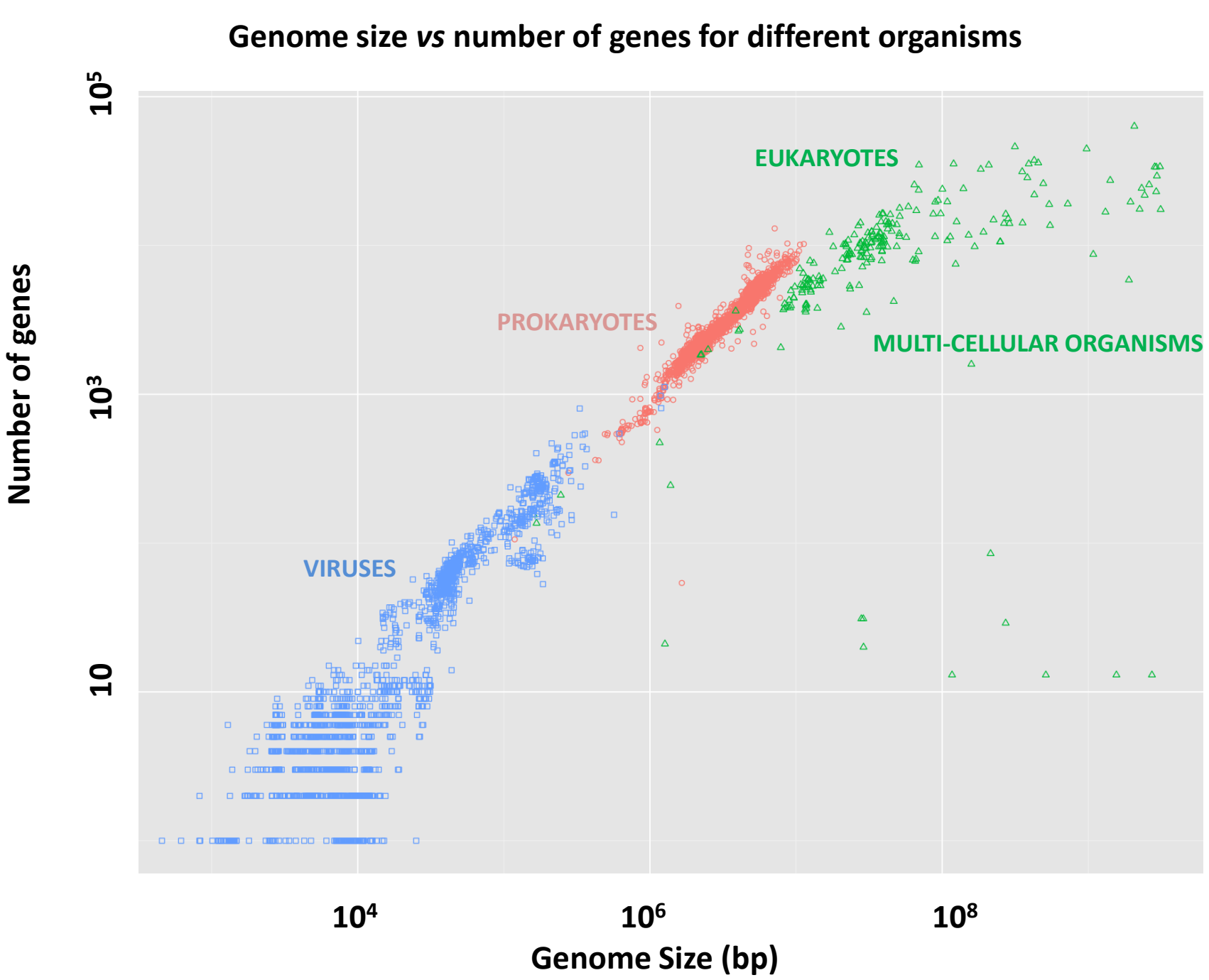
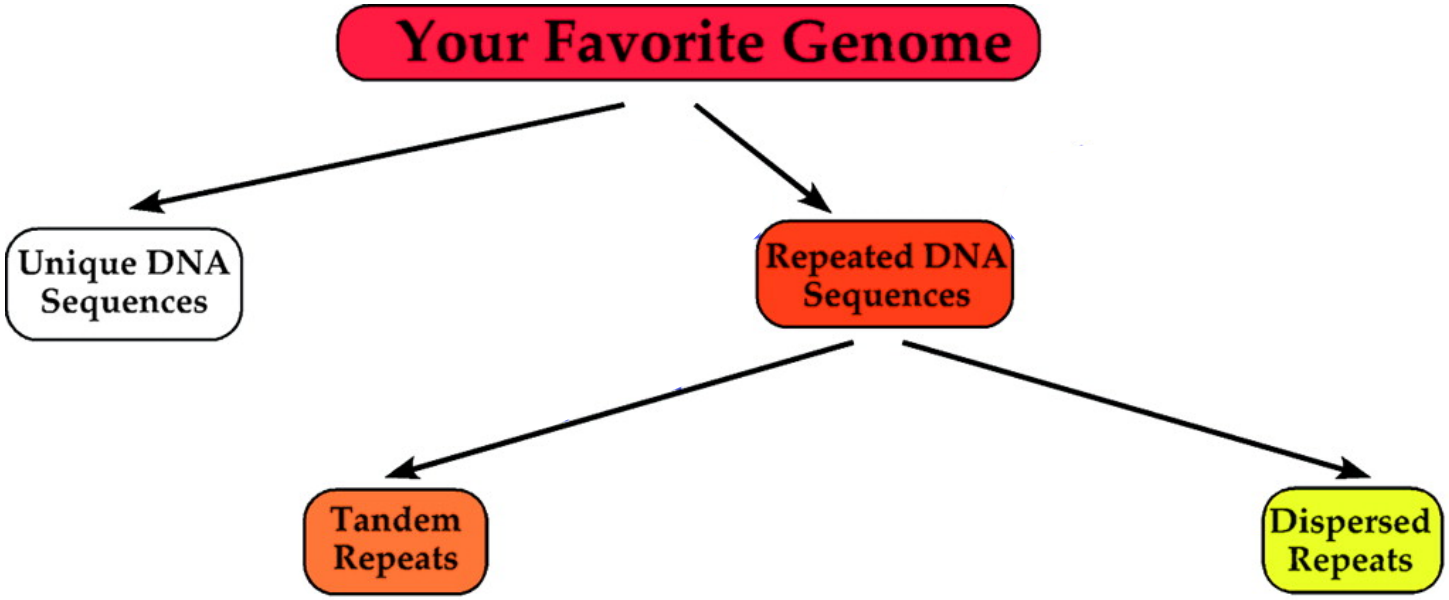
Rencontres Scientifiques des Grands Causses
****

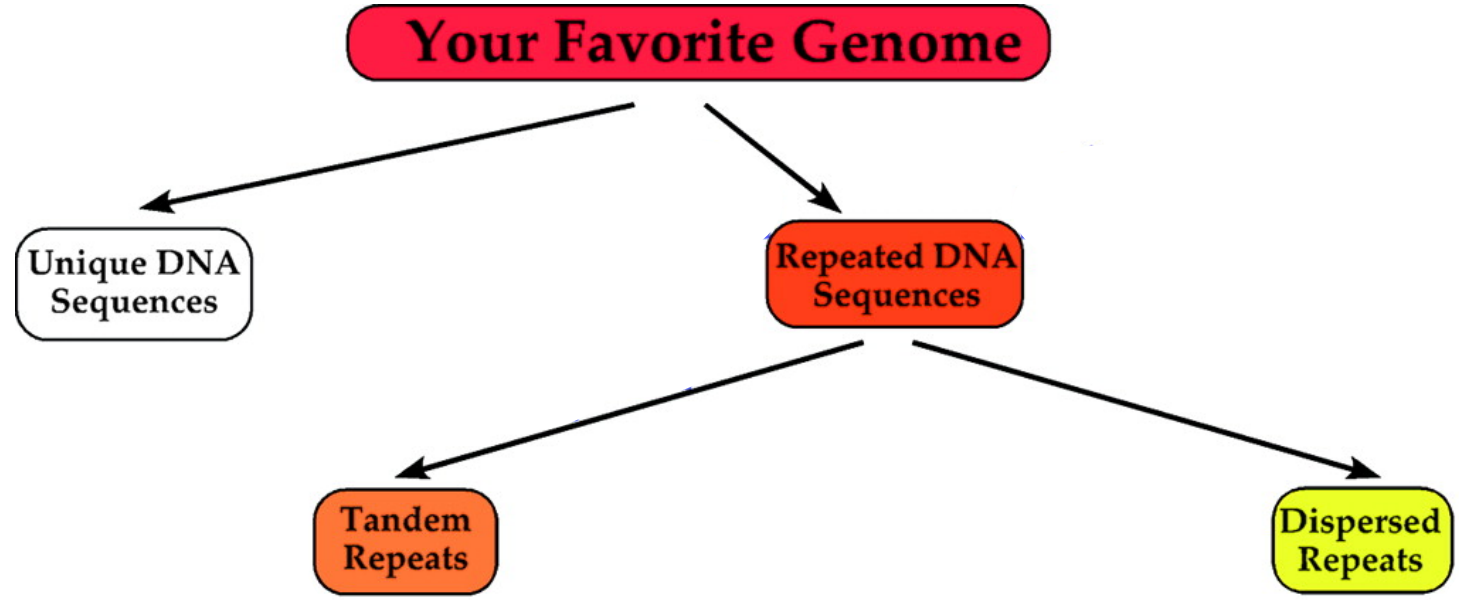
The role of DNA repeats in metazoans genomes

Mozziconacci J.
LPTMC, Université Pierre et Marie Curie

Cournac A. Koszul R.
Institut Pasteur


Genome size *vs* number of genes for different organisms
Number of genes
10^5
10^3
10
EUKARYOTES
PROKARYOTES
MULTI-CELLULAR ORGANISMS
VIRUSES
10^4
10^6
10^8
Genome Size (bp)


Your Favorite Genome
Unique DNA Sequences
Repeated DNA Sequences
Tandem Repeats
Dispersed Repeats



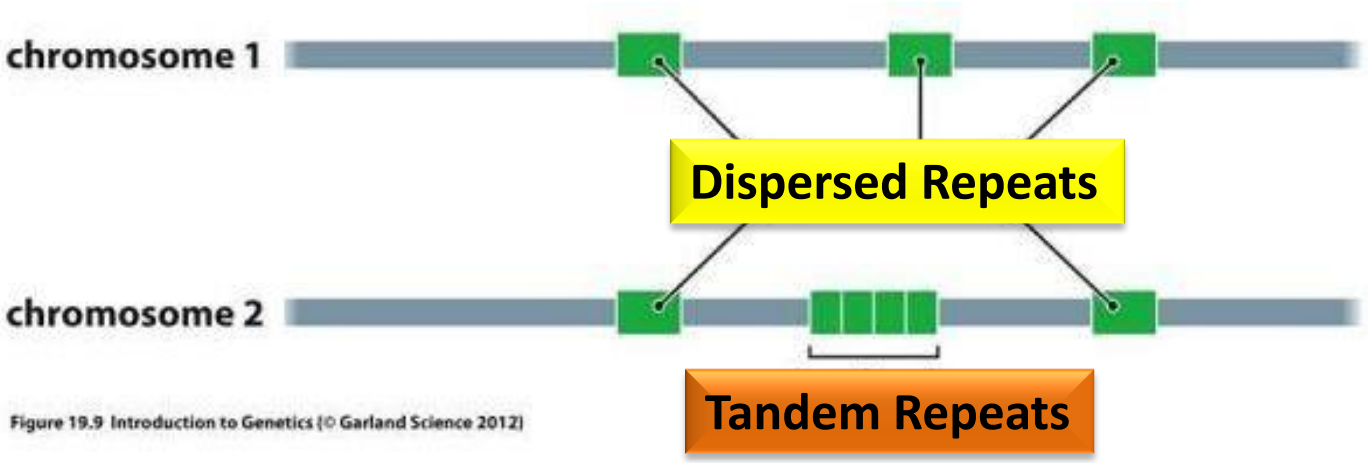


Figure 19.9 Introduction to Genetics (© Garland Science 2012)

