# Master replication origins at the heart of the organisation and fragility of the human genome

Alain Arneodo

*Laboratoire Joliot-Curie / Laboratoire de Physique*
*Ecole Normale Supérieure de Lyon*
*46 allée d'Italie, 69364 Lyon Cedex 07, France*
**Alain.Arneodo@ens-lyon.fr**

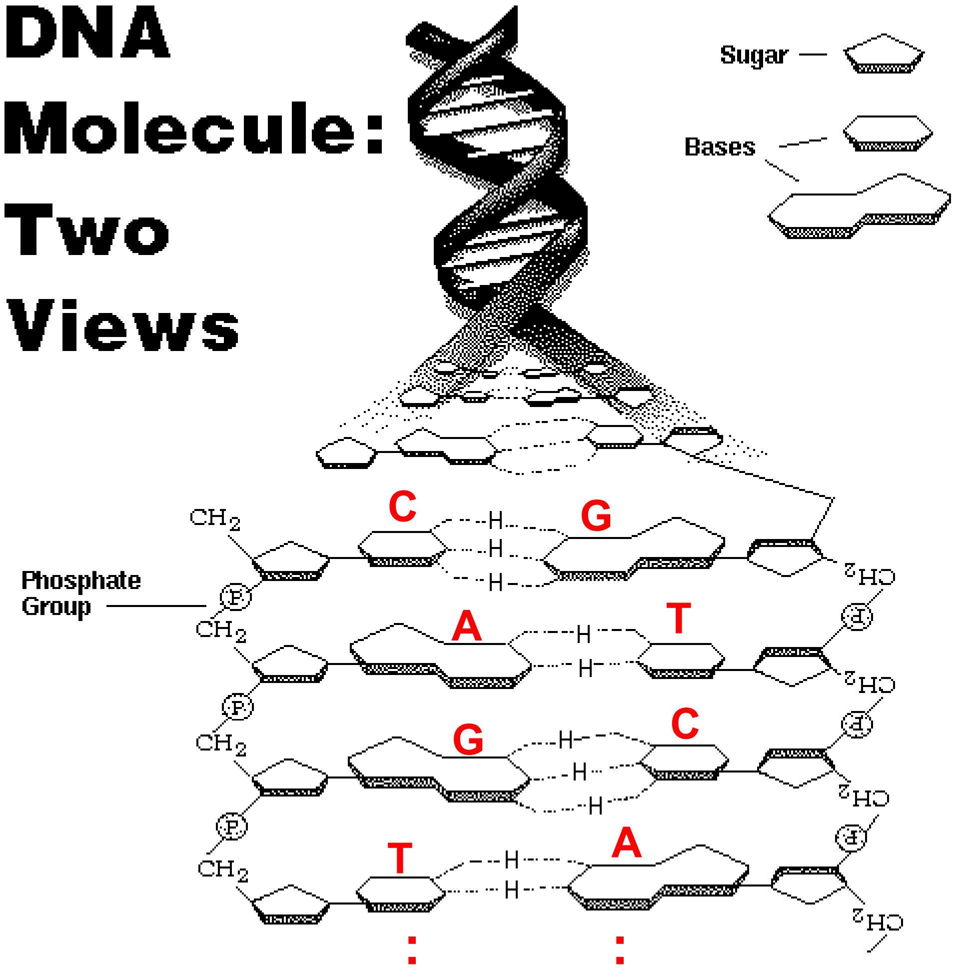
Benjamin Audit
Edward-Benedict Brodie of Brodie
Samuel Nicolay
Philippe Saint-Jean
Cédric Vaillant
Lamia Zaghloul

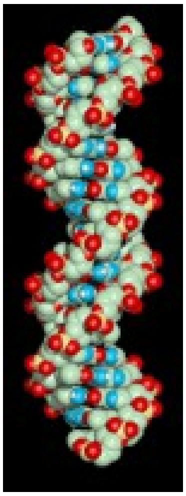
ENS de Lyon, France

Yves d'Aubenton-Carafa
Maxime Huvet
Claude Thermes
Marie Touchon

CGM, Gif-sur-Yvette, France

# DeoxyriboNucleic Acid



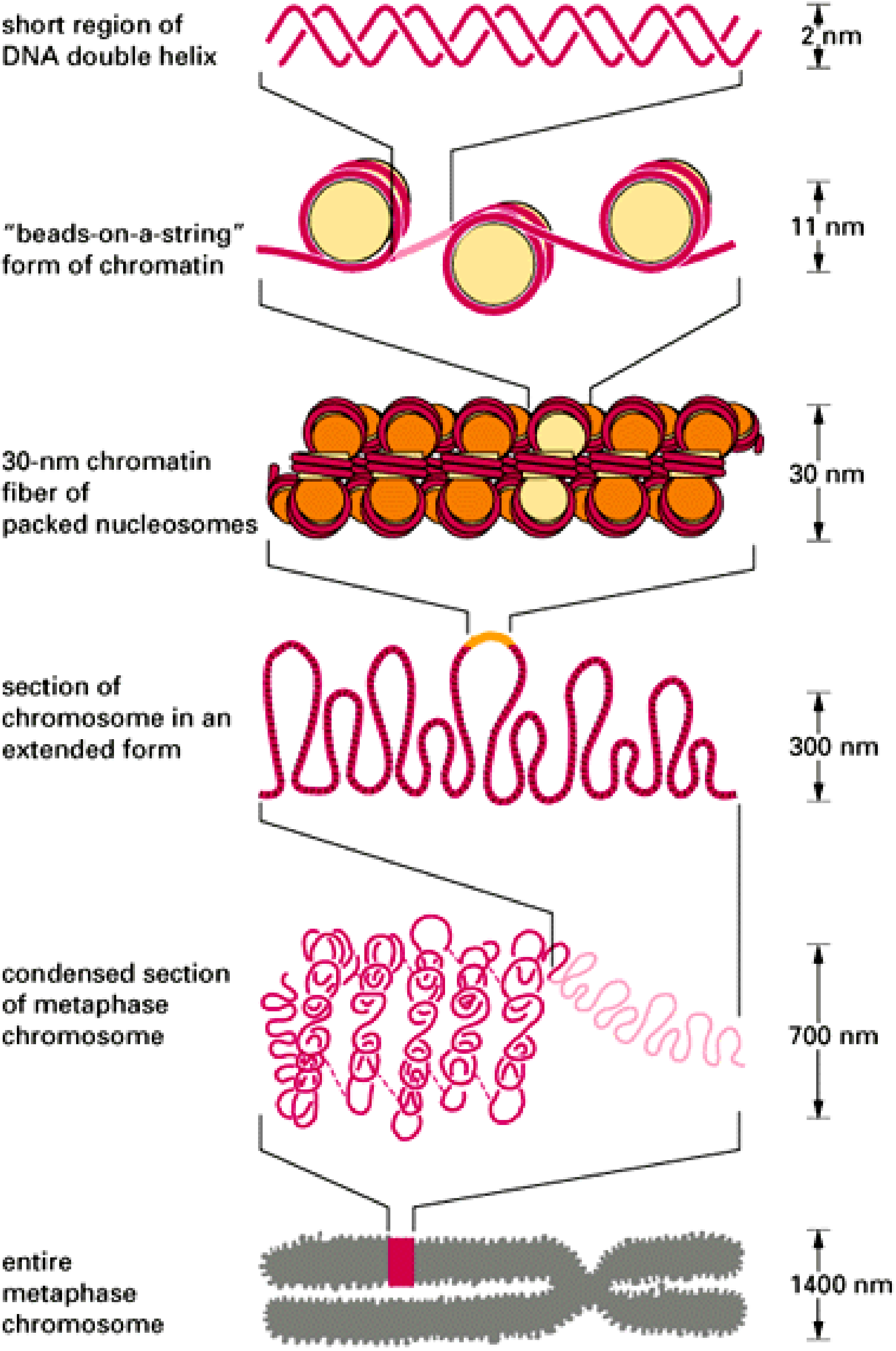


- Double helix macromolecule
- Each strand consists of an oriented sequence of four possible nucleotides:
  Adenine, Thymine, Guanine & Cytosine
- Complementary strands:
  [A]=[T] & [G]=[C] over the sum of both strands

# Sequencing projects result in 4 letter texts :

```
gtcagtttcctgaggcgggtcgggacccaggcgtgagactggagtctgcc
caggggcccagctgagccagcctcctcgtcagctgcttgggccgccagga
cgccgccgggggtgcgccgcgcttccctggatggggtgcccccactcccc
tcggagccccagggagacccccgaactcagctcctctcaggggtgccag
ggggacccctcaaactccactccccgcaggttcctggggagacgcccct
gctcgattcccctcagggtcccagggagacccccctaattcagctcctctc
aggggtactgggggacctctcgagctccactcccatcagggtcccaggga
gacccccaactatgctcaggggtcccagggagatgccagcaccccaact
ccgcttccctgggccccccctccccttacagctcaacttccctcgagagt
ctggggctggggctccgttcagttcttgagtccccttccctcggggtgtc
ccgggccgcccacccccacactgtctgtgattccccaaggcgcgggtct
cgggccgcagcctgttccacgttctgctgctcgttcttttctggctcctt
gctttcgaaggagagaaggaggccttcgtttccagtcttttgcctttttc
taatggagccctgcttttccttccgtgtcccttcaggctacttctgccag
gtttctatttttcattctttattatgacttcgcccaaaatattcttgact
tctattgagaaggattcgggggtctatttcttattcggaggcgtgtgctt
aagttccaaacagatgaggattttccagttaatccttctggggtgactta
ttgcttaatgccaccatagccagaaaatggactctcagtgtccgaaactg
cattcggctctgaagtgtctgtccttgtcacctcttgcaatgtttcgcgg
cgggaagcctgcactcgccgacgctgacgtaactgtttctgtctttcagg
tctacagcctcctgtgggtgggcgatattgacatatactttatttctata
tatgttatgaactcaatatttcttgcagcgggtctgctgataataagata
tgcctactctgcgagtctggaagccatcttaagcttaccctgtatgtgcc
ccatgcatctcttccgttacacggctcctgagttgacacctgtgtgataa
actggtaatagcaagtaaactgttttcttgtgctctgtaagctgctctag
caaattatctaggaggaggtggtcttggaaacccctgatttataagcggg
cagtcagcagtacacgtggcccagaatcgtgattggcatttgaagtgggg
gcagtagggtgggactgagcccttcacctgtggggtctgccctgctcaag
gcagtgtcagaattgaagtgaaatgttggacggtcggtgtccagagagtt
ggagaactggtttgtgtgtaaaaactnacatatttagggtcagaagtatg
                         ...
```

# HIERARCHICAL STRUCTURE OF EUCARYOTIC DNA



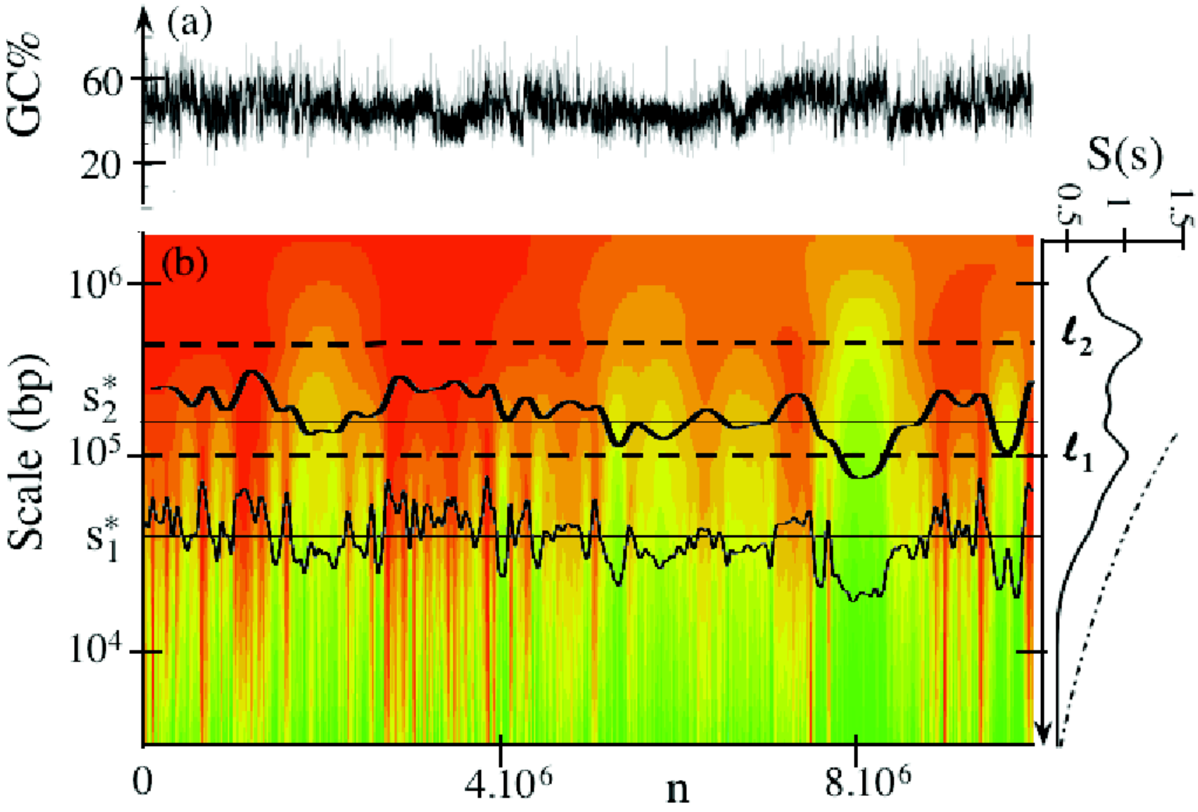
**NET RESULT : EACH DNA MOLECULE HAS BEEN PACKAGED INTO A MITOTIC CHROMOSOME THAT IS 50.000x SHORTER THAN ITS EXTENDED LENGTH**

# LARGE SCALE REPRESENTATION OF GENOMIC SEQUENCES

Nicolay, Phys. Rev. Lett. (2004)

## Space-Scale Representation of the GC Content with a Smoothing Gaussian Filter

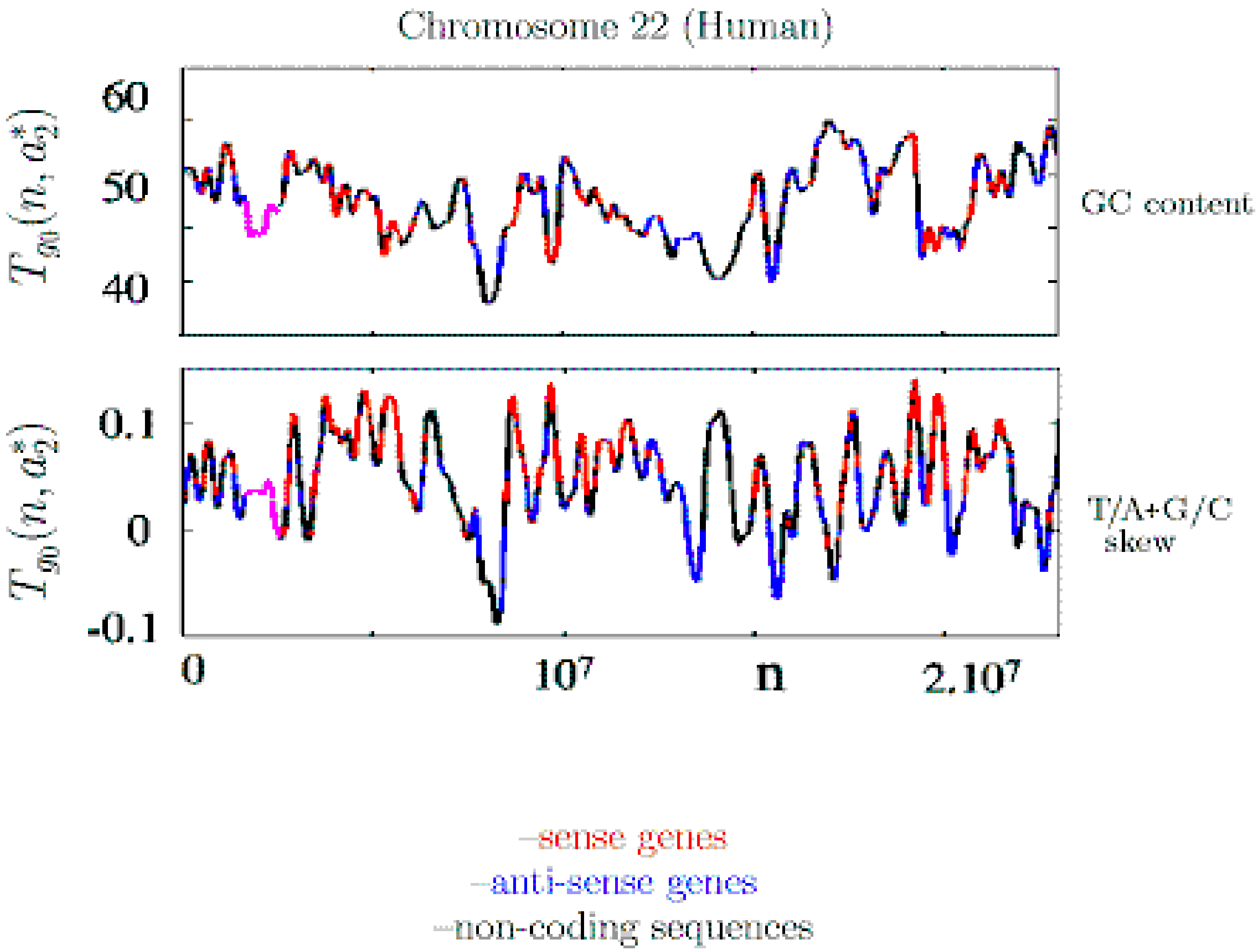
Chromosome 22 (Human)



Filtering scales: $a_1^* = 40$kb, $a_2^* = 160$kb

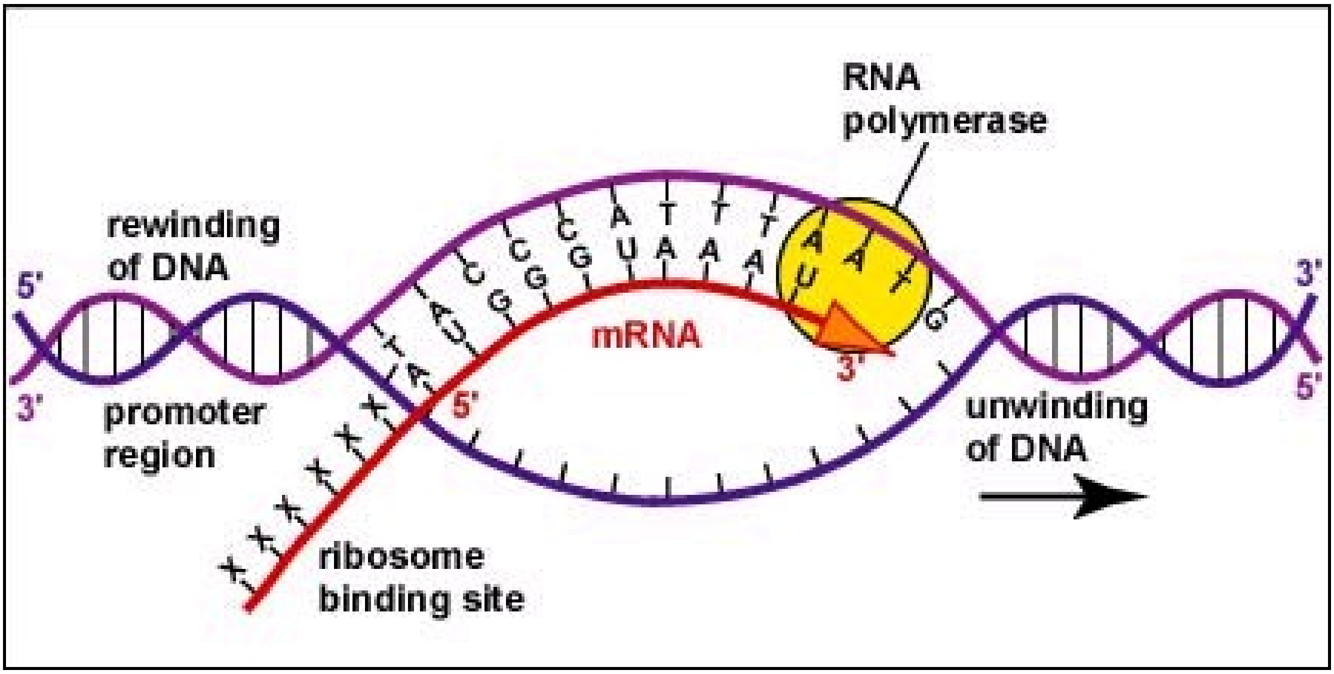
Space-scale content: $S(a) = \sum_n |T_{\psi_M}(n,a)|$,
where $\psi_M$ is the Morlet wavelet

# Strand Compositional Asymmetry

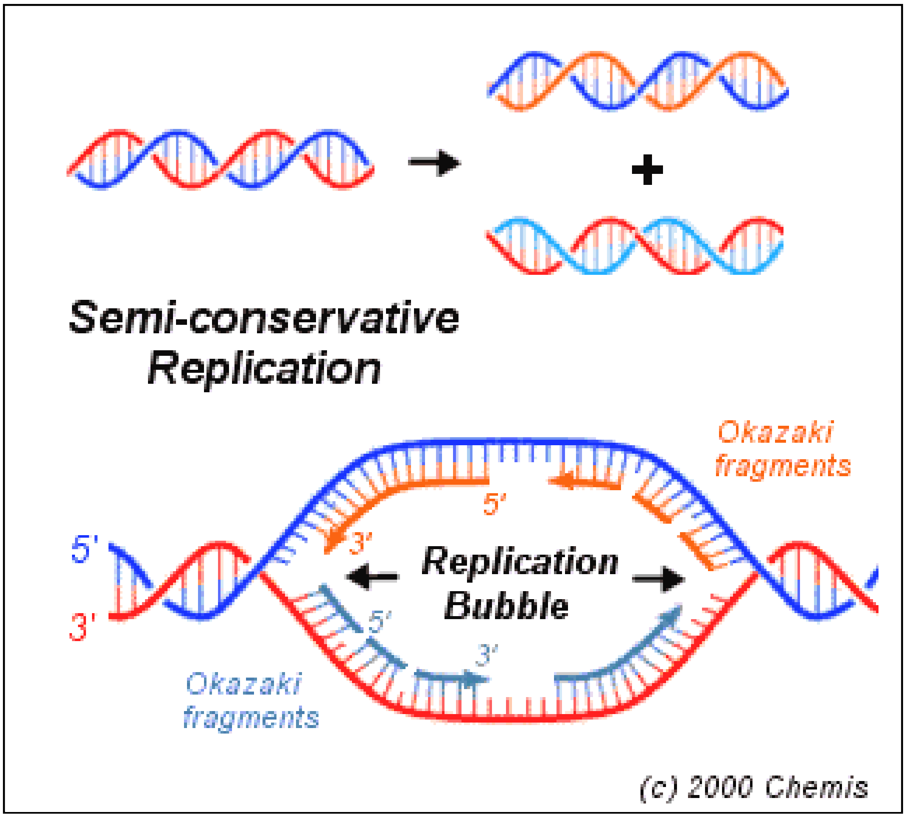


Filtering scales: $a_1^* = 40\text{kb}$, $a_2^* = 160\text{kb}$

# Transcription



# Replication



*Opening of the double helix with a different environment for each strand => asymmetrical process*

