DNA replication timing analysis of human chromosome 22 at high resolution and different developmental states

Eric J. White,[1] Olof Emanuelsson,[2] David Scalzo,[3] Thomas Royce,[2] Steven Kosak[3], Edward Oakeley,[6] Sherman Weissman,[4] Mark Gerstein,[2] Mark Groudine,[3,5,7] Michael Snyder,[1,2,7] and Dirk Schübeler[6,7]

[1] Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, Connecticut 06520-8103, USA

[2] Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 065-8114, USA

[3] Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA

[4] Department of Genetics, Yale University, New Haven, Connecticut 06520-8005, USA

[5] Department of Radiation Oncology, University of Washington School of Medicine, Seattle, Washington 98195, USA

[6] Friedrich Miescher Institute for Biomedical Research, 4058 Basel, Switzerland

[7] Corresponding authors

Mark Groudine
Fred Hutchinson Cancer Research Center
1100 Fairview Ave. N
Seattle WA, 98109
Phone: (206) 667-4497
Fax: (206) 667-5894
Email: markg@fhcrc.org

Michael Snyder
KBT 926
Department of MCDB
Yale University
P.O. Box 208103
New Haven, CT 06520-8103
Phone: (203) 432-3510
Fax: (203) 432-3597
Email: michael.snyder@yale.edu

Dirk Schübeler
Friedrich-Miescher-Institute for Biomedical Research
Maulbeerstrasse 66
CH-4058 Basel
Switzerland
Phone: ++41-61-69-78269
Fax: ++41-61-69-73976
Email: dirk@fmi.ch

ABSTRACT

Duplication of the genome during S-phase of the cell cycle does not occur simultaneously; rather, different sequences are replicated at different times. The replication timing of specific sequences can change during development; however, the determinants of this dynamic process are poorly understood. To gain insights into the contribution of developmental state, genomic sequence and transcriptional activity to replication timing, we investigated the timing of DNA replication at high resolution along an entire human chromosome (chromosome 22) in two different cell types. The pattern of replication timing was correlated with respect to annotated genes, gene expression, novel transcribed regions of unknown function, sequence composition and cytological features. We observed that chromosome 22 contains regions of early and late replicating domains of 100kb-2Mb, many (but not all) of which are associated with previously described chromosomal bands. In both cell types, expressed sequences are replicated earlier than nontranscribed regions. However, several highly transcribed regions replicate late. Overall, the DNA replication timing profiles of the two different cell types are remarkably similar, with only nine regions of difference observed. In one case this difference reflects the differential expression of an annotated gene that resides in this region. Novel transcribed regions with low coding potential exhibit a strong propensity for early DNA replication. While the cellular function of such transcripts is poorly understood, our results suggest that their activity is linked to the replication timing program.

INTRODUCTION

Eukaryotic chromosomes initiate DNA replication from origins that fire at different times during S-phase of the cell cycle. Autoradiography experiments have revealed that mammalian DNA replication origins are spaced from 50-330 kb apart (1-3). Moreover, cytological and molecular studies have established that different chromosomal regions replicate at different times throughout S-phase (reviewed in (4)).

Several studies have correlated the timing of DNA replication with chromosomal features. Microscopic analysis of metaphase chromosome suggested that gene-rich R-bands replicate early and gene-poor G-bands replicate late (5, 6). While the exact nature of G and R band staining is unclear, R bands tend to be GC rich and G bands GC poor. Thus the differential replication timing of G and R bands is consistent with the recent report that replication timing correlates with GC content (7). However, whether R and G bands always correlate with DNA replication timing is not known.

Analysis of a few mammalian loci has revealed that alterations in transcriptional activity can coincide with changes in replication timing (4). Thus it has been speculated that replication timing and gene expression are functionally linked. A more comprehensive correlation of the timing of DNA replication with gene expression using DNA microarrays has been performed in *Drosophila* and human cells using cDNAs and BAC arrays, respectively (7, 8). These studies have demonstrated that expressed regions tend to replicate early in metazoan genomes. In both studies, however, only one cell type was analyzed and expression analysis was limited to annotated genes. Since transcriptional profiles are cell-type specific, we have investigated the relationship

between differential gene expression and replication timing in two cell types of different developmental origin.

Recent studies have revealed that many regions of the genome that lack known genes are nonetheless transcribed into polyadenylated RNA (9, 10), and it has been estimated that 30-50% of chromosomal transcription in a tissue or cell type is derived from such regions (11). Many of these Transcriptionally Active Regions (TAR) have low coding potential and thus do not represent classical protein encoding genes. The function of these RNAs is under intense investigation, and their abundance challenges much of our current thinking about transcriptional and posttranscriptional regulation. Importantly, it is not known how TAR activity relates to DNA replication.

The detection of TAR transcripts was made possible through the development of genomic microarrays, which cover large genomic regions with small tiled probes regardless of previously assigned function. Using such microarray, we have conducted an unbiased high-resolution analysis of DNA replication timing across human chromosome 22. DNA replication timing was compared with sequence characteristics including chromosome cytology, gene expression and TAR activity. Moreover, we analyzed replication timing across a large genomic DNA region in cells of different developmental origin. Our studies indicate that 1) although in general, transcribed regions are replicated early in S phase, many exceptions are noted; 2) most of chromosome 22 shows no cell type specific replication timing, while a small number of regions show cell-type replication timing differences; and 3) novel transcribed regions are positively correlated with early replication and tend to replicate even earlier than annotated genes.

MATERIALS AND METHODS

Cell culture: Human primary lung fibroblasts HFL-1 (ATCC No. CCL-153, (12)) were obtained from ATCC and cultured in DMEM medium containing 10% FCS at 37° C and 5% CO_2. Human B lymphoblastoid cells NC-NC were obtained from DSM, Germany (DSM-ACC 120,(13)) and cultured in RPMI + 10% FCS at 37°C 5% CO_2.

