

A complete protocol for genome-scale analysis of replication timing in mammalian cells

Tyrone Ryba (✉ tryba@bio.fsu.edu)

Florida State University

Dana Battaglia

Florida State University

Benjamin D. Pope

Florida State University

Ichiro Hiratani

National Institute of Genetics

David M. Gilbert

Florida State University

Method Article

Keywords: Replication timing, stem cells, ESCs, human, mouse, development

Posted Date: September 2nd, 2010

DOI: <https://doi.org/10.1038/nprot.2010.151>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Analyses of replication timing in various cell types have yielded insights into genome organization and repackaging events during development, suggesting an important role for the timing program itself or 3D genome organization in regulating developmental gene expression¹⁻³. Although the mechanisms that specify the timing and placement of origin firing in higher eukaryotes remain a mystery, all eukaryotes have a defined replication timing program that is largely conserved between closely related species⁴, including human and mouse^{3,5}. Here, we describe details of the methodology used to analyze replication timing genome-wide in mammalian cells. In this protocol, asynchronously cycling cells are pulse labeled with the nucleotide analog 5-bromo-2-deoxyuridine (BrdU) and then sorted into S-phase fractions based on DNA content using flow cytometry. BrdU-labeled DNA from each fraction is immunoprecipitated, amplified, differentially labeled, and co-hybridized to a whole-genome CGH microarray. To address the analysis bottleneck that often follows genomic experiments, the second half of the protocol covers methods to analyze array data produced using the R framework for statistical computing⁶⁻⁸, including normalization, scaling, and data quality measures, loess (local polynomial) smoothing of replication timing values, segmentation of data into domains, and assignment of timing values to gene promoters. After these preprocessing steps, we cover various common downstream analyses, such as k-means and hierarchical clustering, and methods to relate changes in the replication program to gene expression and other genetic and epigenetic datasets.

Additional details and descriptions of the method can be found at www.ReplicationDomain.org/Protocol.

Reagents

- 1mg/mL BrdU (5-bromo-2'-deoxyuridine) (Sigma Aldrich, B5002) (All solutions are in ddH₂O unless otherwise indicated)
- Cell culture medium appropriate for cell type
- 1 X PBS (10 X Phosphate Buffered Saline, 137 mM NaCl, 2.7mMKCl, 10mM Na₂HPO₄, 2mM KH₂PO₄, diluted with dH₂O)
- 0.2X Trypsin-EDTA (diluted with PBS from a 1X solution, Mediatech 25-053-CI) or Accutase (Innovative Cell Technologies AT104)
- Fetal Bovine Serum (FBS)
- DAPI
- 1 mg/mL Propidium Iodide (PI, Sigma P4179-100MG dissolved in ddH₂O and filtered)
- 10 mg/mL RNase A
- SDS-PK buffer (dH₂O with 50mM Tris-HCl, pH8.0/ 10mM EDTA/ 1M NaCl/ 0.5%SDS)
- 20mg/mL Proteinase K (dissolved in ddH₂O)
- 20mg/mL Glycogen (Fermentas RO561)
- Tris-saturated Phenol-chloroform
- Chloroform

- Isopropanol
- 70% ethanol
- 100% ethanol
- 1 X TE
- 10X IP Buffer \(\text{dH}_2\text{O}\) with 0.1M sodium phosphate, pH 7.0 / 1.4M NaCl / 0.5% Triton X-100)
- 1 X IP Buffer \(\text{10 X IP buffer diluted in ddH}_2\text{O}\)
- Anti-BrdU antibody \(\text{BD Biosciences Pharmigen, Cat. \#555627; 0.5mg/ml diluted to 12.5 } \mu\text{g/ml in PBS}\)
- 10 M ammonium acetate
- Rabbit anti-mouse IgG \(\text{Sigma, Cat \#M-7023}\)
- Anti-Mouse IgG-AlexaFluor488 \(\text{Invitrogen/Molecular Probes, Cat \#A-11029}\)
- Digestion Buffer \(\text{dH}_2\text{O}\) with 50mM Tris-HCl, pH 8.0 / 10mM EDTA / 0.5% SDS)
- Standard PCR and electrophoresis equipment and reagents
- PCR primers for BrdU-IP quality verification \(\text{steps 45-49}\)
- 2 mg/mL glycogen
- Anti-BrdU antibody \(\text{BD Biosciences Pharmigen, Cat. \#555627, undiluted}\)
- 0.1M HCl / 0.5% Triton X-100 in ddH₂O
- 0.1M Sodium Tetraborate, \(\text{Na}_2\text{B}_4\text{O}_7 \cdot 10\text{H}_2\text{O}\), pH 8.5
- 0.5% Tween20 / 1% BSA in PBS
- 0.1% Triton X-100 in PBS
- 0.5% Triton X-100 in PBS
- GenomePlex Complete Whole Genome Amplification Kit
- GenomePlex WGA Reamplification Kit
- QIAquick PCR Purification Kit
- NimbleGen Dual-Color DNA Labeling Kit \(\text{cat. no. 05223547001}\)
- NimbleGen Hybridization Kit \(\text{cat. no. 05583683001}\)
- NimbleGen Wash Buffer Kit \(\text{cat. no. 05584507001}\)

Reagent setup

- SDS-PK buffer : To make 50mL, combine 34mL ddH₂O, 2.5mL 1M Tris-HCl pH8.0, 1mL 0.5 M EDTA, 10mL 5 M NaCl and 2.5mL 10% SDS. Warm to 56°C before use to completely dissolve SDS.
- 10X IP Buffer: To make 50mL, combine 28.5mL ddH₂O, 5mL 1M Sodium Phosphate pH 7.0, 14 mL 5M NaCl, and 2.5mL 10% Triton X-100.
- Digestion Buffer: To make 50mL, combine 44mL ddH₂O, 2.5mL 1M Tris-HCl pH8.0, 1mL 0.5M EDTA, 2.5mL 10% SDS.

Equipment

- Nylon mesh 37 micron \(\text{Small Parts, CMN-0040-D}\)
- 5mL round bottom polystyrene tube \(\text{Falcon 352054}\)

- 15mL round bottom tube (Falcon 2059)
- FACS Aria cell sorter (or comparable sorter)
- Hemacytometer
- Vortex Genie
- Sonicator
- Appropriate NimbleGen arrays, mixers, hybridization system, and microarray scanner

Procedure

Part I: Generation of raw replication timing datasets

PROTOCOL 1. BrdU labeling and staining of cells for FACS

This protocol can be performed using PI staining and ethanol fixation prior to sorting (**option A**), or, for cells that break or clump in ethanol, by collecting cells following pulse labeling with BrdU and treating with a DAPI staining solution containing Triton X-100 to permeabilize cell membranes (**option B**). The drawback of option B is that cells need to be sorted immediately following staining.

A. Labeling and PI staining of cells for FACS following ethanol fixation - TIMING 3.5 h

1. For adherent cells, remove cell culture medium and replenish with fresh medium. Add BrdU at a final concentration of 50 μ M. For suspension cells, add BrdU to the cell culture medium at a final concentration of 50 μ M.
2. Incubate cells for two hours in a carbon dioxide incubator at 37°C, 5% CO₂.
3. For adherent cells, rinse gently with ice-cold PBS twice. For suspension cells, collect cells in a 15mL tube and proceed directly to step 6.
4. Detach adherent cells using 0.2X Trypsin-EDTA for 2-3 minutes or Accutase for 3-6 minutes.
CRITICAL STEP Incubate cells at 37 °C with the enzyme treatment and/or use gentle trituration if necessary to achieve a single cell suspension, as this is necessary for accurate FACS sorting.
5. Add 5mL of cell culture medium (containing FBS if trypsin has been used) to the cell culture dish or flask, pipette gently, and transfer contents to a 15mL round bottom tube.
6. Count the number of cells collected using a hemacytometer.
7. Centrifuge at approximately 200 *g* for 5 minutes.
8. Aspirate supernatant carefully and resuspend cells in 2.5 mL of ice-cold PBS containing 1% FBS.
9. Add 7.5 mL of ice-cold 100% ethanol dropwise while gently vortexing.
CRITICAL STEP Note that gentle vortexing is required.
10. Seal the cap of the 15 mL tube with parafilm and mix gently but thoroughly.
11. **PAUSE POINT** Proceed to the next step to begin PI staining or stop here and store tubes in the dark at -20°C indefinitely.
12. Resuspend the BrdU-labeled, ethanol fixed cells by tapping and inverting the tube.
13. Transfer 4 × 10⁶ - 8 × 10⁶ cells to a 5 mL polystyrene round bottom tube.

