Extraction of Human DNA Replication Timing Patterns from Discrete Microarray Data

Anindya Dutta[1], Neerja Karnani[1], Ankit Malhotra[1], Gabriel Robins[2], and Christopher M. Taylor[1,2,3,4]

1-Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, VA 22908
2-Department of Computer Science, University of Virginia, Charlottesville, VA 22904
3-Department of Computer Science, University of New Orleans, New Orleans, LA 70148
4-Research Institute for Children, Children's Hospital, New Orleans, LA 70118
{adq8,nk6t,am3cp,gr3e}@virginia.edu,{taylor}@cs.uno.edu

Abstract. Effective reproduction is essential for the survival and proliferation of any organism, from the birth of new offspring to the reproduction of individual cells. Each portion of a cell's DNA must be copied exactly once during the replication phase of its cell cycle to ensure viability. In humans, this is achieved by a complex pattern of replication origins and terminations along the chromosomes until the final product is realized. DNA Tiling Microarrays are utilized to assay discrete pools of DNA replicated during different parts of the replication phase. We present a generalized framework for analyzing this discrete timing data to recover a relatively continuous profile of the DNA replication timing. This approach can be used to assay DNA replication timing over a variety of human cell lines or extended to other organisms.

Key words: human replication timing, DNA tiling microarrays

1 Introduction and Related Work

DNA replication is a crucial step in the life cycle of a cell as faithful reproduction of the genetic material is essential for viability of daughter cells [1]. In higher eukaryotes, this process is carried out via the firing of numerous origins of replication along the chromosomes in order to replicate the DNA in a reasonable amount of time. The replication forks emanating from these origins work in parallel to replicate the entire genome, producing a complex schedule of DNA replication timing.

The replication time of individual areas of the genome is of interest for a variety of reasons including the influence of chromatin structure, transcriptional activity, and the possibility of allelic variation in replication timing [2]. Hence, replication timing has been studied in a number of model organisms including

M. Chetty, S. Ahmad, A. Ngom, S. W. Teng (Eds.): PRIB2008, Supplementary Conf. Proc., pp.159-169, 2008.

Saccharomyces cerevisiae [3, 4], *Schizosaccharomyces pombe* [5], and *Drosophila melanogaster* [6, 7]. Microarray technology has played a major role in many of the studies of DNA replication timing [8–10], and more recent studies have extended these techniques to human cell lines [11–16].

One important method used for high resolution studies of DNA replication timing is the isolation of discrete pools of DNA replicated during different parts of S-phase, followed by their hybridization to genome tiling microarrays. We have adopted this method for our work in human DNA replication timing, and developed algorithms to analyze such data effectively and efficiently. In this paper, we present algorithms and techniques for recovery of a relatively continuous profile of DNA replication timing from these discrete pools of replicated DNA.

2 Methods

2.1 Data Collection

The starting point for our analysis is a set of discrete pools of DNA replicated during different parts of S-phase that have each been hybridized to a tiling microarray. In order to harvest enough DNA for the arrays, biologists synchronize a population of cells at the entry point of S-phase. The cells are then released together into the replication phase. Labeling methods are used to isolate the portions of the DNA which replicate during each part of S-phase. This synchronization and release can introduce non-trivial 'synchronization error' whereby each cell c of the population moves with some delay Δ_c with respect to the actual time of release t. Hence, the time at which cell c begins its replication phase is not t, as desired, but actually $t + \Delta_c$. In the case of application and removal of drugs to achieve synchronization, the delay for each individual cell amounts to the amount of time it takes for the cell to recover after the drug has been removed. This can be viewed as a stochastic process.

The magnitude of synchronization error present with a given technique must be accounted for when designing the length of the labeling periods to be used in the experiment. In general, the labeling periods should be made at least twice as long as the expected synchronization error from the chosen technique. We have used instantiations of this experimental design to investigate replication timing in chromosomes 21 and 22 [15], and the ENCODE regions[1] [16]. Below, we present a generalized framework for analyzing this type of experimental data.

Labeling periods begin with the start of S-phase, denoted as 0 hour. The length of each labeling period, L, provides a delicate balance between temporal resolution and resistance to synchronization error. Larger values of L decrease temporal resolution as all DNA replicating within a single time period appears on the same array. However, as smaller values of L approach the expected synchronization error, noise introduced across time period boundaries increases.

