LATE-REPLICATING HETEROCHROMATIN IS CHARACTERISED BY DECREASED CYTOSINE METHYLATION IN THE HUMAN GENOME

Masako Suzuki, Mayumi Oda *et al.*

SUPPLEMENTARY DATA

FIGURE S1

Validation of gene expression microarray data. In panel **(a)** we show correlation coefficients between gene expression microarray and real time RT-PCR data. The data from both cell types showed high correlation values (R^2>0.96). In panel **(b)** we show log intensity density plots and a pair-wise plot for the expression microarray data. The red dots with gene names indicate the genes that we validated by real time RT-PCR.

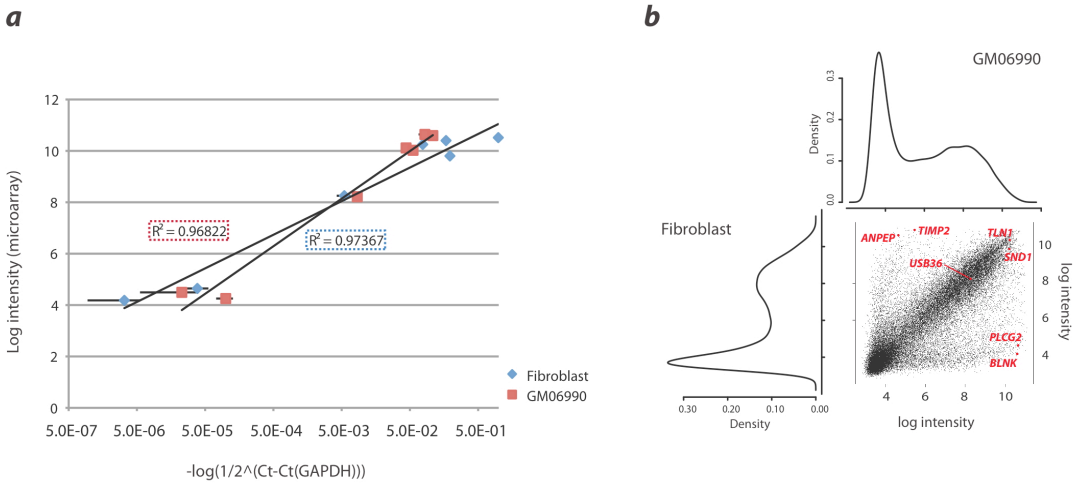

FIGURE S2

The transformation strategy for replication timing data is shown. In panel **(a)**, the sum of read numbers in 1 kb windows for early (y-axis) or later (x-axis) stages of the cell cycle are plotted. The angular transformation that we have recently described [Suzuki, #2132] allows us to represent the replication timing for each window in terms of an angle value. Panel **(b)** shows the relationship between transformed data and raw DNA sequence read data at the same locus illustrated in **Figure 1**; Chr2:137,626,033-234,348,474).

