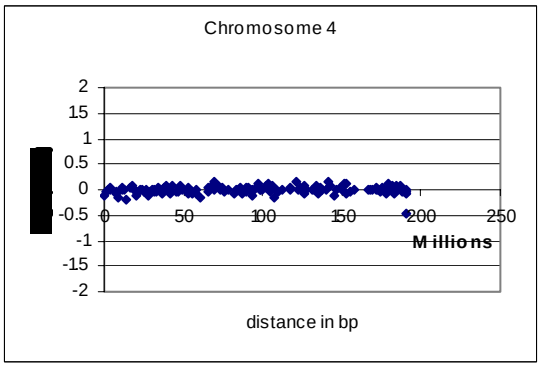

Chromosome Y
2
15
1
0.5
0
-0.5
-1
-15
-2
0 5 10 15 20 25 30
Millions
distance in bp

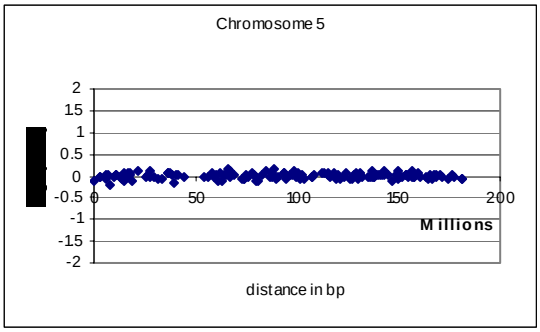
Patient 5:

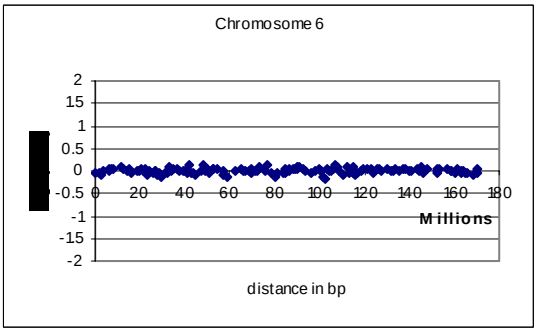

Chromosome 1
2
15
1
0.5
0
-0.5
-1
-15
-2
0 50 100 150 200 250 300
Millions
distance in bp

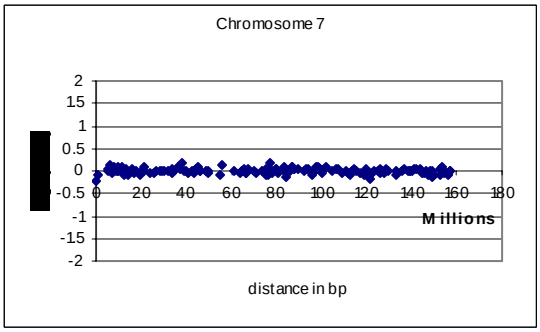

Chromosome 2
2
15
1
0.5
0
-0.5
-1
-15
-2
0 50 100 150 200 250 300
Millions
distance in bp

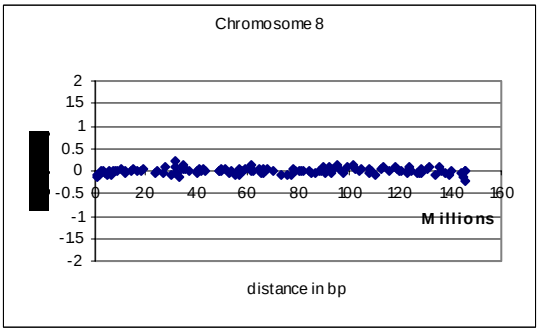

Chromosome 3
2
15
1
0.5
0
-0.5
-1
-15
-2
0 50 100 150 200 250
Millions
distance in bp

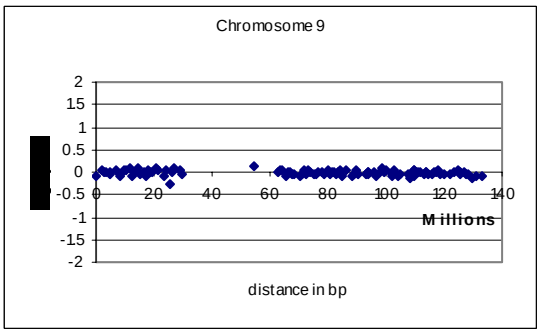

Chromosome 4
2
15
1
0.5
0
-0.5
-1
-15
-2
0 50 100 150 200 250
Millions
distance in bp

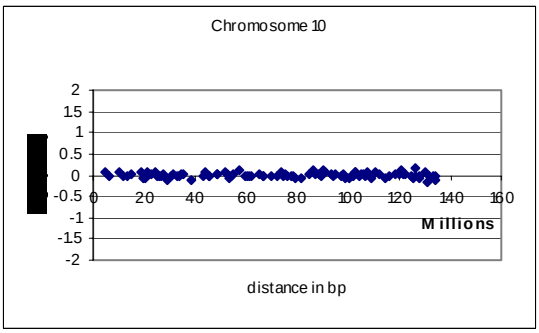

Chromosome 5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
50
100
150
200
Millions
distance in bp

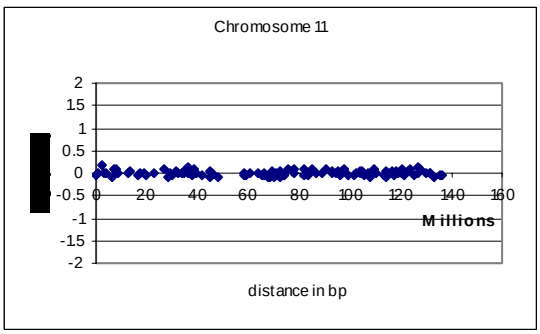

Chromosome 6
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
180
Millions
distance in bp

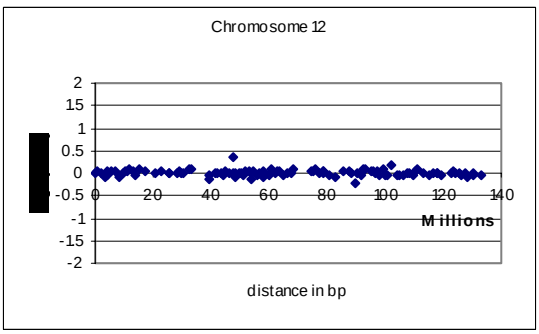

Chromosome 7
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
180
Millions
distance in bp

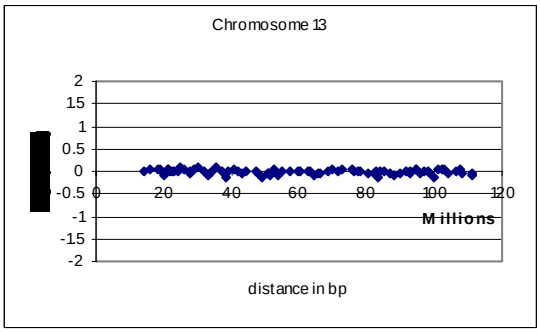

Chromosome 8
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
Millions
distance in bp