The first time period is 0-to-L hours, followed by L-to-2L hours, 2L-to-3L hours, and so on. The length of S-phase in the cell line under consideration guides

[1] The ENCODE regions [12] comprise approximately 1% of the human genome.

the choice of how many time periods to assay. Each time period is labeled and hybridized to its own array set, so the cost of the experiment increases linearly with the number of time periods. For this reason, it is sometimes desirable to use less time periods than would cover the full duration of S-phase since the amount of replicated DNA tends to fall dramatically near the end of the replication phase [15].

2.2 Time of Replication of 50% (TR50) of a Locus

The ultimate goal of a DNA replication timing study is to identify, with as much precision as possible, the actual time during S-phase that a given locus replicated. Under ideal circumstances, a given probe on the array set will display signal[2] in a single time period, with no signal across the other time periods. In this case, the locus in question has replicated sometime during the time period that displays signal. However, this scenario is rare due to synchronization error, tiling array artifacts such as cross-hybridization, and allelic variation in replication timing. The approach we take is to compute a value called the Time of Replication of 50% (TR50) for each probe in the array set. This TR50 value is a linear interpolation of the time at which the cumulative signal across all time periods for the probe passes the 50% point. We denote the signal for probe p in time period X-to-Y as $(\text{X-to-Y})_p$.

The steps to compute the TR50 value for probe p are as follows:

Normalization - Remove baseline signal present across all time periods

```
// Find the minimum signal value of all of the time periods
Minimum = min((0-to-L)_p, (L-to-2L)_p, ...);

// Subtract the minimum signal value from each of the time periods
(0-to-L)_p = (0-to-L)_p - Minimum;
(L-to-2L)_p = (L-to-2L)_p - Minimum;
...
```

Linear Interpolation - Calculate the TR50 value

```
Total = (0-to-L)_p + (L-to-2L)_p + ...; // Sum signal of all time periods

If (Total == 0) // Skip probes with 0 total signal
{ skip this probe; }

// Find the point at which 50% cumulative signal is passed
X = 0; // Start at the first time period
Cum = (0-to-L)_p; // Start with the first time period's signal
While (Cum < (Total / 2.0)) // Check for 50% of total signal
```

[2] Signal for a probe on the array is maintained as a general concept throughout the paper in order to be applicable to both arrays that have only perfect match (PM) probes and arrays that pair a mis-match (MM) probe with each PM probe. In the latter case any negative signals created by the MM probe having higher intensity than the PM probe are truncated to 0. No negative signals are allowed on the array.

```
{
        X = X + L; // Move to the next time period
        Cum = Cum + (X-to-X+L)_p; // Add next time period's signal
}

// Perform linear interpolation (X is the beginning of the time period
// where the cumulative signal surpassed 50% of the total signal)
TR50 = X + L * ((0.5 * Total - (Cum - (X-to-X+L)_p)) / (X-to-X+L)_p);
```

2.3 Temporal Specificity and Allelic Variation

The TR50 value provides an estimate of the time when the majority of replication occurs for a given locus in cases where all alleles at the locus replicate synchronously. This is called temporally specific replication (TSR). However, it has been well documented that different alleles at a given locus can replicate asynchronously [17–26]. This phenomenon, which we denote as temporally non-specific replication (TNSR), can produce a misleading result for the TR50 value. With TNSR, the TR50 value gives the average replication time over all alleles, which can produce a value at a time when no allele was being replicated. For this reason, it is important to identify and separate TNSR probes from TSR probes, which we do via our Temporal Specificity Algorithm.

Many normal cell lines are diploid in nature, having two copies of each chromosomal locus. However, HeLa cells, which we have used in some of our work, typically exhibit three copies of each chromosomal locus [16]. Tetraploidy, having four homologous sets of chromosomes, is common in plants and appears in some insects, amphibians, and reptiles [27]. We have generalized our Temporal Specificity Algorithm for application to cell lines that exhibit N copies of each chromosomal locus. Though there can be exceptions to the general ploidy in any given cell line, N should be set to the most prevalent occurrence of copy number in the cell line. For cell lines that exhibit more than one very common copy number, the larger value should be chosen for N. Having N larger than the actual copy number will perform more accurate classification than when the value of N is less than the actual copy number.

