

Figure 6.5: (Previous page) Replication timing ratio (blue) and Expression level ( $\text{Log}_{10}$  -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 correlation 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).

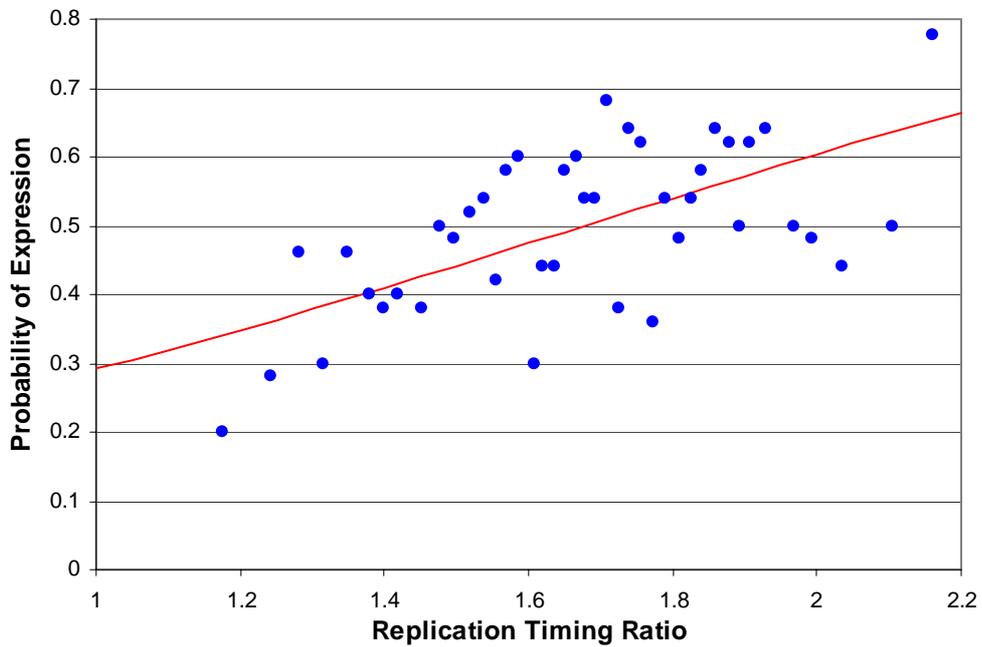


Figure 6.6: Correlation between replication timing and probability of expression for windows of 50 clones. The red line shows the logistic regression given by the equation  $y = e^{-2.13 + 1.31x} / (1 + e^{-2.13 + 1.31x})$ .

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)