Spectralkaryotyping (SKY): Cells were fixed in 3:1 methanol:acetic acid and dropped onto glass slides, which were hybridized with spectral karyotyping reagent according to manufacturer's specifications (Applied Spectral Imaging, Inc.). Metaphase spreads were analyzed with SkyView software.

BrdU labeling: Bromodeoxyuridine triphosphate (BrdU, Sigma) was added to the media of logarithmically growing cultures to a final concentration of 50 µM. Sixty minutes after addition of BrdU, cells were washed twice in cold PBS, resuspended in 2.5 ml PBS, fixed by slowly adding 7.5 ml of cold ethanol and stored at –20°C.

FACS sorting and DNA immunoprecipitation: Cell preparation and fluorescence activated cell sorting (FACS) were essentially as described (14) with only minor modifications. Cells were sorted into S-phase fractions based on DNA content. Two gates were sorted, each representing roughly the first and last third of S-phase. 15,000 cells were collected from each gate directly into lysis buffer in the absence of salmon sperm DNA. DNA was purified as described (14), sonicated, denatured and immunoprecipitated with a monoclonal antibody specific for BrdU (Becton Dickinson).

PCR amplification and fluorescent labeling of BrdU enriched DNA: Amplification of the denatured and immunoprecipitated DNA was performed according to (15) with minor

modification (8). Size distribution and fluorescence of the product was confirmed by agarose gel electrophoresis followed by a fluorescence scan (Typhoon, Molecular Dynamics). Eight PCR reactions were performed for each of the fractions (early and late) and processed as described (16).

Control PCR: Primers to control for abundance in the early and late fraction were designed to amplify products of ~400 bp. Sequences are available on request. Reactions were performed with 2-5 ng DNA and 34 cycles using standard PCR conditions. PCR products were separated by agarose gel electrophoresis and visualized by ethidium bromide gel staining.

RNA extraction*:* Total RNA was Trizol extracted from asynchronously growing cells of each type according to the manufacturer's protocol (Invitrogen).

Expression profiling on Affymetrix GeneChip arrays: 10 μg of total RNA was reverse-transcribed and labeled using the Affymetrix cDNA and IVT kit as per manufacturer's instructions (Affymetrix), and hybridized to HG U133A GeneChips with washing and staining in an Affymetrix Fluidics Station 450. Samples were scanned in an Affymetrix GeneChip 3000 scanner. Each cell type was processed in triplicate.

Expression profiling on chromosomal array: 20μg of total RNA were reverse transcribed with M-MLV reverse transcriptase, using oligo (dT) primers in the presence of amino-allyl dUTP. The resulting cDNA was labeled with Cy5 according to the Amino Allyl cDNA Labeling Kit (Ambion) and hybridized to the microarray slides.

DNA microarray hybridiazation and analysis: Chromosomal microarrays were prehybridized with 80 μl prehybridization buffer (10X Denhardts, 5X SSC, 0.1% SDS,

1% BSA, 25% formamide v/v) for 2-3 hours at 42ºC. Coverslips were removed and microarrays were washed with H_20 and dried by centrifugation at 600 x g for 5 minutes. Cy3 and Cy5 labeled probes were resuspended in 80 µl hybridization buffer for 10 minutes at 37ºC and placed at 100ºC for 1 minute. 10 µg of labeled DNA was hybridized in duplicate to each of the three slides of the chromosome 22 microarray and hybridized for 16 hours at 42º C. Coverslips were removed in wash buffer Solution 1 (0.57x SSC and 0.3% SDS) and washed by submerging 20 times into Solution 1 and 20 times in Solution 2 (0.057x SSC). Microarrays were then dried by centrifugation at 600 x g for 5 minutes and scanned immediately with an Axon 4000A scanner and images analyzed with GenePix Pro 3.0 (Axon industries). All microarray data is available at www.csb.yale.edu/people/gerstein/olof/orirep/. A modified version of the ExpressYourself (17) array analysis package was used for background correction and intensity normalization. Intra-slide normalization was performed as described (18). Inter-slide scaling was done for each channel separately. To compare both cell lines the intensities on the different slides were adjusted to have identical mean and standard deviations ((17) and http://array.mbb.yale.edu/analysis/). The log2 ratios of early versus late intensities were calculated for each feature and mean and SD of replicates were calculated. Fragments with an SD above 0.7 were removed. Only array features with calculable early/late replications ratios obtained from at least half of the microarrays were plotted. Replication profiles of chromosome 22 were generated as described (8). Outliers were handled by a standard LOESS procedure (18).

RNA analysis from Affymetrix microarray: Expression values were estimated using the

Robust Multichip Analysis (RMA algorithm) (19). Subsequent analysis was performed using GeneSpring 6.2 (Silicon Genetics). Genes were required to have a minimum expression level of 50 in at least one cell line, differ in expression by at least 2-fold and pass a t-test ($p < 0.05$) with a Benjamini and Hochberg multiple-testing correction. The resulting datasets are available at the ArrayExpress archive of the EBI (http://www.ebi.ac.uk/arrayexpress/).

RNA analysis from chromosomal microarray: Only features that were measurable (non flagged) in at least 3 of the 4 repeat experiments were analyzed for expression. A fragment was considered expressed if the average foreground over background ratio was above 2.5 and the signal above 200 pixels. RNA expression versus DNA replication timing plots were performed as described (8). The chromosome-wide expression profile was performed as described (9).

RESULTS

Construction of DNA replication timing profiles for two cell types

To construct a high resolution DNA replication profile of a mammalian chromosome and compare its replication timing in different cell lines, we utilized a human chromosome 22 genomic DNA microarray (9). This DNA microarray contains continuously tiled fragments which represent >90% of nonrepetitive chromosome 22 DNA; the average length of fragments is ~800 base pairs. The array is ordered from the centromere to the end of the long arm of chromosome 22. To examine potential changes in DNA replication timing and gene expression in distinct epigenetic backgrounds, cells derived from two discrete tissues were analyzed. We used primary fibroblast cells (HFL-1, (12)) and the lymphoblastoid cell line, NC-NC (13). To exclude that differences in replication timing between the two cell lines might be a consequence of chromosomal abnormalities, we performed detailed spectral karyotyping (SKY) (19). In both cases, normal autosomal karyotypes without rearrangements or translocations were observed (Figure 1).