14. Centrifuge at approximately 200 *g* for 5 minutes.
15. Decant supernatant carefully
16. Add 2 mL of PBS with 1% FBS and mix well by tapping the tube.
17. Centrifuge at approximately 200 *g* for 5 minutes.
18. Decant supernatant carefully.
19. Resuspend cell pellet in PBS with 1% FBS to achieve a solution of 3×10^6 cells/mL.
20. Add 1mg/mL propidium iodide to a final concentration of 50 μg mL.
21. Add 10 mg/mL RNase A to a final concentration of 250 μg /mL.
22. Tap the tube to mix and incubate for 20 to 30 minutes at room temperature in the dark.
23. Filter cells by pipeting them through 37 micron nylon mesh into a 5 mL polystyrene round bottom tube.
24. Keep samples on ice in the dark and proceed directly to FACS sorting.

B. BrdU labeling and DAPI staining of cells for FACS - TIMING 3 h

1. Remove cell culture medium from growing cells and replace with cell culture medium containing BrdU at a final concentration of 50uM.
 2. Incubate cells for two hours in a carbon dioxide incubator at 37°C, 5% CO₂.
 3. Rinse adherent cells gently with ice-cold PBS twice.
 4. Detach adherent cells using 0.2X Trypsin-EDTA for 2-3 minutes or Accutase for 3-6 minutes.
- CRITICAL STEP** Incubate cells at 37 °C with the enzyme treatment and/or use gentle trituration if necessary to achieve a single cell suspension, as this is necessary for accurate FACS sorting
5. Add 5mL of cell culture medium \ (containing FBS if trypsin has been used) to the cell culture dish or flask, pipette gently, and transfer contents to a 15mL tube. Count the number of cells collected using a hemacytometer.
 6. Centrifuge at approximately 200 *g* for 5 minutes.
 7. Aspirate the supernatant carefully.
 8. Add 5mL of ice-cold PBS and pipette gently but thoroughly.
 9. Centrifuge at approximately 200 *g* for 5 minutes.
 10. Aspirate supernatant carefully.
 11. Resuspend the cell pellet in DAPI staining solution \ (1 X PBS with 0.1% Triton X-100 and 2 μg /mL DAPI) to achieve a solution of $5 \times 10^6 - 10 \times 10^6$ cells/mL.
 12. Filter cells by pipeting them through 37 micron nylon mesh into a 5 mL polystyrene round bottom tube.
 13. Keep samples on ice in the dark and proceed directly to FACS sorting.

PROTOCOL 2. FACS sorting fractions of S-phase - TIMING 1 h

This protocol describes the use of flow cytometry to sort cells based on DNA content. Electrostatic sorting is performed as described here on a BD \ (Becton-Dickinson) FACS Aria flow cytometer. During FACS analysis forward and side scatter analyses should be used to select an appropriate population of cells

free of doublets or cell debris, both of which can hinder accurate sorting of desired populations. Lasers used in this protocol include 488 Blue to detect PI or 407 Violet to detect DAPI in cells that have been stained for DNA content. Two separate fractions of S phase, early and late, are chosen to be collected, but more can be collected if desired^{9,10}.

1. Run sample on FACS Aria cell sorter \ (Any comparable cell sorter can be used)

CRITICAL STEP It is very important to keep all samples chilled on ice or at 4°C during FACS analysis to avoid cell cycle progression in the absence of BrdU. Protect samples from light.

2. Use forward and side scatter information to select the desired population of cells to be included in the sort.

3. Create a histogram that plots cell count on the y-axis and DNA content \ (fluorochrome intensity) on the x-axis. See **Fig.1A**

4. Select two distinct S-phase populations to be sorted into separate fractions as indicated in Fig.1.

5. Store cell immediately on ice in the dark until all samples have been sorted. Cells are sorted into fresh 5ml round bottom tubes and kept at 4°C during the sort.

6. Centrifuge at 400g for 10 minutes at 4°C. Alternatively, if fewer than 150,000 cells have been collected for each fraction, proceed directly to step 8.

7. Decant supernatant gently, only once.

CRITICAL STEP The cell pellet can easily become loose. Some residual sheath fluid can be left in the tube.

8. Add 1 mL of SDS-PK buffer containing 0.2mg/mL Proteinase K and 0.05mg/mL glycogen for every 100,000 cells collected and mix vigorously by tapping the tube.

9. Incubate samples in a 56°C water bath for 2 hours.

10. Mix sample thoroughly and aliquot 200µl, equivalent to approximately 20,000 cells, per 1.5 mL tube.

11. Stop here and store samples at -20°C in the dark or proceed directly to step 12.

PAUSE POINT Samples can be stored for at least 6 months before use.

12. Add 200µl SDS-PK buffer with 0.05mg/mL glycogen to each aliquot and proceed directly to BrdU immunoprecipitation.

PROTOCOL 3. BrdU immunoprecipitation - TIMING 2 d

This protocol begins with cells that have been pulse labeled with BrdU and sorted into different cell cycle stages by FACS. Here DNA from these cells is sonicated into fragments ranging from 250bp to 2kb and then immunoprecipitated using an anti-BrdU antibody. Immunoprecipitated DNA is then purified and ready for use in further steps of replication timing analysis for a time period of up to one month. If samples have been stored at -20°C prior to beginning the immunoprecipitation, thaw samples in a 56°C water bath and add 200µl of SDS-PK Buffer pre-warmed to 56°C with 0.05mg/mL glycogen to each sample prior to performing the phenol-chloroform extraction in step 13.

13. Extract once with phenol-chloroform, collecting the upper phase in a 1.5mL tube.

14. Extract once with chloroform, collecting the upper phase in a 1.5mL tube.

15. Add 1 volume of isopropanol and mix well.
16. Store at -20°C for 20 minutes.
17. Centrifuge at 16.1 *g* for 30 minutes at 4°C.
18. Discard the supernatant and add 750µl 70% ethanol to the pellet
19. Centrifuge at 16.1 *g* for 5 minutes at 4°C, then remove all ethanol and let the pellet dry.
20. Resuspend the pellet in 500µl 1x TE.

PAUSE POINT Stop here and store overnight at 4°C or proceed directly to step 21.

21. Sonicate DNA to an average size of ~0.7-0.8 kb.
22. Incubate sample at 95°C for 5 minutes to heat denature.
23. Cool sample on ice for 2 minutes.
24. Add 60µl of 10X IP Buffer to a clean 1.5mL tube.
25. Add the denatured DNA to the tube containing 60µl 10X IP Buffer.
26. Add 40µl of 12.5 µg/ml anti-BrdU antibody.
27. Incubate for 20 minutes at room temperature with constant rocking.

CRITICAL STEP Foil tubes to keep samples in the dark.

28. Add 20µg of rabbit anti-mouse IgG.

CRITICAL STEP Foil tubes to keep samples in the dark.

29. Incubate for 20 minutes at room temperature with constant rocking.
30. Centrifuge at 16.1 *g* for 5 minutes at 4°C
31. Remove supernatant completely.

CRITICAL STEP If pellet becomes loose, then centrifuge sample again in order to completely remove the supernatant without disturbing the pellet. Several centrifugations may be necessary to completely remove supernatant.

32. Add 750µl of 1X IP Buffer that has been chilled on ice.

33. Centrifuge at 16.1 *g* for 5 minutes at 4°C.

34. **CRITICAL STEP** Remove supernatant completely, as in step 31.

35. Resuspend the pellet in 200µl digestion buffer with freshly added 0.25mg/mL proteinase K.

PAUSE POINT Incubate samples overnight at 37°C.