# Symmetrical properties of the strands: “Parity Rule type 2”

$$[A] = [T] \quad \& \quad [G] = [C]$$

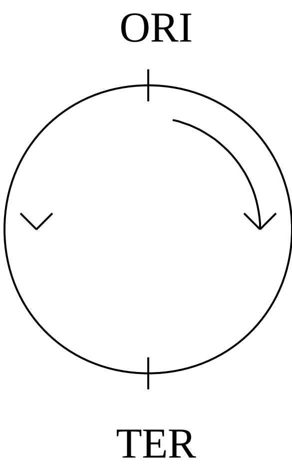
***in each strand***

Deviations from this property estimated by the compositional skews

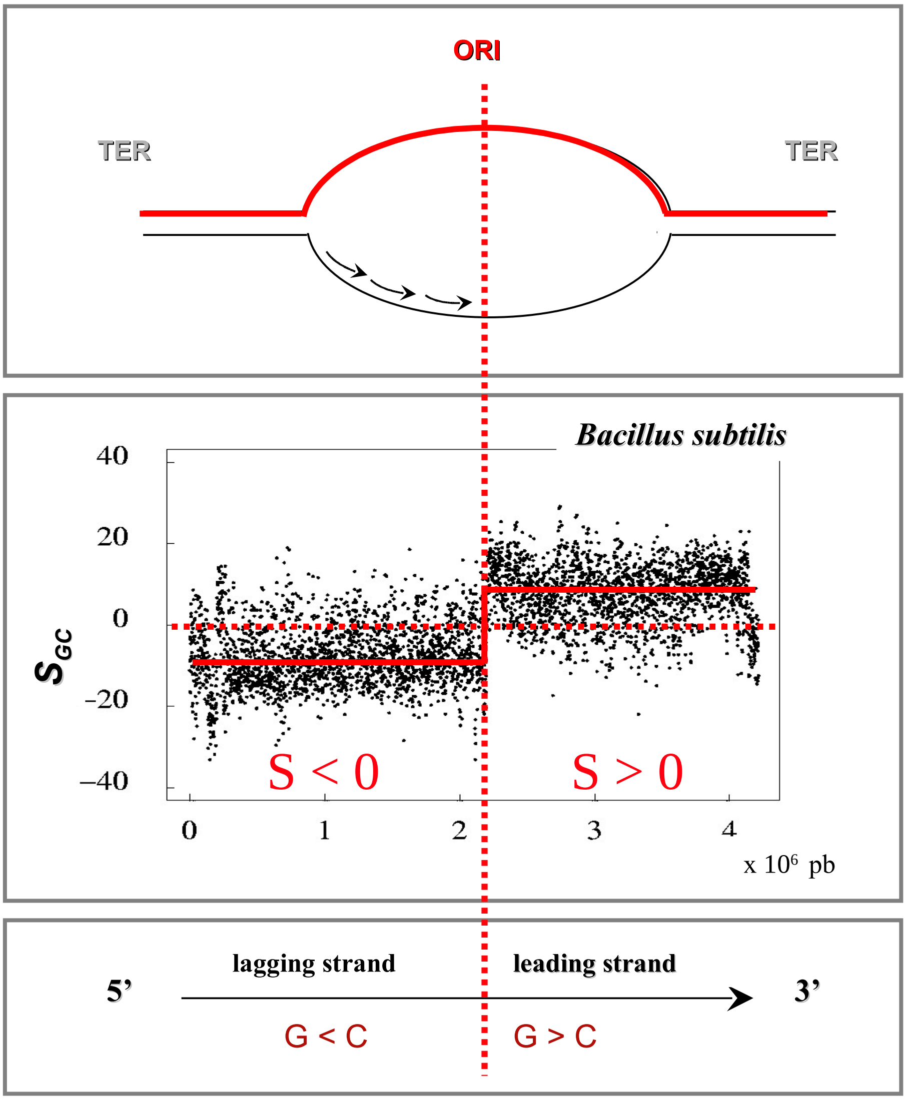
$$S_{CG} = \frac{[C] - [G]}{[C] + [G]}$$

$$S_{AT} = \frac{[A] - [T]}{[A] + [T]}$$

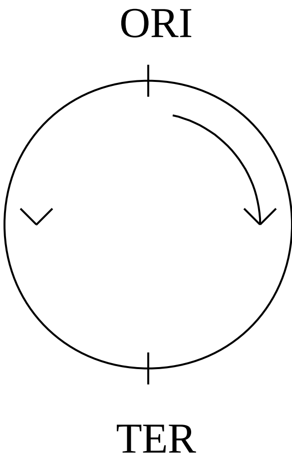
***Compositional skew due to local biases in a strand in the course of biological mechanisms***

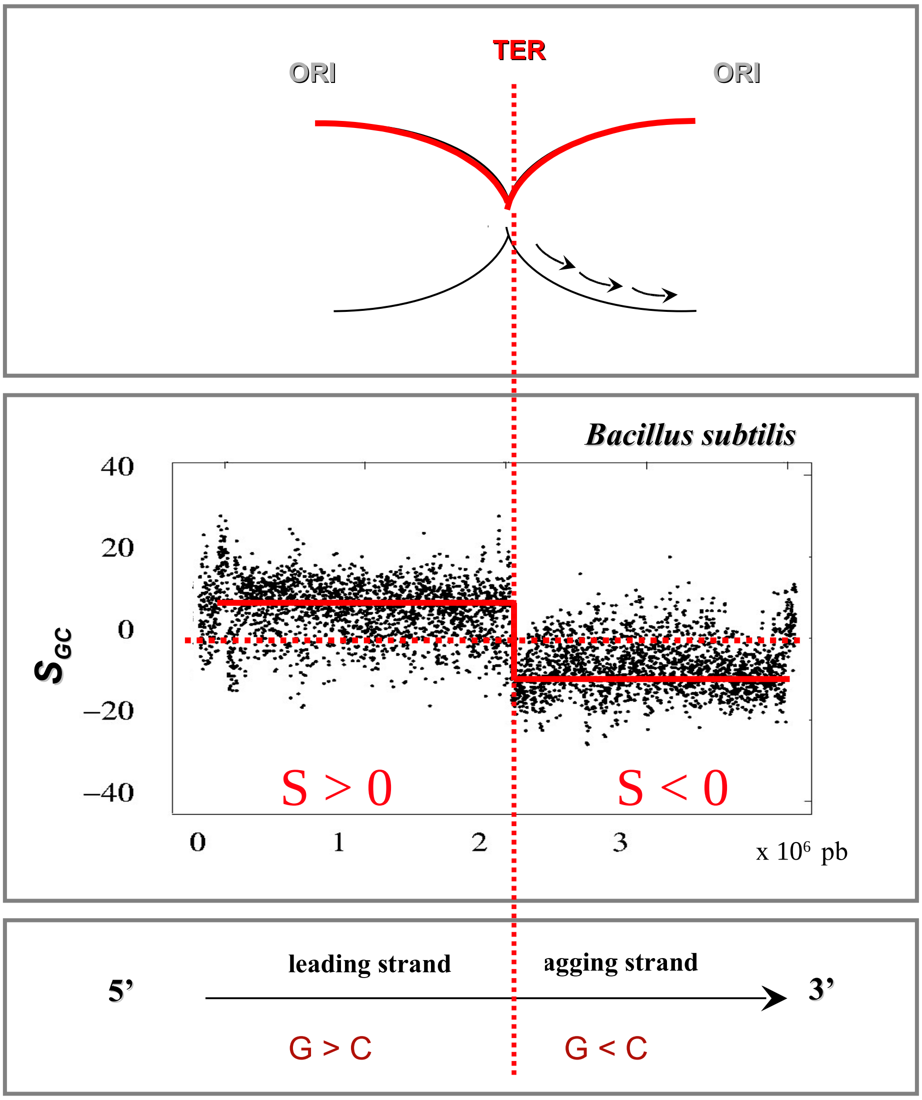


# Replication bias in bacteria:
# An upward jump at the origin



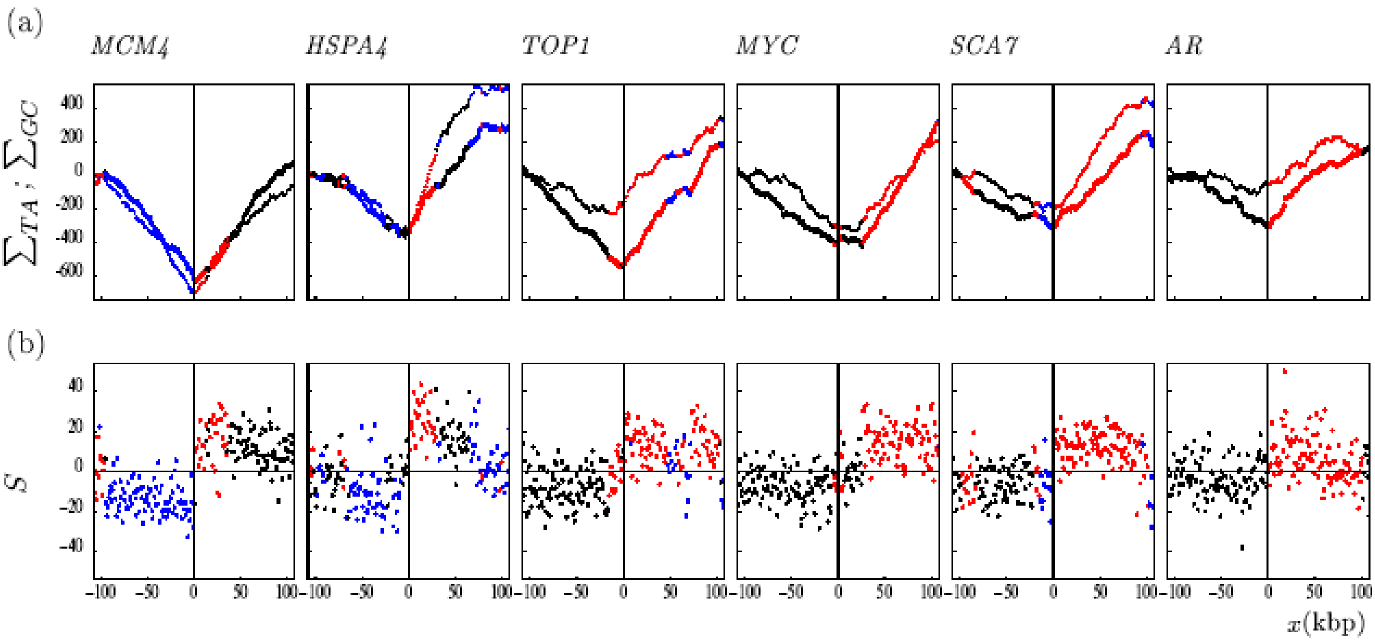
# Replication bias in bacteria:
# A downward jump at the terminus





# Skew profiles around human replication origins

Brodie of Brodie, Phys. Rev. Lett. (2005)
Touchon, PNAS (2005)



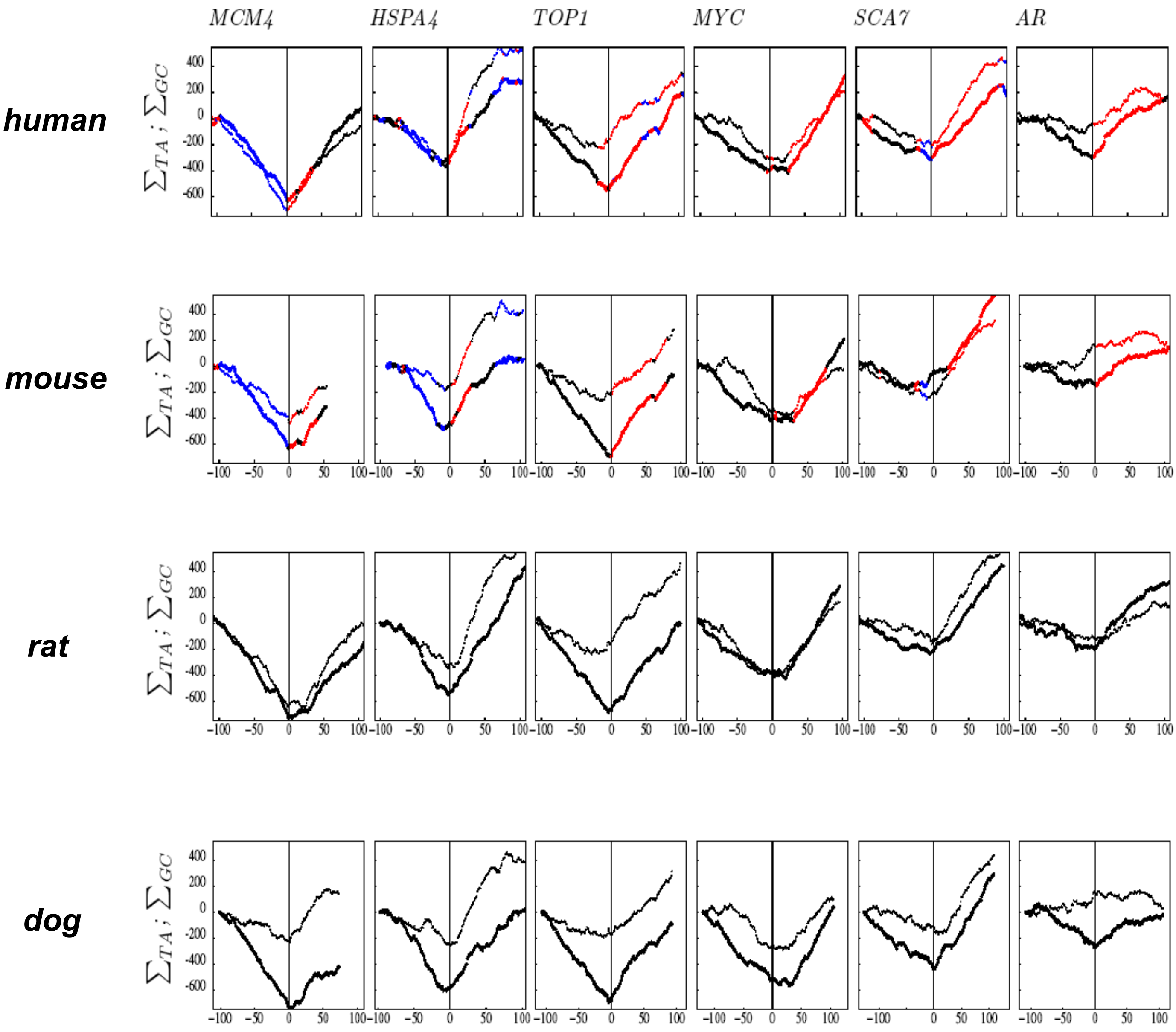
**upward jumps : $\Delta S$ ~ 24%**

**Can these profiles be explained by transcription only ?**

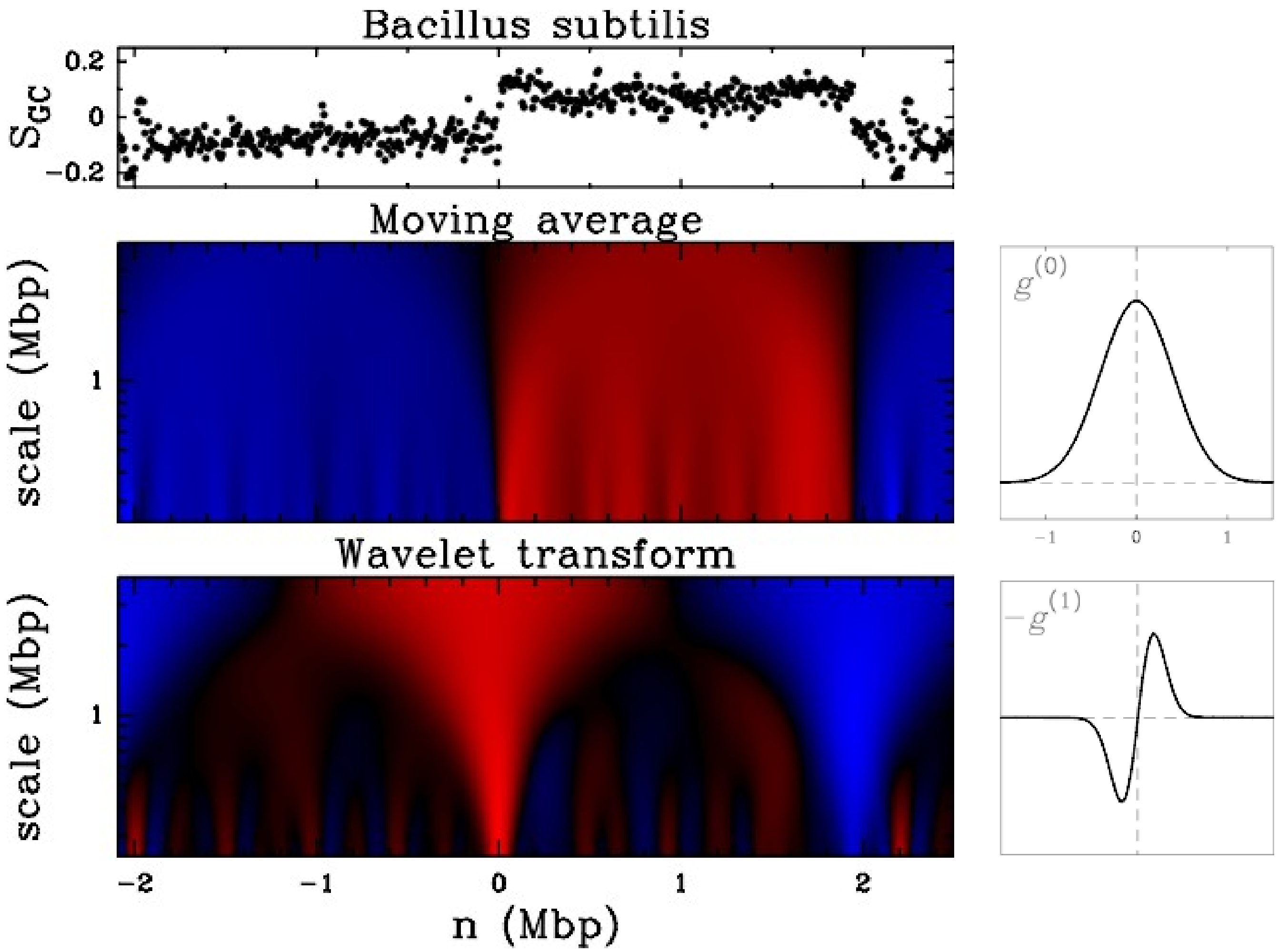
**Do intergenic regions show upward transitions of the skew *S* ?**

# Conservation of profiles in mammalian genomes

Touchon, PNAS (2005)

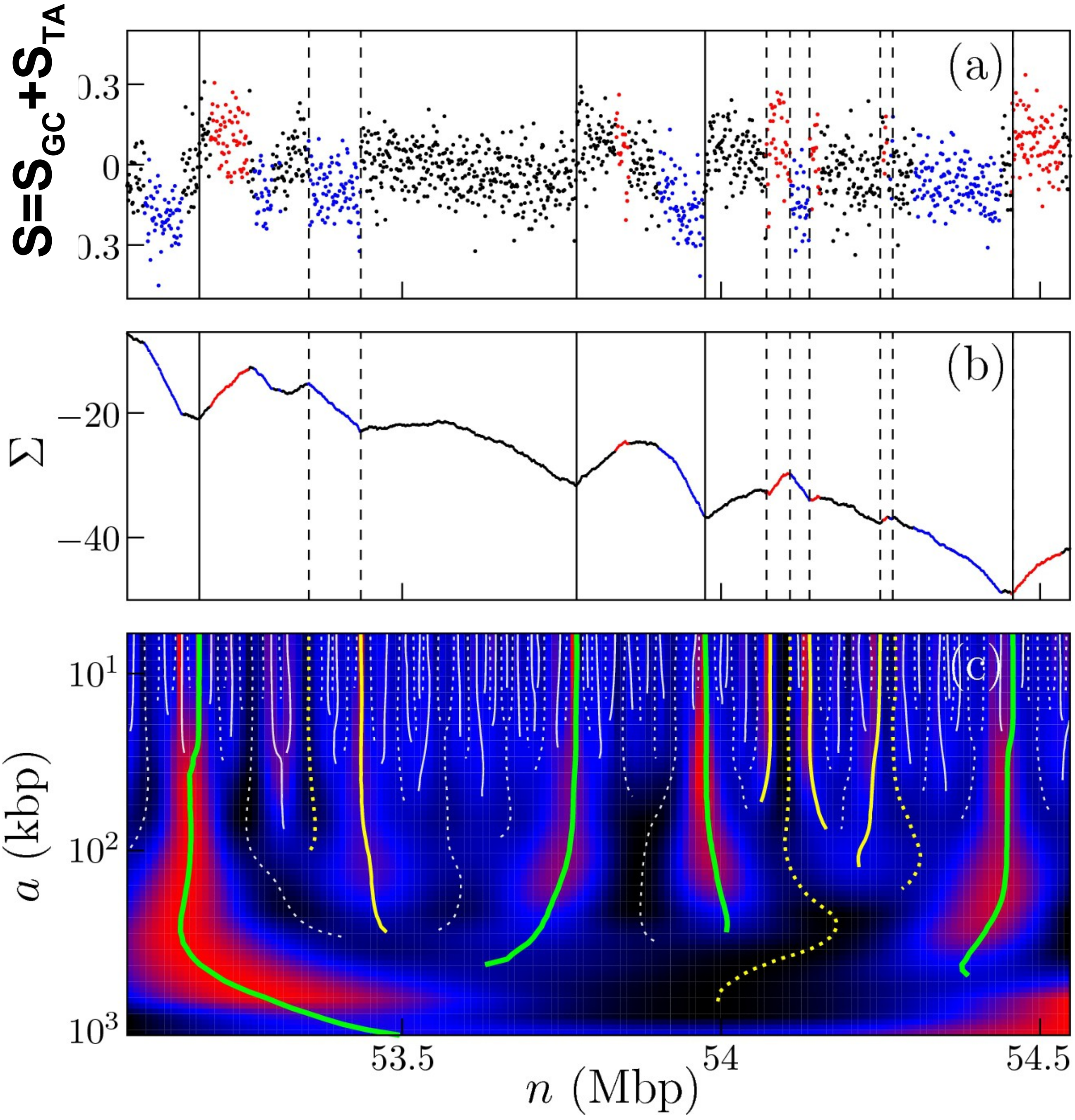


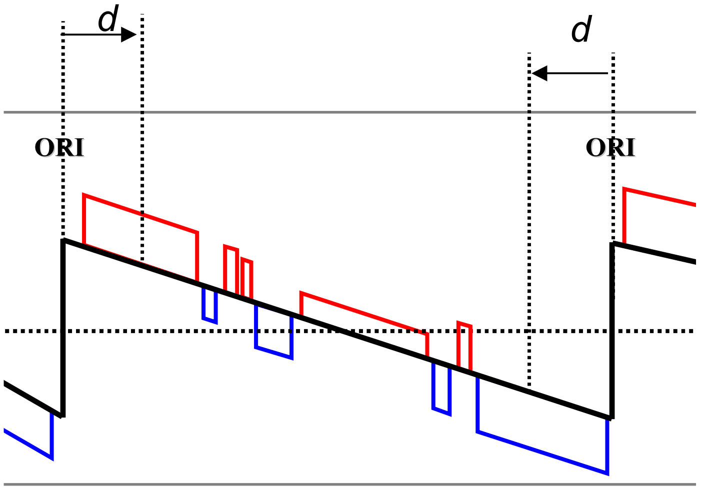
# Multi-scale detection of jumps in skew profiles using the wavelet transform