The steps of the Temporal Specificity Algorithm are as follows:

Normalization - This step is the same as in the TR50 calculation[3]

```
Total = (0-to-L)_p + (L-to-2L)_p + ...; // Sum signal of all time periods

If (Total == 0) // Skip probes with 0 total signal
{ skip this probe; }

// Find the maximum sum of all sets of two adjacent time periods
Maxsum = max((0-to-L)_p+(L-to-2L)_p, (L-to-2L)_p+(2L-to-3L)_p, ...);

// Find the maximum signal value of all of the time periods
```

[3] In practice the TR50 calculation and Temporal Specificity Algorithm are computed together, but they are presented separately here for clarity.

```
Maximum = max((0-to-L)_p, (L-to-2L)_p, ...);

// Find the maximum sum of all sets of two adjacent time periods that
// does not include the maximum signal value in either time period
Maxsumnot = 0;
X = L; // start X at the beginning of the 2nd time period
While ((X-to-X+L)_p exists)
{
        If (((X-L-to-X)_p < Maximum) and ((X-to-X+L)_p < Maximum))
        { // Neither time period includes the maximum signal
                Maxsumnot = max(Maxsumnot, (X-L-to-X)_p + (X-to-X+L)_p);
        }
}

If (Maxsum > (1 - 1/N) * Total) // Are all alleles replicating together?
{ classify probe as TSR; }
Else If (Maxsumnot >= (1/N) * Total) // Is at least one allele separate?
{ classify probe as TNSR; }
Else // Isolated signal is not strong enough to represent an allele.
{ classify probe as TSR; }
```

This classification scheme might seem arcane at first because it has been evolved over a number of attempts to classify the probes correctly. The final algorithm was arrived at after a thorough combinatorial analysis of the possible positions of replicating alleles with respect to time periods and their boundaries. We elucidate the reasoning behind each part of the algorithm in detail below.

Our original attempts to classify probes focused on the signal of each time period individually. However, due to the presence of synchronization error in the population, loci that replicate near the boundary of two adjacent time periods can contribute significant signal to both. This causes such loci to appear to undergo TNSR, even though the alleles may actually replicate together near the boundary. To address this issue we adopted the strategy of summing adjacent time periods. The sum of two adjacent time periods gives a view of the replication that occurs in either time period or on the boundary between them.

The first step of the classification algorithm is to determine if there is strong evidence that all alleles replicated together. The candidate set of adjacent time periods is selected by finding the maximum sum of signal for any set of two adjacent time periods. If this sum exceeds (1 - $1/N$) of the total signal, we classify the probe as TSR. This implies that less than $1/N$ of the total signal is contained in the other time periods. With N alleles at the locus, each individual allele is expected to contribute $1/N$ of the total signal across all time periods.

The second step is only performed if the first step failed to yield strong evidence for all alleles replicating together. In the second step, we look for evidence that at least one allele is replicating apart from a time period with the maximum signal value. We already know (since the first step failed) that at least $1/N$ of the signal is isolated from the two adjacent time periods that contained the maximum sum. The objective here is to determine if the signal that is isolated from the maximum signal value is concentrated enough to represent at least one allele.

To test this, we find the maximum sum of two adjacent time periods that does not include a time period with the maximum signal. Note that the maximum signal does not have to appear in one of the two time periods that contributed to the sum in the first step. Hence this test is subtly unrelated to the first. If this sum is at least $1/N$ of the total signal, then there is evidence for at least one allele replicating apart from the majority of signal. Namely, the evidence is for an allele to be replicating in one of the two time periods that produced this sum or on the boundary between them. In this case the probe is classified as TNSR.

Lastly, if the second test fails to yield evidence for an allele replicating apart from the majority of signal, then we consider the remaining scattered signal to be due to array artifacts and classify the probe as TSR.

2.4 Segregation of Temporally Specific and Temporally Non-specific Area

The probe data computed by the TR50 and Temporal Specificity Algorithm is very noisy due to cross-hybridization and other microarray artifacts. To address this, we take advantage of the fact that the replication mechanism provides us with spatial locality for replicated segments. As a replication fork proceeds, it causes adjacent loci on the chromosomes to replicate at similar times until the fork stalls or meets DNA that has already been replicated.