36. Add 100µl of fresh digestion buffer with freshly added 0.25mg/mL proteinase K.

37. Incubate samples for 60 minutes at 56°C.

38. Extract once with phenol-chloroform, collecting the upper phase in a 1.5mL tube.

39. Extract once with chloroform, collecting the upper phase in a 1.5mL tube.

40. Add 0.625µl of 20mg/mL glycogen, 100µl of 10 M ammonium acetate, and 750µl of 100% ethanol and mix well.

41. Store at -20°C for 20 minutes.

42. Centrifuge at 16.1 *g* for 30 minutes at 4°C.

43. Remove supernatant, rinse pellet with 70% ethanol and dry.

44. Resuspend the pellet in 80µl 1X TE.

PAUSE POINT Store DNA at 4°C for up to one month.

Quality control check of early and late S-phase DNA

To ensure quality, BrdU-IPs are screened by PCR amplification using primers specific to DNA markers of known relative replication time (i.e. early or late). As PCR results can vary between aliquots of the same sample and replication timing changes during development^{2,3}, consistency across multiple samples from the same cell type is the most reliable way to assess the quality of IP samples. Among the mouse sequences listed below, mitochondrial DNA replicates throughout the cell cycle¹¹ and is equally represented in early and late S-phase fractions. Alpha-globin, Pou5f1 and Mmp15 are generally early replicating markers, whereas beta-globin, Zfp42, Dppa2, Ptn, Mash1, and Akt3 are generally late replicating markers. Zfp42 and Dppa2 are early replicating in ESCs whereas they are late replicating in all somatic cell types examined to date. Among the human sequences listed below, mitochondrial DNA is equally represented in early and late S-phase fractions, while alpha-globin, MMP15 and BMP1 are generally early replicating markers. PTGS2, NETO1, SLITRK6, ZFP42 and DPPA2 are generally late replicating. Markers amplified, sizes and primer sequences are as follows:

Mouse test regions

Intergenic mitochondrial DNA, 346 bp

Forward 5'-GACATCTGGTTCTTACTTCA-3'

Reverse 5'-GTTTTTGGGGTTTGGCATT-3'

Alpha-globin, 439 bp

Forward 5'-AAGGGGAGCAGAGGCATCA-3'

Reverse 5'-AGGGCTTGGGAGGGACTG-3'

Beta-globin, 369 bp

Forward 5'-CAGTAAGCCACAGATCCTATTG-3'

Reverse 5'-CCCATAGTGACTATTGACTGTG-3'

Pou5f1, 194 bp

Forward 5'-CCCTCCCTAAGTGCCAGTTTCT-3'

Reverse 5'-GTAATCGCCCTCAGCAGTGTCT-3'

Mmp15, 360 bp

Forward 5'-AACAGAAGGCCTGCCTTGAC-3'

Reverse 5'-TGCATAGCACGACAGCATTG-3'

Zfp42, 211 bp

Forward 5'-TGAGATTAGCCCCGAGACTGAG-3'

Reverse 5'-CGTCCCCTTTGTCATGTACTION-3'

Dppa2, 199 bp

Forward 5'-CCACAGGAAGACAGGAAGCAGT-3'

Reverse 5'-AGCCAGACAGGAGCCCTAGAGT-3'

Ptn, 230 bp

Forward 5'-CTGGAATGAGTTACTGACGGGG-3'

Reverse 5'-CTGGCCCCACTGTGTAATAAGC-3'

Mash1, 182 bp

Forward 5'-GAAGATGAGCAAGGTGGAGACG-3'

Reverse 5'-AGTAGGACGAGACCGGAGAACC-3'

Akt3, 173 bp

Forward 5'-GAAGTGTGGGTTGAACCTCTGG-3'

Reverse 5'-GCACCCTCTCCACTGTTCTGAT-3'

Human test regions

Intergenic mitochondrial DNA, 168 bp

Forward 5'-CCTAGGAATCACCTCCCATTCC-3'

Reverse 5'-GTGTTTAAGGGTGGCTAGGG-3'

Alpha-globin, 257 bp

Forward 5'-GACCCTCTTCTCTGCACAGCTC-3'

Reverse 5'-GCTACCGAGGCTCCAGCTTAAC-3'

Beta-globin, 241 bp

Forward 5'-CCTGAGGAGAAGTCTGCCGTTA-3'

Reverse 5'-GAACCTCTGGGTCCAAGGGTAG-3'

MMP15, 249 bp

Forward 5'-CAGGCCTCTGGTCTCTGTCATT-3'

Reverse 5'-AGAGCTGAGAAACCACCACCAG-3'

BMP1, 177 bp

Forward 5'-GATGAAGCCTCGACCCCTAGAT-3'

Reverse 5'-ACCCGTCAGAGACGAACTTGAG-3'

hPTGS2, 230 bp

Forward 5'-GTTCTAGGCTGGTGTCCCATTG-3'

Reverse 5'-CTTTCTGTACTGCGGGTGGAAC-3'

hNETO1, 286 bp

Forward 5'-GGAGGTGGAATGCTAGGGACTT-3'

Reverse 5'-GCTGAGTGTGGCCTTAAGAGGA-3'

hSLITRK6, 281 bp

Forward 5'-GGAGAACATGCCTCCACAGTCT-3'

Reverse 5'-GTCCTGGAAGTTGAGTGGATGG-3'

hZFP42, 233 bp

Forward 5'-CTTGTGGGGACACCCAGATAAG-3'

Reverse 5'-AACCACTCCAGGCAGTAGTGA-3'

hDPPA2, 168bp

Forward 5'-AGGTGGACAGCGAAGACAGAAC-3'

Reverse 5'-GGCCATCAGCAGTGCCTAAAC-3'

PROTOCOL 4. PCR method for quality control of BrdU-immunoprecipitation - TIMING 5 h

45. Prepare enough PCR master-mix to screen all IP samples with each primer set. Each reaction requires:

9.63 uL nuclease free water

1.25 uL 10X Buffer

0.25 uL 10mM dNTPs

0.06 uL X U/mL Taq Polymerase

0.31 uL F/R 20uM combined primers*****

46. Aliquot 11.5 uL master-mix per tube and add 1 uL IP sample.

47. Run samples in thermocycler with an initial denaturation step of 94°C for 2 min, followed by 38 cycles of 94°C for 45 sec, 60°C for 45 sec and 72°C for 2 min, and then a final elongation step of 72°C for 5 min.

48. Add 2.5 uL 6x loading dye to every 12.5 uL reaction and load 6 uL onto 1.5% agarose gel.

49. Score each IP based on anticipated enrichment of amplicon DNA.

CRITICAL STEP Before proceeding, verify sample quality with corresponding primer sets listed above.

50. If several IPs of the same sample and S-phase fraction pass the screening, pool equal amounts of each IP to a final volume of 50 uL (e.g. If two IP pass, combine 25 uL of each to pool).

***** Mitochondrial primer sets should be used at 1.0 uM concentration instead of 0.5 uM. For this master-mix add 0.63 uL F/R 20uM combined primers and 9.31 uL nuclease-free water per reaction.

PROTOCOL 5. Whole Genome Amplification - TIMING 8 h

Once purified by immunoprecipitation and screened for sample quality, BrdU incorporated DNA must be amplified to obtain sufficient amounts for array hybridization. If multiple samples pass PCR screening, DNA from parallel immunoprecipitations can be pooled together and used as the starting material for whole genome amplification. Otherwise, a single screened immunoprecipitation is used. Whole genome amplification (WGA) is performed using GenomePlex Complete Whole Genome Amplification Kit (Sigma Catalog Number WGA2) followed by GenomePlex WGA Reamplification Kit (Sigma Catalog

Number WGA3). Amplified samples are then loaded onto a gel to determine size range and screened once more by PCR to ensure no bias was introduced during amplification.

51. Precipitate DNA fractions by adding 1.25 uL 2 mg/mL glycogen, 20 uL 10M ammonium acetate and 150 uL ethanol to each 50 uL IP sample (if pooling multiple samples, 50 uL total volume should still be used). Mix well, let stand at -20°C for 20 min, then centrifuge for 30 min at max speed and 4°C.

52. Rinse pellets with 70% ethanol, air dry and resuspend in 10 uL nuclease free water.

53. Transfer the 10uL samples to 0.2 mL PCR tubes and follow GenomePlex Complete Whole Genome Amplification Kit (Sigma Catalog Number WGA2) procedure from the Library Preparation step (i.e. skip Fragmentation)¹².