The protocol for constructing a DNA replication-timing profile is described in Figure 2. Briefly, asynchronously growing cells were pulse labeled with 5-bromodeoxyuridine (BrdU) for one hour, ethanol fixed and sorted by FACS analysis according to DNA content (20). A representative FACS plot for each cell type is shown. Cells from the first and last third of S-phase were collected. Following DNA isolation, sonication and denaturation, BrdU-containing DNA was enriched by immunoprecipitation. BrdU labeled DNA from the first third of S-phase was amplified by

PCR with Cy5 end labeled primers, and immunoprecipitated DNA from the last third was labeled with Cy3 (8). Equal amounts of the two labeled probes were co-hybridized to the chromosome 22 genomic DNA microarray. In addition, "dye swap" experiments, in which early replicating DNA was labeled with Cy3 and late replicating DNA was labeled with Cy5, were performed to account for any influence of the individual fluorophores on experimental results. In total, 3 and 4 independent experiments were performed for HFL-1 and NC-NC, respectively. Each independent experiment consisted of four separate hybridizations; two of which were "dye swaps".

The resulting array data was analyzed using the ExpressYourself program, and features were normalized by total fluorescence intensity normalization (17, 18). To construct a chromosomal map of replication timing, the log ratio of early/late replication was calculated and plotted according to genomic position. Early replicating regions are indicated as a positive log ratio, while late replicating regions are indicated by a negative log ratio. In this analysis, neighboring genomic probes tended to have comparable ratios, indicating that they are replicated at similar times in S phase. This observation was validated by autocorrelation analysis, which revealed significant autocorrelation (>0.05) for up to 58 and 35 fragments for HFL-1 and NC-NC lymphoid cells, respectively (Figure 3A). Given an average spacing of 2 kilobases from the start of one fragment to the next fragment, this corresponds to an autocorrelation over 116 and 70 kb for HFL-1 and NC-NC cells, respectively. As a control, we pooled early and late replicating DNA and subsequently labeled and then hybridized this DNA to the array. As shown in Figure 3B, no considerable enrichment of early or late replicating DNA was observed with this

sample. Thus, we conclude that the observed enrichments in early and late replication timing are significant. As a separate control, we performed single-gene PCR of a subset of sequences. This analysis revealed that early replicating regions detected by microarray are correspondingly abundant in the early S phase fraction when assayed by PCR, while late replicating regions are more abundant in the late S fraction (Figure 3C). To generate the complete profile for both lines as shown in Figure 4, LOESS smoothing was performed to fit the replication timing dataset along a best fit (21).

DNA replication timing profile of primary human lung fibroblasts

Hybridizations of labeled BrdU immunoprecipitated DNA to the chromosome 22 microarray produced a high resolution replication timing profile (Figure 3). Analysis of the replication timing profile of HFL-1 fibroblast cells revealed that both early and late replicating regions are present throughout chromosome 22. A total of 24 early replicating and 24 late replicating regions were identified (defined as segments above or below the base line); the sizes of these regions vary from ~100 kilobases to 2 megabases. Of these regions, 14 early and 9 late replicating regions are statistically significant, when a standard deviation cutoff (both above and below the median replication timing ratio) is applied (Figure 4).

DNA replication timing patterns often, but not always, correlate with chromosome cytology

Correlation of the different replicating regions in human lung fibroblasts with

chromosome cytology reveals that several late replicating regions overlap with expected G bands at the 850 band resolution (22). In particular, regions 10.0-13.0 Mb and 15.5-19.0 Mb exhibit strong overlap. These two regions also show a corresponding decrease in percentage of chromosomal GC content (Figure 4). Two other G-bands located at 5.2-7.2 Mb and 31-33.1 Mb occur in regions of the chromosome where the replication profile decreases and nears the baseline, suggesting that not all G-bands replicate late. Interestingly, these regions do not have a lowered GC content. Indeed, it has been observed that the degree of GC content directly correlates with G-band replication timing (3, 23). In addition the G-band located at 23.4 – 25.9 Mb was not clearly identified as late replicating. However, our DNA microarray contains few features in the region between 25-26 Mb and our replication timing profile identified a late replicating, GC poor region around 24 Mb at the centromeric end of this G-band. As expected, pericentromeric DNA, composed of large regions of heterochromatin replicates late in S-phase. Finally, consistent with previous data (7, 24), the subtelomeric region replicates early in S-phase.

Comparison of DNA replication timing profiles for the lymphoblast and fibroblast cell lines

Analysis of the DNA replication timing of lymphoblastoid NC-NC cells, revealed extensive similarity with that of the fibroblasts. Similar to the fibroblast line, 26 regions of both early and late replication timing were observed with 15 and 13 peaks above and below one standard deviation from the baseline, respectively. A t-test of the replication profiles indicated that over 99% of the chromosome fragments replicated at a

similar time in lymphoblast and fibroblast cells. Interestingly, nine regions that differed in replication timing ($p < 0.05$) were observed. Six of these regions replicated earlier in the HFL-1 cells than the NC-NC cells, while the remaining three replicated later. Thus, the timing of replication in several chromosomal regions is specific for a distinct cell type.

Correlation of DNA replication timing with transcription

Total RNA was prepared from both cell lines, reverse transcribed with oligo (dT) primers, labeled with Cy5, and hybridized to the chromosome 22 microarray. The results were analyzed both globally and as a function of chromosome position.