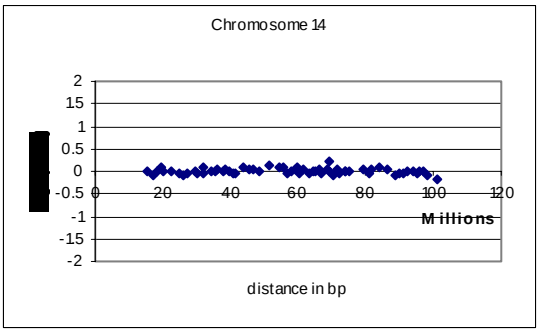

Chromosome 9
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
Millions
distance in bp

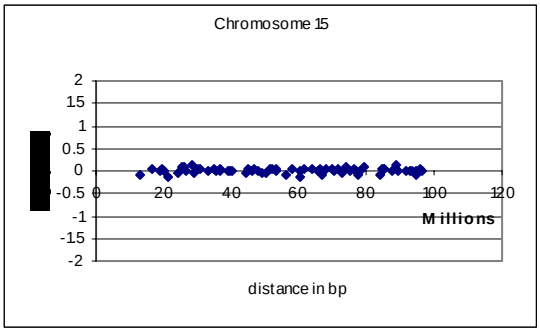

Chromosome 10
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
Millions
distance in bp

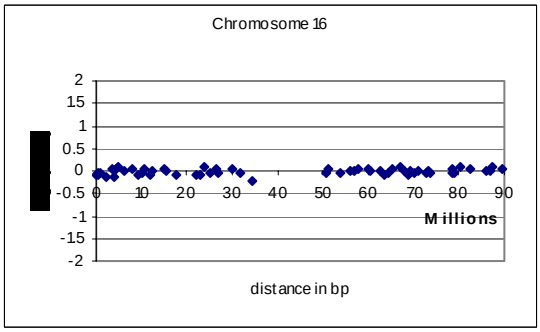

Chromosome 11
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
Millions
distance in bp

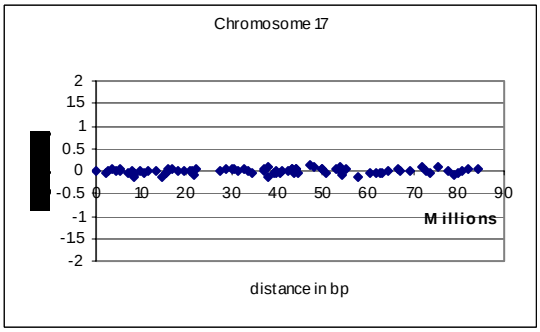

Chromosome 12
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
Millions
distance in bp

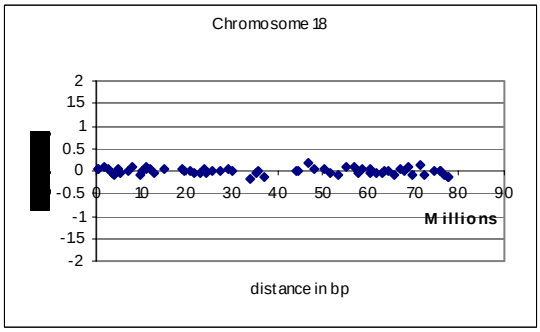

Chromosome 13
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

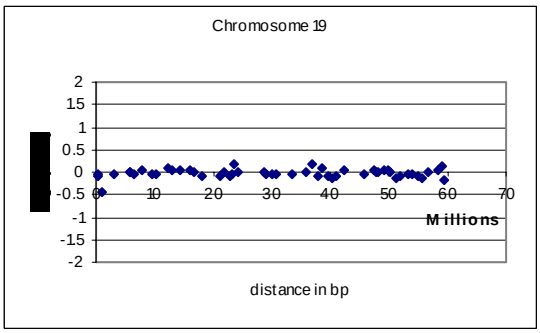

Chromosome 14
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

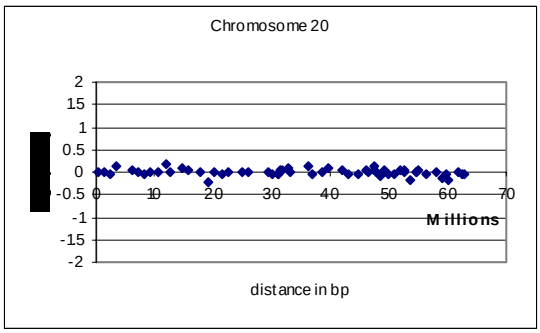

Chromosome 15
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

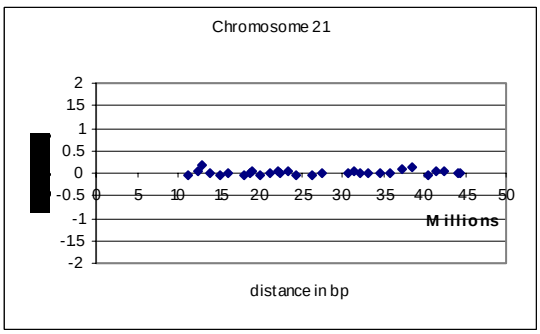

Chromosome 16
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp

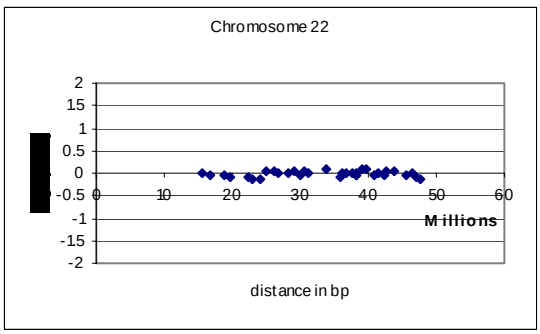

Chromosome 17
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp

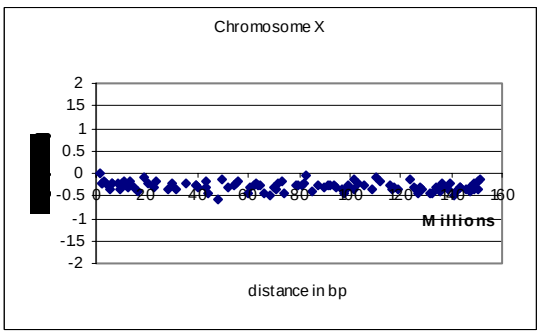

Chromosome 18
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp

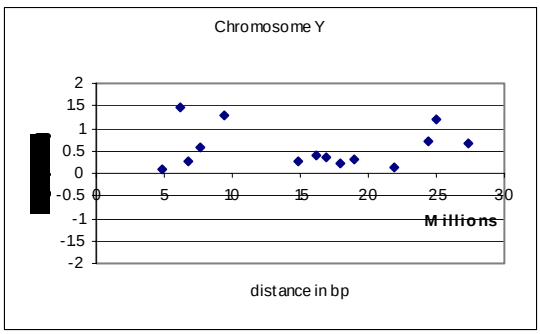

Chromosome 19
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
Millions
distance in bp