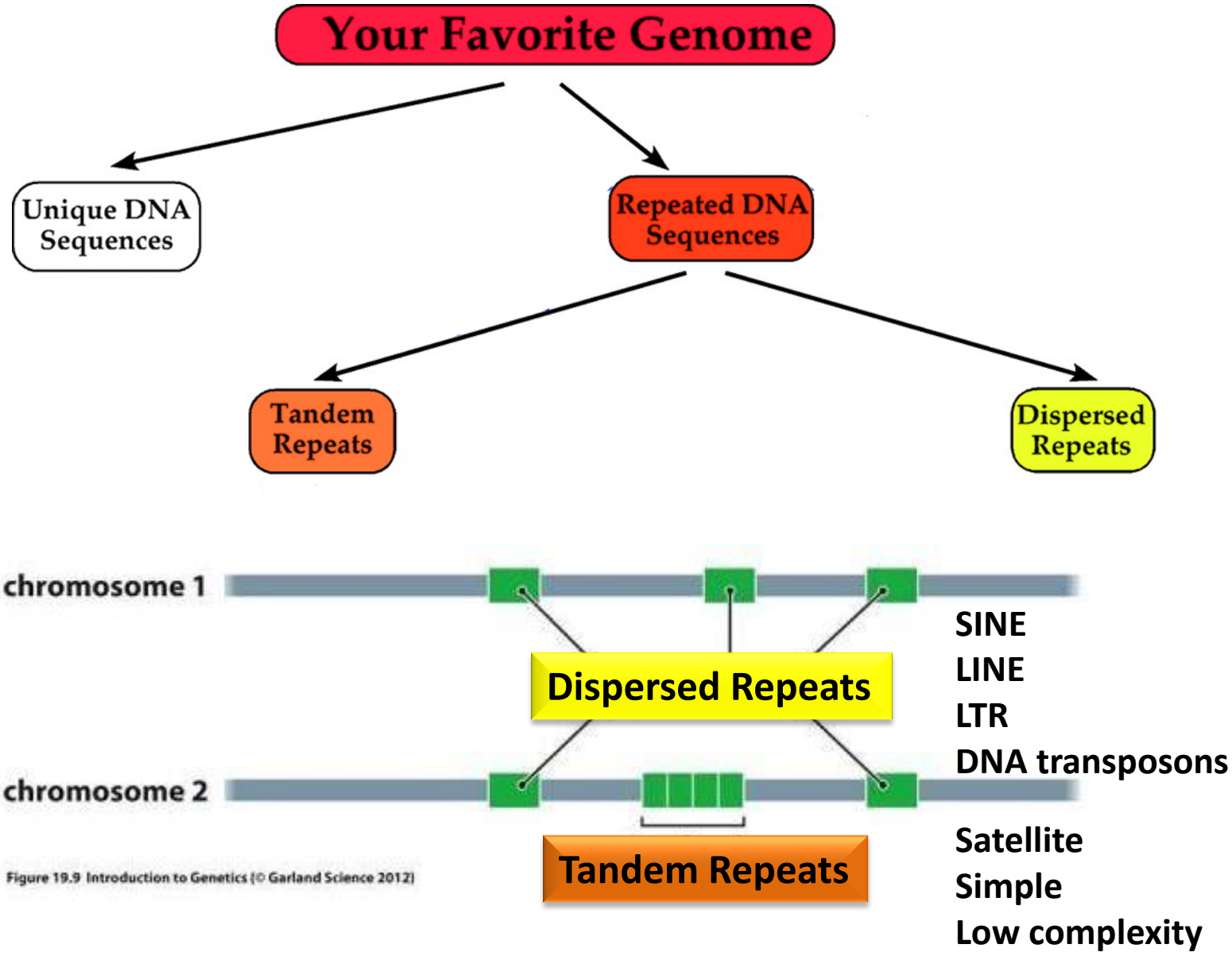
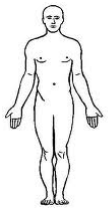


Figure 19.9 Introduction to Genetics (© Garland Science 2012)

There are **1395** different repetitive elements in the human genome (Repbase)

Your Favorite Genome



Adapted from Richard *et al.* Microbiol Mol Biol Rev. 2008

Unique DNA Sequences

30 % (only 2% are coding genes)

Repeated DNA Sequences

55 % of the genome

Tandem Repeats

15 % of the genome

Dispersed Repeats

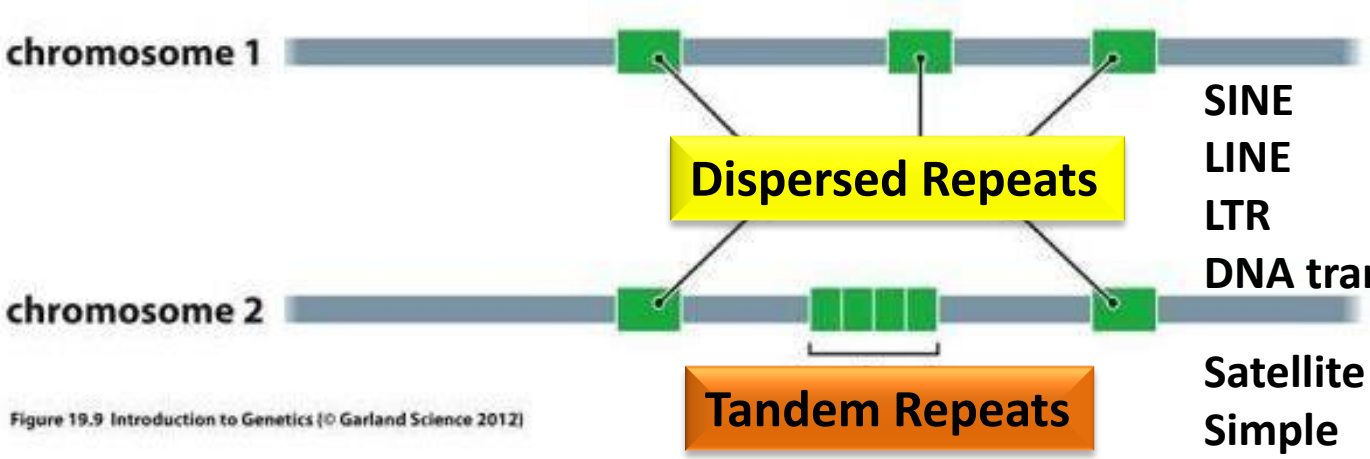


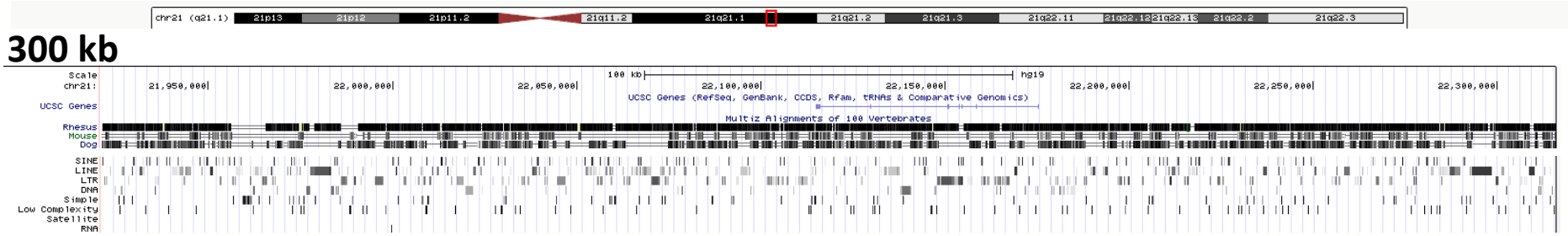
Figure 19.9 Introduction to Genetics (© Garland Science 2012)

SINE
LINE
LTR
DNA transposons

Satellite
Simple
Low complexity

There are **1395** different repetitive elements in the human genome (Repbase)

Example of repetitive DNA elements in the genomic context



Example of repetitive DNA elements in the genomic context

Example of repetitive DNA elements in the genomic context

Why are repeated elements so common in metazoan and plants genomes ?

1) **They are just junk/selfish elements**

2) **They have a function and are selected by evolution**

1956

In a recent brief review (McClintock, 1956), a description was given of types of elements carried in the maize chromosomes that serve to control gene action and to induce, at the site of the gene, heritable modifications affecting this action. These elements were initially discovered because they do not remain at one position in the chromosome complement. They can appear at new locations and disappear from previously determined locations. The presence of one such element at or near the locus of a known gene may affect the action of this gene. In so doing, it need not alter the action potentials of the genic substances at the locus. Therefore, these elements were called controlling elements. It was also shown that controlling elements fall into groups, the members of each operating as an integrated system in the control of gene action.

Evolution of the mammalian transcription factor binding repertoire via transposable elements

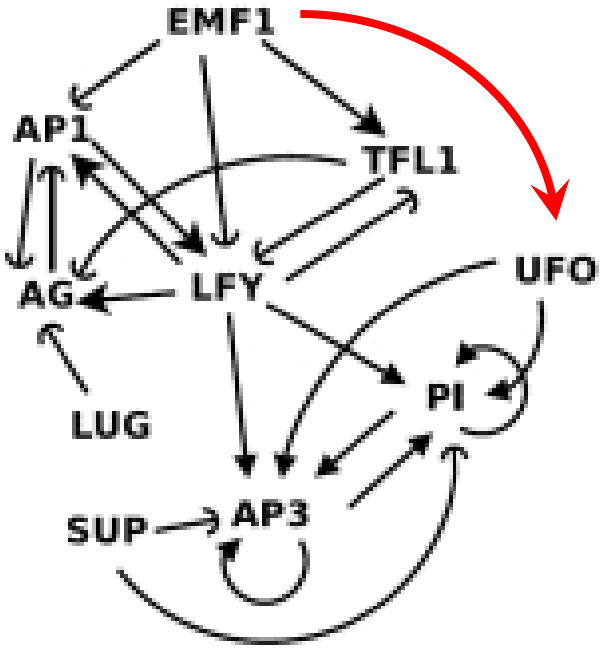
Guillaume Bourque,[1,5] Bernard Leong,[1] Vinsensius B. Vega,[1] Xi Chen,[2] Yen Ling Lee,[3] Kandhadayar G. Srinivasan,[3] Joon-Lin Chew,[2] Yijun Ruan,[3] Chia-Lin Wei,[3] Huck Hui Ng,[2] and Edison T. Liu[4]

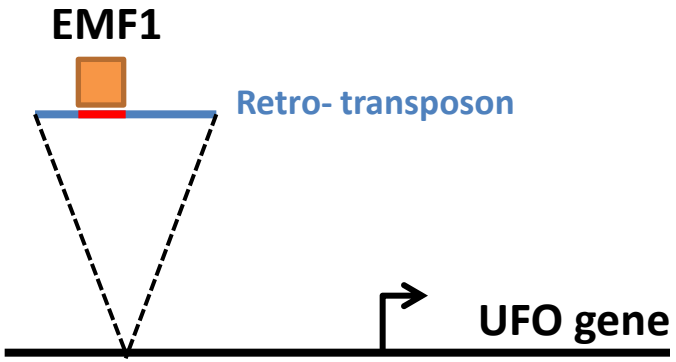
[1]Computational and Mathematical Biology, Genome Institute of Singapore, Singapore 138672, Singapore; [2]Stem Cell and Developmental Biology, Genome Institute of Singapore, Singapore 138672, Singapore; [3]Genome Technology and Biology, Genome Institute of Singapore, Singapore 138672, Singapore; [4]Cancer Biology and Pharmacology, Genome Institute of Singapore, Singapore 138672, Singapore

Identification of lineage-specific innovations in genomic control elements is critical for understanding transcriptional regulatory networks and phenotypic heterogeneity. We analyzed, from an evolutionary perspective, the binding regions of seven mammalian transcription factors (ESR1, TP53, MYC, RELA, POU5F1, SOX2, and CTCF) identified on a genome-wide scale by different chromatin immunoprecipitation approaches and found that only a minority of sites appear to be conserved at the sequence level. Instead, we uncovered a pervasive association with genomic repeats by showing that a large fraction of the bona fide binding sites for five of the seven transcription factors (ESR1, TP53, POU5F1, SOX2, and CTCF) are embedded in distinctive families of transposable elements. Using the age of the repeats, we established that these repeat-associated binding sites (RABS) have been associated with significant regulatory expansions throughout the mammalian phylogeny. We validated the functional significance of these RABS by showing that they are over-represented in proximity of regulated genes and that the binding motifs within these repeats have undergone evolutionary selection. Our results demonstrate that transcriptional regulatory networks are highly dynamic in eukaryotic genomes and that transposable elements play an important role in expanding the repertoire of binding sites.

Genome research, 2008

Re-wiring gene regulatory networks





Flowering gene netwok in A. Thaliana (see Mendoza and Alvarez-Buylla, 1998; Mendoza *et al.*, 1999; Espinosa-Sotoa *et al.*, 2004).