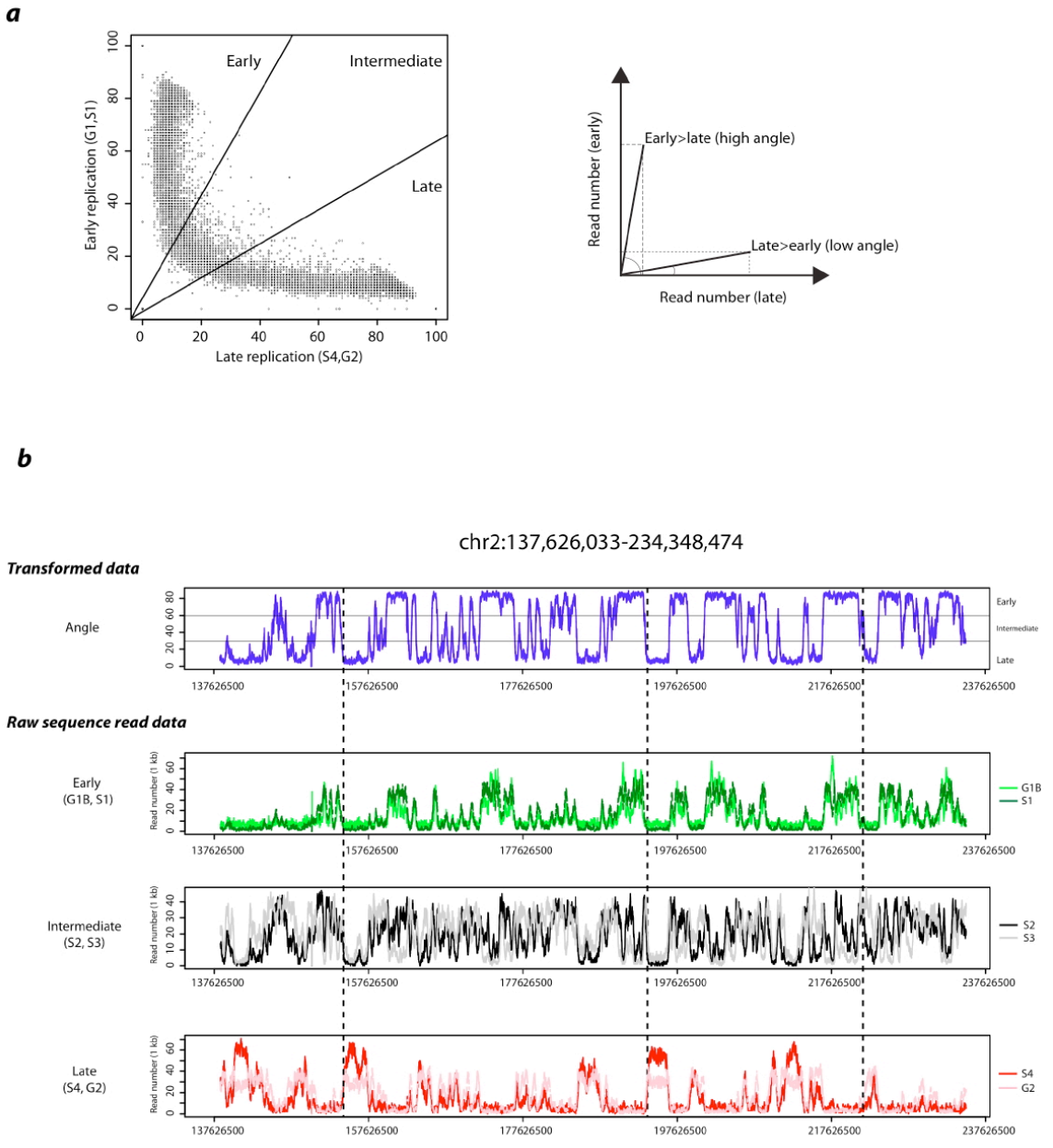

FIGURE S3

DNA hypermethylation occurs predominantly in actively-transcribed gene bodies. Cytosine methylation status in both promoter regions (-2 kb to +2 kb from annotated transcription start sites) and gene bodies (+2 kb from transcription start site to the end of the gene) was calculated as the average HpaII/Msp $\log_2$ ratio within the gene. Expression status is indicated by the averaged signal intensity from expression microarray data. Gene bodies show a clear shift towards increased methylation (values <0) when that gene is actively transcribed whereas promoters are hypomethylated (values >0) whether the associated gene is expressed or not.

