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.