Biol. Rev. (2005), **80**, pp. 227–250. © 2005 *Cambridge Philosophical Society*
doi:10.1017/S1464793104006657 *Printed in the United Kingdom*

Why repetitive DNA is essential to genome function

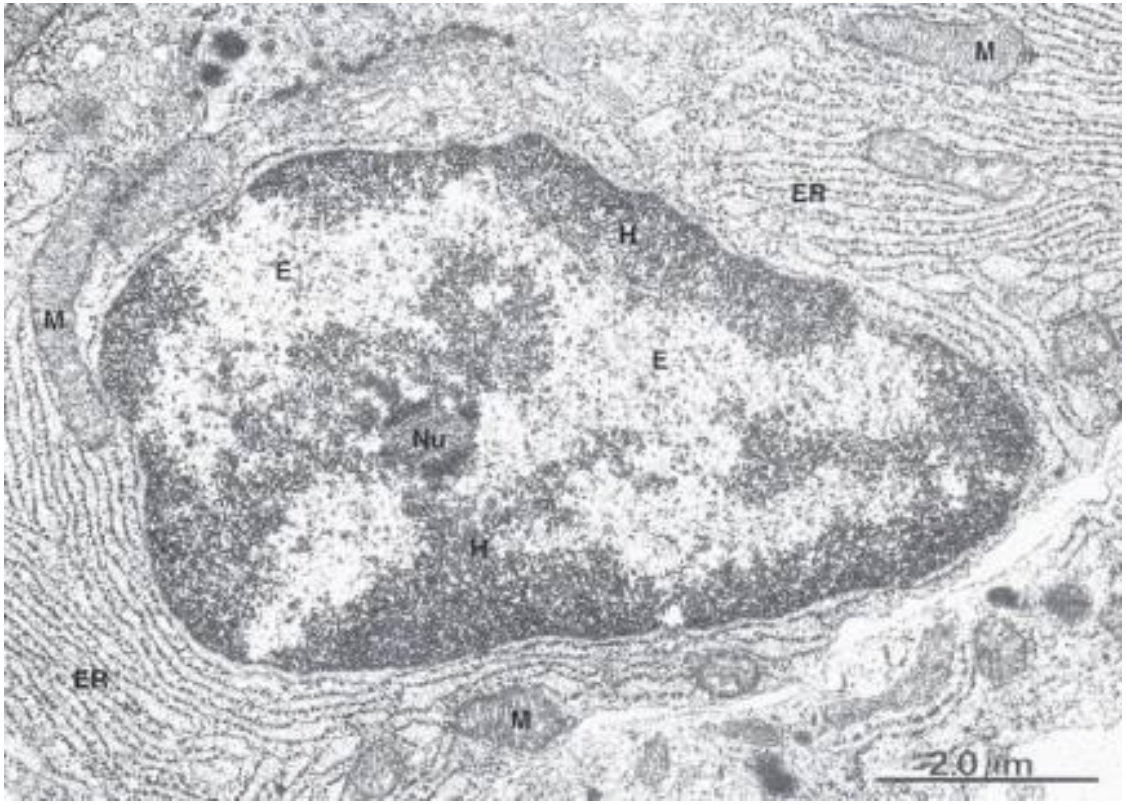
James A. Shapiro[1,*] and Richard von Sternberg[2,3]

[1] *Department of Biochemistry and Molecular Biology, University of Chicago*, 920 *E. 58th Street, Chicago, IL* 60637, *USA* (E-mail: jsha@uchicago.edu)
[2] *National Center for Biotechnology Information – GenBank Building* 45, *Room* 6*AN*.18*D*-30, *National Institutes of Health, Bethesda, Maryland* 20894 (E-mail: sternber@ncbi.nlm.nih.gov)
[3] *Department of Systematic Biology, NHB*-163, *National Museum of Natural History, Smithsonian Institution, Washington, D.C.*, 20013-7012 (E-mail: Sternberg.Richard@NMNH.SI.EDU)

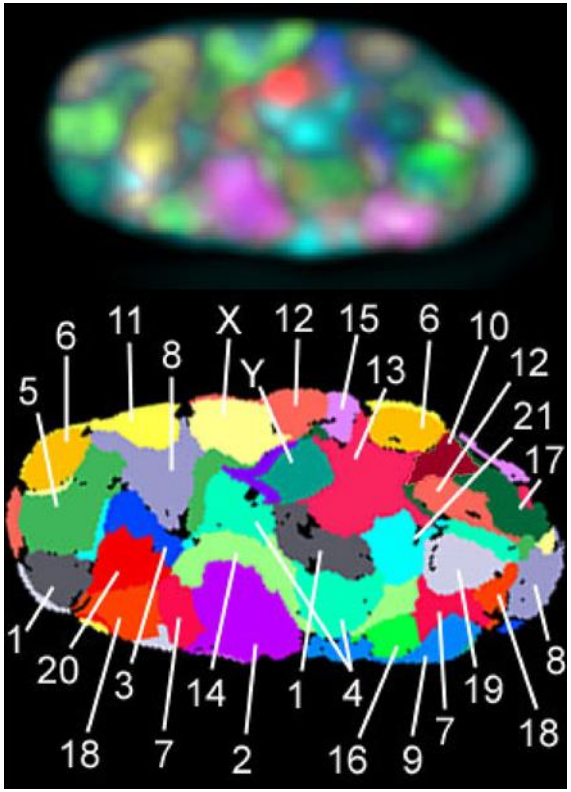
(*Received* 9 *March* 2004; *revised* 24 *September* 2004; *accepted* 29 *September* 2004)

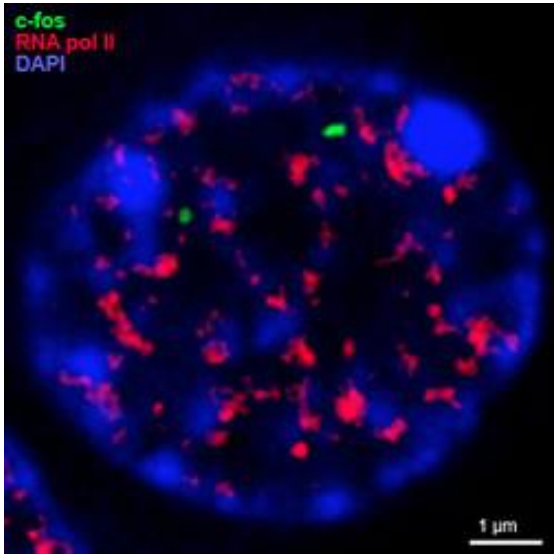
ABSTRACT

There are clear theoretical reasons and many well-documented examples which show that repetitive DNA is essential for genome function. Generic repeated signals in the DNA are necessary to format expression of unique coding sequence files and to organise additional functions essential for genome replication and accurate transmission to progeny cells. Repetitive DNA sequence elements are also fundamental to the cooperative molecular interactions forming nucleoprotein complexes. Here, we review the surprising abundance of repetitive DNA in many genomes, describe its structural diversity, and discuss dozens of cases where the functional importance of repetitive elements has been studied in molecular detail. In particular, the fact that repeat elements serve either as initiators or boundaries for heterochromatin domains and provide a significant fraction of scaffolding/matrix attachment regions (S/MARs) suggests that the repetitive component of the genome plays a major architectonic role in higher order physical structuring. Employing an information science model, the 'functionalist' perspective on repetitive DNA leads to new ways of thinking about the systemic organisation of cellular genomes and provides several novel possibilities involving repeat elements in evolutionarily significant genome reorganisation. These ideas may facilitate the interpretation of comparisons between sequenced genomes, where the repetitive DNA component is often greater than the coding sequence component.



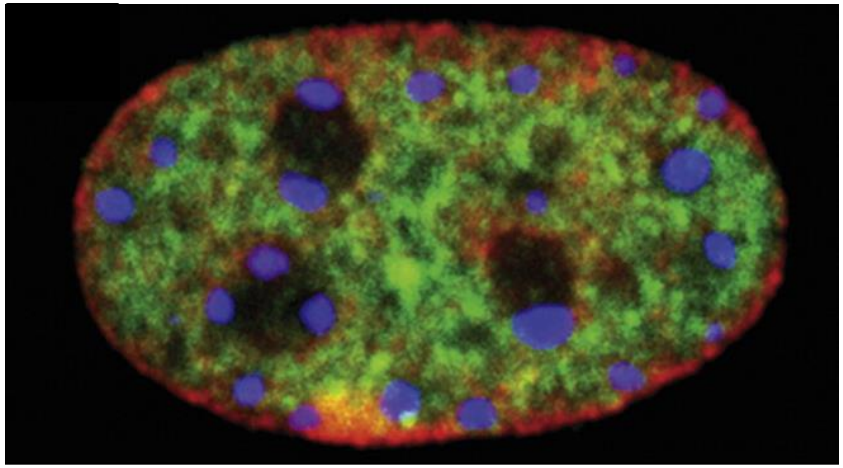
T Cremer





A Riccio

Nuclear structure seen with the microscope



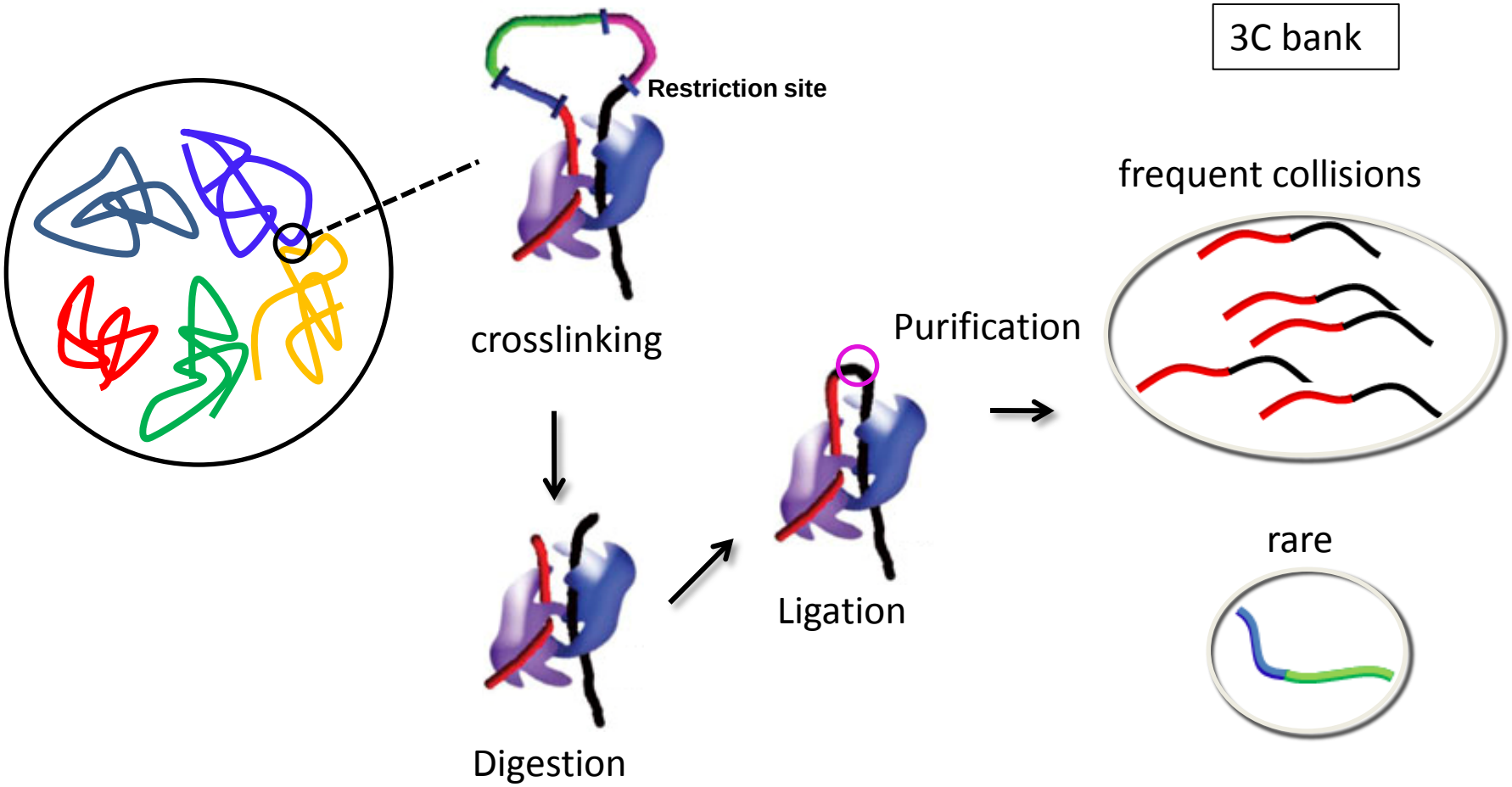
Nuclear architecture of Mouse fibroblasts (Solovei et al., Cell, 2009).

constitutive Heterochromatin
Satellite

Heterochromatin LINE

Euchromatin SINE

Measuring chromosomal contact experimentally: The Chromosome Conformation Capture experiment

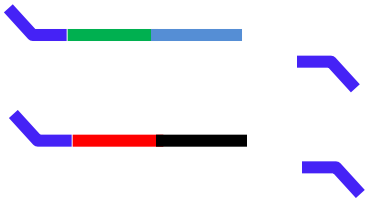


Axel Cournac, Marie-Nelly, Marbouty, Koszul, Mozziconacci, 2012
Duan et al., 2010
Dekker et al., l 2002