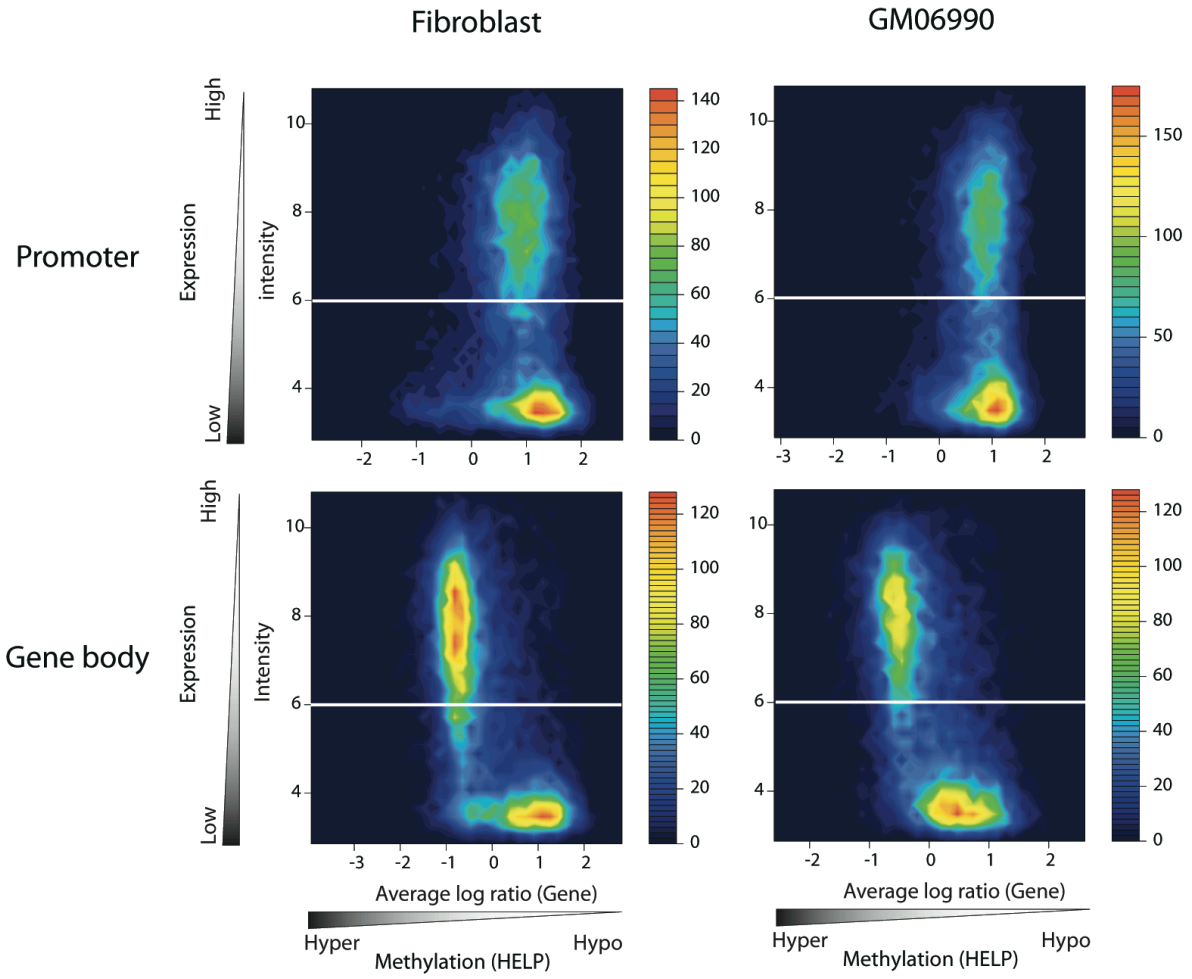

FIGURE S4

The genome-wide data were validated using quantitative techniques at a number of individual loci. In the upper panel, the *TLN1* gene is found to be expressed in both the GM06990 and fibroblast cells (as shown by the real time RT-PCR data, right panel), with HELP patterns showing promoter hypomethylation and gene body methylation. Quantitative bisulphite MassArray confirms the HELP data at four loci in this gene. The lower panel shows a gene (*BLNK*) expressed only in GM06990 cells, with HELP data again indicating promoter hypomethylation but with increased gene body methylation only in the expressing cell type. Bisulphite MassArray again confirms the HELP patterns of promoter hypomethylation and differences in gene body methylation.