Chromosome 20
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
Millions
distance in bp

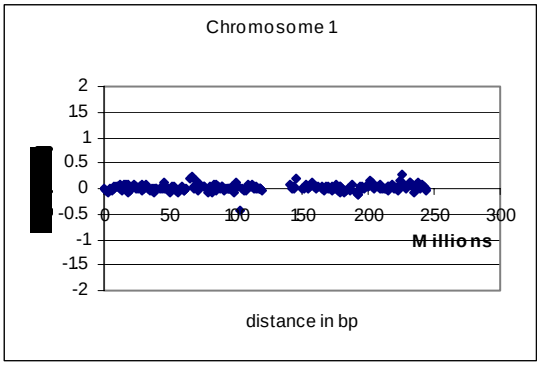

Chromosome 21
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
5
10
15
20
25
30
35
40
45
50
Millions
distance in bp

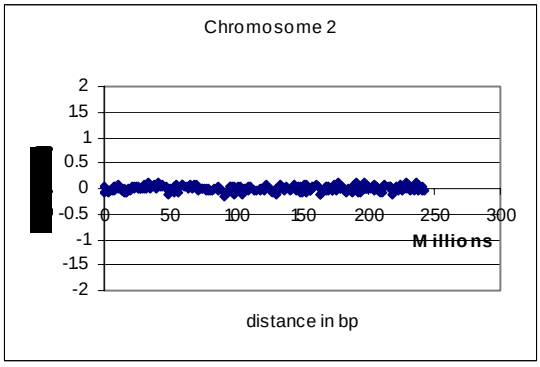

Chromosome 22
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
Millions
distance in bp

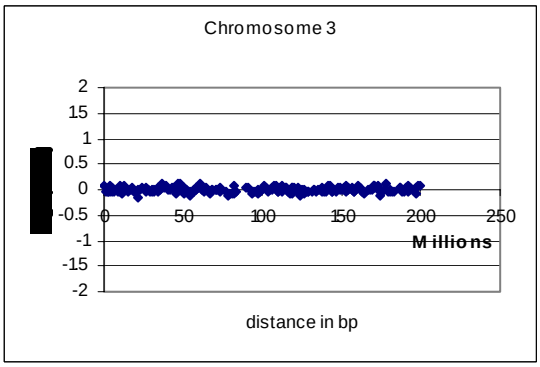

Chromosome X
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
Millions
distance in bp

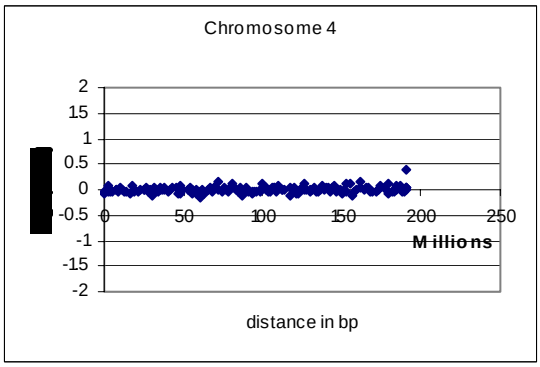

Chromosome Y
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
5
10
15
20
25
30
Millions
distance in bp

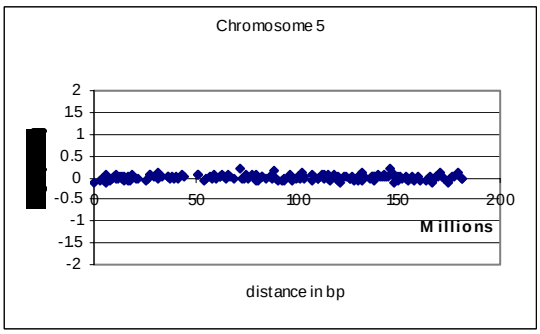
Patient 6:

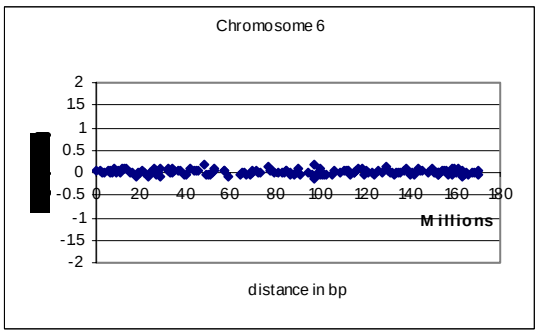

Chromosome 1
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
50
100
150
200
250
300
Millions
distance in bp

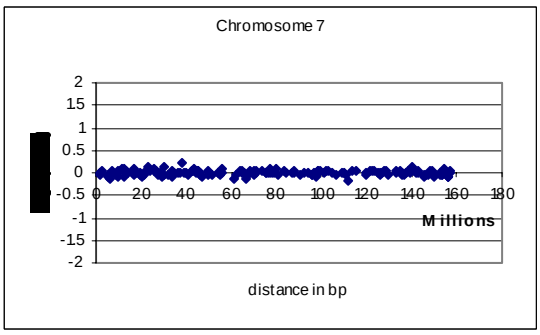

Chromosome 2
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
50
100
150
200
250
300
Millions
distance in bp

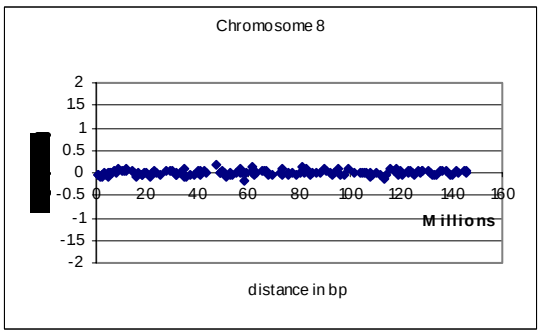

Chromosome 3
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
50
100
150
200
250
Millions
distance in bp

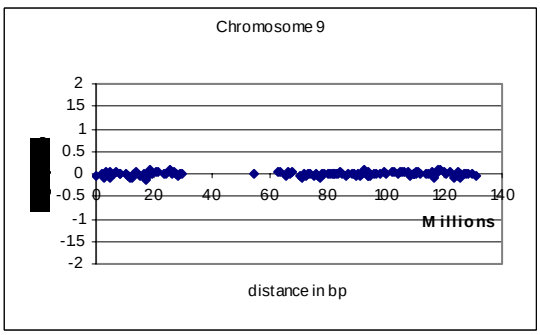

Chromosome 4
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
50
100
150
200
250
Millions
distance in bp