For the global analysis, microarray hybridization results from probes of annotated genes (Sanger 2.3 and Ensembl version 18.34.1) were binned into groups of 50 similar replication timing ratios and the percentage of features which showed at least a 2.5 fold increase in fluorescence from RNA over background (with a cutoff of 200 pixels) in each bin was calculated and plotted as a likelihood of expression relative to DNA replication timing (Figure 5). A logistic regression curve was fitted to the data points. The strongly positive slope of the fibroblast and lymphoblast regression curves indicates that early replicated DNA is more likely to be expressed than is later replicated DNA. Interestingly, in both cell types, very late replicating DNA is a clear exception to this trend, as a higher percentage of this very late replicating fraction is transcribed (note the increase at –0.3 in Figure 5).

Novel transcribed sequences replicate early

In addition to known genes, we were interested in investigating the DNA replication timing of previously unidentified and unannotated novel transcribed sequences. These transcriptionally active regions (TAR) constitute more than a third of chromosomal transcription. Most of the transcripts have low coding potential, and neither their function nor whether they behave similar to annotated genes with respect to DNA replication timing is known. To address this question, we plotted the total percent of RNA expression against DNA replication timing separately for annotated and unannotated sequences (Figure 5). This global analysis revealed that TAR tend to replicate earlier than non-transcribed sequences suggesting that their activity is linked to replication timing similarly to that of annotated genes. Interestingly, in both cell types the correlation of TAR and early replication is more pronounced than in the case of annotated genes.

Differential replication timing of an immunoglobulin locus correlates with expression

In addition to the global comparison, we examined in detail those regions that showed differential replication timing or expression in the two cell types. In the majority of differentially replicated regions, we did not detect significant differences in gene expression. However, the immunoglobulin lambda-like polypeptide 3 locus (IGLL3) is an

exception. This gene was expressed higher in the lymphoblastoid cells than in the fibroblasts, and the region in which it lies (9.0-9.5 Mb) is replicated significantly earlier in the lymphoblastoid cells. Thus, in this instance, replication timing correlates with differential gene expression. To investigate this further, we performed the "reverse" experiment and compared the replication timing of 55 genes that were differentially expressed as measured in both our chromosomal array and Affymetrix cDNA arrays (Materials and Methods). These genes showed no differences in replication timing at any resolution investigated, suggesting that in the genomic regions analyzed, changes in expression are associated with changes in replication at only a small subset of loci.

Highly transcribed late replicating regions

We also examined DNA replication timing and expression as a function of chromosome position. DNA replication timing and total chromosomal RNA expression for both cell types was plotted as a sliding window of the percentage of transcribed microarray fragments in 100 kb, as previously described (9) (Figure 4). Significant novel transcription covering an extensive region was found in the late replicating region at 18.0 Mb and between 11.0-11.5 Mb (9). Although, in general, gene expression correlates well with early replication timing along the chromosome, ~6 regions of late replicating DNA demonstrated a high percentage of transcribed sequences. The regions, between 10-10.5, 18.0-18.2 and 23.0-23.5 Mb, replicated very late, yet were rich in annotated, expressed genes. In addition, the late replicating regions at ~14.9, 17.8 and 24.1 Mb were also heavily transcribed. Thus, although our experiments revealed a global correlation

between expression and early replication, there are specific regions that show an opposite trend, indicating that early replication is not a simple consequence of transcriptional activity and vice versa.

DISCUSSION

The regulation of replication timing, as well as its function in transcriptional regulation is not well understood and current models are largely based on the examination of a few genes. To examine this relationship more comprehensively we have generated a high resolution DNA replication timing map for a mammalian chromosome and in addition compared replication timing between two cell types of different developmental origin. Our analysis reveals that human chromosome 22 contains 24-26 early and late replication regions ranging in size from 100 kb-2Mb. Many of these regions agree well with early and late replicating R and G bands and in general, we found the G bands of lowest GC to be the latest replicating, consistent with the work of others (3, 23).

Previous studies in both *Drosophila* and human cells have demonstrated that early timing of DNA replication correlates with gene expression on a global scale. Our study supports this finding, but also extends this observation to previously unidentified TARs, which reside outside of annotated genes and constitute up to one-third of nuclear transcription (11). Interestingly we find that TARs replicate even earlier than annotated genes. Transcribed regions have an open chromatin conformation (25), and it is tempting to speculate that TAR activity might facilitate early replication, perhaps by mediating a chromatin structure necessary to initiate early origin firing (26). Alternatively, early replication itself could influence transcriptional activity of these sequences.

While our global analysis of transcribed sequences revealed a correlation between transcribed sequences and replication timing, examination of all chromosomal sequences reveals that some highly transcribed regions are replicated late. Only a few

examples of this class of genes have been reported previously for mammalian cells (27, 28). Currently, the basis of this exception is not known, and we can only hypothesize that late replication might be involved or required for the proper regulation of this subset of sequences.

Gene expression has been observed to correlate with DNA replication timing. As gene expression can differ significantly between cell types in higher eukaryotes, chromosomal regions might be expected to show significant differences in replication timing between different cell types. However, our comparison of DNA replication timing in human fibroblast and lymphoblast cells revealed that replication timing of chromosome 22 is largely indistinguishable between these cell types--only nine areas of significant differences in timing were observed. One of these areas contains IGLL3, an immunoglobulin gene, which is expressed at higher levels and replicates earlier in the lymphoblastoid cells. We also observed many other regions (55) in which gene expression differs significantly between the two cells types that did not exhibit differences in replication timing regardless of the window size examined. Thus, differences in replication timing do not necessarily correlate with differences in gene expression (29).

Can the results we obtained on chromosome 22 be extrapolated to the remainder of the human genome? Chromosome 22 replicates earlier than other chromosomes (7), has a high gene density and contains many housekeeping genes. Their activity in combination with the high TAR activity may promote early replication by mediating a high percentage of active transcription throughout the chromosome, which might mask

potential effects of differentially expressed genes on replication. Consequently, replication timing differences might be more pronounced in chromosomes or chromosomal regions, which have a lower gene-density. This possibility could be addressed by examining the replication timing of other human chromosomes using a similar approach to the one presented here for chromosome 22.