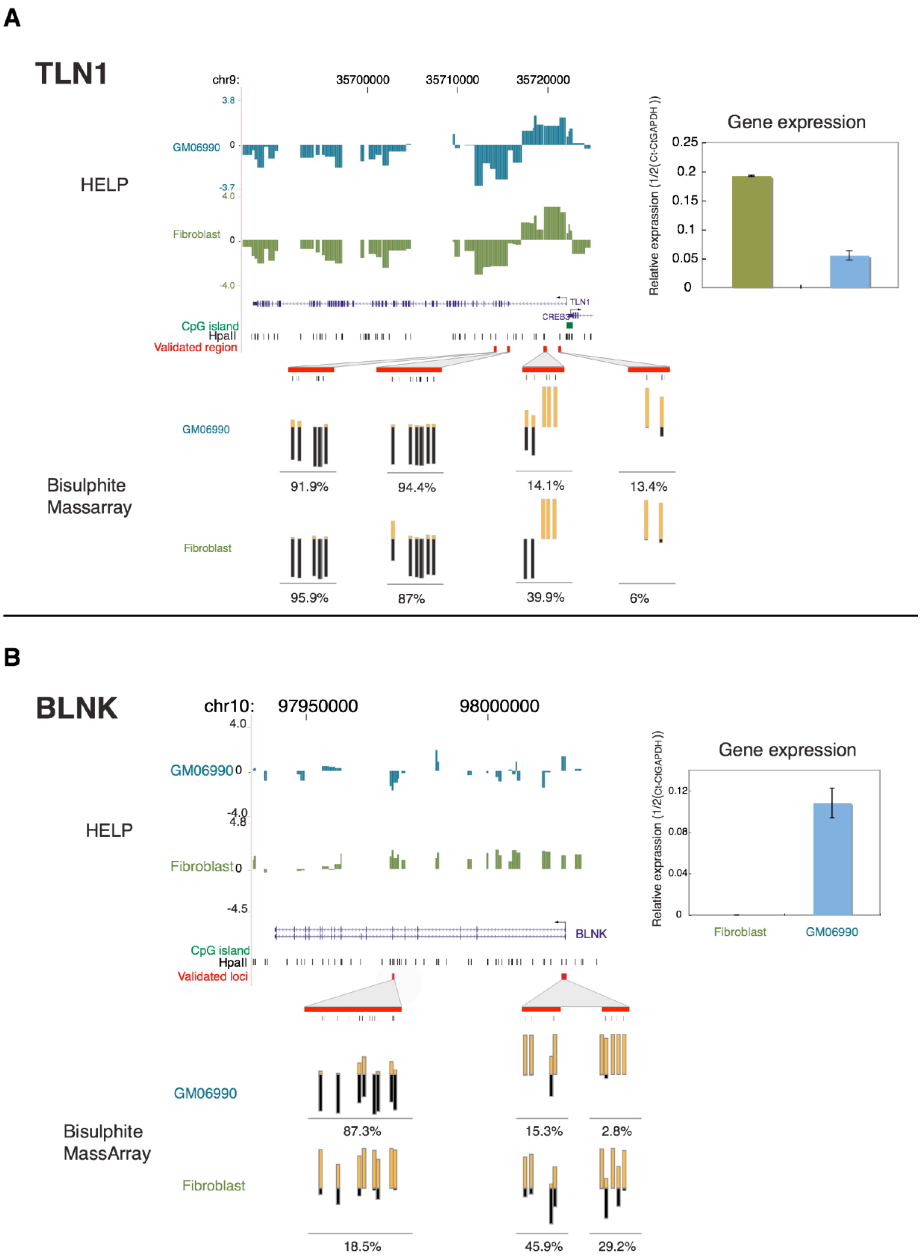

FIGURE S5

The magnitude of the effect of omitting gene bodies on DNA methylation is shown for the fibroblast genome. In panel A, we illustrate the shift towards decreased methylation (HELP ratio values>0) with density plots of HELP data values comparing 100 kb windows throughpout the genome with those omitting genic regions. The influence of gene bodies to increase methylation is marked.

In panel B, we show what happens when the analysis is restricted to gene-containing 100 kb windows, calculating the mean HELP ratio in the window with and without gene body data included. Those that change methylation values from rations <0 (methylated) to >0 (hypomethylated) are identified for each category of replication timing. The shift towards hypomethylation when we exclude gene bodies is most pronounced in early-replicating regions (higher angle values), a complementary means of illustrating the data shown in **Figure 4**.