54. Purify entire WGA products using QIAquick PCR Purification Kit (Qiagen Catalog Number 28106). Elute in 30 uL nuclease free water pre-warmed to 65°C and determine concentration using Nanodrop.

55. Dilute WGA samples to appropriate concentration (we use 1 uL DNA of 20 ng/uL) and follow GenomePlex WGA Reamplification Kit (Sigma Catalog Number WGA3) Reamplification Procedure A.

56. Purify entire reamplified WGA products using QIAquick PCR Purification Kit as done in step 54.

57. Screen purified final products using same PCR method used to screen BrdU-IPs as described in protocol 4.

PROTOCOL 6. S/G1 FACS Sorting (Alternative to Steps 1-57) - TIMING 1 d

In this method, cells are sorted into two fractions, G1 phase and S phase, based on DNA content, and replication timing is derived from the 2-fold copy number increase for early vs. late replicating sequences in pure S-phase populations. DNA analysis using flow cytometry can be performed simply by the use of a single DNA-binding fluorescent dye such as PI or DAPI as originally described. The advantage of this method is that it eliminates the need for BrdU-IP and whole genome amplification steps (described below), which need to be carefully controlled. However, direct comparisons have shown that this method produces a lower signal to noise ratio than protocol 1¹.

58. For adherent cells, remove cell culture medium from exponentially growing cells and replace with cell culture medium containing BrdU at a final concentration of 10 uM. For suspension cells, add BrdU to the cell culture medium at a final concentration of 10µM. In order to obviate the amplification step prior to labeling and array hybridization, start with 6 million cells. One should also prepare a small sample of non BrdU-labeled, EtOH-fixed cells for PI-only control and set aside a small number of BrdU-labeled cells for BrdU-only control.

59. Incubate cells for 30 minutes in a carbon dioxide incubator at 37°C, 5% CO₂

60. Fix as in protocol 1, steps 3-11.

PAUSE POINT Cells can be stored as in protocol 1

61. Aliquot (multiples of) $< 3 \times 10^6$ cells in a 1.5mL tube(s), centrifuge for 5 min., 200 *g* at room temperature. Centrifugation and removal of supernatant is much easier with 1.5 mL tubes as the pellets are very loose.

62. Aspirate supernatant completely with P200. (will be effective if spun down ~3 sec once again to

discard residual supernatant – same for all aspiration procedures below)

63. Loosen pellet by brief vortexing.

64. Add 1ml of 0.1M HCl/0.5% Triton X-100, resuspend by tapping.

65. Incubate 15 min, room temperature in the dark.

66. Centrifuge 5 min, 200 *g* at room temperature, then aspirate supernatant completely.

67. Add 1ml of 0.1M Sodium Tetraborate and resuspend by tapping.

68. Centrifuge 5 min, 200 *g* at room temperature, then aspirate supernatant completely.

69. Add 0.15 µg anti-BrdU antibody in 0.5 ml 0.5% Tween20/1% BSA/PBS and resuspend by tapping.

70. Incubate for 30 min at room temperature in the dark.

71. Centrifuge 5 min, 200 *g* at room temperature, then aspirate supernatant completely.

72. Add 0.5 ml of 0.5% Tween20/1% BSA/PBS.

73. Centrifuge 5 min, 200 *g* at room temperature, then aspirate supernatant completely.

74. Add 1 µg anti-Mouse IgG-AlexaFluor488 in 100 µl 0.5% Tween20/1% BSA/PBS and resuspend by tapping (or when $1-2 \times 10^6$ cells are used, add 0.5 µg anti-Mouse IgG-AlexaFluor488 in 50 µl).

75. Incubate for 30 min at room temperature in the dark.

76. Centrifuge 5 min, 200 *g* at room temperature, then aspirate supernatant completely.

77. Add 0.5 ml of 0.5% Tween20/1% BSA/PBS.

78. Centrifuge 5 min, 200 *g* at room temperature, then aspirate supernatant completely.

79. Resuspend the pellet in 1 ml of 5µg/ml PI in PBS (For “BrdU-only” control, just add PBS).

80. Transfer to a round bottom 5mL tube (i.e. Falcon 2054).

81. Adjust concentration to 2×10^6 /ml by adding 5 µg/ml PI in PBS

82. Bring to flow lab for sorting (on ice, dark)

83. Filter with 37 µm mesh filter.

PROTOCOL 7. Labeling and hybridizing - TIMING 1-3 d

In order to avoid the ambiguity inherent to generalized methods, this protocol is based specifically on NimbleGen products. During our earliest attempts to profile replication timing genome-wide, NimbleGen outperformed other platforms we tested and became our platform of choice. Others have successfully applied this method to additional platforms^{13,14}, including deep sequencing of BrdU-IP DNA¹⁵. Currently, mammalian replication timing data generated from microarray hybridization and deep sequencing is of equal quality^{1,3}, while the microarray method remains more cost-effective and the bioinformatics are considerably less demanding for the typical laboratory. Future advances reducing BrdU-labeling times and sequencing limitations may make this method more worthwhile and accessible¹⁶.

84. Differentially label reamplified early and late WGA DNA fractions with Cy3 and Cy5 dyes following the Sample Labeling Instructions for NimbleGen Dual-Color DNA Labeling Kit (cat. no. 05223547001). Briefly, 1 ug of early or late replicating DNA is labeled with either Cy5 or Cy3 random 7-mer dyes by Klenow reaction, precipitated with isopropanol, and resuspended in nuclease-free water and quantified.

As instructed, combine equal quantities of labeled early and late fraction DNA (specific quantity will depend upon array design).

85. Transfer labeled samples to hybridization buffer, denature at 95°C for 5 min, prepare array(s) with appropriate mixer(s) (specific mixer will depend upon array design), load samples on array(s) and hybridize using NimbleGen Hybridization Kit (cat. no. 05583683001) at 42°C for 24 to 72 hours depending on the array feature size.

86. Following Hybridization, wash array(s) as directed using NimbleGen Wash Buffer Kit (cat. no. 05584507001).

87. Scan array(s) and save images as .tif files using appropriate NimbleGen microarray scanner and software package. Open .tif files in NimbleScan software, overlay grid on the scanned image (i.e. grid cells correspond to probes), and generate two .pair files per experiment, which contain the signal intensity of all probes on the array for either Cy3 or Cy5, for normalization and downstream analysis. Specific instructions for these steps can be found in NimbleGen Arrays User's Guide, CGH Analysis.

88. If desired, arrays can be stripped for reuse using NimbleGen Array Reuse Kit 40 (contact NimbleGen customer service for ordering information). Scan array to confirm successful stripping.

Part II: Normalization and computational analysis of replication timing datasets

In this section of the protocol we focus on computational methods for analyzing replication timing using whole-genome comparative genome hybridization (CGH) microarrays, using the R framework for statistical computing⁸⁻¹⁰ Additional help on using R can be found at <http://cran.r-project.org/>.

For normalization, we use the R package LIMMA (linear models for microarray data), also available with a user interface through the limmaGUI package, for normalization and scaling^{17,18}. Methods for verifying the quality of array data produced are covered in quality control (QC) sections 1-6, summarized in **Fig.3**.

