From the epigenome to the functional and structural nuclear organisation

Benjamin Audit

Equipe SIgnaux, SYstèmes et PHysiquE (SISYPHE)
Laboratoire de Physique de l'ENS-Lyon

ENS de Lyon

Françoise Argoul

Alain Arneodo

Antoine Baker

Pierre Borgnat

Rasha Boulos

Guénola Drillon

Pablo Jensen

Hanna Julienne

Nicolas Tremblay

Cédric Vaillant

Lamia Zaghloul

CGM, Gif-sur-Yvette

Yves d'Aubenton-Carafa

Chun-Long Chen

Claude Thermes

ENS, Paris

Guillaume Guilbaud

Olivier Hyrien

Malik Kahli

Nataliya Petryk

Aurélien Rappailles

CEA, Saclay

Arach Goldar

Heterochromatin and euchromatin
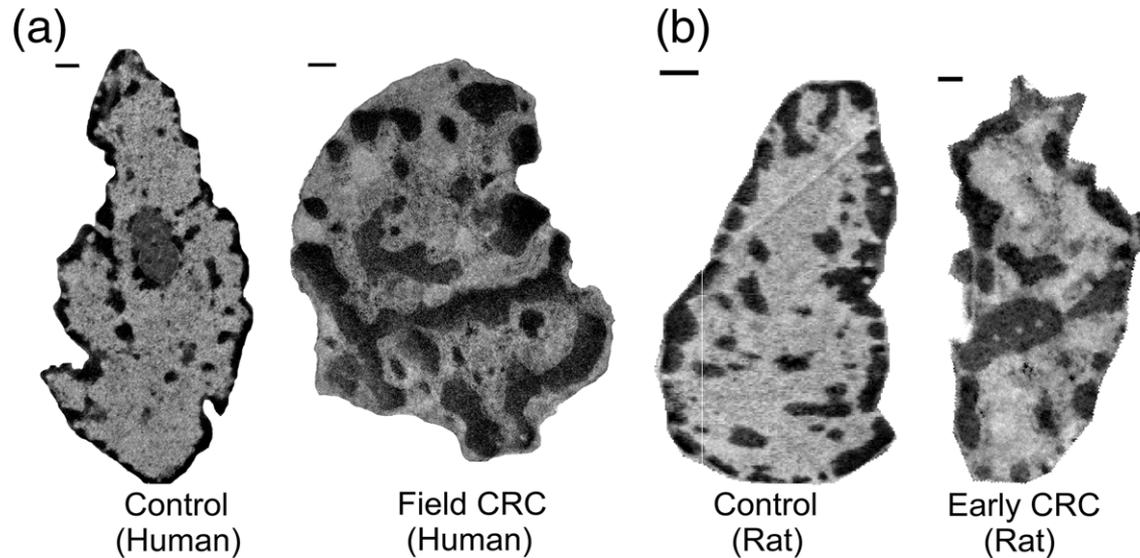
Figure 4 Example TEM micrographs. (a) Histologically normal rectal cell nuclei from control patients and those harboring a pre-cancerous adenoma elsewhere in the colon, representing field CRC. Scale bars correspond to 500 nm. **(b)** Histologically normal colonic cell nuclei from control rats and those treated with azoxymethane for 10 weeks (premalignant time point), representing early CRC. Scale bars correspond to 250 nm.

Cherkezyan, BMC Cancer (2014)

Heterochromatin: dense, transcriptionally silent

Euchromatin: decondensed, transcriptionally active

Organisation of eukaryotic nucleus

Chromosome territories
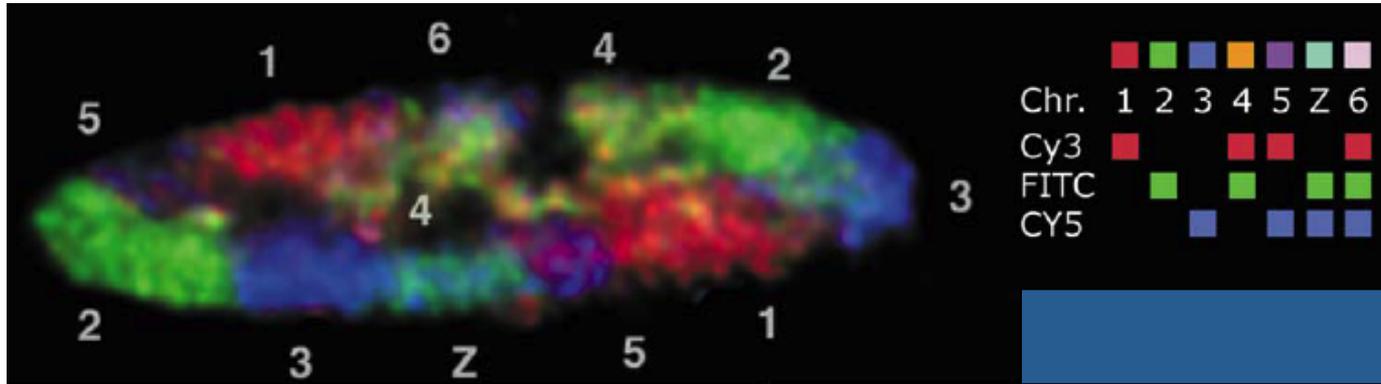
Chicken fibroblast nucleus where the 7 macro-chromosome are revealed by FISH

Regionalisation of nuclear functions
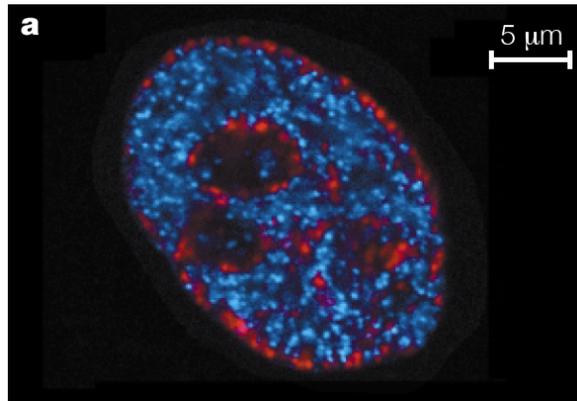
Human neuroblastoma

HeLa Cell

Early replication *(Cy3)*

Mid- to late- replication *(Cy5)*
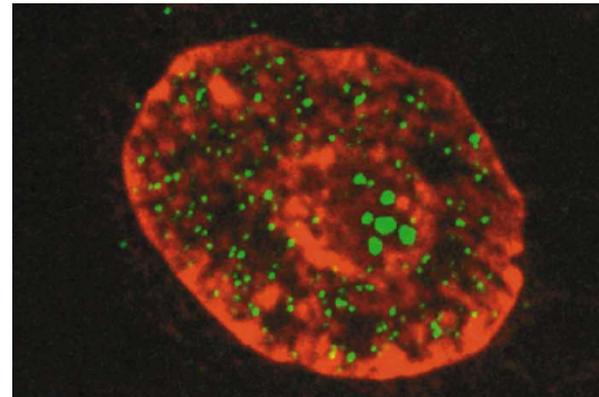
RNA (5-bromo-UTP / FITC)

DNA (TOTO-3)