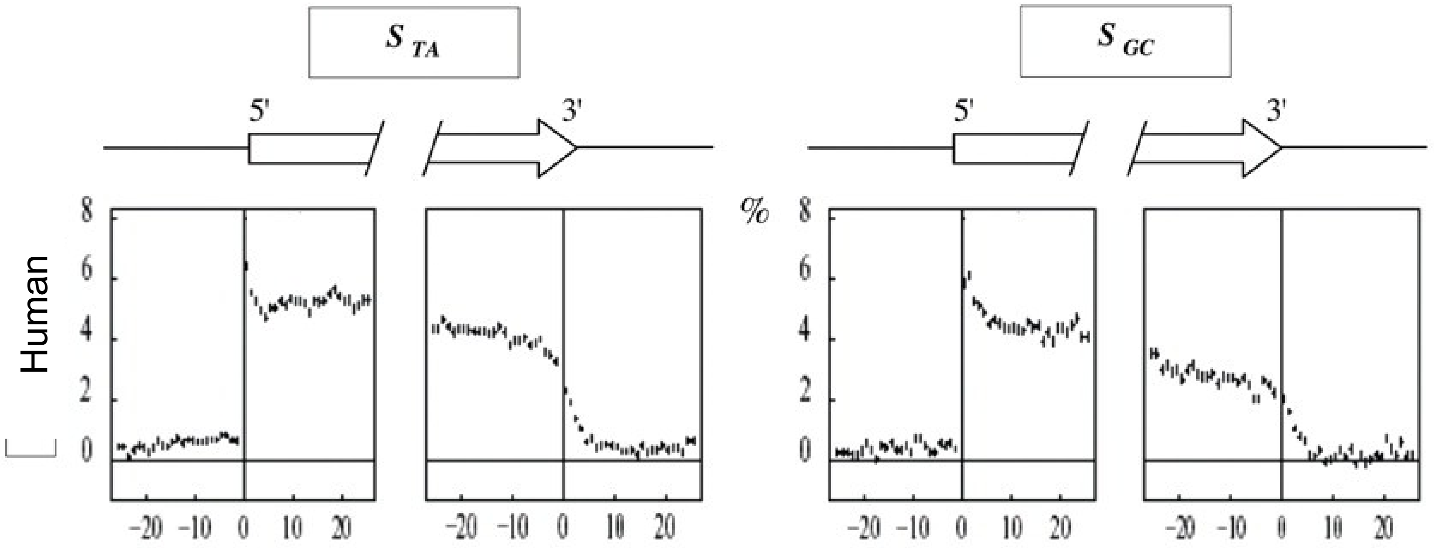
# Wavelet transform of the skew profile

## Human chromosome 6



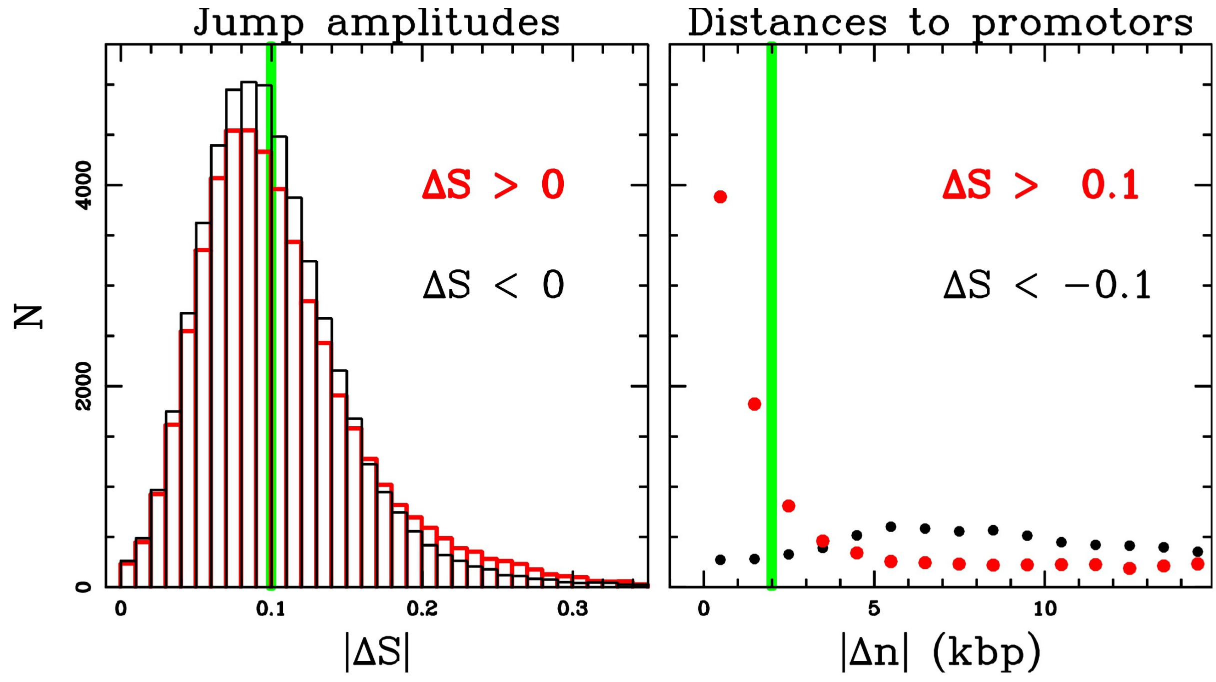


# Transcription bias



# Statistics of upward and downward jumps at scale 10 kbp

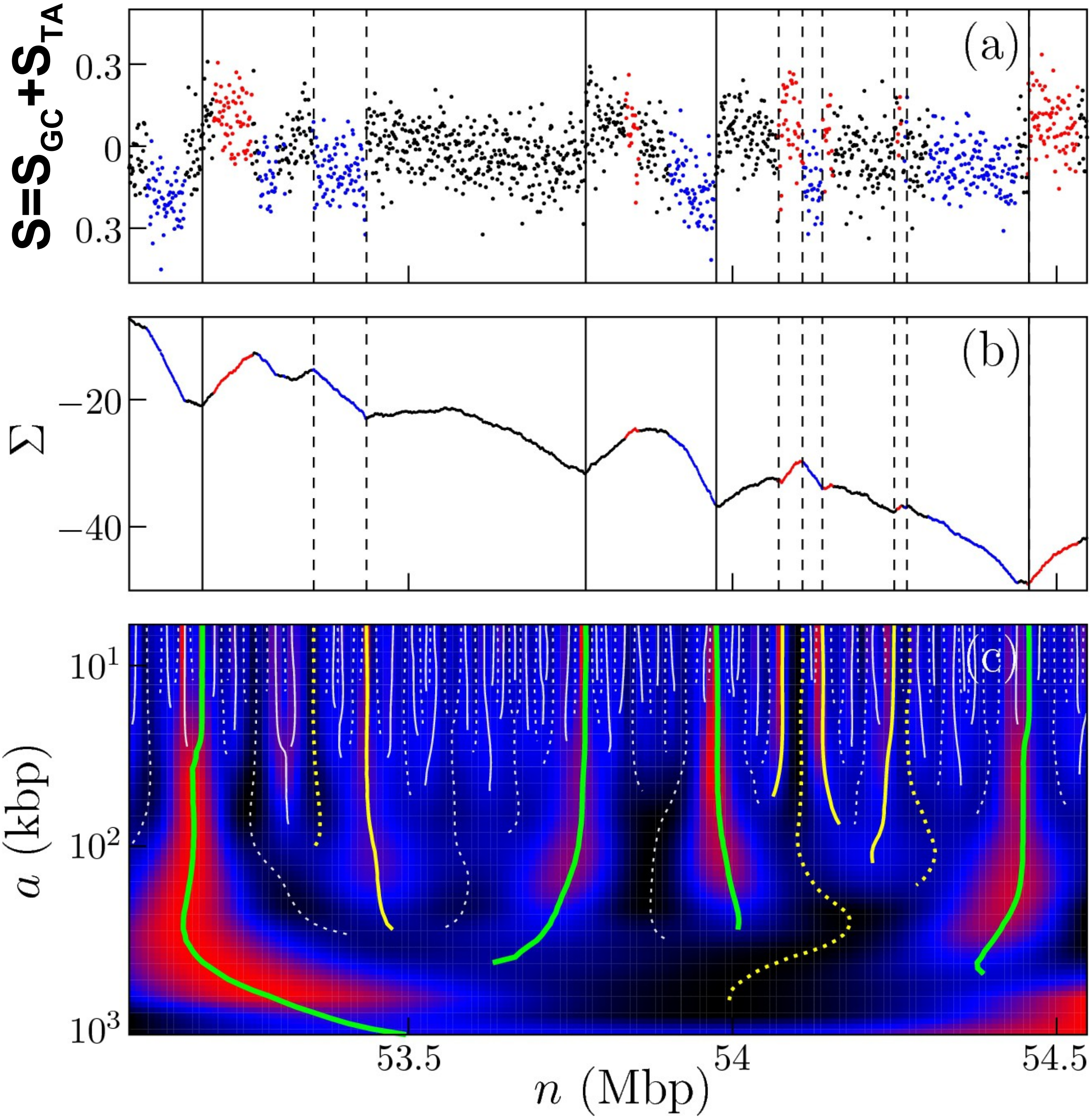
Nicolay, Phys. Rev. E (2007)



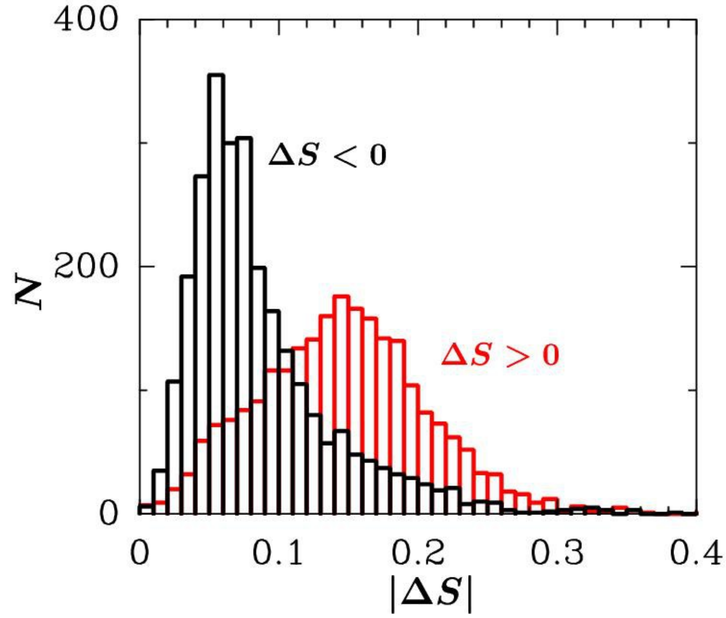
$\Delta S > 0.1$, $\Delta n < 2kbp$: **36%** (7228/20023) significantly biased human genes likely expressed in germ line cells.

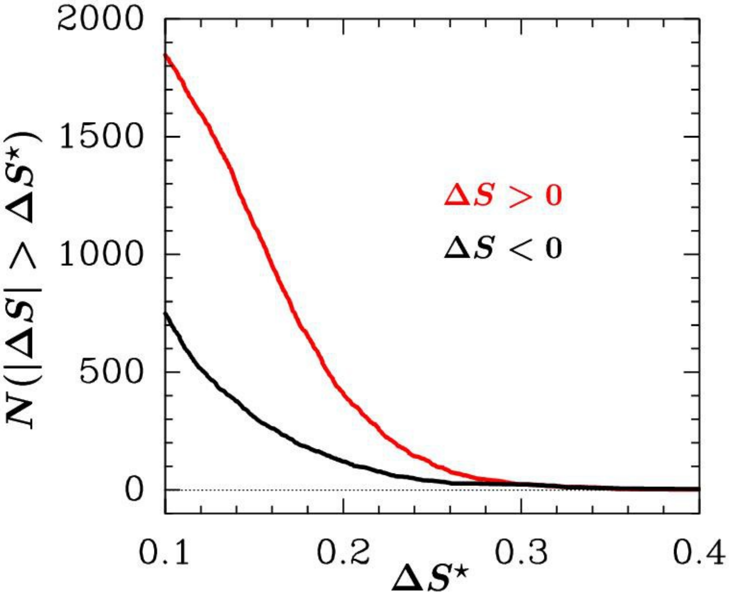
T.I. Lee *et al.*, Cell 125 (2006): **32%** of human ES cell genes bound to Pol II

# Jumps in the skew profiles at scales a > 200kbp

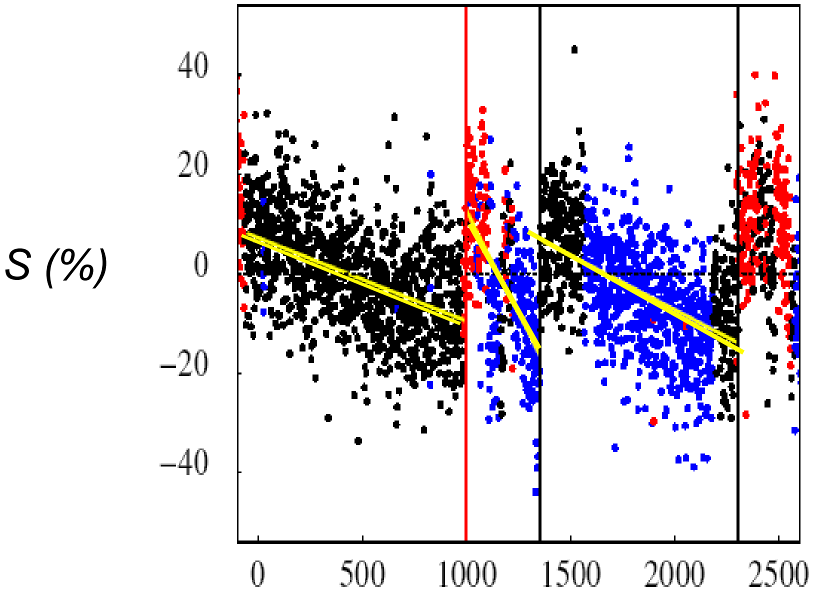


# Statistics of upward and downward jumps in Human skew profiles

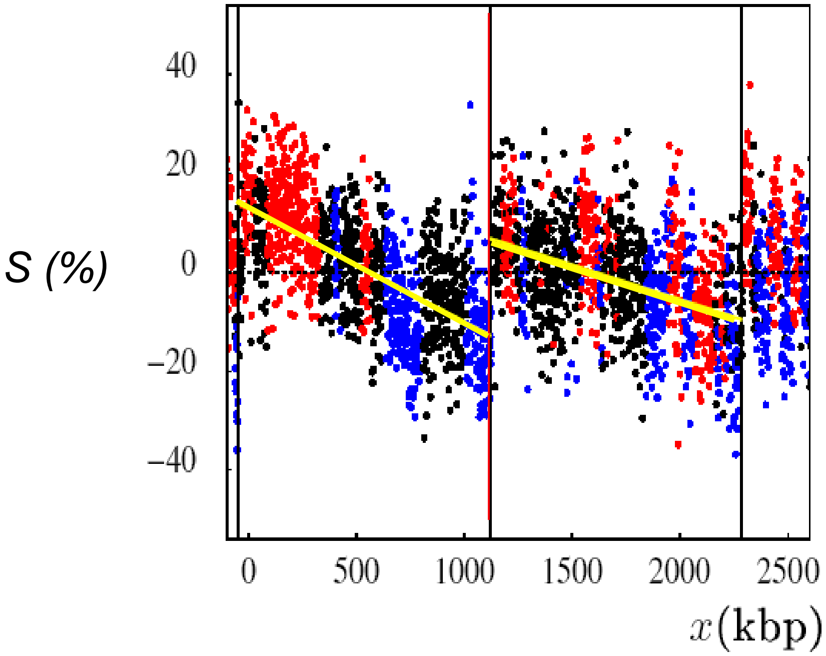




# « Factory roof » profiles



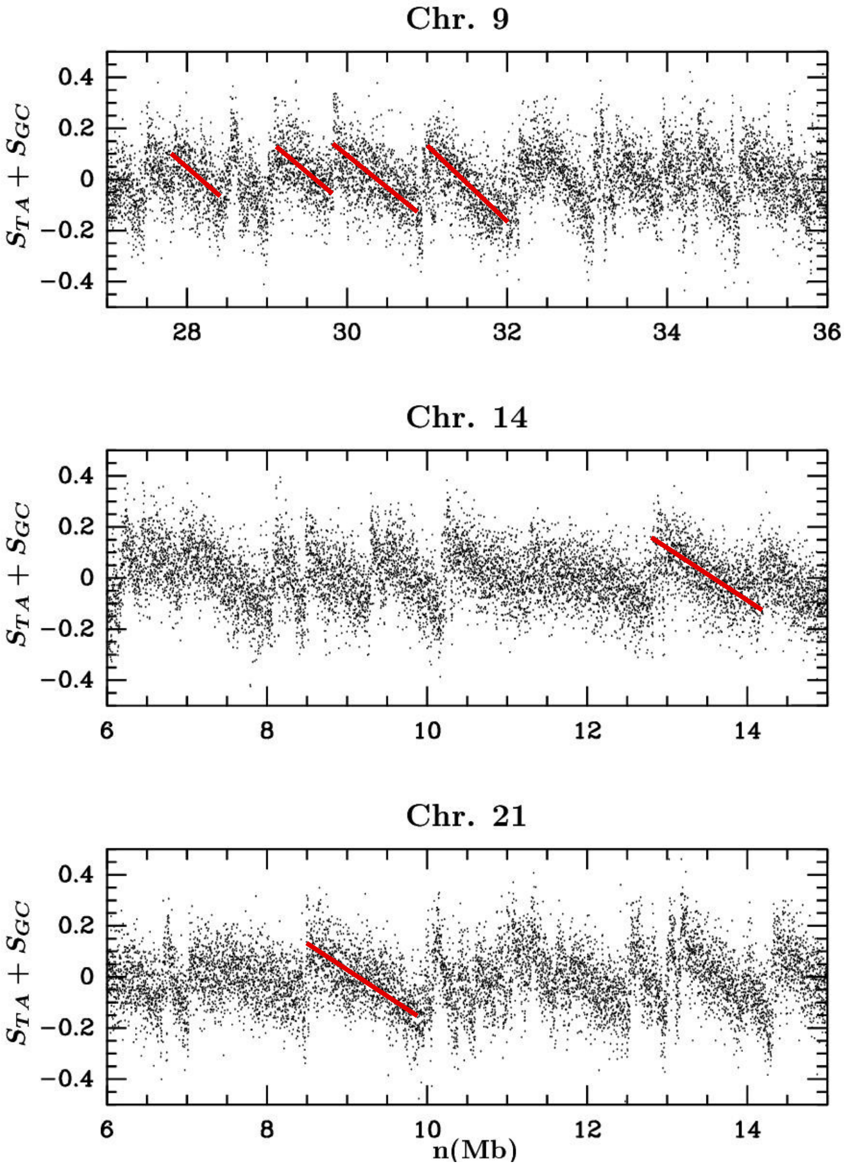
*TOP1*



*MCM4*

# Skew profiles along 9Mb Human contigs

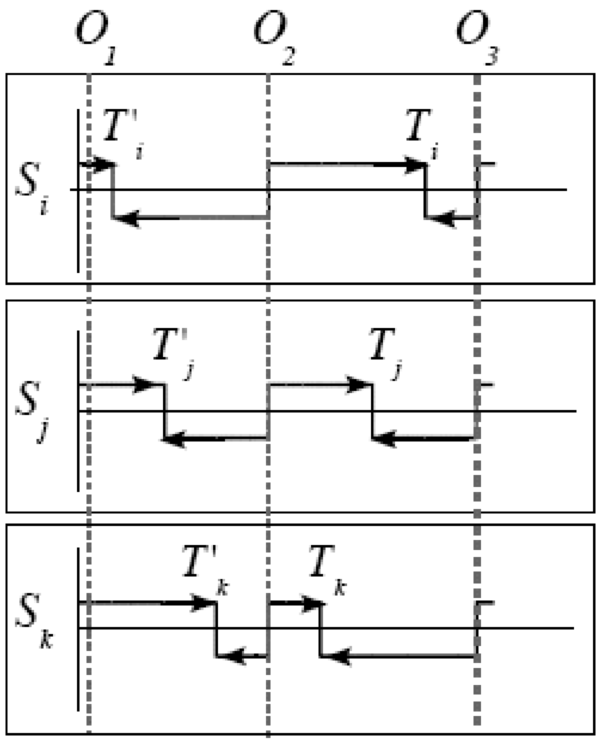
Brodie of Brodie, Phys. Rev. Lett. (2005)
Touchon, PNAS (2005)

