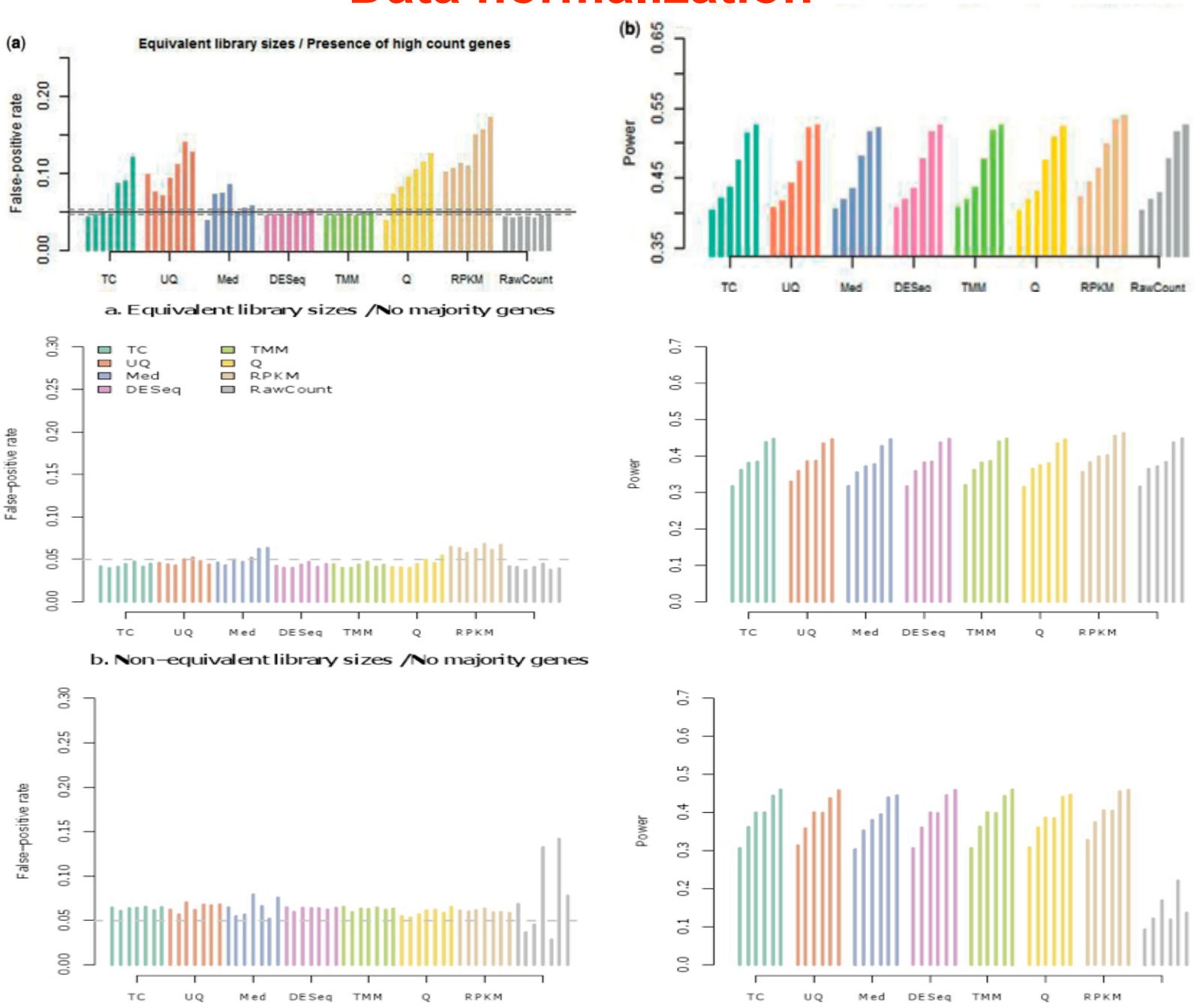
Data normalization



The French StatOmique Consortium. *Brief Bioinform.* 201

Human replication timing program :

determination by deep sequencing & role in genome evolutio

Chun-Long CHEN

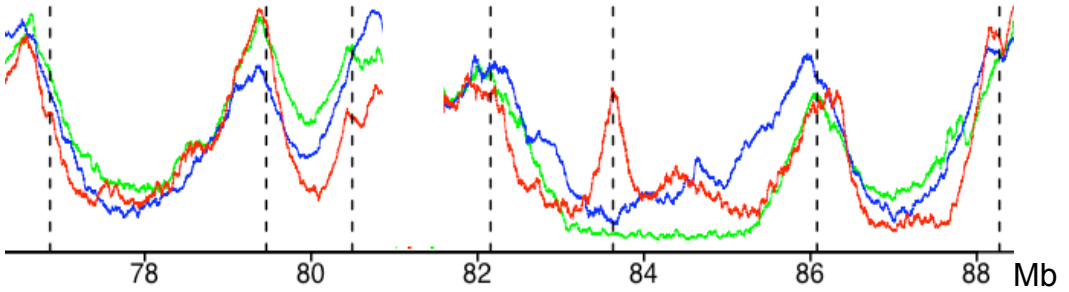
Genome Analysis lab, Centre de Génétique Moléculaire

07/12/12. BioInfo Club. IJM. Par

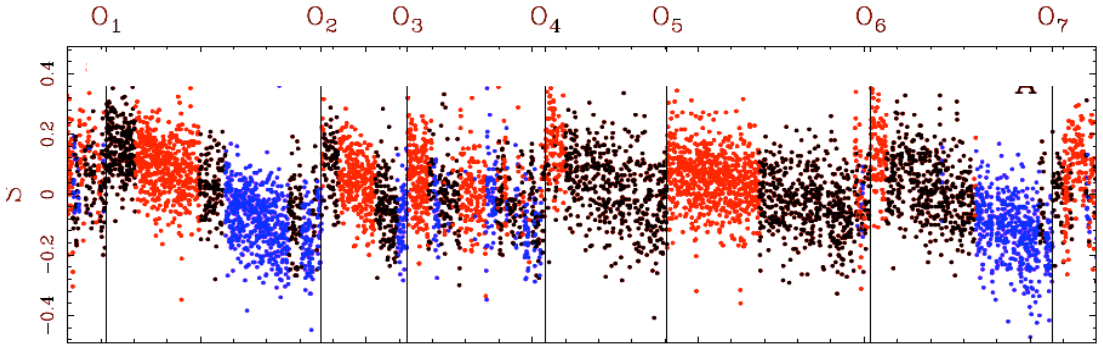
Human replication timing program :

determination by deep sequencing & role in genome evolutio

Replication timing



Nucleotide compositional skew



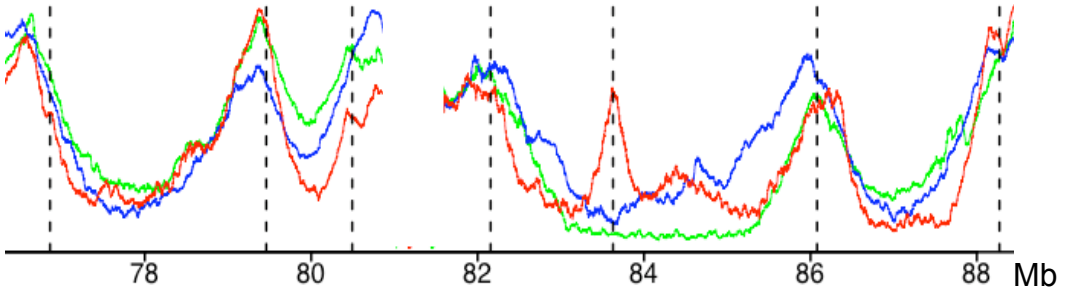
- **What are these beautiful profiles?**
- **What can we learn from these profiles?**

Human replication timing program :

determination by deep sequencing & role in genome evolutio

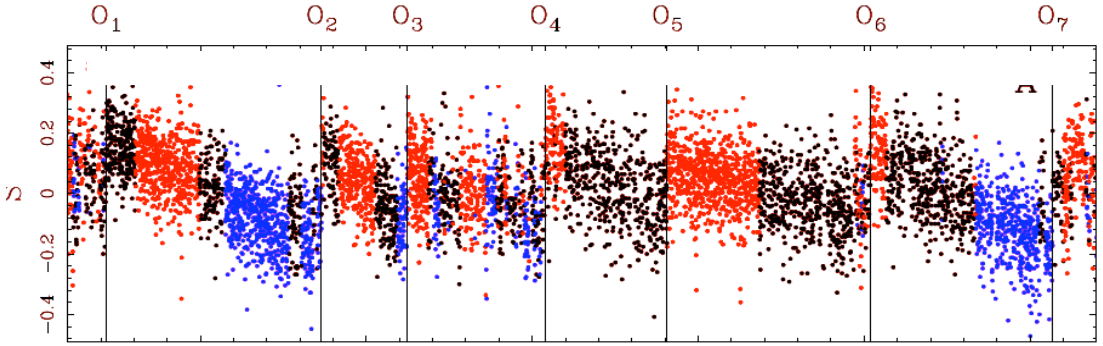
- **Part I: Spatio-temperol replication program of the human genome**

Replication timing



- **Part II: Impact of replication on the evolution and organization of the genome**

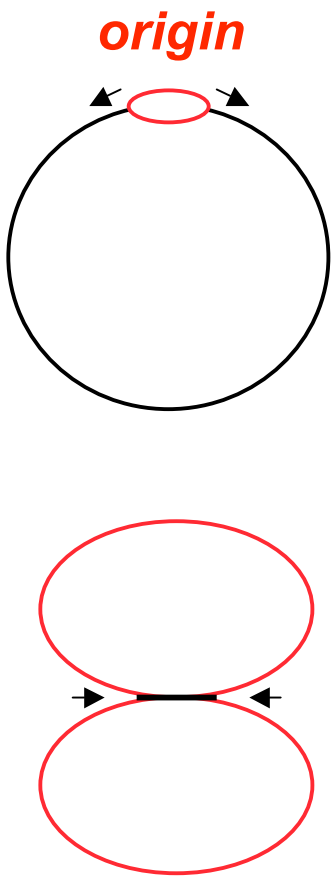
Nucleotide compositional skew



INTRODUCTION

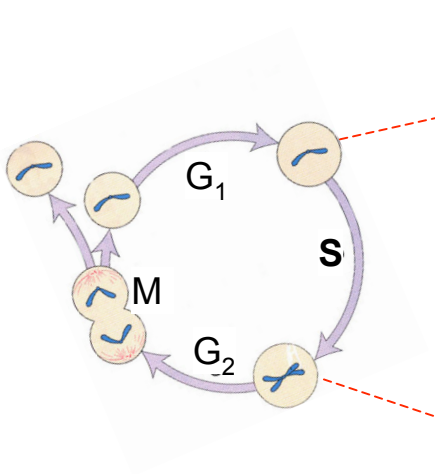
Genome replication program

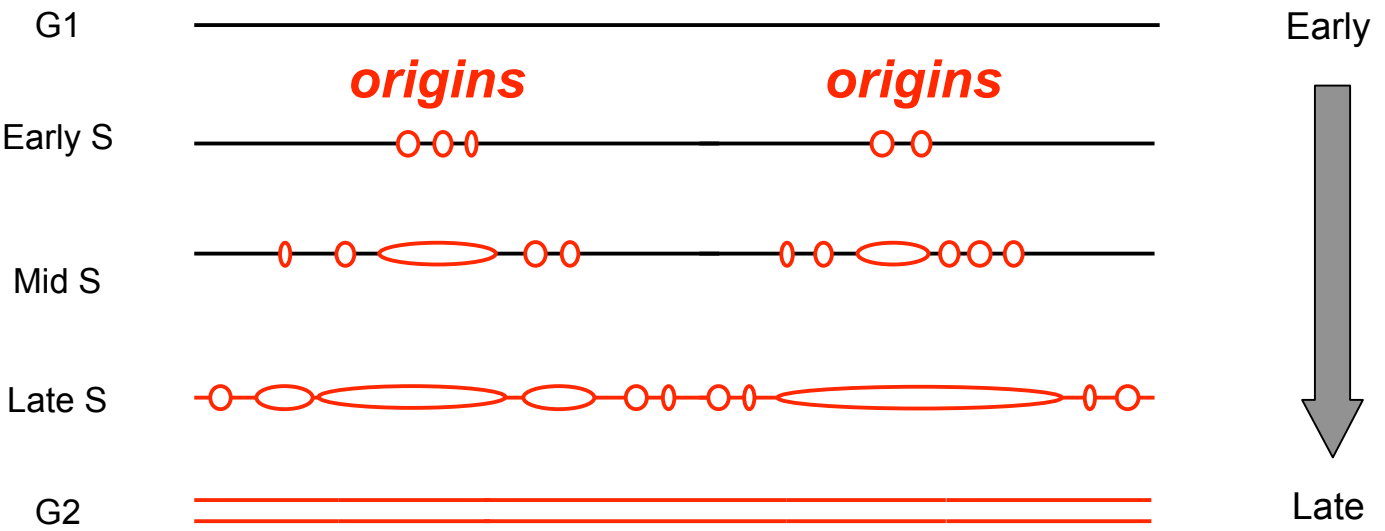
Bacteria:



Genome replication program

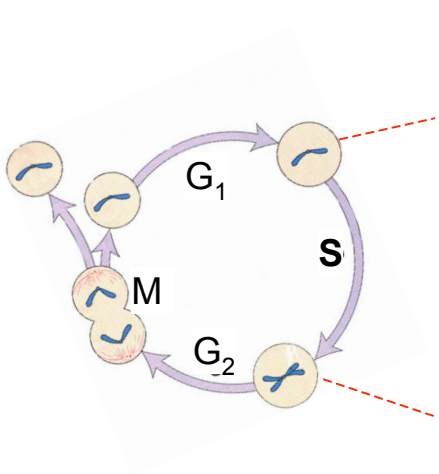
Eukaryote :

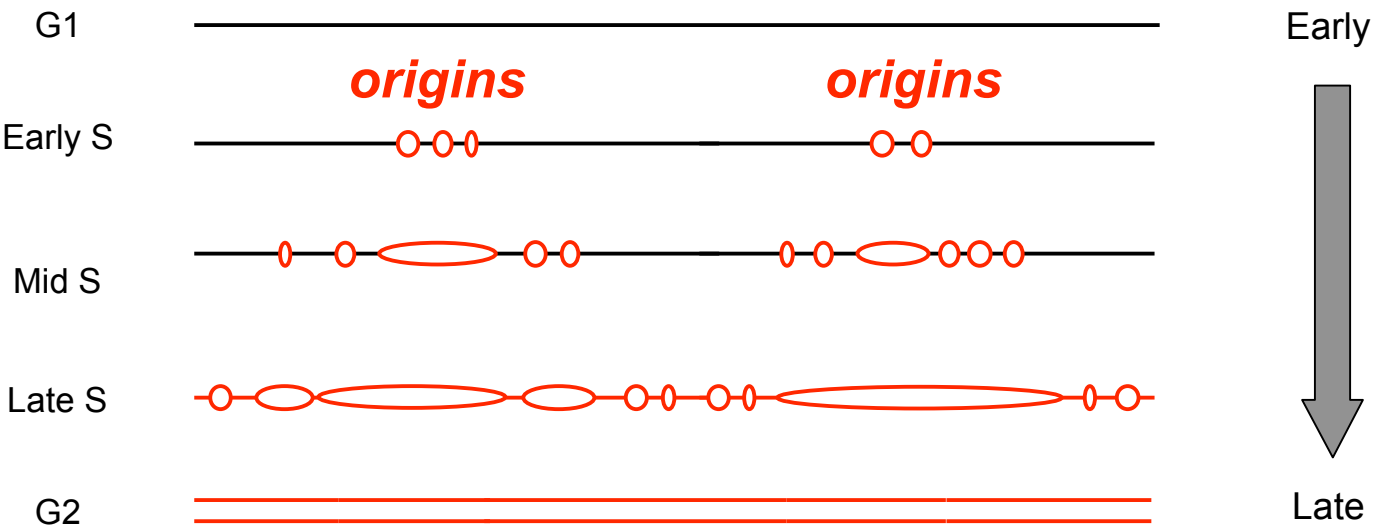




Genome replication program

Eukaryote :



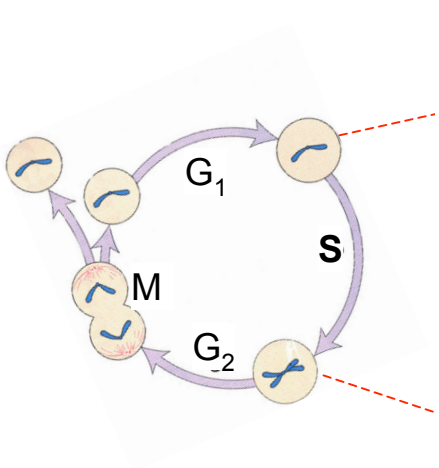


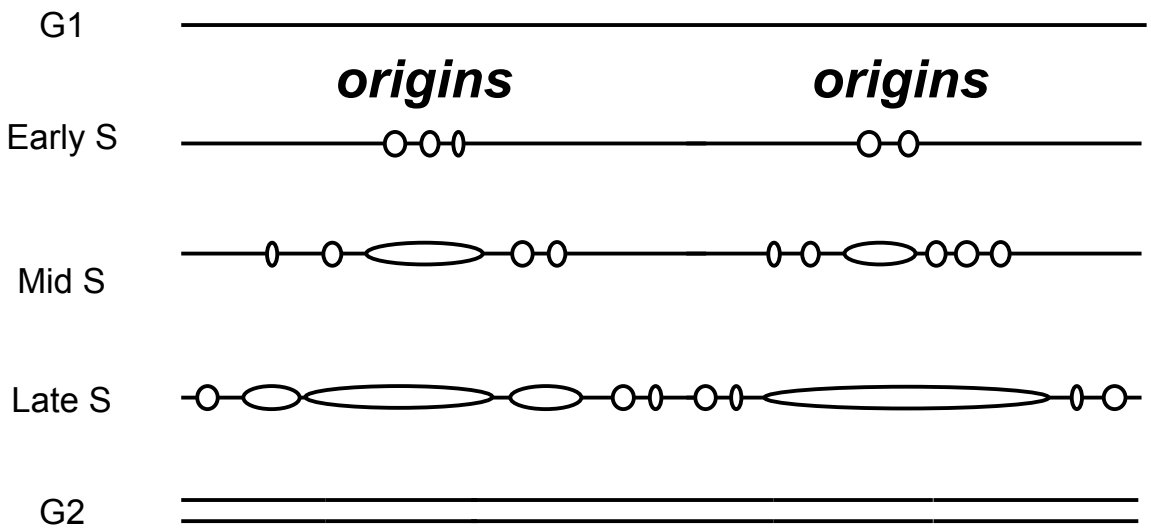
- ***S. cerevisiae :*** ARS regions (~ 125 bp ; 11 bp ACS consensus) ; most origins determined
- ***S. pombe :*** ARS (~ 750 bp; no consensus, but AT-stretch) a large number of origins determined
- **multicellular eukaryotes :** ***replication program is poorly known !***

  Human genome (3.3 10^9 base pairs)

  - replication fork speed: 0.7 – 1.5 kbp/min
  - ~ 30 000 to 50 000 replication orgins : one origin every 50-100 kbp
  - replication is achieved in ~ 8 hours

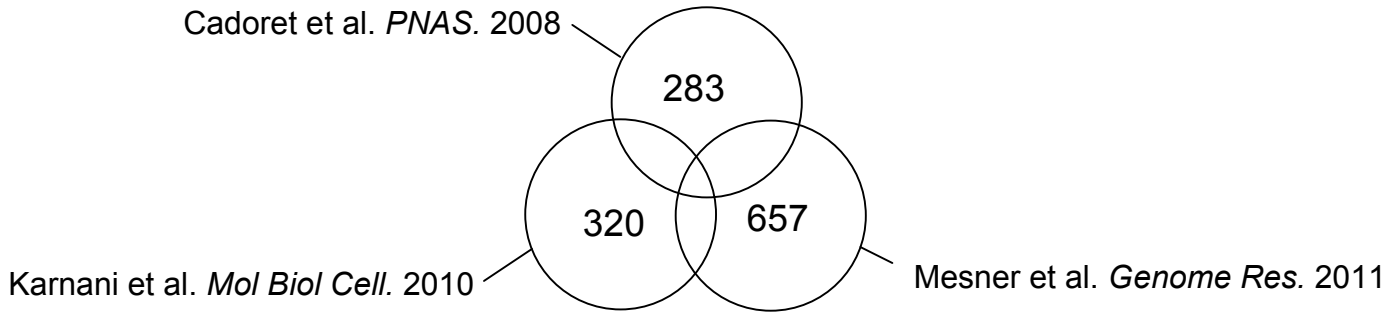
Human replication origin positions are still poorly known





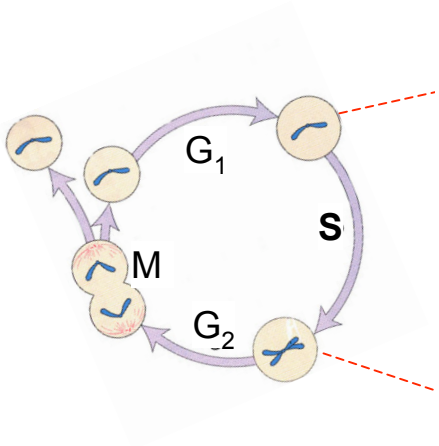
Replication origins

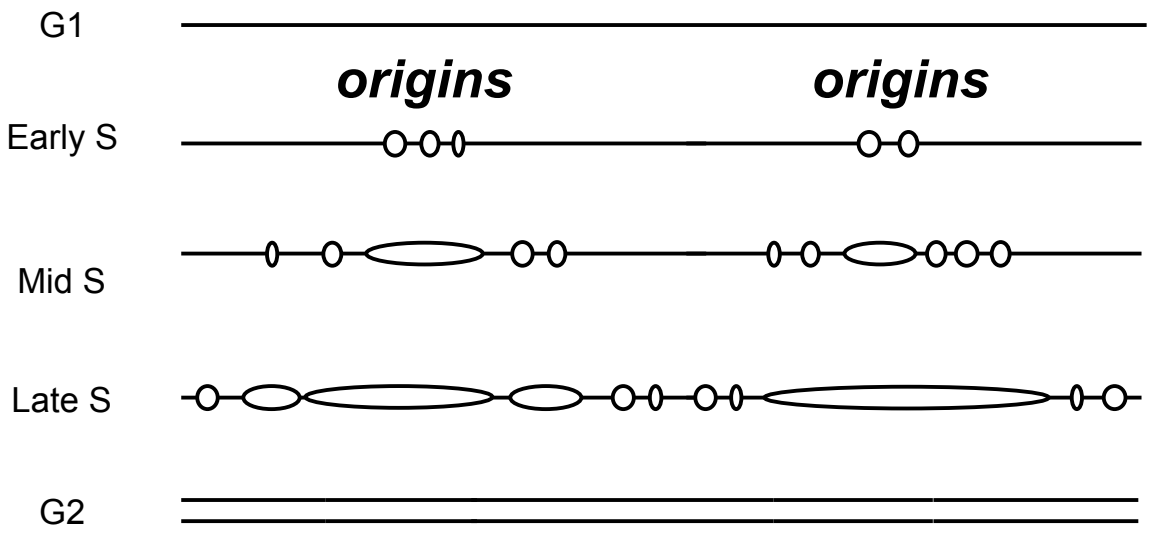
(ENCODE region, 1% of genome)



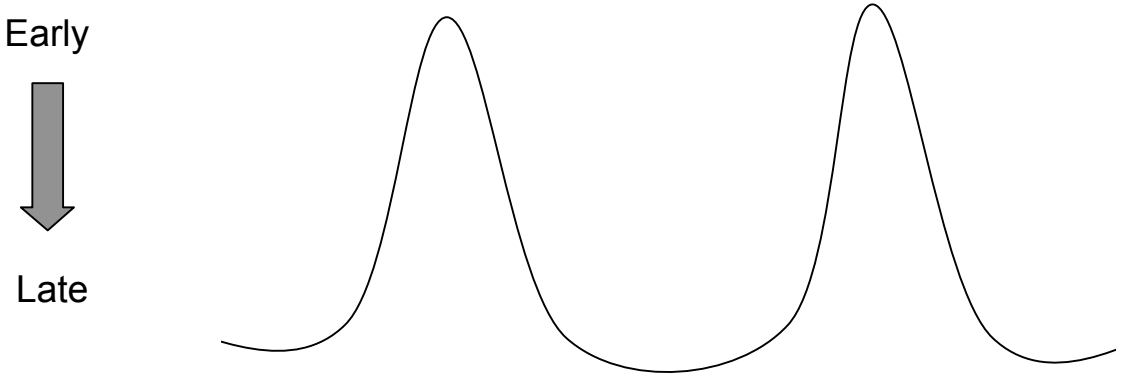
low agreement (<14% overlap) between different studies

But reliable replication timing data





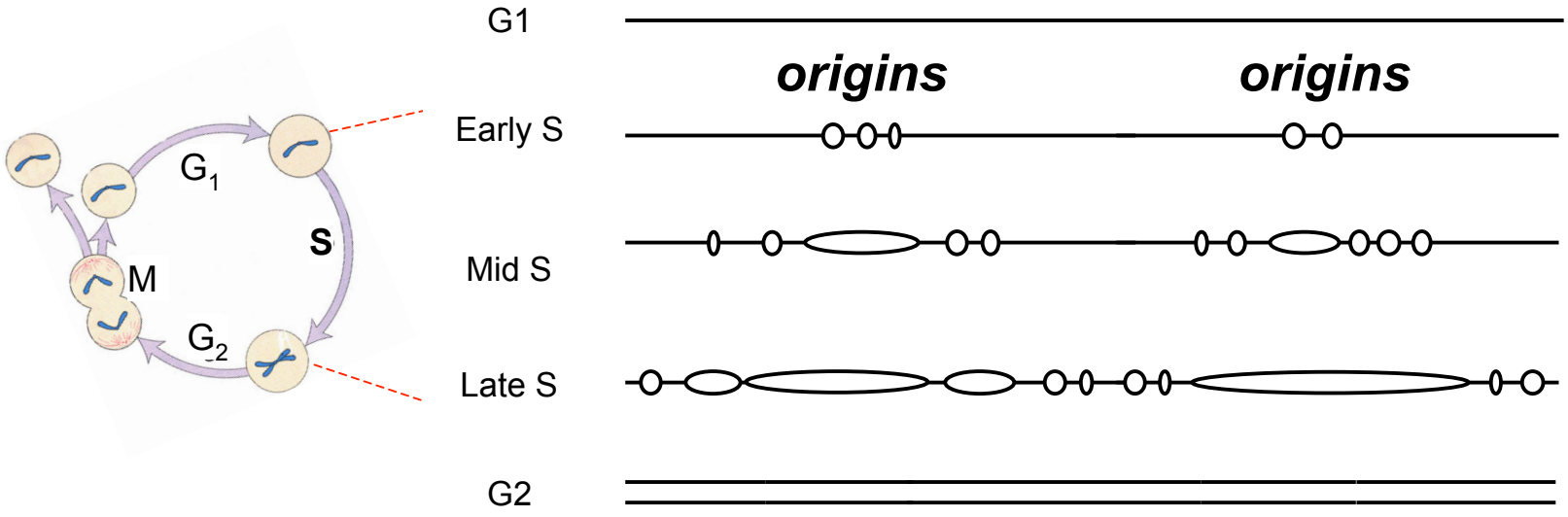
Replication timing



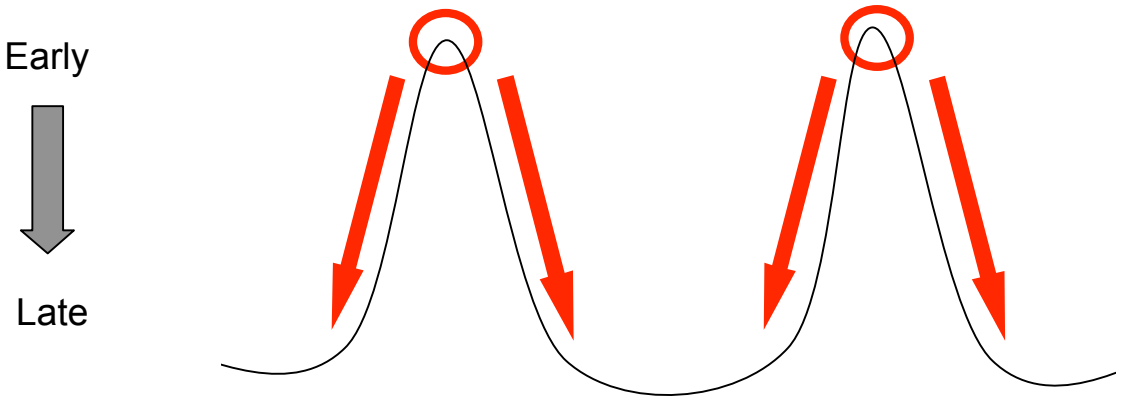
good agreement between different studies

Woodfine et al. *Hum. Mol. Genet.* 2004
White et al. *PNAS.* 2004
Woodfine et al. *Cell Cycle.* 2005
Karpani et al. *Genome Res.* 2007

But reliable replication timing data

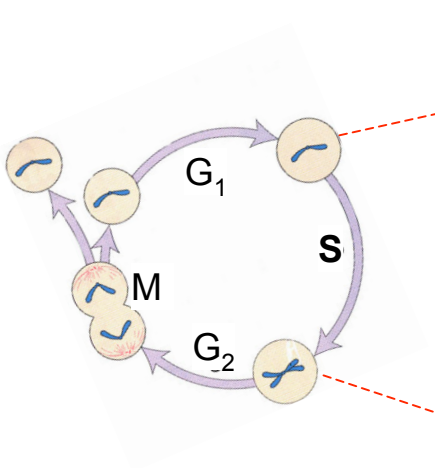


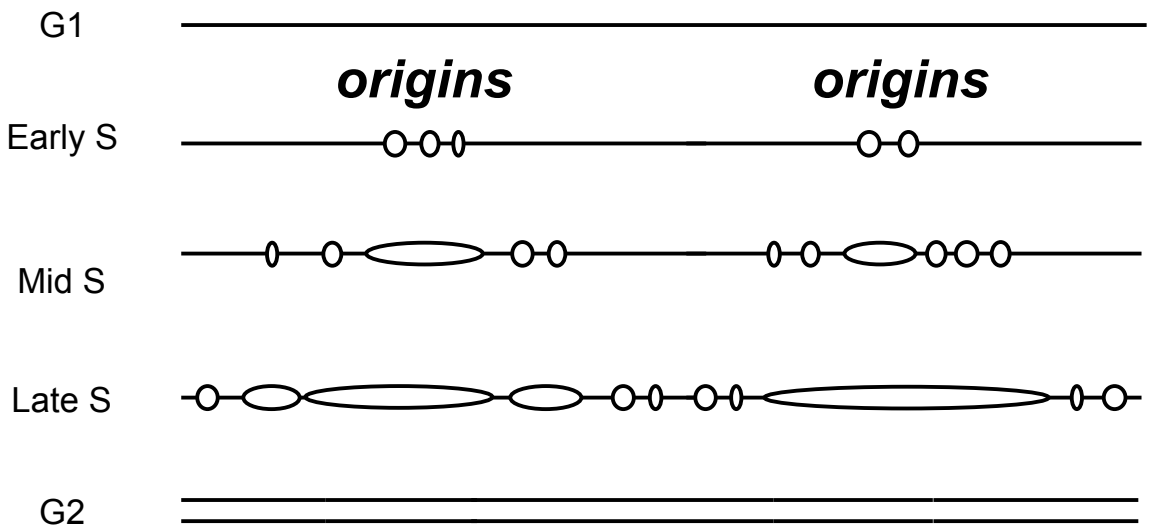
Initiation zones= regions with multiple replication origins



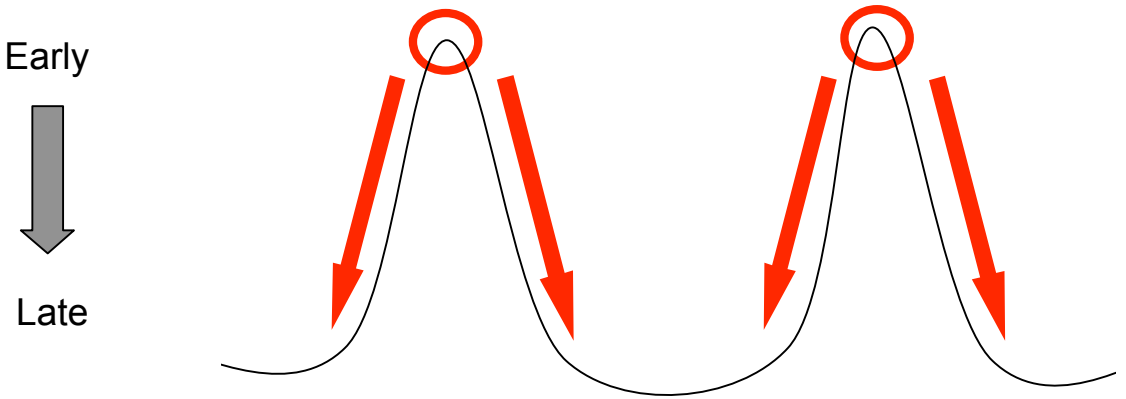
Woodfine et al. *Hum. Mol. Genet.* 2004
White et al. *PNAS.* 2004
Woodfine et al. *Cell Cycle.* 2005
Karpani et al. *Genome Res.* 2007

But reliable replication timing data





Initiation zones= regions with multiple replication origins



Woodfine et al. *Hum. Mol. Genet.* 2004
White et al. *PNAS.* 2004
Woodfine et al. *Cell Cycle.* 2005
Karpani et al. *Genome Res.* 2007

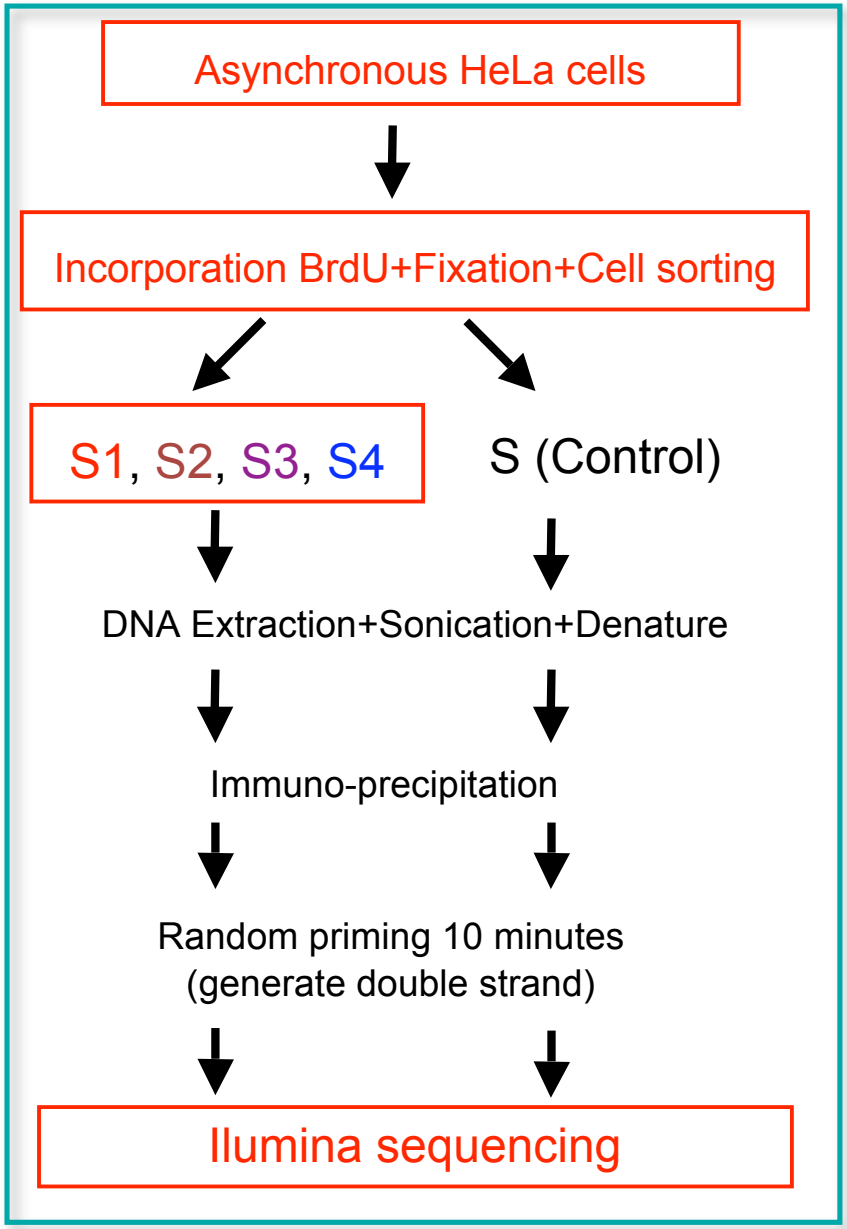
Genome-wide replication timing data are lacking !