We pass a sliding window over each chromosomal sequence in order to generate broad regions of replication. The first task is to segregate TSR regions from TNSR regions. The size of the window used should be chosen to match the expected size of a replication fragment. Replication fragments will vary in size based on the length of time that the responsible replication fork operated in the given time period, so this parameter should be chosen based on what the typical expected size for a replication fragment is. A variety of strategies can be used to estimate this [16], but this is also a tunable parameter. Larger window sizes will attenuate noise in the data more; however, if the window size becomes larger than the replication fragments, then multiple of them can get merged. Smaller window sizes will suffer more from the noise inherent in the microarray data. This parameter can be increased until noise is attenuated at an acceptable level.

We also require a minimum probe density to generate intervals. If too few probes fall into a window, then such an area will not be classified for lack of tiling data. This is another tunable parameter. Setting a higher density decreases coverage of the generated intervals but increases confidence in the classification.

The sliding window is placed at the beginning of each chromosome to start segregation of the regions. As the window moves from probe to probe, the minimum probe density is tested for and when this density is exceeded a TSR interval or TNSR interval is begun based on whether there are more TSR or TNSR probes in the window. In the event of a tie, the window begins when the next probe is reached (which will break the tie). The current interval is ended when the probe density drops below the minimum level or when the TSR to TNSR probe ratio changes direction (in which case a new interval is started).

2.5 TR50 Smoothing

The TR50 values provide a noisy view of the replication timing pattern. In order to get a more continuous estimate of the replication profile, a locally weighted least squares (lowess) smoothing [28] is performed on the set of TSR probes. The smoother is set to consider all probes within the same window size used for the segregation above. Only TR50 values for the TSR probes are used because TR50 values for TNSR probes are unreliable as discussed above.

3 Results

We used the methods described in Section 2 to analyze two technical replicates and one biological replicate of the HeLa cell line (human) using Affymetrix ENCODE tiling arrays [16]. In this section, we report results pertinent to the methods themselves. Throughout this section, Replicate 1 (Rep1) and Replicate 2 (Rep2) refer to two technical replicates (the same biological sample hybridized to two sets of arrays) and Replicate 3 (Rep3) refers to the biological replicate.

Computation on the individual probes (Sections 2.2 and 2.3) performs a normalization step for probes that have no time period with 0 signal. The percentage of probes normalized during this process is shown in Figure 1.

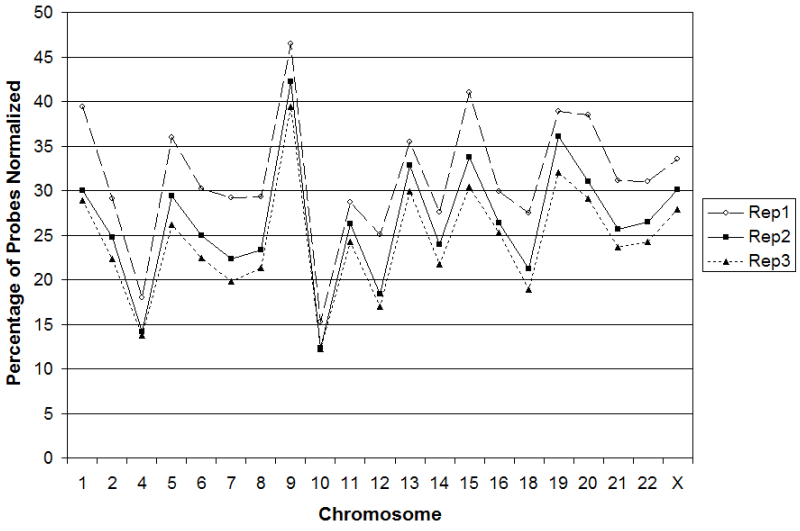

Fig. 1. Percentage of probes normalized for each replicate.

This graph plots the percentage of probes on the array where every time period had positive signal. Such probes were normalized by subtracting the minimum signal from all time periods, in order to remove baseline signal for the

given probe. The three replicates all exhibit the same trend for each chromosome, indicating that the process is indeed removing signal from array artifacts, instead of removing variations in signal between different replicates.

Figure 2 shows the percentage of probes classified as TSR for each replicate on each chromosome. In this case, Rep2 and Rep3 show the same general trend for each chromosome, while Rep1 has a more varied pattern. This underscores the importance of processing the probes through windows in the next steps since two technical replicates (Rep1 and Rep2) show varied results at the probe level.