ACKNOWLEDGEMENTS

We thank M. Smith for critical reading of the manuscript and Christiane Wirbelauer for technical assistance. E.W. is supported by a Ruth L. Kirchstein NIH-HGRI postdoctoral fellowship, HG002715. O.E. is supported by a Knut and Alice Wallenberg Foundation postdoctoral fellowship. S.K. is a fellow of the Jane Coffin Childs Memorial Fund for Biomedical Research. Work in the Snyder, Gerstein and Weissman labs was supported by a NIH-HGRI Center for Excellence in Genome Sciences grant, HG02357. M. G. is supported by NIH grants DK44746 and HL65440. E. O. and D. S. are supported by the Novartis Research foundation.

REFERENCES

1. Huberman, J. A. & Riggs, A. D. (1968) *J. Mol. Bio.* **32,** 327-341.
2. Hand, R. (1978) *Cell* **15,** 317-25.
3. Dutrillaux, B., Couturier, J., Richer, C. L. & Viegas-Pequinot, E. (1976) *Chromosoma* **58,** 51-61.
4. Cimbora, D. M. & Groudine, M. (2001) *Cell* **104,** 643-6.
5. Drouin, R., Holmquist, G. P. & Richer, C. L. (1994) *Adv. Hum. Genet.* **22,** 47-115.
6. Federico, C., Saccone, S. & Bernardi, G. (1998) *Cytogenet. Cell Genet.* **80,** 83-88.
7. Woodfine, K., Fiegler, H., Beare, D. M., Collins, J. E., McCann, O. T., Young, B. D., Debernadi, S., Mott, R., Dunham, I. & Carter, N. P. (2004) *Hum. Mol. Genet.* **13,** 191-202.
8. Schubeler, D., Scalzo, D., Kooperberg, C., Steensel, B. v., Delrow, J. & Groudine, M. (2002) *Nat. Genet.* **32,** 438-442.
9. Rinn, J. L., Euskirchen, G., Bertone, P., Martone, R., Luscombe, N. M., Hartman, S., Harrison, P. M., Nelson, F. K., Miller, P., Gerstein, M., Weissman, S. & Snyder, M. (2003) *Genes Dev.* **17,** 529-40.
10. Kapranov, P., Cawley, S. E., Drenkow, J., Bekiranov, S., Strausberg, R. L., Fodor, S. P. & Gingeras, T. R. (2002) *Science* **296,** 916-9.
11. Kampa, D., Cheng, J., Kapranov, P., Yamanaka, M., Brubaker, S., Cawley, S., Drenkow, J., Piccolboni, A., Bekiranov, S., Helt, G., Tammana, H. & Gingeras, T. R. (2004) *Genome Res.* **14,** 331-42.
12. Bruel, S. D., Bradley, K. H., Hance, A. J., Schafer, M. P., Berg, R. A. & Crystal, R. G. (1980) *J. Bio. Chem.* **255,** 5250-60.
13. MacLeod, R. A. & Bryant, P. E. (1992) *Mutagenesis* **7,** 285-90.
14. Cimbora, D. M., Schubeler, D., Reik, A., Hamilton, J., Francastel, C., Epner, E. M. & Groudine, M. (2000) *Mol. Cell. Bio.* **20,** 5581-5591.
15. Lieb, J. D., Liu, X., Botstein, D. & Brown, P. O. (2001) *Nat. Genet.* **28,** 327-34.
16. van Steensel, B., Delrow, J. & Henikoff, S. (2001) *Nat. Genet.* **27,** 304-8.
17. Luscombe, N. M., Royce, T. E., Bertone, P., Echols, N., Horak, C. E., Chang, J. T., Snyder, M. & Gerstein, M. (2003) *Nucleic Acids Res.* **31,** 3477-3482.
18. Quackenbush, J. (2002) *Nat. Genet.* **32,** 496-501.
19. Schrock, E., DuManoir, S., Veldman, T., Schoell, B., Wienberg, J., Ferguson-Smith, M. A., Ning, Y., Ledbetter, D. H., Bar-Am, I., Soenksen, D., Garini, Y. & Reid, T. (1996) *Science* **273,** 494-497.
20. Hansen, R. S., Canfield, T. K., Lamb, M. M., Gartler, S. M. & Laird, C. D. (1993) *Cell* **73,** 1403-9.
21. Cleveland, W. S. & Devlin, S. J. (1988) *J. Am. Statistical Assoc.* **83,** 596-610.
22. Francke, U. (1994) *Cytogenet. Cell Genet.* **6,** 206-219.
23. Saccone, S. & Bernardi, G. (2002) *Meth. Cell Sci.* **23**.
24. Smith, Z. E. & Higgs, D. R. (1999) *Hum. Mol. Genet.* **8,** 1373-86.

25. Schubeler, D., MacAlpine, D. M., Scalzo, D., Wirbelauer, C., Kooperberg, C., van Leeuwen, F., Gottschling, D. E., O' Neill, L. P., Turner, B. M., Delrow, J., Bell, S. P. & Groudine, M. (2004) *Genes. Dev.* **18,** 1263-71.
26. Gilbert, D. (2002) *Curr. Opin. Cell Bio.* **14,** 377-383.
27. Hatton, K. S., Dhar, V., Brown, E. H., Iqbal, M. A., Stuart, S., Didamo, V. T. & Schildkraut, C. L. (1988) *Mol. Cell. Bio.* **8,** 2149-2158.
28. Gartler, S. M., Goldstein, L., Tyler-Freer, S. E. & Hansen, R. S. (1999) *Hum. Mol. Genet.* **8,** 1085-1089.
29. Azuara, V., Brown, K.E., Williams, R.R., Webb, N., Dillon, N., Festenstein, R., Buckle, V., Merkenschlager, M., & Fisher, A. G. (2003) *Nat. Cell Bio.* **5,** 668 - 674.