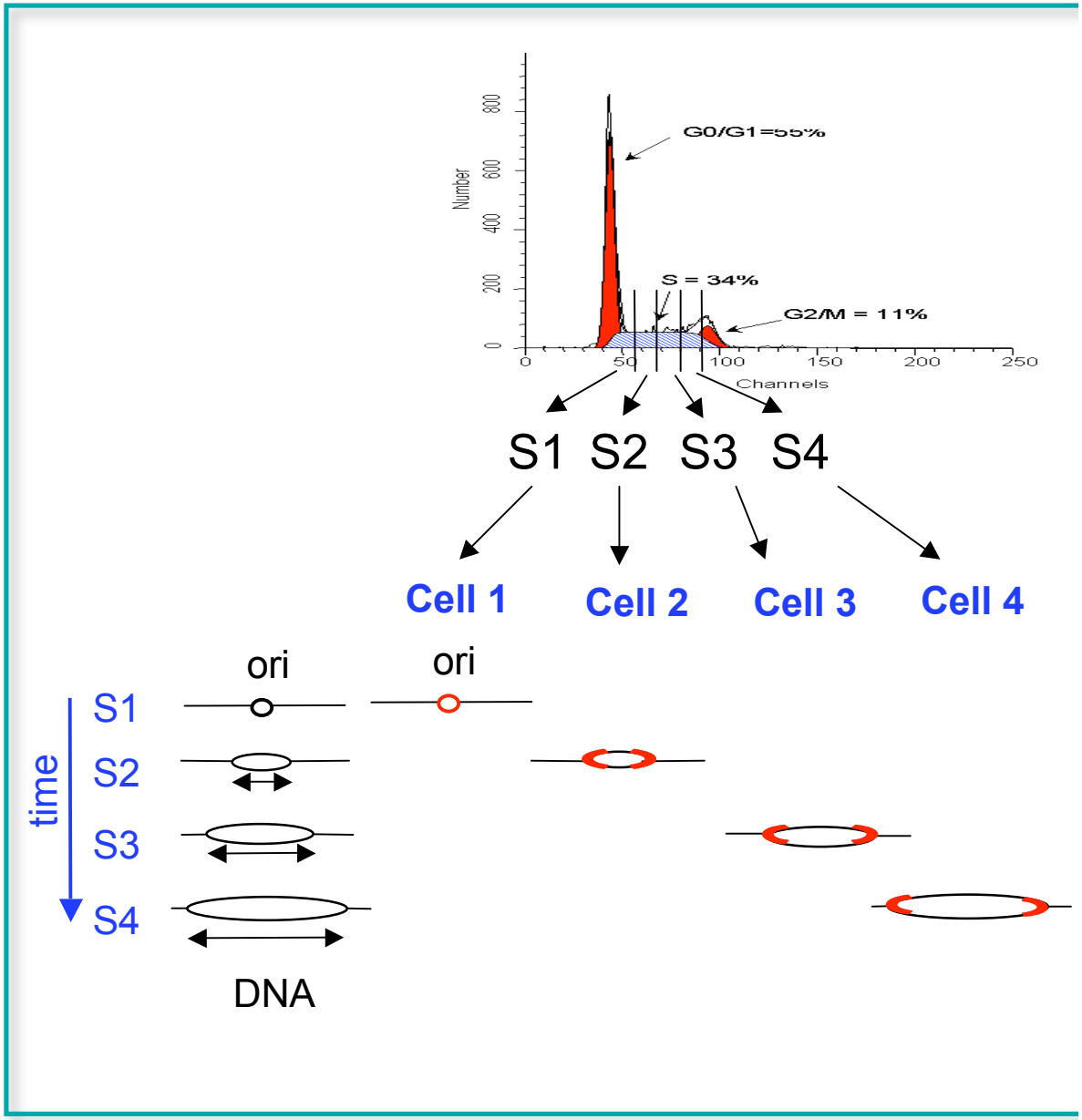
PART I

DETERMINATION OF REPLICATION TIMING PROFILE ALONG ENTIRE HUMAN GENOME

Determination of replication timing profile by massive sequencing of neo-replicated DNA

(Collaborators: *A. Rappailles, G. Guilbaud, O. Hyrien*, ENS Paris)

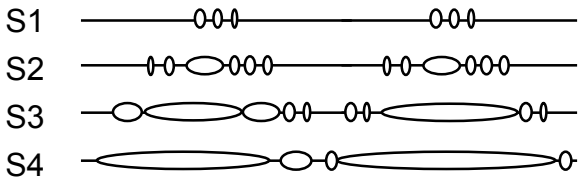




Chen CL, et al. *Genome Res.* (2010) 20:

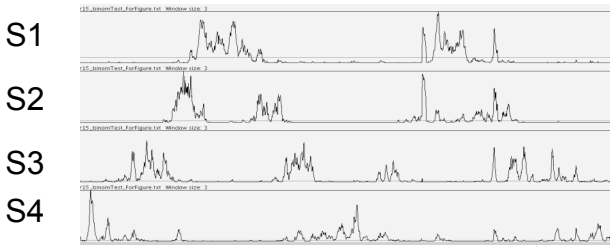
Determination of replication timing profile by massive sequencing of neo-replicated DNA

Deep sequencing
of neo-replicated DNA
cells sorted in 4 fractions of S phase

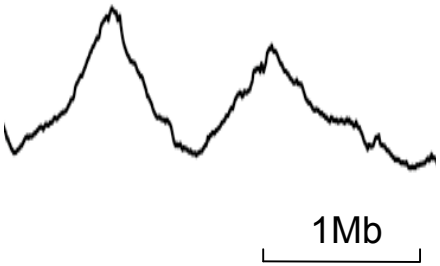


Background evaluation
Normalization

Density profiles



Replication timing profile



Early

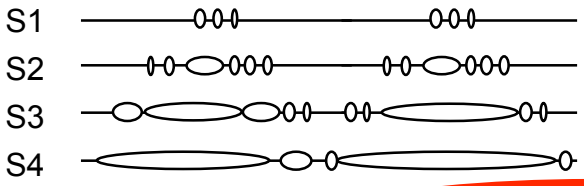
Late

One of the first high resolution replication timing profiles of the human genome

Chen CL, et al. *Genome Res.* (2010) 20:

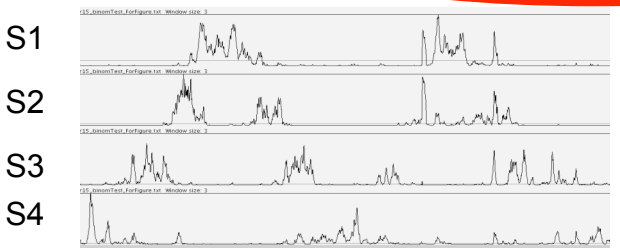
Determination of replication timing profile by massive sequencing of neo-replicated DNA

Deep sequencing
of neo-replicated DNA
cells sorted in 4 fractions of S phase

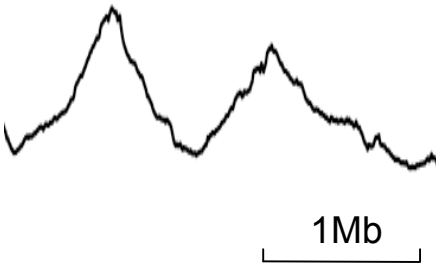


Background evaluation
Normalization

Density profiles



Replication timing profile



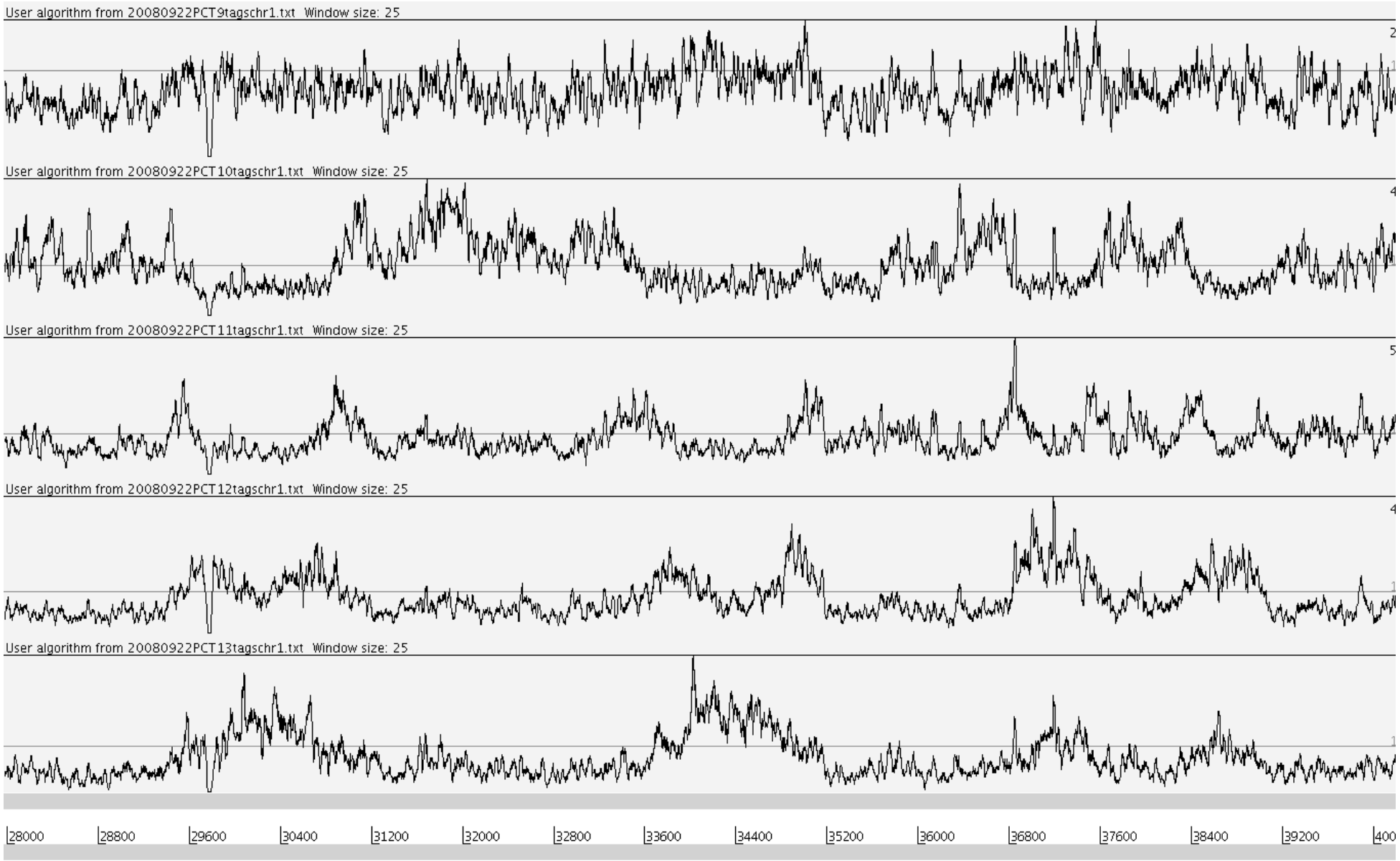
Early

Late

One of the first high resolution replication timing profiles of the human genome

Chen CL, et al. *Genome Res.* (2010) 20:

Raw tag density profiles

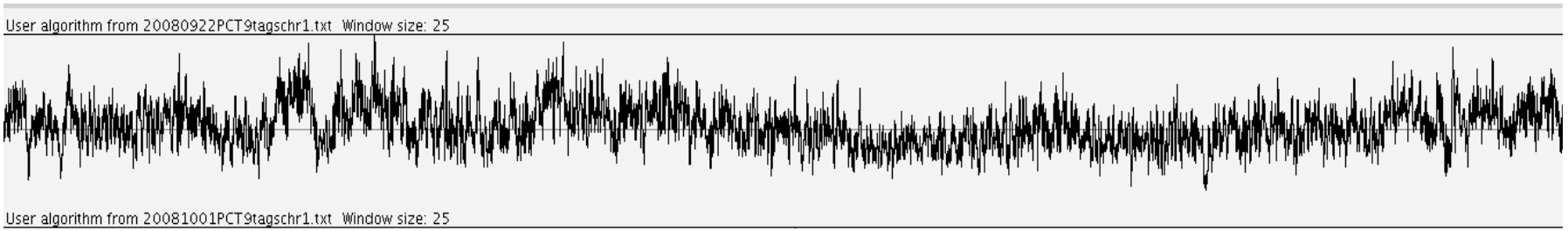


A considerable variation was observed even for the control sample

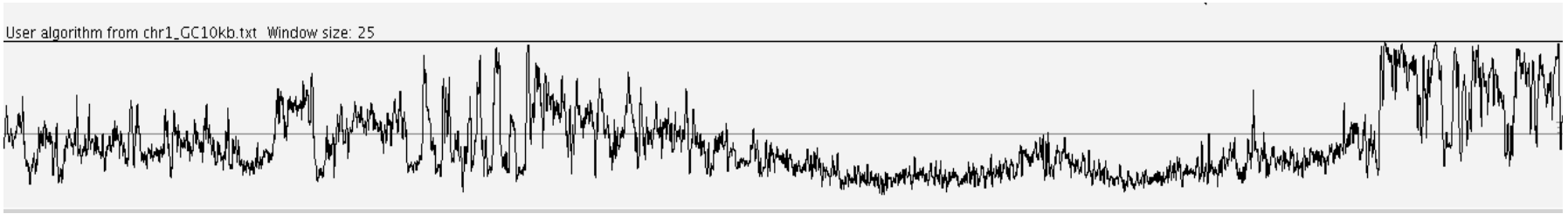
Correction of sequencing bias

Variation of tag density associated with GC content

Density profile (Control)

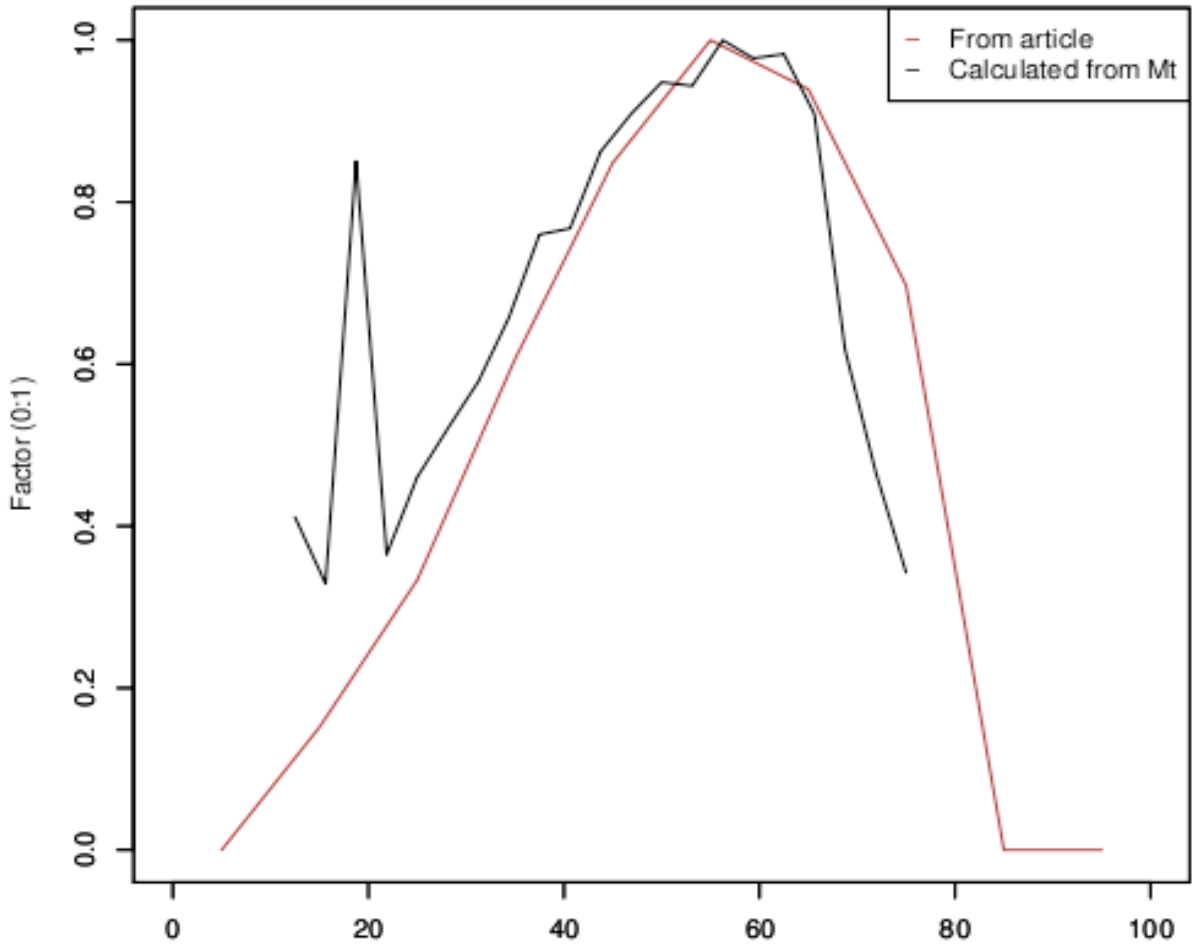


GC%



chr1(167Mb-202M

Variation of tag density at different GC% (computed in the 32nt tag)



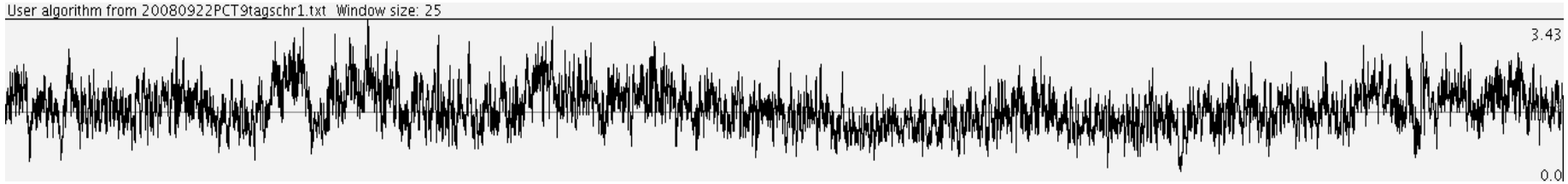
Published data

This study

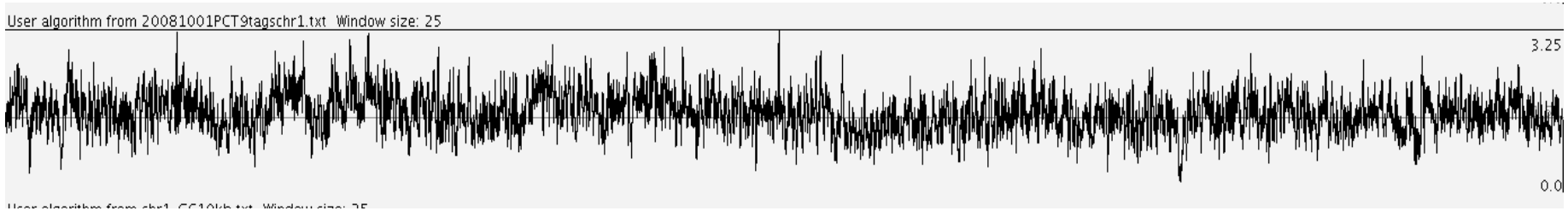
Published data from:
Hillier LW. et al. *Nat Methods.*(2008) 5:1

Profile of the tag density with correction of the GC% bias

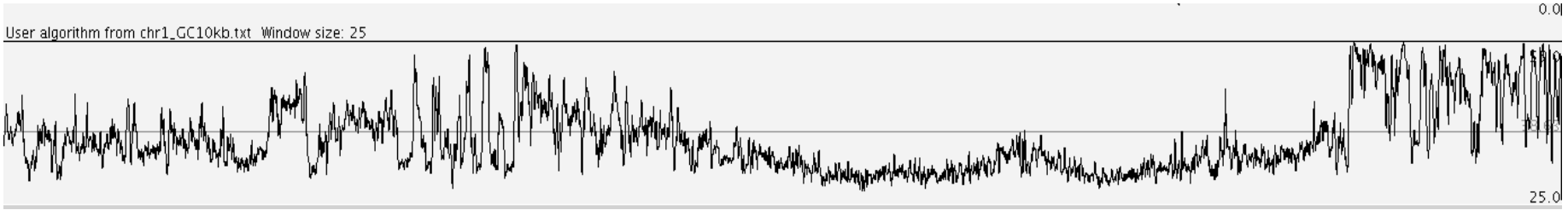
Before



After



GC%

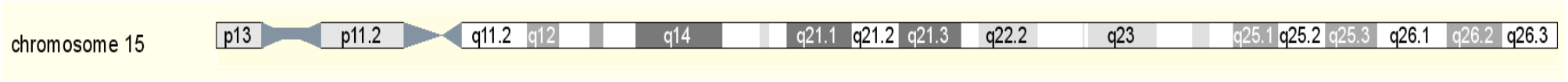


chr1(167Mb-202Mb)

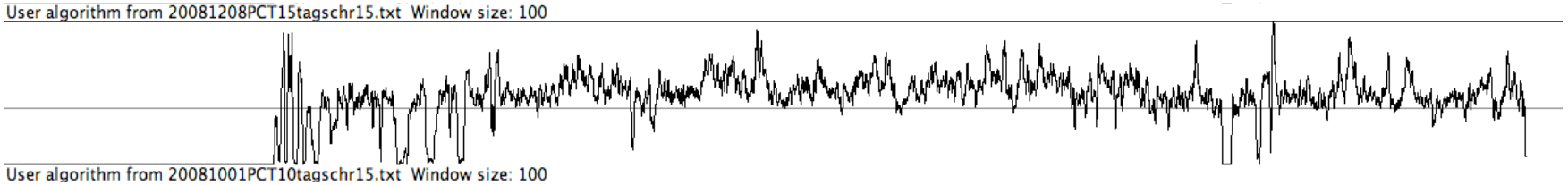
After correction, the tag density profile displays less dependent on the GC%

Data normalization

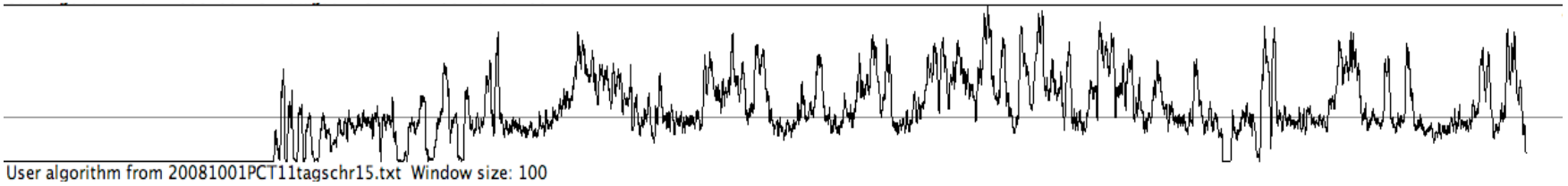
Compute the enrichment of neo-replicated DNA for each sample



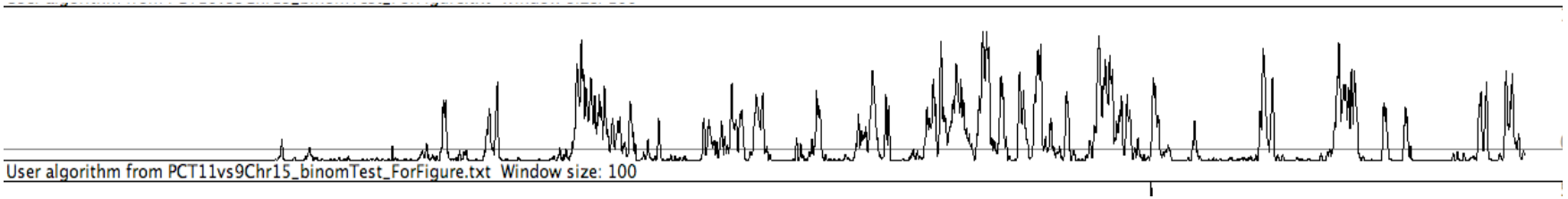
Control (S)



Signal (S1)

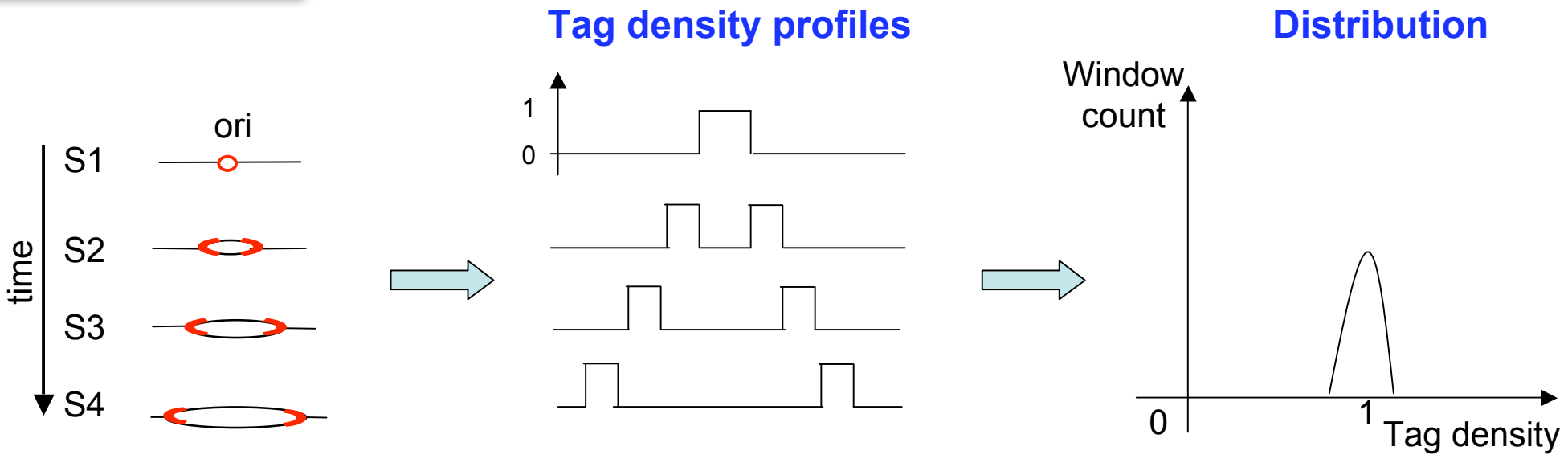


Enrichment
Signal/Control



10Mb

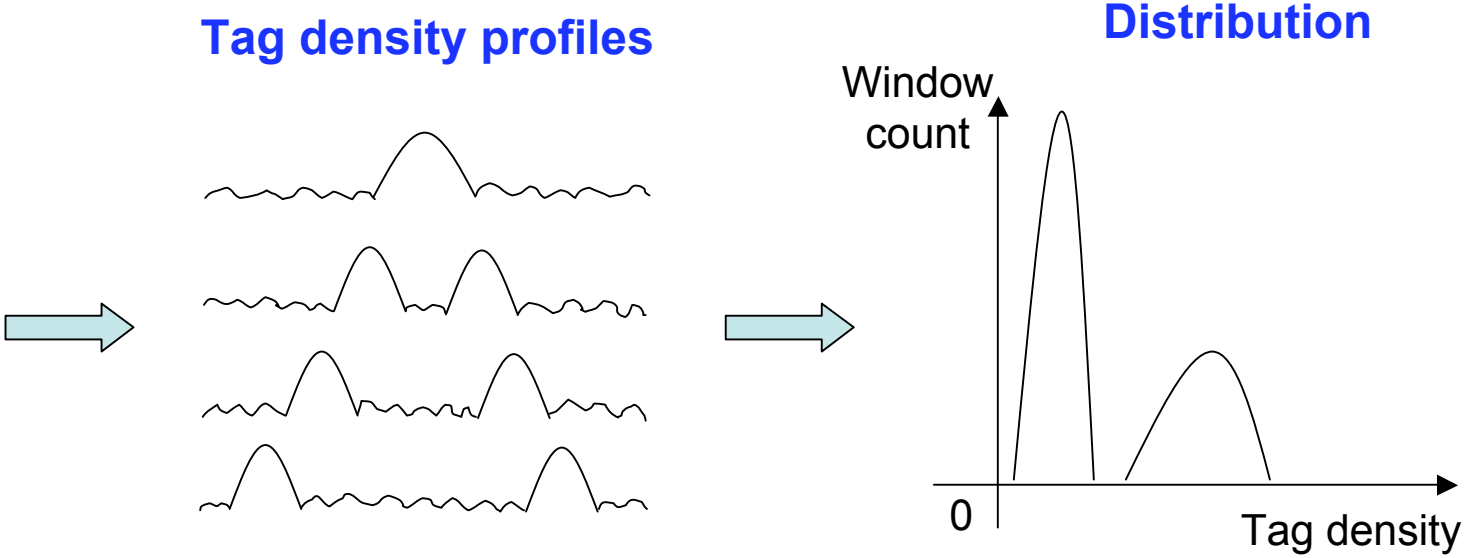
Ideal case:



Expected:

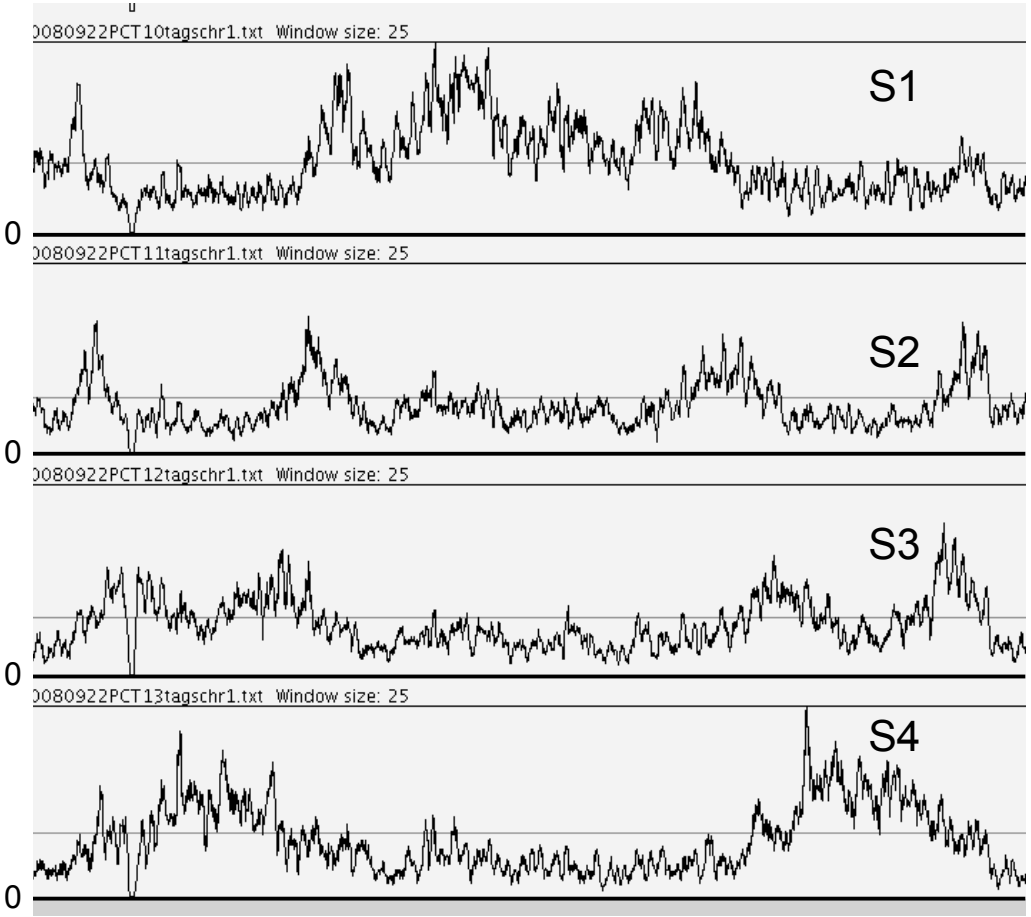
- Variation between cells
- ChIP are not 100% specific
- FACS are not 100% specific