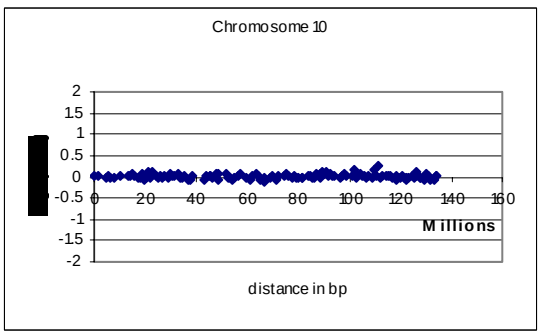

Chromosome 5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
50
100
150
200
Millions
distance in bp

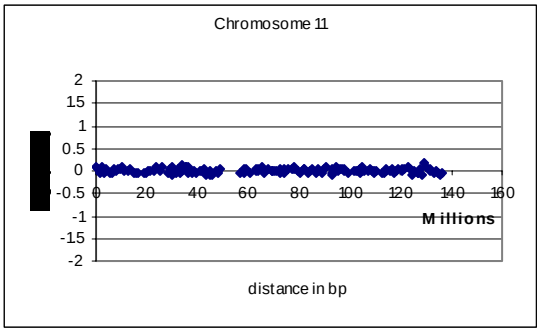

Chromosome 6
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
180
Millions
distance in bp

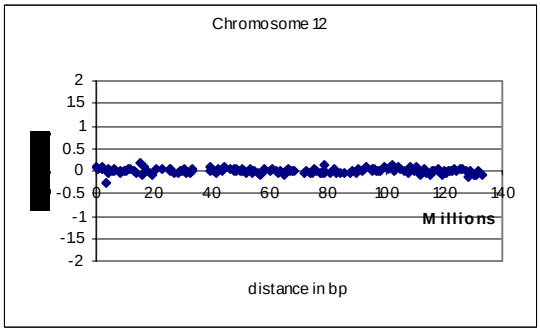

Chromosome 7
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
180
Millions
distance in bp

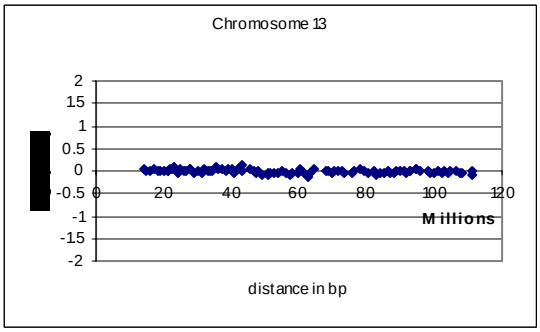

Chromosome 8
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
Millions
distance in bp

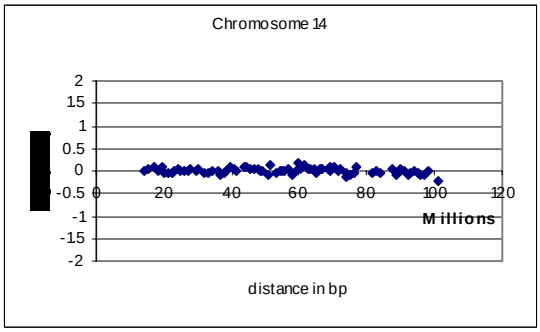

Chromosome 9
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
Millions
distance in bp

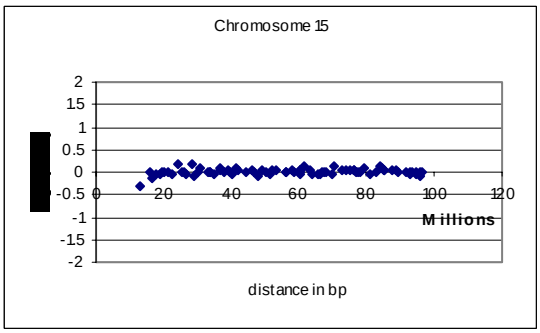

Chromosome 10
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
Millions
distance in bp

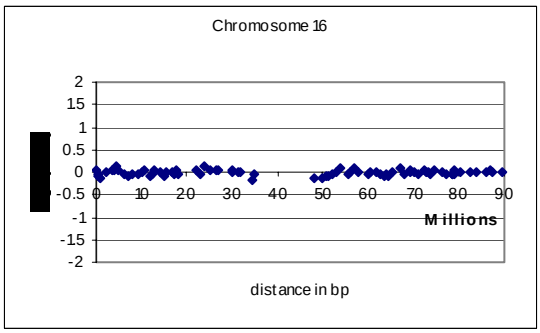

Chromosome 11
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
160
Millions
distance in bp

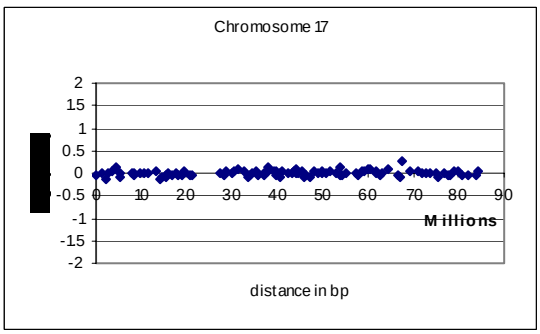

Chromosome 12
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
140
Millions
distance in bp

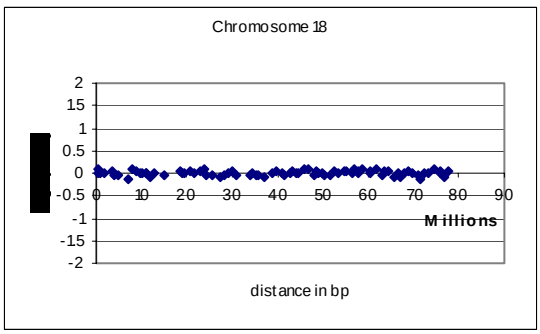

Chromosome 13
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

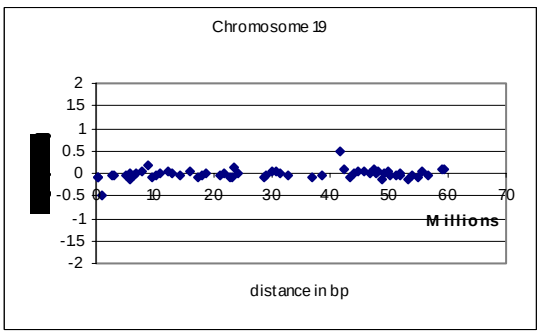

Chromosome 14
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

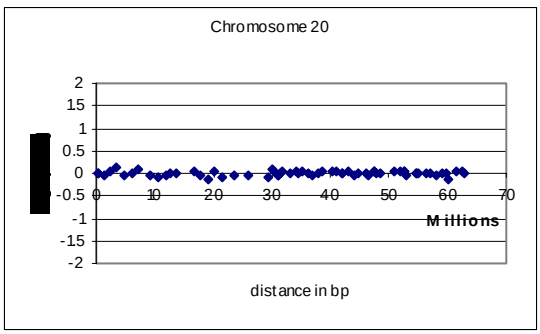

Chromosome 15
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
120
Millions
distance in bp