FIGURE LEGENDS

Figure 1. **Karyotypic analysis of primary fibroblast (HFL-1) and lymphoblastoid cell line (NC-NC)**. Spectralkaryotyping was used to identify chromosome number and integrity. No abnormality was detected on the autosomes, while a subfraction of the female NC-NC cells lacks the inactive x-chromosome as previously described (13).

Figure 2. **Experimental strategy**. Histogram plots of DNA content for each cell type. Asynchronously replicating cells were labeled with BrdU for 1 hour and sorted by FACS for DNA content. DNA from cells from the first and last thirds of S-phase was extracted and newly replicated DNA was immunoprecipitated with αBrdU antibody. Immunoprecipitated DNA was labeled with Cy3 or Cy5 and hybridized to the chromosome 22 DNA microarray.

Figure 3**. Control experiments**. *(A)* Autocorrelation analysis to measure the chromosomal extent of similar replication timing. The autocorrelation function is calculated for an increasing number of neighboring sequences. Positive autocorrelation exists if neighboring spots tend to be alike, here replicate at a similar time in S-phase. Shown is the autocorrelation plot of replication timing ratios prior to smoothing (see Figure 4) to identify the lengths of significantly similar (>0.05) DNA replication timing. *(B)* DNA replication timing profile of HFL-1 compared to control hybridization. The DNA replication timing profile of HFL-1 (green) is plotted with the mixed early and late

S-phase DNA control (grey). Shown are the first 20 Mbp of chromosome 22 and the chromosomal positions of amplicons used in single-gene controls. *(C)* Single gene control PCR of early and late replicating regions for the HFL-1 cell type. Amplicon names refer to microarray features from which ~400 bp fragments were amplified. In each case the control reaction confirms early or late replication timing as revealed by the microarray analysis.

Figure 4. **DNA replication profiles for HFL-1 and NC-NC**. DNA replication timing ratios (ratios are $\log_2$ transformed) were plotted according to chromosomal position. Shown is the non-repetitive part of chromosome 22 extending from the centromere (upper left) to the telomere of the q-arm (lower right). Data was smoothed by LOESS smoothing and the best fit plots are shown for HFL-1 (green line) and NC-NC (red line). The black mark along the baseline indicates regions of DNA replication timing that were present on the microarray and compared between both cell types. Associated orange hash marks indicate regions of significant difference ($p<0.05$) in DNA replication timing. Percentage of chromosomal GC content is plotted as an orange line near the top of the plot as a sliding 100 Kb window. Percentage of RNA expression for both cell types is plotted just below GC content as a sliding 100 Kb window along the chromosome for HFL-1 (cyan line) and NC-NC (grey). Annotated genes that are expressed in HFL-1 and NC-NC are indicated as blue squares above (HFL-1) and below (NC-NC) the baseline. Locations of cytological G-bands are indicated by grey boxes along the chromosome position axis (22).

Figure 5. **Relation of replication timing to transcription and CG content.** mRNA expression and novel transcription (TAR) correlate with early DNA replication timing. Microarray features were binned in groups of 50 features with similar replication timing ratios. The percentage of expressed features in each bin was calculated and plotted according to DNA replication timing ratio. A logistic regression curve was fitted to the data points. Red data points represent only those fragments overlapping with annotated genes and blue data points represent only those fragments that do not overlap with annotated genes

Figure 1 White et al. (7cm X 1 column)

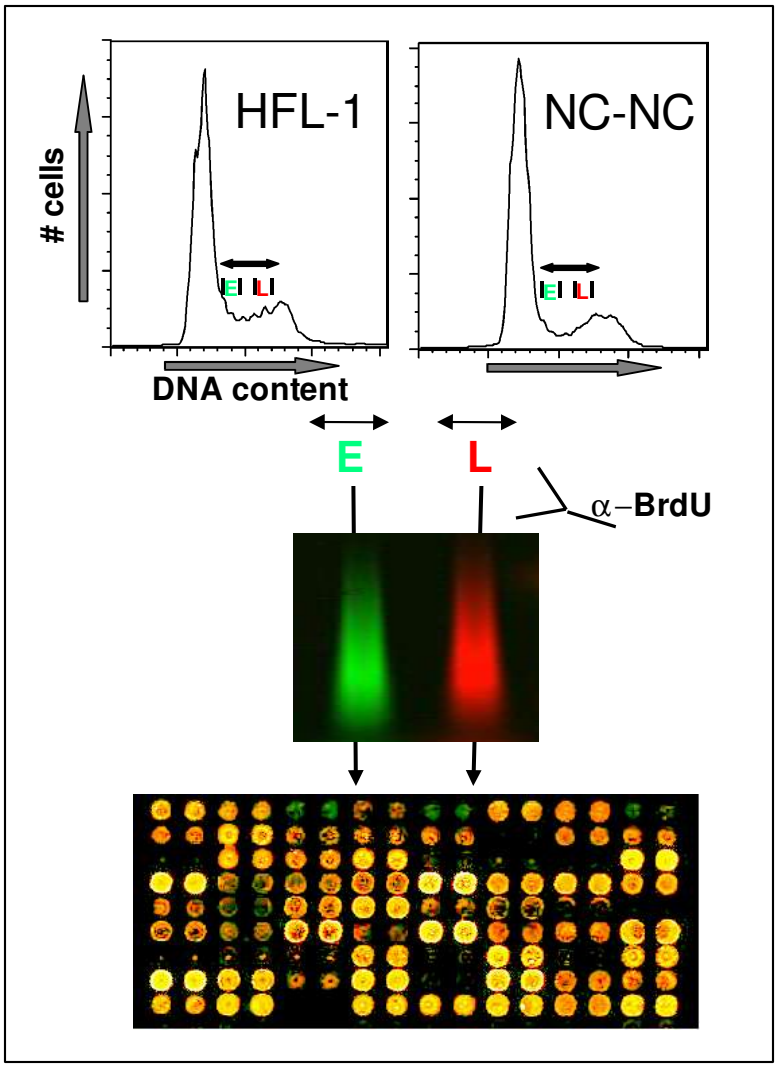

Figure 2 White et al (7.5cm X 1 column)