Measuring chromosomal contact experimentally:
The Chromosome Conformation Capture experiment

3C bank

frequent collisions

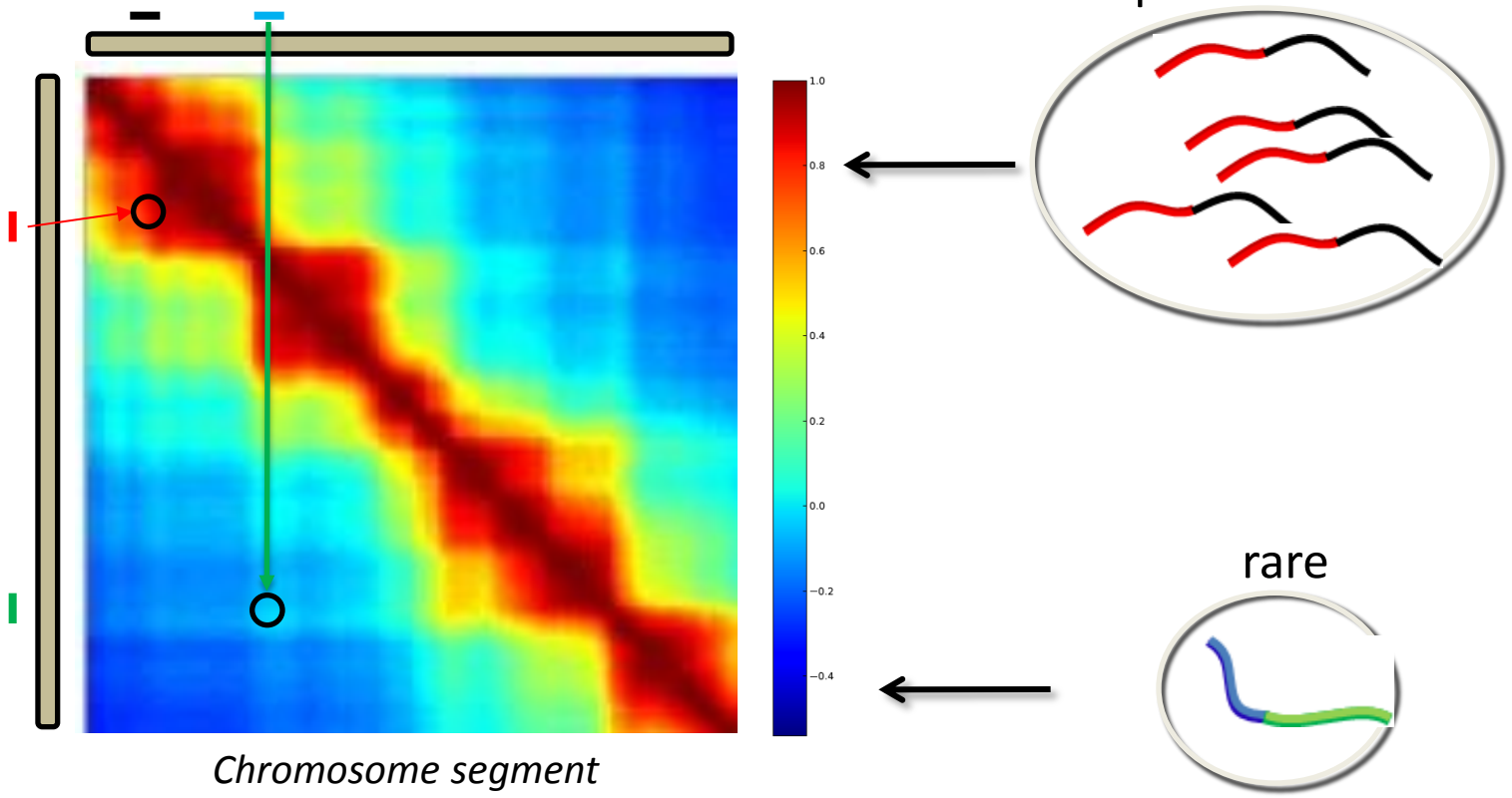


Illumina pair-end sequencing
couting the hits
statistical analysis

rare

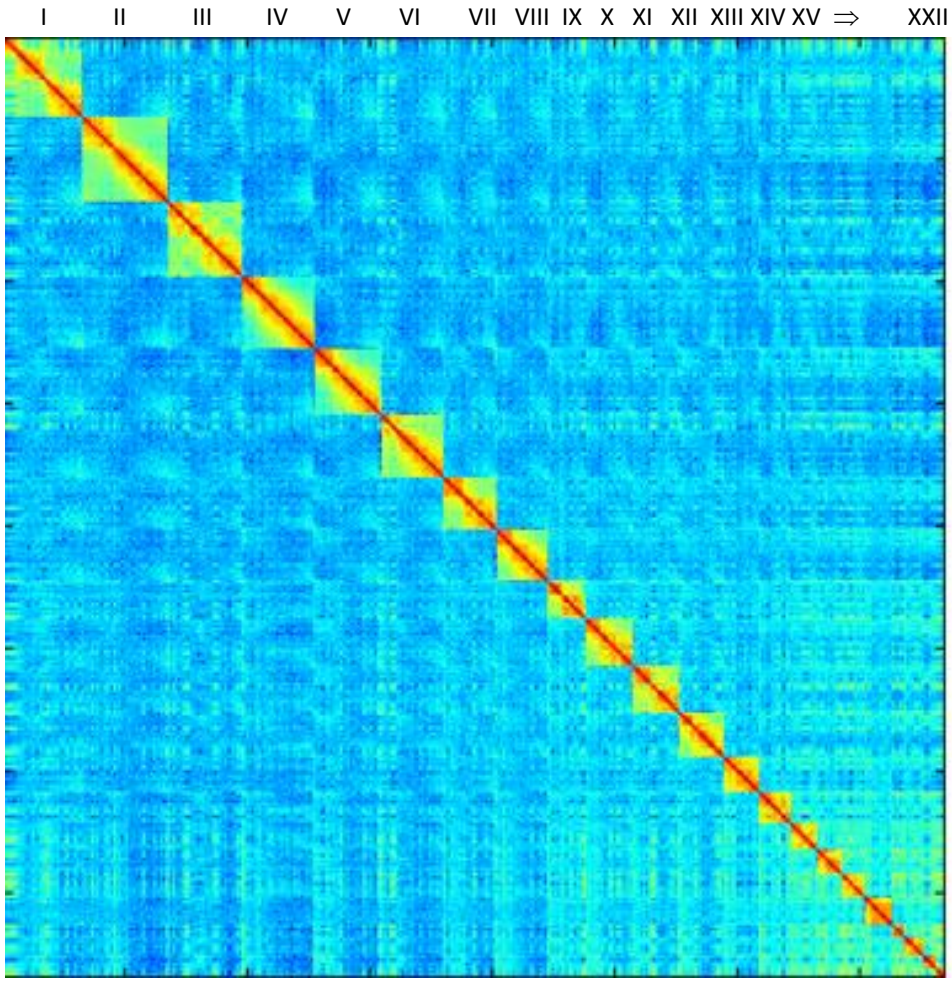
Cournac et al., 2012
Duan et al., 2010
Rodley et al., 2009
Dekker et al., 2002

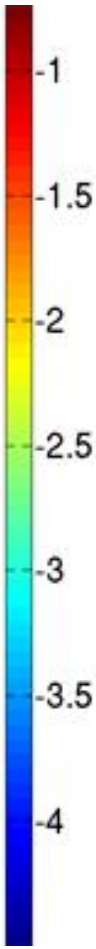
3C library



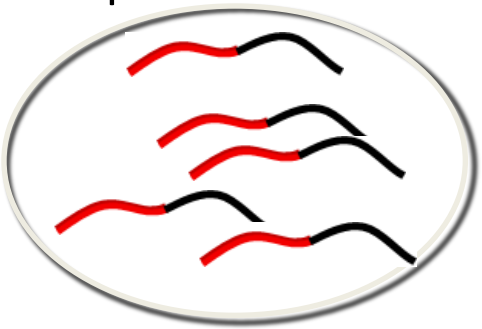
Lieberman-Aiden et al., 2009; Dixon et al., 2011; Khalor et al., 2012; Dekker et al., 2002

3C library





frequent collisions



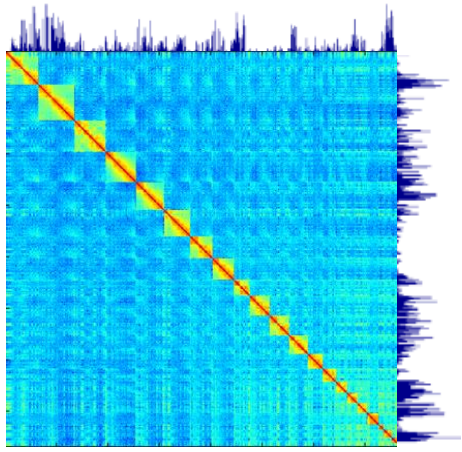
rare

Human genome, hESC

(data from Dixon et al., 2011, reprocessed Cournac and Mozziconacci)

Lieberman-Aiden et al., 2009; Dixon et al., 2011; Khalor et al., 2012; Dekker et al., 2002

Co-localization Score (CS) of all the different repetitive elements found in the human genome.

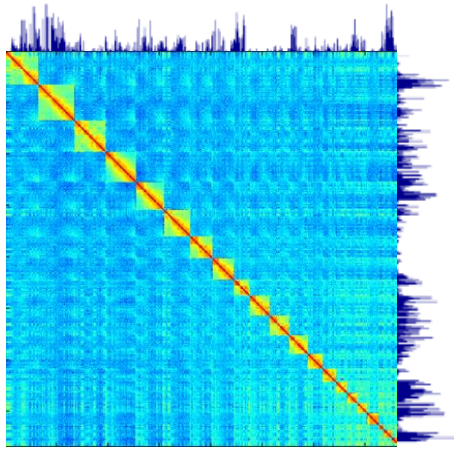


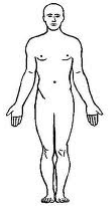
$CS_{AluJB} = \langle Mij \rangle_{ij\ containing\ AluJB}$

- 30 kbp bins (80.000 bins)
- Inter-chromosomal contacts

Cournac, Koszul and Mozziconacci, in revision

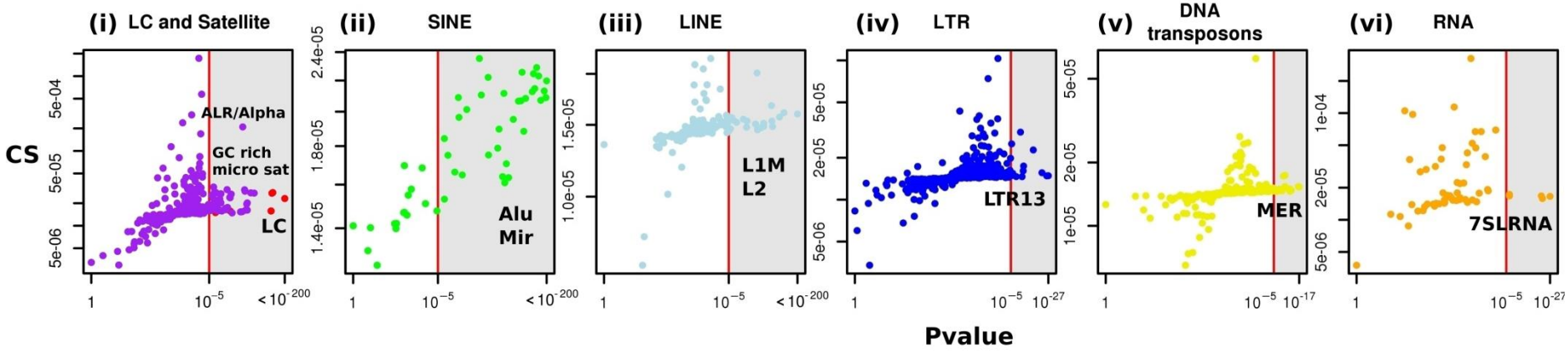
Co-localization score of all the different repetitive elements found in the human genome.