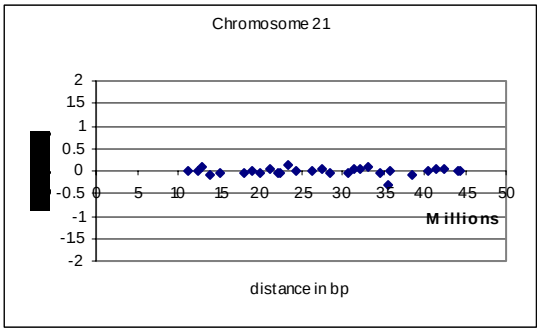

Chromosome 16
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp

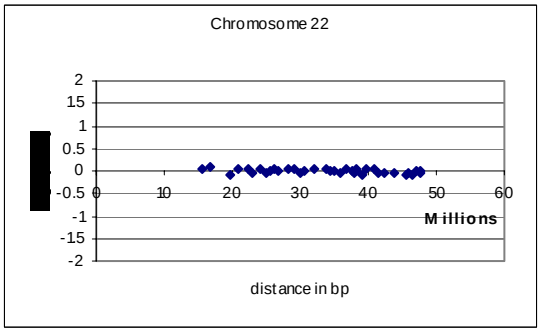

Chromosome 17
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp

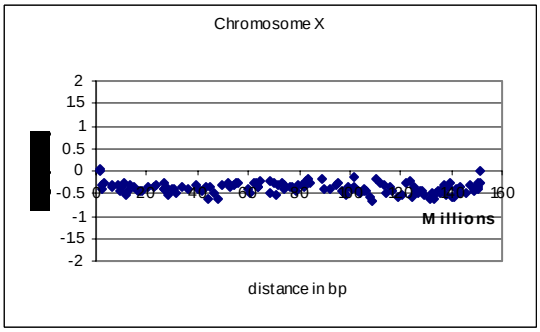

Chromosome 18
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
80
90
Millions
distance in bp

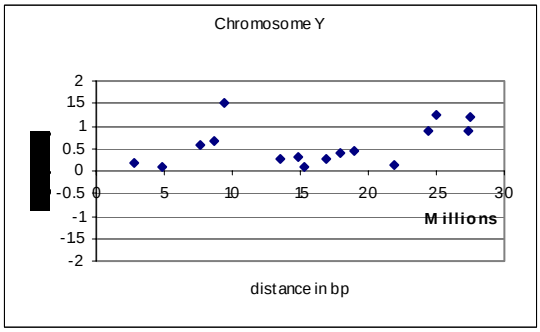

Chromosome 19
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
Millions
distance in bp


Chromosome 20
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
70
Millions
distance in bp


Chromosome 21
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
5
10
15
20
25
30
35
40
45
50
Millions
distance in bp


Chromosome 22
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
Millions
distance in bp


Chromosome X
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
Millions
distance in bp


Chromosome Y
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
5
10
15
20
25
30
Millions
distance in bp

Appendix 11: Clones known to report an incorrect copy number change on the 1Mb resolution array

Autosomes reporting an unexpected linear ratio

Clone	Chr	Position (bp)
bA326G21	1	143533568
bA5K23	1	159201779.5
dJ1108M17	1	104815713
dJ97P20	1	167462445
bA32C20	2	128082133
bA400O18	2	184873490
963K6	4	191632378
bA94E2	5	18186914.5
dJ159G19	6	80412976
dJ93N13	6	32596730.5
bA17M8	8	136548513.5
bA350F16	8	46413667
bB445N5	10	38414815
bA13E1	10	48233085
221K18	12	131110572
bA25J23	13	78143302
bA279F15	13	55656472
820M16	14	104124908
bA2F9	15	18507931
bA161M6	16	1085464
dJ843B9	17	43741286
bA220N20	17	44390565.5
bA416K7	17	45289524.5
bA294I20	19	86263
bA50L23	22	19845427.5

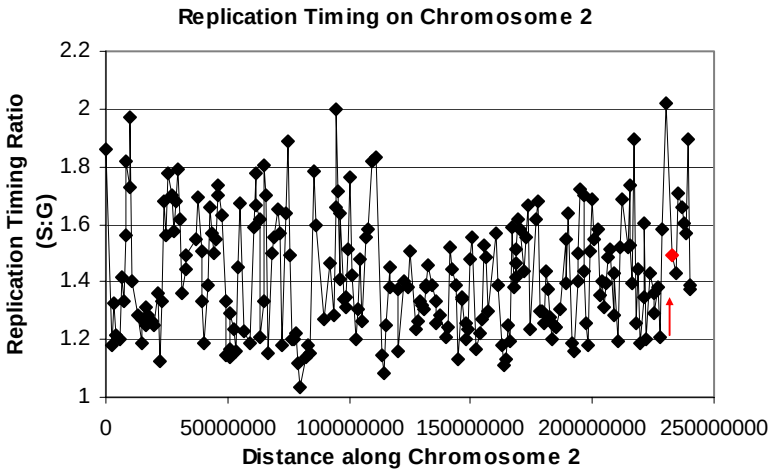
X Clones reporting an unexpected linear ratio

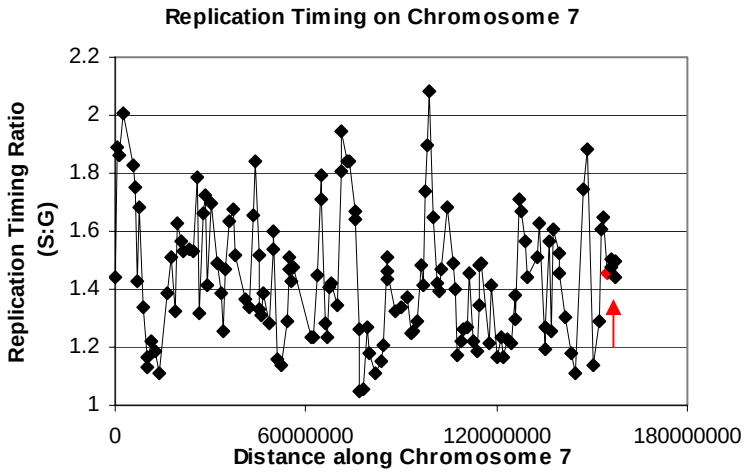
Clone	Chr	Position (bp)
98C4	X	490000
bA155F12	X	1834539.5
bA457M7	X	2000055.5
bA418N20	X	2326367.5
bA483M24	X	5959670
bA323F16	X	6434519
bA431J24	X	15946832
bA2J15	X	16774468
bA268G12	X	25943688.5
bA163L4	X	39575156.5
bA56H2	X	49574586.5

bB188A5	X	53052560
bA445O16	X	53863026
dJ966K21	X	54681885
dJ323B6	X	60874153.5
bB130F17	X	61742036.5
dJ583H20	X	66927306.5
bB260P4	X	68630583
bA236O12	X	70675479
dJ411B6	X	71885166.5
dJ875J14	X	72373814
dJ93L7	X	80445116.5
dJ225D2	X	80937973.5
bB166C10	X	85629051
bA122L9	X	86374786.5
bA156J23	X	87618715.5
dJ421I20	X	98011402
dJ312P4	X	100360142.5
dJ290B4	X	108330281
dJ93I3	X	109797538
bA434C1	X	111084399
dJ428A13	X	121210464
bA218L14	X	148668646
225F6	X	149149818