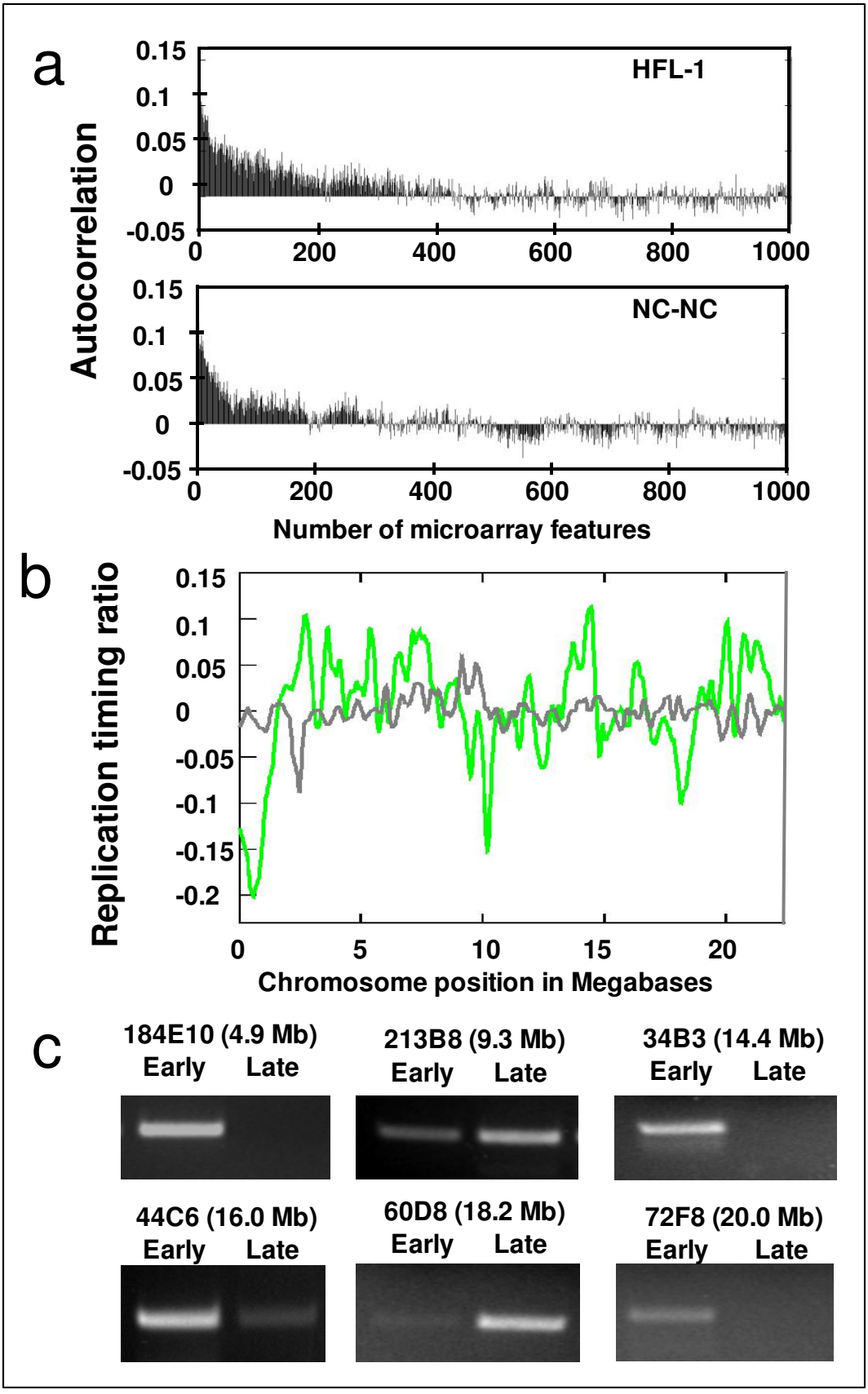

Figure 3 White et al. (11 cam X 1 column)