⋮

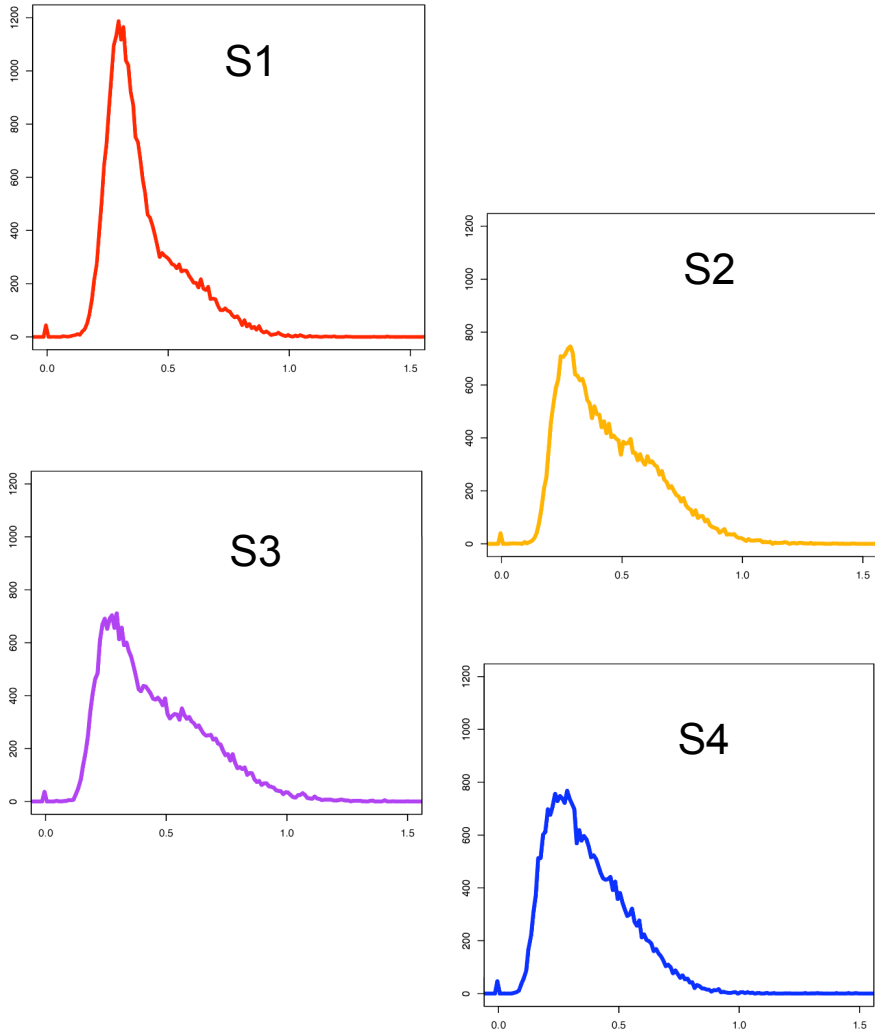


Real data:

Tag density profiles

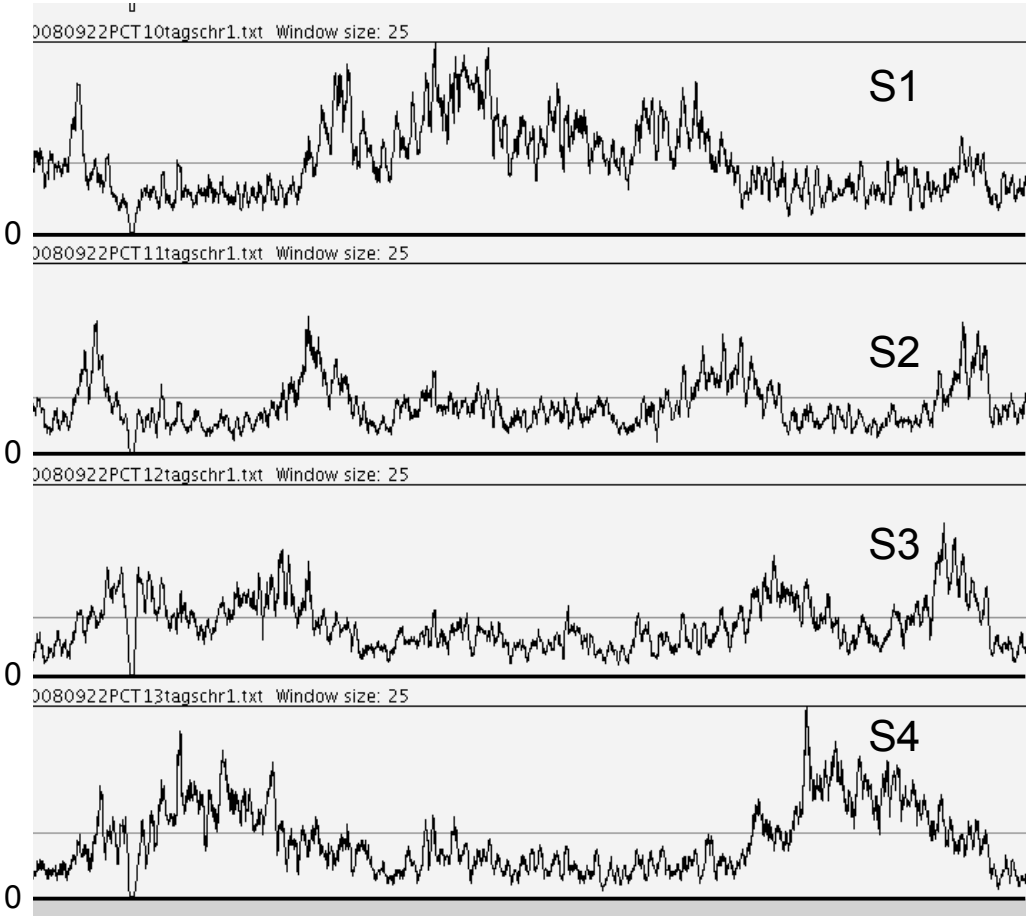


Distribution

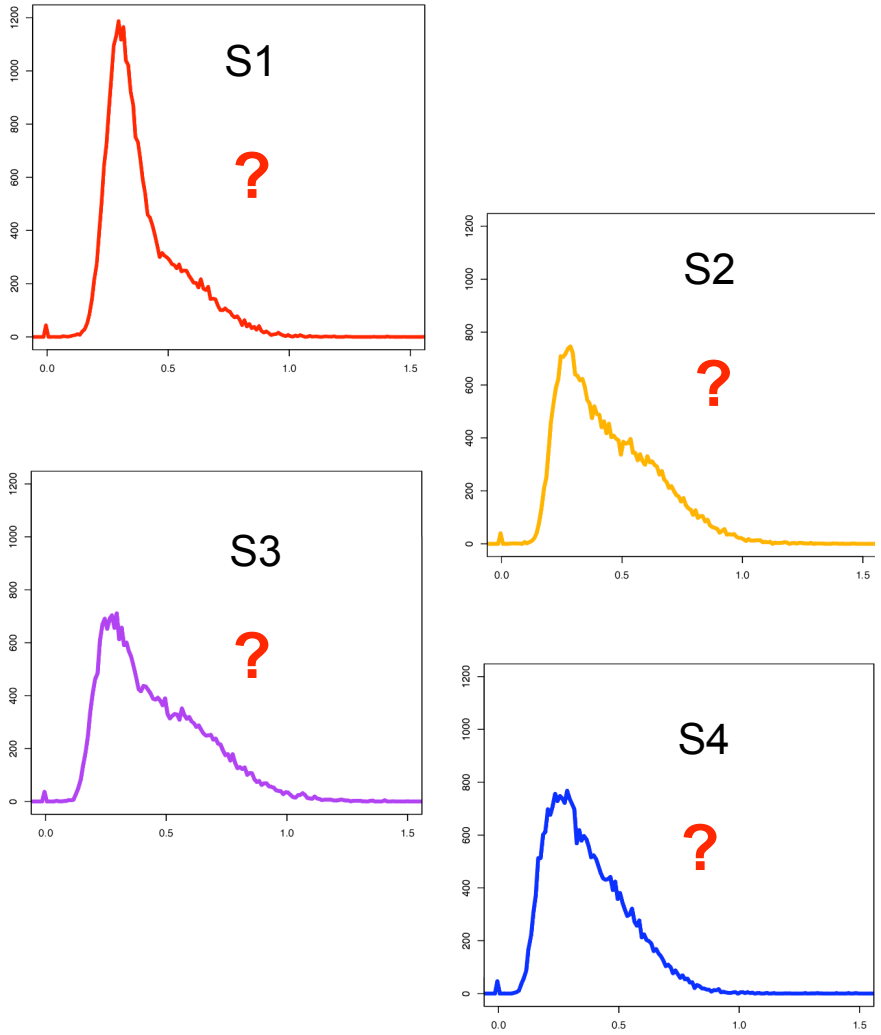


Real data:

Tag density profiles

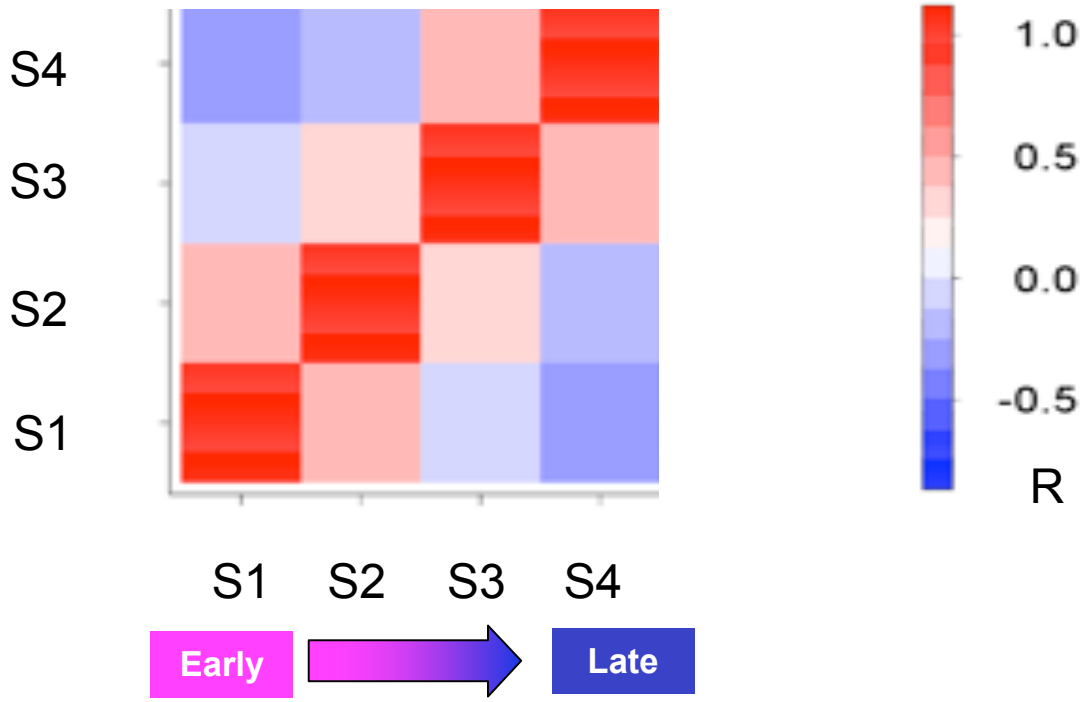


Distribution



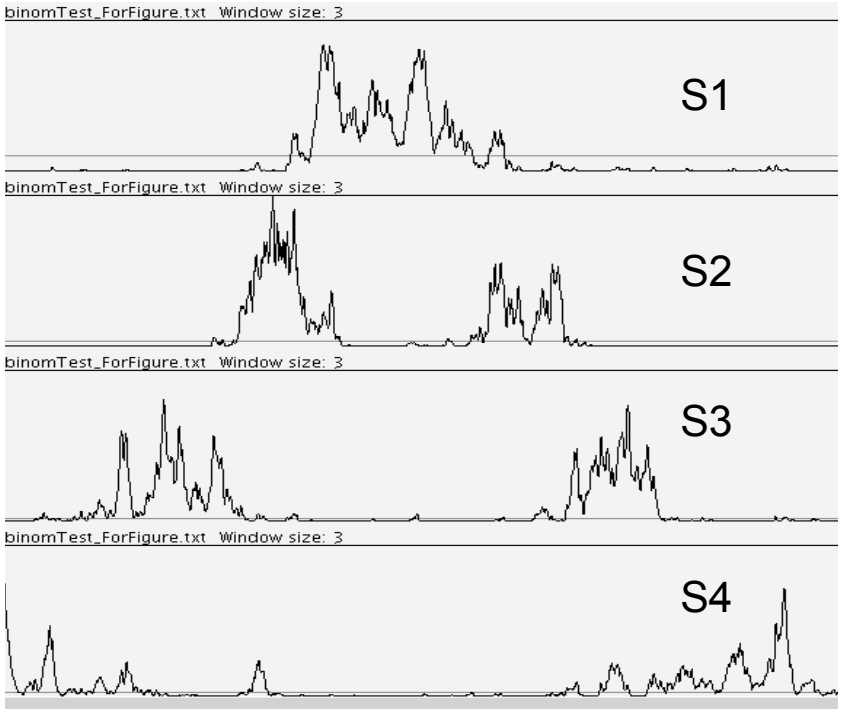
How to separate the signal from the background for each sample?

Genome wide pairwise correlations between S_i samples



Positive correlations are observed only between neighboring S_i fractions

Process to estimate the background values

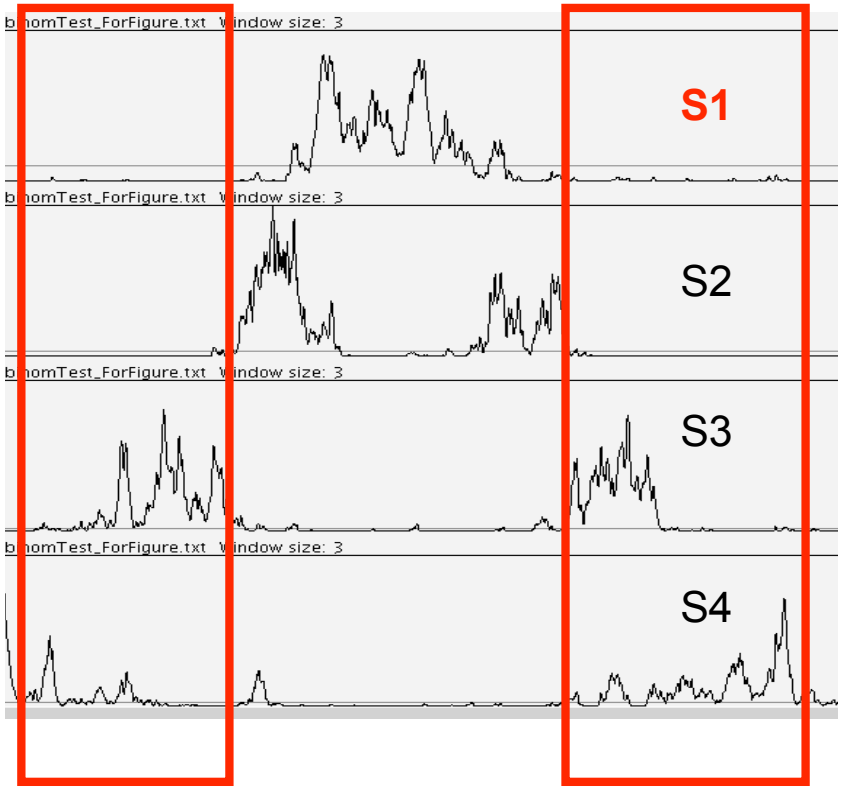


Initial background values

↓ *Probability test Si vs S*

Probability enrichment profiles

Process to estimate the background values



Background regions of S1:
significant enrichment in S3 or S4 but not S2

Initial background values

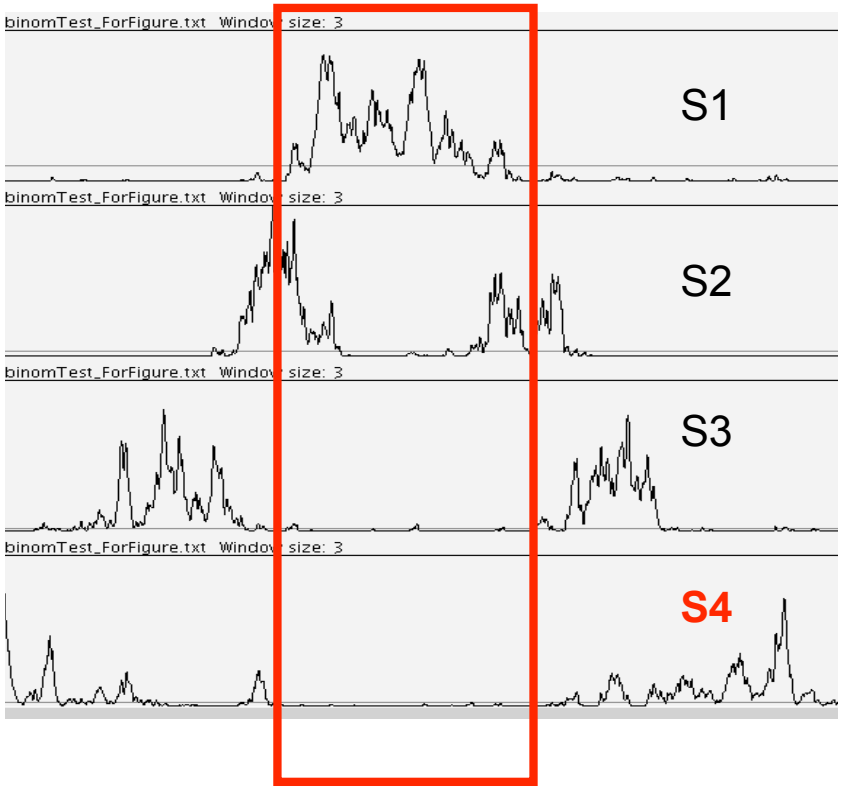
↓ *Probability test Si vs S*

Probability enrichment profiles

↓

Define the background regions

Process to estimate the background values



Background regions of S4:
significant enrichment in S1 or S2 but not S3

Initial background values

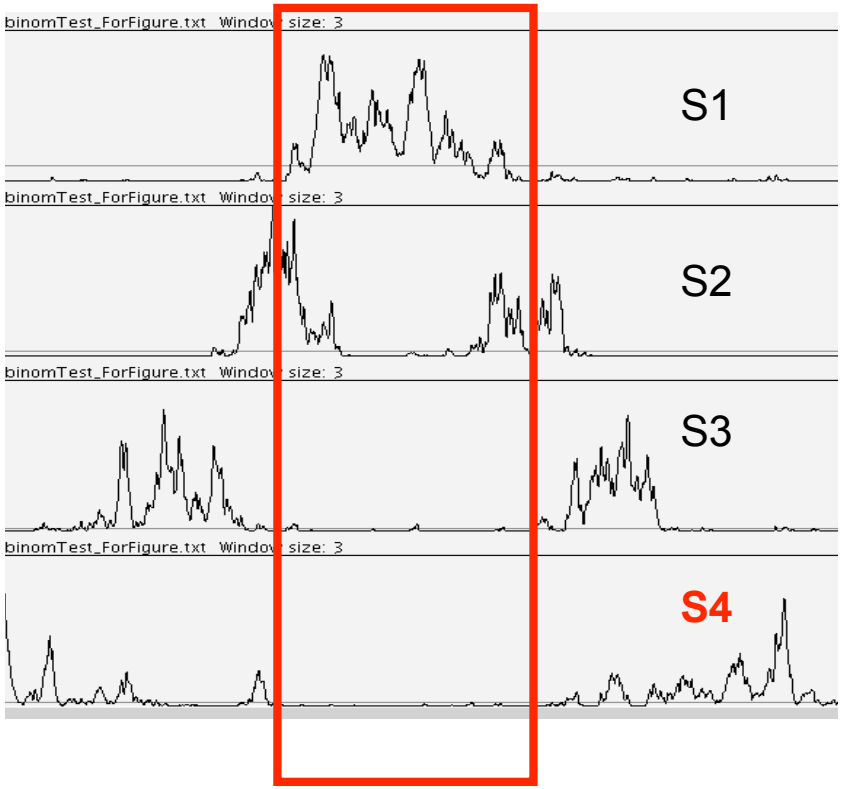
↓ *Probability test Si vs S*

Probability enrichment profiles

↓

Define the background regions

Process to estimate the background values



Background regions of S4:
significant enrichment in S1 or S2 but not S3

Initial background values

↓ *Probability test Si vs S*

Probability enrichment profiles

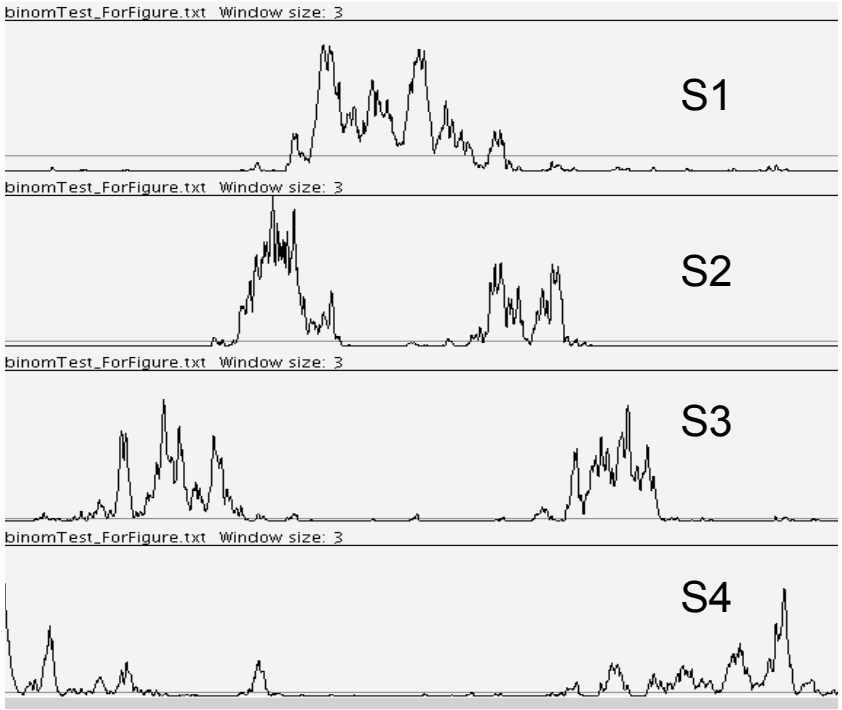
↓

Define the background regions

↓

Compute the new background values

Process to estimate the background values



Initial background values

Probability test Si vs S

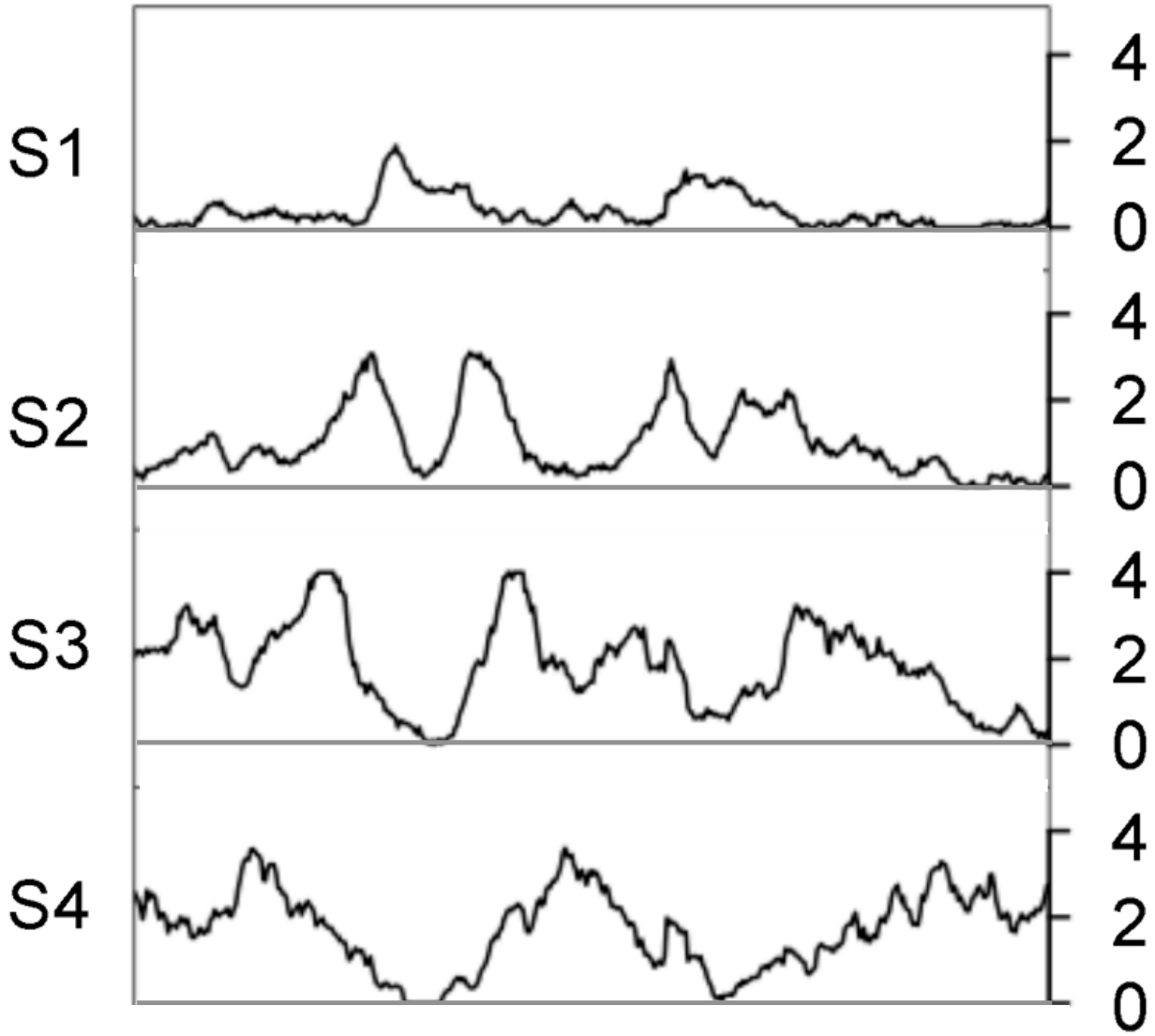
Probability enrichment profiles

Define the background regions

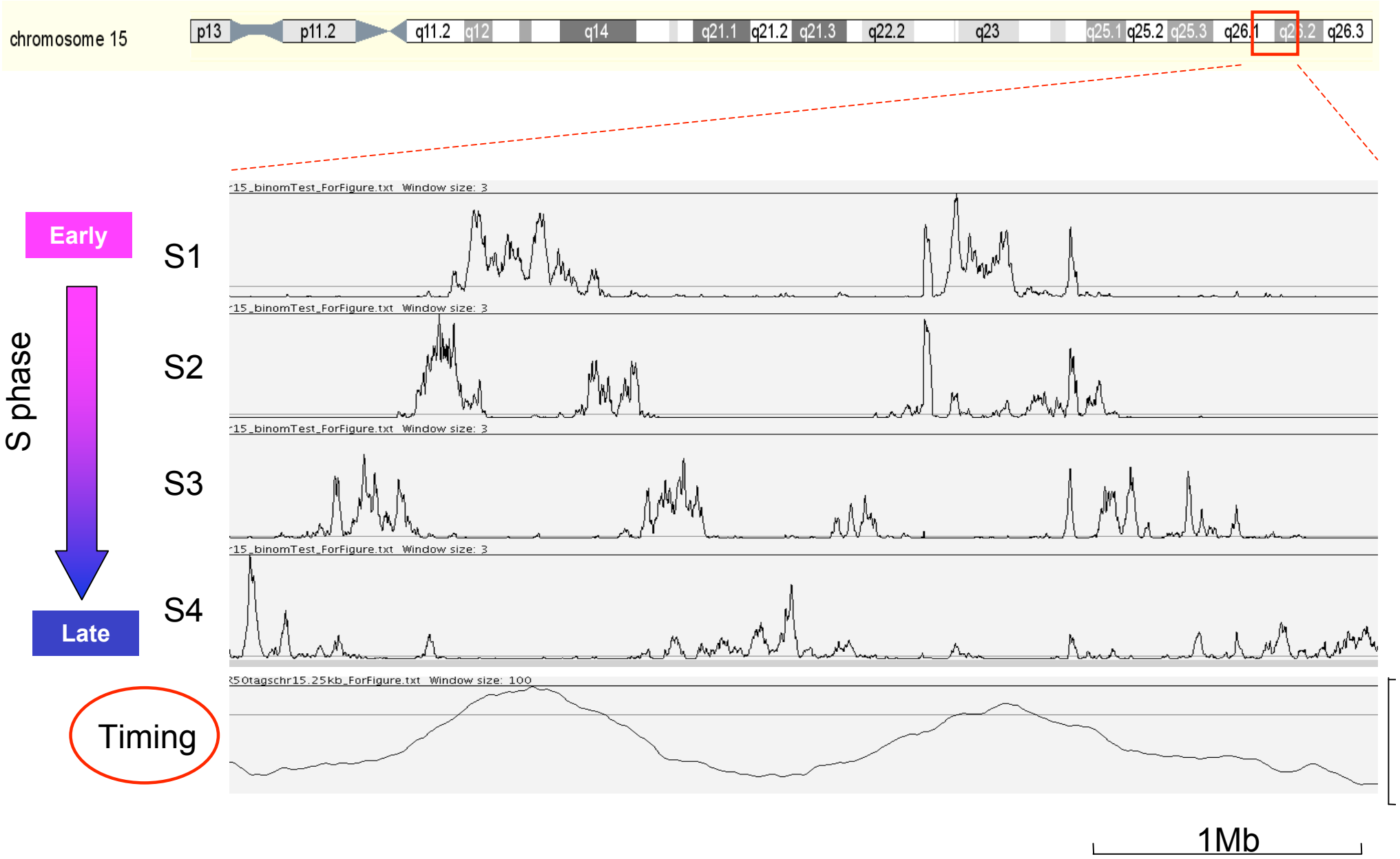
Compute the new background values

Repeat the processes up to get stable values for normalization.

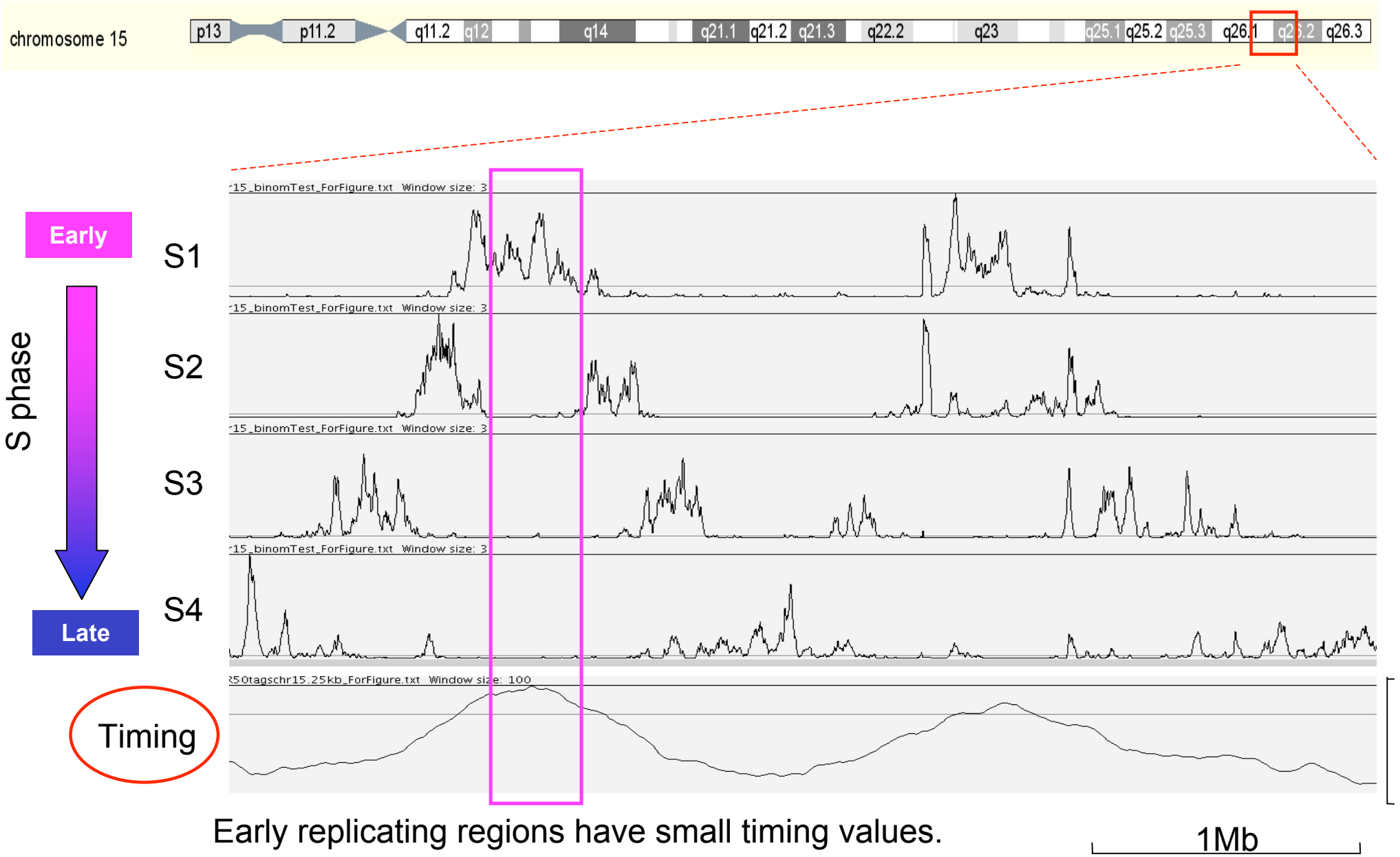
Results: Tag density profiles after normalization