Chromatin conformation capture for the human genome
Hi-C data from Lieberman-Aiden, Science 326 (2009)
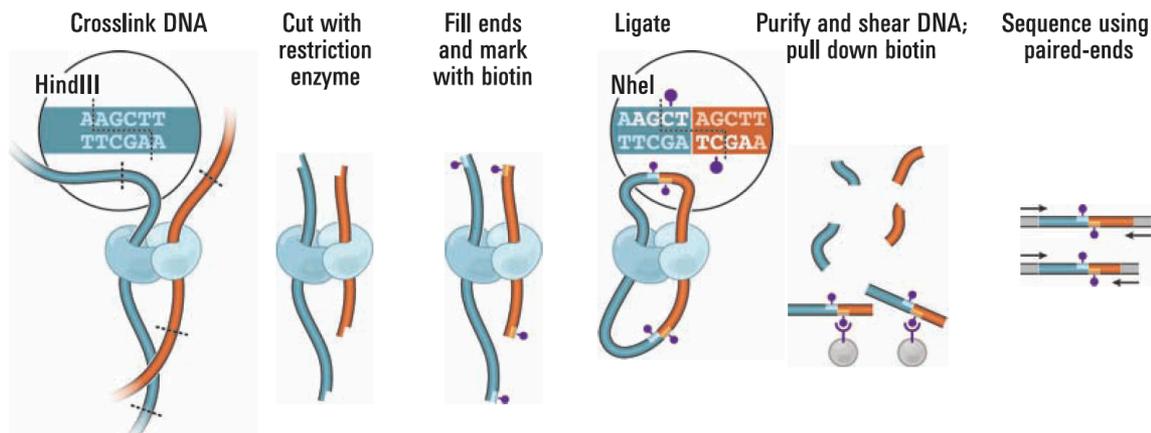
Crosslink DNA
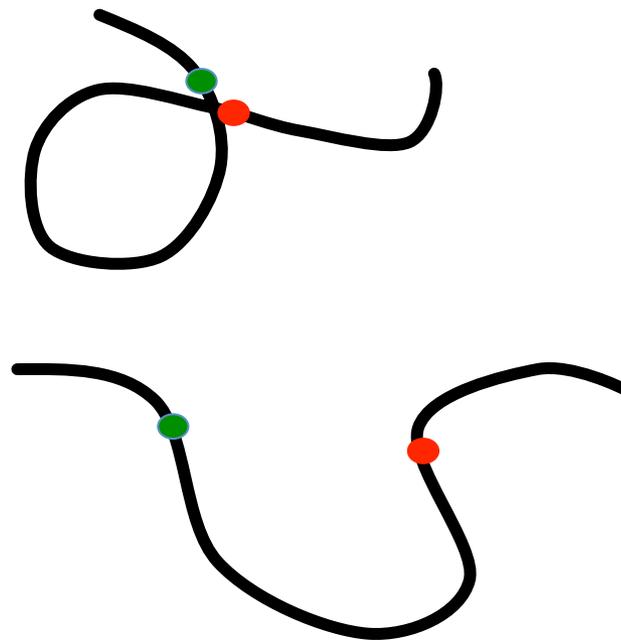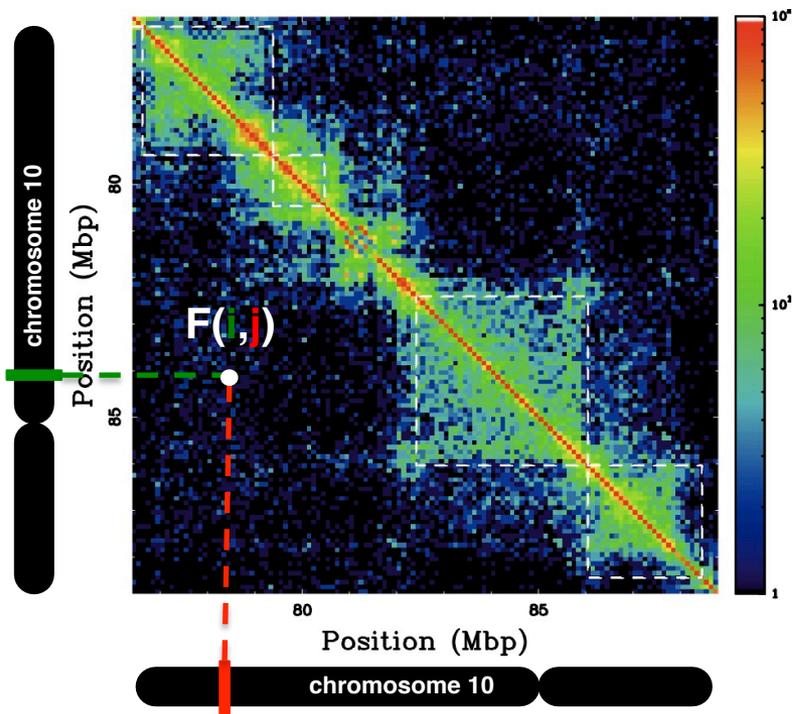
HindIII

AAGCTT
TTCGAA

Cut with restriction enzyme

Fill ends and mark with biotin

Ligate

NheI

AAGCT AGCTT
TTCGA TCGAA

Purify and shear DNA; pull down biotin

Sequence using paired-ends

Co-localisation frequency matrix

chromosome 10

Position (Mbp)

80

85

F(i, j)

Position (Mbp)