Panel C shows results of an analysis similar to that of Aran *et al.* [Aran, 2011 #2311], illustrating that the removal of gene bodies from the dataset reduces the correlation strength between DNA methylation and replication timing but a substantial correlation remains, indicating that transcription-targeted DNA methylation is a substantial but not sole contributor to this correlation.

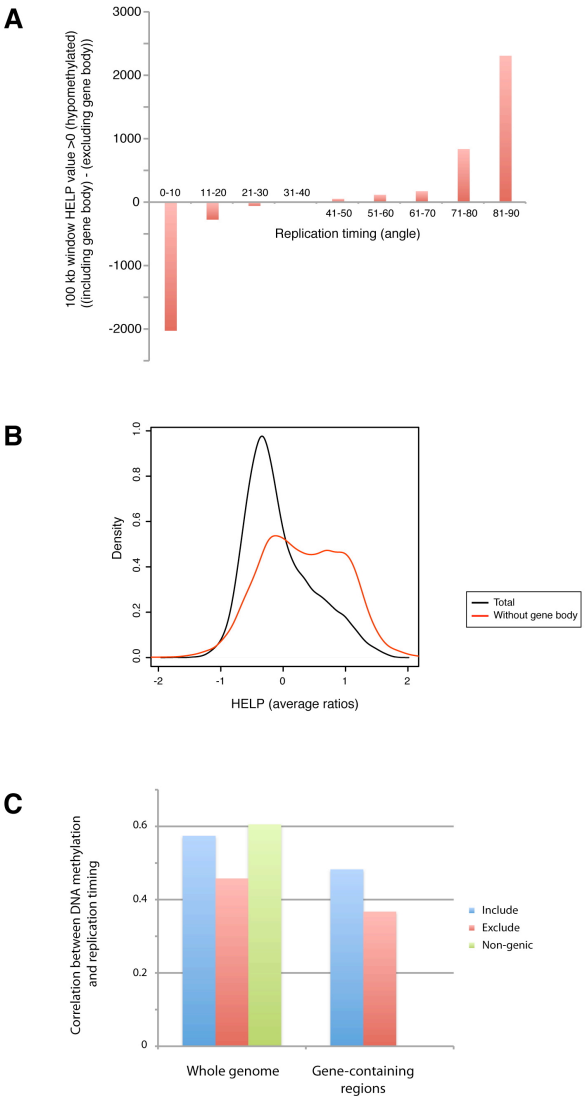

FIGURE S6

Using 100 kb sliding windows we find that actively-transcribed regions are replicated earlier than silenced genes, but that a small subset of early-replicating regions contain relatively inactive loci. In late-replicating regions, the genes present are mostly inactive.