1.) Create RGL files from the original NimbleGen .pair files. These "Red-Green lists" contain columns for both red (Cy5) and green (Cy3) channel signal intensities. Note that the genomic position column may need to be set as numeric manually with classes[x] = "numeric" (where x is the column number containing position information) after line 2.

```
1| tab5rows <- read.delim("318990_4L1210LymphoblastP1_532.pair", header = TRUE, nrows = 5, skip=1)
2| classes <- sapply(tab5rows, class)
3| mLymph1Cy3 <- read.delim("318990_4L1210LymphoblastP1_532.pair", header=T, nrows=390000,
comment.char = "", colClasses=classes, skip=1)
4| mLymph1Cy5 <- read.delim("318990_4L1210LymphoblastP1_635.pair", header=T, nrows=390000,
comment.char = "", colClasses=classes, skip=1)
5| mLymph1 <- data.frame(S_Cy5=mLymph1Cy5[,10],S_Cy3=mLymph1Cy3[,10])
6| mLymph2 <- data.frame(S_Cy5=mLymph2Cy5[,10],S_Cy3=mLymph2Cy3[,10])
```

```
7| write.table\(mLymph1, file="318990_4L1210LymphoblastP1.rgl.txt", row.names=F, quote=F, sep="\t",  
eol="\r\n")
```

```
8| write.table\(mLymph2, file="319048_4L1210LymphoblastP1-2.rgl.txt", row.names=F, quote=F, sep="\t",  
eol="\r\n")
```

2.) Create a "targets" text file that describes the target files for normalization. We will name this file "T.txt". Note that, to be read correctly, the file should be tab-delimited and contain only one carriage return at the end of the final line. Place this file in the same directory as the raw .pair files and .rgl files generated above.

```
SlideNumber Name FileName Cy3 Cy5  
1 mLymph1 318990_4L1210LymphoblastP1.rgl.txt late early  
2 mLymph2 319048_4L1210LymphoblastP1-2.rgl.txt early late
```

3.) Install the LIMMA package. Install a current version of the LIMMA package according to the instructions at <http://bioinf.wehi.edu.au/limma/> , or using the command line interface:

```
9| source\( "http://www.bioconductor.org/biocLite.R" )  
10| biocLite\( "limma" )  
11| biocLite\( "statmod" )
```

4.) Perform loess and scale normalization using LIMMA. Loess normalization `\(normalizeWithinArrays)` corrects the internal dependence of red-green ratios on their intensity independently for each array, and is examined further in sections QC#1 and QC#3. Scale normalization `\(normalizeBetweenArrays)` equalizes the distribution of timing values between multiple samples for comparisons.

```
12| library\(limma)  
13| setwd\( "D:\\RT project\\Raw datasets" )  
14| t = readTargets\( "T.txt", row.names="Name" )  
15| r = read.maimages\( t, source="generic", columns=list\( R="S_Cy5", G="S_Cy3" ) )  
16| MA.l = normalizeWithinArrays\( r, method="loess" )  
17| MA.q = normalizeBetweenArrays\( MA.l, method="scale" )
```

QC # 1 – Distribution of spot intensities for red and green channels

The distributions of red and green spot intensities should be fairly well aligned , and have tails with high `\` (above 211, or 2048) signal. Experiments in which signal intensity drops off more sharply will often show higher levels of noise in the final dataset `\(Fig.3A)`.

```
18| plotDensities\( r )  
19| plotDensities\( MA.l )  
20| plotDensities\( MA.q )
```

QC #2 – Distribution of ratio values before and after normalization

Boxplots of the Cy5/Cy3 ratios for each experiment may be slightly different before normalization (and after within-array normalization), but 1st and 3rd quartiles (the box boundaries) of all experiments should be equal after between-array normalization (Fig.3B).

```
21| boxplot(MA.l$M~col(MA.l$M), names=colnames(MA.l$M))
22| boxplot(MA.q$M~col(MA.q$M), names=colnames(MA.q$M))
```

QC #3 – Ratio vs. intensity plots

Create MA plots to examine dependence of ratios on signal intensities. Points will often be skewed to low Cy5/Cy3 ratios at low intensities due to photobleaching of Cy5, but should be corrected after within-array loess normalization in LIMMA (Fig.3C).

```
23| plotMA(r, array=1) # Raw data, replicate 1
24| plotMA(MA.l, array=1) # After within-array normalization
```

5.) Create an intermediate file containing the normalized datasets. At this stage we can write a file containing the results of normalization and remove the other data from memory.

```
25| write.table(MA.q$M, file="Loess_mLymph_112909.txt", quote=F, row.names=F, sep="\t")
26| rm(r, MA.l, MA.q, mLymphCy3, mLymphCy5, mLymph1, mLymph2); gc(reset=T)
```

6.) Assign position and chromosome information to the normalized datasets. This can be accomplished using the original .pair files, which typically contain this information in columns "POSITION" and "SEQ_ID" respectively.

```
27| tab5rows <- read.table("Loess_mLymph_112909.txt", header = TRUE, nrows = 5)
28| classes <- sapply(tab5rows, class)
29| RT = read.table("Loess_mLymph_112909.txt", header=T, nrows=390000, comment.char = "",
colClasses=classes)
30| tab5rows <- read.delim("318990_4L1210LymphoblastP1_635.pair", header = T, nrows = 5, skip=1)
31| classes <- sapply(tab5rows, class)
32| a = read.delim("318990_4L1210LymphoblastP1_635.pair", header=T, nrows=390000, comment.char
= "", colClasses=classes, skip=1)
33| RT$CHR = a$SEQ_ID
34| RT$POSITION = a$POSITION
```

For datasets containing a different format of SEQ_ID column (e.g., "chr11:1-134452384", use the following lines to parse chromosome and probe position information from the PROBE_ID column:

```
35| b = a$PROBE_ID
36| b = as.character(b)
```

```

37| x = strsplit(b, "FS")
38| y = unlist(x) # chr [1:1439379] "CHR10" "057882816" "CHR10"
39| y1 = y[c(TRUE, FALSE)] # chr [1:719690] "CHR10" "CHR10" "CHR10"
40| y2 = y[c(FALSE, TRUE)] # chr [1:719689] "057882816" "104954995"
41| y2 = as.numeric(y2) # num [1:719690] 5.79e+07 1.05e+08 7.31e+07
42| RTa= data.frame(CHR=y1,POSITION=y2, RT, stringsAsFactors=F)

```

7.) Plot timing values across a chromosome. This serves to verify the orientation for early/late domains, as well as the overall technical quality of the experiments (Fig.2A).

```

43| RTb = subset(RT, RT$CHR=="chr1")
44| par(mar=c(3.1,4.1,1,1),mfrow=c(2,1))
45| plot(RTb[,4]~RTb$POSITION,pch=19,cex=0.2,col="grey",ylim=c(-3,3))
46| plot(RTb[,5]~RTb$POSITION,pch=19,cex=0.2,col="grey",ylim=c(-3,3))

```

Using known regions of early or late replication, we can now verify that the timing values are properly oriented. If not, we reverse them by multiplying the appropriate data columns by -1.

```

47| RT[,c(4,5)] = RT[,c(4,5)] * -1 # Reverses early/late orientation

```

8.) Average replicate datasets. This is quite straightforward:

```

48| RT$mLymphAve = (RT[,4] + RT[,5]) / 2

```

QC #4 – Autocorrelation function (ACF)

Since nearby loci should replicate almost synchronously, the correlation of timing between neighboring probes, or autocorrelation, is a useful measure of overall data quality. High-quality datasets will have a correlation between nearest neighbor timing values of $R = 0.60-0.80$. Before checking the autocorrelation, it is important to ensure that the data is sorted by chromosome and genomic position, as below:

```

49| RT = RT[order(RT$CHR, RT$POSITION),]

```

Next, plot the autocorrelation and output the second (nearest neighbor) autocorrelation to the console:

```

50| acf(RT[,4],lag=1000)$acf [2] # Replicate 1: R = 0.742
51| acf(RT[,5],lag=1000)$acf [2] # Replicate 2: R = 0.665
52| acf(RT$mLymphAve, lag=1000)$acf [2] # Averaged 1 and 2: R = 0.762

```

QC #5 – Array image to check for spatial artifacts

Spatial artifacts in the scanned images should not much affect timing values in any particular location in the genome, but may reduce the overall signal to noise ratio of the experiment. To check for spatial artifacts, examine the original .tiff images for signs of regional bias.