80

85

chromosome 10

A dichotomous view of chromatin organization and topology at Mb resolution

A Observed

Chr 14

B Observed/Expected

Chr 14

G Chr14

Coverage

Genes

H3K27me3 — Repressed

H3K36me3 — Active

DNAseI — DNA Accessibility

Eigenvector

Chr 14

A dichotomous view of chromatin organization and topology at Mb resolution

Ryba, Genome Research (2010)
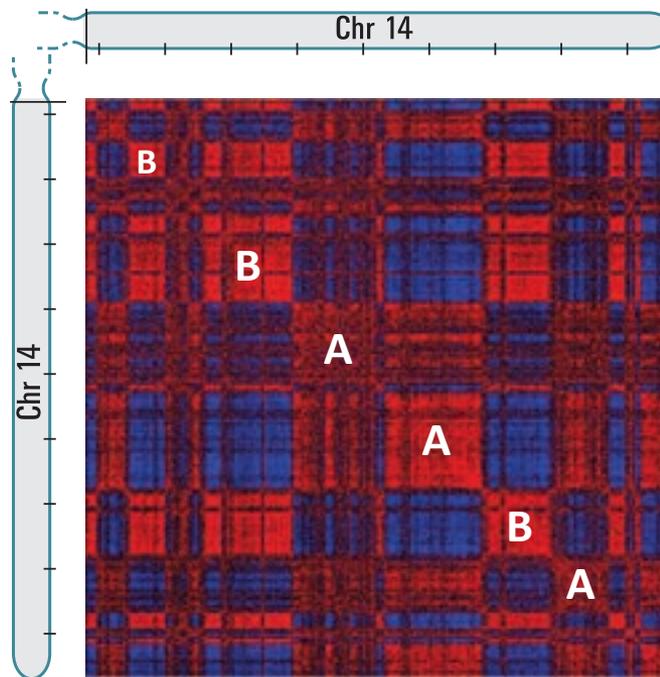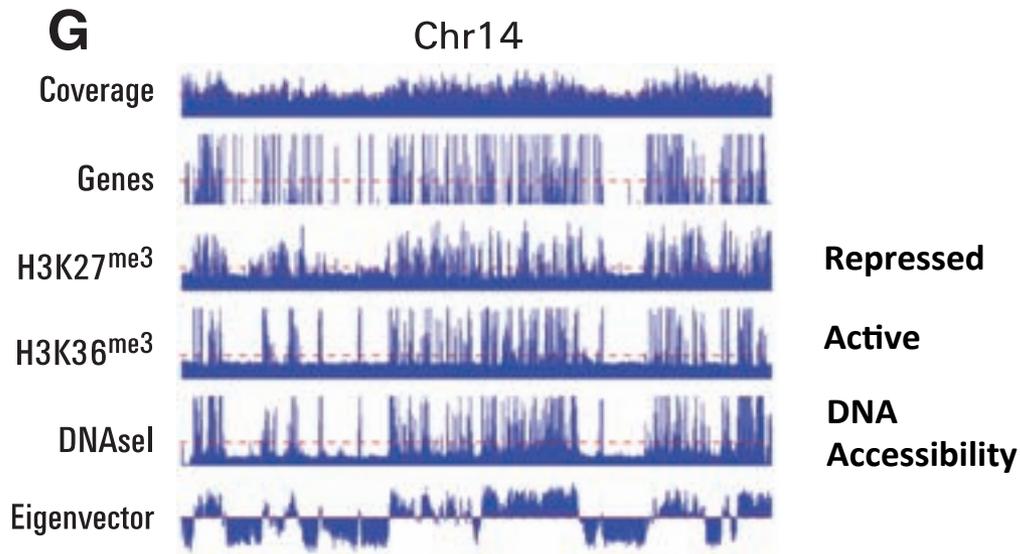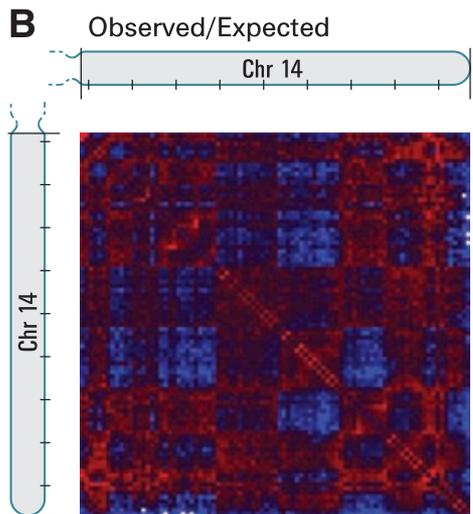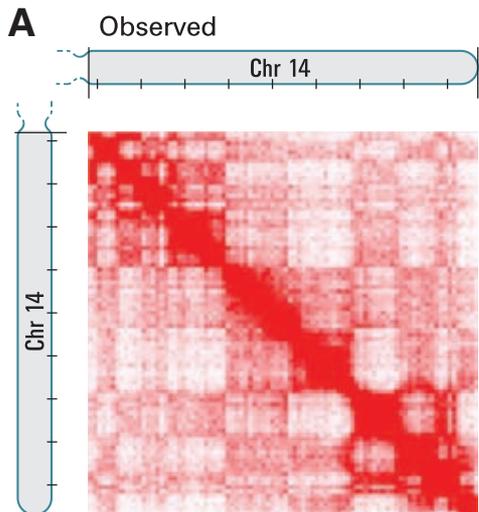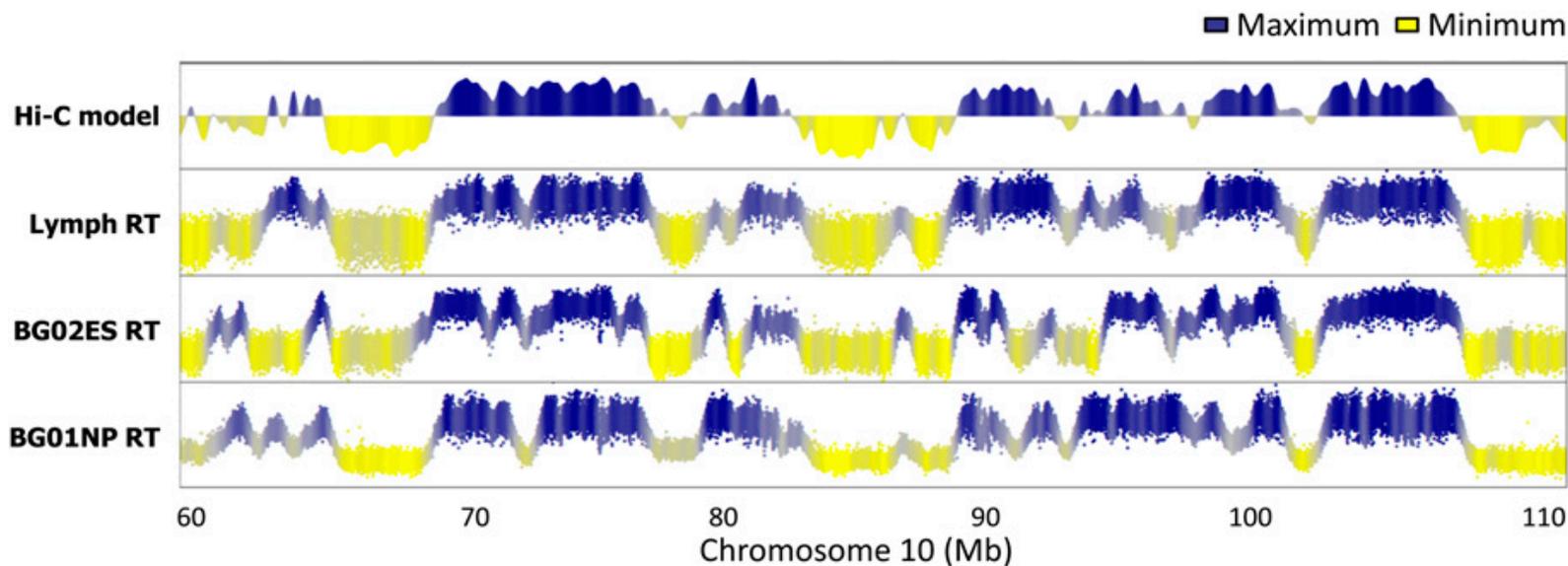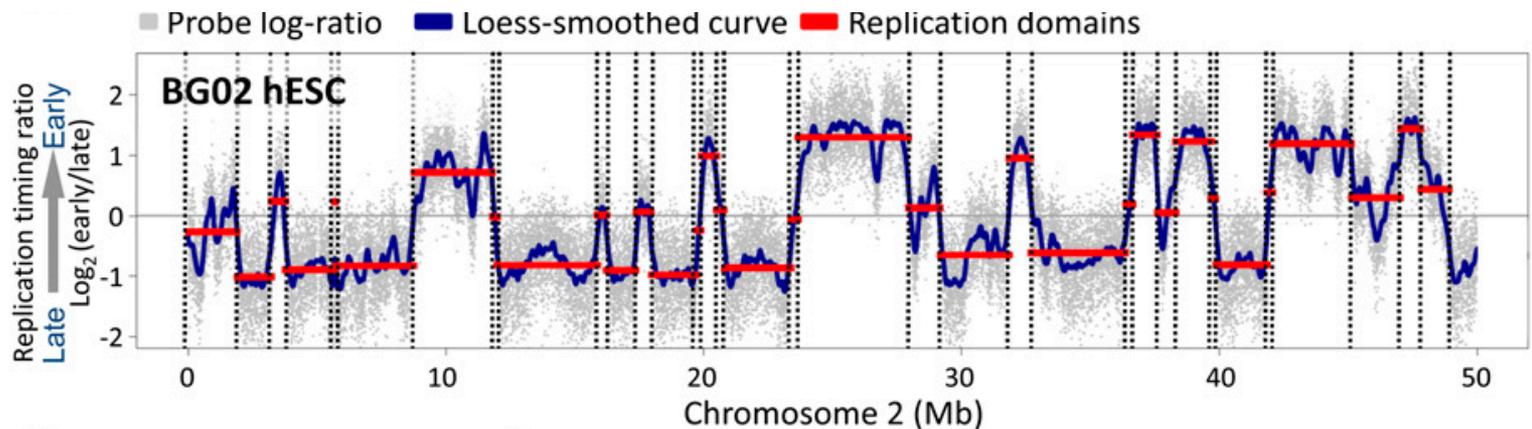
Human chromatin is characterised by 4 epigenetic states at scale 100 kbp