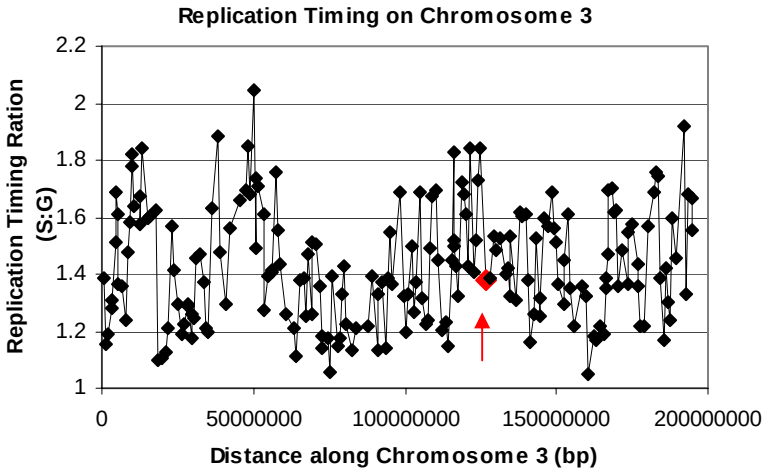
Appendix 12: Position of Chromosomal Breakpoints on the Replication Timing Profiles Location of breakpoints are indicated by red clones and red arrows.

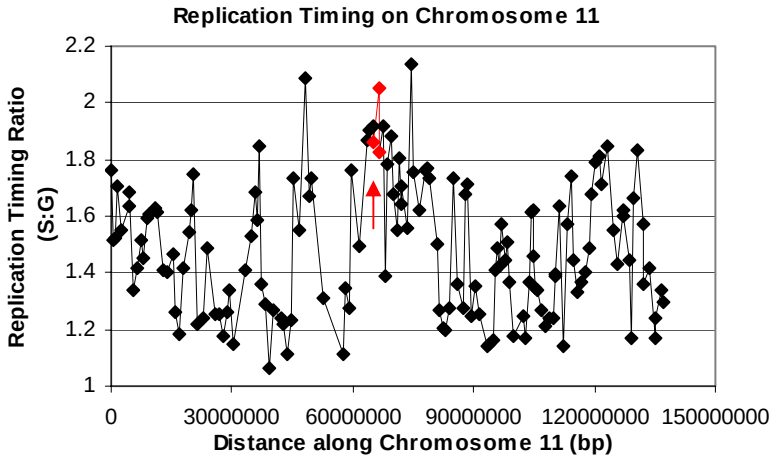
t(2:7)a



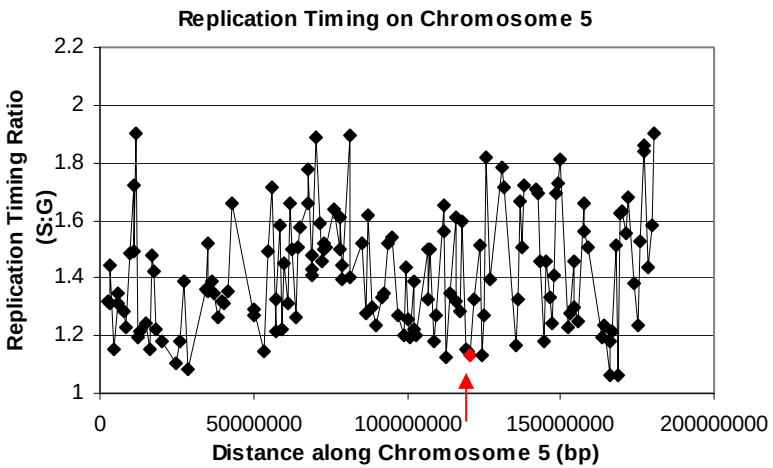


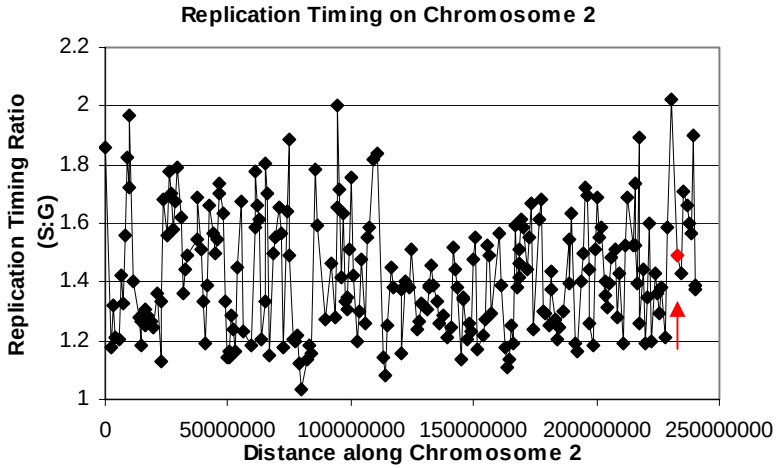
(3:11)



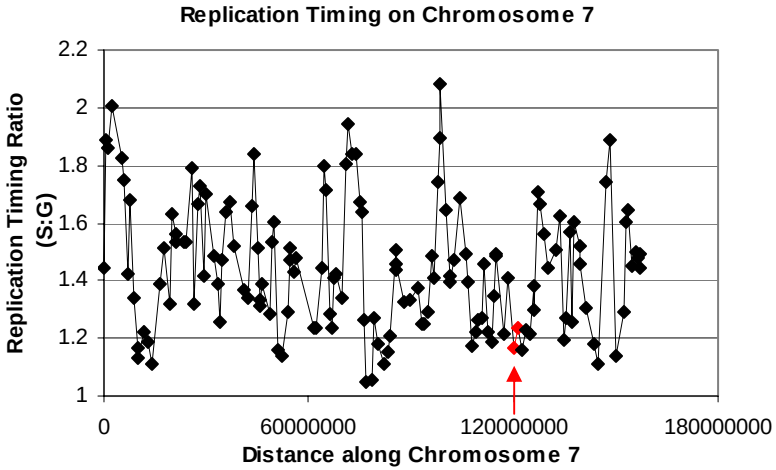


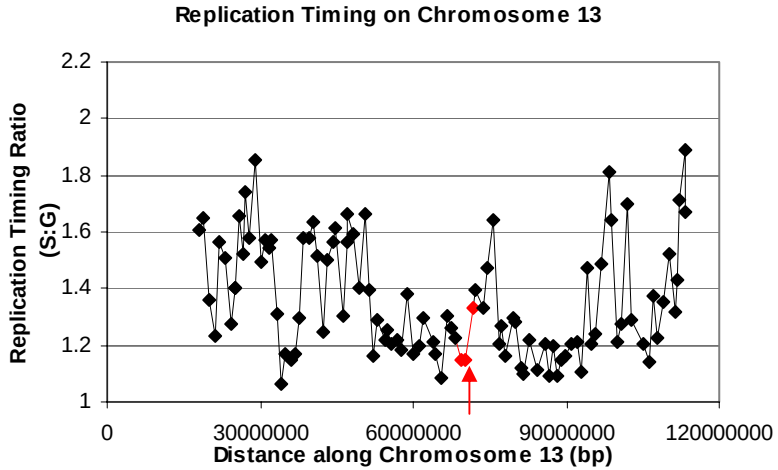
t(2:5)