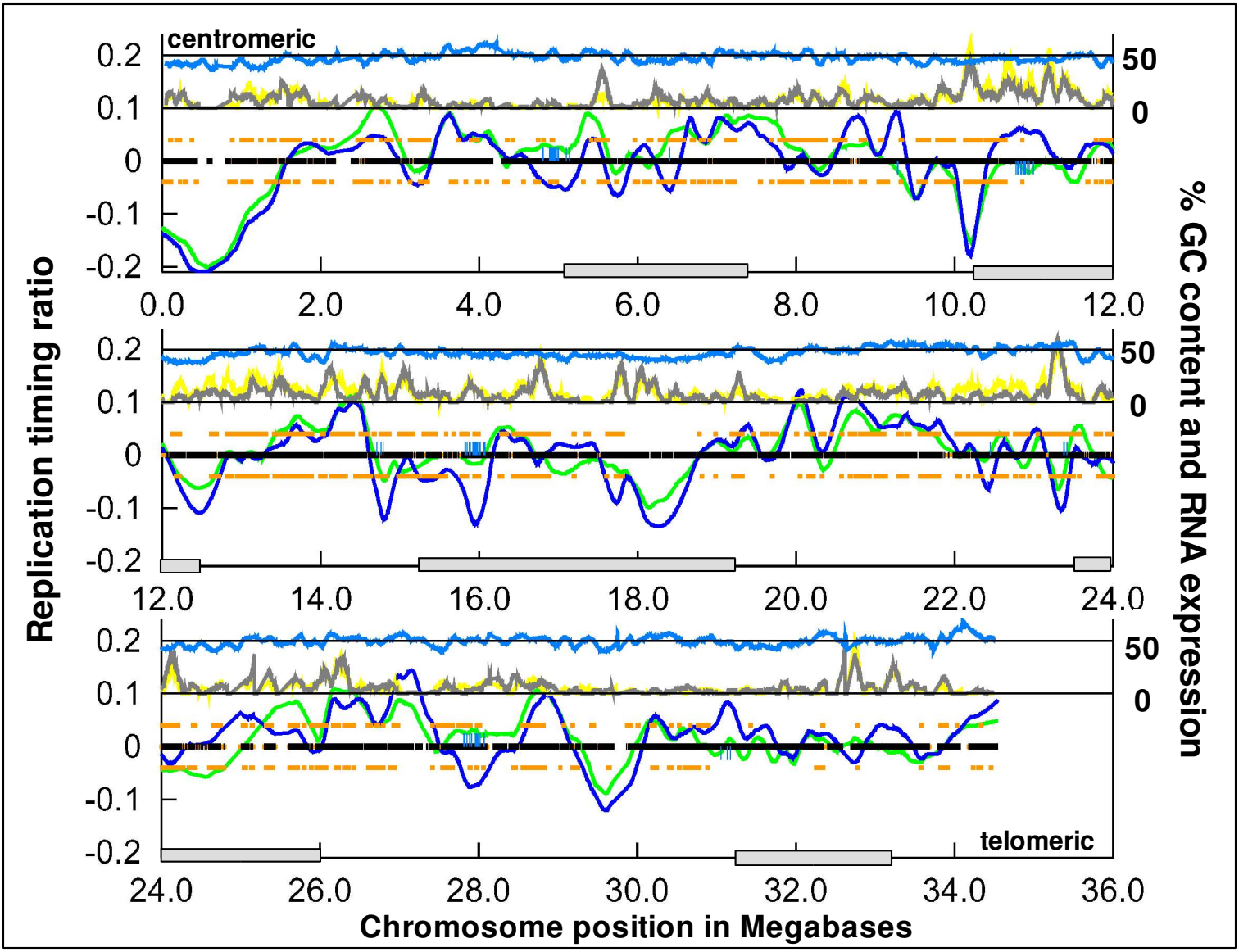

Figure 4 White et al (10 cm X 2 column).

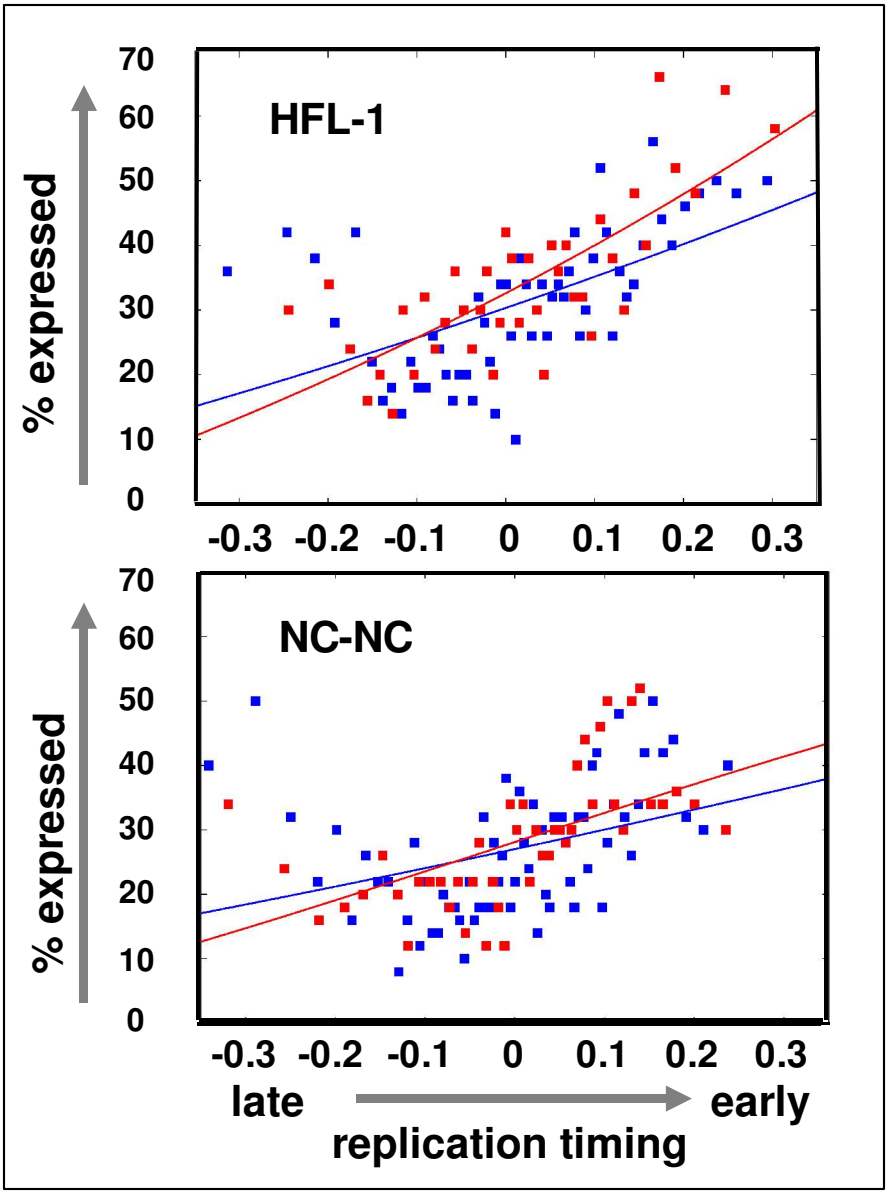

Figure 5 White et al. (8 cm X 1 column)