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