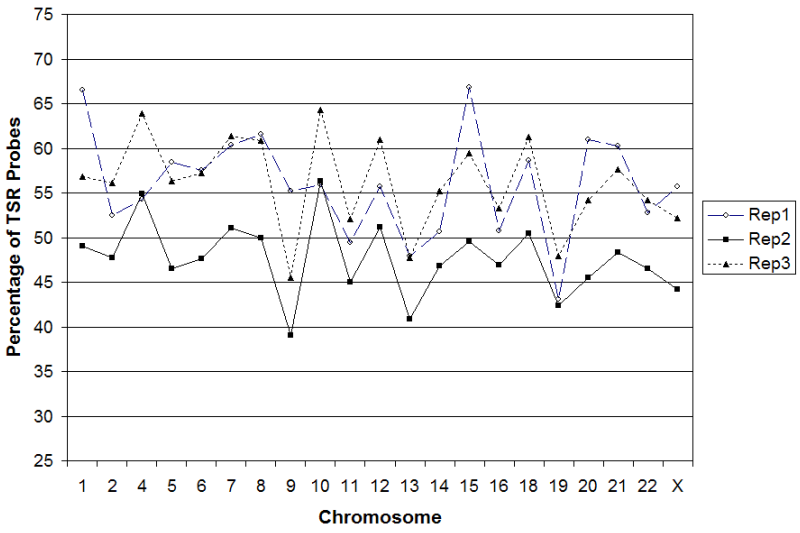

Fig. 2. Percentage of probes classified as temporally specific (TSR) per replicate.

The segregation and TR50 smoothing was done in a window of size 60,000 base pairs. This was chosen based on a profiling calculation of the expected size of replication fragments in the experiment under consideration [16].

Figure 3 shows the TR50 data for a region on chromosome 21 with the smoothed TR50 curve overlaid. Segregation of TSR regions from TNSR regions was performed with a minimum probe density of 25%. This required at least 600 probes to fall inside of the sliding 60,000 base pair window in order to generate intervals (each probe tiles 25 base pairs). The segregation intervals are shown in Figure 3 above the TR50 data.

The TR50 data at the probe level is quite noisy, but a pattern can be seen in the data where tightly grouped probes produce darker areas in the graph. The smoothed TR50 curve follows these trends closely. There is a late replicating domain (broad peak) in the graph which is surrounded by early replicating DNA. These domains have proven to be quite interesting, as the broad peak of late replication is associated with low gene density, low transcriptional activity, and a high level of repressive histone marks [2]. Further, the early replicating

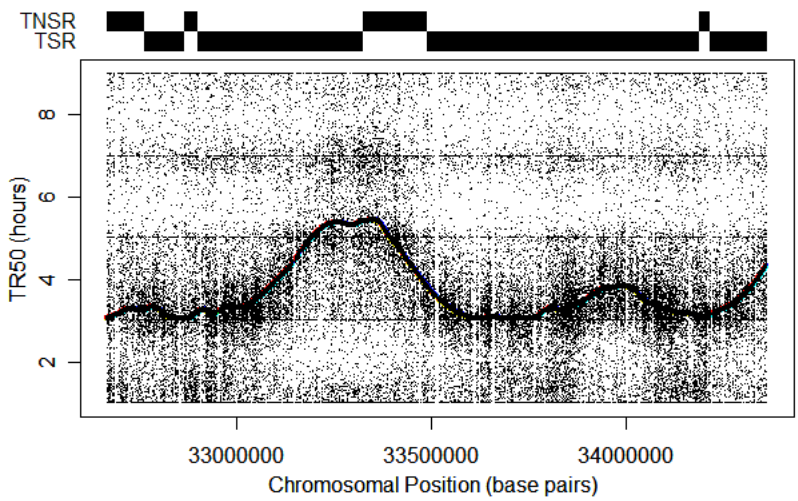

Fig. 3. Replication profile over a region of chromosome 21. Each dot in the graph corresponds to the TR50 value of a single probe. The smoothed TR50 curve is overlaid. The segregation of TSR and TNSR regions is shown above the graph.

domains surrounding this broad peak are associated with high gene density, high transcriptional activity, and high levels of activating histone marks. The troughs in the replication curve allude to possible sites of replication origin, while the peaks could be indicative of sites of fork termination. Notice from the segregation at the top of the figure that there is a large section of TNSR on the right half of the broad peak of late replication. This area is associated with high levels of both repressive and activating histone marks. We have found that on average roughly 20% of the ENCODE regions undergo TNSR [16].