Principal Component Analysis using 13 epigenetic marks

Correlation matrix
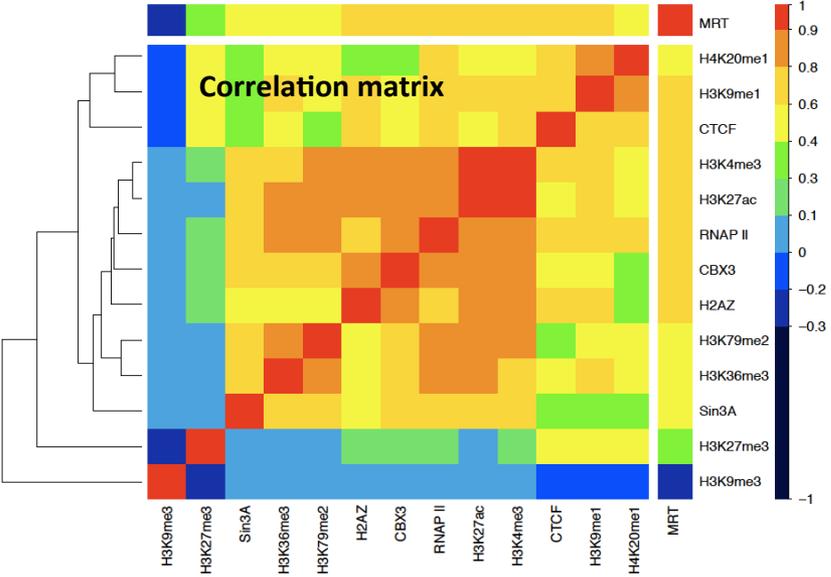
K-means clustering in 3D space (PC1, PC2, PC3)
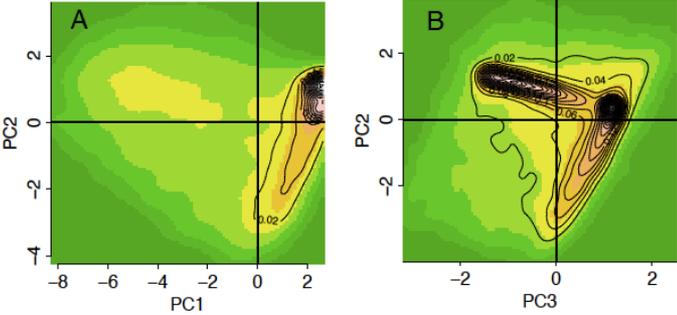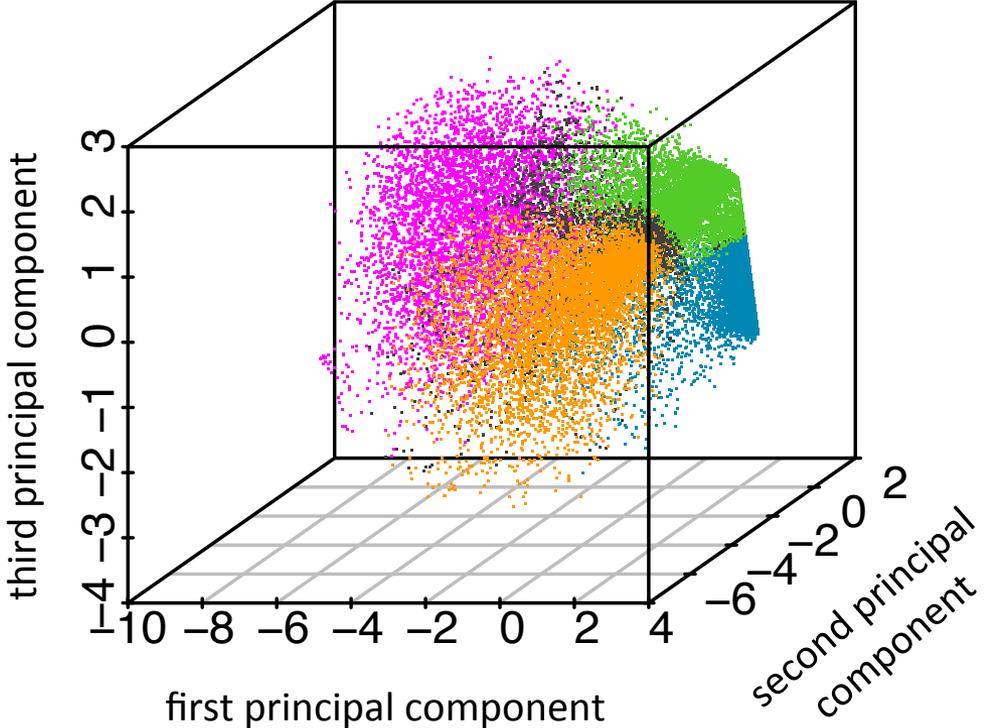
80% of the variance is explained by the first 3 principal components
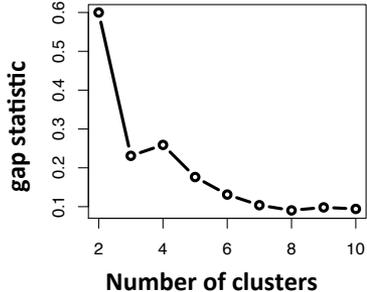
Characterisation of chromatin states in K562

1 - Transcriptionally active chromatin

2 - Domain of gene repression by Polycomb

3 - Silent unmarked chromatin

4 - HP1 heterochromatin
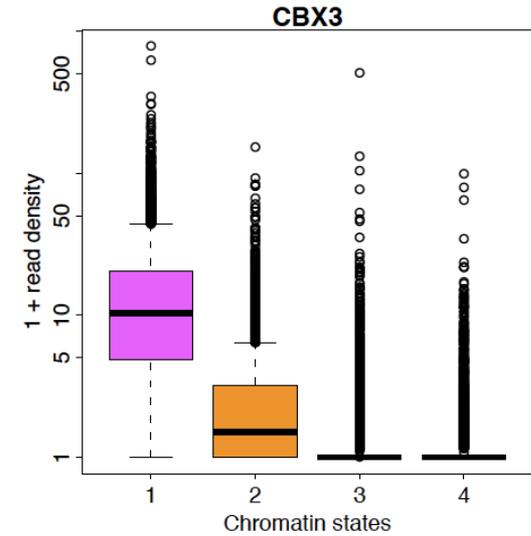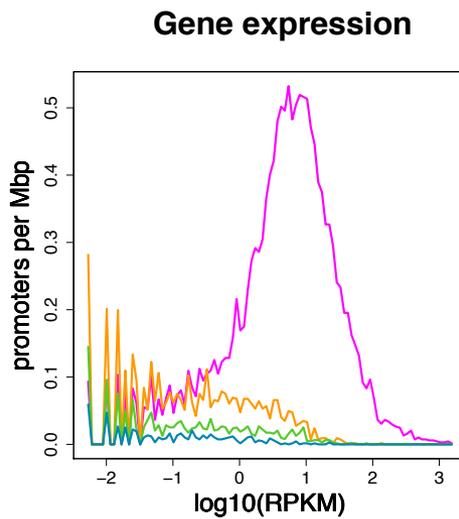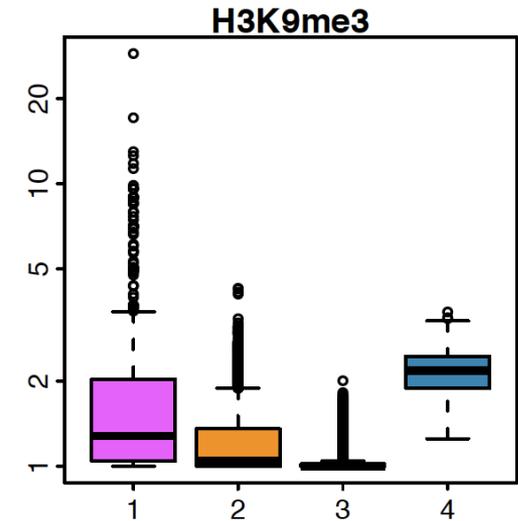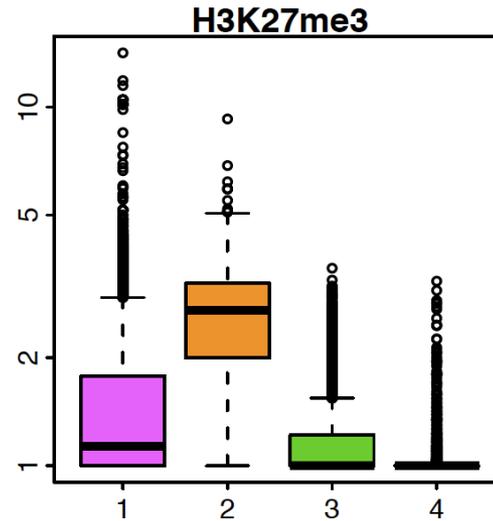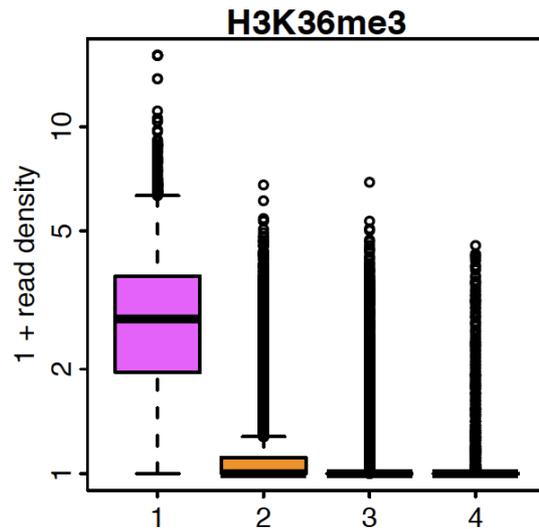
4 chromatin states in all differentiated human cell types