Analysis of static properties of the timing program in a given cell type

1.) Loess smoothing. Loess is a locally weighted smoothing method useful for deriving an overall replication profile from noisier raw data. The smoothness of the fitted line can be set by a bandwidth parameter, which we typically set to the equivalent of 300 kilobases for studies of human and mouse replication timing. Note that optimal smoothing span is system-specific and should be determined empirically by finding the smallest span that reproduces most of the features between replicate profiles.

```
53| chrs = levels\ (RT$CHR); str\ (chrs) # Create a list of all chromosomes
54| chrs = chrs\ [-22] # Remove chrY_random
55| AllLoess = NULL
56| for \ (chr in chrs) \ { # For each chromosome,
57| RTI = NULL # Create a variable to store loess-smoothed values
58| RTb = subset\ (RT, RT$CHR==chr) # Subset the dataset to a single chromosome
59| lspan = 300000/\ (max\ (RTb$POSITION)-min\ (RTb$POSITION)) # Set smoothing span
60| cat\ ("Current chromosome: ", chr, "\n") # Output current chromosome/dataset
61| RTIa = loess\ (RTb$X318990_4L1210LymphoblastP1.rgl ~ RTb$POSITION, span = lspan)
62| RTIb = loess\ (RTb$X319048_4L1210LymphoblastP1.2.rgl ~ RTb$POSITION, span = lspan)
63| RTIc = loess\ (RTb$mLymphAve ~ RTb$POSITION, span = lspan)
64| RTI = data.frame\ (CHR=RTb$CHR, POSITION=RTb$POSITION, RTIa$fitted, RTIb$fitted, RTIc$fitted)
65| AllLoess = rbind\ (AllLoess, RTI)
66| }
67| x = as.data.frame\ (AllLoess)
68| names\ (x)\ [4:5] = c\ ("x300smo_4L1210LymphoblastR1", "x300smo_4L1210LymphoblastR2")
```

We can now check the results of loess smoothing using the following **(Fig. 2B)**:

```
69| RTc = subset\ (RT, CHR=="chr1")
70| LSc = subset\ (LS, CHR=="chr1")
71| par\ (mar=c\ (2.2,5.1,1,1),mfrow=c\ (3,1), col="grey", pch=19, cex=0.5, cex.lab=1.8, xaxs="i")
72| plot\ (RTc$X318990_4L1210LymphoblastP1.rgl~RTc$POSITION, ylab="mLymph R1", xaxt="n")
73| lines\ (LSc$x300smo_4L1210LymphoblastR1~LSc$POSITION, col="blue3", lwd=3)
74| plot\ (RTc$X319048_4L1210LymphoblastP1.2.rgl~RTc$POSITION, ylab="mLymph R2", xaxt="n")
75| lines\ (LSc$x300smo_4L1210LymphoblastR2~LSc$POSITION, col="blue3", lwd=3)
76| plot\ (RTc$mLymphAve~RTc$POSITION, xlab="Coordinate \ (bp)", ylab="mLymph ave")
77| lines\ (LSc$x300smo_4L1210LymphoblastAve~LSc$POSITION, col="blue3", lwd=3)
```

QC #6 – Correlations between replicate experiments

Once loess smoothing is completed and the technical quality of the array data is established, we can compare biological replicate experiments to establish the relative level of biological similarity between

samples. To isolate biological rather than array quality differences, we typically use loess-smoothed averaged-replicate data, rather than individual, raw or normalized data, for these correlations:

```
78| cor(x[,c(4:6)])
```

```
Rep1 Rep2 Ave
```

```
Lymphoblast Rep1 1.000 0.978 0.995
```

```
Lymphoblast Rep2 0.978 1.000 0.994
```

```
Lymphoblast Ave 0.995 0.994 1.000
```

2.) Segmentation. Segmentation is a useful way to define the boundaries of replication domains and determine their average timing. Biologically, these segments appear to correspond to domains of coordinately regulated, synchronously firing origins that may be part of replication factories. We perform segmentation using the DNACopy algorithm designed by Venkatraman et al.¹⁷.

```
79| library(DNACopy)
```

```
80| mLymph = CNA(RT$mLymphAve, RT$CHR, RT$POSITION, data.type="logratio", sampleid = "mLymph")
```

```
81| Seg.mLymph = segment(mLymph, nperm=10000, alpha=1e-15, undo.splits="sdundo", undo.SD=1.5, verbose=2)
```

```
82| write.table(Seg.mLymph$output, row.names=F, quote=F, file="Mouse lymphoblast segments.txt", sep="\t")
```

The number of segments assigned can be examined using `str(Seg.mLymph$output)`, and visualized using various functions built into DNACopy (Fig. 2C).

```
83| par(ask=T,mar=c(3.1,4.1,1,1)) # Set figure margins
```

```
84| plot(Seg.mLymph, plot.type="c") # Plot each chromosome separately
```

```
85| plot(Seg.mLymph, plot.type="s") # Plot overview of all chromosomes
```

```
86| plot(subset(Seg.mLymph,chromlist="chr2"), pch=19, pt.cols=c("gray","gray"), xmaploc=T, ylim=c(-3,3), main="")
```

Datasets with different technical quality should be adjusted to the same autocorrelation level before segmentation to avoid assigning additional domains to higher-quality datasets. For this purpose, we adapt the "jitter" function of Chambers et al.¹⁸, substituting Gaussian noise for uniformly distributed noise.

```
88| addNoise = function(x, factor = 1, amount = NULL) {
```

```
89| if (length(x)==0)
```

```
90| return(x)
```

```
91| if (!is.numeric(x))
```

```
92| stop("'x' must be numeric")
```

```

93| z <- diff\(r <- range\(x\[is.finite\(x)])
94| if \(z==0)
95| z <- abs\(r \[1] )
96| if \(z==0)
97| z <- 1
98| if \(is.null\(amount)) \{
99| d <- diff\(xx <- unique\(sort.int\(round\(x, 3 -
100| floor\(\log10\(z))))))
101| d <- if \(length\(d)
102| min\(d)
103| else if \(xx != 0)
104| xx/10
105| else z/10
106| amount <- factor/5 * abs\(d)
107| }
108| else if \(amount==0)
109| amount <- factor * \(z/50)
110| x + stats::rnorm\(length\(x), 0, amount)
111| }
112| RT$mLymphR1noise = addNoise\(as.numeric\(RT$mLymphR1), factor=300, amount=.848)
113| RT$mLymphR2noise = addNoise\(as.numeric\(RT$mLymphR2), factor=300, amount=.968)
114| acf\(RT$mLymphR1noise)$acf \[2] #0.8274651
115| acf\(RT$mLymphR2noise)$acf \[2] #0.8274654

```

3.) Domain size analysis (boxplots of sizes). After segmentation, the sizes of replication domains can be extracted from the segmented dataset to examine average sizes examined together or separately for domains with early or late timing.

```

116| LymphR1 = Seg.mLymphR1$output; LymphR2 = Seg.mLymphR2$output
117| LymphR1$size = LymphR1$loc.end - LymphR1$loc.start
118| LymphR2$size = LymphR2$loc.end - LymphR2 $loc.start
119| LymphR1Early = subset\(LymphR1, LymphR1$seg.mean > 0)
120| LymphR2Late = subset\(LymphR2, LymphR2$seg.mean < 0)
121| boxplot\(LymphR1Early$size, LymphR2Early$size, LymphR1Late$size, LymphR2Late$size)

```

Analysis of dynamic changes in the timing program

To examine changes in the replication program during differentiation, several useful methods leverage the segmentation and loess smoothing methods introduced above. Since no single method is sufficient to fully describe the type, degree, and distribution of timing changes during development, we cover several complementary ways to measure these properties and explore the relationships between cell types.