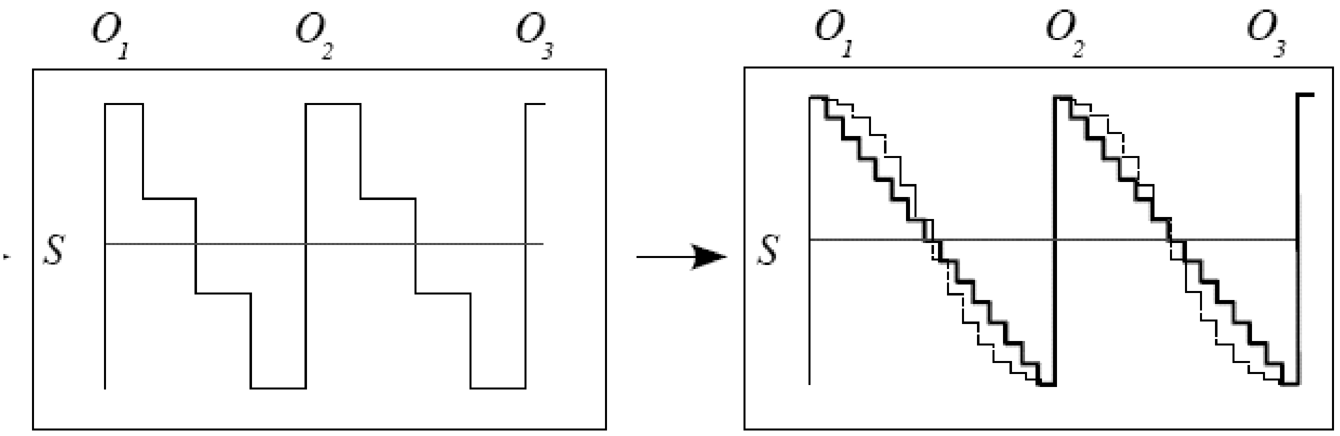


# Model of Replication

## Well positionned replication origins, separated by diffuse termini

Brodie of Brodie, Phys. Rev. Lett. (2005)
Touchon, PNAS (2005)

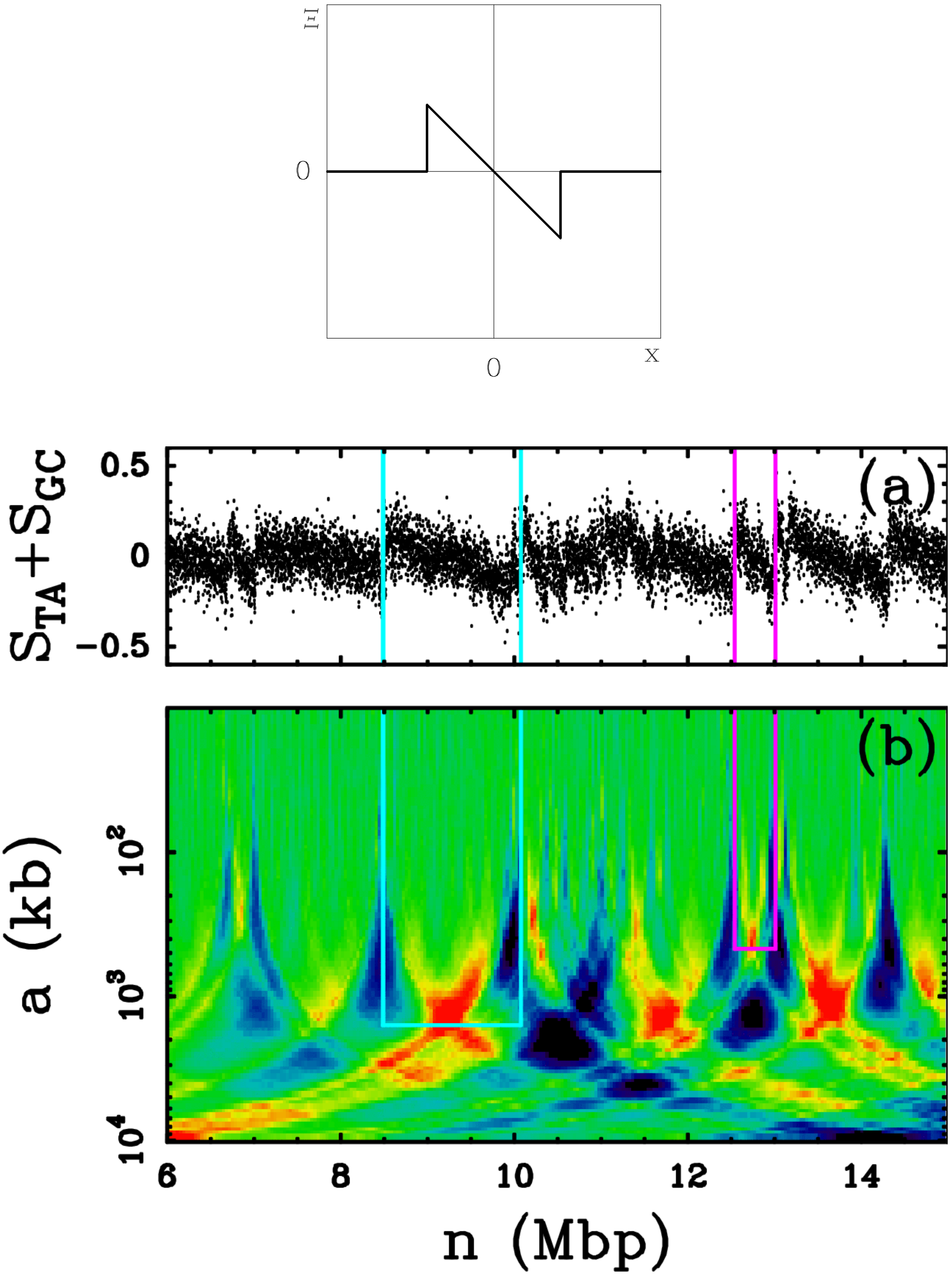




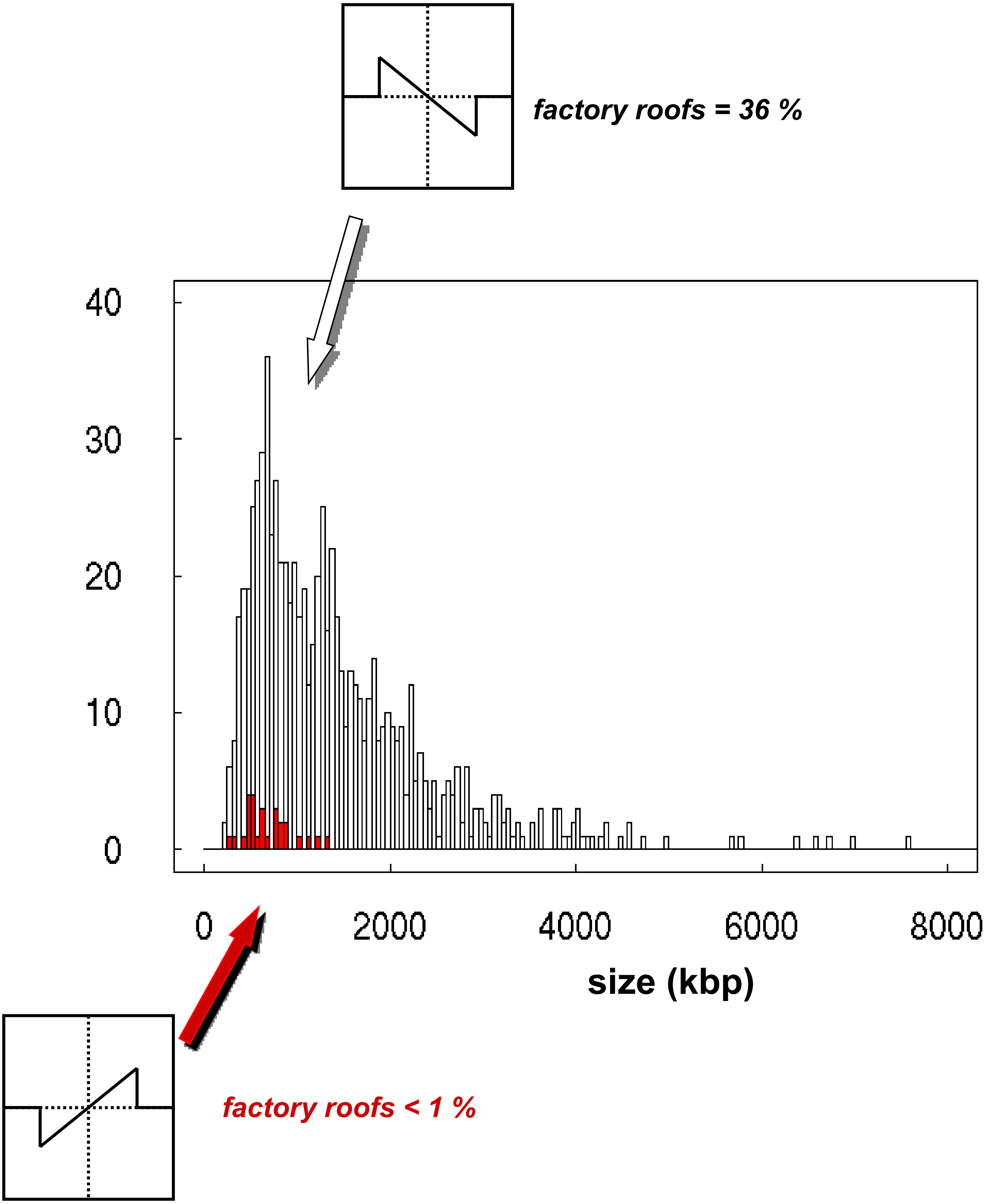
**In germ-line cells, termination sites are regularly dispersed between adjacent origins**

# Replication domain detection using an adapted analyzing wavelet

Audit, Phys. Rev. Lett. (2007)
Huvet, Genome Res. (2007)



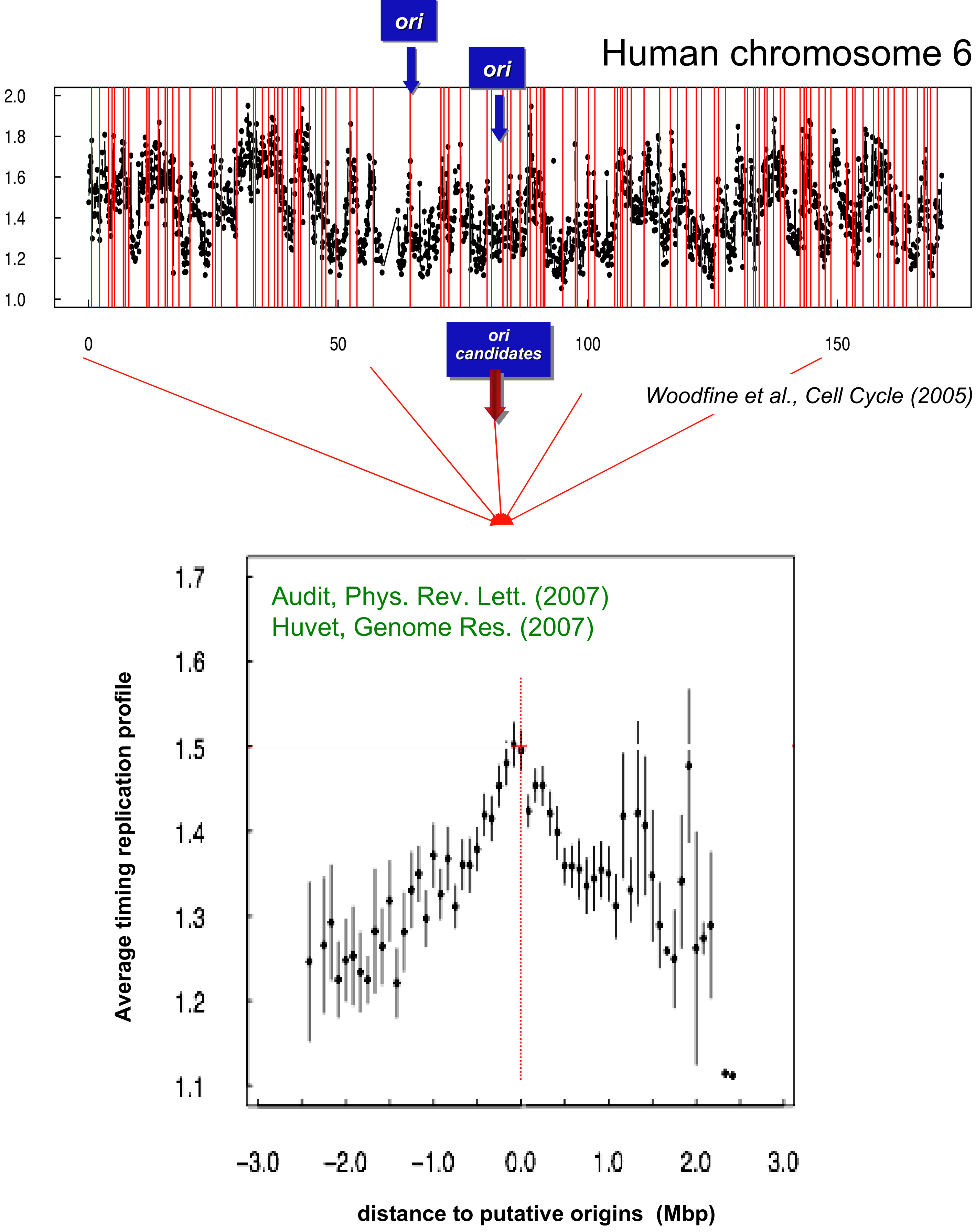
# ASYMMETRY OF HUMAN GENOME



*factory roofs = 36 %*

*factory roofs < 1 %*

# Human chromosome 6



**ori**

**ori**

**ori candidates**

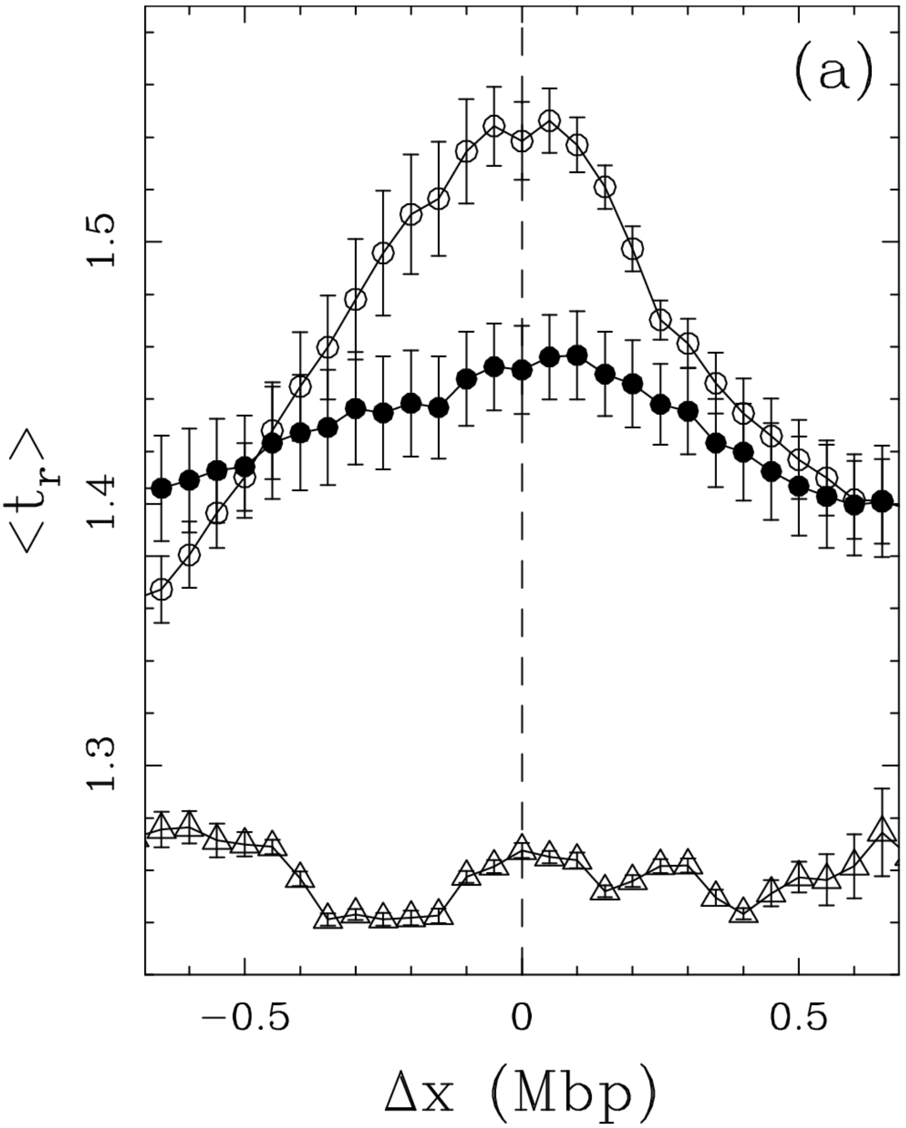
*Woodfine et al., Cell Cycle (2005)*

Audit, Phys. Rev. Lett. (2007)
Huvet, Genome Res. (2007)

**The detected jumps are replicated earlier than surrounding regions**

# The predicted origins replicate earlier than their surrounding regions

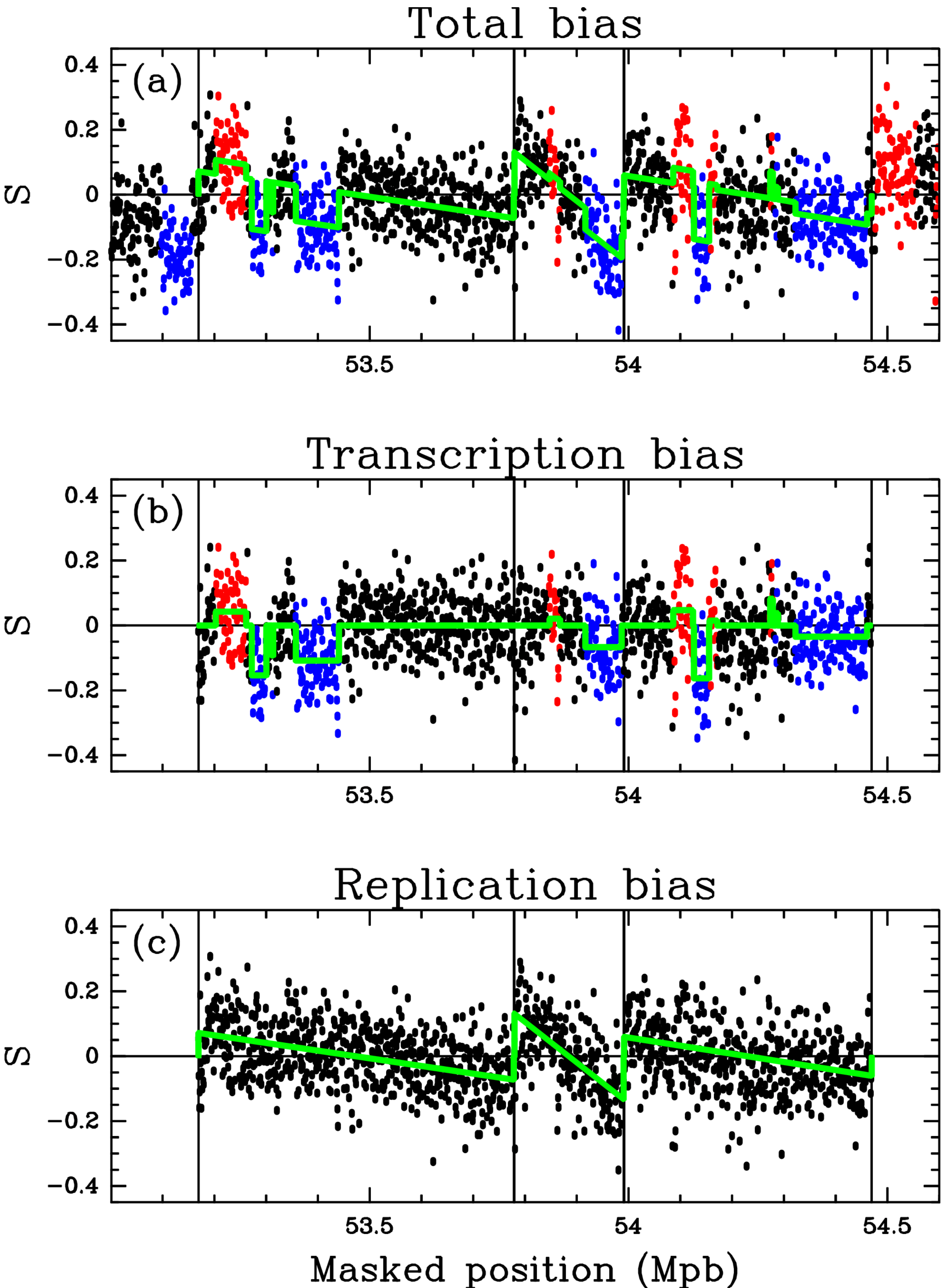
Audit, Phys. Rev. Lett. (2007)
Huvet, Genome Res. (2007)



 : Average timing ratio profile around the 83 putative chr6 origins
○ : 20 origins with well defined local maxima
△: 10 late replicating origins
Timing ratio data from *Woodfine et al., Cell Cycle (2005)*