All of the algorithms and techniques that we have presented to generate the replication timing profile run in linear time with respect to the size of the tiling array set used. This efficiency is achieved by using incrementally updating sliding windows. The linear runtime will allow for the general methods presented here to be utilized for whole genome analysis with moderate computational resource requirements. The replication timing profile constructed (displayed in Figure 3) produces a relatively continuous view of the replication timing in addition to identifying TNSR regions where inter-allelic variation of replication timing occurs.

4 Conclusion

We have presented a generalized framework and algorithms for analyzing a common type of DNA replication timing assay using tiling arrays. We have also discussed techniques for choosing parameters for analysis of a given replication timing array set. This approach overcomes the noise present in such tiling array

data to reconstruct a relatively continuous replication timing profile and identify areas of temporally non-specific replication. The algorithms developed have linear time complexity in the size of the tiling array set so that the approach can be used for whole genome analysis in a variety of organisms requiring only a moderate expenditure of computational resources. Lastly, we have discussed an example of the framework being applied to a set of DNA replication timing data over a small portion of the human genome. In the future, we intend to utilize this approach to analyze replication timing over the full human genome.

5 Acknowledgements and Data Availability

This work was supported by funding from the National Human Genome Research Institute (NHGRI) of the National Institutes of Health (NIH). Raw microarray data for the timing experiments discussed is available in ArrayExpress with accession number E-MEXP-708. Segregation and TR50 data are available as tracks in the UCSC genome browser:

http://genome.ucsc.edu/cgi-bin/hgTrackUi?g=encodeUvaDnaRepSeg
http://genome.ucsc.edu/cgi-bin/hgTrackUi?g=encodeUvaDnaRepTr50

References

1. Bell, S.P., Dutta, A.: DNA Replication in Eukaryotic Cells. Annu. Rev. Biochem. 2002;71:333–74.
2. ENCODE Project Consortium: Identification and Analysis of Functional Elements in 1447(7146):799–816
3. Raghuraman, M.K., Winzeler, E.A., Collingwood, D., Hunt, S., Wodicka, L., Conway, A., Lockhart, D.J., Davis, R.W., Brewer, B.J., Fangman, W.L.: Replication Dynamics of the Yeast Genome. Science. 2001 Oct 5;294(5540):115–21.
4. Lucas, I.A., Raghuraman, M.K.: The Dynamics of Chromosome Replication in Yeast. Curr. Top. Dev. Biol. 2003;55:1–73.
5. Kim, S.M., Huberman, J.A.: Regulation of Replication Timing in Fission Yeast. EMBO J. 20001 Nov 1;20(21):6115–26.
6. Schubeler, D. Scalzo, D., Kooperberg, C., van Steensel, B., Delrow, J., Groudine, M.: Genome-wide DNA Replication Profile for Drosophila Melanogaster: a Link Between Transcription and Replication Timing. Nat. Genet. 2002 Nov;32(3):438–42.
7. MacAlpine, D.M., Rodriguez, H.K., Bell, S.P.: Coordination of Replication and Transcription Along a Drosophila Chromosome. Genes. Dev. 2004 Dec 15; 18(24):3094–105.
8. McCune, H.J., Donaldson, A.D.: DNA Replication: Telling Time with Microarrays. Genome Biol. 2003;4(2):204. Epub 2003 Jan 30.
9. MacAlpine, D.M., Bell, S.P.: A Genomic View of Eukaryotic DNA Replication. Chromosome Res. 2005;13:309–26.
10. Donaldson, A.D.: Shaping Time: Chromatin Structure and the DNA Replication Programme. Trends Genet. 2005 Aug;21(8):444–9.
11. Woodfine, K., Fiegler, H., Beare, D.M., Collins, J.E., McCann, O.T., Young, B.D., Debernardi, S., Mott, R. Dunham, I., Carter, N.P.: Replication Timing of the Human Genome. Hum. Mol. Genet. 2004 Jan 15;13(2):191–202.