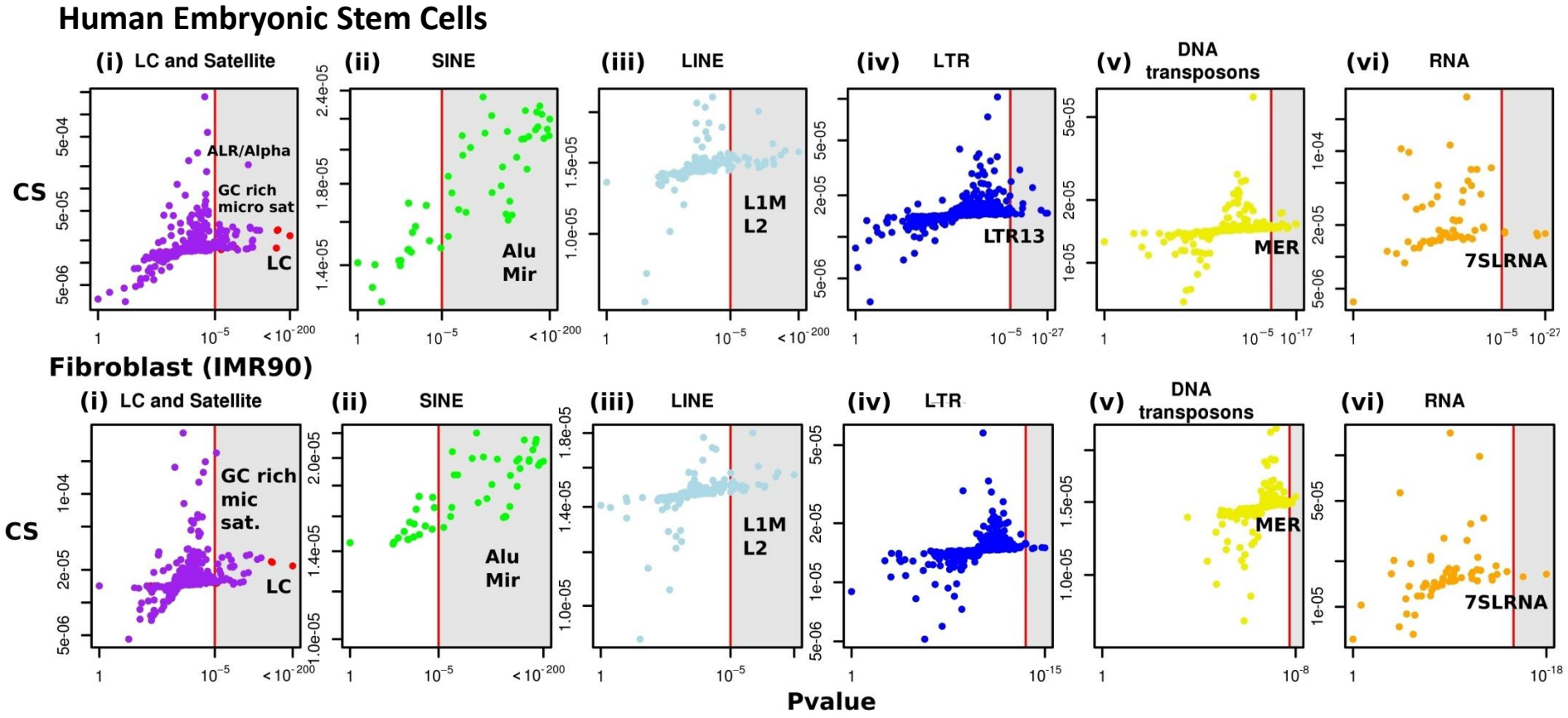
$$CS_{AluJB} = \langle Mij \rangle_{ij \text{ containing AluJB}}$$

Pvalues are determined using random swapping of repeats positions under various null models



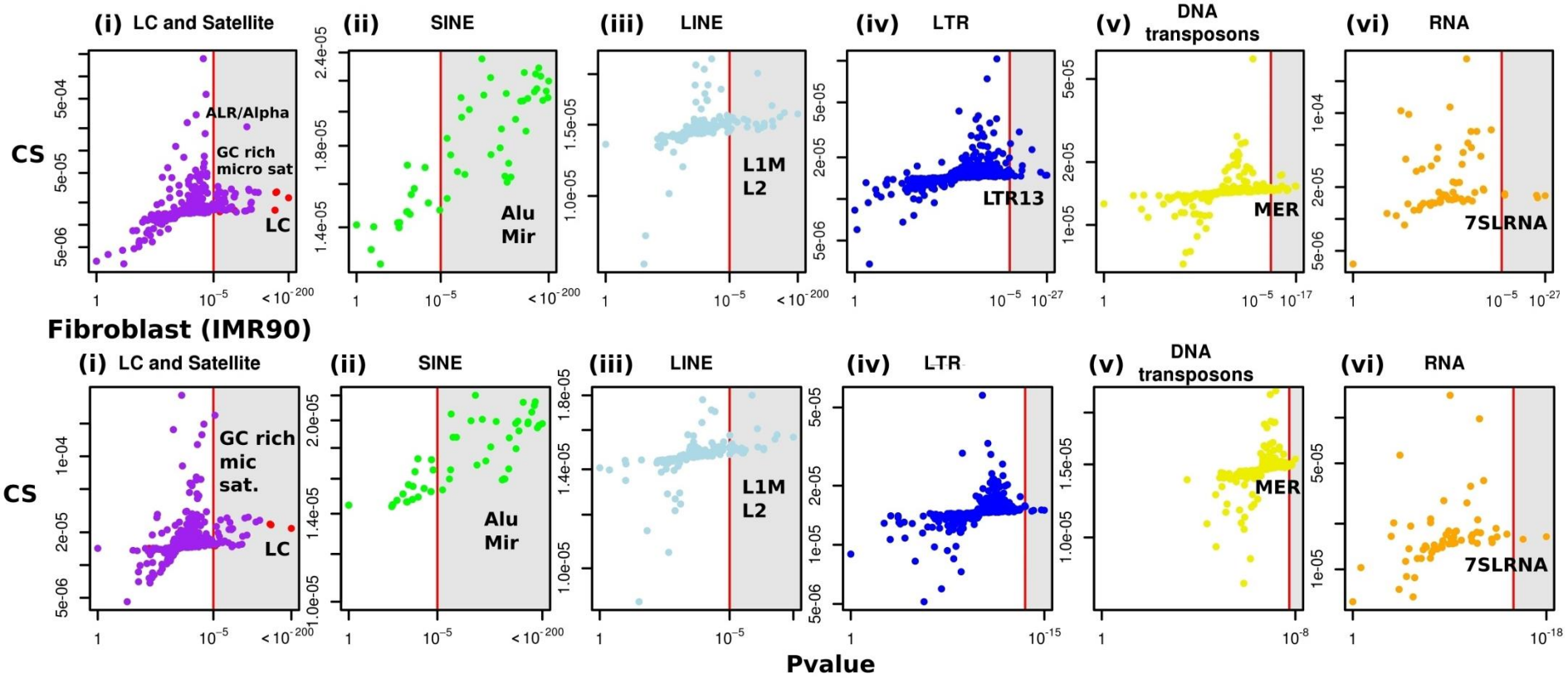
Cournac, Koszul and Mozziconacci, in revision

Comparison between two different cell types



Comparison between two different cell types

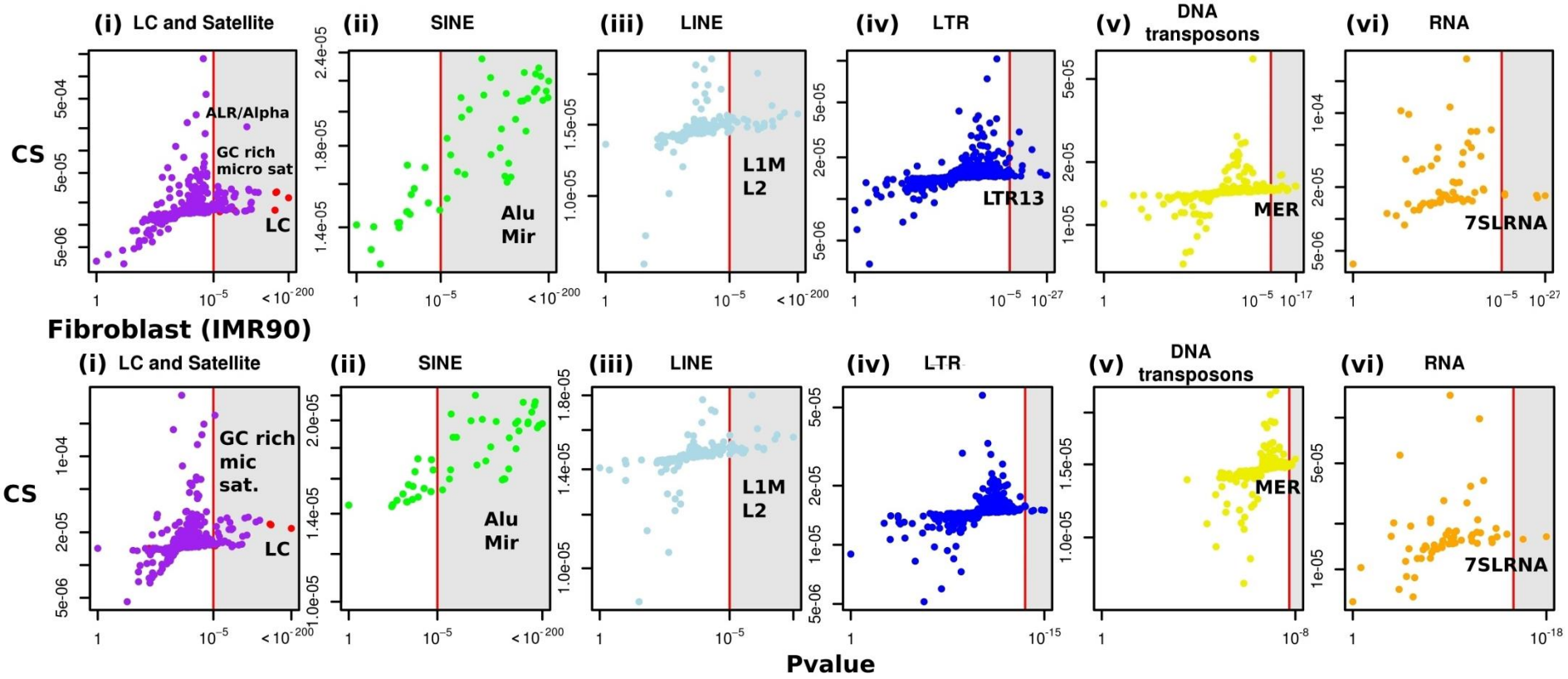
Human Embryonic Stem Cells



(ATGGTG)n	(CAGA)n	(CAGG)n	(CATG)n	(CCCCCA)n	(CCCTG)n	(CGGGGG)n					
(GGGAGA)n	(TCTCCC)n	(TGAA)n	(TTAA)n	(TTAGGG)n	(TTA)n	(GTGTG)n					
Charlie1a	Charlie18a	Charlie22a	Charlie4a	Charlie4z							
HAL1	Harlequin-int	HY3	HERVK9-int								
L1MA9	L1MD	L1MD2	L1ME2z	L1PA13	L1PA15	L3					
LTR10C	LTR13A	LTR56	LTR5B	LTR8							
MER102a	MER113	MER117	MER1A	MER20	MER20B	MER2A	MER21A	MER21B	MER21C	MER41A	
MER41B	MER58B	MER74B	MER91B	MER96							
MLT1F	MLT1J	MLT1B									
SVA_F	THE1B	U6	X3_LINE								

Comparison between two different cell types

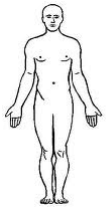
Human Embryonic Stem Cells

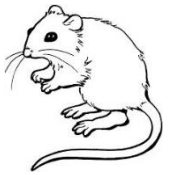


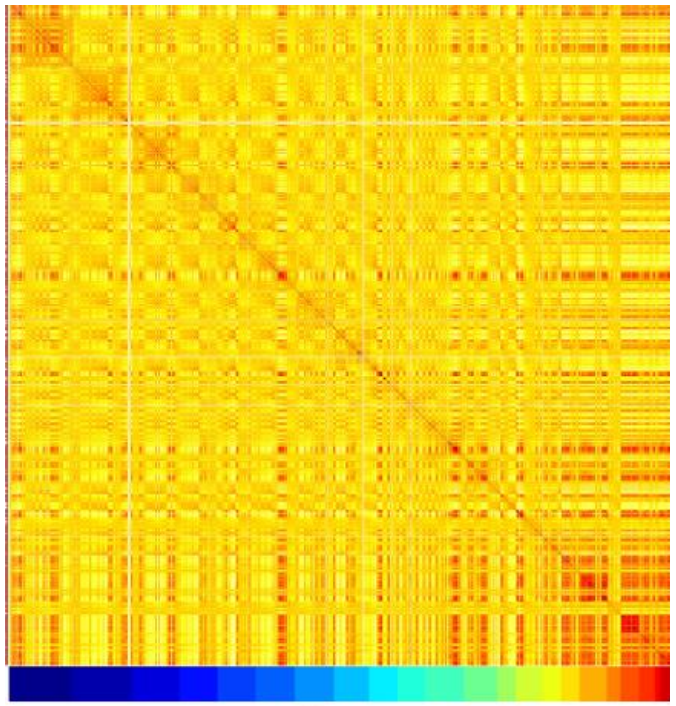
(ATGGTG)n	(CAGA)n	(CAGG)n	(CATG)n	(CCCCCA)n	(CCCTG)n	(CGGGGG)n					
(GGGAGA)n	(TCTCCC)n	(TGAA)n	(TTAA)n	(TTAGGG)n	(TTA)n	(GTGTG)n					
Charlie1a	Charlie18a	Charlie22a	Charlie4a	Charlie4z							
HAL1	Harlequin-int	HY3	HERVK9-int								
L1MA9	L1MD	L1MD2	L1ME2z	L1PA13	L1PA15	L3					
LTR10C	LTR13A	LTR56	LTR5B	LTR8							
MER102a	MER113	MER117	MER1A	MER20	MER20B	MER2A	MER21A	MER21B	MER21C	MER41A	
MER41B	MER58B	MER74B	MER91B	MER96							
MLT1F	MLT1J	MLT1B									
SVA_F	THE1B	U6	X3_LINE								