Hmec - Mammary Epithelial Cells

K562 - myelogenous leukemia line

Monocd14ro1746 - Monocyte CD14+

GM12878- lymphoblastoid cell line
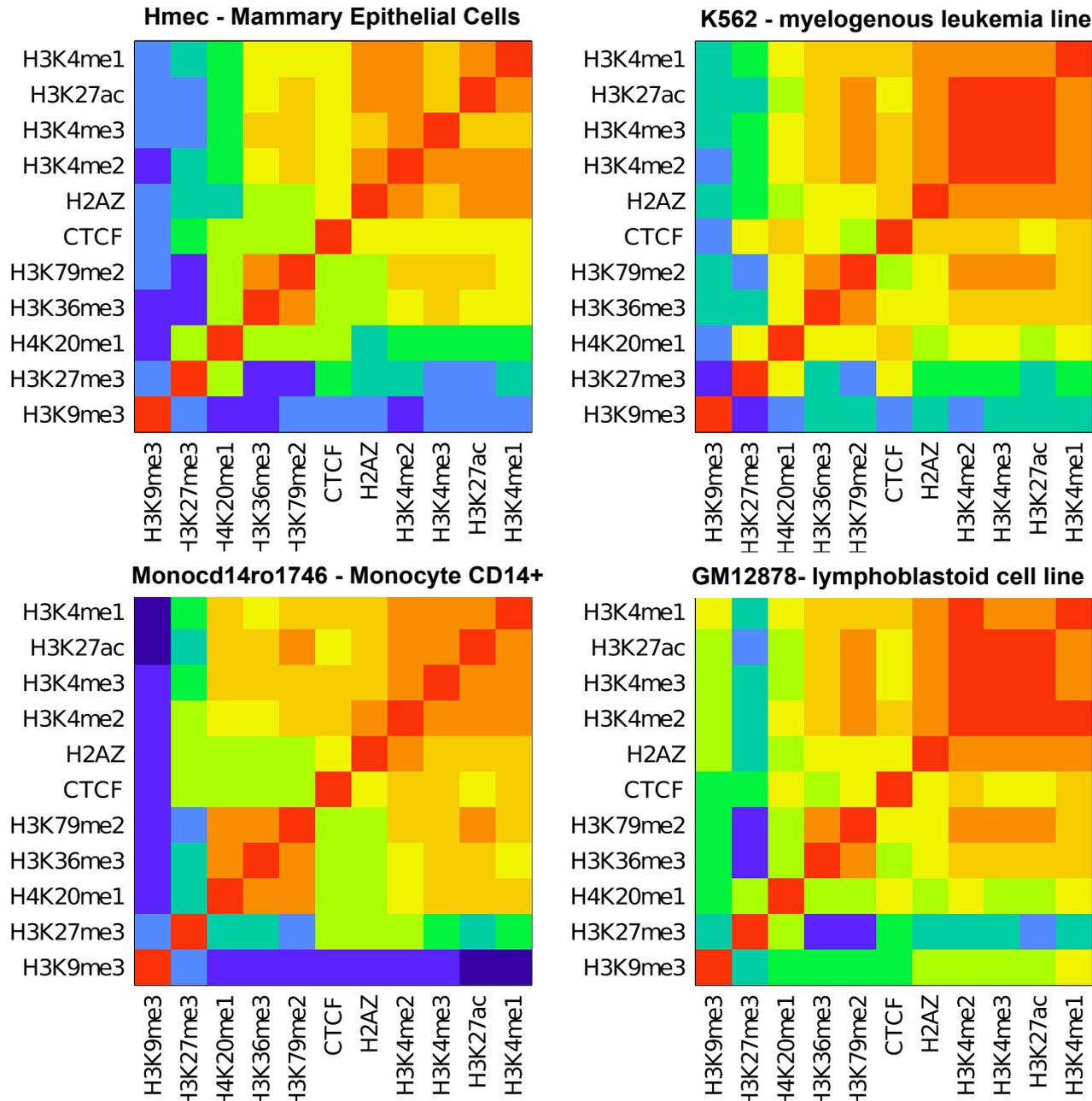
Chromatin states are replicated at different times
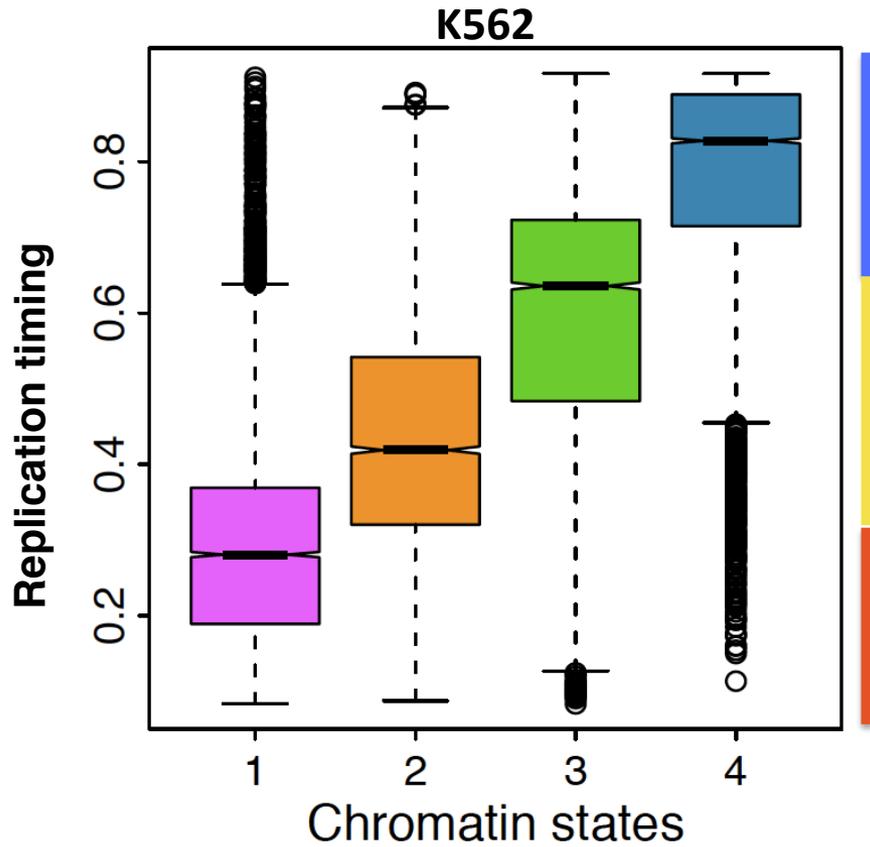
Early / **Mid** / **Late** replication
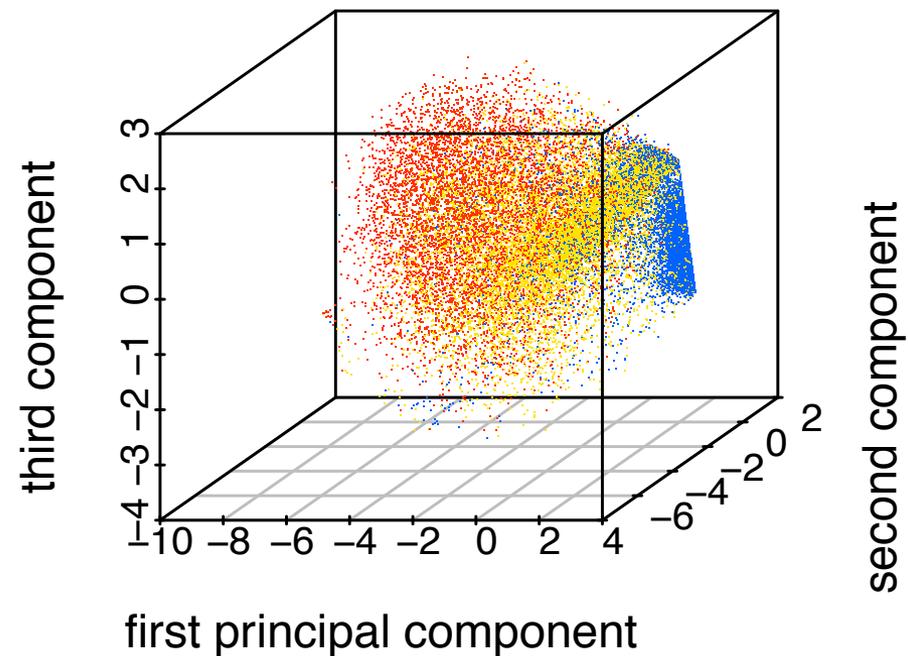
Embryonic stem cell specific chromatin organization
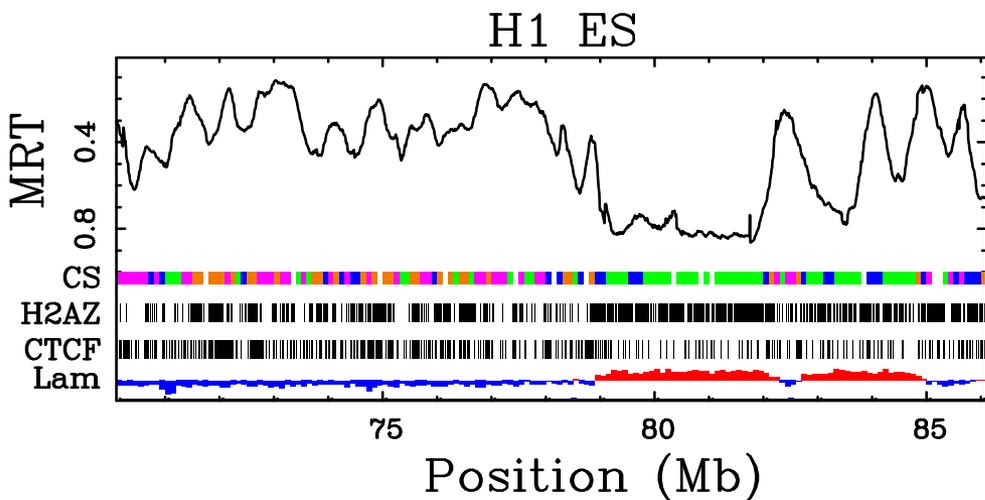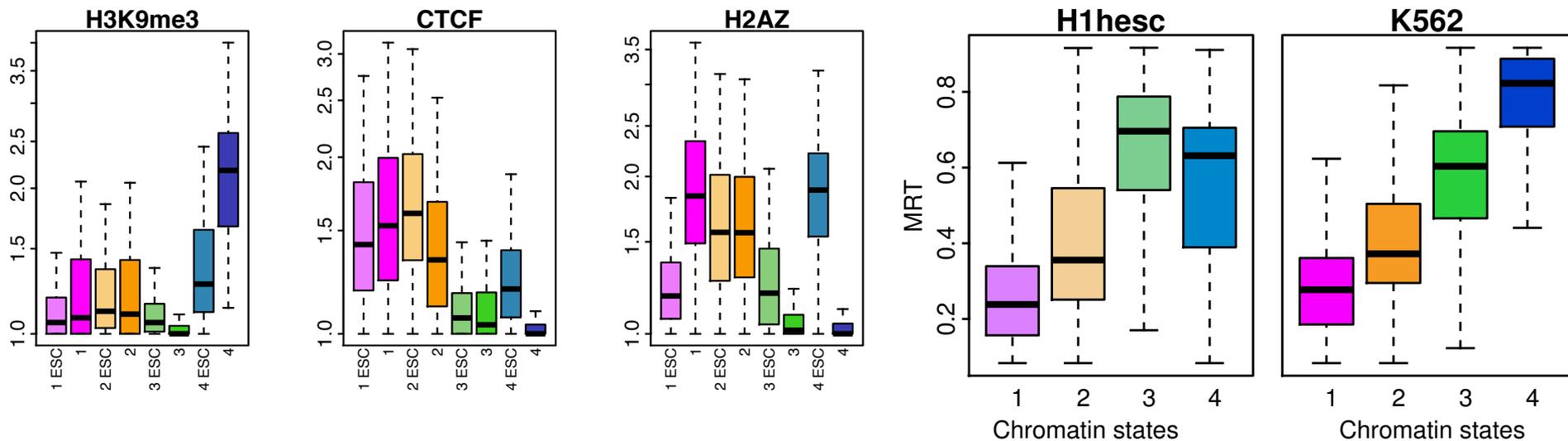
Highly dynamic heterochromatic state in embryonic stem cell