# Disentangling replication and transcription contributions to skew profiles

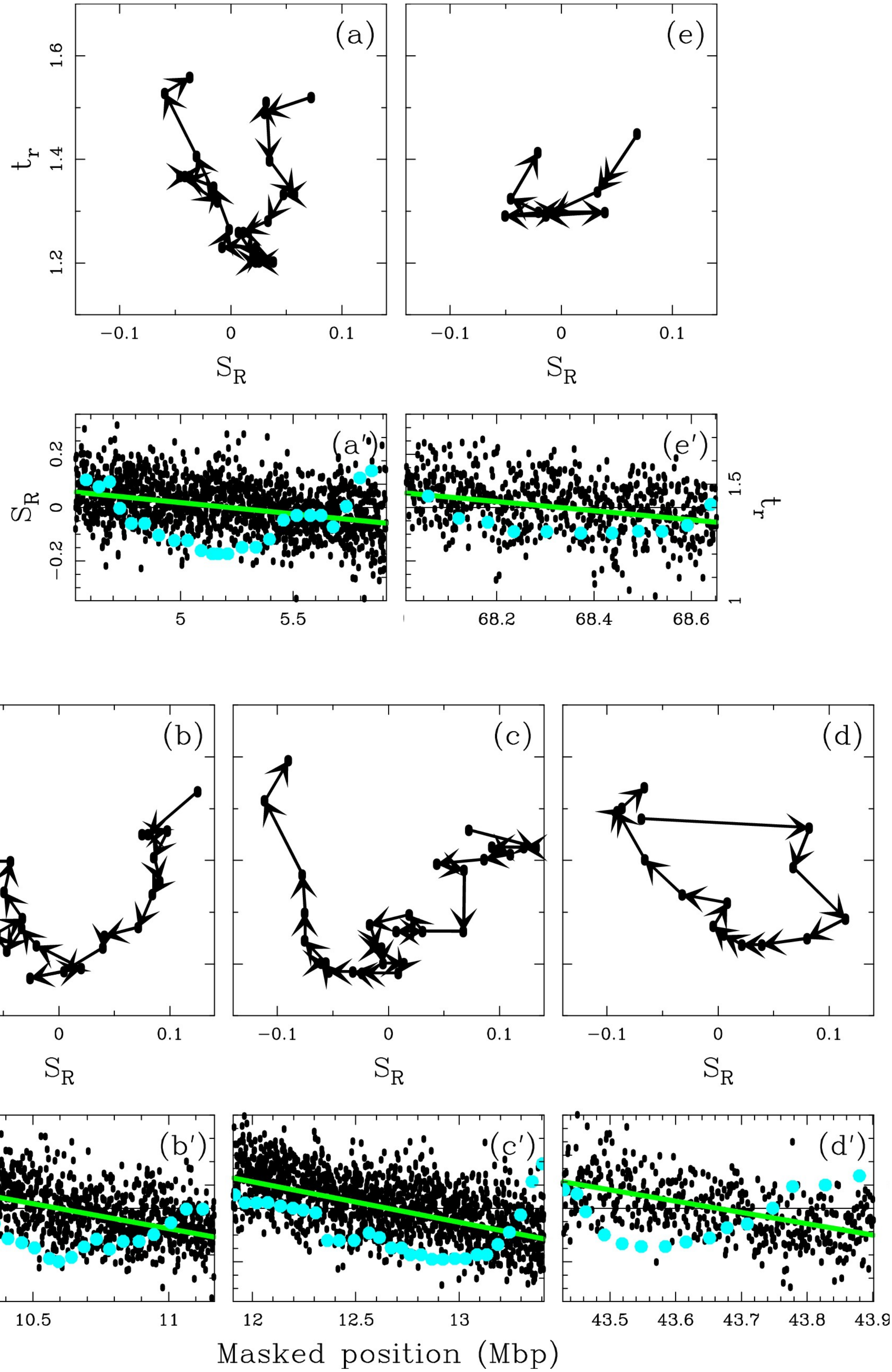
Audit, Phys. Rev. Lett. (2007)



Total bias

Transcription bias

Replication bias


(a)
(e)
t_r
1.6
1.4
1.2
−0.1
0
0.1
S_R
(a′)
(e′)
S_R
0.2
0
−0.2
t_r
1.5
1
5
5.5
68.2
68.4
68.6
(b)
(c)
(d)
(b′)
(c′)
(d′)
10.5
11
12
12.5
13
43.5
43.6
43.7
43.8
43.9
Masked position (Mbp)

# Genome-wide high-resolution (kbp) experimental verification of putative replication origins using massive sequencing technology

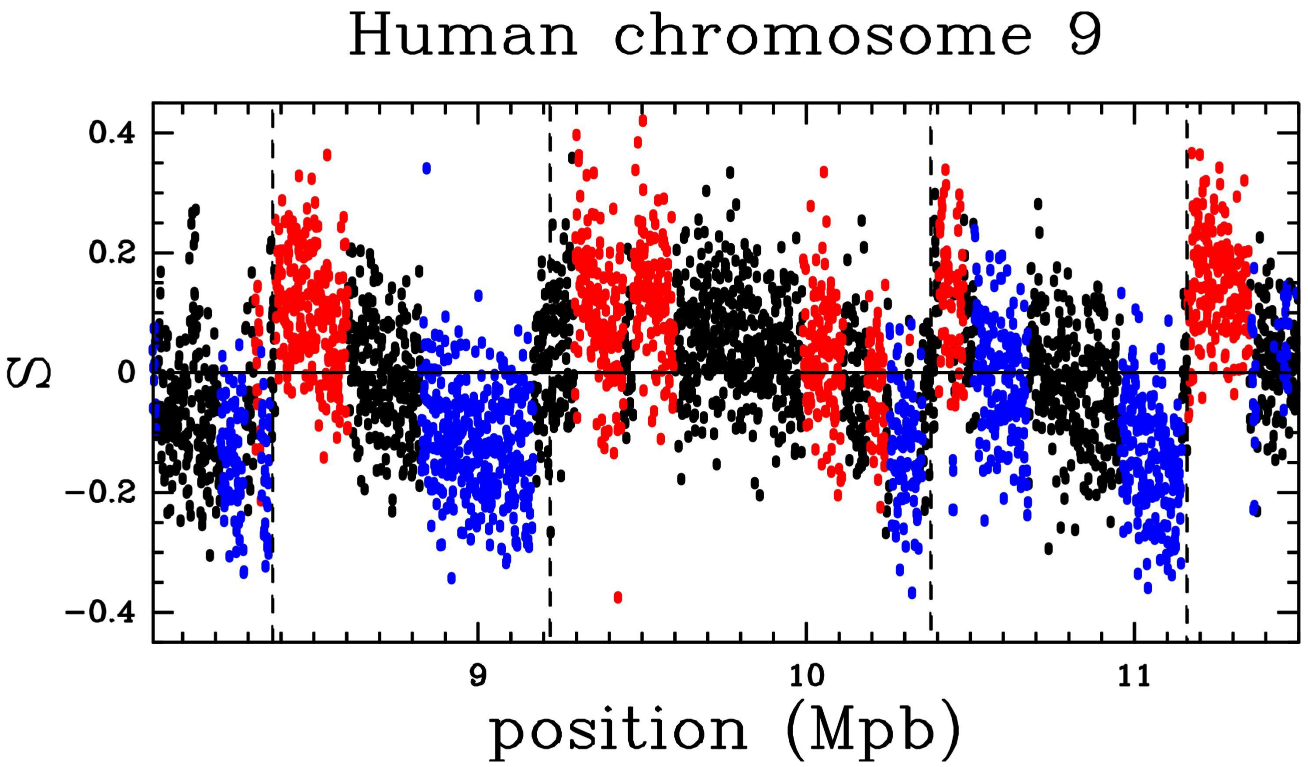
**Olivier Hyrien's group - ENS - Paris**

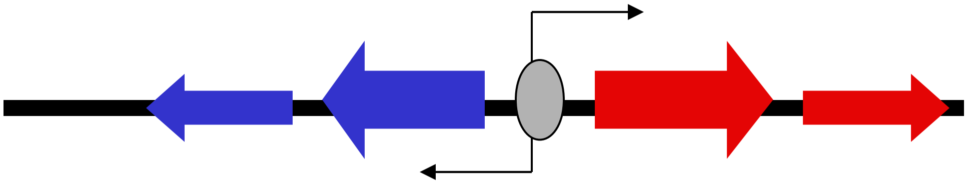
## Abundance of small nascent replication strands

## Replication timing profile

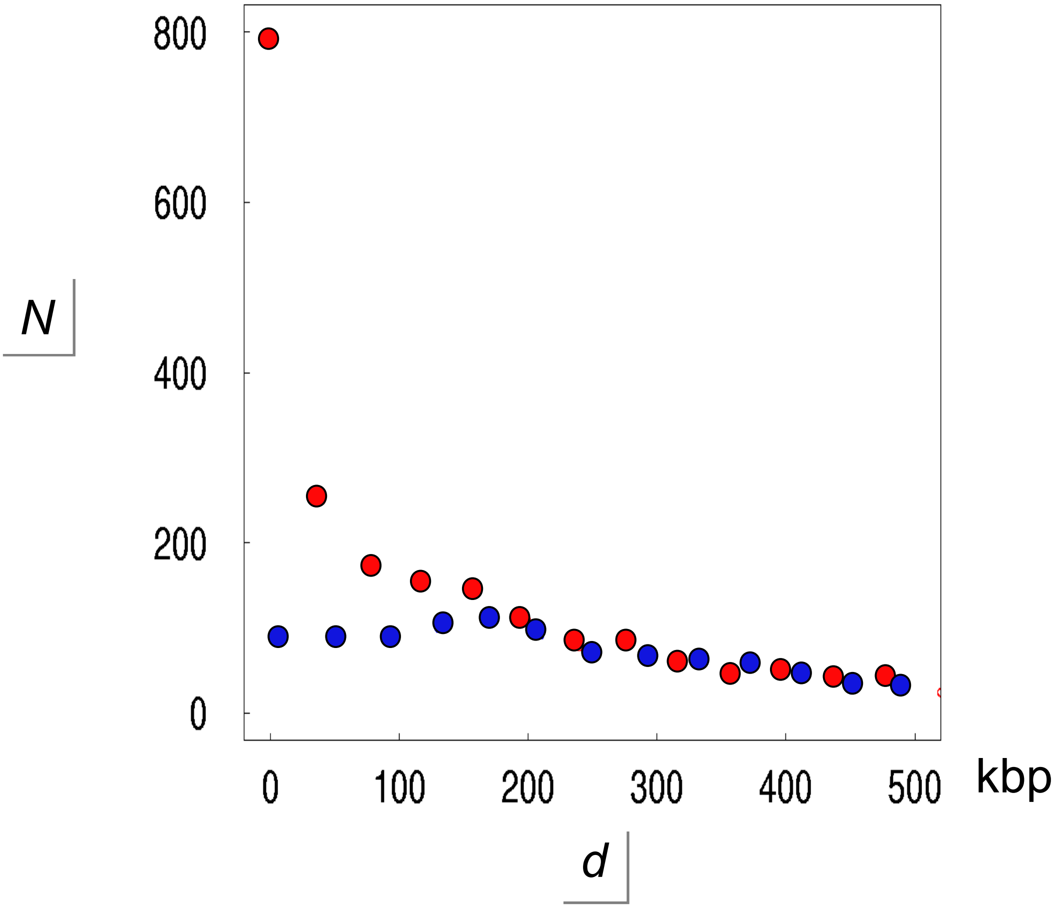
# Genome organisation around putative replication origins

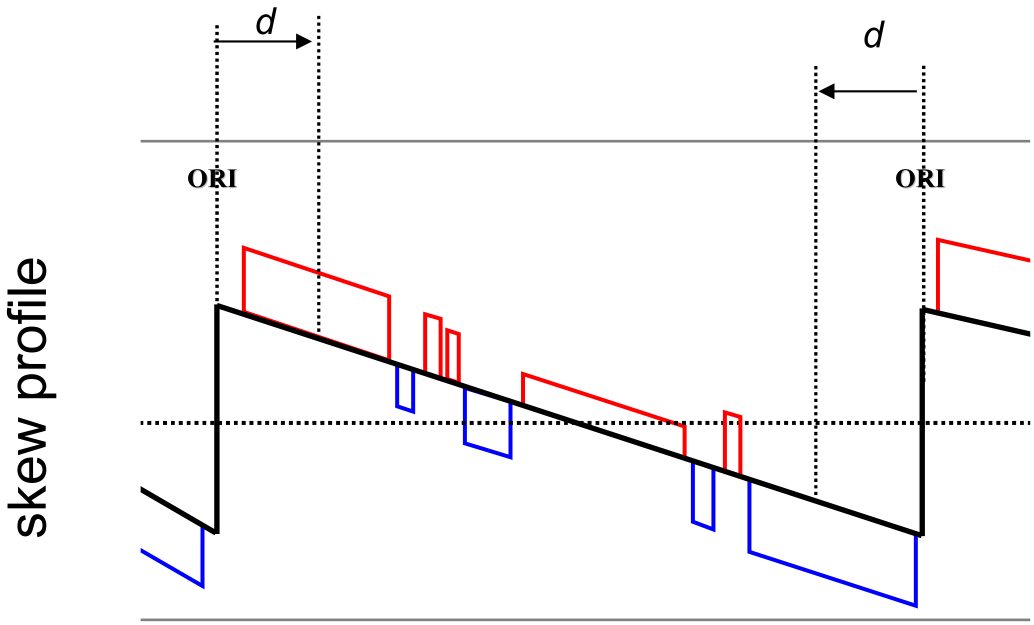
Huvet, Genome Res. (2007)



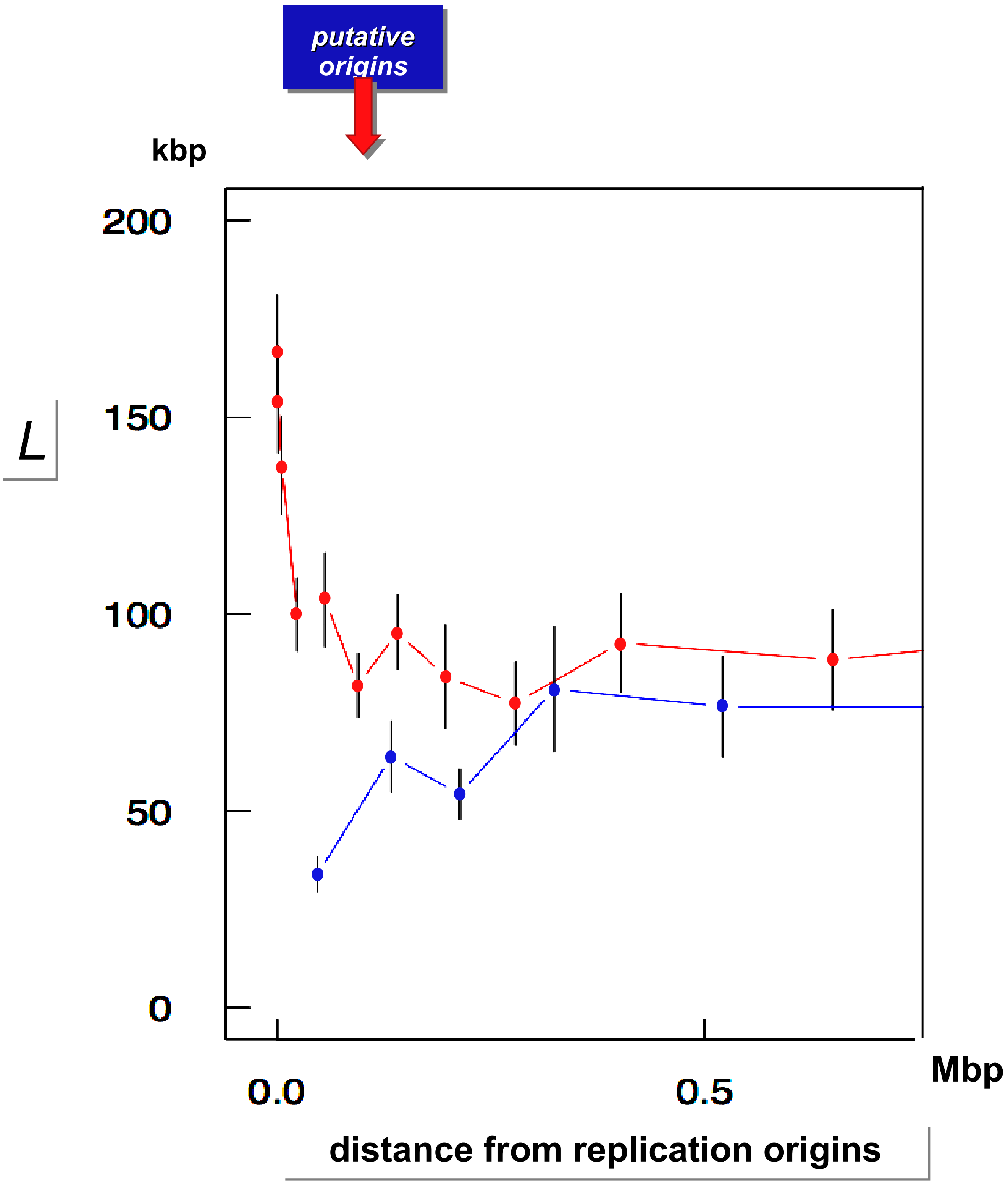


# Gene distribution around predicted replication origins

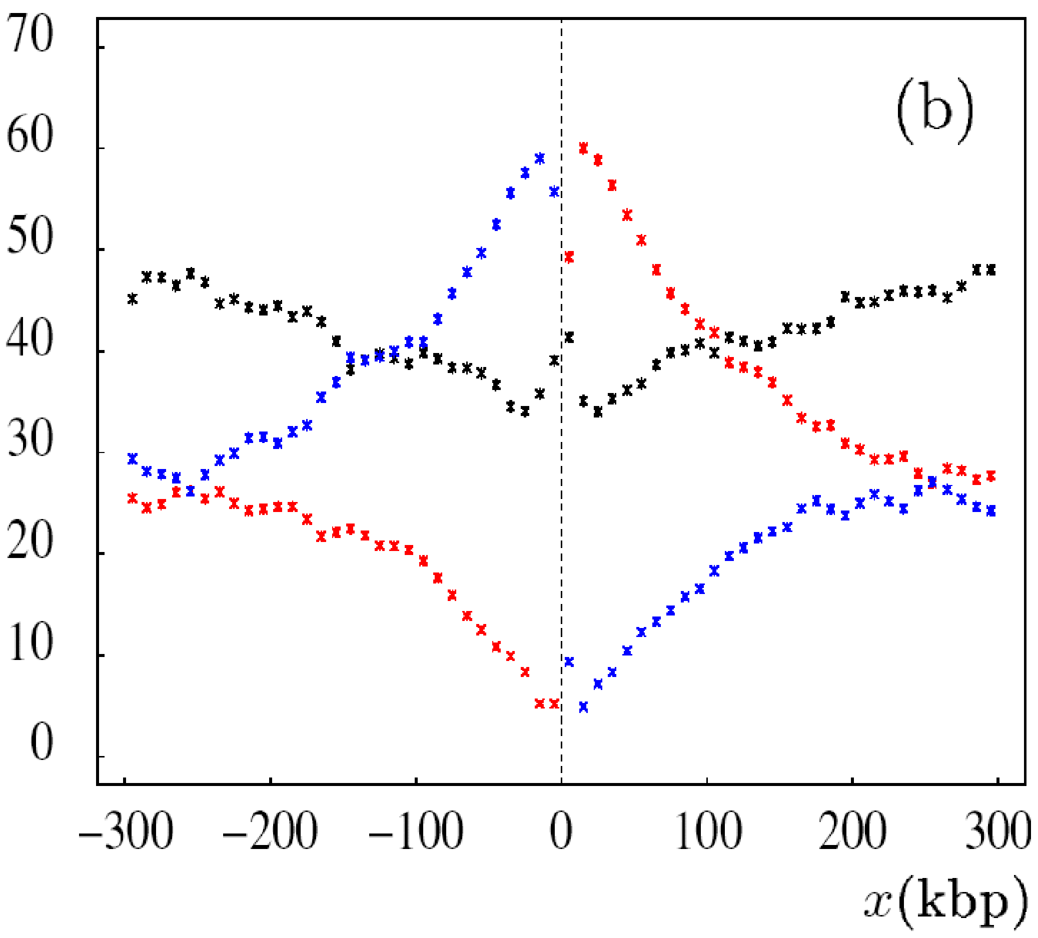




# Distribution of gene length around predicted replication origins



# Genome organisation around putative replication origins

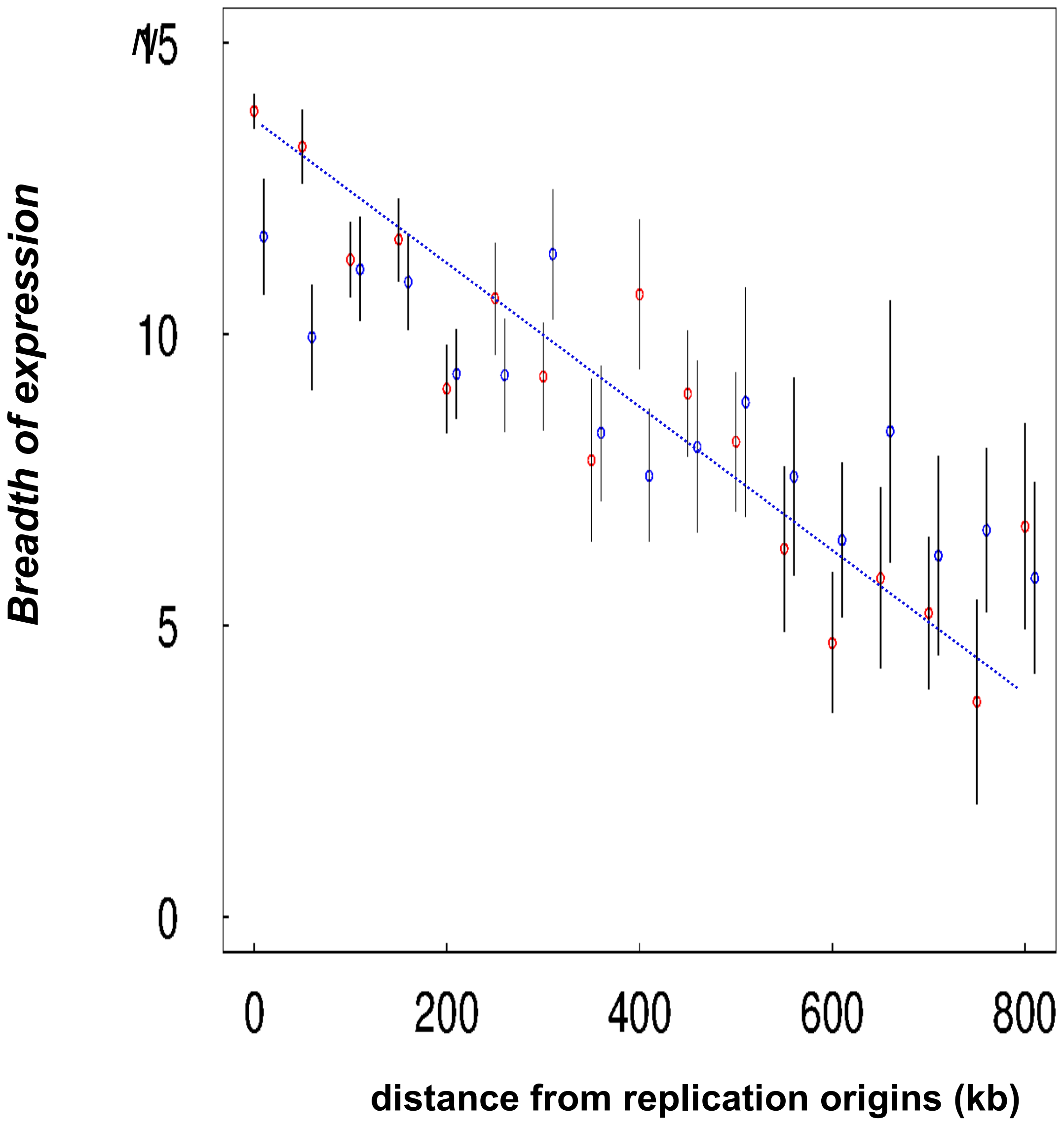


**transcribed regions around putative origins (± 20 kb) :**

- **55 % are oriented as replication fork progression**
- **5% are in opposed orientation**
- **35% are intergenic**