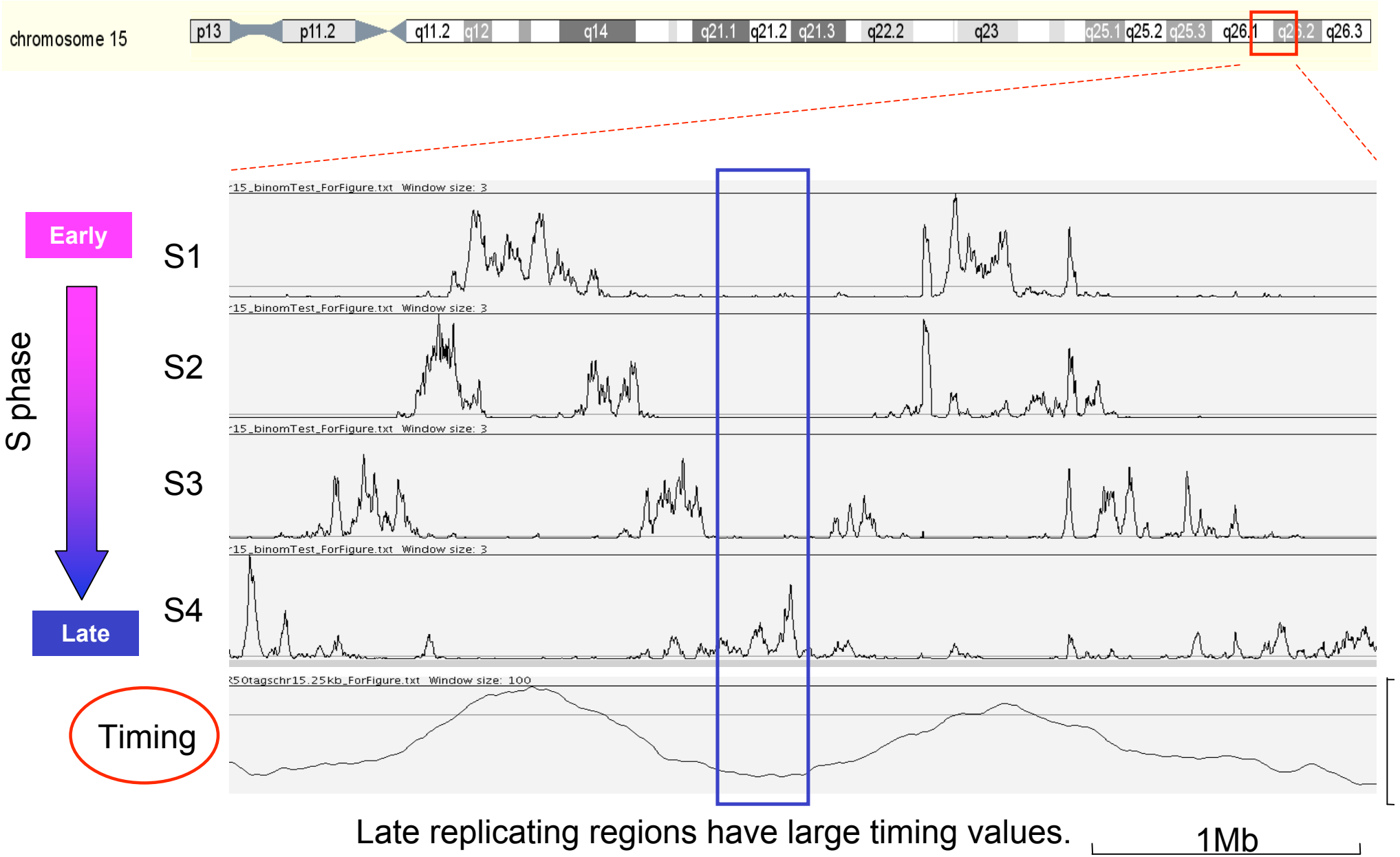
Computation of S50 (time for 50% replication) for each defined window

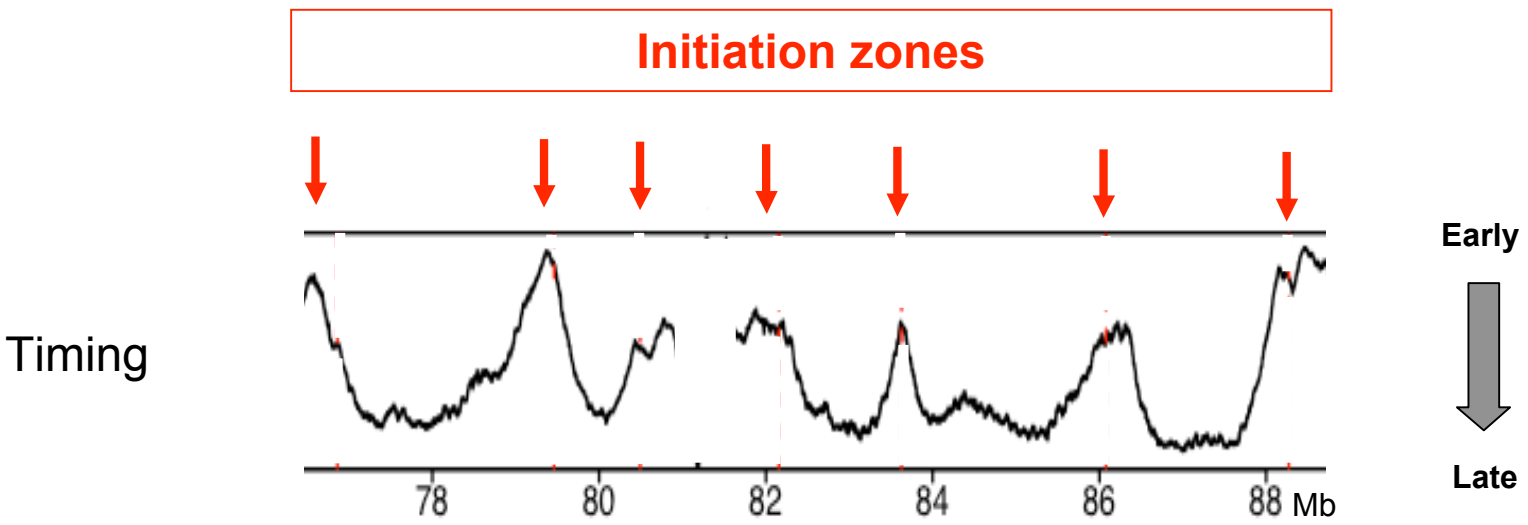


Computation of S50 (time for 50% replication) for each defined window



Computation of S50 (time for 50% replication) for each defined window

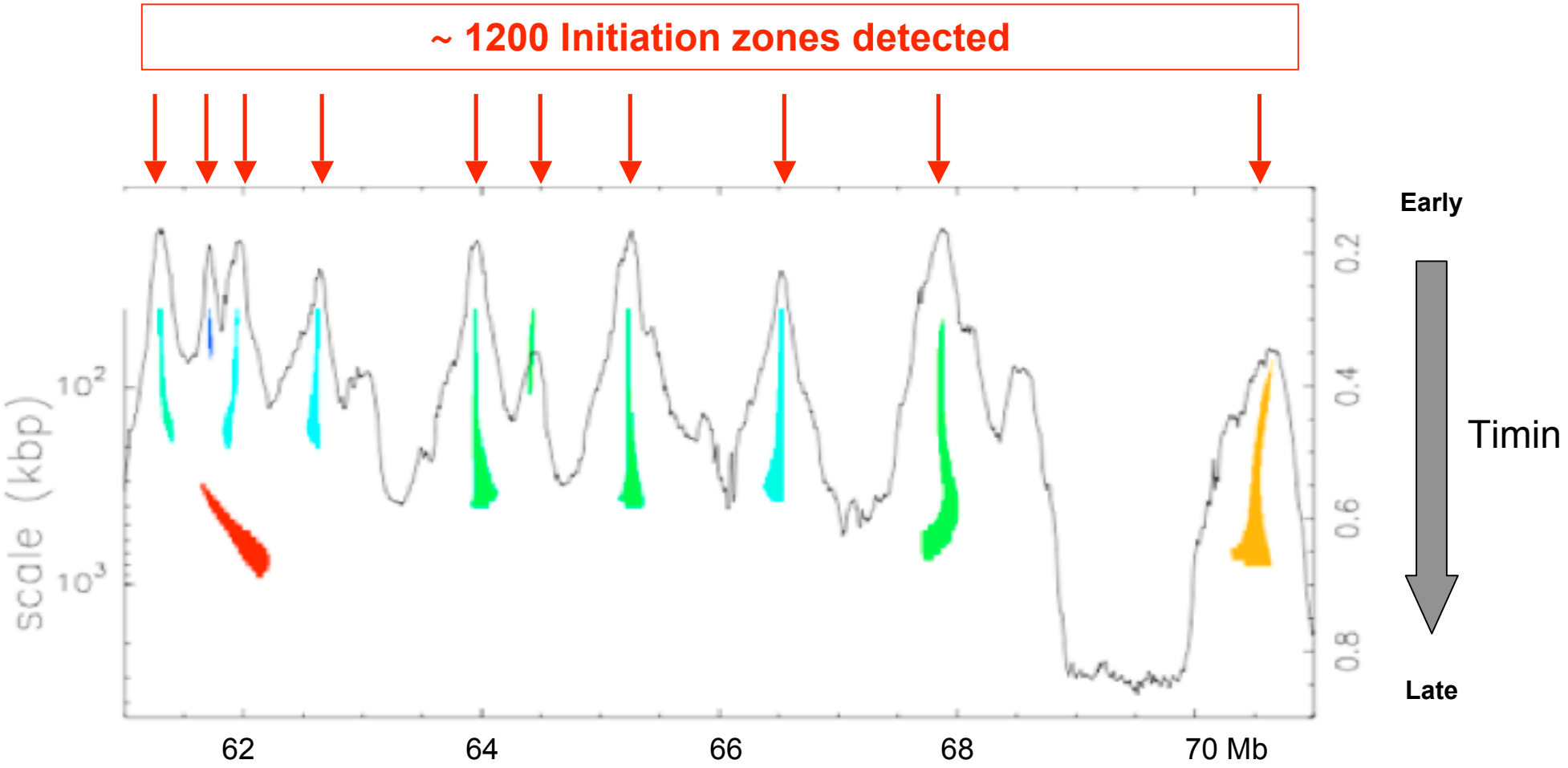




Detect the peaks in the timing profile to obtain the replication initiation zones

Detection of replication initiation zones

Collaborators: *B. Audit and A. Arneodo* (ENS Lyon)

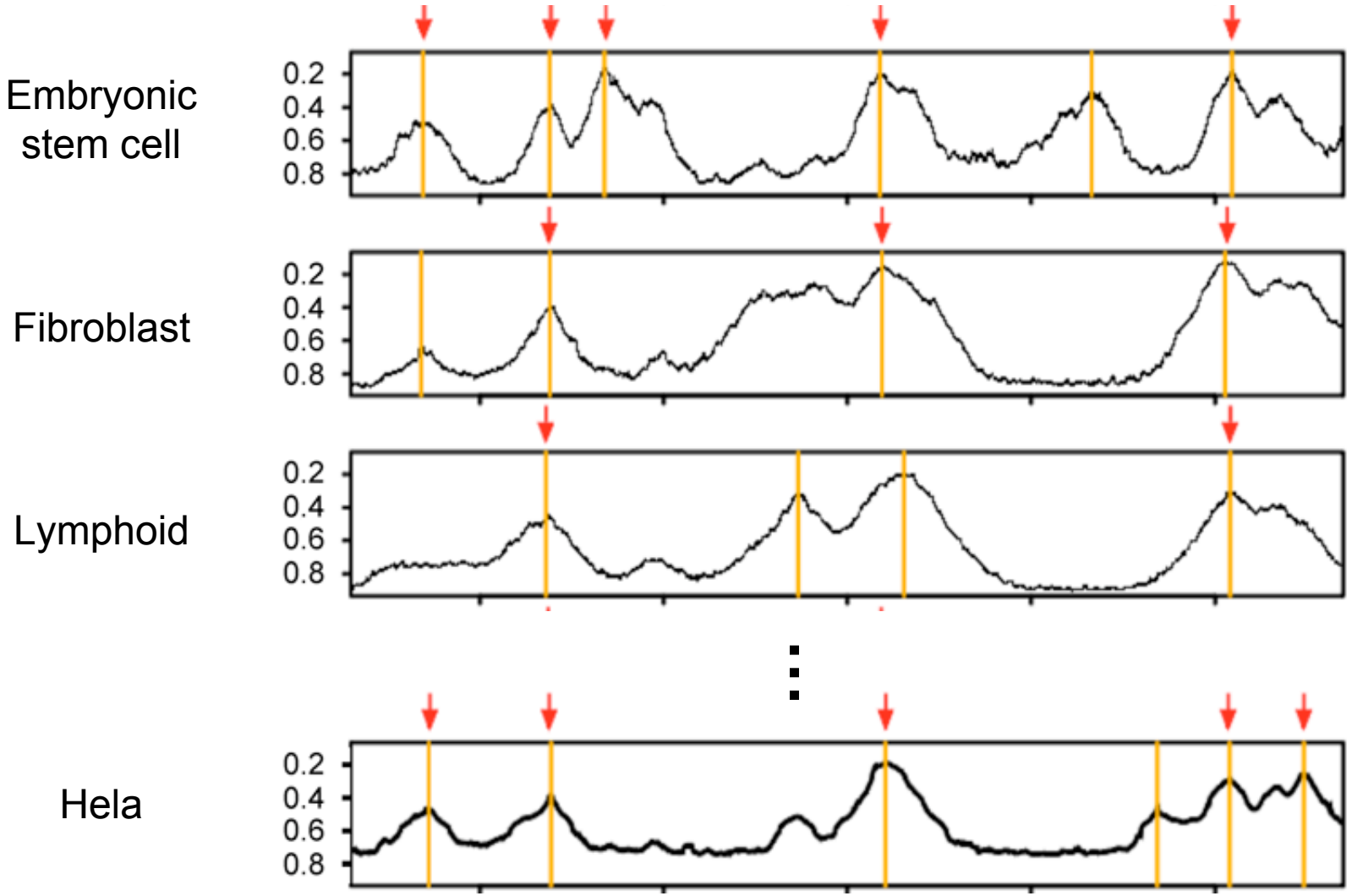


Wavelet-based method (multiple-scale analysis)

CHEN CL. et al. *Mol. Biol. Evol.* (2011) 28:2

Are these replication initiation zones conserved amongst different cell types?

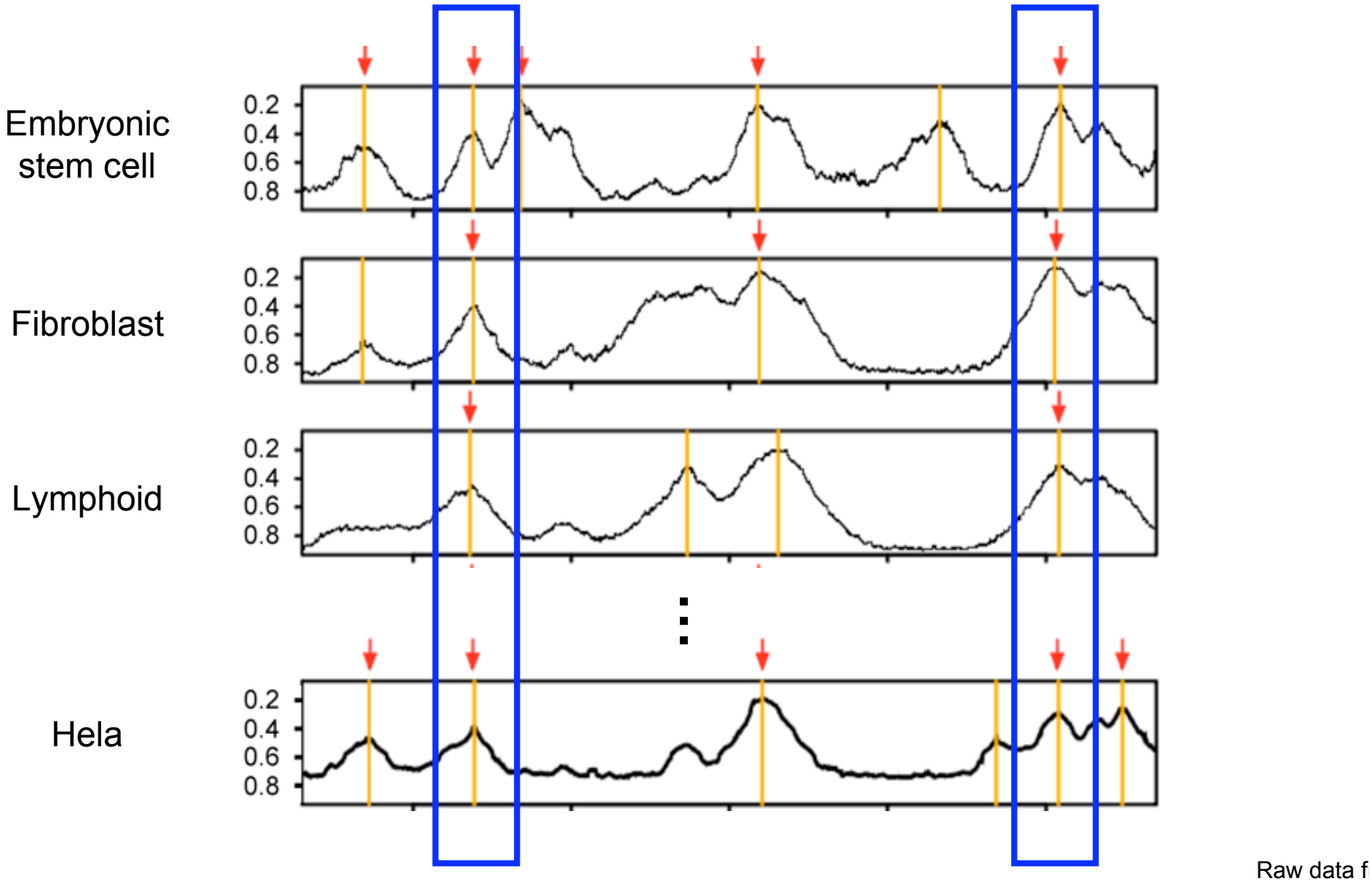
Comparison of initiation zones of different cell types



Raw data f

Hansen RS. et al. *PNAS*. (2010) 107
Desprat R. et al. *Genome Res*. (2009) 19:2

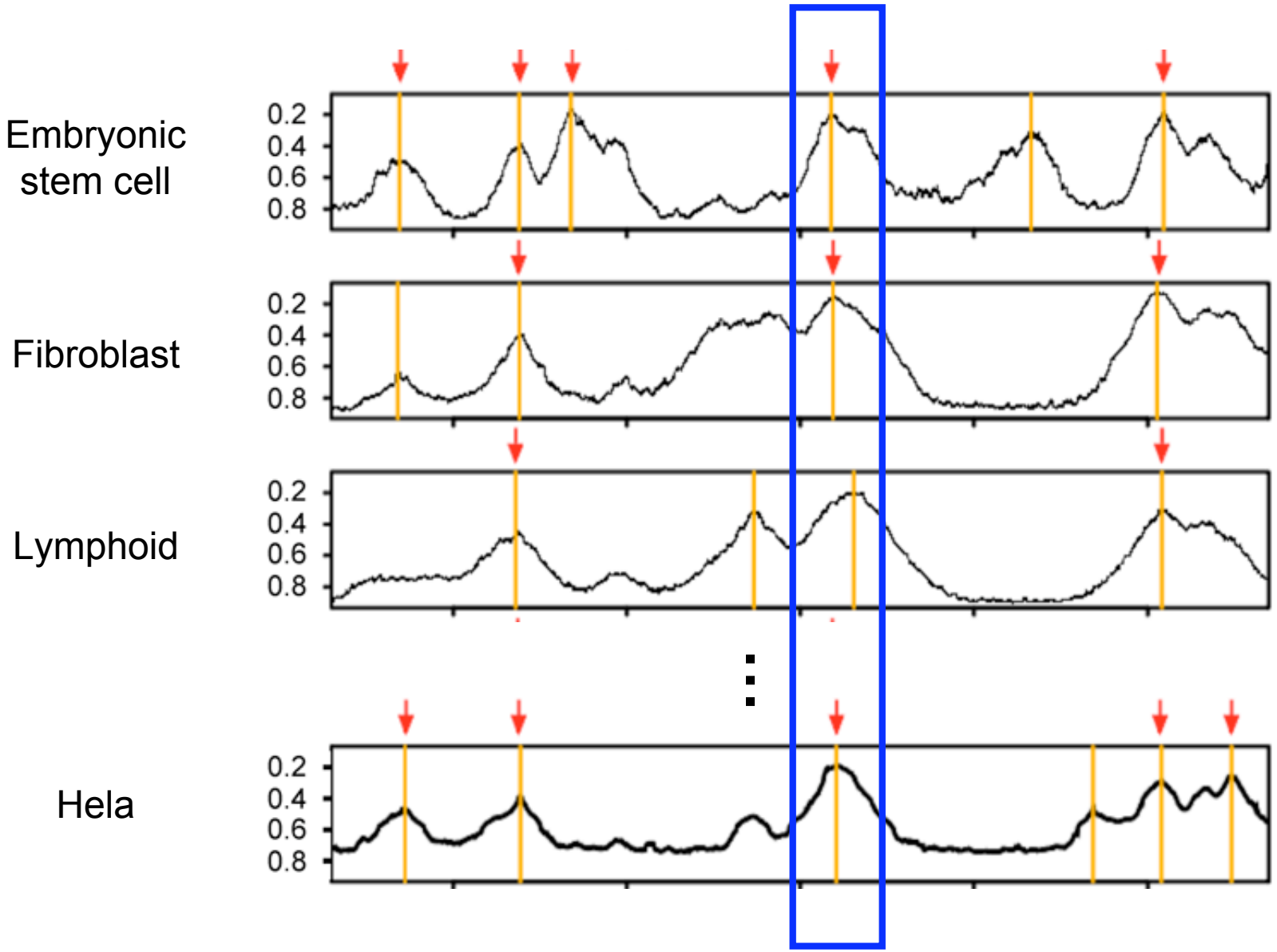
Comparison of initiation zones of different cell types



Raw data f

Hansen RS. et al. *PNAS*. (2010) 107
Desprat R. et al. *Genome Res*. (2009) 19:2

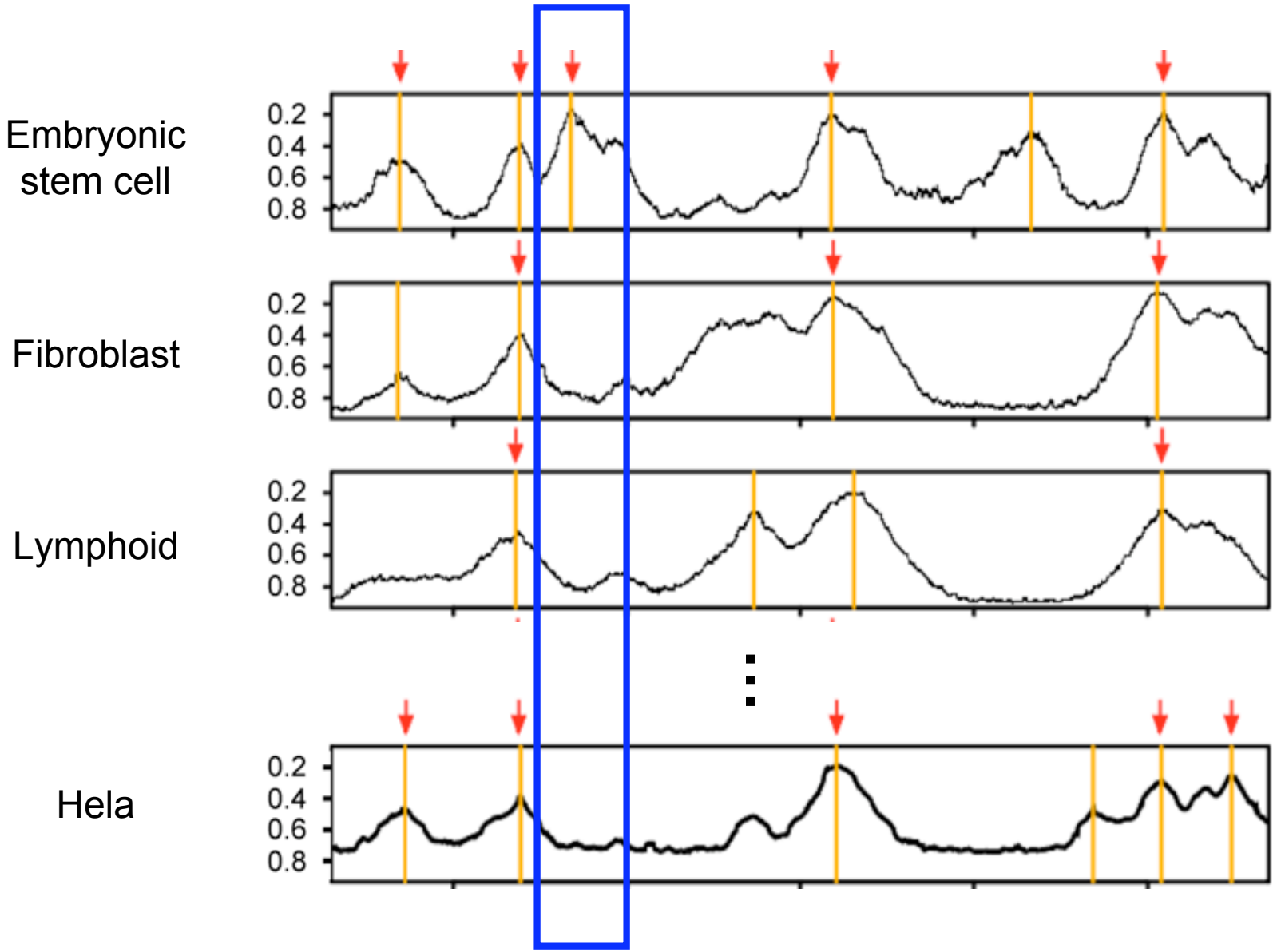
Comparison of initiation zones of different cell types



Raw data f

Hansen RS. et al. *PNAS*. (2010) 107
Desprat R. et al. *Genome Res*. (2009) 19:2

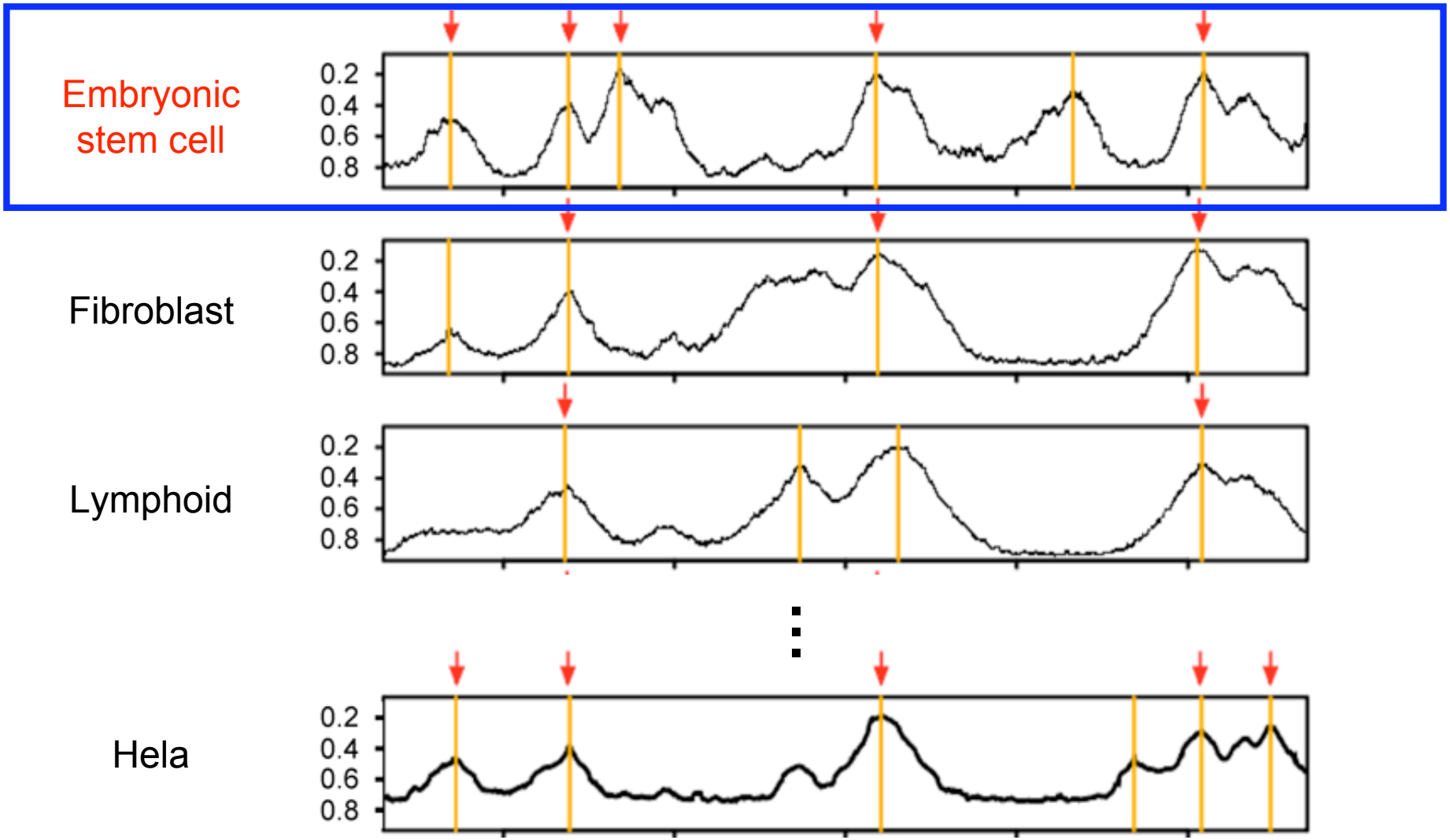
Comparison of initiation zones of different cell types



Raw data f

Hansen RS. et al. *PNAS*. (2010) 107
Desprat R. et al. *Genome Res*. (2009) 19:2

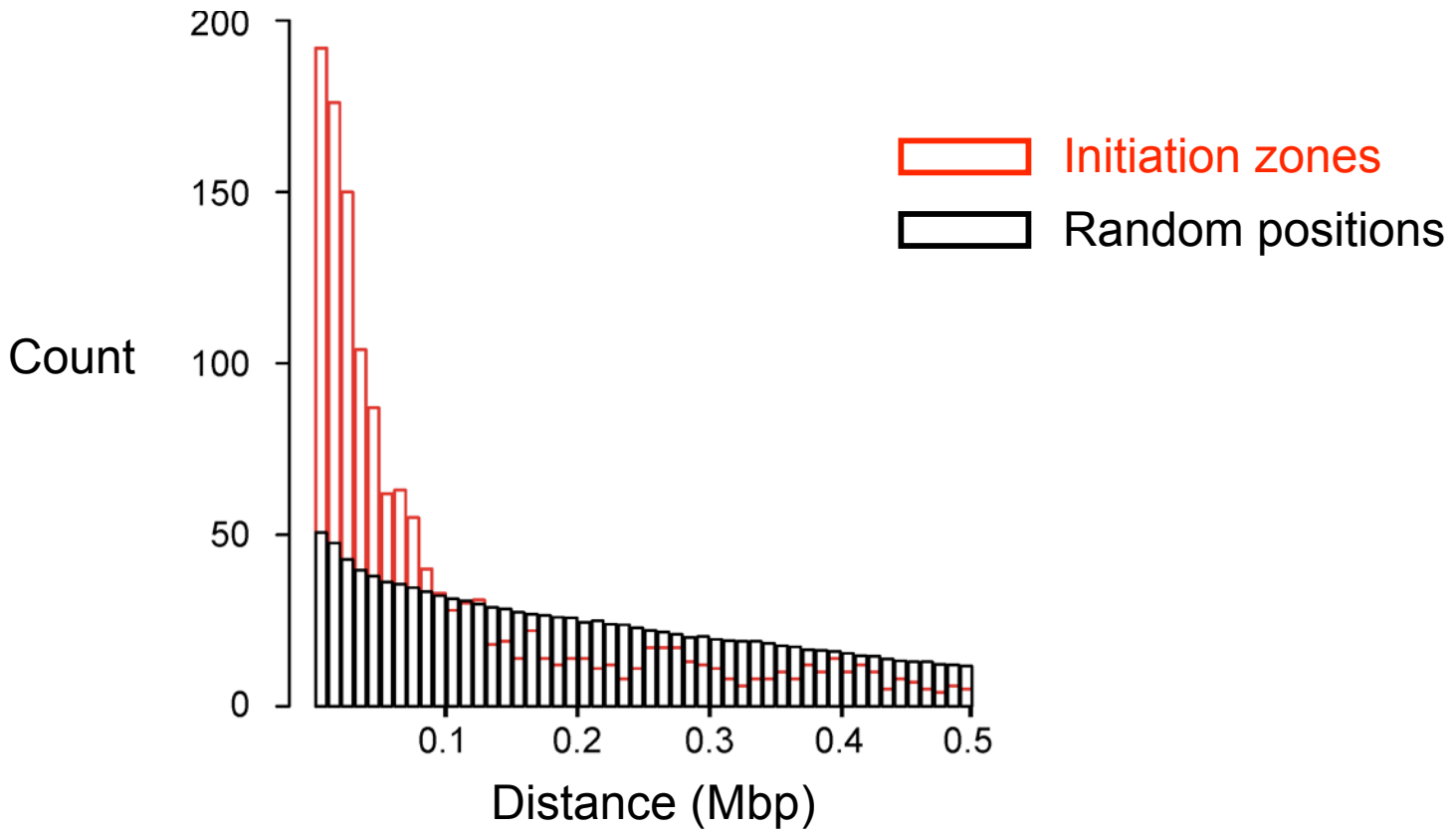
Statistical evaluation : comparison with null distribution of simulation



Raw data f

Hansen RS. et al. *PNAS*. (2010) 107
Desprat R. et al. *Genome Res*. (2009) 19:2

Statistical evaluation : comparison with null distribution of simulation



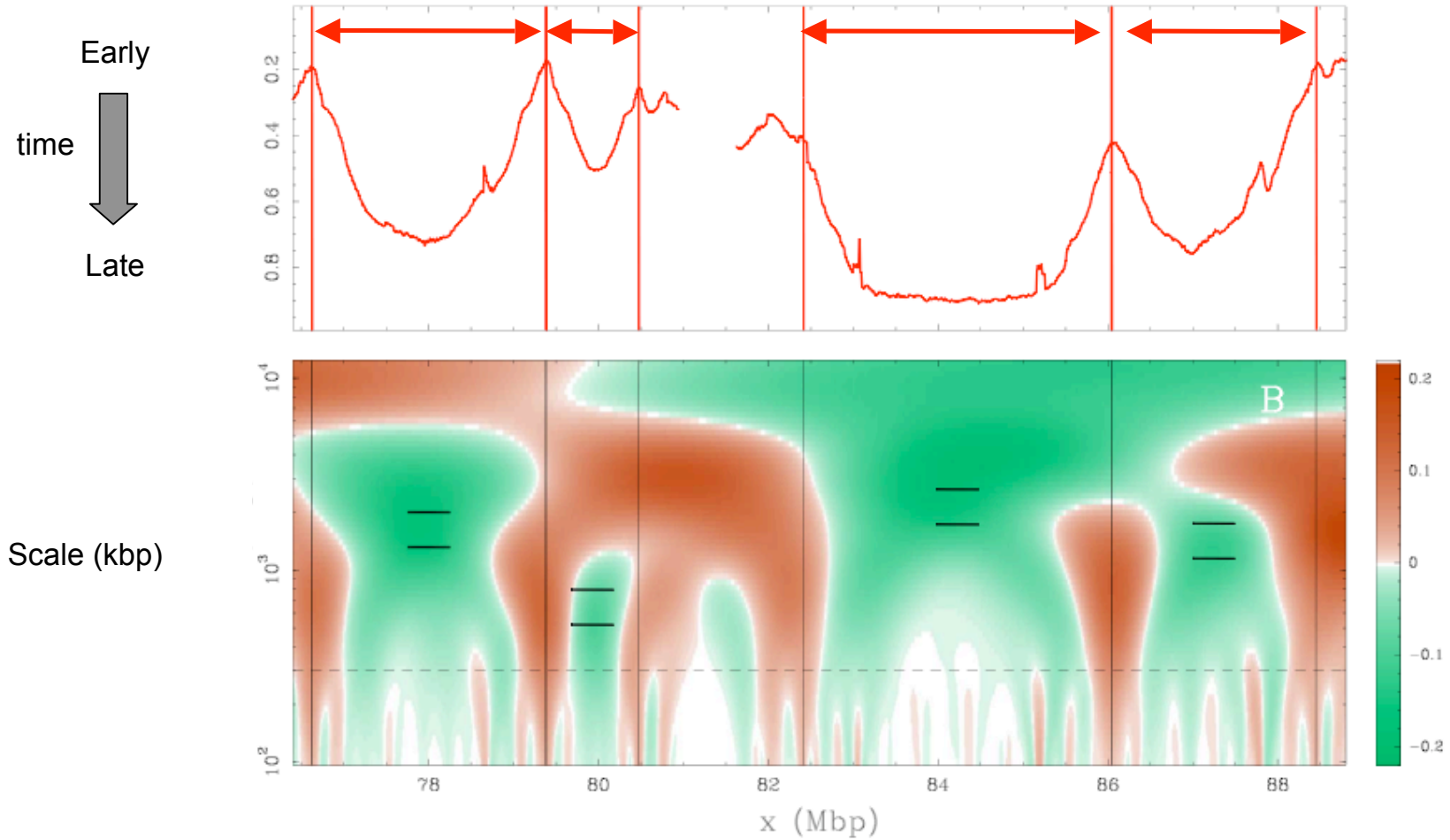
Replication initiation zones of one cell type are significantly associated with replication initiation zones of other cell types