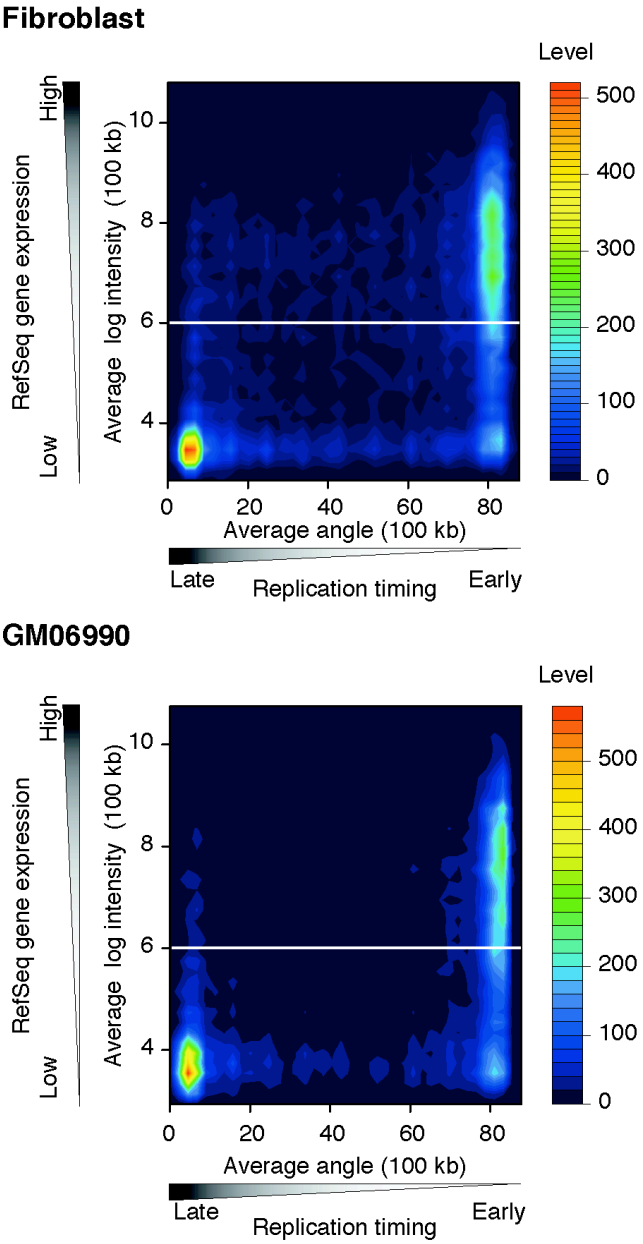

FIGURE S7

When DNase hypersensitivity is compared with DNA methylation (left), it is clear that the more methylated loci (values<0) have greater DNase hypersensitivity (increased numbers of DNase hypersensitive peaks). When DNA replication timing is tested, the earlier-replicating regions are associated with greater DNase hypersensitivity. As nuclease sensitivity is an indicator of chromatin compartments, this serves as a further indication that the more nuclease-sensitive euchromatin contains the more methylated DNA of the genome.

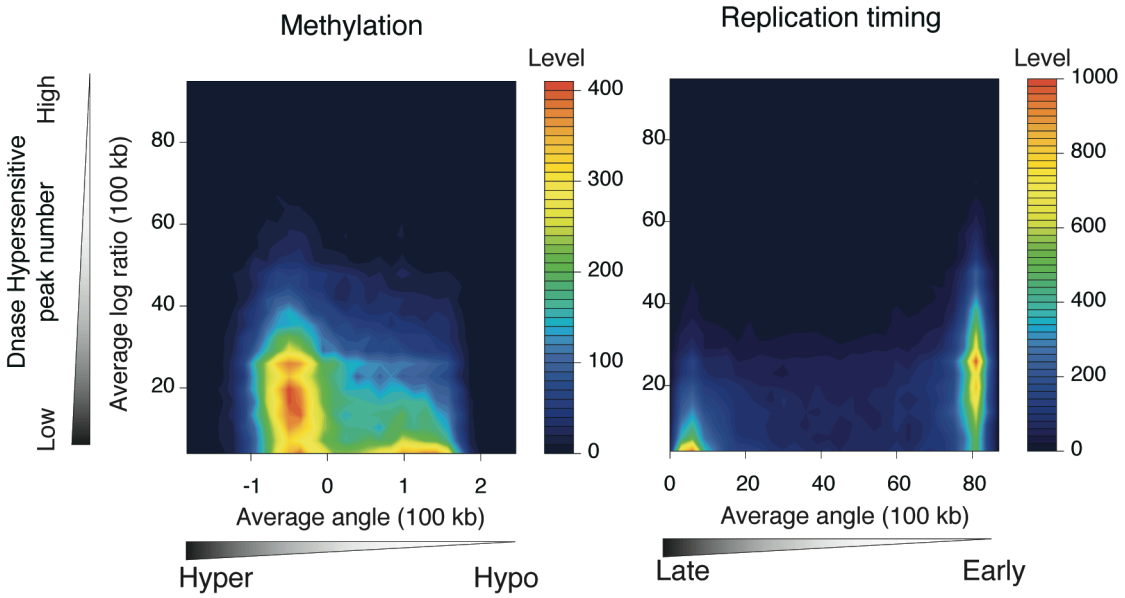

FIGURE S8

To illustrate how SOM data would look when there is little correlation with the outcome of interest, we generated a SOM using RefSeq gene and HpaII-amplifiable fragment (HAF) numbers as vectors. HAFs are the loci that are capable of being interrogated by the HELP assay, generating fragments between 50-2,000 bp. Overlaying DNA replication timing characteristics following the generation of this SOM shows no clear partitioning, indicative of the absence of correlation between these genomic variables and replication timing.