Replication Timing on Chromosome 5
Replication Timing Ratio (S:G)
2.2
2
1.8
1.6
1.4
1.2
1
0
50000000
100000000
150000000
200000000
Distance along Chromosome 5 (bp)

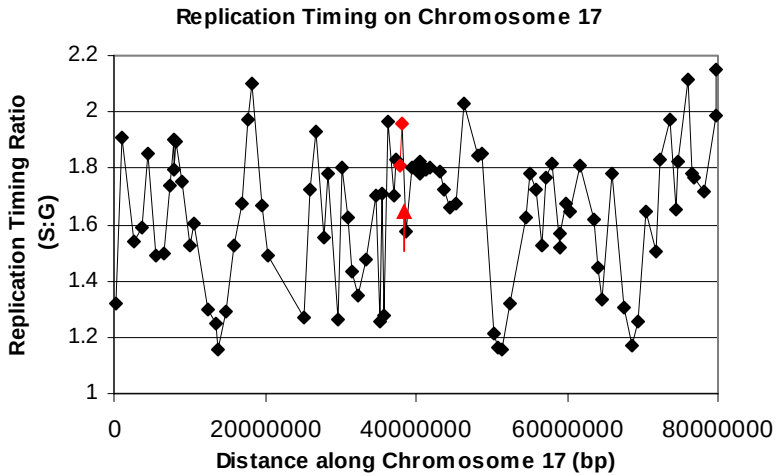

Replication Timing on Chromosome 2
Replication Timing Ratio (S:G)
2.2
2
1.8
1.6
1.4
1.2
1
0
50000000
100000000
150000000
200000000
250000000
Distance along Chromosome 2

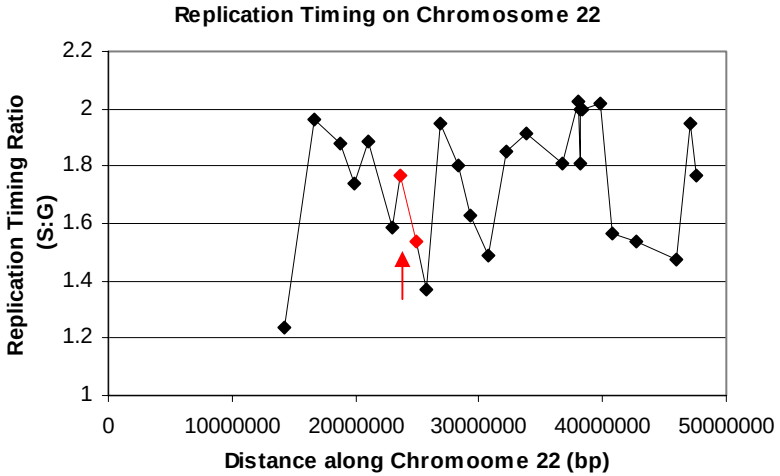
t(7:13)


Replication Timing on Chromosome 7
Replication Timing Ratio (S:G)
2.2
2
1.8
1.6
1.4
1.2
1
0
60000000
120000000
180000000
Distance along Chromosome 7

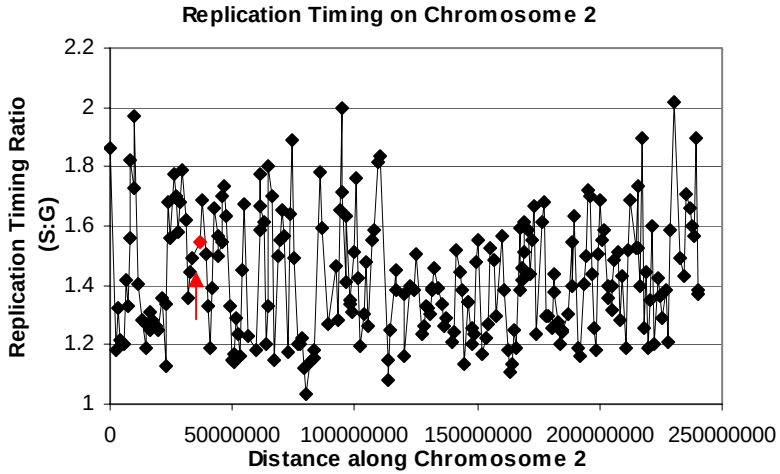

Replication Timing on Chromosome 13
Replication Timing Ratio (S:G)
2.2
2
1.8
1.6
1.4
1.2
1
0
30000000
60000000
90000000
120000000
Distance along Chromosome 13 (bp)

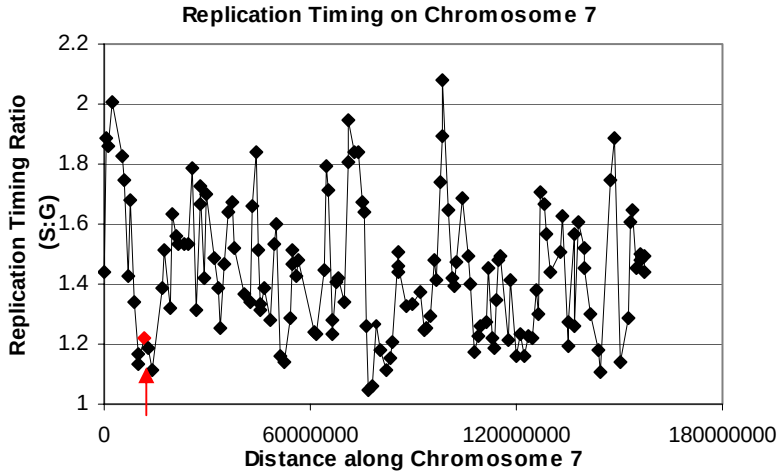
t(17:22)


Replication Timing on Chromosome 17
Replication Timing Ratio (S:G)
2.2
2
1.8
1.6
1.4
1.2
1
0
20000000
40000000
60000000
80000000
Distance along Chromosome 17 (bp)