Oct4
Nanog
Both

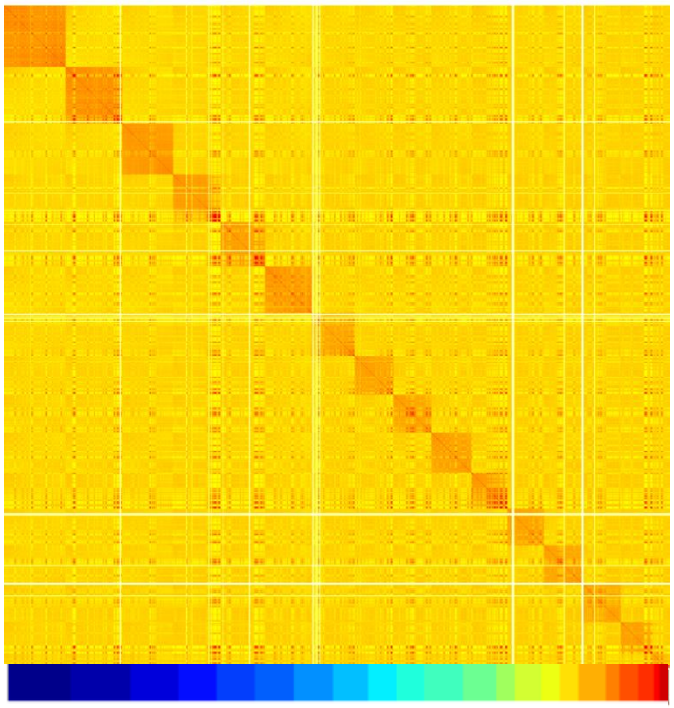
Comparison between mouse and human genome 3D organization





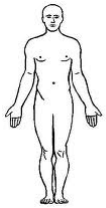


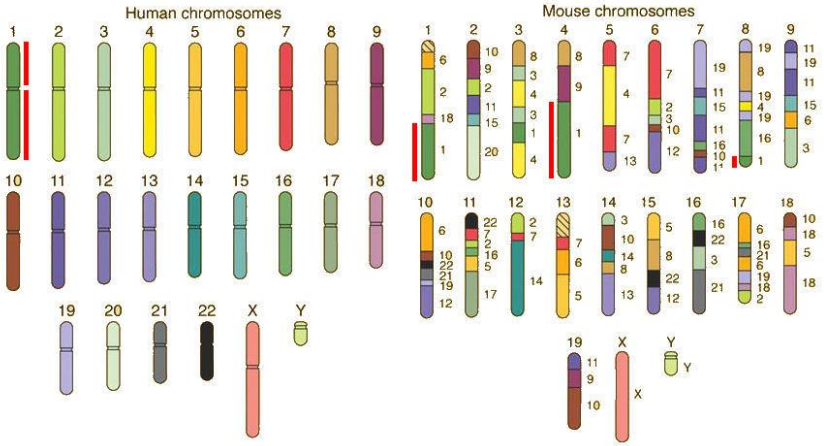
Human chromosomes

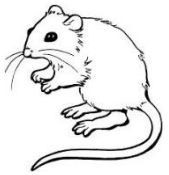


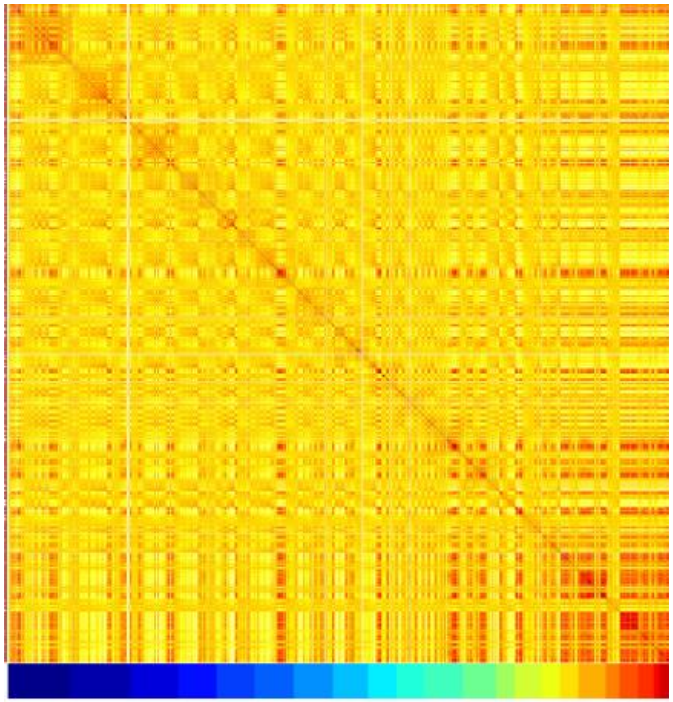
Mouse chromosomes

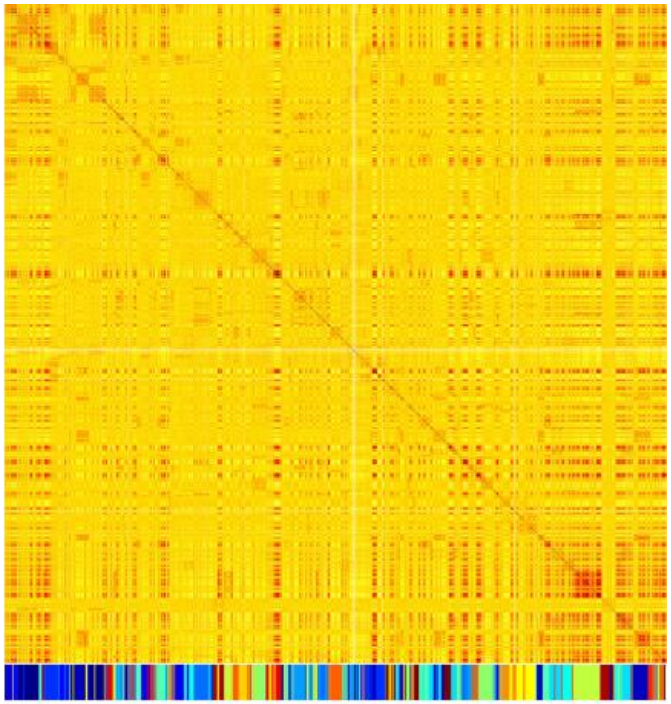
Whereas syntenic blocs are reshuffled in the mouse vs human genomes,
The 3D structure is conserved



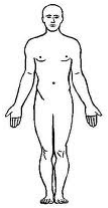


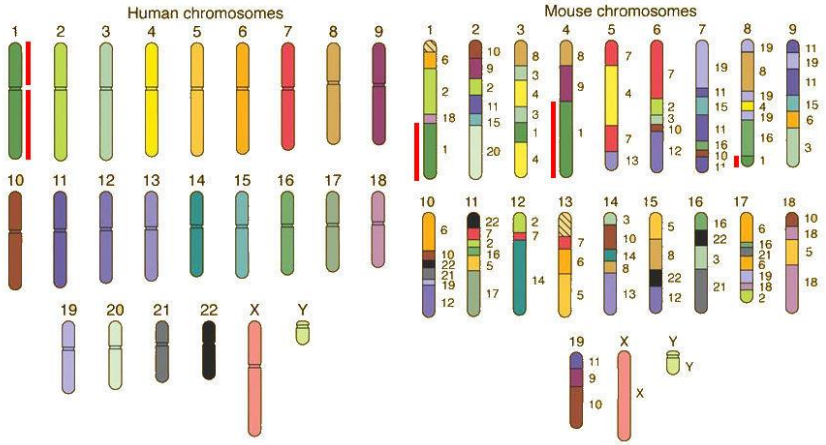


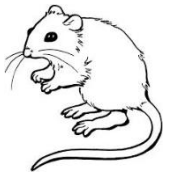




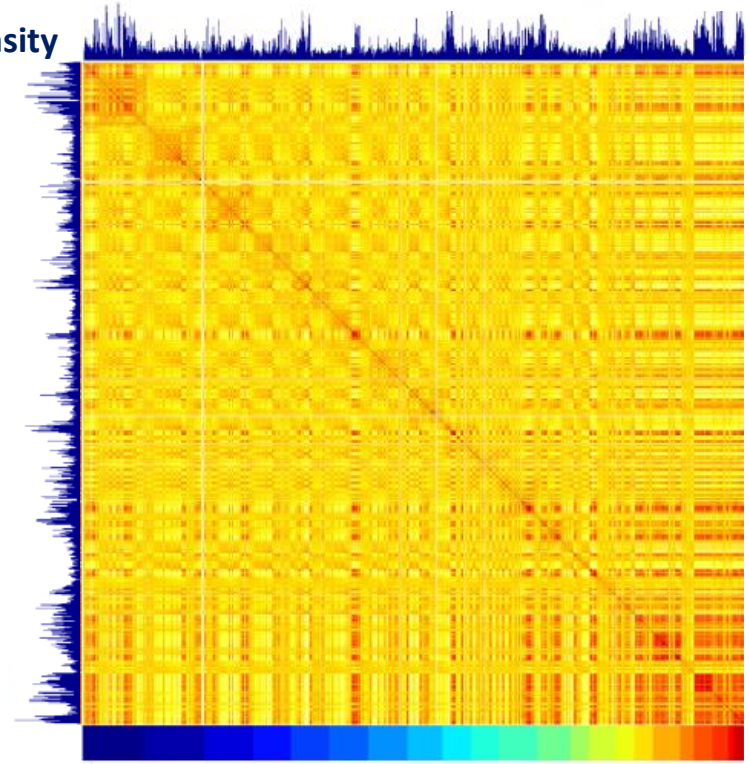
The conserved 3D structure correlates with the MIR density