Genome-wide segmentation of the human genome into 4 chromatin states

Chromatin states 1+2 and 3+4 colocalize

1+2

3+4

E

26% of the genome
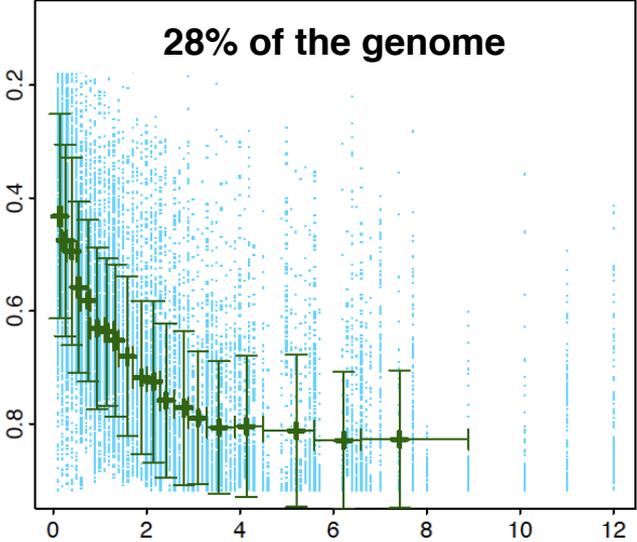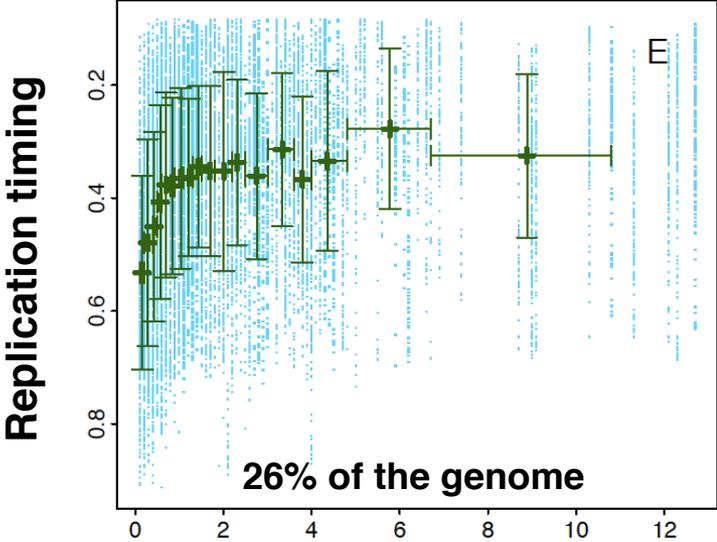
28% of the genome