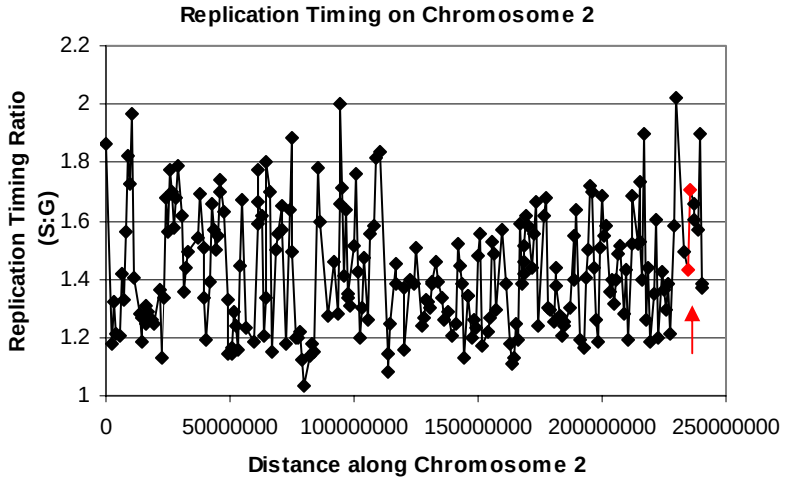

Replication Timing on Chromosome 22
Replication Timing Ratio (S:G)
2.2
2
1.8
1.6
1.4
1.2
1
0
10000000
20000000
30000000
40000000
50000000
Distance along Chromoome 22 (bp)

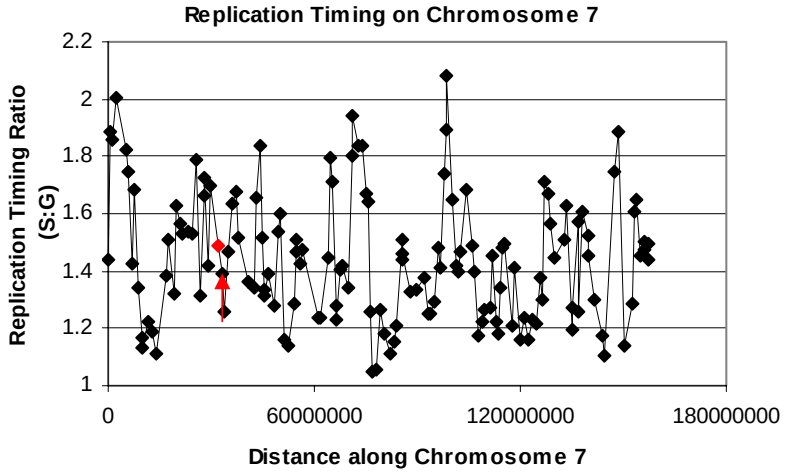
t(2:7)b


Replication Timing on Chromosome 2
Replication Timing Ratio (S:G)
2.2
2
1.8
1.6
1.4
1.2
1
0
50000000
100000000
150000000
200000000
250000000
Distance along Chromosome 2

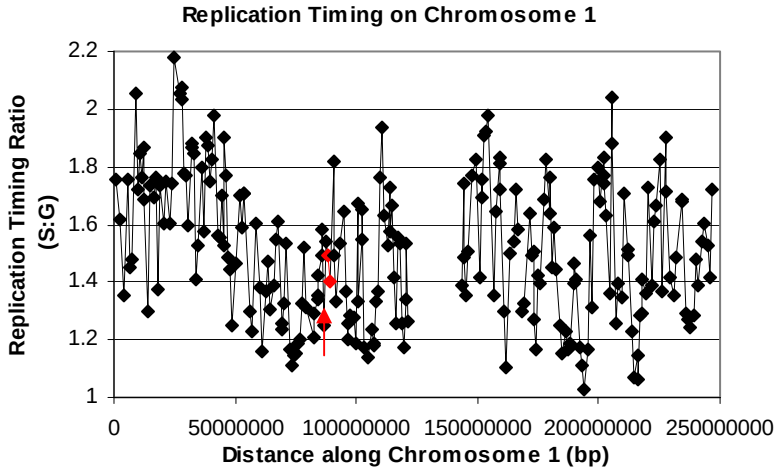

Replication Timing on Chromosome 7
Replication Timing Ratio (S:G)
2.2
2
1.8
1.6
1.4
1.2
1
0
60000000
120000000
180000000
Distance along Chromosome 7

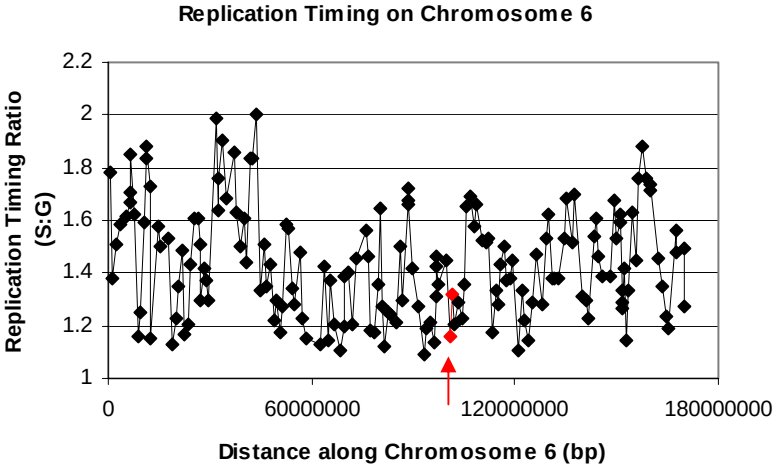
t(2:7)c


Replication Timing on Chromosome 2
Replication Timing Ratio (S:G)
2.2
2
1.8
1.6
1.4
1.2
1
0
50000000
100000000
150000000
200000000
250000000
Distance along Chromosome 2


Replication Timing on Chromosome 7
Replication Timing Ratio (S:G)
2.2
2
1.8
1.6
1.4
1.2
1
0
60000000
120000000
180000000
Distance along Chromosome 7

t(1:6)


Replication Timing on Chromosome 1
Replication Timing Ratio (S:G)
2.2
2
1.8
1.6
1.4
1.2
1
0
50000000
100000000
150000000
200000000
250000000
Distance along Chromosome 1 (bp)


Replication Timing on Chromosome 6
Replication Timing Ratio (S:G)
2.2
2
1.8
1.6
1.4
1.2
1
0
60000000
120000000
180000000
Distance along Chromosome 6 (bp)

Appendix 13: The significance of a correlation co-efficient.

Value of Coefficient (r)	**Meaning**
0.00-0.19	A very weak correlation
0.20-0.39	A weak correlation
0.40-0.69	A modest correlation
0.70-0.89	A strong correlation
0.90-1.00	A very strong correlation

Table from ‘Practical Statistics for Field Biologists’ (Fowler 1998).

Appendix 14: Publications arising from this work.

‘The Replication timing of the human genome.’ Woodfine et al. Human Molecular Genetics