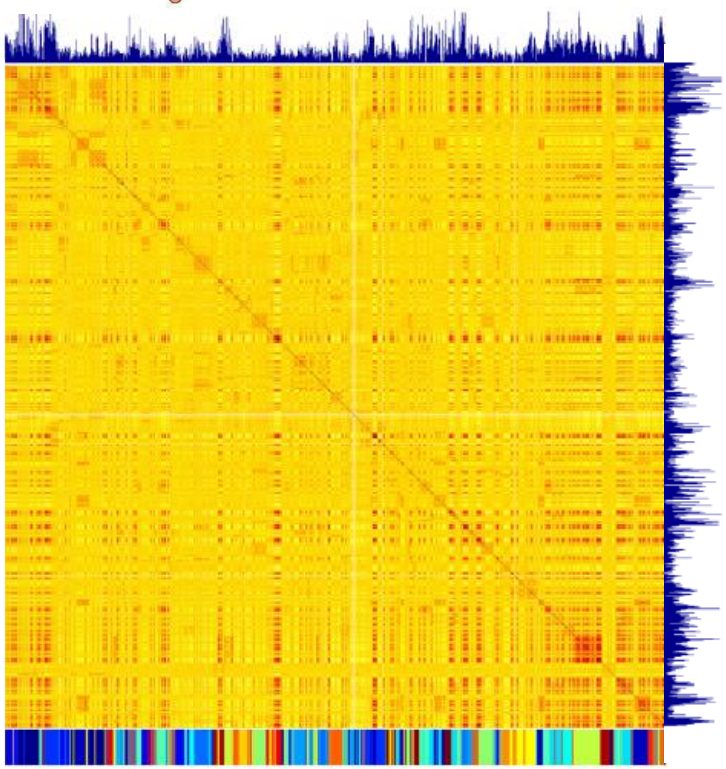




MIR density

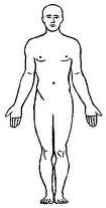


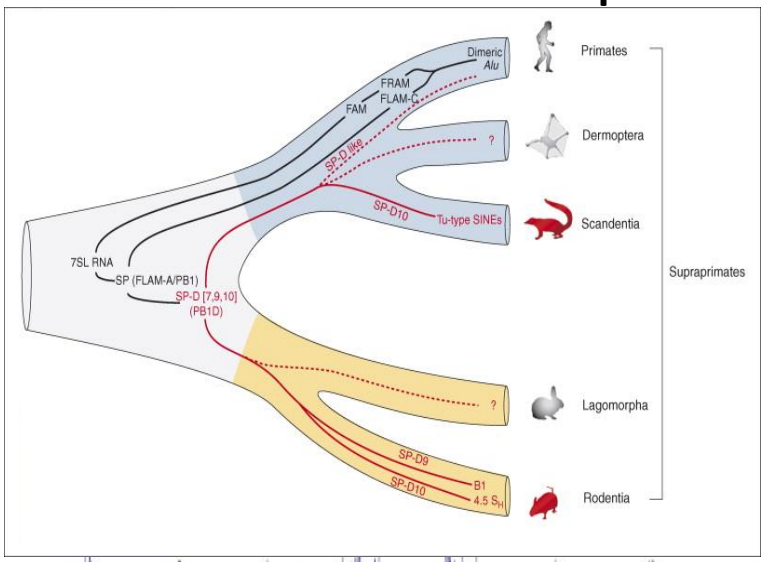
Human chromosomes

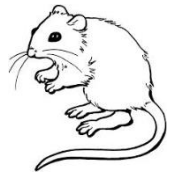


Mouse chromosomes

Human Alu VS rodent B1 sequences

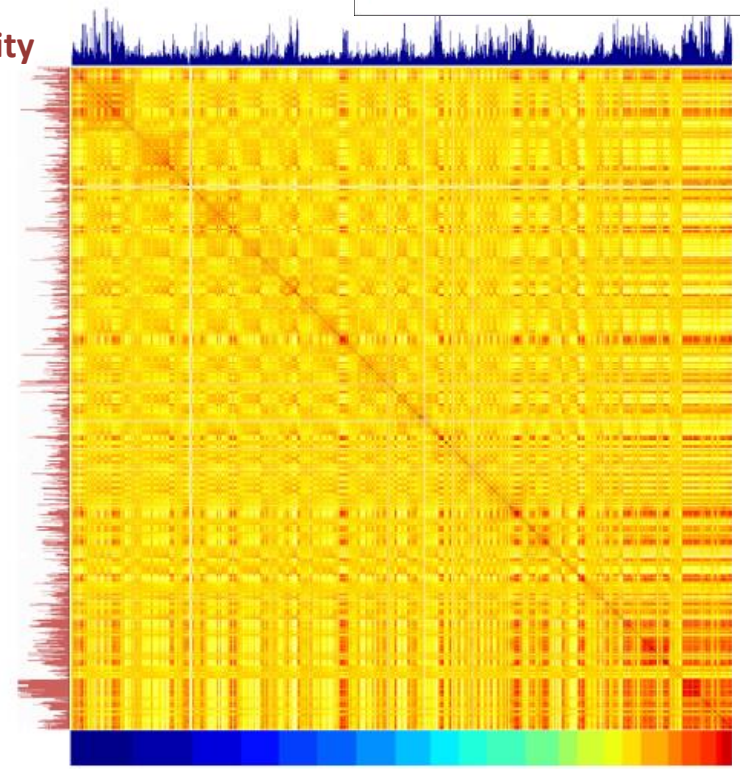






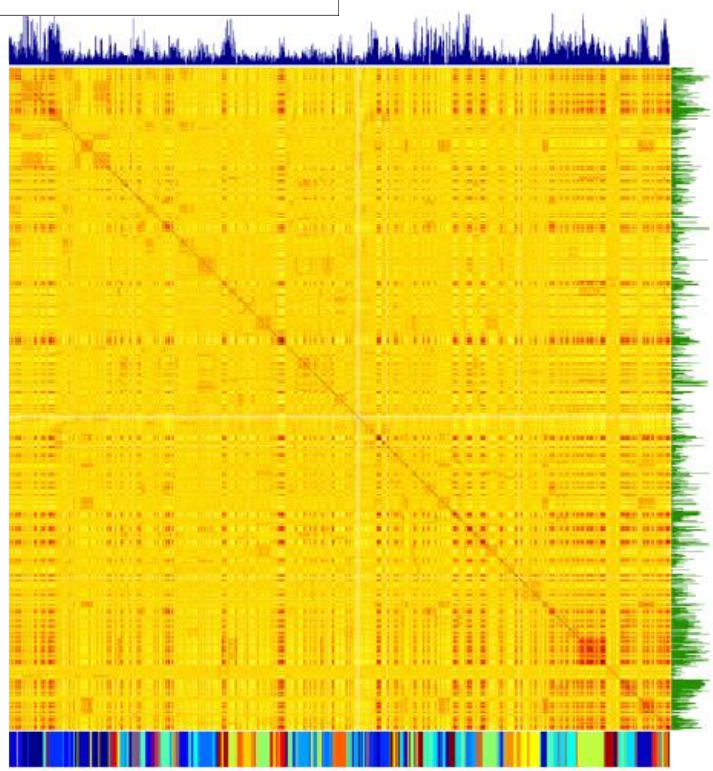
Trends Genet. 2007 Apr;23(4):158-61.
Kriegs JO, Churakov G, Jurka J, Brosius J, Schmitz J.

Alu density



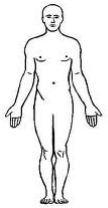
Human chromosomes

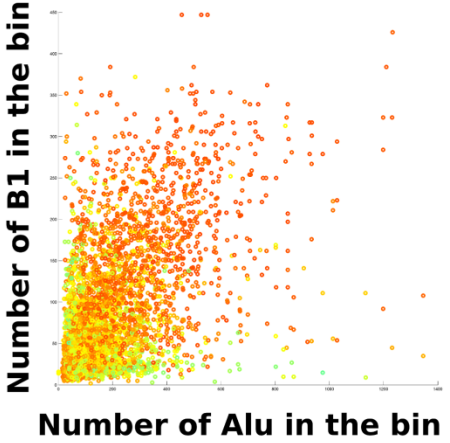
B1 density

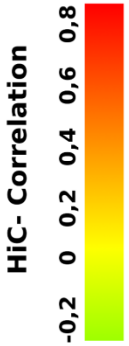


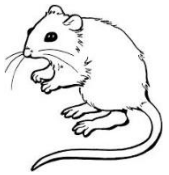
Mouse chromosomes

Differences in the 3D organization of the two genomes correlated with different SINE distribution