# Gene expression around predicted replication origins

ESTs



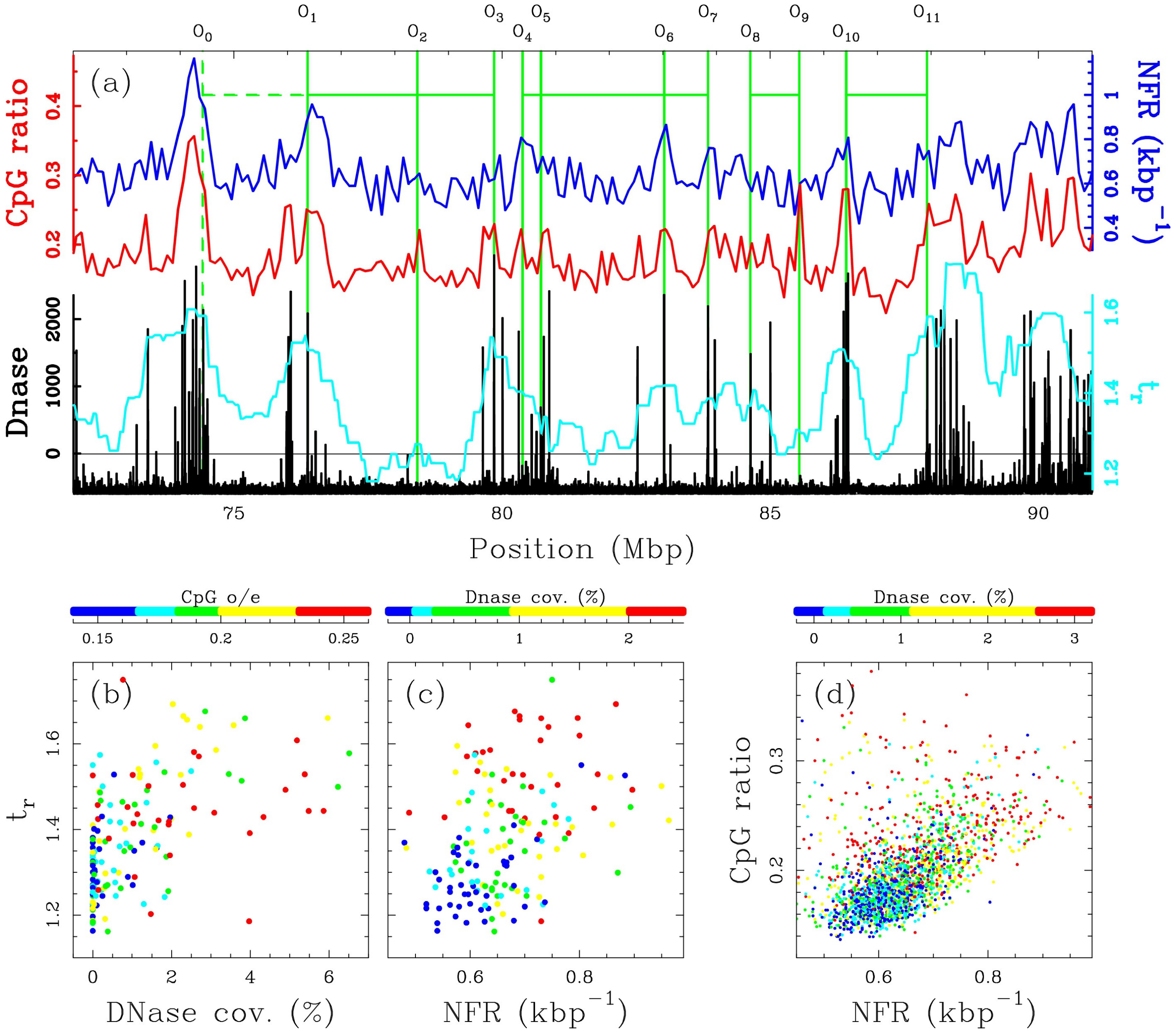
# Model for the large-scale functional organisation of the human genome

Huvet, Genome Res. (2007)

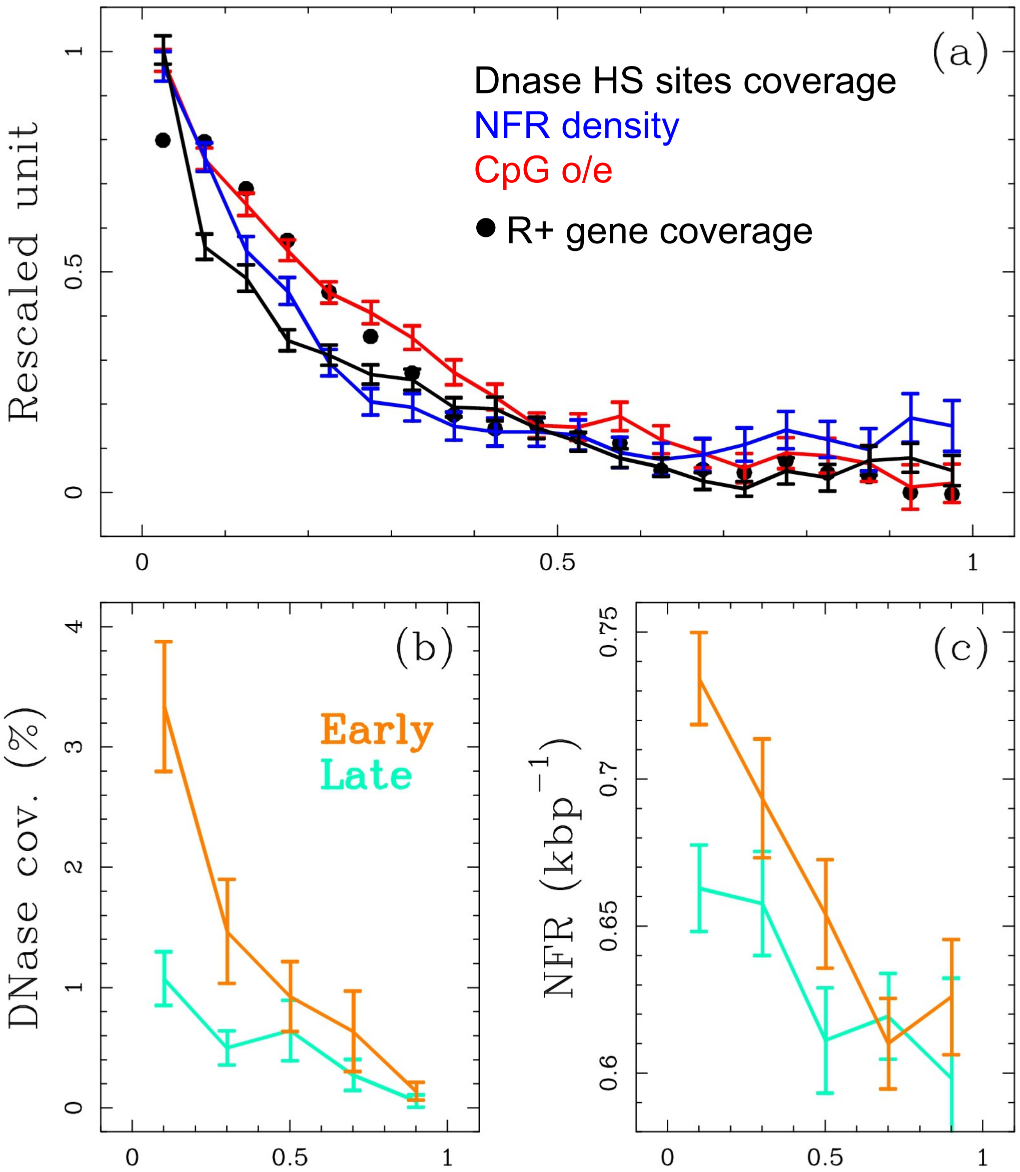
# Genomic DNA codes for open chromatin around "master" replication origins

Audit, Nucleic Acids Res. (2009)

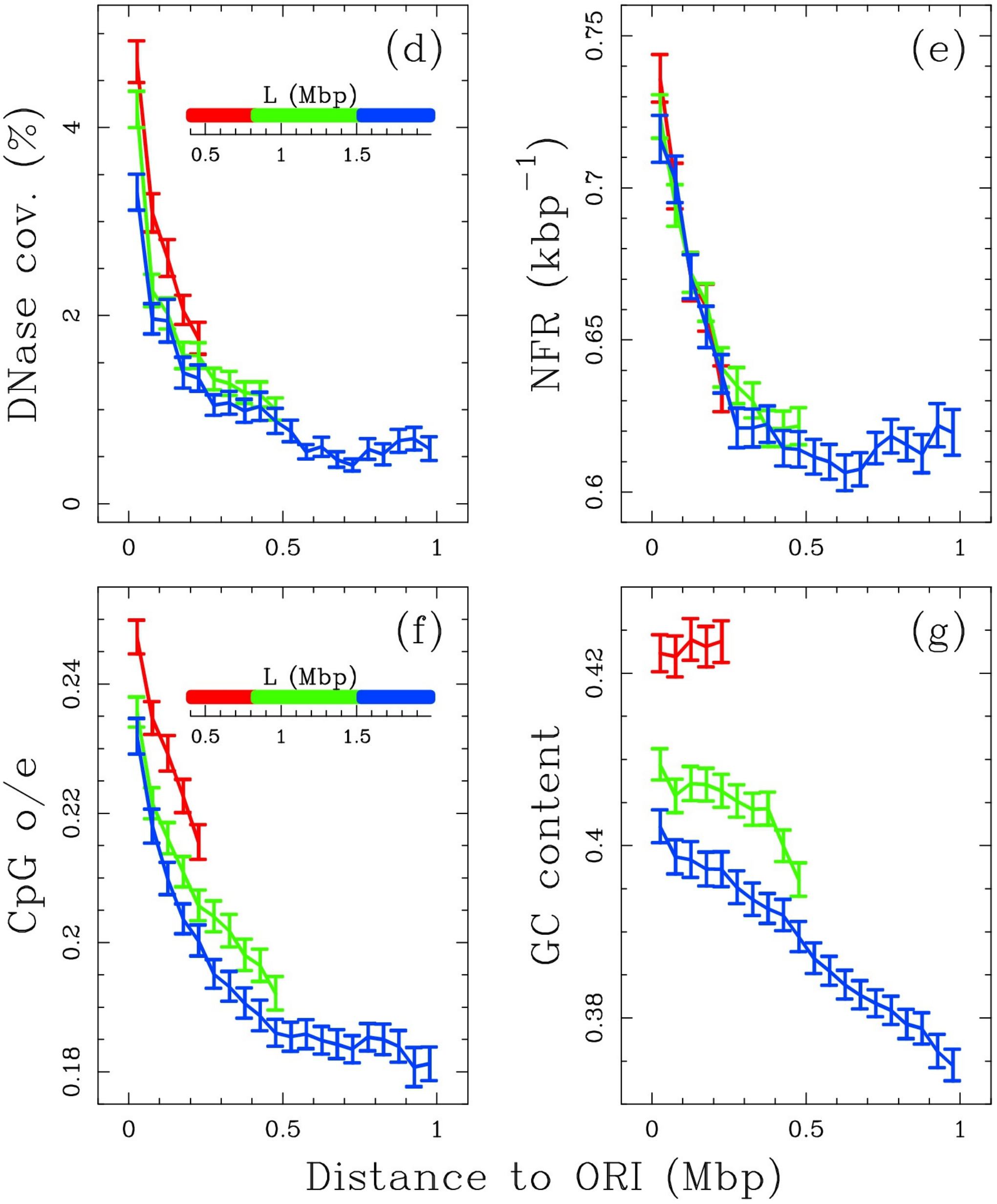
Human chromosome 6



# Genomic DNA codes for open chromatin around “master” replication origins

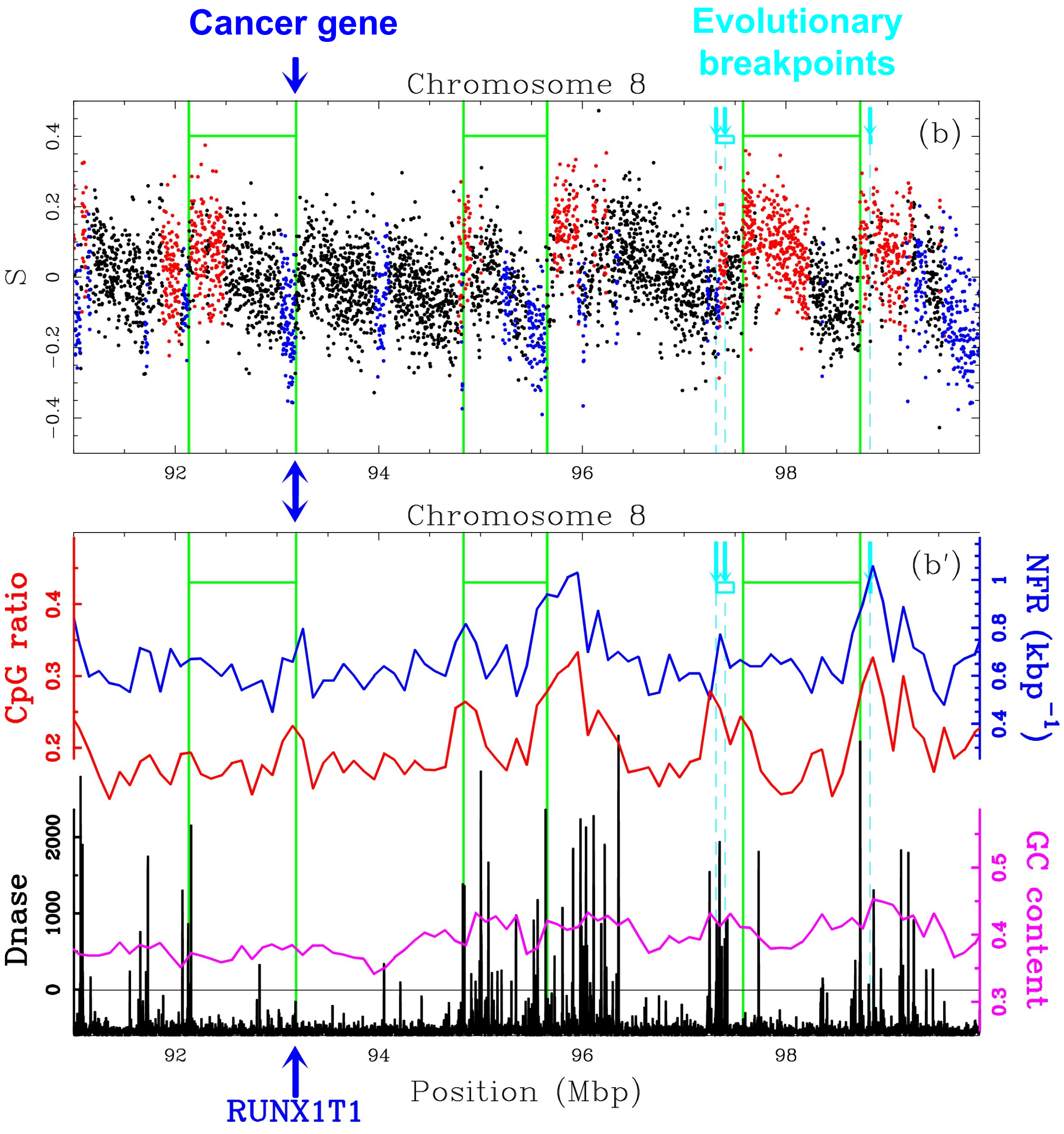


# Genomic DNA codes for open chromatin around "master" replication origins



# Fragility of the human genome around “master” replication origins

Audit, Nucleic Acids Res. (2009)
Lemaitre, BMC Genomics (2009)

