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.

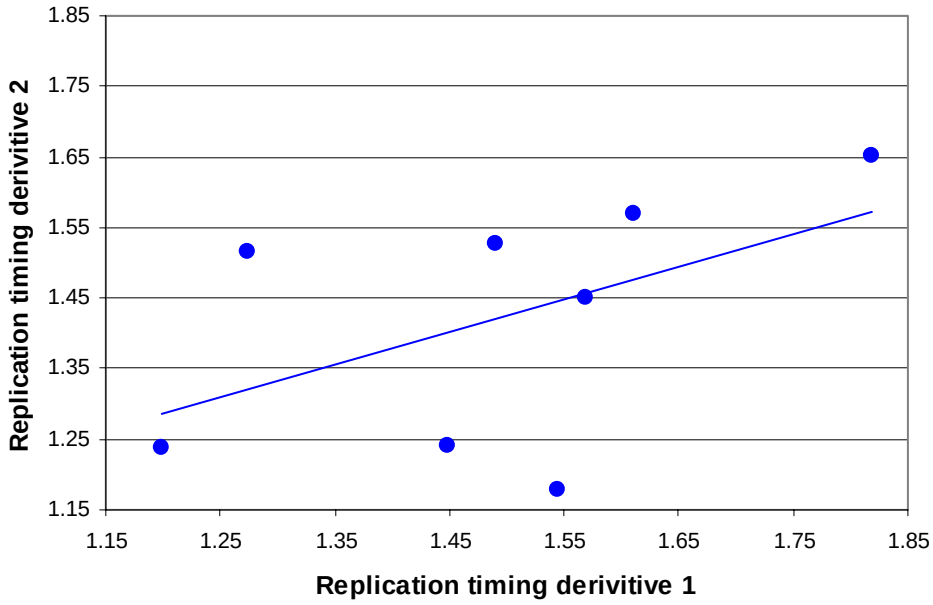


Figure 6.14: The correlation between the replication timing of the first and second breakpoint on a congenital translocation.

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)

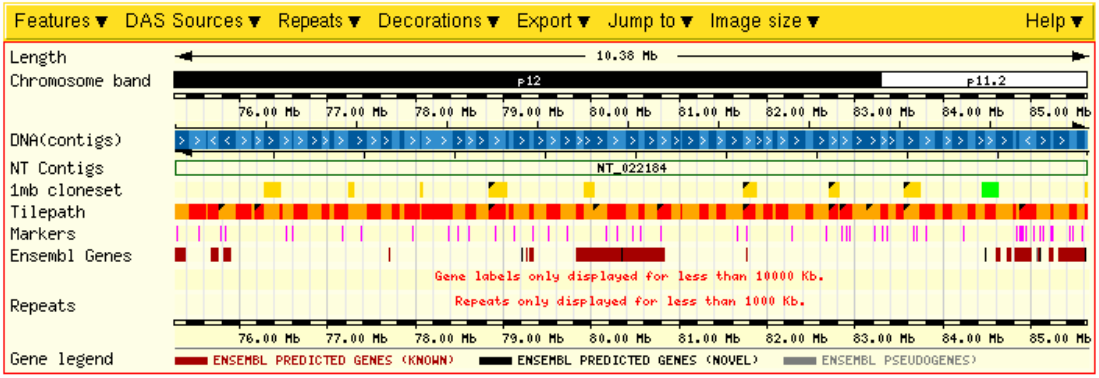
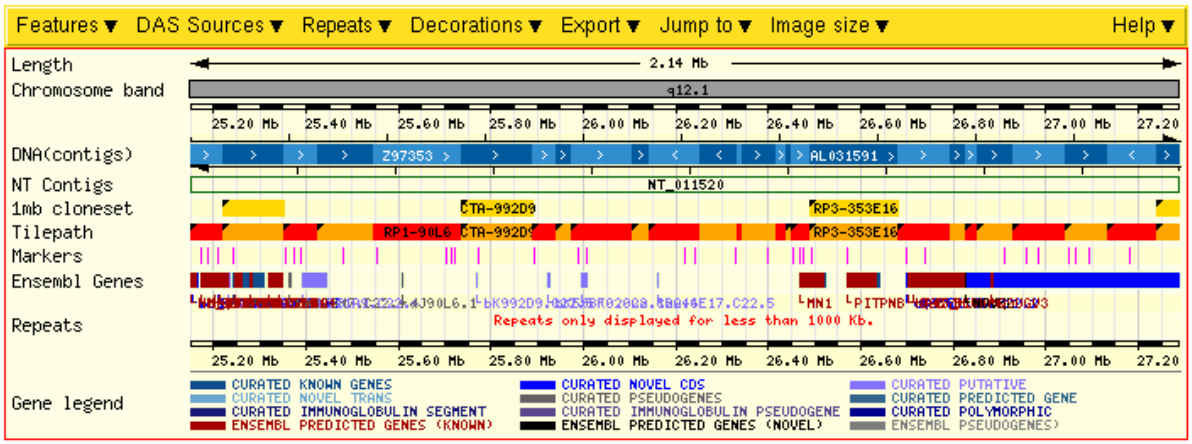