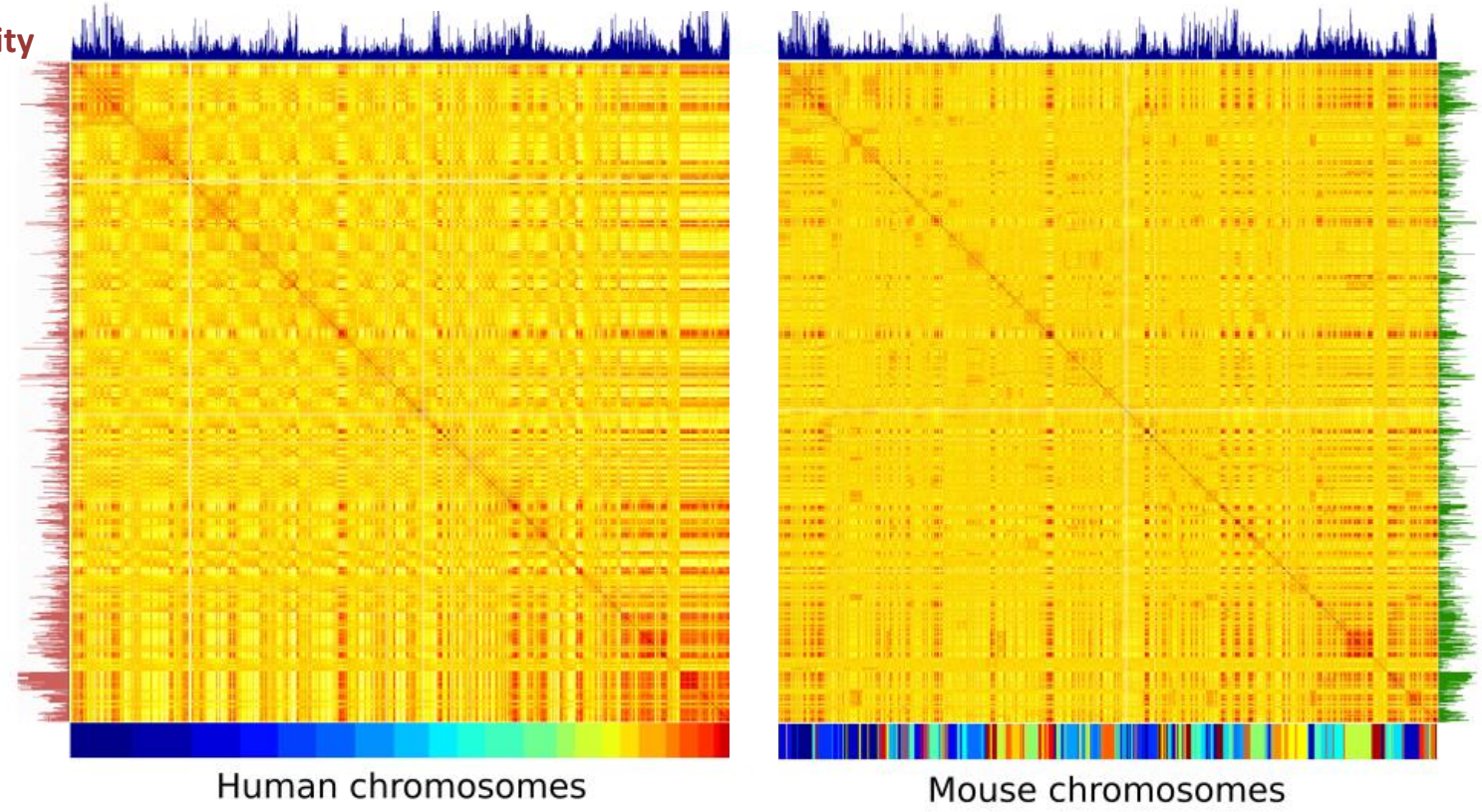




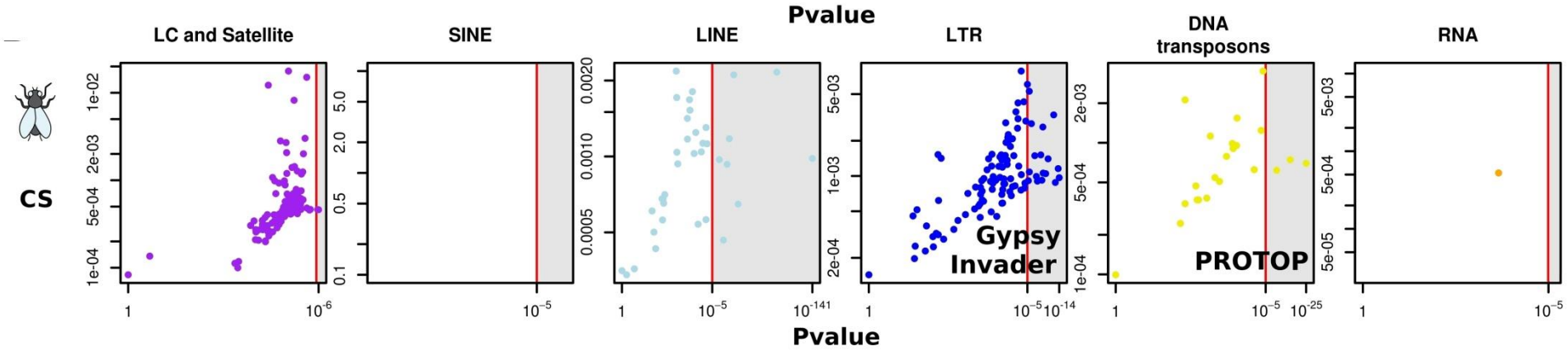


Alu density

B1 density

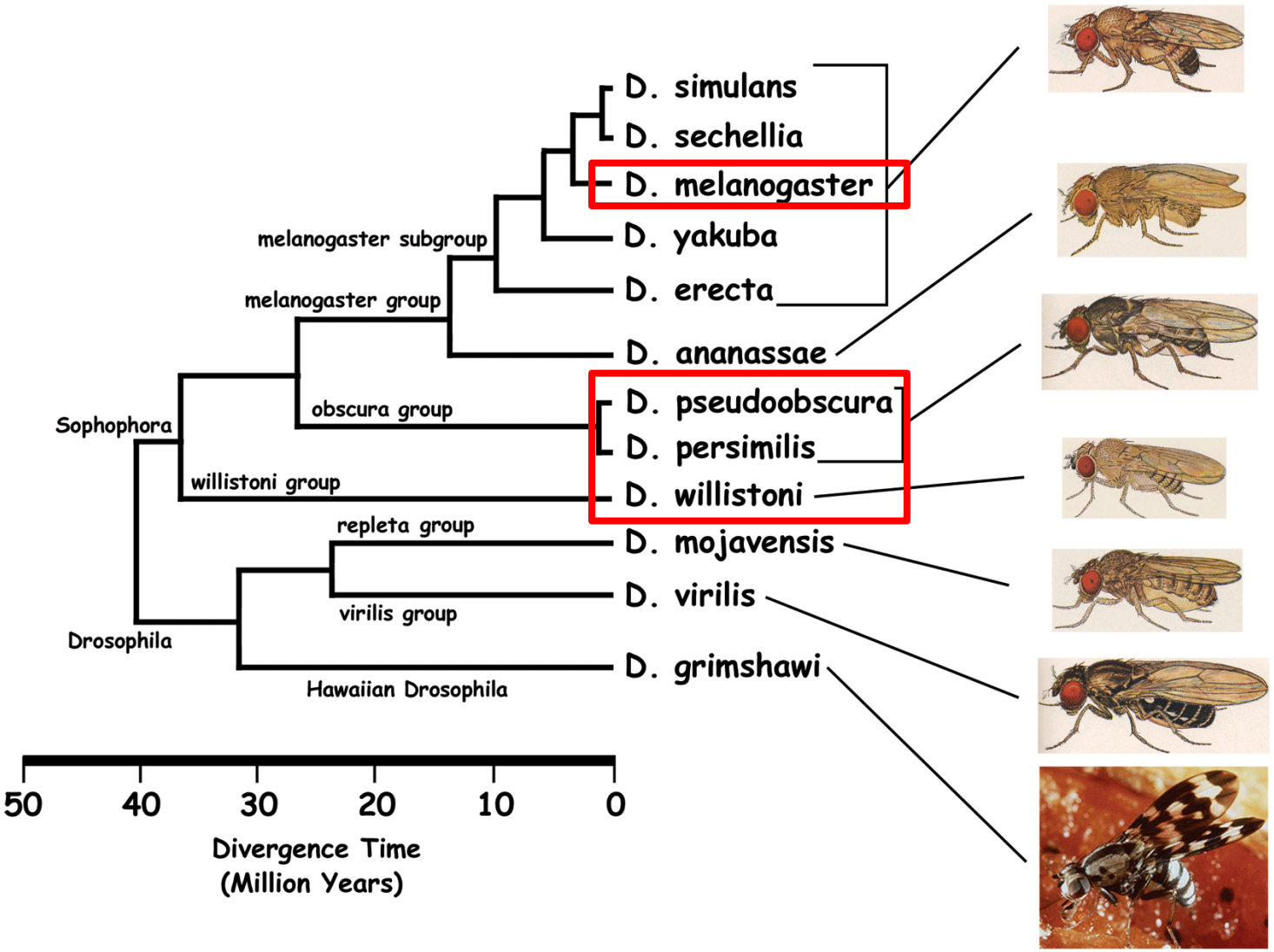


Co-localization score of all the different repetitive elements in the Drosophila genome.

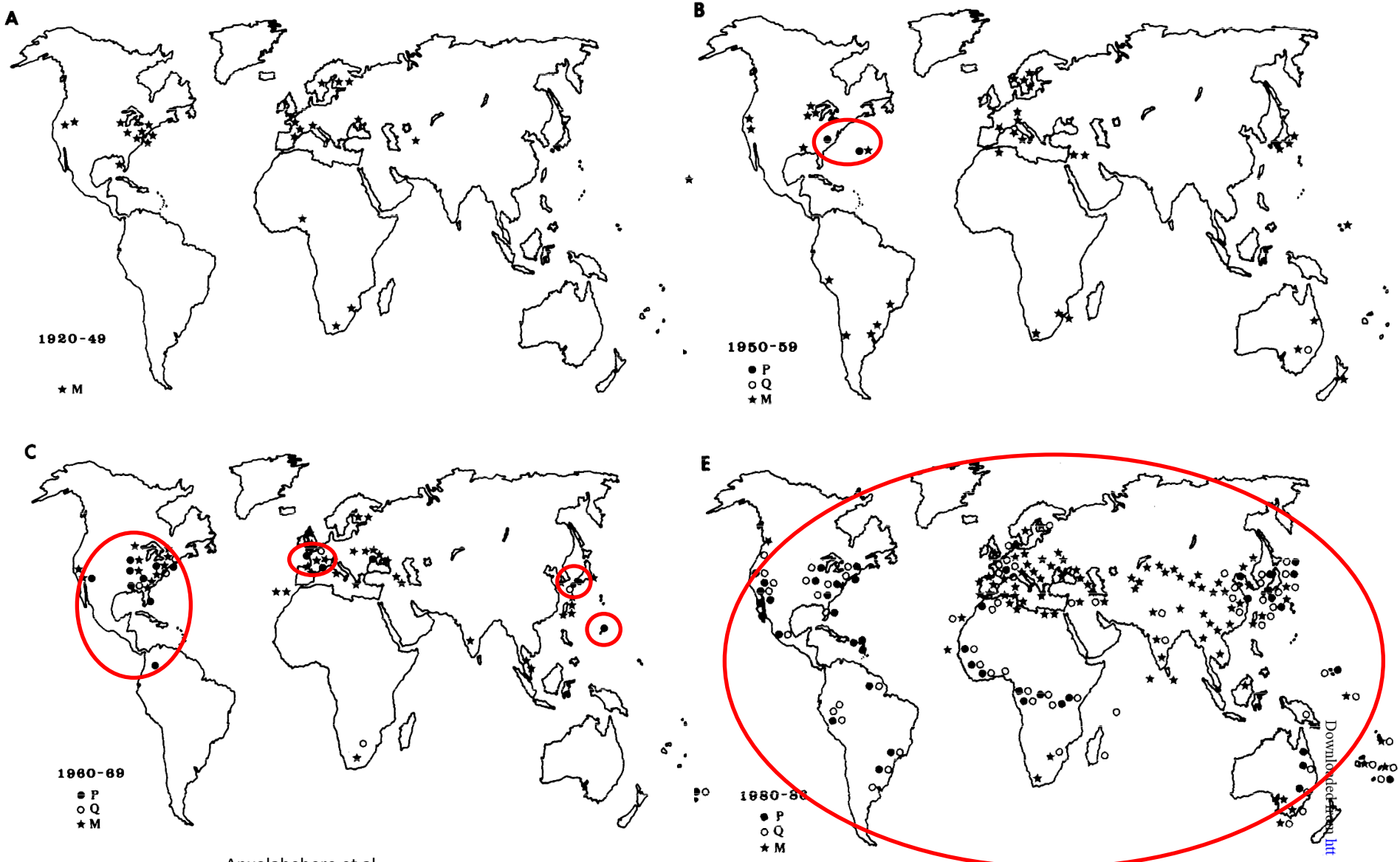


Data from Dixon et al, Cell, 2012

Element P

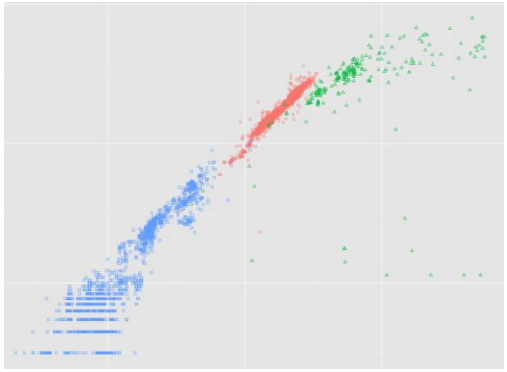

D. simulans
D. sechellia
D. melanogaster
D. yakuba
D. erecta
D. ananassae
D. pseudoobscura
D. persimilis
D. willistoni
D. mojavensis
D. virilis
D. grimshawi
melanogaster subgroup
melanogaster group
obscura group
willistoni group
repleta group
virilis group
Sophophora
Drosophila
Hawaiian Drosophila
50
40
30
20
10
0
Divergence Time
(Million Years)

Element P



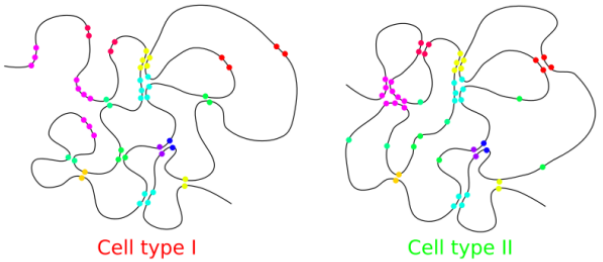
Anxolabehere et al.

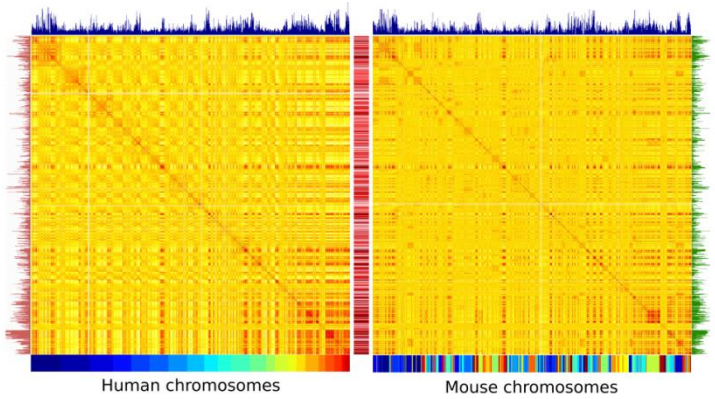
Summary



The genome of metazoans is full of DNA repeats.

<u>Role in cell differentiation:</u>
DNA repeats organize the genome folding in a cell type and transcription factor specific manner





<u>Role in evolution:</u>
Retro-transposition waves are able to reshape the genome folding in different organisms.

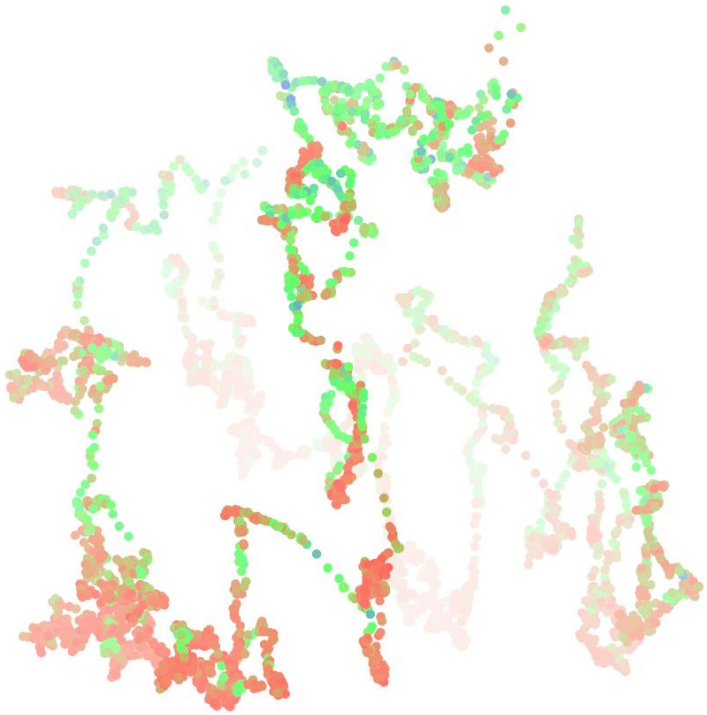
Acknowledgments

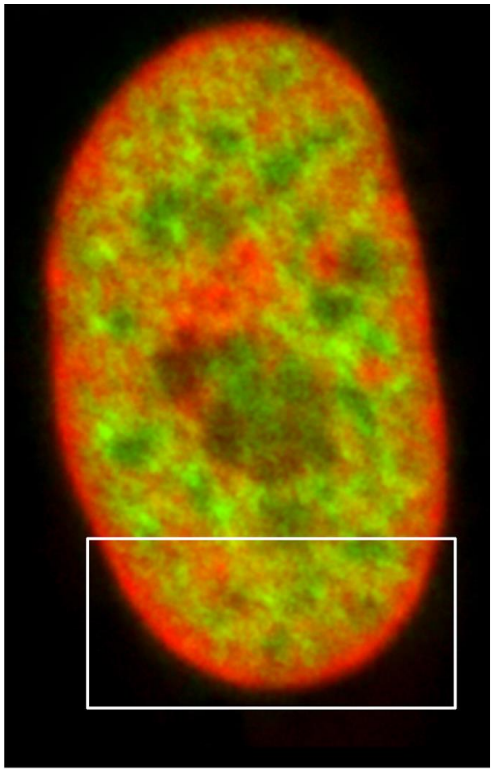
UPMC
J.M. Victor
M. Barbi
A. Lesne
R. Cortini
B. Caré
J. Riposo

ENS
T. Mora

Pasteur
R. Koszul
A. Cournac
M. Marbouty

MNHN
J.B. Boulé
C. Escudé
J. Ollion
C. Lavelle





Alu

DNA

T. Cremer

Ongoing work with A. Lesne, A Counac and J. Riposo