12. ENCODE Project Consortium: The ENCODE (ENCyclopedia Of DNA Elements) Project. Science. 2004 Oct 22;306(5696):636–40.
13. White, E.J., Emanuelsson, O., Scalzo, D., Royce, T., Kosak, S., Oakeley, E.J., Weissman, S., Gerstein, M., Groudine, M., Snyder, M., Schubeler, D.: DNA Replication-timing Analysis of Human Chromosome 22 at High Resolution and Different Developmental States. Proc. natl. Acad. Sci. U.S.A. 2004 Dec 21;101(51):17771–6.
14. Woodfine, K., Beare, D.M., Ichimura, K., Debernardi, S., Mungall, A.J., Fiegler, H., Collins, V.P., Carter, N.P., Dunham, I.: Replication Timing of Human Chromosome 6. Cell Cycle. 2005 Jan;4(1):172–6.
15. Jeon, Y., Bekiranov, S., Karnani, N., Kapranov, P., Ghosh, S., MacAlpine, D., Lee, C., Hwang, D.S., Gingeras, T.R., Dutta, A.: Temporal Profile of Replication of Human Chromosomes. Proc. Natl. Acad. Sci. U.S.A. 2005 May 3;102(18):6419–24.
16. Karnani, N., Taylor, C., Malhotra, A., Dutta, A.: Pan-S Replication Patterns and Chromosomal Domains Defined by Genome-tiling Arrays of ENCODE Genomic Areas. Genome Res. 2007 Jun;17(6):865–76.
17. Kitsberg, D., Selig, S., Brandeis, M., Simon, I., Keshet, I., Driscoll, D.J., Nicholls, R.D., Cedar, H.: Allele-specific Replication Timing of Imprinted Gene Regions. Nature. 1993 Jul 29;364(6436):459–63.
18. Knoll, J.H., Cheng, S.D., Lalande, M.: Allele Specificity of DNA Replication Timing in the Angelman/Prader-Willi Syndrome Imprinted Chromosomal Region. Nat. Genet. 1994 Jan;6(1):41–6.
19. Boggs, B.A., Chinault, A.C.: Analysis of Replication Timing Properties of Human X-chromosomal Loci by Fluorescence in Situ Hybridization. Proc. Natl. Acad. Sci. U.S.A. 1994 Jun 21;91(13):6083–7.
20. Carothers, A.D., Bickmore, W.A.: Models of DNA Replication Timing in Interphase Nuclei: An Exercise in Inferring Process from State. Biometrics. 1995 Jun;51(2):750–5.
21. Bickmore, W.A., Carothers, A.D.: Factors Affecting the Timing and Imprinting of Replication on a Mammalian Chromosome. J. Cell. Sci. 1995 Aug;108 (Pt 8):2801–9.
22. Kawame, H., Gartler, S.M., Hansen, R.S.: Allele-specific Replication Timing in Imprinted Domains: Absence of Asynchrony at Several Loci. Hum. Mol. Genet. 1995 Dec;4(12):2287–93.
23. Greally, J.M., Starr, D.J., Hwang, S., Song, L., Jaarola, M., Zemel, S.: The Mouse H19 Locus Mediates a Transition Between Imprinted and Non-imprinted DNA Replication Patterns. Hum. Mol. Genet. 1998 Jan;7(1):91–5.
24. Simon, I., Tenzen, T., Reubinoff, B.E., Hillman, D., McCarrey, J.R., Cedar, H.: Asynchronous Replication of Imprinted Genes is Established in the Gametes and Maintained During Development. Nature. 1999 Oct 28;401(6756):929–32.
25. Mostoslavsky, R., Singh, N., Tenzen, T., Goldmit, M., Gabay, C., Elizur, S., Qi, P., Reubinoff, B.E., Chess, A., Cedar, H., Bergman, Y.: Asynchronous Replication and Allelic Exclusion in the Immune System. Nature. 2001 Nov 8;414(6860):221–5.
26. Kagotani, K., Takebayashi, S., Kohda, A., Taguchi, H., Paulsen, M., Walter, J., Reik, W., Okumura, K.: Replication Timing Properties within the Mouse Distal Chromosome 7 Imprinting Cluster. Biosci. Biotechnol. Biochem. 2002 May;66(5):1046–51.
27. Griffiths, A.J., Wessler, S.R., Lewontin, R.C., Carroll, S.B.: Introduction to Genetic Analysis. W.H. Freeman. 2007 Feb 16;0-7167-6887-9.
28. Jacoby, W.G.: Statistical Graphics for Univariate and Bivariate Data. Sage Publications Inc. 1997 Feb 24;0-7619-0083-7.