Distribution of chromatin states in K562 replication U-domains
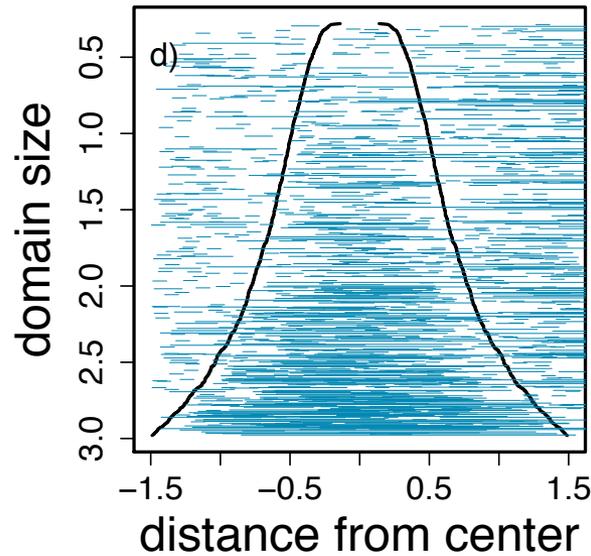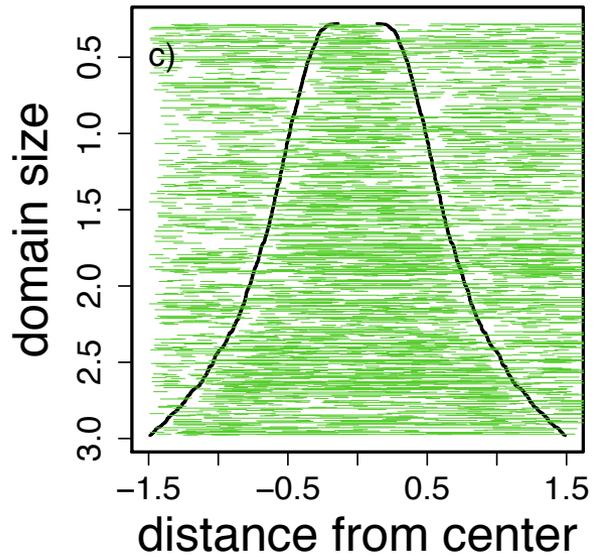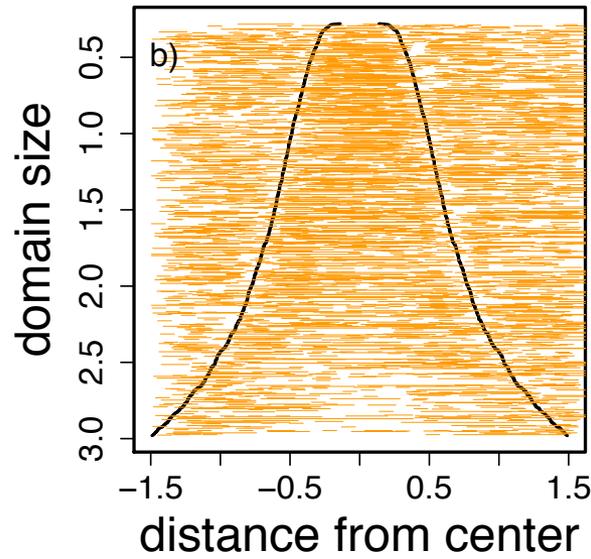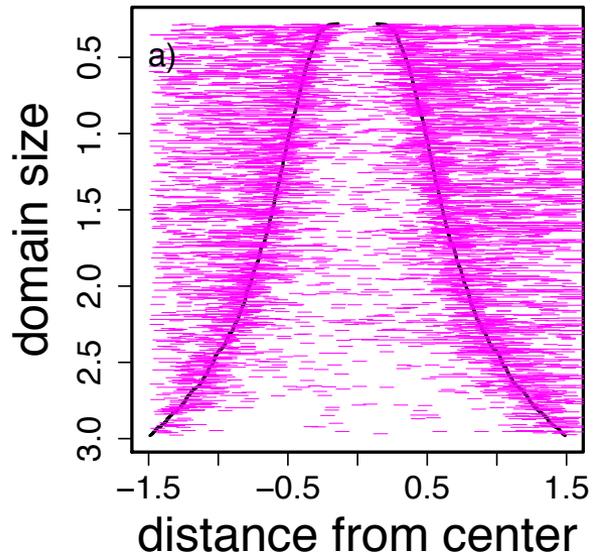
46% of the genome is covered by U-domains
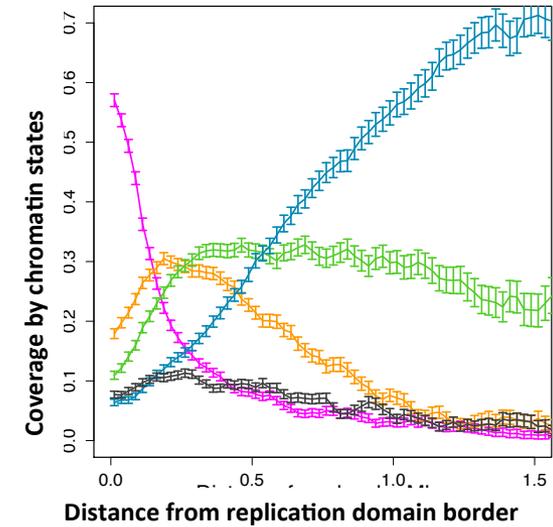
Crosslink DNA | Cut with restriction enzyme | Fill ends and mark with biotin | Ligate | Purify and shear DNA; pull down biotin | Sequence using paired-ends

Co-localisation frequency matrix

F(i, j)

chromosome 10

Position (Mbp)

Position (Mbp)

chromosome 10

A

Contact probability (log)

Distance (Mb)

— Intrachromosomal, Chr 1
···· Interchromosomal, Chr 1–Chr 10
···· Interchromosomal, Chr 1
···· Interchromosomal, Chr 1–Chr 21
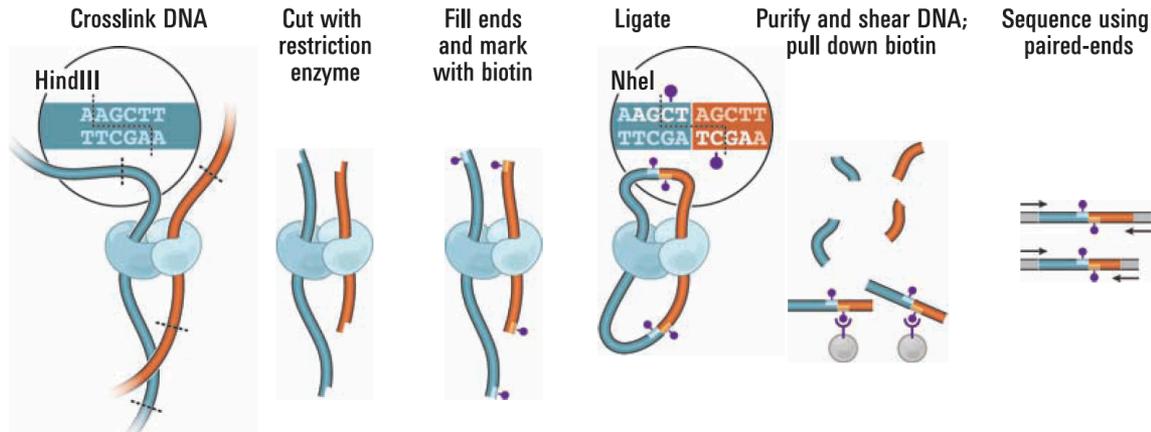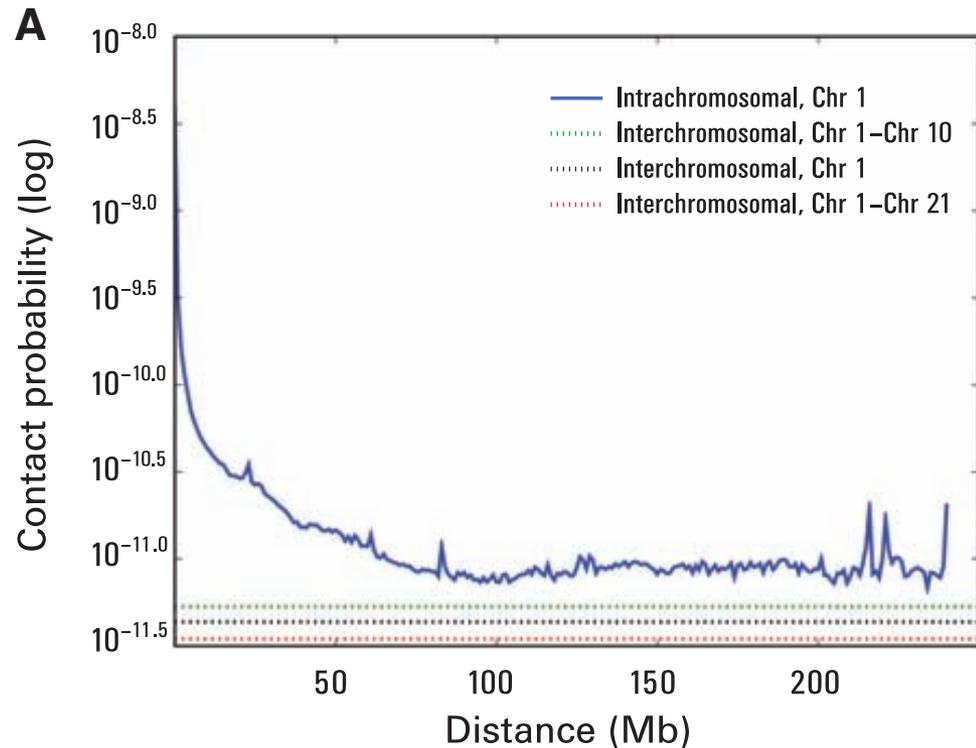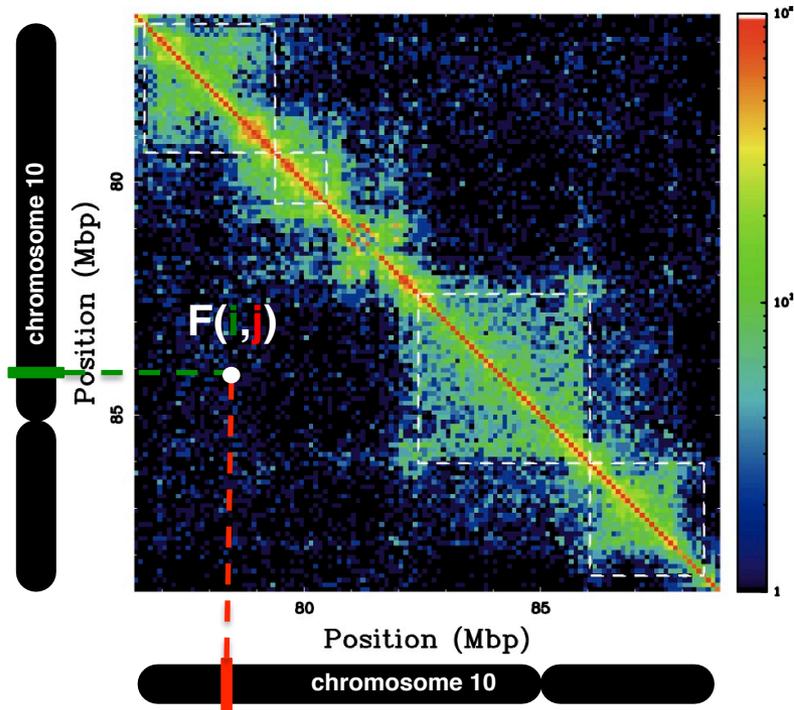
Packing of chromatin is consistent with the behavior of a fractal globule