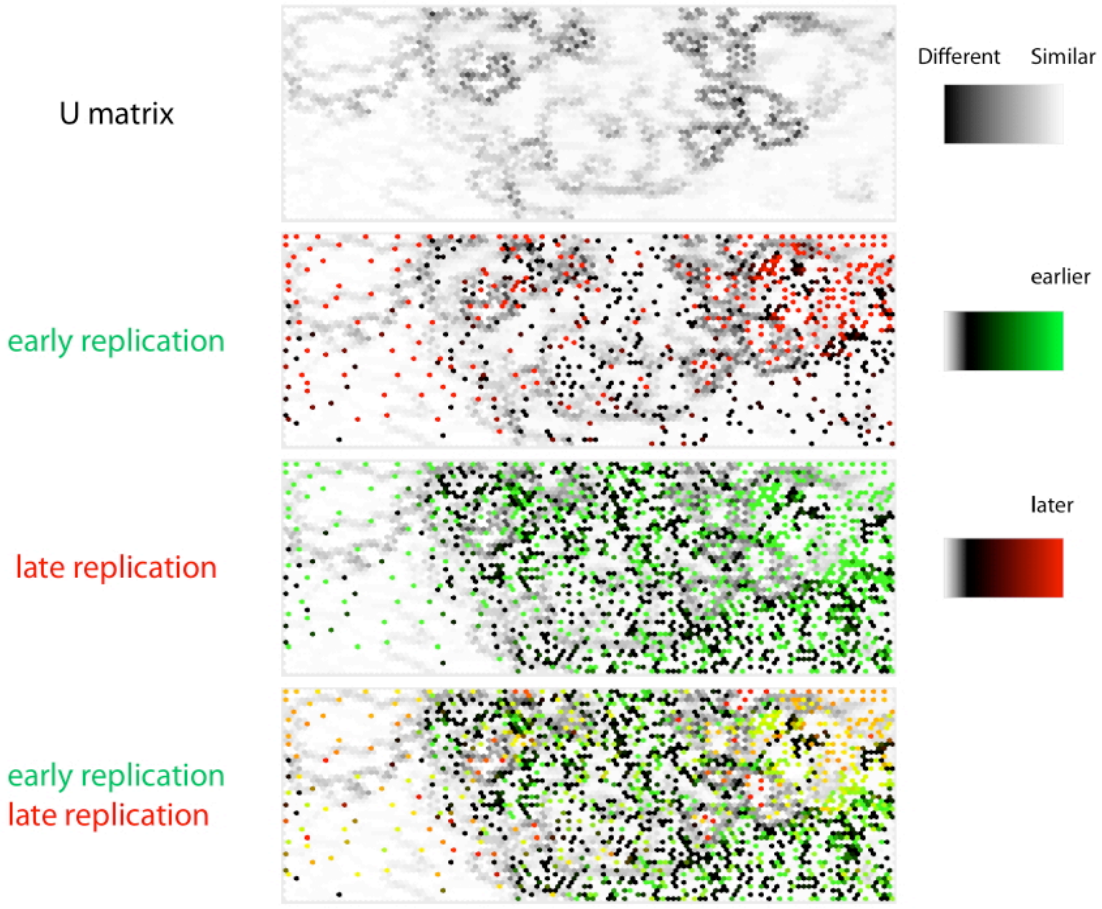

FIGURE S9

Analysis of data from the pilot phase of the ENCODE project [, 2007 #1675] from the GM06990 cell line, correlated with replication timing data from the same cells [Hansen, #2142]. In panel **(a)**, histone modifications are plotted as immunoprecipitation/input intensity ratios from ChIP-chip experiments and averaged in 100 kb windows (y-axis), compared with average replication timing for the same windows. Only the histone H3 lysine 9 trimethylation (H3K9me3) modification is enriched in late-replicating regions. Panel **(b)** shows the distributions of replication timing, cytosine methylation and H3K9me3 ChIP-chip data for this 1% of the human genome, notably demonstrating a shift towards earlier replication compared with the genome as a whole.