4.) Percent changes analysis. To quantify the amount of the genome changing timing between two or more cell types, we recommend applying an empirical cutoff for biologically measurable timing changes to quantify these genome-wide. As most methods for quantifying timing changes are sensitive to scale differences, datasets should be scaled and normalized together in LIMMA prior to analysis.

```
122| RTd1 = RT$mLymphR1 - RT$mLymphR2
123| mLength = length(RTd1) # Total number of probes
124| s = 0.67 # Cutoff for significant changes
125| sum(abs(RTd1)>s)/mLength # Percentage changing, R1 vs. R2
126| sum(RTd1 < -s)/mLength # EtoL subset: 1.6%
127| sum(RTd1 > s)/mLength # LtoE subset: 1.3%
```

5.) Clustering approaches. Our two most common methods are hierarchical clustering, which agglomerates experiments from the most similar pairs of replication profiles to gradually more inclusive clusters, and k-means clustering, which partitions a set of timing profiles into a pre-set k number of similar patterns. For k-means clustering, we use `Cluster19` and `TreeView` (<http://rana.lbl.gov/EisenSoftware.htm>). For hierarchical clustering, we use the "pvclust" package in R²⁰. Since many clustering algorithms are computationally intensive, and to improve the precision of individual RT measurements, we average timing datasets into 200kb windows prior to clustering:

```
128| mLymph.R1 = NULL; mLymph.R2 = NULL
129| for (x in 1:10994) {
130| z1 = x * 35 # For 385k arrays,
131| z2 = (x+1) ** 35 # 5.8kb median spacing ** 35 = 203kb
132| mLymph.R1[x] = mean(RT$mLymphR1[z1:z2])
133| mLymph.R2[x] = mean(RT$mLymphR2[z1:z2])
134| cat("Current window: ", x, "\n")
135| }
136| RT = data.frame(mLymph.R1, mLymph.R2)
137| library(pvclust)
138| cluster.bootstrap <- pvclust(RT, nboot=1000, method.dist="abscor")
139| plot(cluster.bootstrap)
140| pvrect(cluster.bootstrap)
```

6.) Properties of RT switching domains. To compute domains that switch to earlier or later replication timing during development, we first subtract the normalized (not loess smoothed) values of the two experiments to be compared, then segment the resulting profile of timing changes as shown below:

```
141| dRT = CNA(RT$NPCave-RT$ESCave, RT$CHR, RT$POSITION, data.type="logratio", sampleid= "NPC-ESC dRT")
142| Seg.dRT = segment(dRT, nperm=10000, alpha=1e-15, undo.splits = "sdundo", undo.SD=1.5,
```

```

verbose=2)
143| dRTdom = Seg.dRT$output
144| quantile\(dRTdom$seg.mean, probs = c\((0.05, 0.95)) # Top 5%
145| quantile\(dRTdom$seg.mean, probs = c\((0.40, 0.60)) # Middle 20%
146| LtoEdom = subset\(dRTdom, dRTdom$seg.mean > 1.28552)
147| EtoLdom = subset\(dRTdom, dRTdom$seg.mean < -1.32328)
148| middleDom = subset\(dRTdom, dRTdom$seg.mean > -0.14808 & dRTdom$seg.mean < 0.23698)

```

7.) Assignment of replication timing values to gene promoters. The smoothed data object produced by loess is valuable for calculating timing \((or other) values at a shared set of genomic coordinates, which is especially helpful for comparing datasets from different array platforms or assigning timing data to NCBI's RefSeq gene promoter locations \((<http://www.ncbi.nlm.nih.gov/RefSeq/>) as outlined below:

```

149| chrs = levels\((RefSeq$CHR) # List of chromosomes to be analyzed
150| AllSm = NULL # Store smoothed data for all chromosomes
151| ChrSm = NULL # Store smoothed data for current chromosome
152| for\((chr in chrs) \{
153| RTc = subset\((RT, CHR==chr)
154| RSc = subset\((RefSeq, CHR==chr)
155| cat\("Current chromosome: ", chr, "\n")
156| lspan = 300000/\((max\((RTc$POSITION)-min\((RTc$POSITION))
157|
158| smLym1 = loess\((RT$mLymphR1 ~ RT$POSITION, span = lspan)
159| smLym2 = loess\((RT$mLymphR2 ~ RT$POSITION, span = lspan)
160| smLym3 = loess\((RT$mLymphAve ~ RT$POSITION, span = lspan)
161|
162| Lym1 = predict\((smLym1, RSc$TSS)
163| Lym2 = predict\((smLym2, RSc$TSS)
164| Lym3 = predict\((smLym3, RSc$TSS)
165|
166| ChrSm = data.frame\((CHR=chr,POSITION= RSc$TSS, Lym1, Lym2, Lym3)
167| AllSm = rbind\((AllSm, ChrSm)
168| }
169| write.table\((AllSm, "Mouse lymphoblast RT at RefSeq gene positions.txt", quote=F, sep="\t")

```

8.) Assignment of histone and other epigenetic marks to gene promoters. Through the ENCODE project and other initiatives, a wide array of epigenetic datasets have been made, and similar data repositories. Just as timing can be assigned to gene promoters as outlined above, datasets from GEO \((<http://www.ncbi.nlm.nih.gov/geo/>) or the UCSC Genome Browser \((<http://genome.ucsc.edu/>) can be assigned to any set of genomic loci, including gene promoters or the boundaries of replication domains. Below is a generic method for assigning values from epigenetic mark datasets to promoter regions

surrounding gene transcription start sites. However, note that the specific methods of quantification should be informed by the biological properties of the marks to be quantified and the questions to be addressed.

```
170| chrs = levels\(Marks$CHR); AllGenes = NULL; AllHist = NULL
171| for \(chr in chrs) \{
172| RSc = subset\(RefSeq, CHR==chr)
173| RTc = subset\(Marks, CHR==chr)
174| for\(m in 1:nrow\(RSc)) \{
175| if\(RSc\[m,]$Strand=="+") \{
176| RTcSub = subset\(RTc, \(RTc$Start < RSc\[m,]$txStart +500) & \(RTc$Start > RSc\[m,]$txStart -
177| 2500))
177| AllHist = rbind\(AllHist, apply\(RTcSub, 2, max)\[3:12])
178| AllGenes = rbind\(AllGenes, RSc\[m,]$Gene)
179| }
180| if\(RSc\[m,]$Strand=="-") \{
181| RTcSub = subset\(RTc, \(RTc$Start < RSc\[m,]$txEnd +2500) & \(RTc$Start > RSc\[m,]$txEnd - 500))
182| AllHist = rbind\(AllHist, apply\(RTcSub, 2, max)\[3:12])
183| AllGenes = rbind\(AllGenes, RSc\[m,]$Gene)
184|
185| }
186| cat\( "Chromosome:", chr, " Gene:", m, "/", nrow\(RSc), "\n")
187| }
188| }
189| OutFile = data.frame\(cbind\(AllGenes, AllHist), stringsAsFactors=F)
190| write.table\(OutFile, file="Histone modifications at RefSeq gene
positions.txt", row.names=F, quote=F, sep="\t")
```

9.) Integration of epigenetic mark values over replication domains (density vs. timing) Given that the magnitude of correlations between genetic properties generally increases when measured in larger windows, it is important to quantify these relationships in windows consistent with biologically regulated unit sizes. The method below can be used to compute correlations between domainwide replication timing values and the average level of epigenetic marks within timing domains segmented above.

```
191| Seg.RT = read.table\( "Lymph-1 segments.txt",header=T)
192| chrs = levels\(Seg.RT$chrom)
193| MarksData = NULL # Store averaged mark data in domains
194| RTData = NULL
195| dom = 0
196| for\(chr in chrs) \{
197| Seg.RTb = subset\(Seg.RT, Seg.RT$chrom==chr)
```

```

198| MarksB = subset\(Marks, Marks$CHR==chr)
199| for \(i in 1:dim\(Seg.RTb) \[1]) \{
200| # Find subset of marks in RT domain
201| cat\(("Current chr:", chr, " Domain:", i, "\n")
202| MarksD = subset\(MarksB, MarksB$Start > Seg.RTb\[i,]$loc.start & MarksB$Start < Seg.RTb\[i,]$loc.end)
203| MarksD = MarksD\[3:12]
204| MarksD\[1:10] = MarksD\[1:10] - MarksD\[1]
205| MarksData = rbind\(MarksData, apply\(MarksD,2, "mean"))
206| dom = dom + 1
207| }
208| }
209| cor\(Seg.RT$seg.mean, data.frame\(MarksData))
210| plot\(Seg.RT$seg.mean, data.frame\(MarksData) \[1])

```

References

1. Hiratani, I. et al. Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS Biology* **6**, e245 \ (2008).
2. Hiratani, I. et al. Genome-wide dynamics of replication timing revealed by in vitro models of mouse embryogenesis. *Genome Research* **20**, 155-69 \ (2010).
3. Ryba, T. et al. Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome Research* **20**, 761-70 \ (2010).
4. Pope, B.D., Hiratani, I. & Gilbert, D.M. Domain-wide regulation of DNA replication timing during mammalian development. *Chromosome Research* **18**, 127-36 \ (2010).
5. Yaffe, E. et al. Comparative analysis of DNA replication timing reveals conserved large-scale chromosomal architecture. *PLoS Genetics* **6**, e1001011 \ (2010).
6. R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-2673 \ (R Foundation for Statistical Computing: 2008).
7. Gentleman, R.C. et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology* **5**, R80 \ (2004).
8. Ihaka, R. & Gentleman, R. R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics* **5**, 299 \ (1996).
9. Gilbert, D.M. Temporal order of replication of *Xenopus laevis* 5S ribosomal RNA genes in somatic cells. *Proceedings of the National Academy of Sciences of the United States of America* **83**, 2924-8 \ (1986).