CHEN CL et al. *Mol. Biol. Evol.* (2011) 28:23

Detection of U-shaped domains in replication timing profile

A. Baker, B. Audit, A. Arneodo

U-shaped timing domains

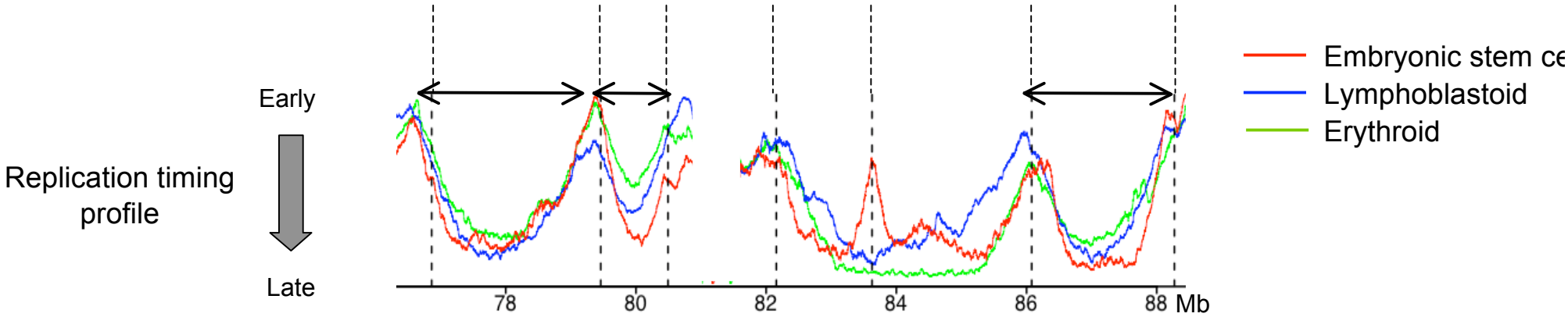


Wavelet-based method

Detect >1000 U-shaped timing domains (average size ~1Mb)
covering half of the genome in each of 7 examined cell types

Audit A. et al. *Nat. Protoc.* In pr

A larger fraction of replication domains are conserved in different cell types



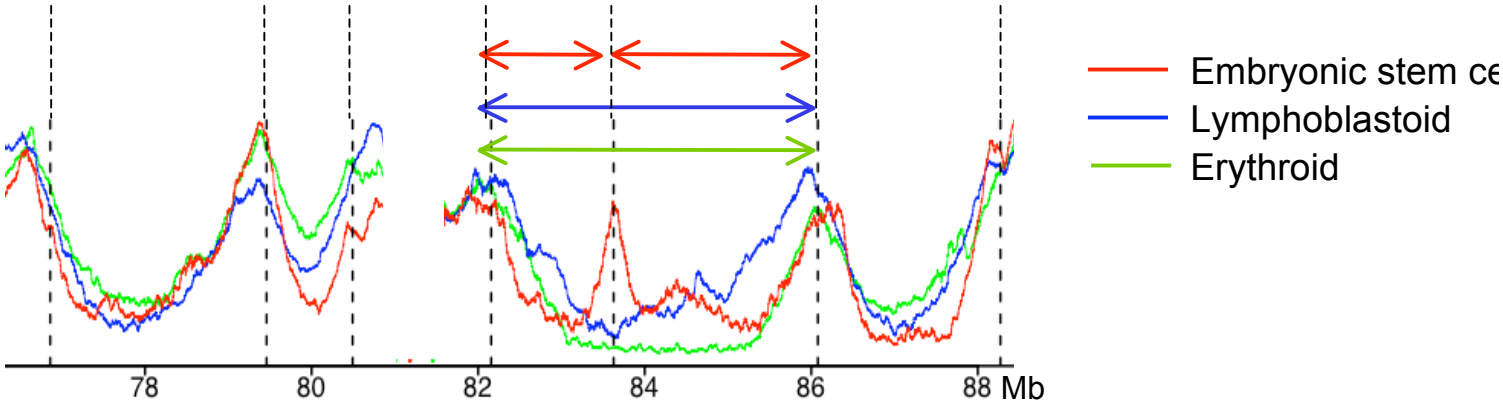
U-shaped timing domains

But also vary in some cases

Replication timing profile

Early

Late



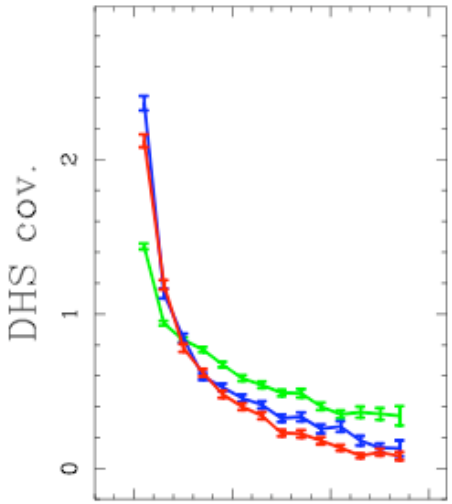
U-shaped timing domains

How is this replication program controlled?

Higher order chromatin structure?

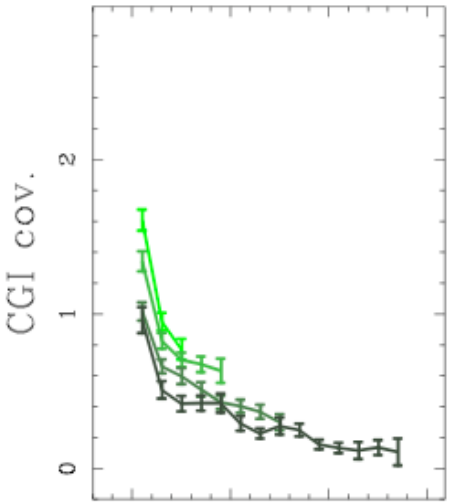
Replication domain corresponds to specific chromatin organization

HYPERSENSITIVITY TO DNASE I



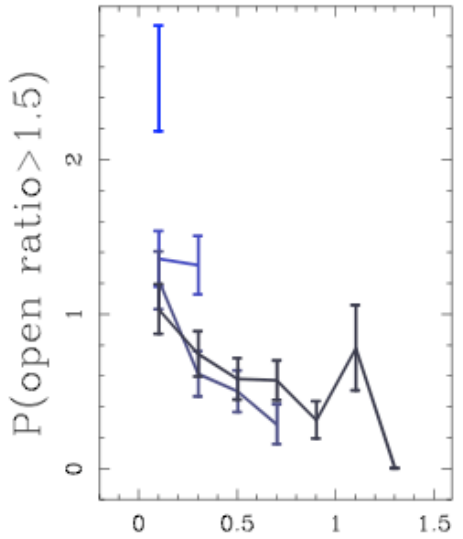
Distance to U-domain borders (Mbp)

CpG ISLAND



Distance to U-domain borders (Mbp)

OPEN CHROMATIN



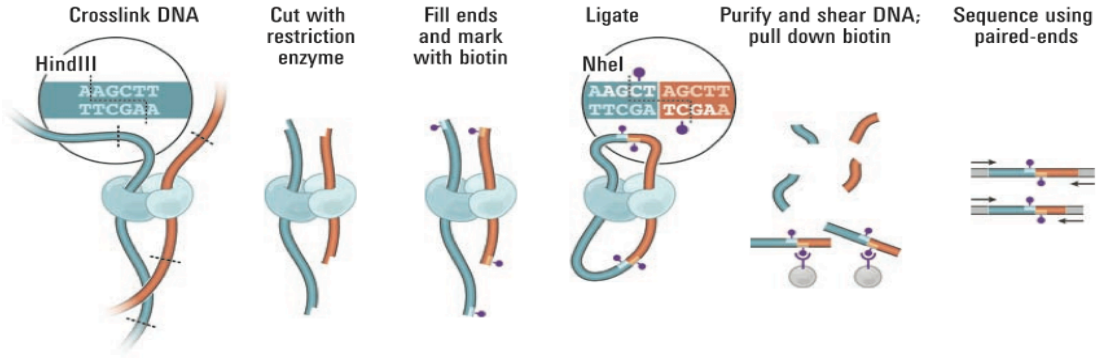
Distance to U-domain borders (Mbp)

Gradient of open chromatin structure along replication U-domains

Baker A et al. *PLoS. Comput. Biol*. (2012) 8:e10024

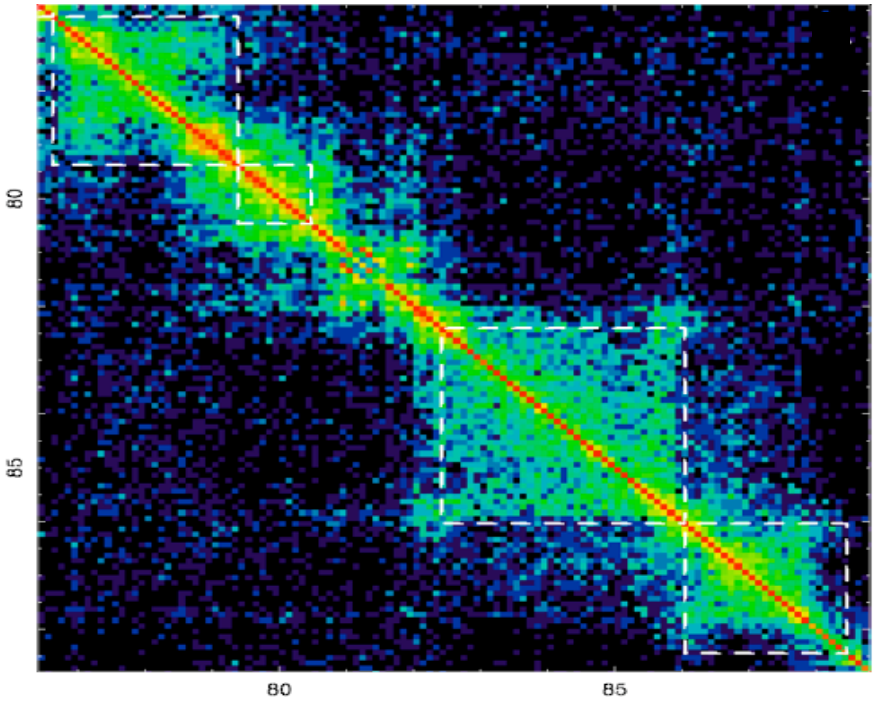
Replication domain corresponds to specific chromatin organization

Capturing
Chromatin
Conformation



Chromatin interaction (Hi-C) matrix

(spatial proximity map)

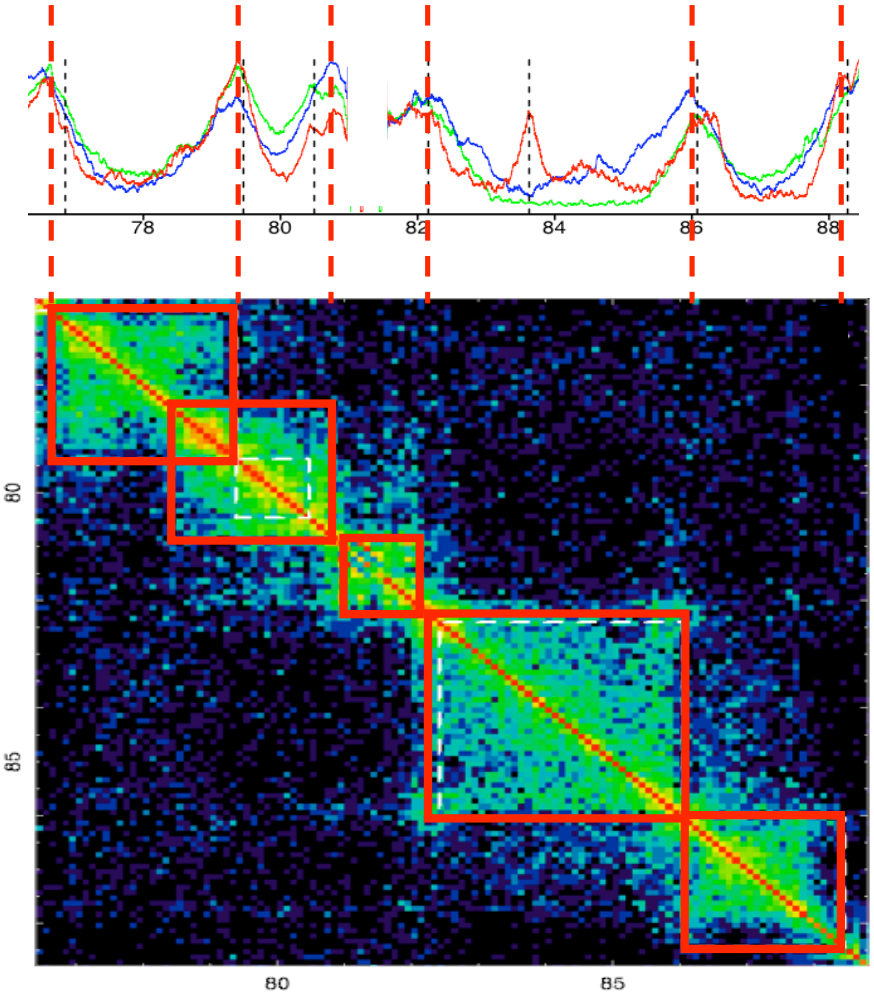


(data of erythroid from Lieberman-Aiden et al *Science.* 2

Replication domain corresponds to specific chromatin organization

Replication domains correspond to self-interacting units

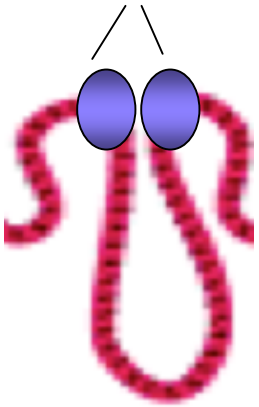
Timing



Chromatin interaction (Hi-C) matrix

(spatial proximity map)

Timing-domain borders

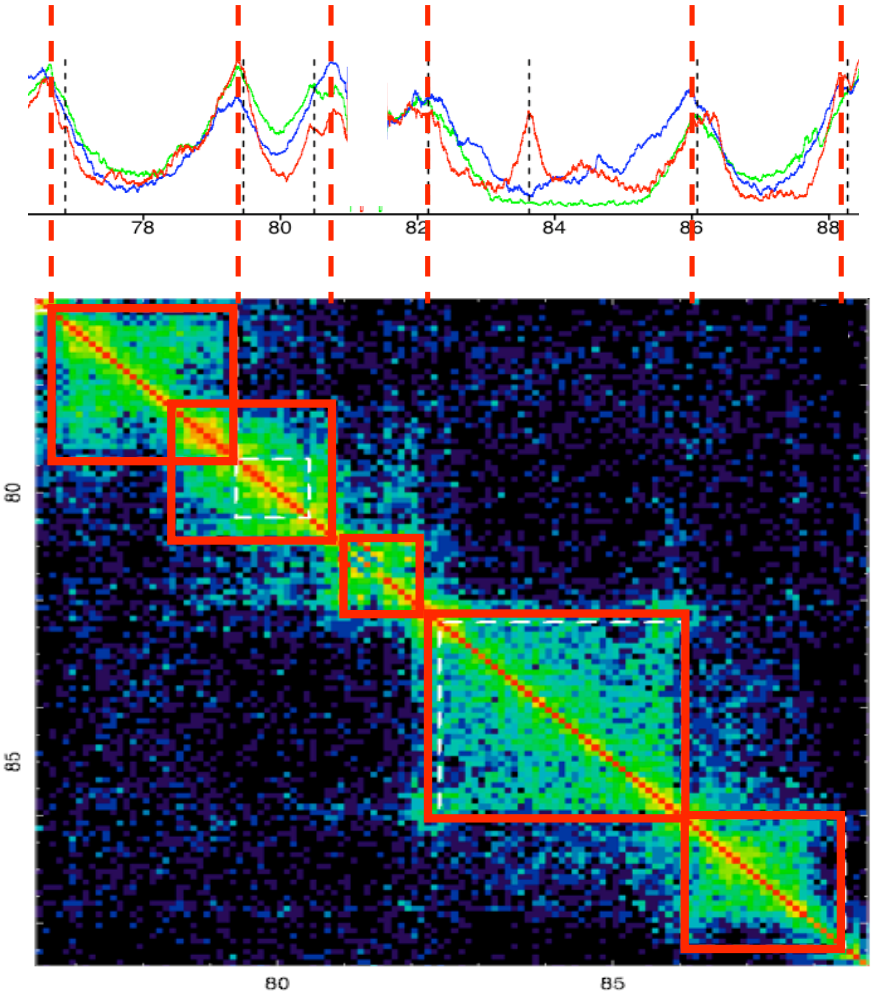


Baker A et al. *PLoS. Comput. Biol*. (2012) 8:e10024
Moindrot B. et al. *Nucleic. Acids. Res.* (2012) 40:94

Replication domain corresponds to specific chromatin organization

Replication domains correspond to self-interacting units

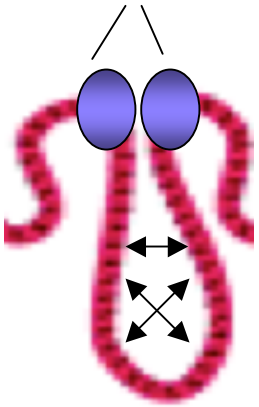
Timing



Chromatin interaction (Hi-C) matrix

(spatial proximity map)

Timing-domain borders



Baker A et al. *PLoS. Comput. Biol*. (2012) 8:e10024
Moindrot B. et al. *Nucleic. Acids. Res.* (2012) 40:94

Conclusions of Part I

- A method was developed and used to obtain one of the first high resolution replication timing profile of human genome by massive sequencing;
- A large fraction of genome displays conserved replication program in different cell types;
- U-shaped timing domains are widely observed in human cell types;
- U-domains corresponds to higher order chromatin organization.

PART II

IMPACT OF REPLICATION ON THE EVOLUTION AND ORGANIZATION OF THE GENOME

- Replication time as a major determinant of mammalian neutral substitution rates

CHEN et al. *Genome Res*. 2010

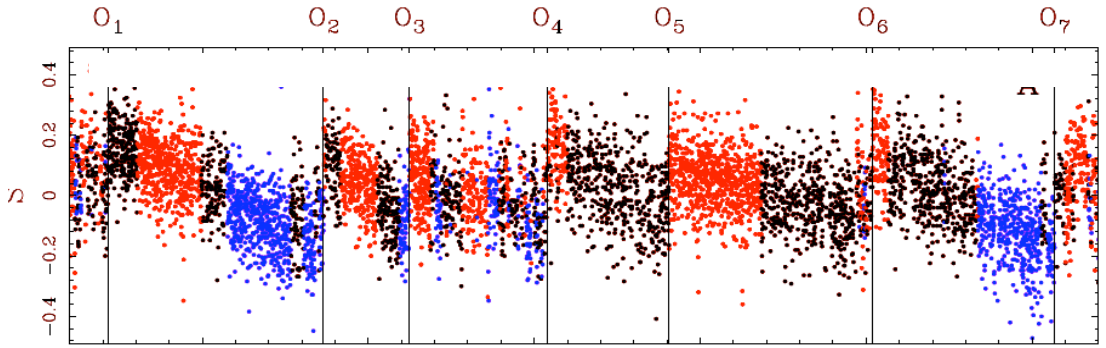
- Impact of replication on the evolution of human genome nucleotide composition

CHEN et al. *Mol Biol Evol*. 2011

PART II

IMPACT OF REPLICATION ON THE EVOLUTION AND ORGANIZATION OF THE GENOME

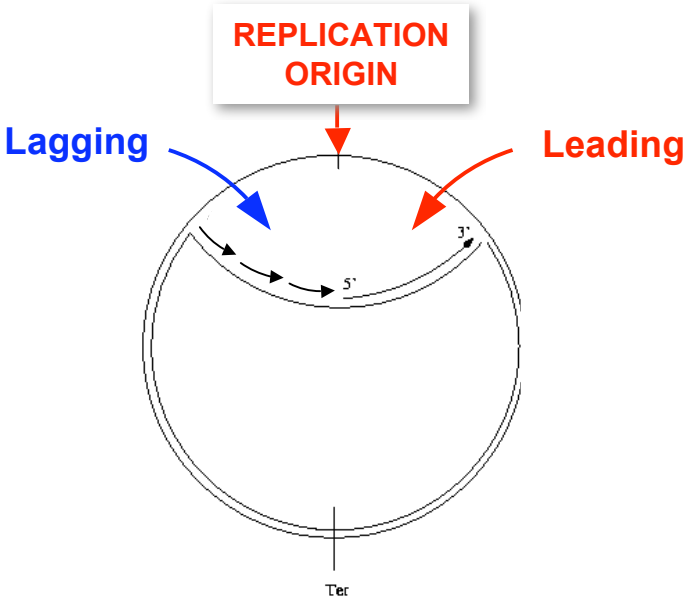
Nucleotide compositional skew



- Impact of replication on the evolution of human genome nucleotide composition

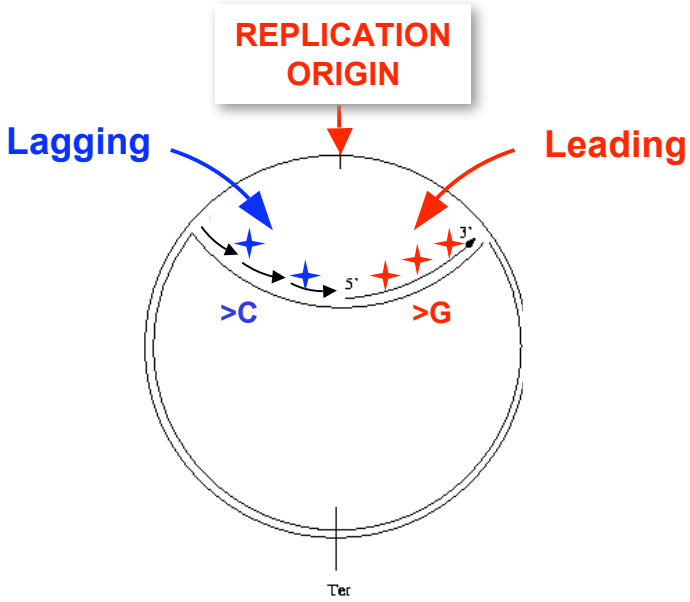
CHEN et al. *Mol Biol Evol*. 2011

Eubacteria:



Eubacteria:

Substitution rates
differ between
leading and lagging strands



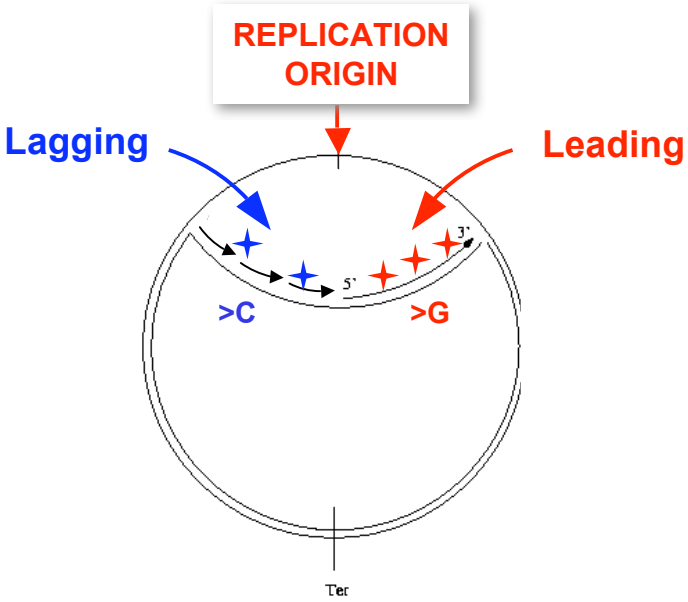
Francino & Ochman *Trends Genet.* 199
Frank & Lobry *Gene.* 199
Mrazek & Karlin *PNAS.* 199

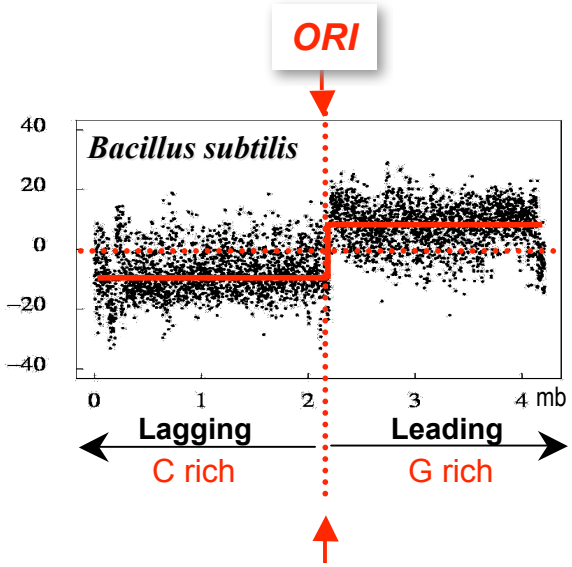
Eubacteria:

Substitution rates
differ between
leading and lagging strands

Compositional skew

$$S_{GC} = \frac{n_G - n_C}{n_G + n_C}$$



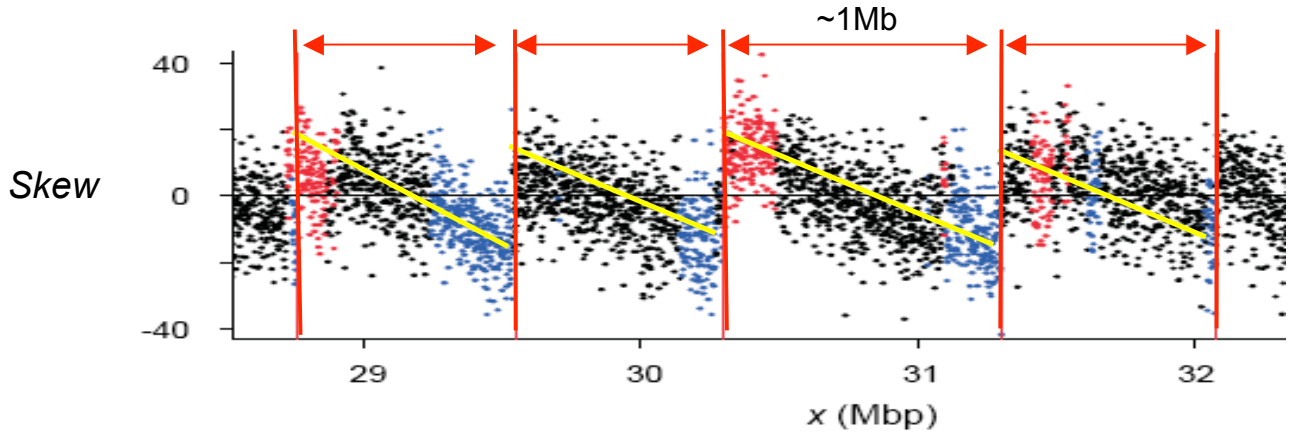


Upward jump
that allows to predict replication origin

Francino & Ochman *Trends Genet.* 199
Frank & Lobry *Gene.* 199
Mrazek & Karlin *PNAS.* 199

Human:

N-domains : > 1/3 of the genome

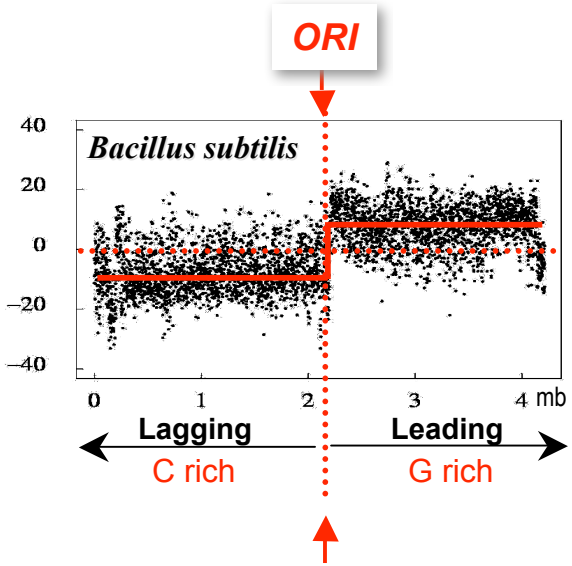


Touchon et al. *PNAS.* 20

Huvet et al. *Genome Res.* 20

Compositional skew

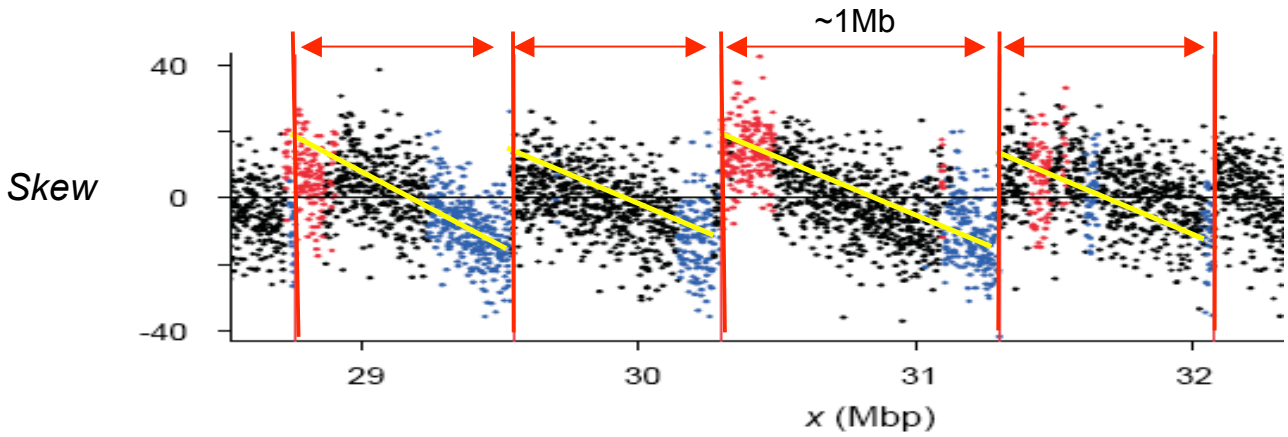
$$S_{GC} = \frac{n_G - n_C}{n_G + n_C}$$



Upward jump
that allows to predict replication origin

Human:

N-domains : > 1/3 of the genome



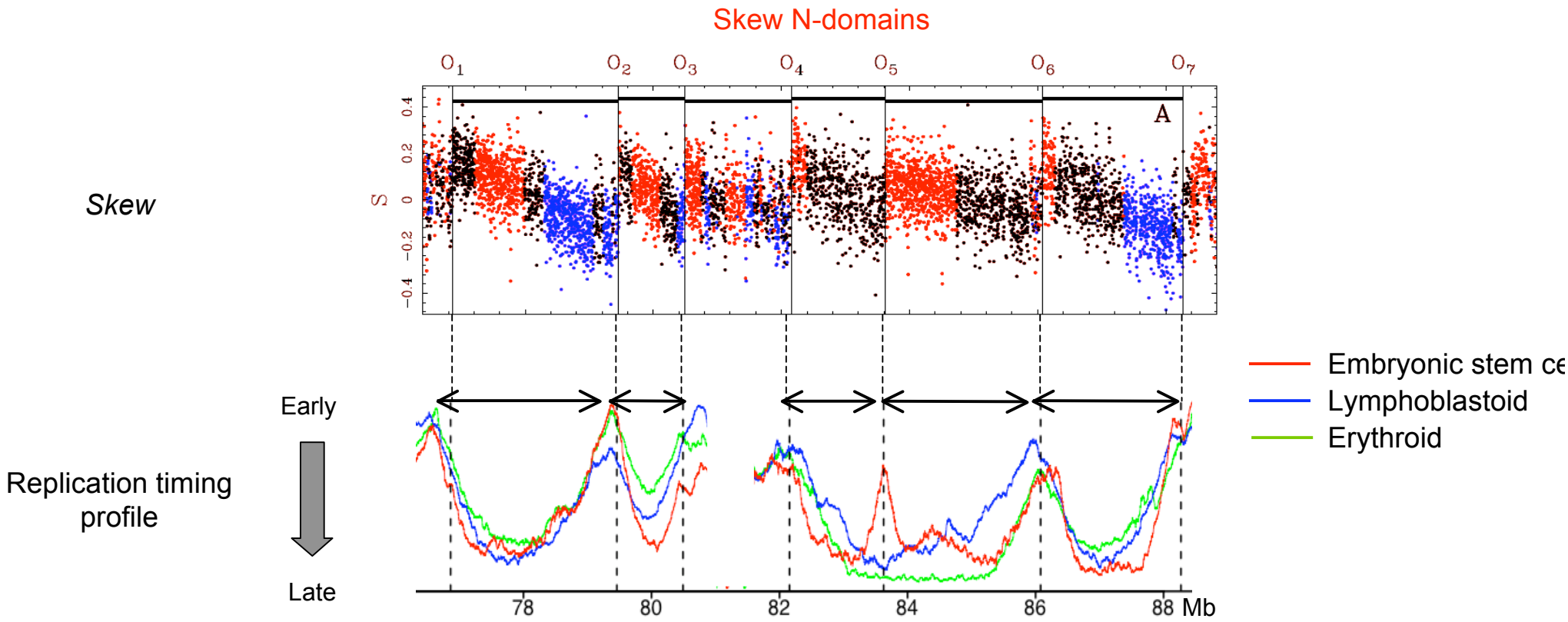
Touchon et al. *PNAS.* 20

Huvet et al. *Genome Res.* 20

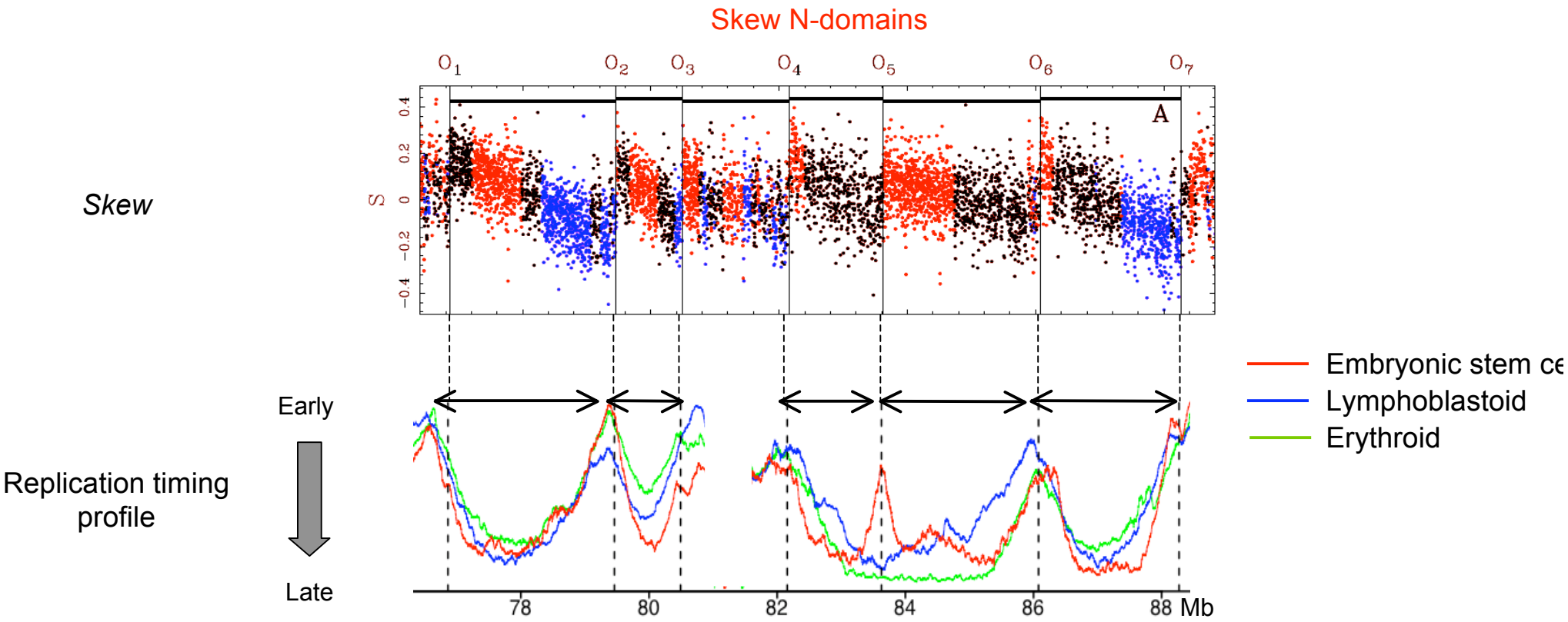
Do N-domains result from replication?