Figure 6.15: Ensembl view of a region of chromosome 2 (75-86Mb) to illustrate the position of the 1Mb clones in relation to genes in this region. The tracks of Ensembl show the cytogenetic location of the area studied and the representation of sequenced DNA contigs. The '1Mb cloneset' track shows the 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.

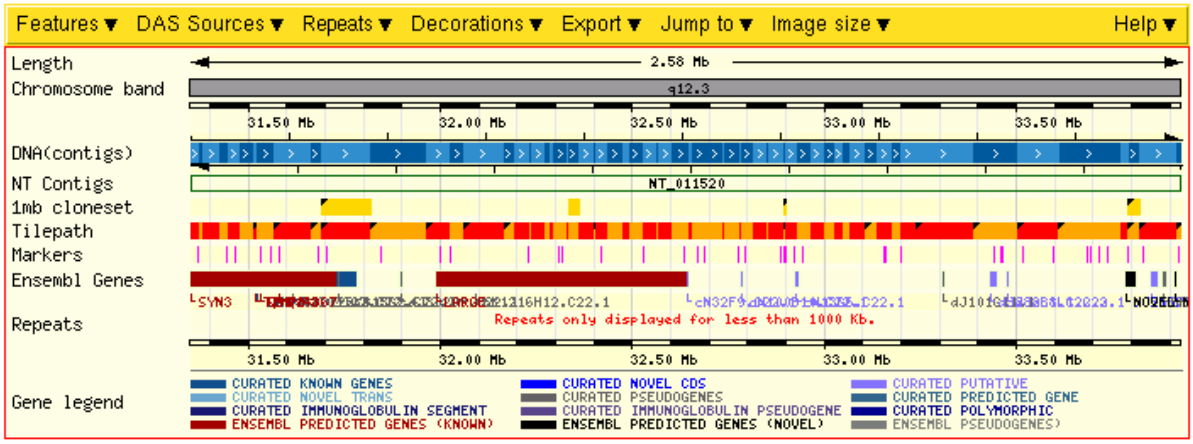
The chromosome 22 tile path array provides complete coverage of the whole q arm of chromosome 22 and ensures that the transcriptional activity of all chromosome 22 genes can be correlated with replication timing.

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

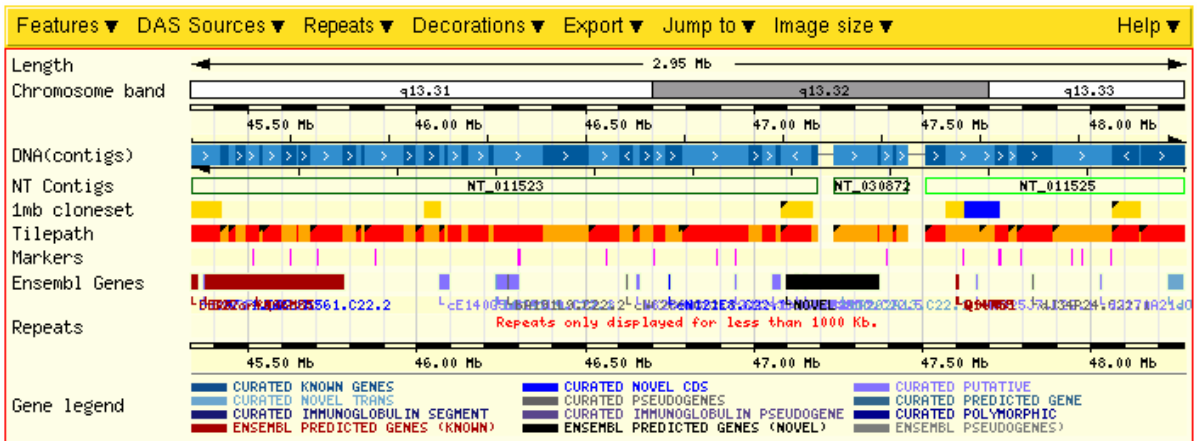


Figure 6.16: Ensembl pages illustrating those regions that are late replicating and under represented on the Affymetrix array are gene poor. A: Region 1, 25150000-27290000bp, B: Region 2, 31350000-33930000bp, C: Region 3 45330000-48280000bp. The tracks are as described in the legend to Figure 6.15.

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