10. Gilbert, D.M. & Cohen, S.N. Bovine papilloma virus plasmids replicate randomly in mouse fibroblasts throughout S phase of the cell cycle. *Cell* **50**, 59-68\ (1987).
11. Aladjem, M.I. et al. Replication initiation patterns in the beta-globin loci of totipotent and differentiated murine cells: evidence for multiple initiation regions. *Molecular and Cellular Biology* **22**, 442-52\ (2002).
12. O'Geen, H. et al. Comparison of sample preparation methods for ChIP-chip assays. *BioTechniques* **41**, 577-80\ (2006).
13. Schübeler, D. et al. Genome-wide DNA replication profile for *Drosophila melanogaster*: a link between transcription and replication timing. *Nature Genetics* **32**, 438-42\ (2002).
14. Schwaiger, M. et al. Chromatin state marks cell-type- and gender-specific replication of the *Drosophila* genome. *Genes & Development* **23**, 589-601\ (2009).
15. Hansen, R.S. et al. Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 139-44\ (2010).
16. Pombo, A. & Gilbert, D.M. Nucleus and gene expression: the structure and function conundrum. *Current Opinion in Cell Biology* \ (2010). doi:10.1016/j.ceb.2010.04.010
17. Venkatraman, E.S. & Olshen, A.B. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* **23**, 657-63\ (2007).
18. Chambers, J.M. *Software for Data Analysis: Programming with R.* \ (Springer Publishing Company, Incorporated: 2008).
19. Eisen, M.B. et al. Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 14863-8\ (1998).
20. Suzuki, R. & Shimodaira, H. Pvcust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**, 1540-2\ (2006).

Figures

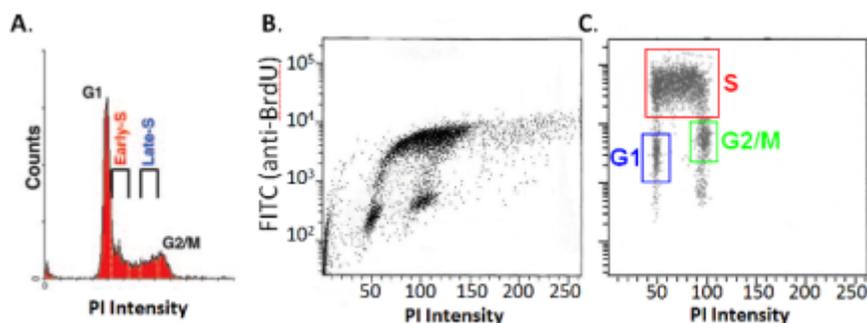


Figure 1

Figure 1. FACS sorting and analysis. **A**. A typical cell cycle profile for a mammalian fibroblast population obtained during FACS analysis by plotting cell count vs. DNA content. In this example, cellular DNA was stained with PI, so DNA content is represented by PI intensity. A G1 peak, representing cells with 2N DNA content, and a G2/M peak, representing cells that have undergone replication and therefore possess a 4N DNA content are labeled. The area between these two peaks is representative of cells in S phase, and can be sorted into two fractions, as indicated here, to obtain early and late S phase samples. **B**. A typical non-corrected BrdU/PI plot. Note how the plot is skewed to the right due to spectral overlap. **C**. A corrected BrdU/PI plot. Sorting windows for nicely separated G1, S and G2/M phases of the cell cycle are indicated.

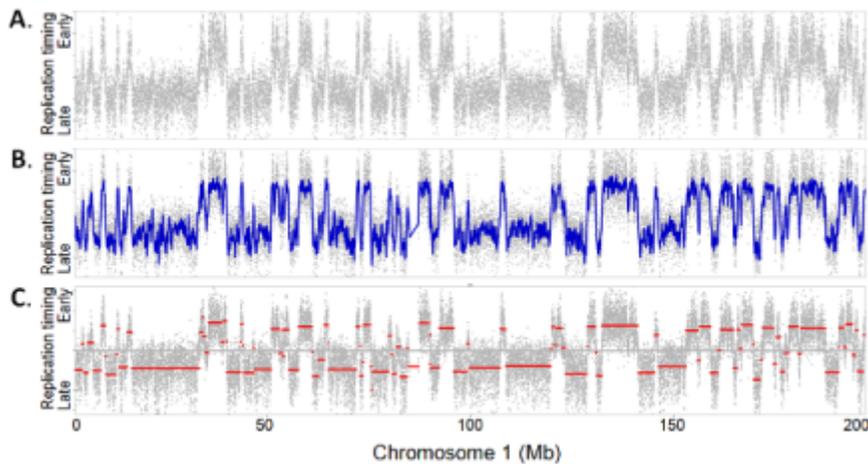


Figure 2

Figure 2. Raw, loess smoothed, and segmented replication timing data. **A**. Replication timing values across chromosome 1. Individual $\log_2(\text{Cy5}/\text{Cy3})$ probe intensities are plotted in grey (y-axis) against their position on chromosome 1 (x-axis). **B**. Raw (grey) and loess-smoothed (blue) timing values along chromosome 1. **C**. Raw data (grey) overlaid with segmented timing domains (red) along chromosome 2.

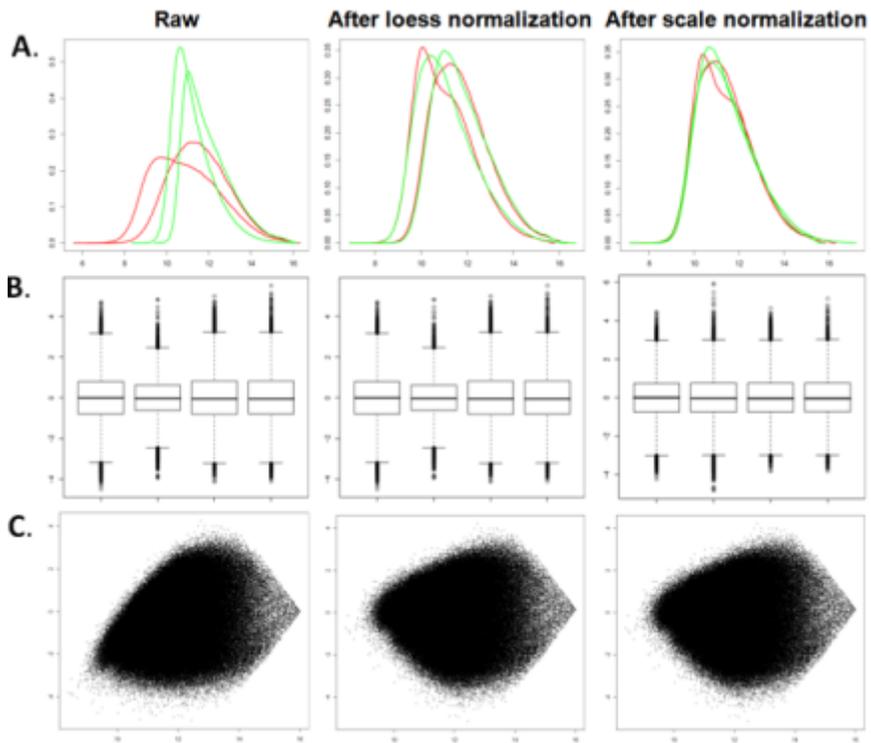


Figure 3

Figure 3. Quality control measures for replication timing datasets. **A.** Distributions of Cy5 (red) and Cy3 (green) channel intensities before and after loess and scale normalization for two mouse lymphoblast replicate experiments. **B.** Boxplots of replication timing values for raw, loess-, and scale-normalized data in four experiments. **C.** Timing ratio (y-axis) vs. probe intensity (x-axis) plots for data before and after loess and scale normalization for one mouse lymphoblast sample.