If yes, what kind of replication program generates the N-shape?

COMPARISON OF SKEW PROFILE AND REPLICATION TIMING PROFILES



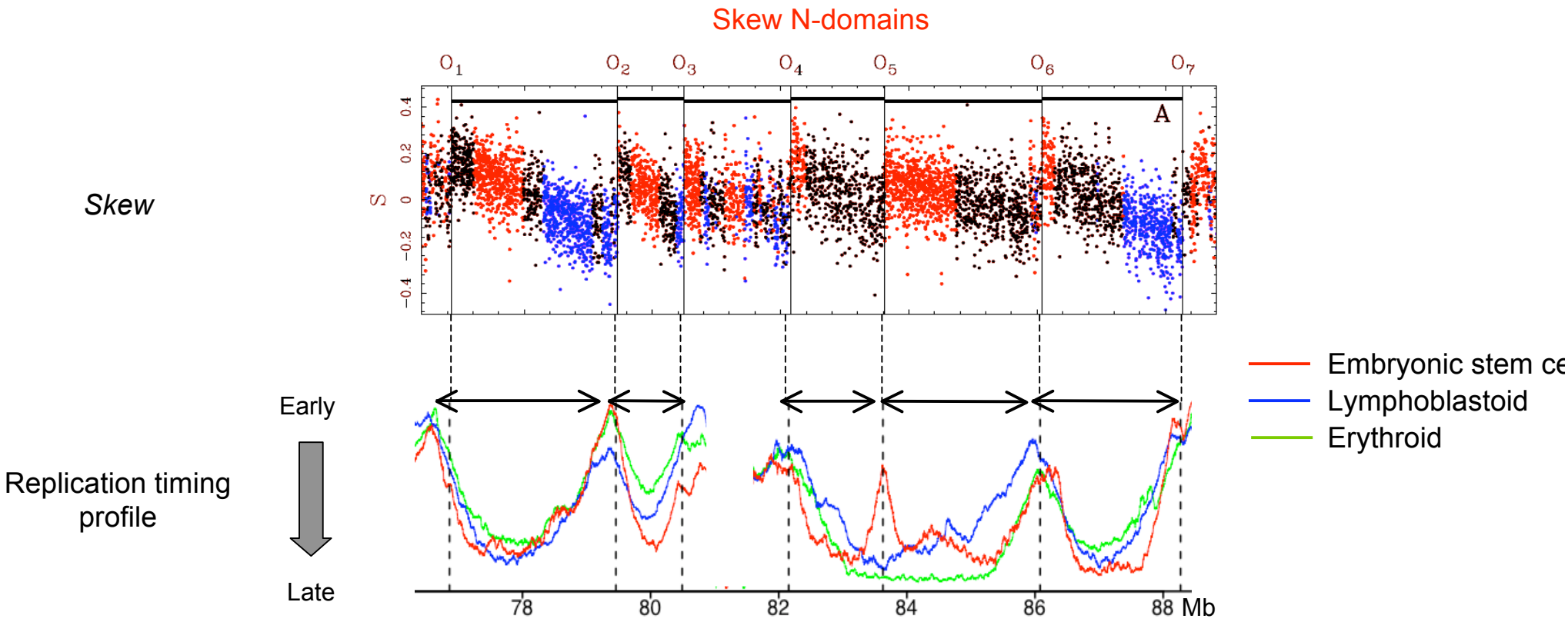
COMPARISON OF SKEW PROFILE AND REPLICATION TIMING PROFILES



- N-domain extremities are significantly associated with replication initiation zones
- Replication starts from N domain borders and propagates to center in late S phase

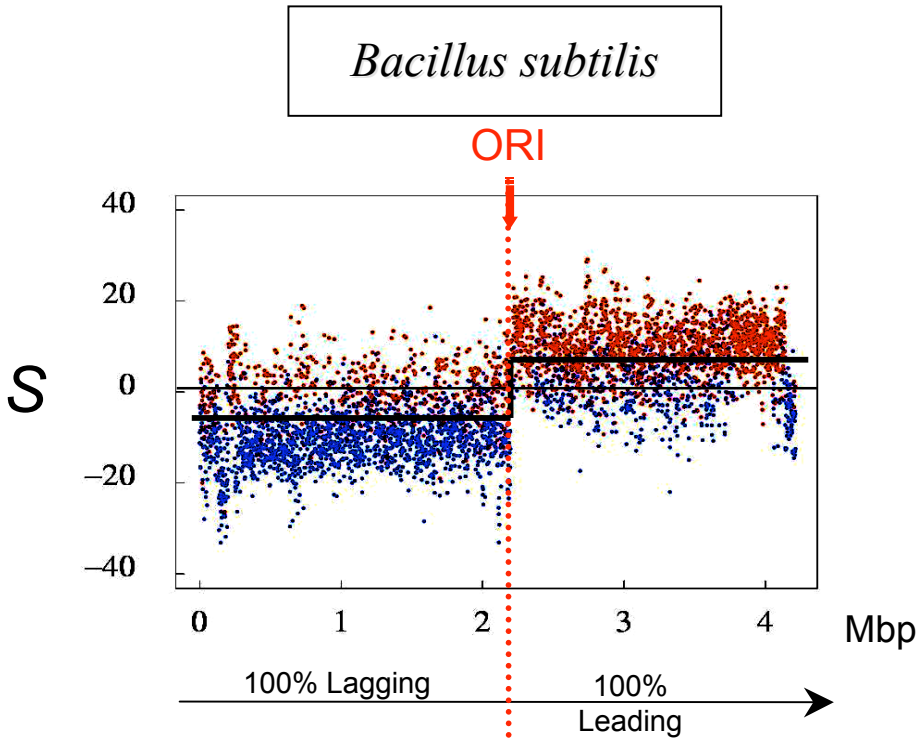
CHEN CL et al. *Mol. Biol. Evol.* (2011) 28:2
Baker A et al. *PLoS. Comput. Biol.* (2012) 8:e10024

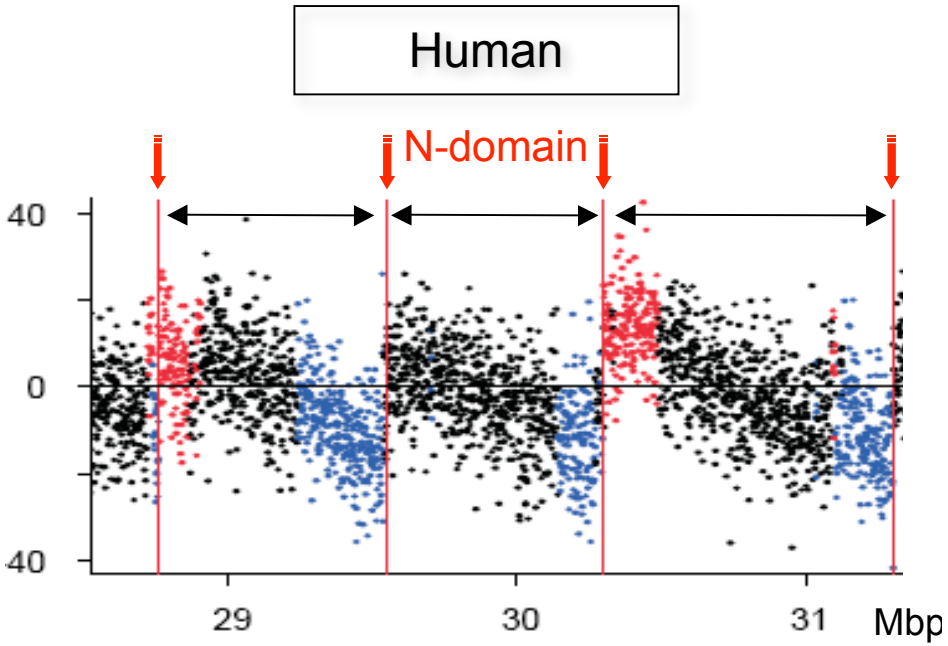
COMPARISON OF SKEW PROFILE AND REPLICATION TIMING PROFILES



N-domains ⟷ replication domains in germline cells

U-domains ⟷ replication domains in somatic cells

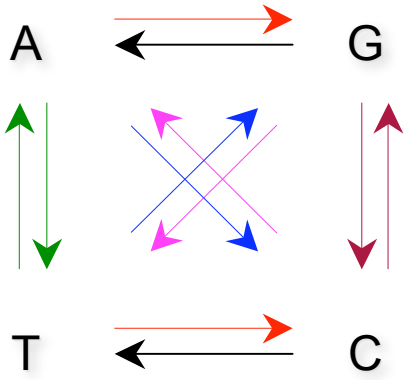


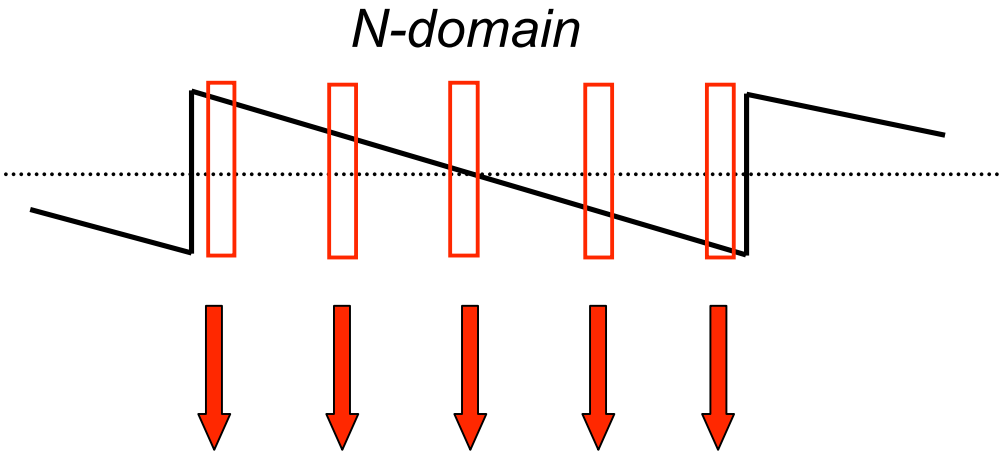


Is the “N” skew pattern generated by asymmetric nucleotide substitution rates ?

?

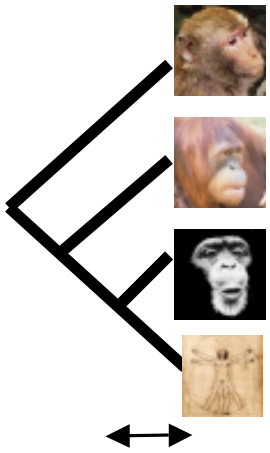
Computation of nucleotide substitution rates





A -> C	?	?	?	?	?
A -> G	?	?	?	?	?
A -> T	?	?	?	?	?
C -> A	?	?	?	?	?
C -> G	?	?	?	?	?
C -> T	?	?	?	?	?
G -> A	?	?	?	?	?
G -> C	?	?	?	?	?
G -> T	?	?	?	?	?
T -> A	?	?	?	?	?
T -> C	?	?	?	?	?
T -> G	?	?	?	?	?

Computation of nucleotide substitution rates



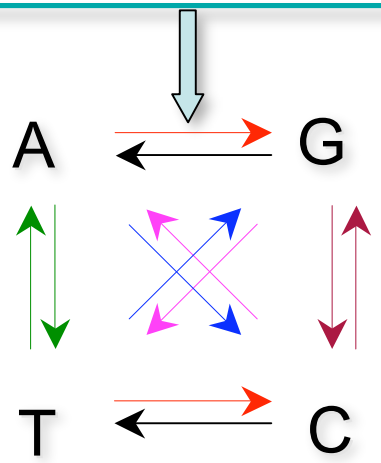
Macaca mulatta
(rhesus monkey)
Pongo pygmaeus abelii
(orangutan)
Pan troglodytes
(chimpanzee)
Homo sapiens
(human)

```
AACTTTCGGTAGTGGATGACATCCCATGTCGTGT
AACTTTCGGTAGCGGATGACATCCCATGTCGTGT
AACTTTCGATAGTGGATGAGATCCCATGTCGTGT
AACTTTCGTTAGTGGATGACATCCTATGTCGTGT
```

non-informative site

chimp C → G

human C → T



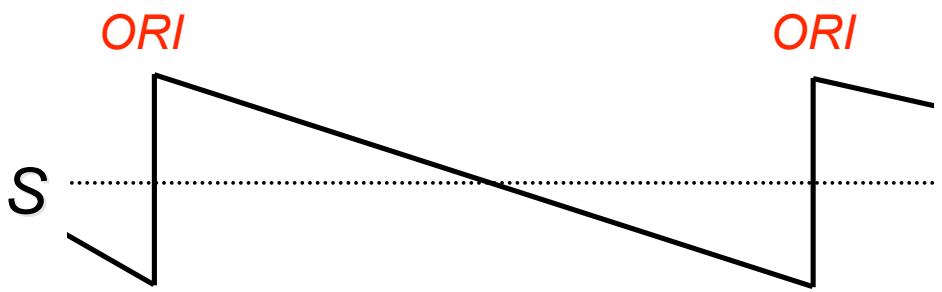
Alignment data retrieved from *Ensei*

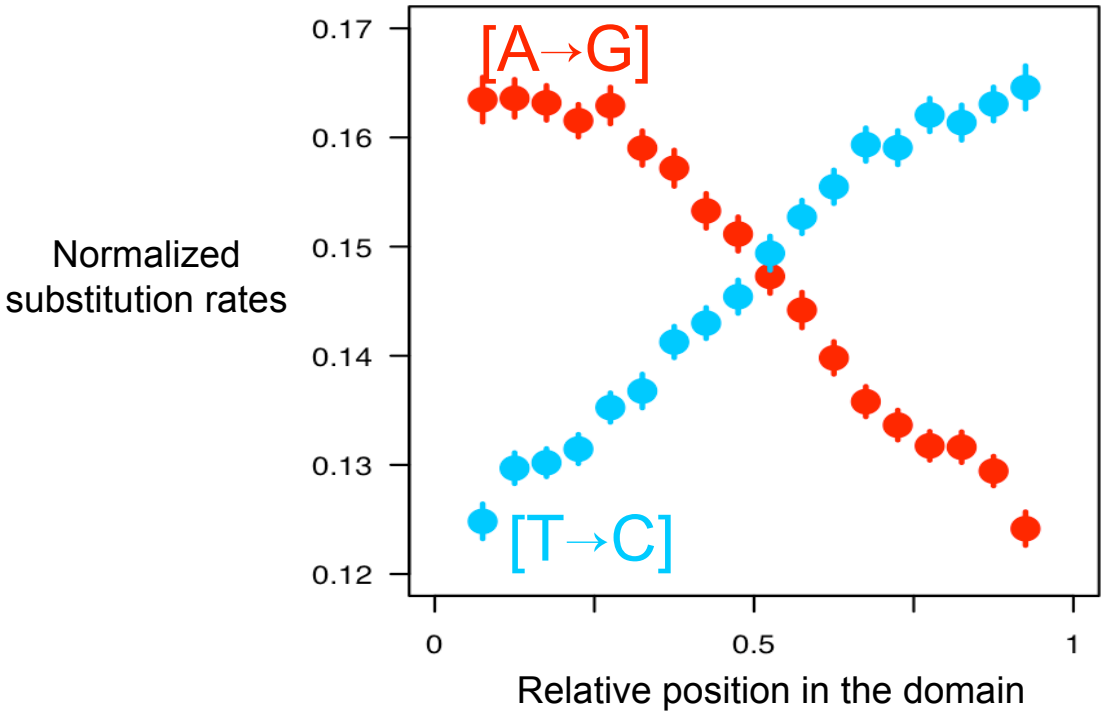
Compare A→G values on the two strands

Compare A→G to T→C on the same strand

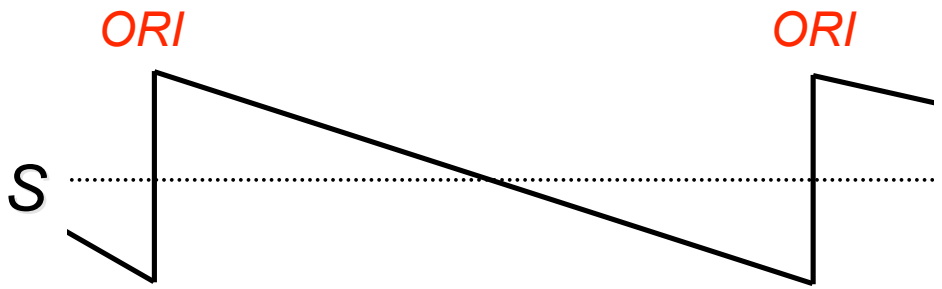
Compare complementary substitution rates on the same strand

Complementary substitution rate along N-domains



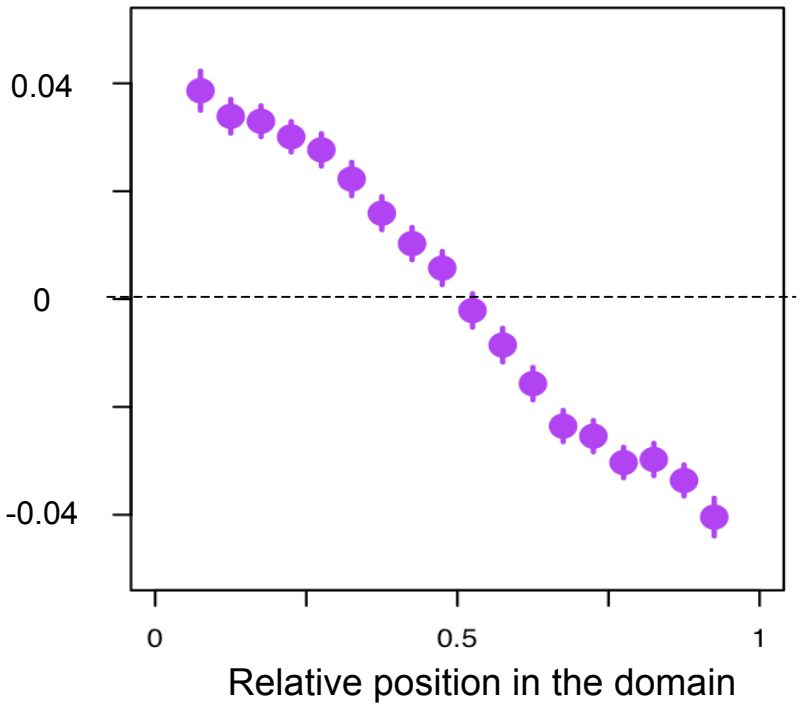


Complementary substitution rate along N-domains



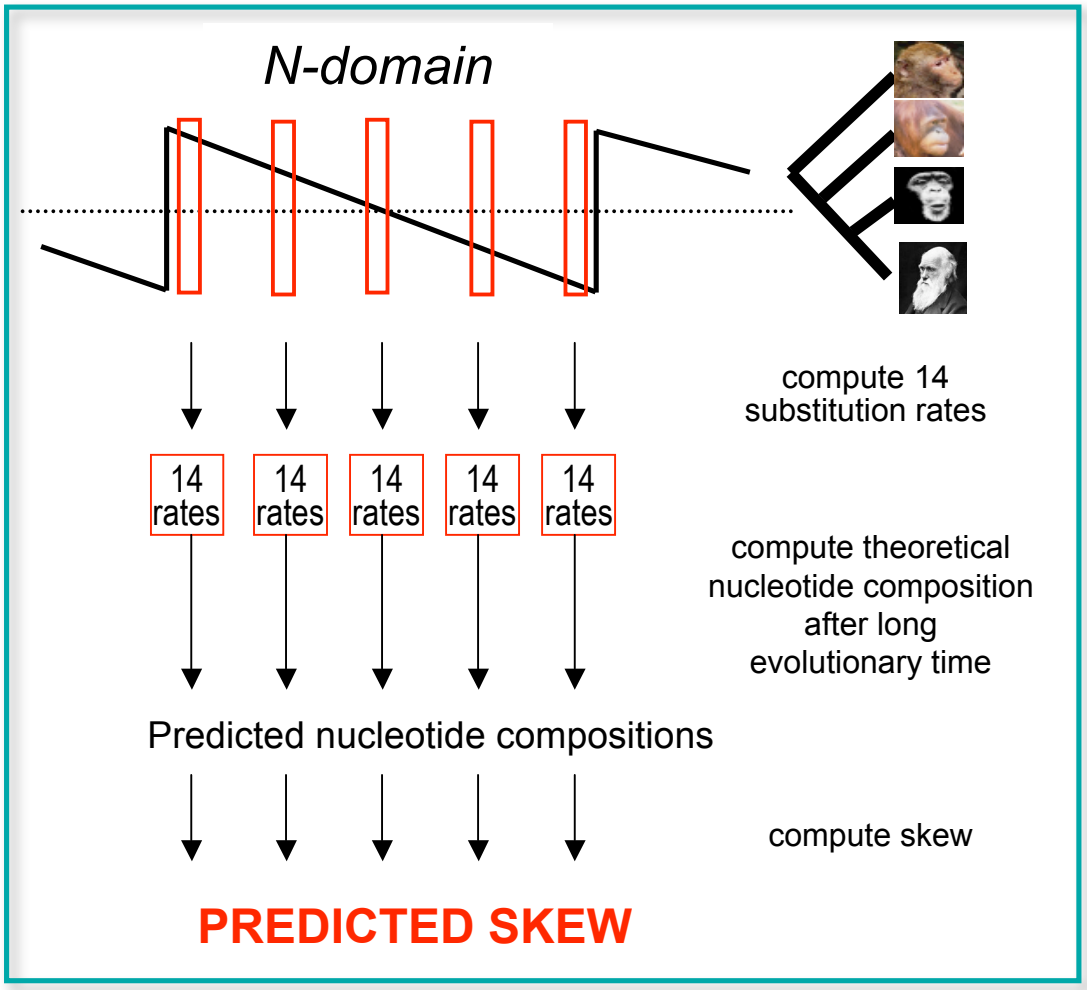
Asymmetry

$\triangle = [A \rightarrow G] - [T \rightarrow C]$



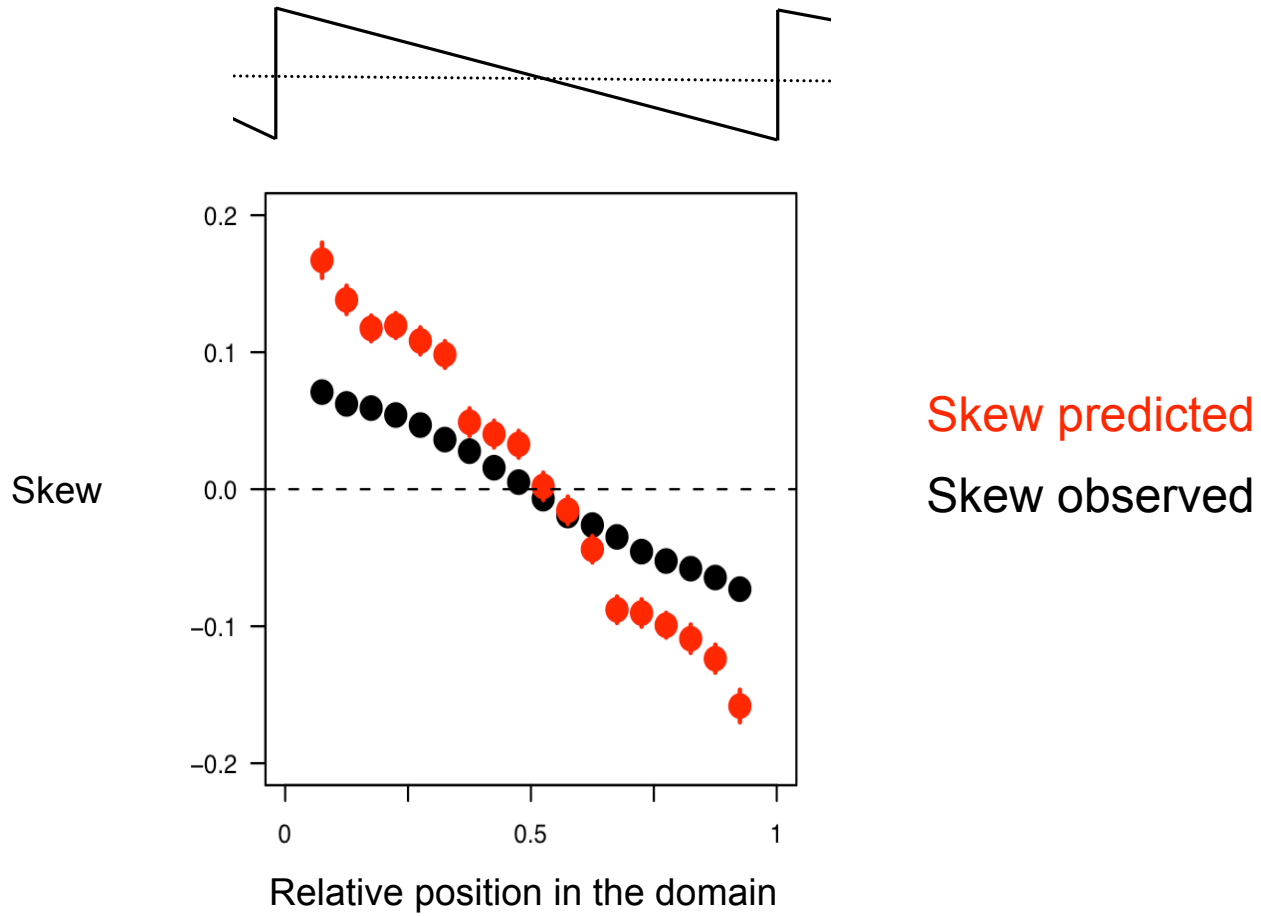
Reproduces perfectly the “N” skew profile

Compute the predicted skew (*S* at equilibrium) along N-domain



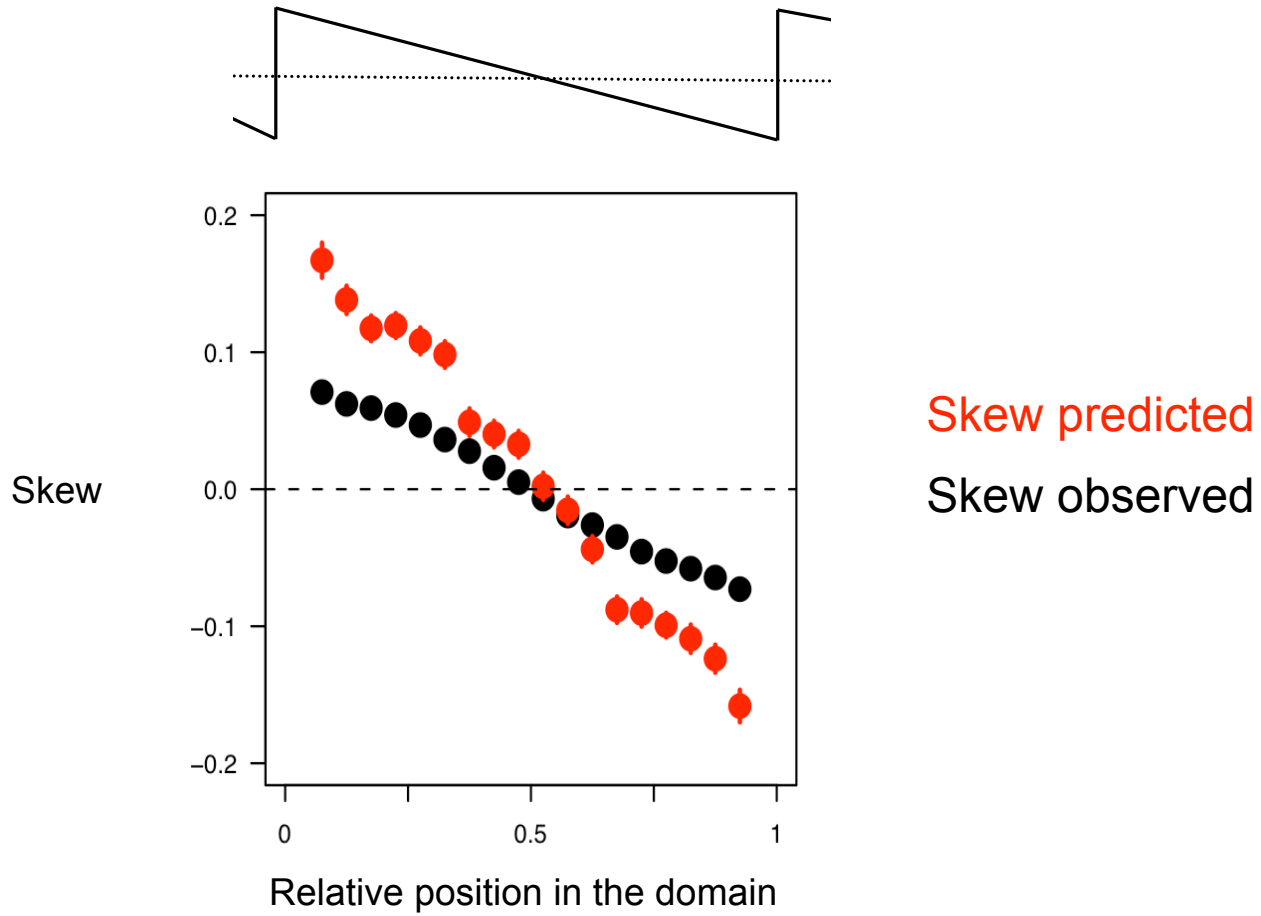
$$
\begin{aligned}
\frac{\partial f_{aa}}{\partial t} &= caf_{ca} + gaf_{ga} + taf_{ta} - acf_{aa} - agf_{aa} - atf_{aa} + caf_{ac} + gaf_{ag} + taf_{at} - acf_{aa} - agf_{aa} - atf_{aa} + \frac{pgpaf_{cg}f_{ga}}{f_{ga}+f_{gc}+f_{gg}+f_{gt}} \\
\frac{\partial f_{ac}}{\partial t} &= caf_{cc} + gaf_{gc} + taf_{tc} - acf_{ac} - agf_{ac} - atf_{ac} + acf_{aa} + gcf_{ag} + tcf_{at} - caf_{ac} - cgf_{ac} - ctf_{ac} + \frac{pgpaf_{cg}f_{gc}}{f_{ga}+f_{gc}+f_{gg}+f_{gt}} - \frac{cpt}{f_{ca}+f_{cc}} \\
\frac{\partial f_{ag}}{\partial t} &= caf_{cg} + gaf_{gg} + taf_{tg} - acf_{ag} - agf_{ag} - atf_{ag} + agf_{aa} + cgf_{ac} + tgf_{at} - gaf_{ag} - gcf_{ag} - gtf_{ag} + \frac{pgpaf_{cg}f_{gg}}{f_{ga}+f_{gc}+f_{gg}+f_{gt}} \\
\frac{\partial f_{at}}{\partial t} &= caf_{ct} + gaf_{gt} + taf_{tt} - acf_{at} - agf_{at} - atf_{at} + atf_{aa} + ctf_{ac} + gtf_{ag} - taf_{at} - tcf_{at} - tgf_{at} + \frac{pgpaf_{cg}f_{gt}}{f_{ga}+f_{gc}+f_{gg}+f_{gt}} + \frac{cptpf}{f_{ca}+f_{cc}+} \\
\frac{\partial f_{ca}}{\partial t} &= acf_{aa} + gcf_{ga} + tcf_{ta} - caf_{ca} - cgf_{ca} - ctf_{ca} + caf_{cc} + taf_{ct} - acf_{ca} - agf_{ca} - atf_{ca} + pgpaf_{cg} \\
\frac{\partial f_{cc}}{\partial t} &= acf_{ac} + gcf_{gc} + tcf_{tc} - caf_{cc} - cgf_{cc} - ctf_{cc} + acf_{ca} + gcf_{cg} + tcf_{ct} - caf_{cc} - cgf_{cc} - ctf_{cc} - \frac{cptpf_{cc}f_{cg}}{f_{ca}+f_{cc}+f_{cg}+f_{ct}} \\
\frac{\partial f_{cg}}{\partial t} &= acf_{ag} + gcf_{gg} + tcf_{tg} - caf_{cg} - cgf_{cg} + agf_{ca} + cgf_{cc} + tgf_{ct} - gcf_{cg} - gtf_{cg} - cptpf_{cg} - pgpaf_{cg} \\
\frac{\partial f_{ct}}{\partial t} &= acf_{at} + gcf_{gt} + tcf_{tt} - caf_{ct} - cgf_{ct} - ctf_{ct} + atf_{ca} + ctf_{cc} + gtf_{cg} - taf_{ct} - tcf_{ct} - tgf_{ct} + \frac{cptpf_{cc}f_{cg}}{f_{ca}+f_{cc}+f_{cg}+f_{ct}} \\
\frac{\partial f_{ga}}{\partial t} &= agf_{aa} + cgf_{ca} + tgf_{ta} - gaf_{ga} - gcf_{ga} - gtf_{ga} + caf_{gc} + gaf_{gg} + taf_{gt} - acf_{ga} - agf_{ga} - atf_{ga} - \frac{pgpaf_{cg}f_{ga}}{f_{ga}+f_{gc}+f_{gg}+f_{gt}} \\
\frac{\partial f_{gc}}{\partial t} &= agf_{ac} + cgf_{cc} + tgf_{tc} - gaf_{gc} - gcf_{gc} - gtf_{gc} + acf_{ga} + gcf_{gg} + tcf_{gt} - caf_{gc} - cgf_{gc} - ctf_{gc} - \frac{pgpaf_{cg}f_{gc}}{f_{ga}+f_{gc}+f_{gg}+f_{gt}} - \frac{cptp}{f_{ca}+f_{cc}} \\
\frac{\partial f_{gg}}{\partial t} &= agf_{ag} + cgf_{cg} + tgf_{tg} - gaf_{gg} - gcf_{gg} - gtf_{gg} + agf_{ga} + cgf_{gc} + tgf_{gt} - gaf_{gg} - gcf_{gg} - gtf_{gg} - \frac{pgpaf_{cg}f_{gg}}{f_{ga}+f_{gc}+f_{gg}+f_{gt}} \\
\frac{\partial f_{gt}}{\partial t} &= agf_{at} + cgf_{ct} + tgf_{tt} - gaf_{gt} - gcf_{gt} - gtf_{gt} + atf_{ga} + ctf_{gc} + gtf_{gt} - taf_{gt} - tcf_{gt} - tgf_{gt} - \frac{pgpaf_{cg}f_{gt}}{f_{ga}+f_{gc}+f_{gg}+f_{gt}} + \frac{cptpf_{g}}{f_{ca}+f_{cc}+} \\
\frac{\partial f_{ta}}{\partial t} &= atf_{aa} + ctf_{ca} + gtf_{ga} - taf_{ta} - tcf_{ta} - tgf_{ta} + caf_{tc} + gaf_{tg} + taf_{tt} - acf_{ta} - agf_{ta} - atf_{ta} \\
\frac{\partial f_{tc}}{\partial t} &= atf_{ac} + ctf_{cc} + gtf_{gc} - taf_{tc} - tcf_{tc} - tgf_{tc} + acf_{ta} + gcf_{tg} + tcf_{tt} - caf_{tc} - cgf_{tc} - ctf_{tc} - \frac{cptpf_{tc}f_{cg}}{f_{ca}+f_{cc}+f_{cg}+f_{ct}} \\
\frac{\partial f_{tg}}{\partial t} &= atf_{ag} + gtf_{gg} - taf_{tg} - tcf_{tg} - tgf_{tg} + agf_{ta} + cgf_{tc} + tgf_{tt} - gaf_{tg} - gcf_{tg} - gtf_{tg} + cptpf_{cg} \\
\frac{\partial f_{tt}}{\partial t} &= atf_{at} + ctf_{ct} + gtf_{gt} - taf_{tt} - tcf_{tt} - tgf_{tt} + atf_{ta} + ctf_{tc} + gtf_{tg} - taf_{tt} - tcf_{tt} - tgf_{tt} + \frac{cptpf_{tc}f_{cg}}{f_{ca}+f_{cc}+f_{cg}+f_{ct}}
\end{aligned}
$$