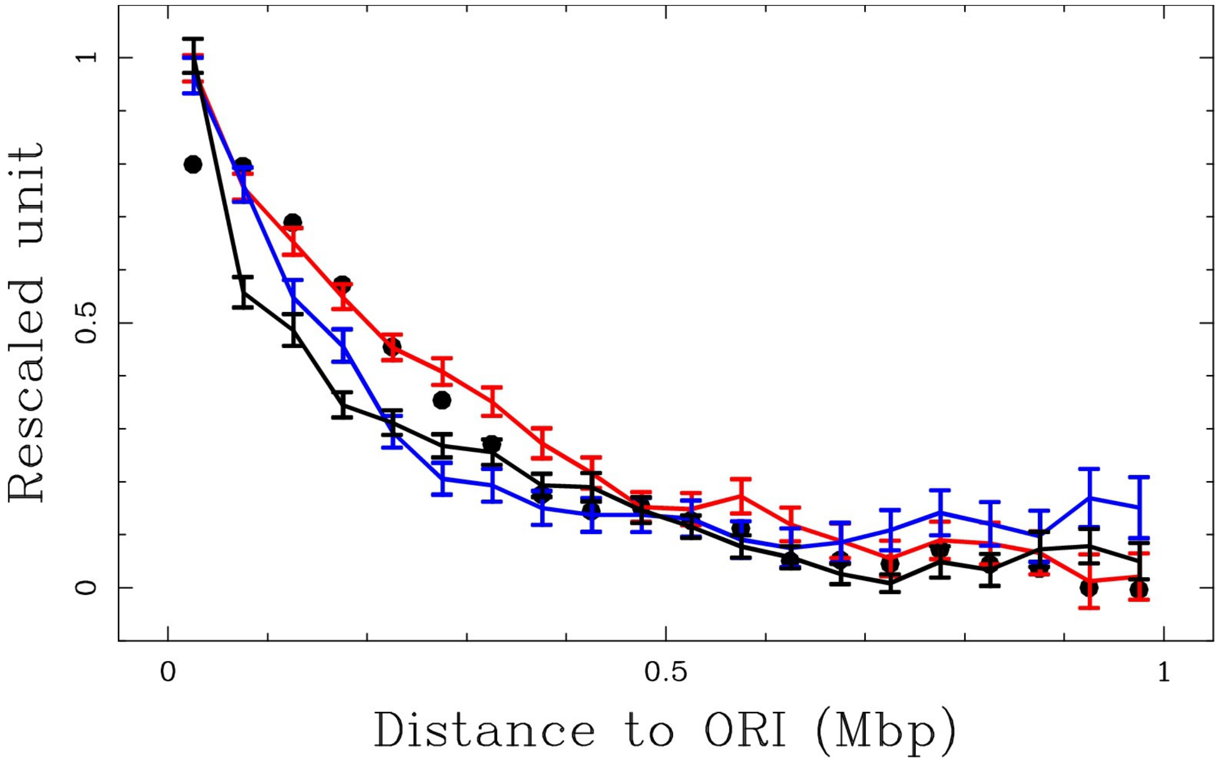


**Cancer gene**

**Evolutionary breakpoints**

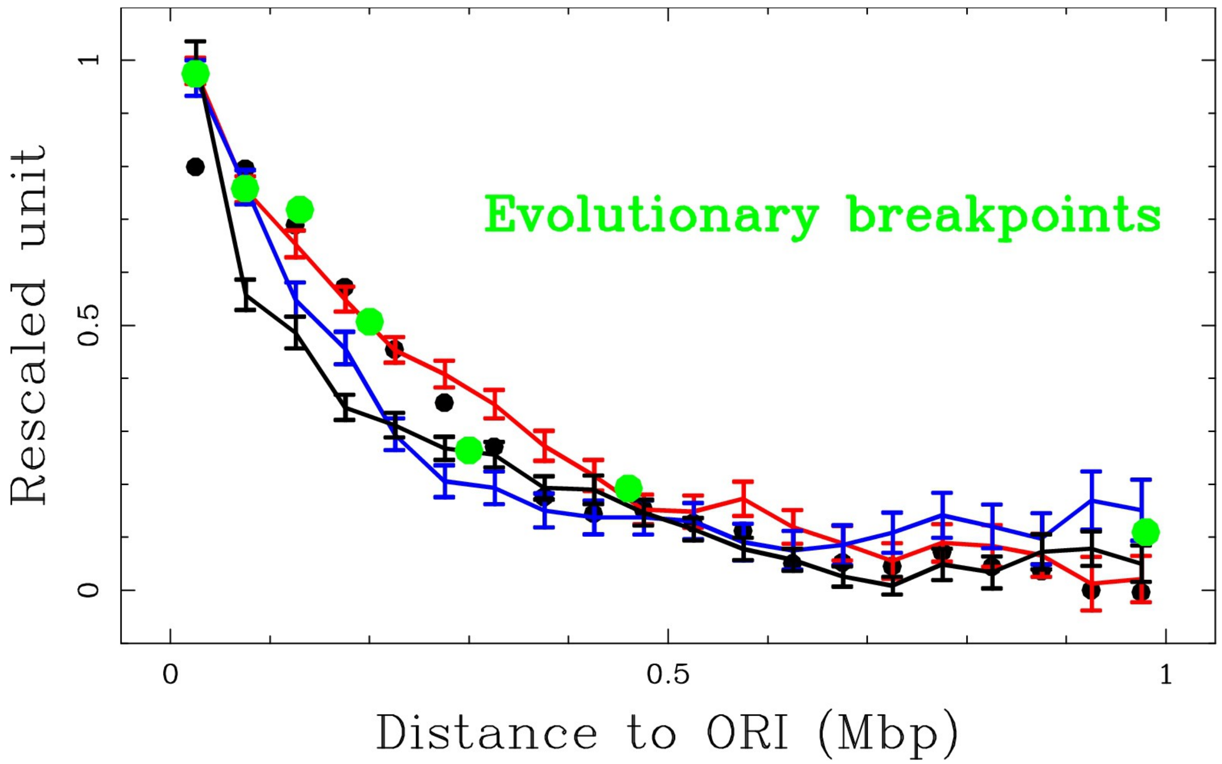
**RUNX1T1**

# Genomic DNA codes for **fragile open chromatin** around “master” replication origins



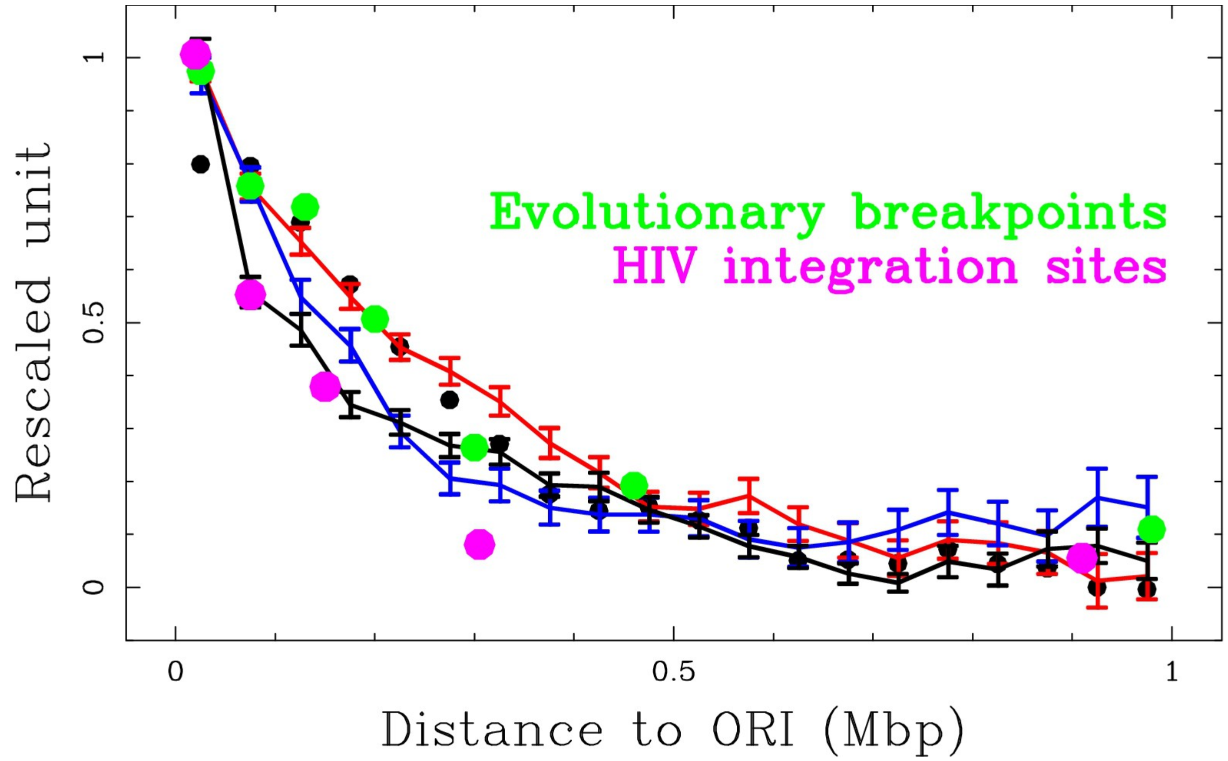
Dnase HS sites coverage
NFR density
CpG o/e
● R+ gene coverage

# Genomic DNA codes for fragile open chromatin around “master” replication origins



Dnase HS sites coverage
NFR density
CpG o/e
● R+ gene coverage

# Genomic DNA codes for **fragile open chromatin** around "master" replication origins



Dnase HS sites coverage

NFR density

CpG o/e

● R+ gene coverage

# A cascading domino model for the spatio-temporal replication program in mammalian cells

# REFERENCES

***DNA replication timing data corroborate in silico human replication origin predictions***

B. AUDIT, S. NICOLAY, M. HUVET, M. TOUCHON, Y. D'AUBENTON-CARAFA, C.THERMES & A. ARNEODO. Phys. Rev. Lett. **99**, 248102 (2007).

***Human gene organization driven by the coordination of replication and transcription***

M. HUVET, S. NICOLAY, M. TOUCHON, B. AUDIT, Y. D'AUBENTON-CARAFA, A. ARNEODO.& C. THERMES. Genome Res. **17**, 1278-1285 (2007).

***Bifractality of human DNA strand-asymmetry profiles results from transcription.***

S. NICOLAY, E.B. BRODIE OF BRODIE, M. TOUCHON, B. AUDIT, Y. D'AUBENTON-CARAFA, C. THERMES & A. ARNEODO. Phys. Rev. E **75**, 032902 (2007).

***DNA in chromatin: from genome-wide sequence analysis to the modeling of replication in mammals.***

A. ARNEODO, Y. D'AUBENTON-CARAFA, B. AUDIT, E.B. BRODIE OF BRODIE, S. NICOLAY, P.ST-JEAN, C. THERMES, M. TOUCHON, & C. VAILLANT. Adv. Chem. Phys. **135**, 203 (2007).

***Replication-associated strand asymmetries in mammalian genome : towards detection of replication origins***

M. TOUCHON, S. NICOLAY, B. AUDIT, E.B. BRODIE OF BRODIE, Y. D'AUBENTON-CARAFA, A. ARNEODO & C. THERMES, Proc. Natl. Acad. Sci.USA **102**, 9836-9841 (2005)

***From DNA sequence analysis to modeling replication in the Human genome***

E.B. BRODIE OF BRODIE, S. NICOLAY, M. TOUCHON, B. AUDIT, Y. D'AUBENTON-CARAFA, C. THERMES & A. ARNEODO, Phys. Rev. Lett. **94**, 248103 (2005)

***Transcription-coupled and splicing-coupled strand asymmetries in eukaryotic genomes.***

M. TOUCHON, A. ARNEODO, Y. D'AUBENTON-CARAFA & C. THERMES, Nucleic Acids Res. **32**, 4969 (2004)

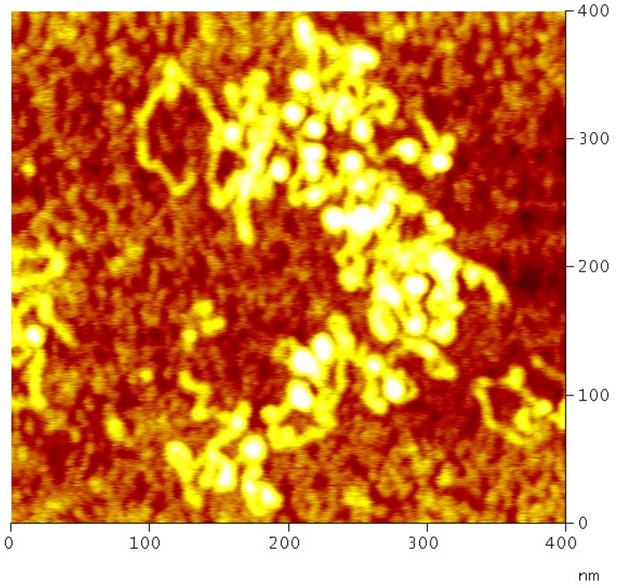
***Low Frequency rhythms in human DNA sequences : a key to the organization of gene location and orientation?***

S. NICOLAY, F. ARGOUL, M. TOUCHON, Y. D'AUBENTON-CARAFA, C. THERMES & A. ARNEODO, Phys. Rev. Lett. **93**, 108101 (2004)

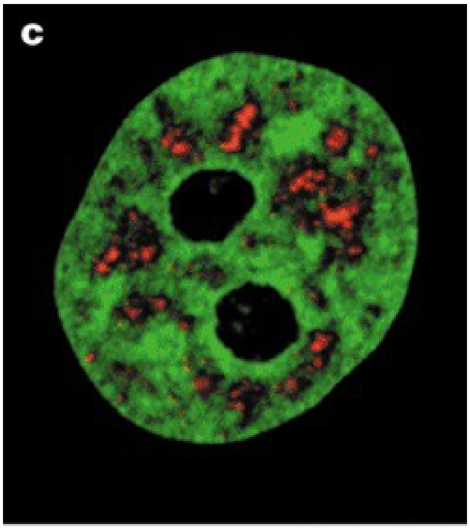
***Transcription-coupled TA and GC strand asymmetries in the human genome.***

M. TOUCHON, S. NICOLAY, A. ARNEODO, Y. D'AUBENTON-CARAFA & C. THERMES, FEBS Letters **555**, 579 (2003)

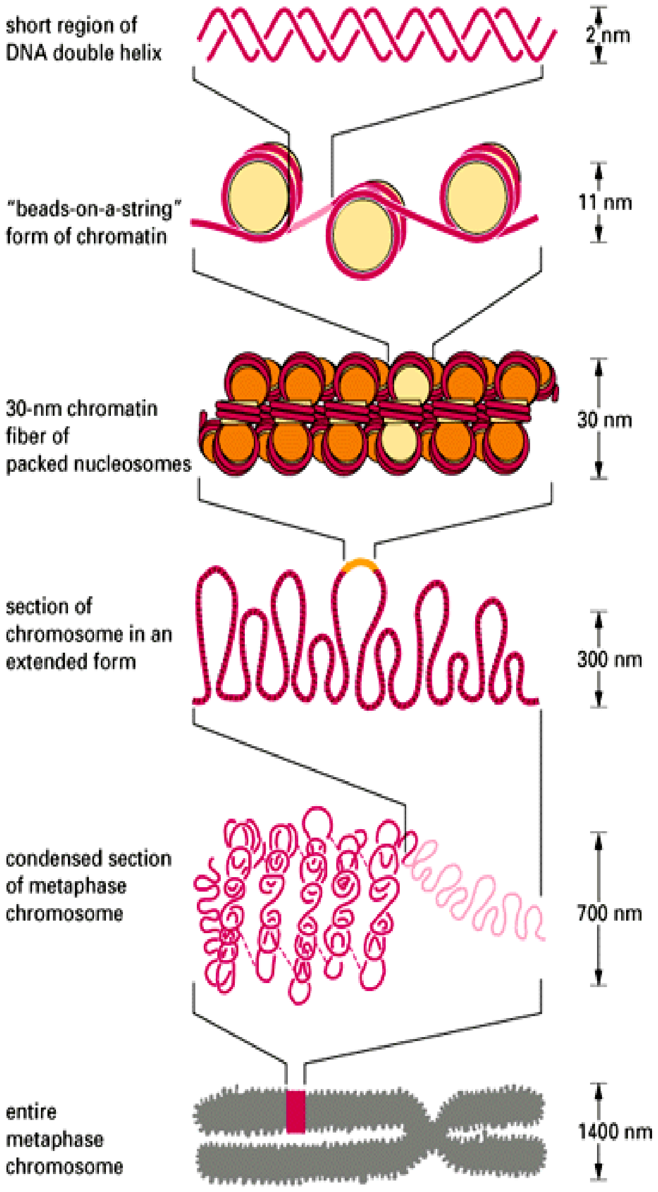
## Native Chromatin fiber



## Eukaryotic Nucleus



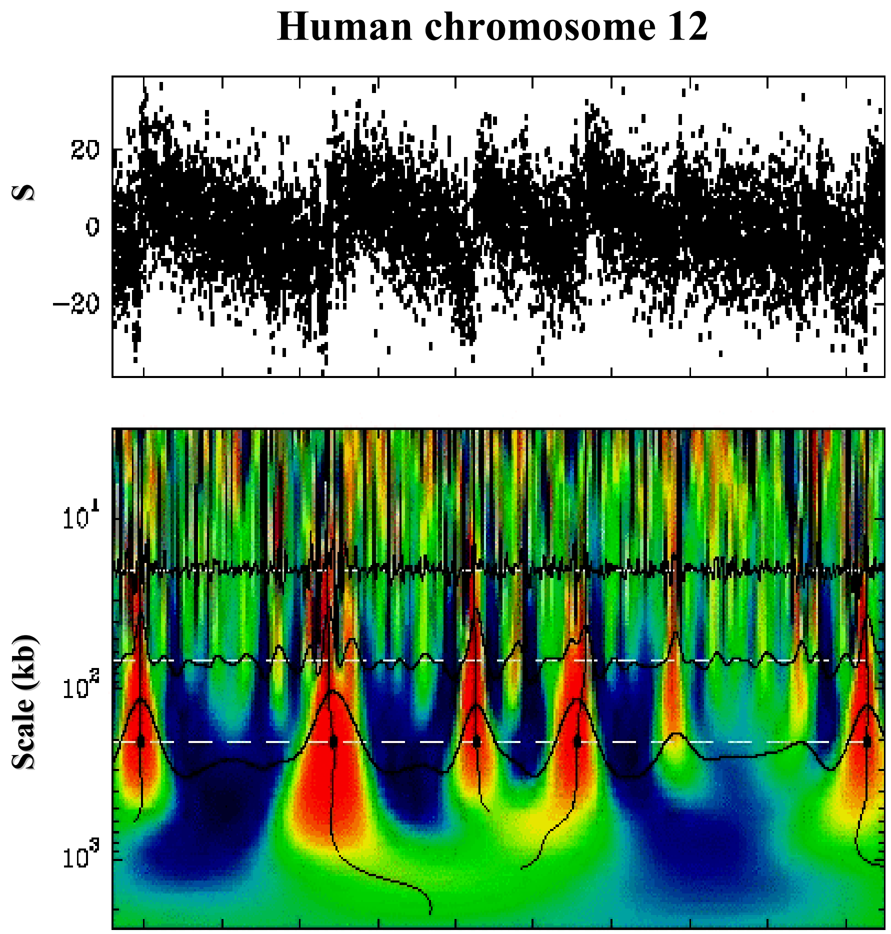
Nature Reviews | Genetics



# Laboratoire Joliot-Curie

# Multi-scale detection of jumps in human skew profile

## Wavelet based analysis

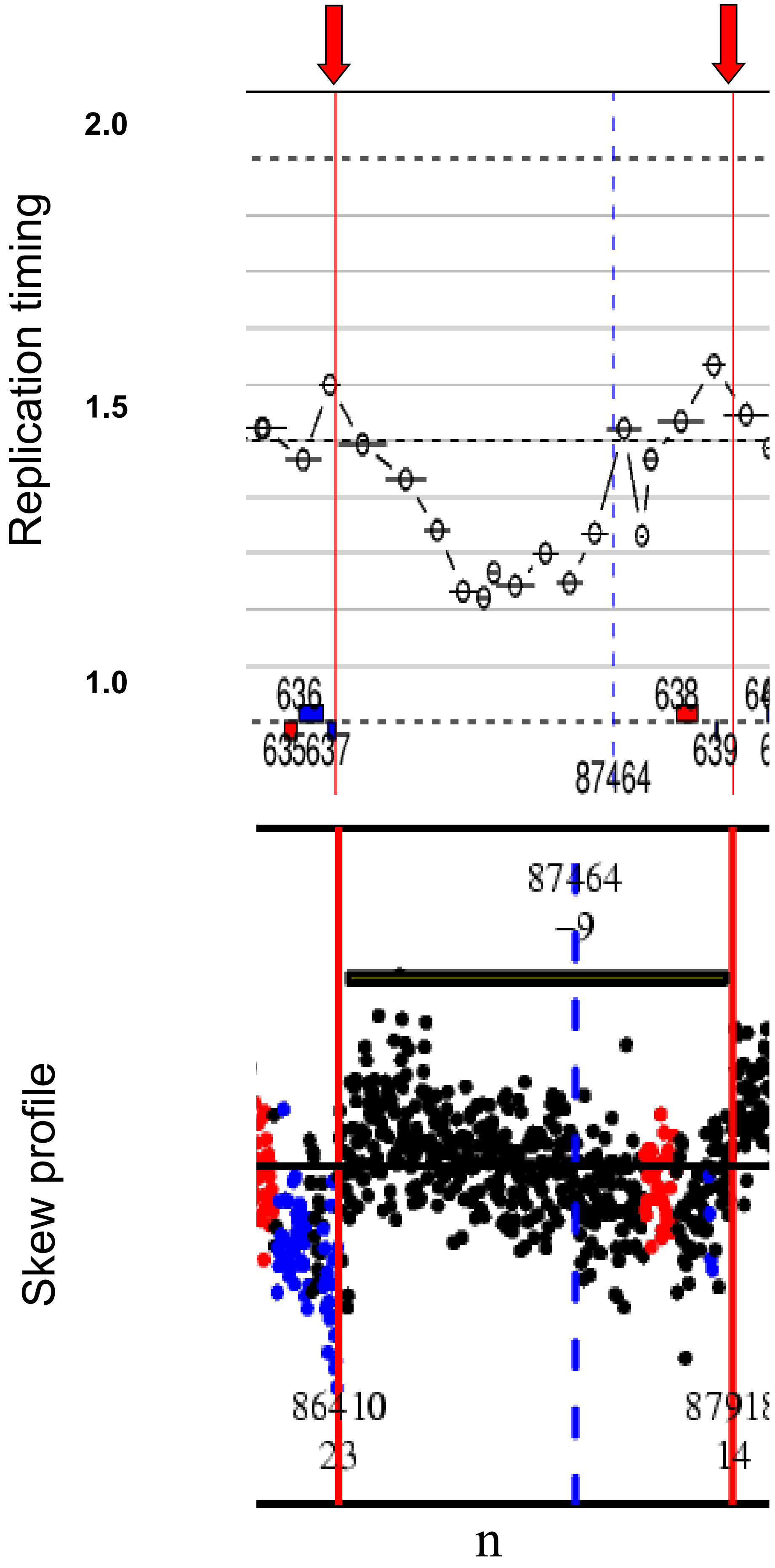


position of transitions

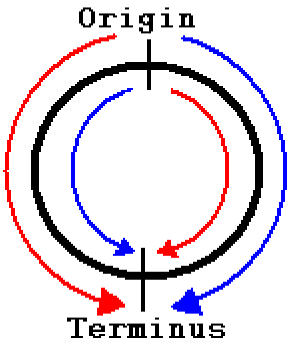
identification of transitions

Scale analysis : 200 kbp

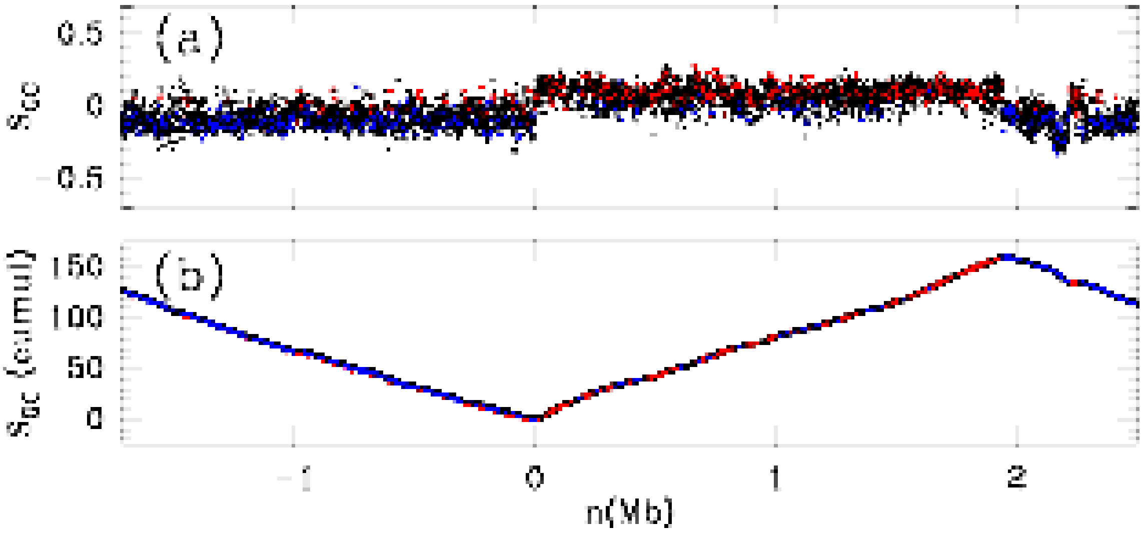
Detection of 2415 upward jumps and of a similar number of downward jumps


Replication timing
2.0
1.5
1.0
636
635637
638
639
87464
Skew profile
87464
−9
86410
23
14
n