COMPOSITION AT EQUILIBRIUM REPRODUCES PERFECTLY THE “N” SKEW PROFILE



N-domains result from mutation asymmetry in germline cells

COMPOSITION AT EQUILIBRIUM REPRODUCES PERFECTLY THE “N” SKEW PROFILE



N-domains result from mutation asymmetry in germline cells

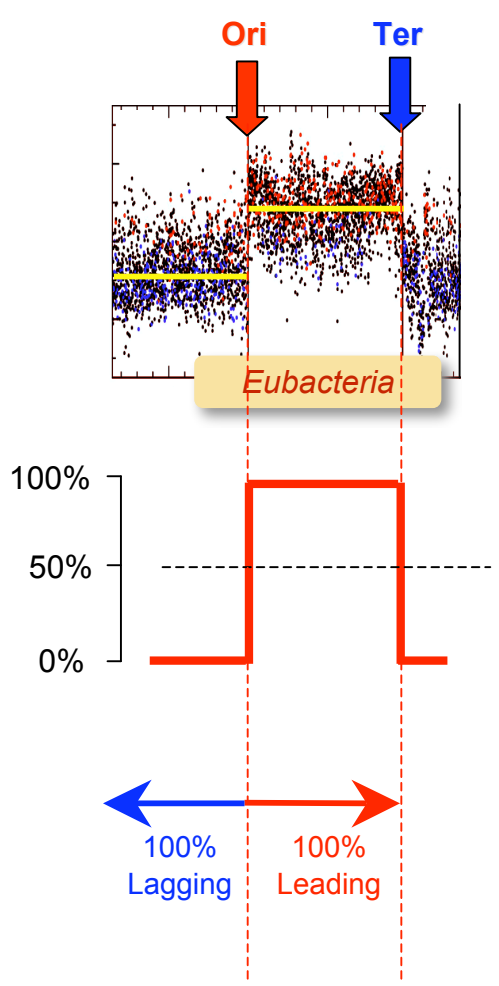
The skew is not at equilibrium. Time to reach the observed skew : 300 – 400 million years

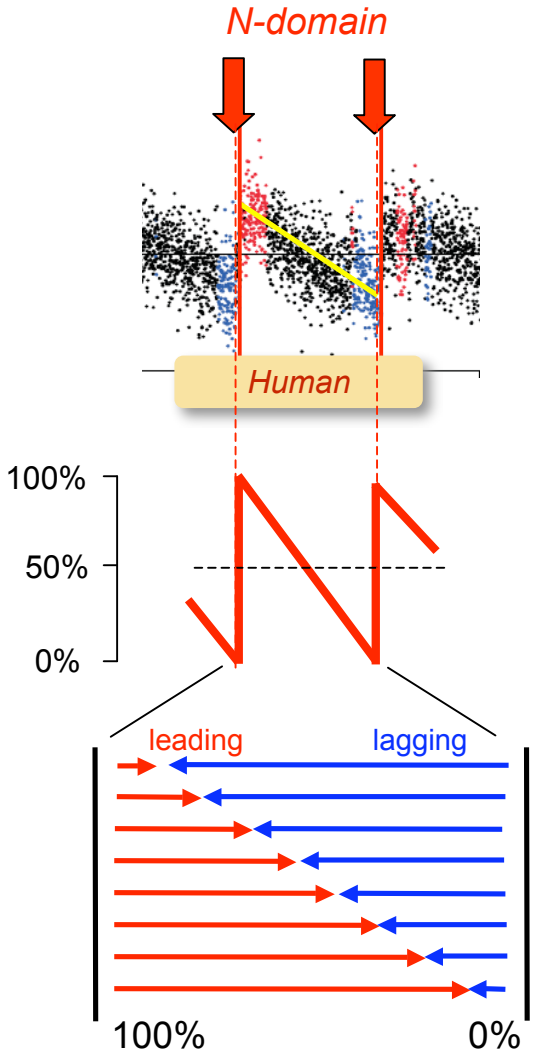
What model of replication can explain the N-pattern ?

MODEL: N-shape results from gradient of replication fork polarity

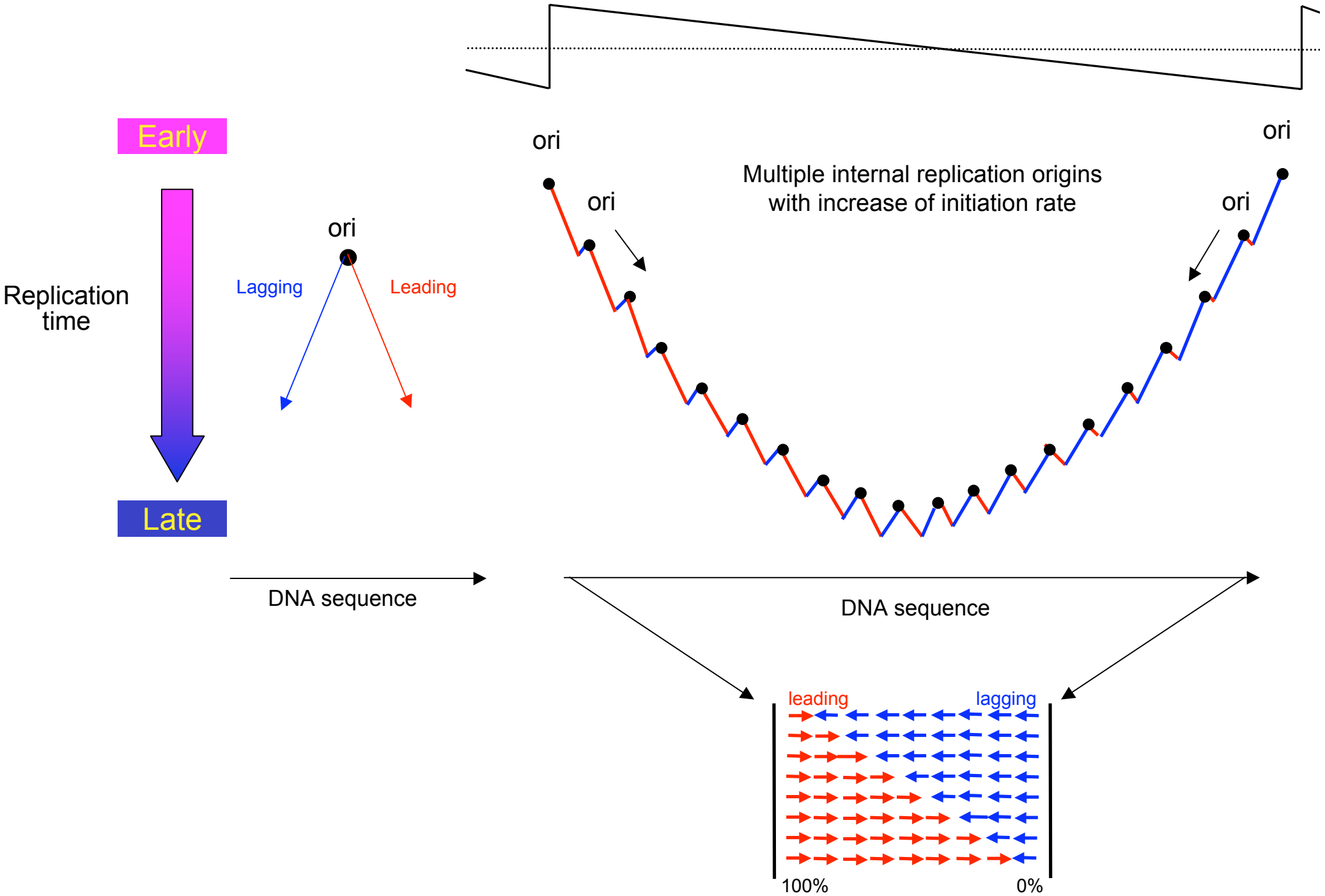
Replication fork polarity

$$\frac{\textit{Leading}}{\textit{Leading + lagging}}$$



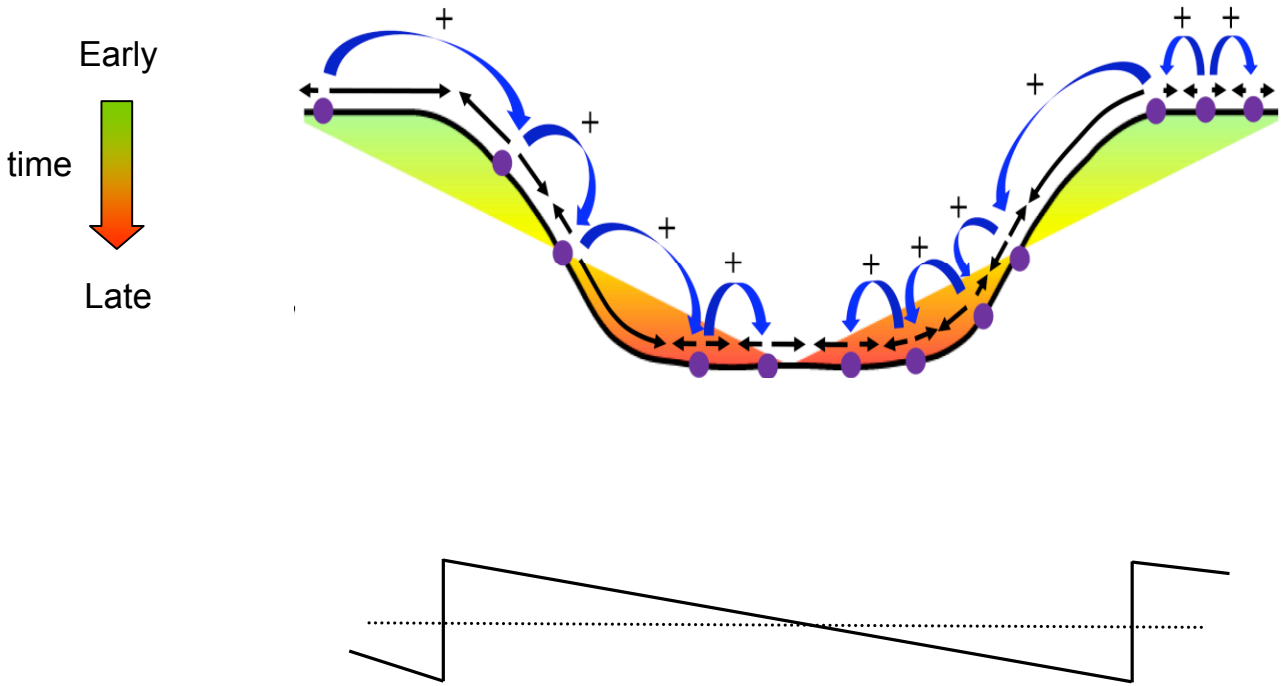


gradient of replication fork polarity


Early
Late
Replication time
ori
Lagging
Leading
DNA sequence
ori
ori
ori
Multiple internal replication origins
with increase of initiation rate
DNA sequence
leading
lagging
100%
0%

Domino replication model

(based on our timing and DNA combing data)

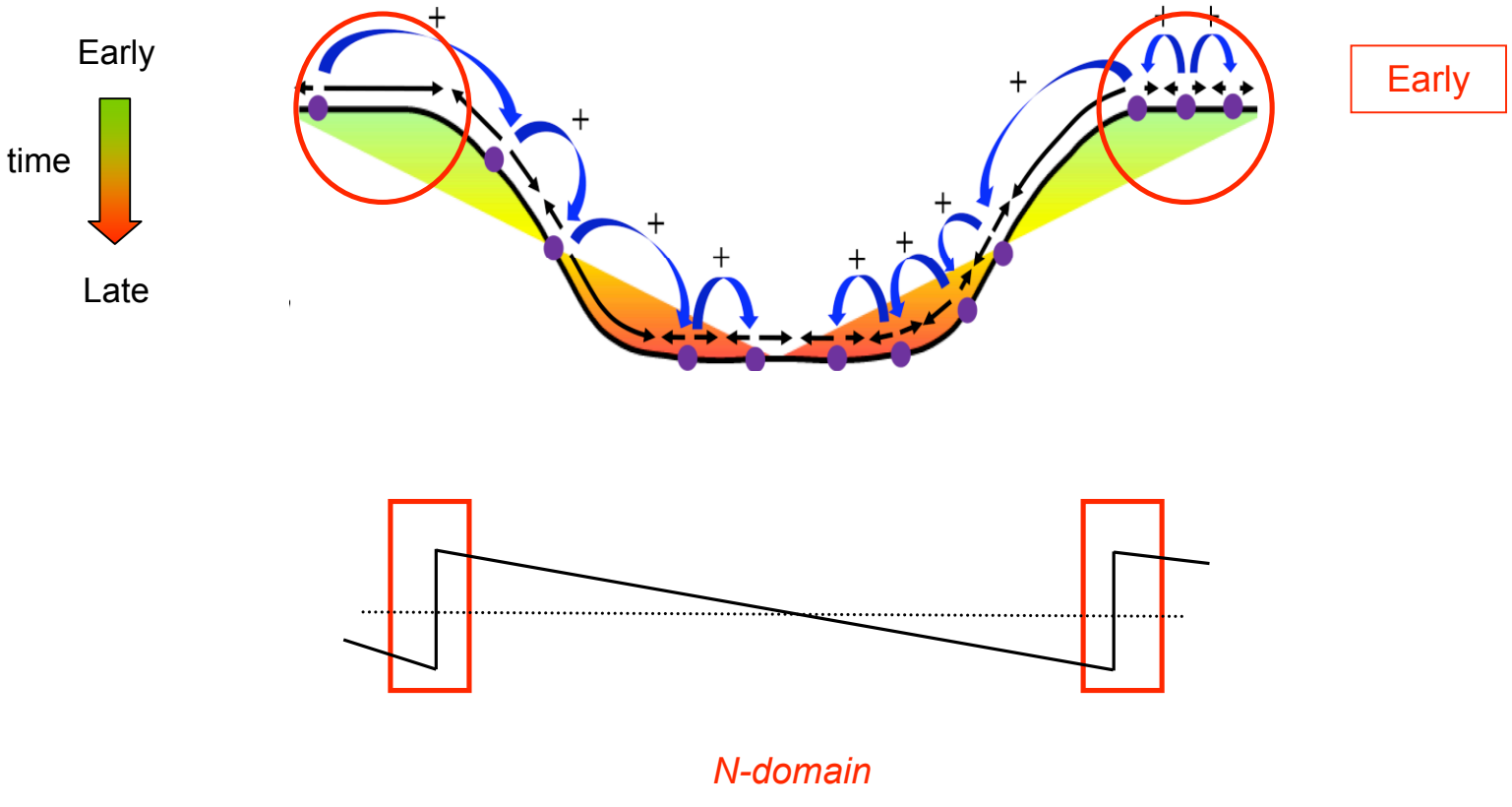


N-domain

Guilbaud et al. *PLoS Comput. Biol.* (2011) 7:e10023
Baker A et al. *PLoS. Comput. Biol.* (2012) 8:e10024

Domino replication model

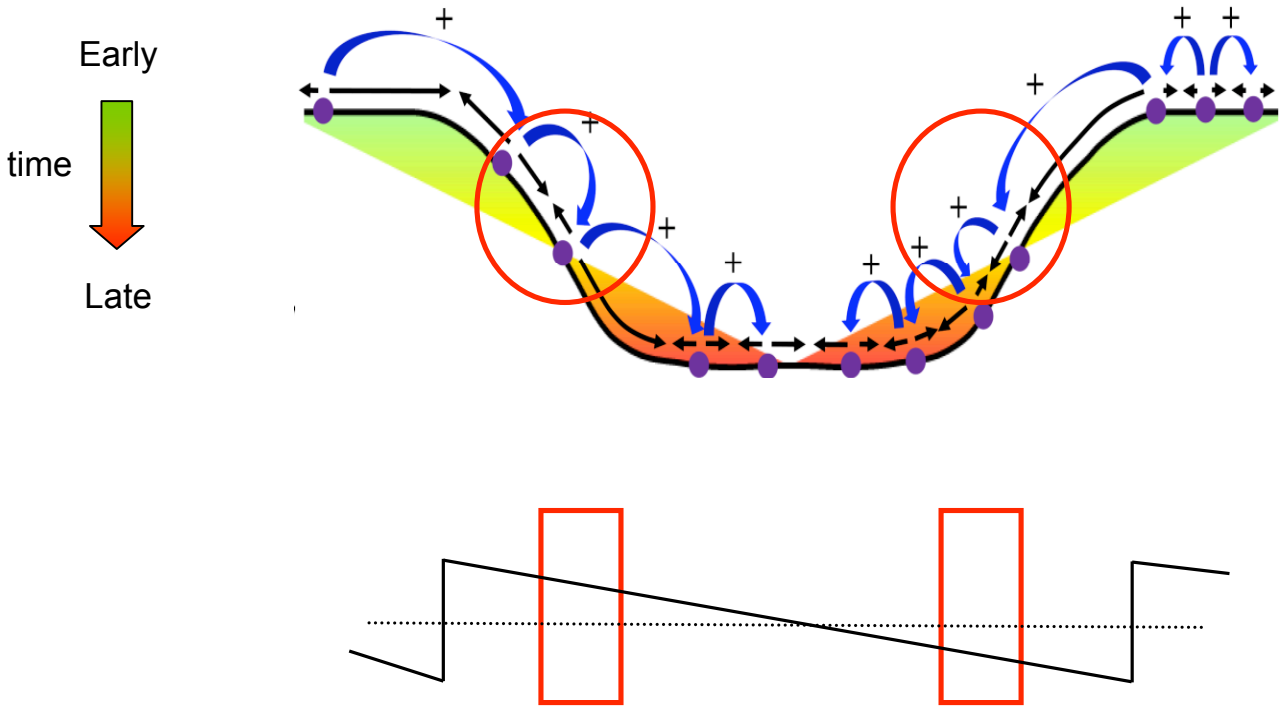
(based on our timing and DNA combing data)



Guilbaud et al. *PLoS Comput. Biol.* (2011) 7:e10023
Baker A et al. *PLoS. Comput. Biol.* (2012) 8:e10024

Domino replication model

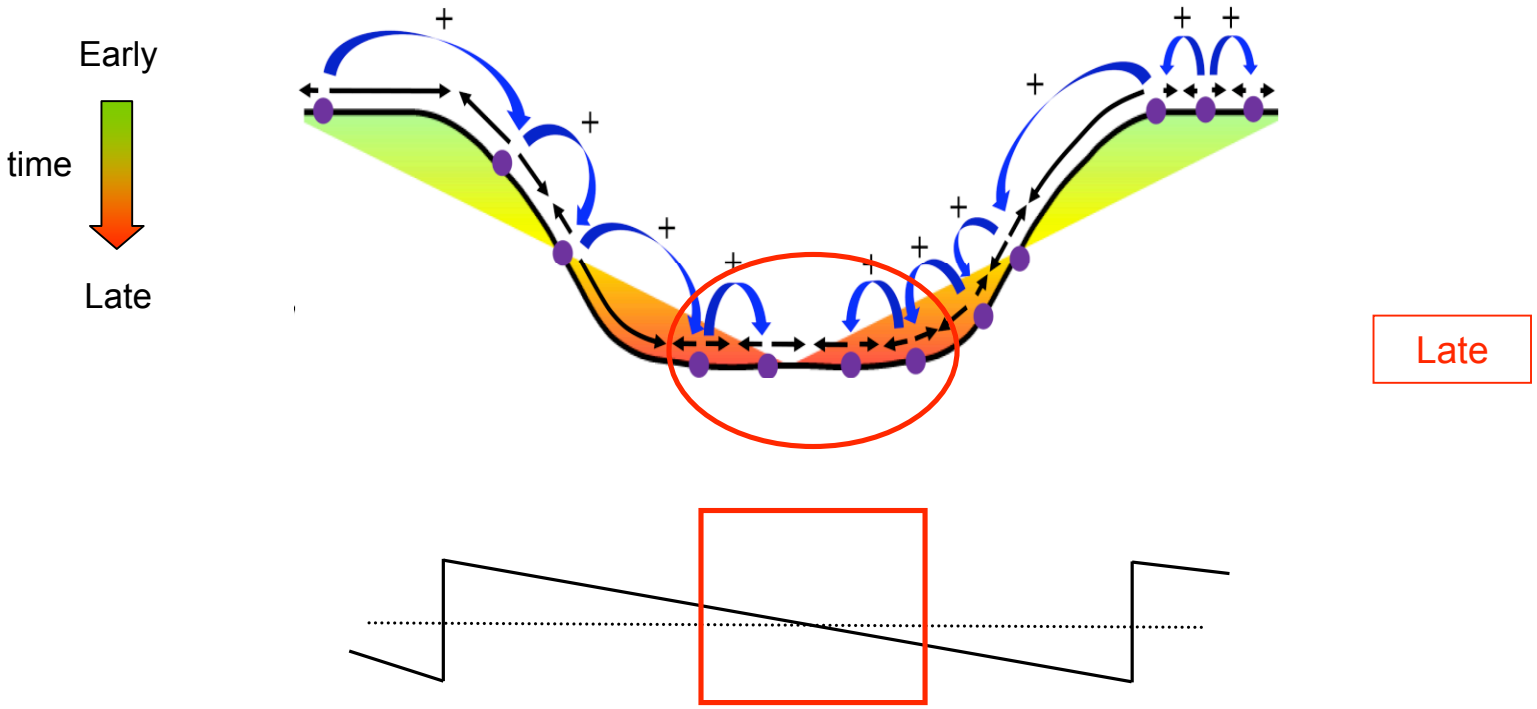
(based on our timing and DNA combing data)



Guilbaud et al. *PLoS Comput. Biol.* (2011) 7:e10023
Baker A et al. *PLoS. Comput. Biol.* (2012) 8:e10024

Domino replication model

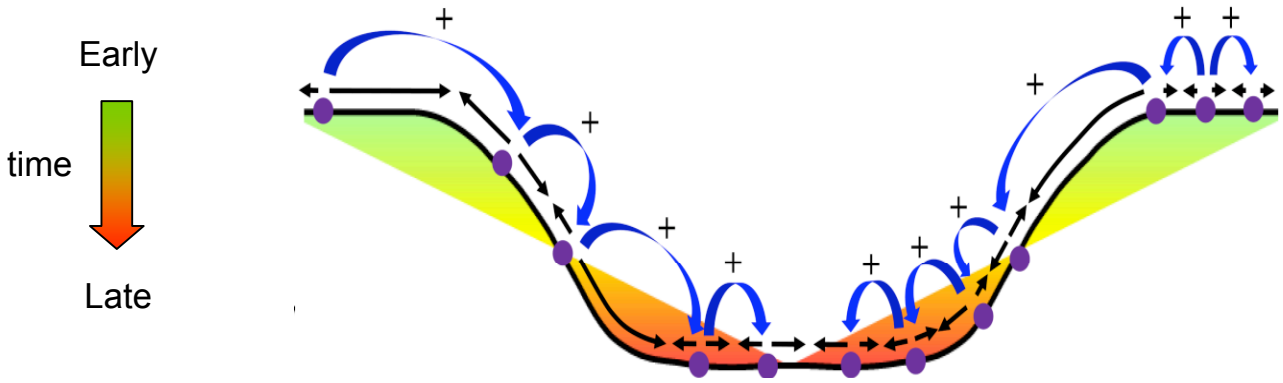
(based on our timing and DNA combing data)



Guilbaud et al. *PLoS Comput. Biol.* (2011) 7:e10023
Baker A et al. *PLoS. Comput. Biol.* (2012) 8:e10024

Domino replication model

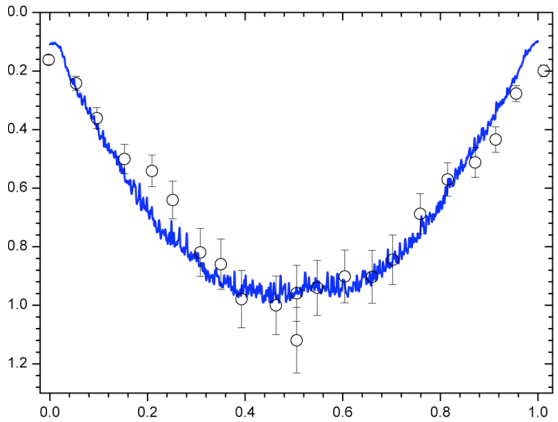
(based on our timing and DNA combing data)



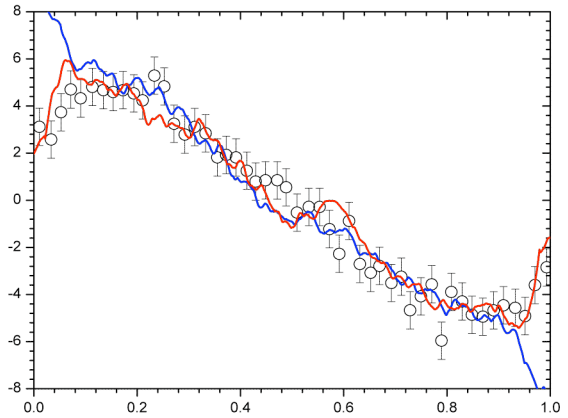
Computational simulation

A. Goldar, O. Hyrien

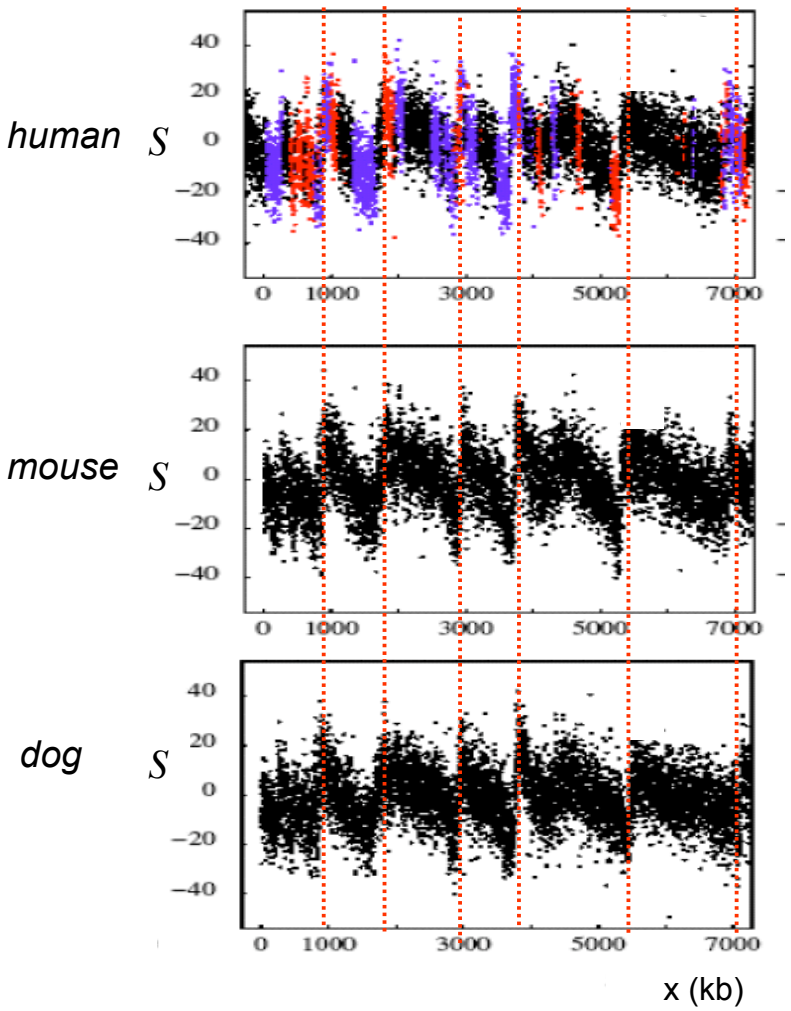
U-shaped replication timing



N-shaped compositional skew

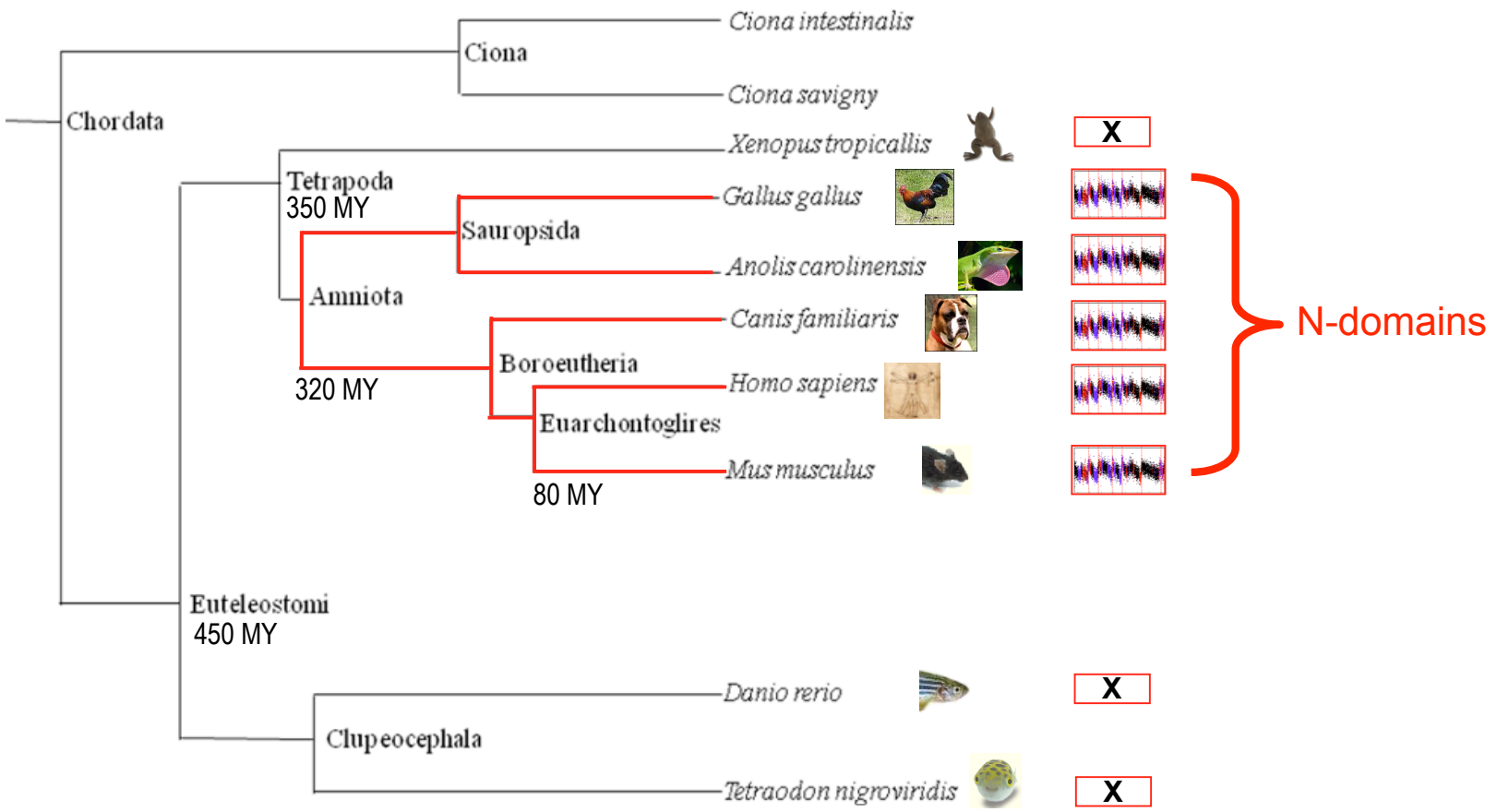


Skew N-domains in mammalian genomes



The N-domains are conserved during mammalian evolution.

N-domains are 320 million years old



This replication program has been conserved since amniota divergence

Conclusions of Part II

- N-shape of skew profile is generated by gradient of replication fork polarity
- Skew N-domains correspond to U-shaped timing domains of germline cells
- We construct a domino-model of replication : replication initiates at master origins and propagates by cascade of secondary initiations associated with a gradient of open chromatin structure
- This replication program has been conserved during mammalian evolution

Acknowledgments and collaborations

Chun-Long Chen
Maud Silvain
Yves d'Aubenton-Carafa
Claude Thermes
(CGM, Gif sur Yvette)

Arach Goldar
(CEA, iBiTec-S, Gif-sur-Yvette)

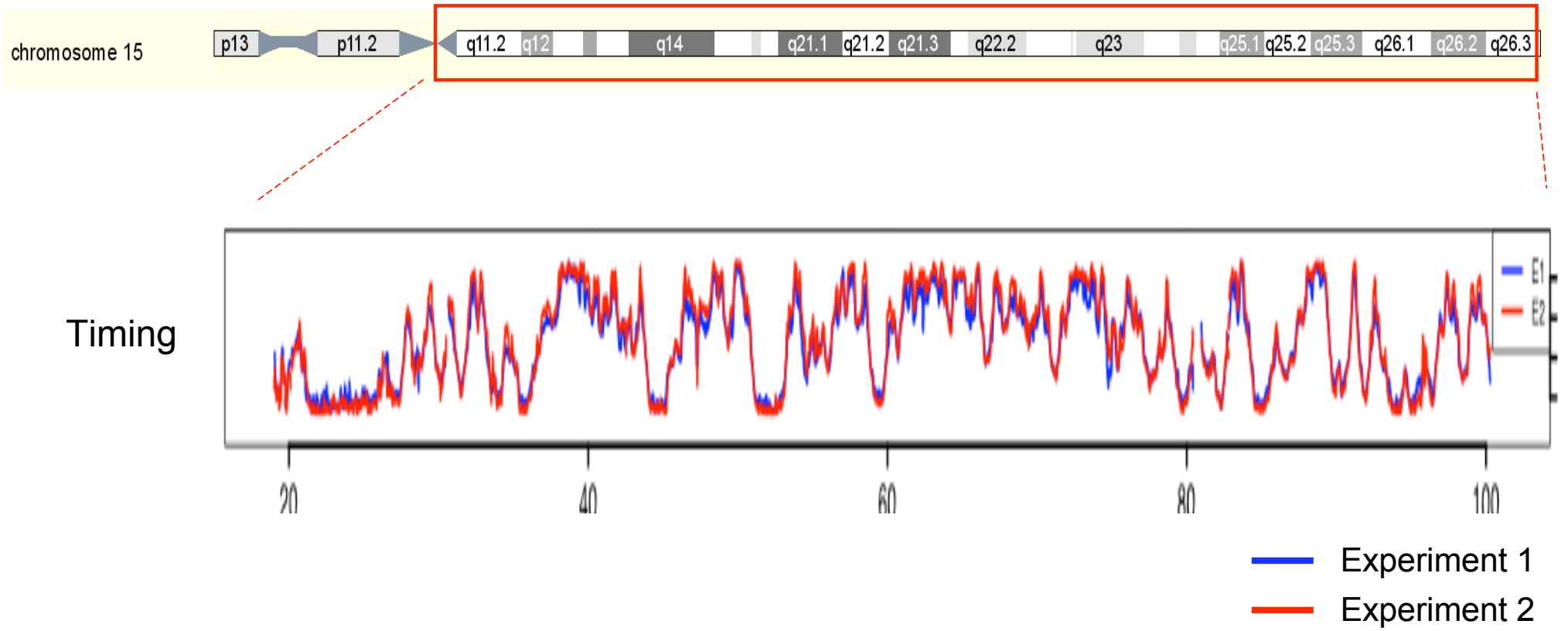
Guillaume Guilbaud
Aurélien Rappailles
Olivier Hyrien
(ENS-Paris)

Benoit Moindrot
Fabien Mongelard
(ENS-Lyon)

Benjamin Audit
Antoine Baker
Alain Arneodo
(ENS-Lyon)

Merci!

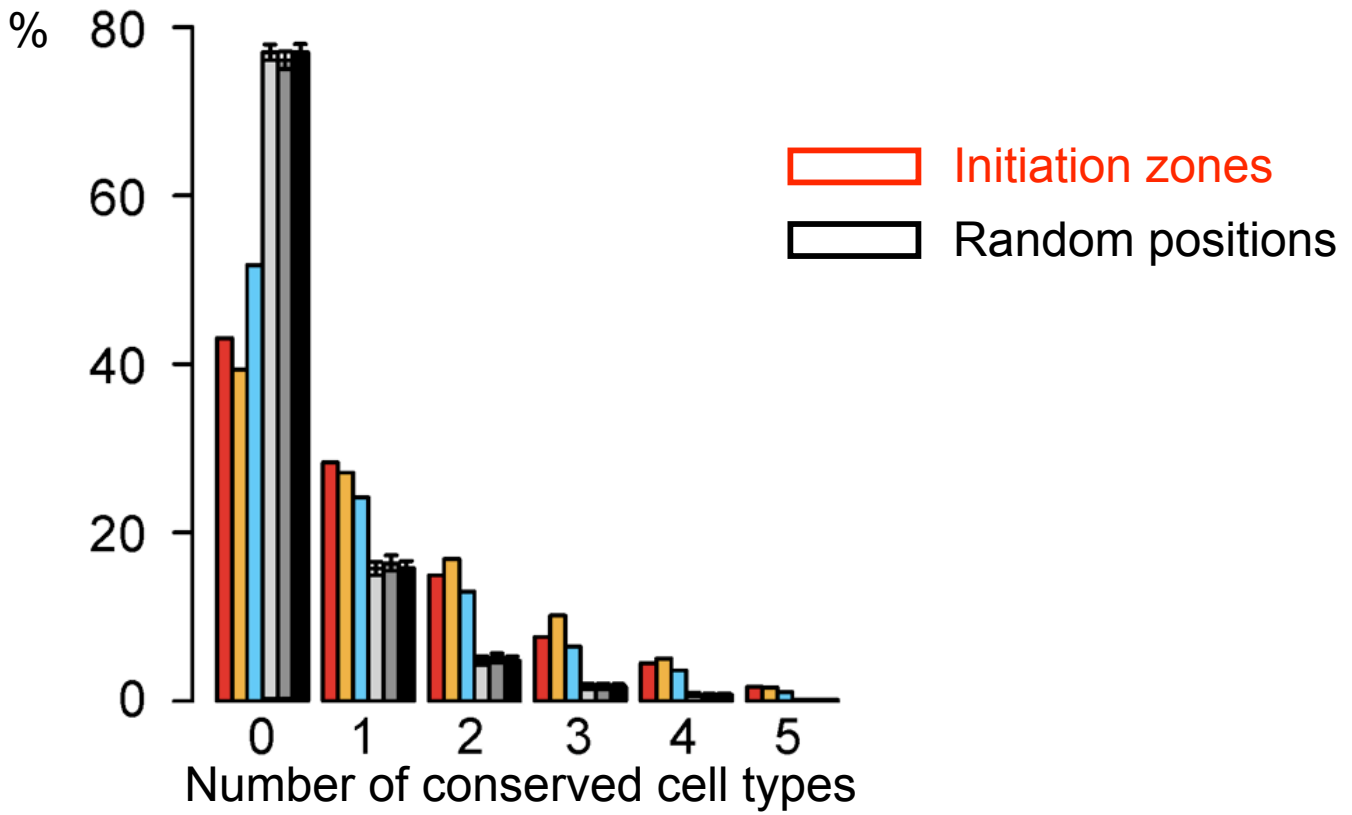
Correlation between two experiments



Replication timing values are highly correlated between two biological duplicates

$R=0.97, P<10^{-15}$

Statistical evaluation : comparison with null distribution of simulation



Replication initiation zones of one cell type are significantly associated with replication initiation zones of other cell types

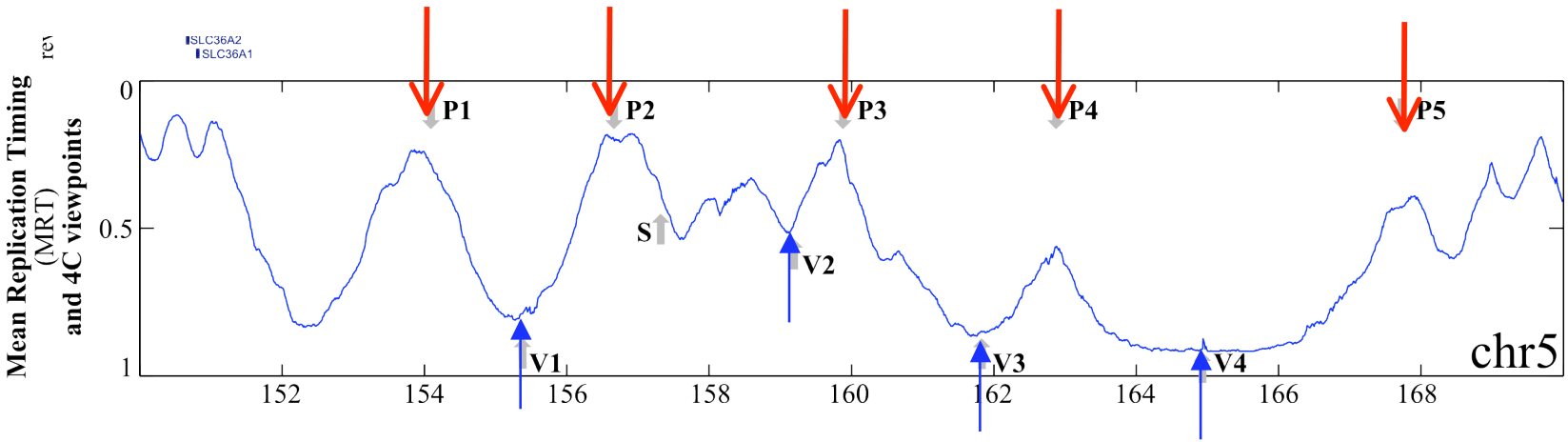
CHEN CL et al. *Mol. Biol. Evol.* (2011) 28:23

3D structure of U-domains

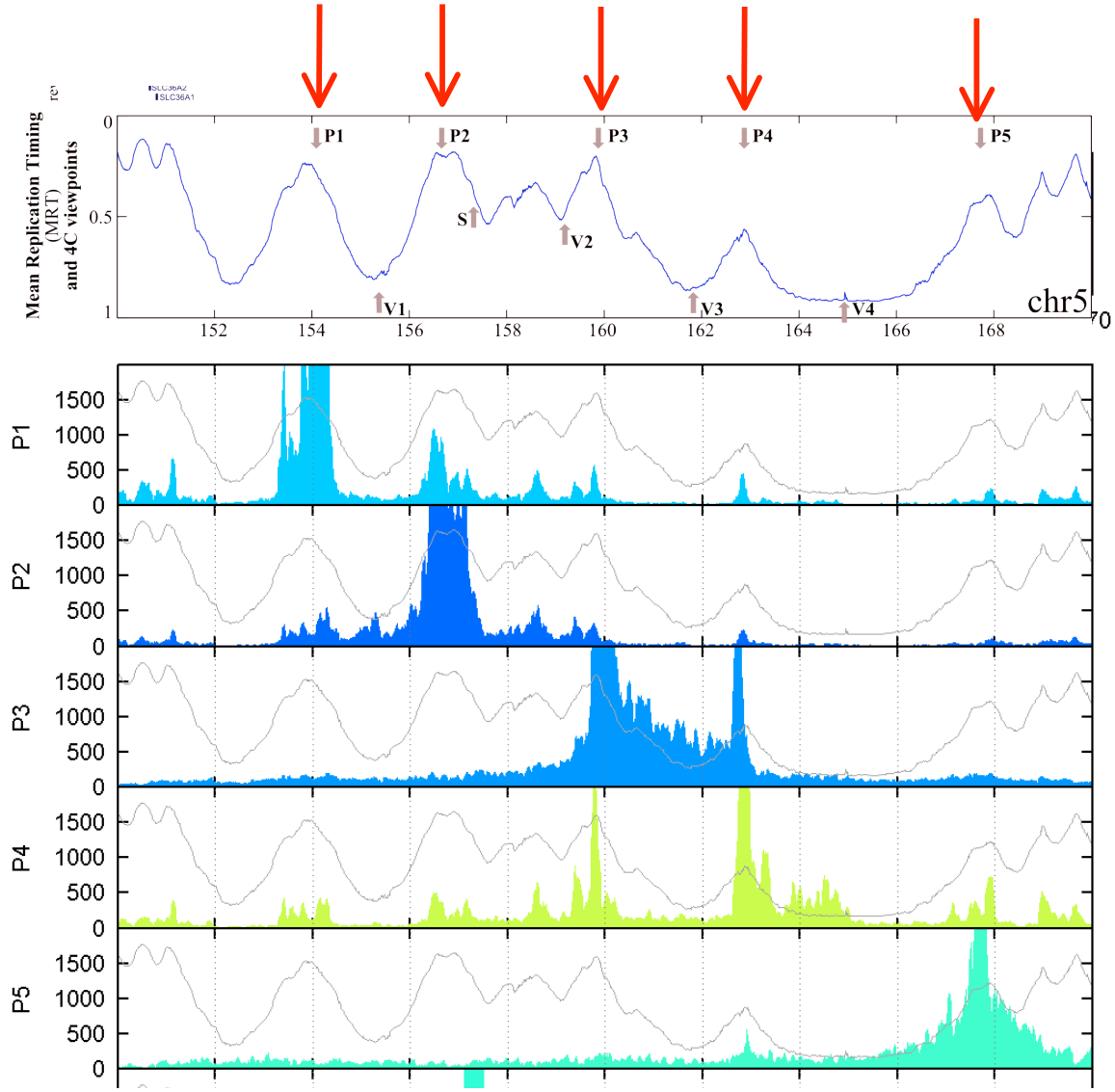
4C experiments (Circularized Chromosome Conformation Capture)

Contacts between DNA fragments

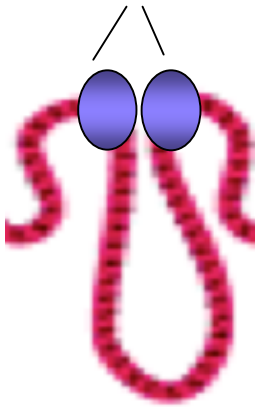
- at N-domain extremities (P1-P5) and other genome regions
- at N-domain centers (V1-V4) and other genome regions



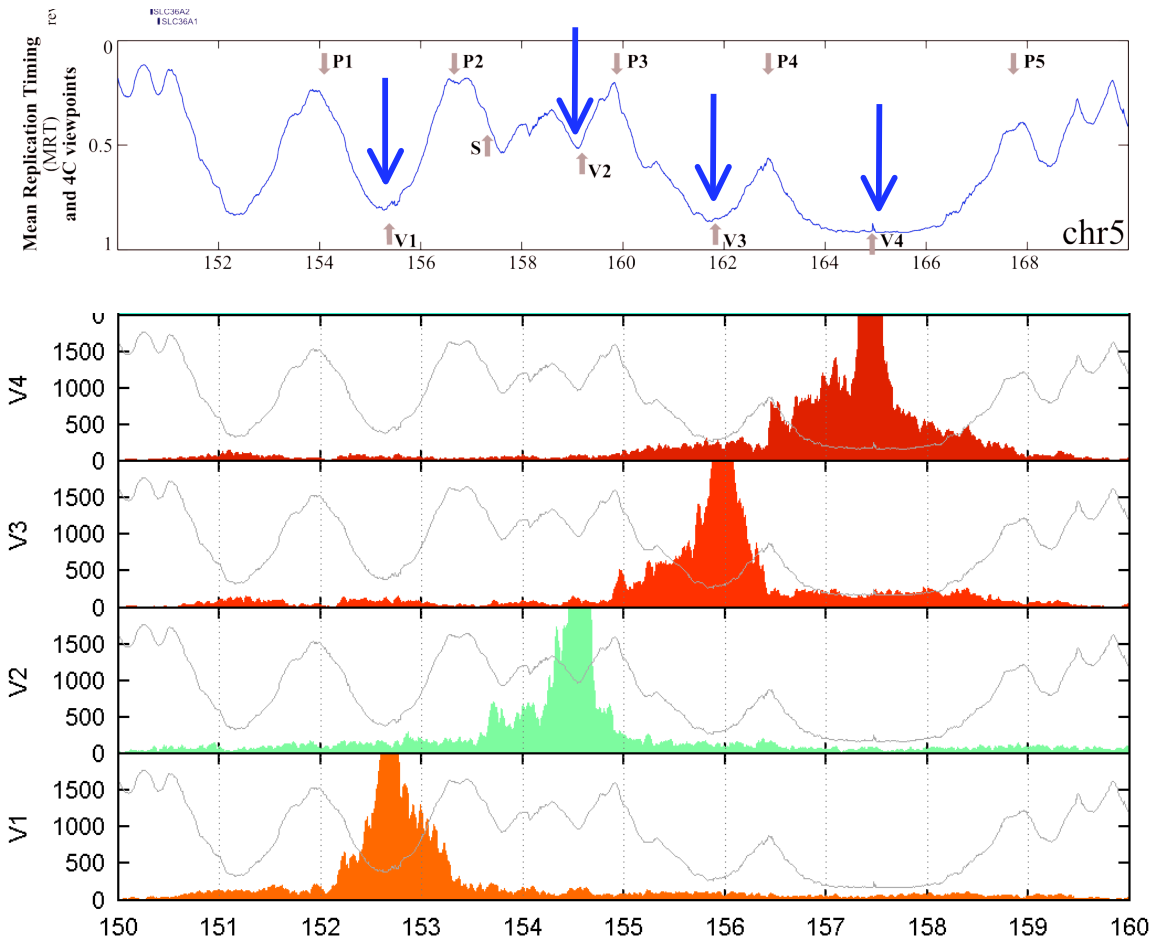
Resting Lymphocytes

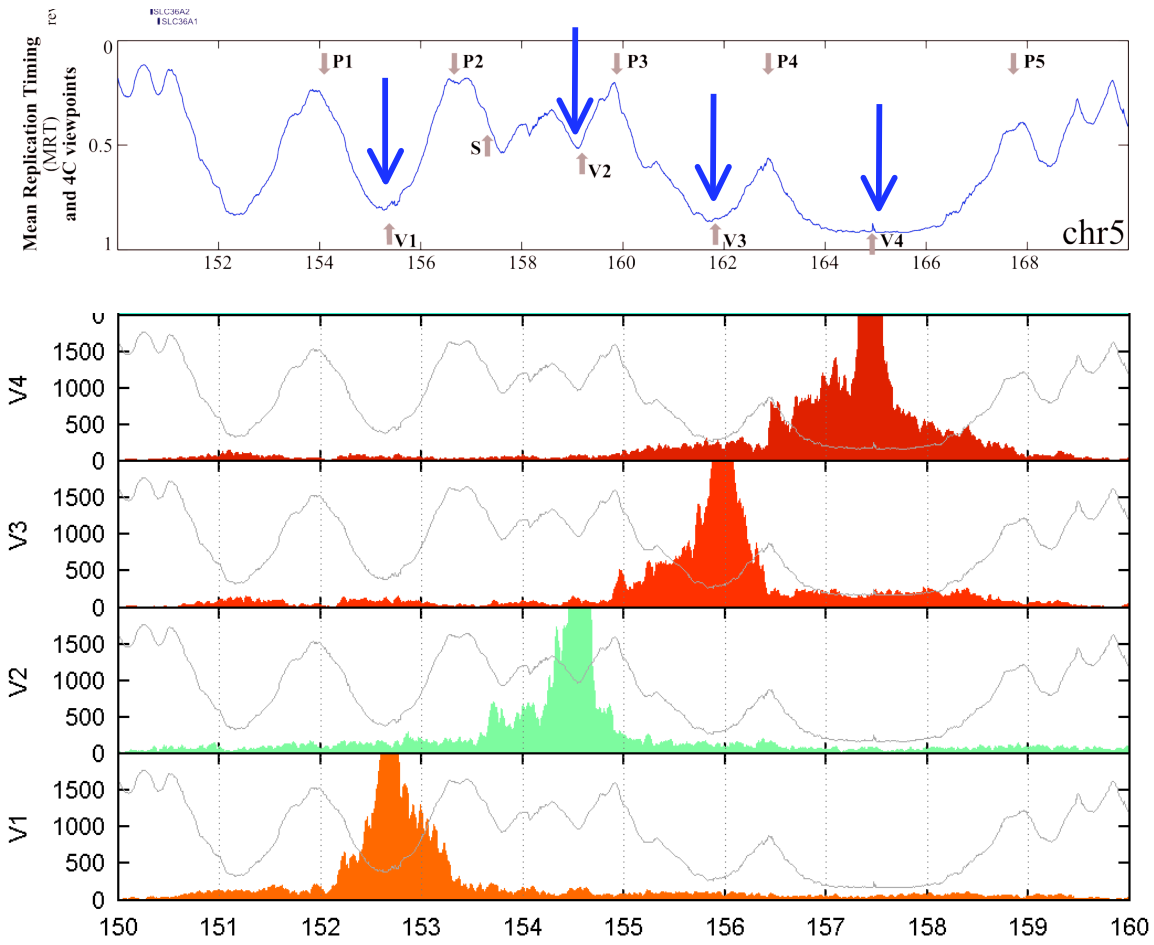


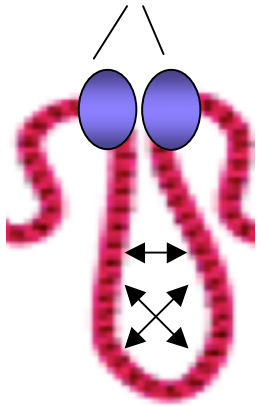
Timing-domain borders



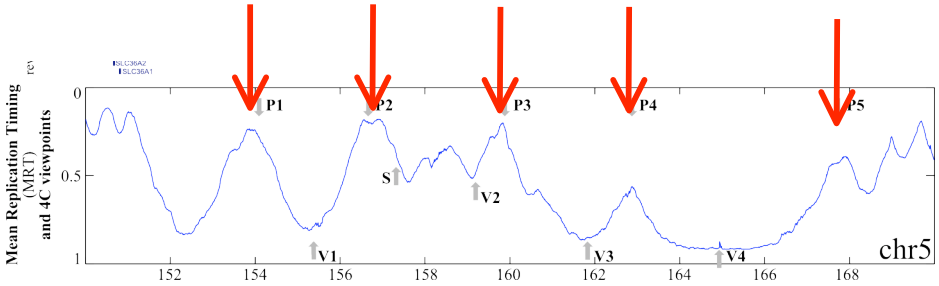
The two domains extremities contact each other

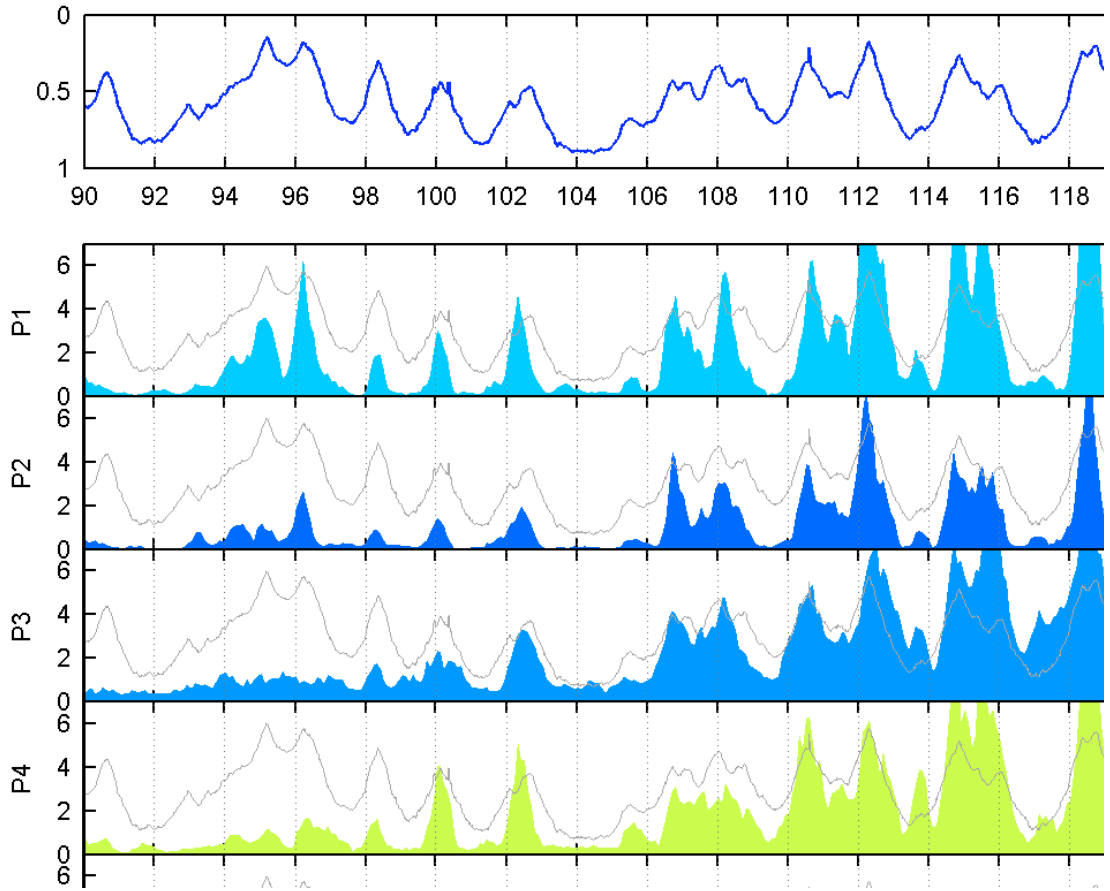

Mean Replication Timing
(MRT)
and 4C viewpoints
0
0.5
1
P1
P2
P3
P4
P5
S
V1
V2
V3
V4
chr5
152
154
156
158
160
162
164
166
168
V4
V3
V2
V1
1500
1000
500
0
150
151
152
153
154
155
156
157
158
159
160





U-domains are self-interacting units



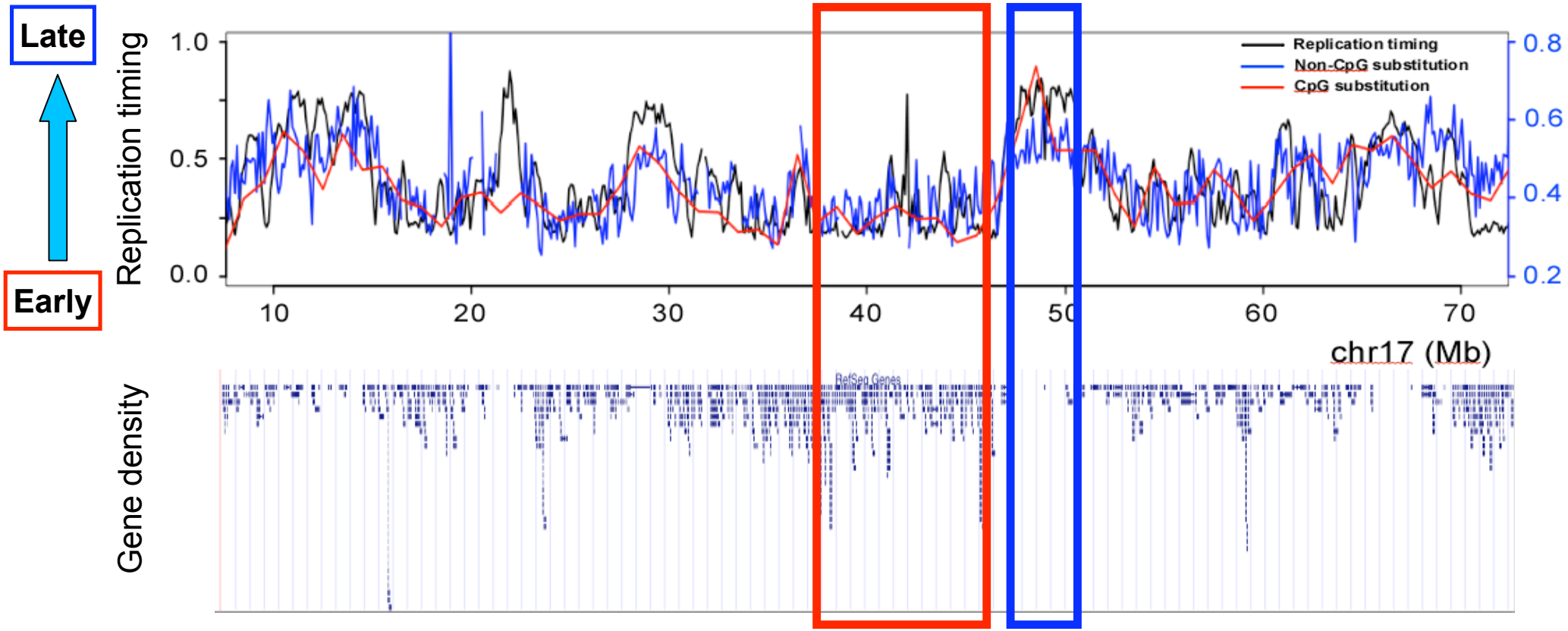


Distant U-domain extremities contact each other at large genomic distance and across chromosomes

↓

higher order chromatin organization

Impact of replication on evolution and organization of the human genome



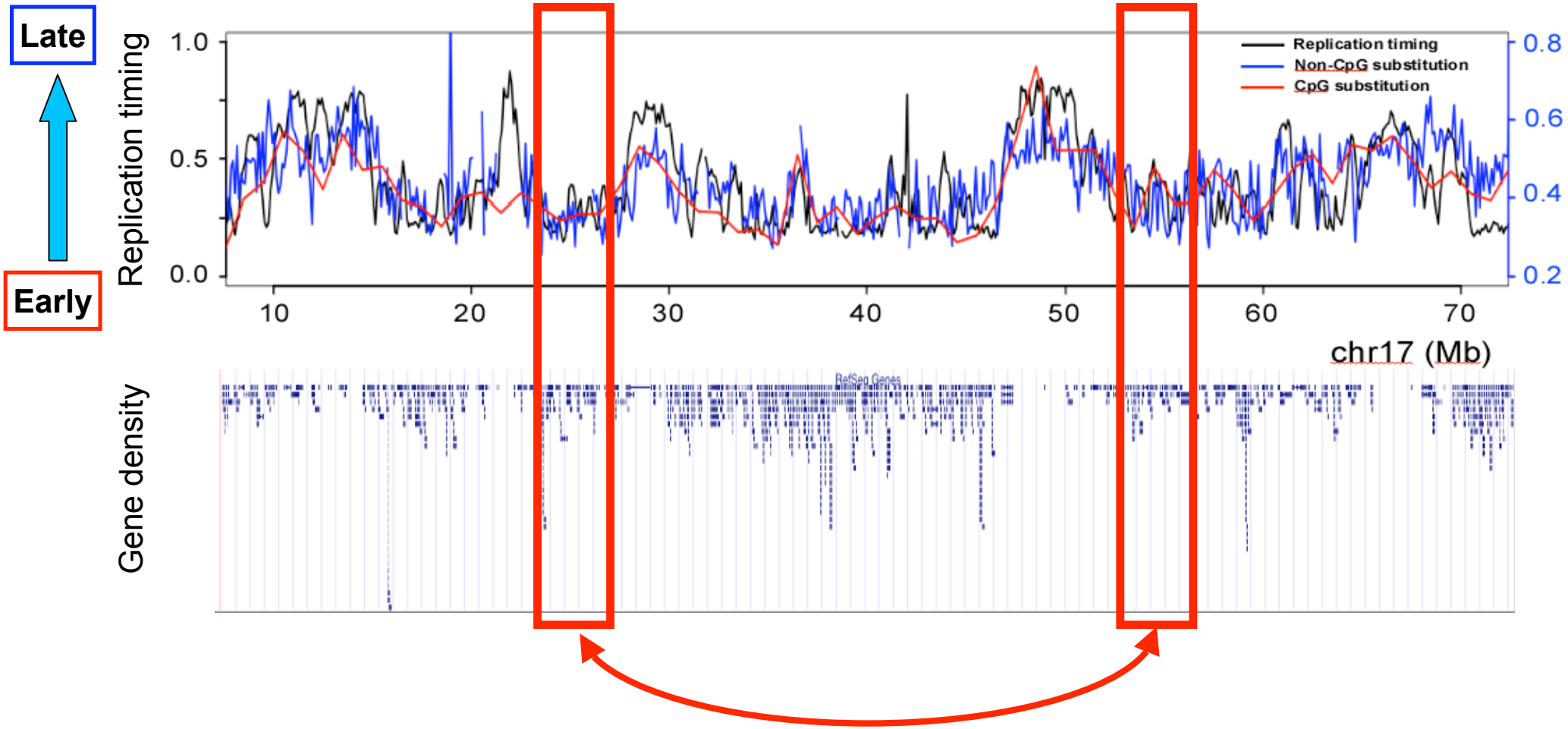
Early replication **Late replication**

- High GC content
- High gene density
- Housekeeping gene
- Low methylation
- Low substitution rate

- Low GC content
- Low gene density
- Tissue-specific gene
- High methylation
- High substitution rate

Watanabe et al. *Hum. Mol. Genet.* 2002
Woodfine et al. *Hum. Mol. Genet.* 2004
Woodfine et al. *Cell Cycle.* 2005
Huvet et al. *Genome Res.* 2007
Stamatoyannopoulos et al. *Nat. Genet.* 2009
Chen CL. et al. *Genome Res.* 2010
Chen CL et al. *Mol. Biol. Evol.* 2011
.....

Impact of replication on evolution and organization of the human genome



Rearrangements are more frequently between regions with close replication timing
long-range chromatin origination associated with replication program

Lemaitre et al. *BMC Genomics.* 2009
Yaffe et al. *PLoS Genet.* 2010
Ryba et al. *Genome Res.* 2010
.....