# Replication and genome organization

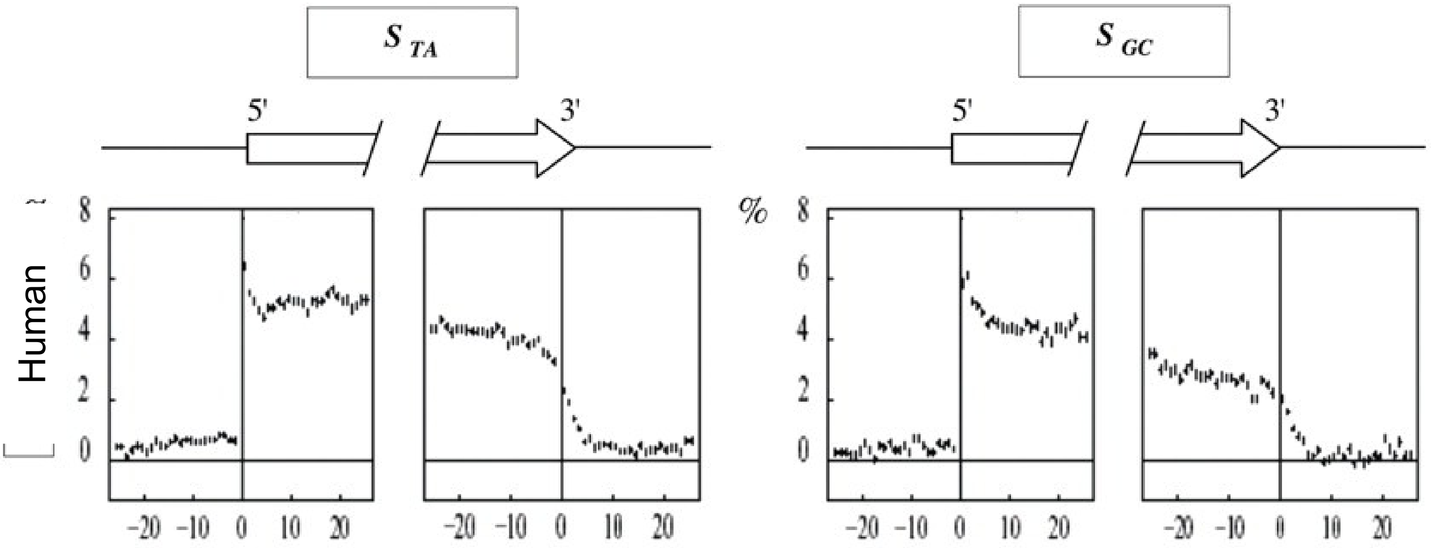


## Replication bias in *Bacillus subtilis*

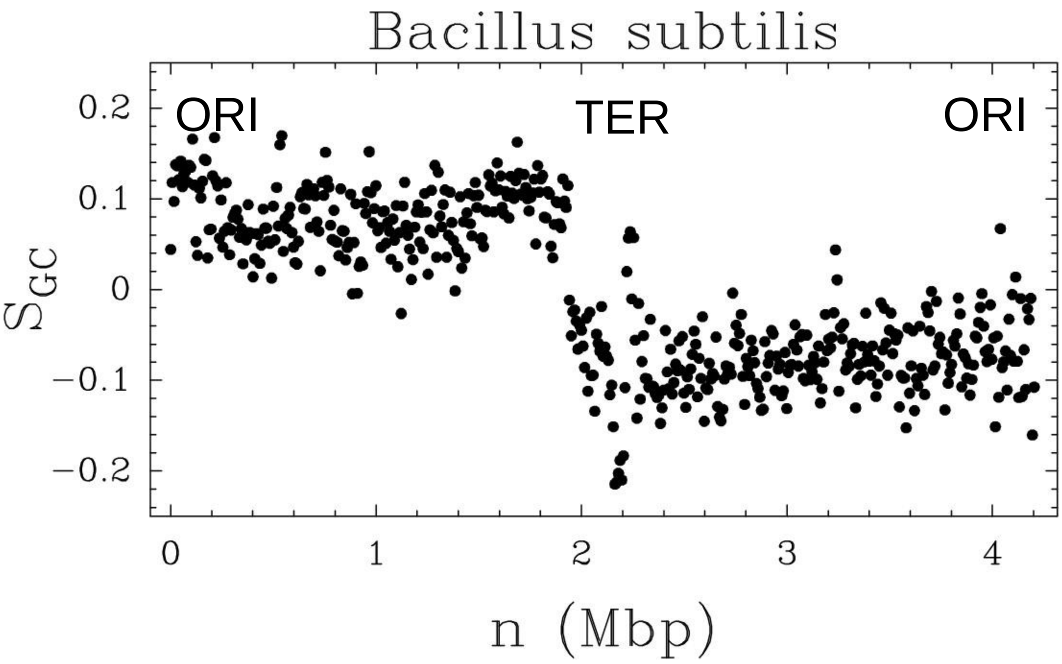


**Sense genes, anti-sense genes**

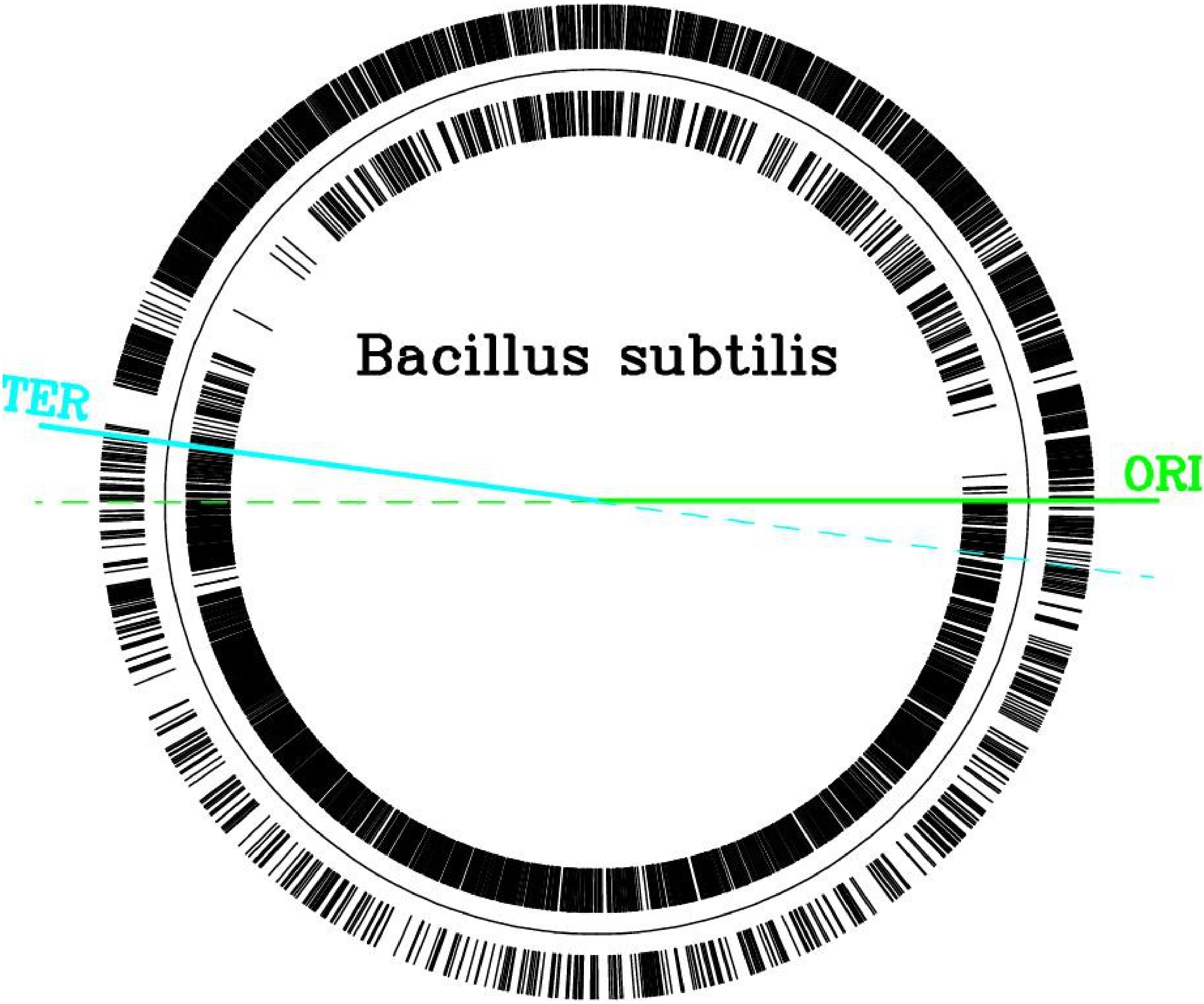
# Transcription bias



# Replication bias in bacteria



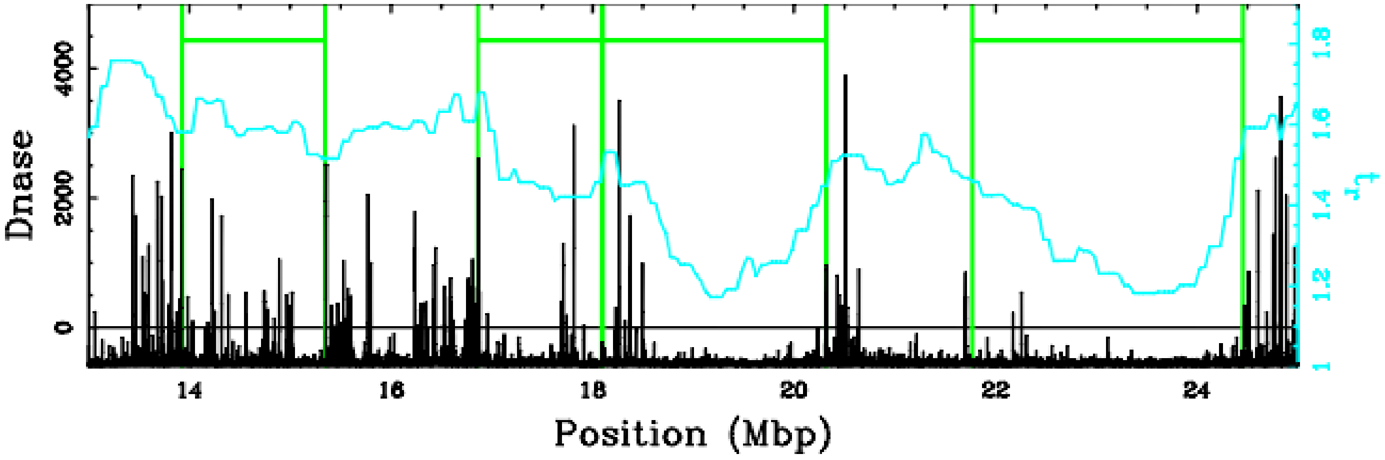
# Replication and genome organization

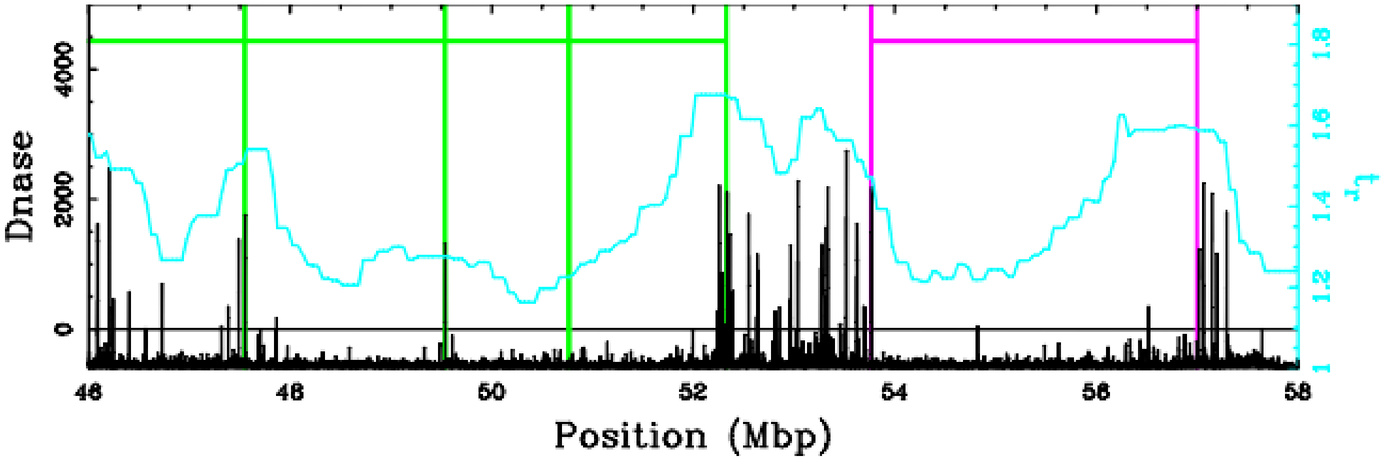


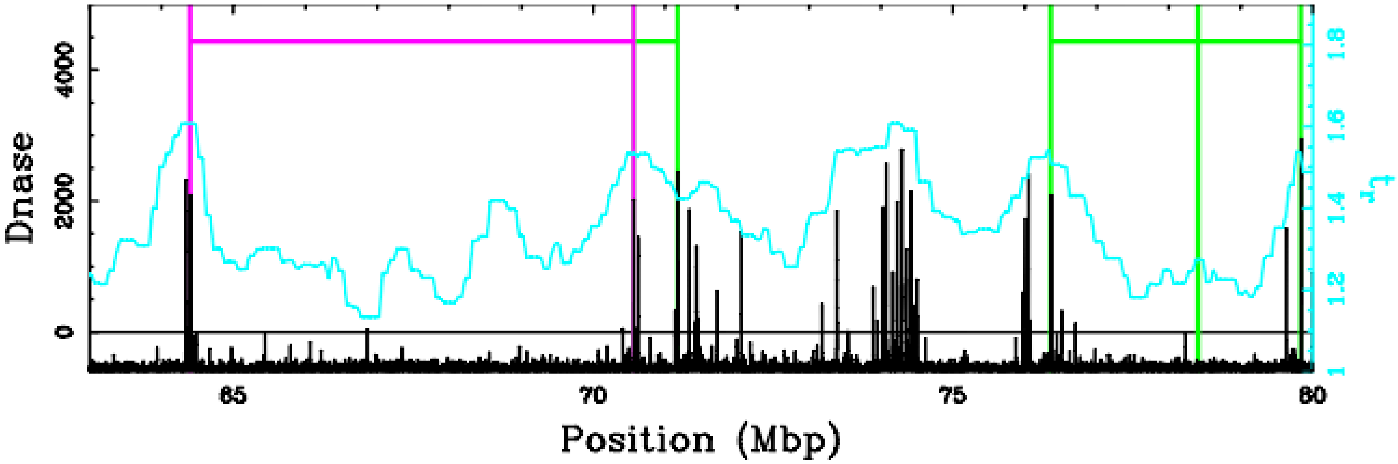
# High-resolution mapping of Open Chromatin

Boyle et al., Cell 132, 311 (2008)

Human chromosome 6



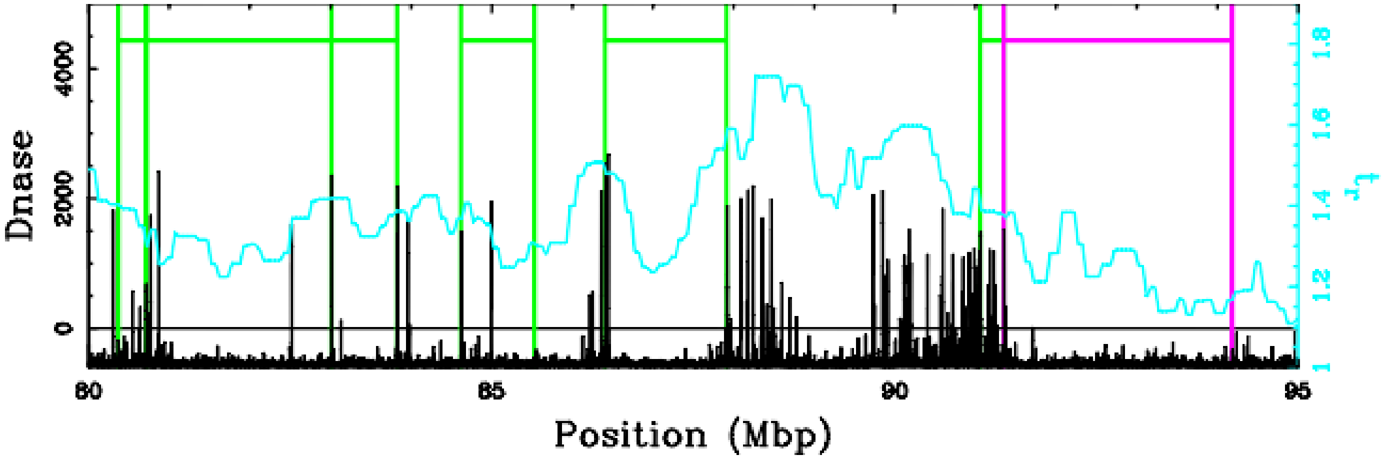


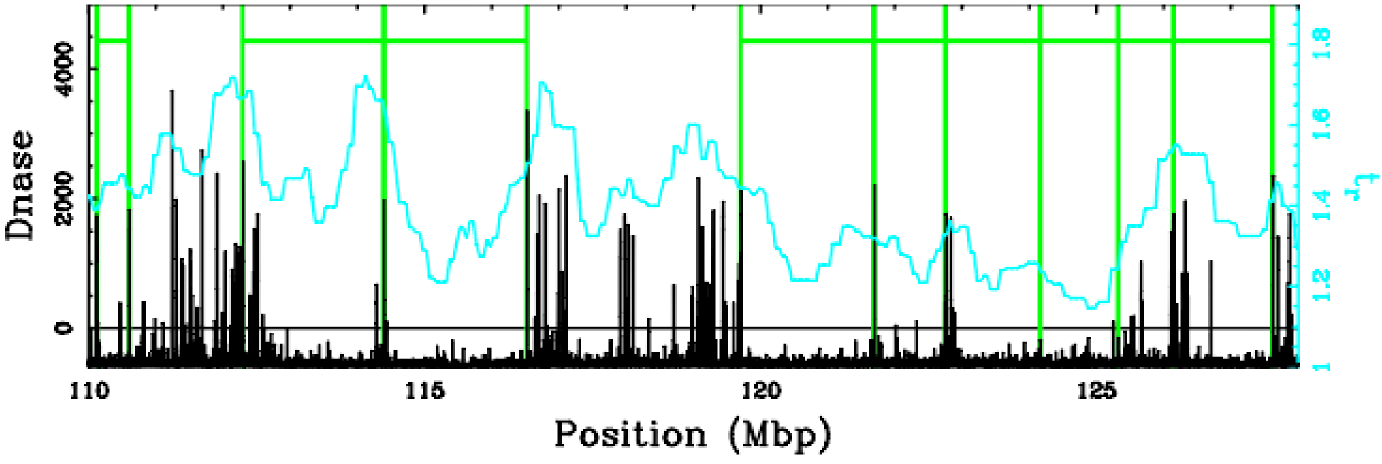


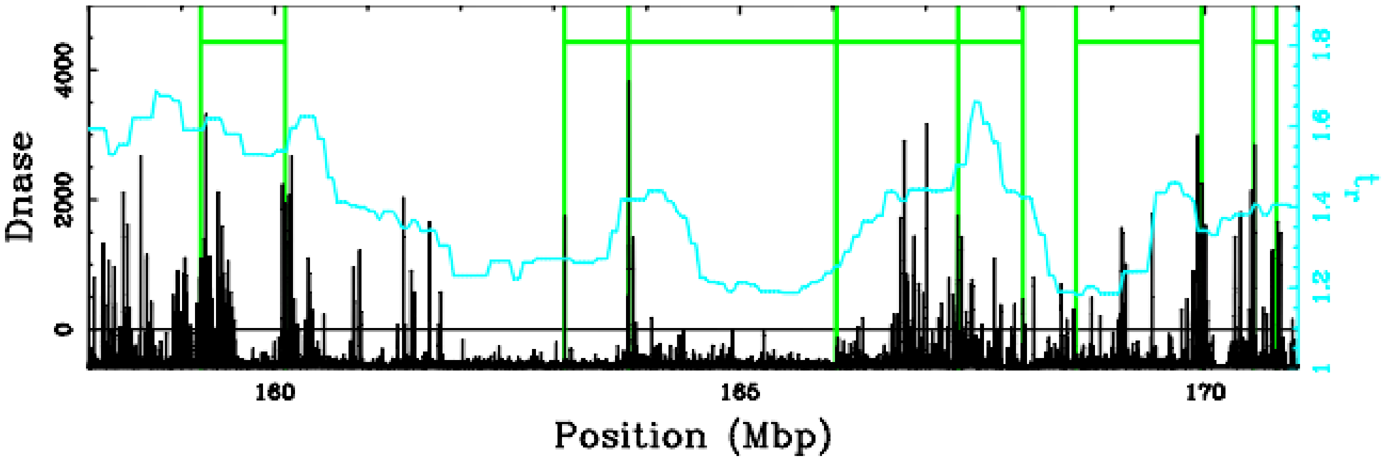
# High-resolution mapping of Open Chromatin

Boyle et al., Cell 132, 311 (2008)

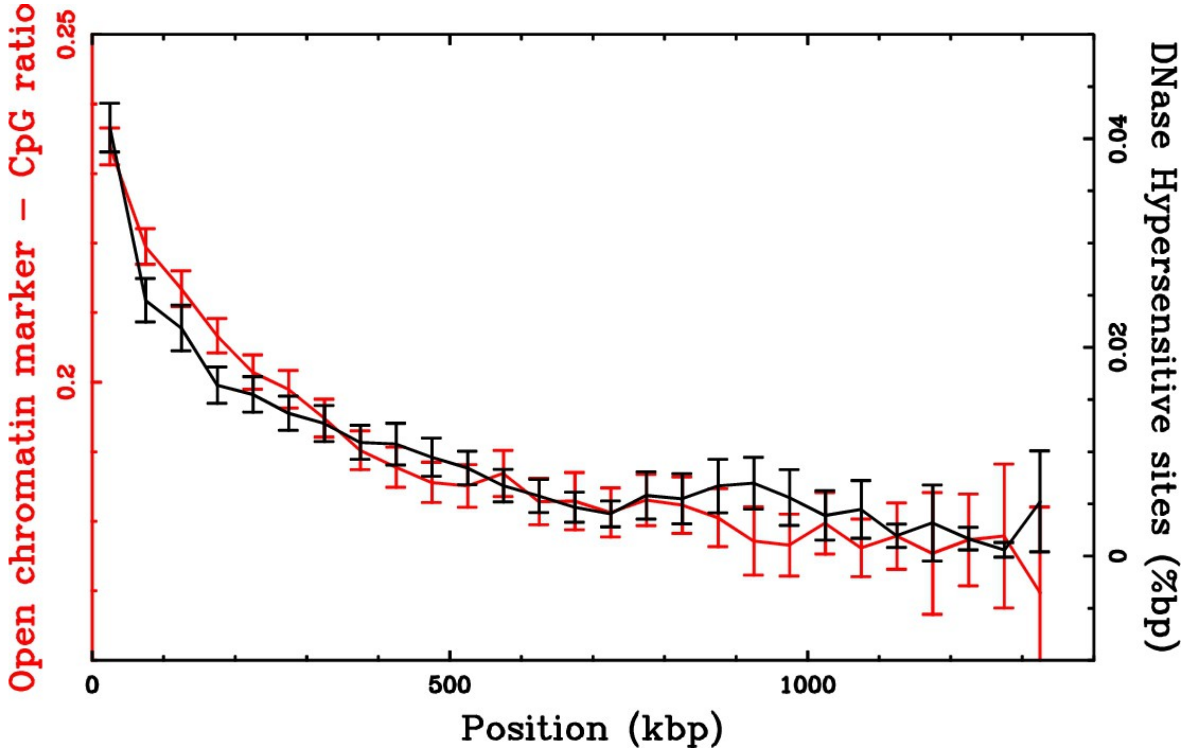
Human chromosome 6







# The sequence is highly predictive of open chromatin around putative replication origins



# The sequence is highly predictive of open chromatin around putative replication origins