A Experiment

B Models

$$F(s) = k \cdot s^{\alpha}$$

C UNFOLDED POLYMER

FOLDED POLYMER

Equilibrium globule   Cross-section view

Fractal globule   Cross-section view

D Nuclear scale

Chromosome territories

Chromosome scale

Open

Closed

Megabase scale
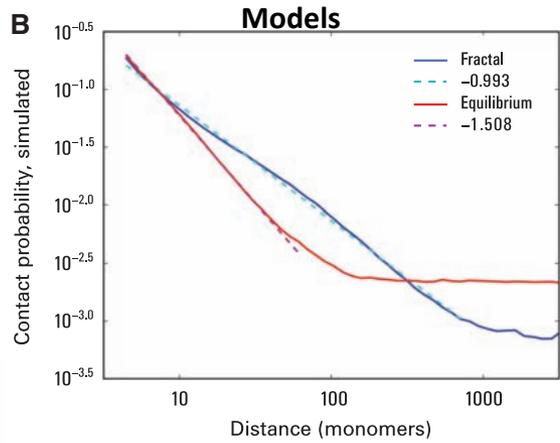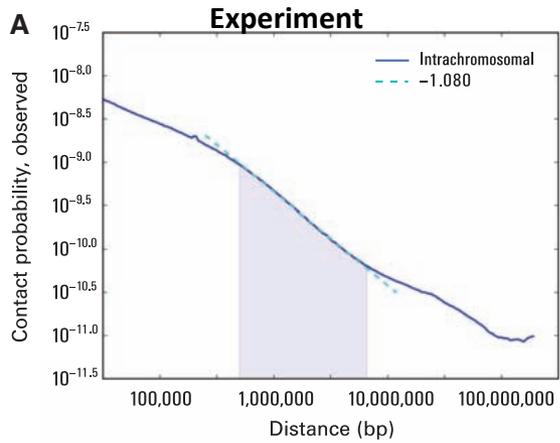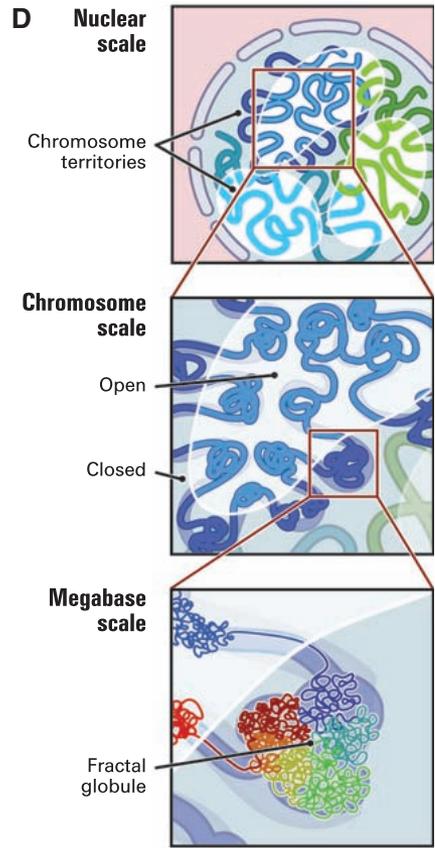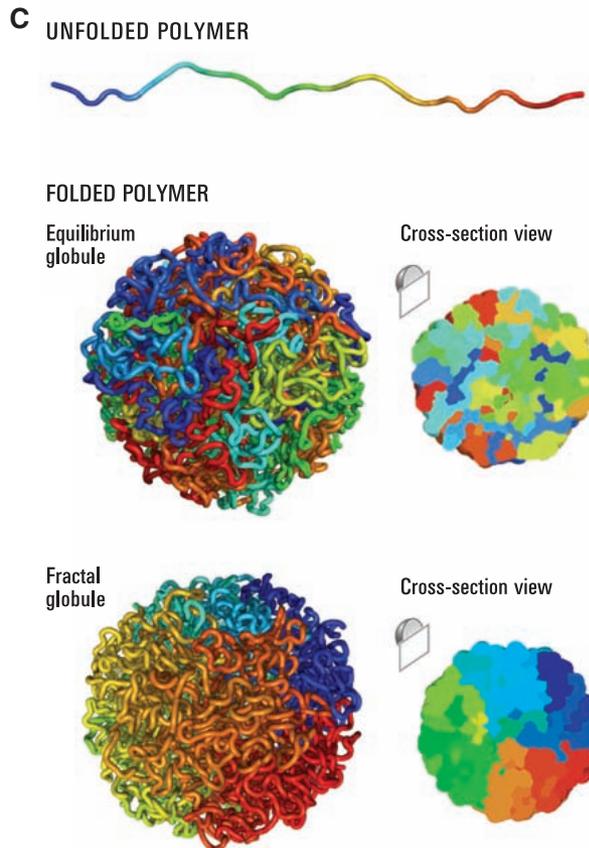
Fractal globule

Non equilibrium fractal globule:
$$\alpha = -1$$

Equilibrium globule:
$$\alpha = -d_f/d_w$$

d_f: geometrical fractal dimension
$d_f = 3$: space-filling state
d_w: dynamical fractal dimension
$d_w = 2$: normal diffusion

$$\alpha = -3/2$$

Structural organization of chromatin domains
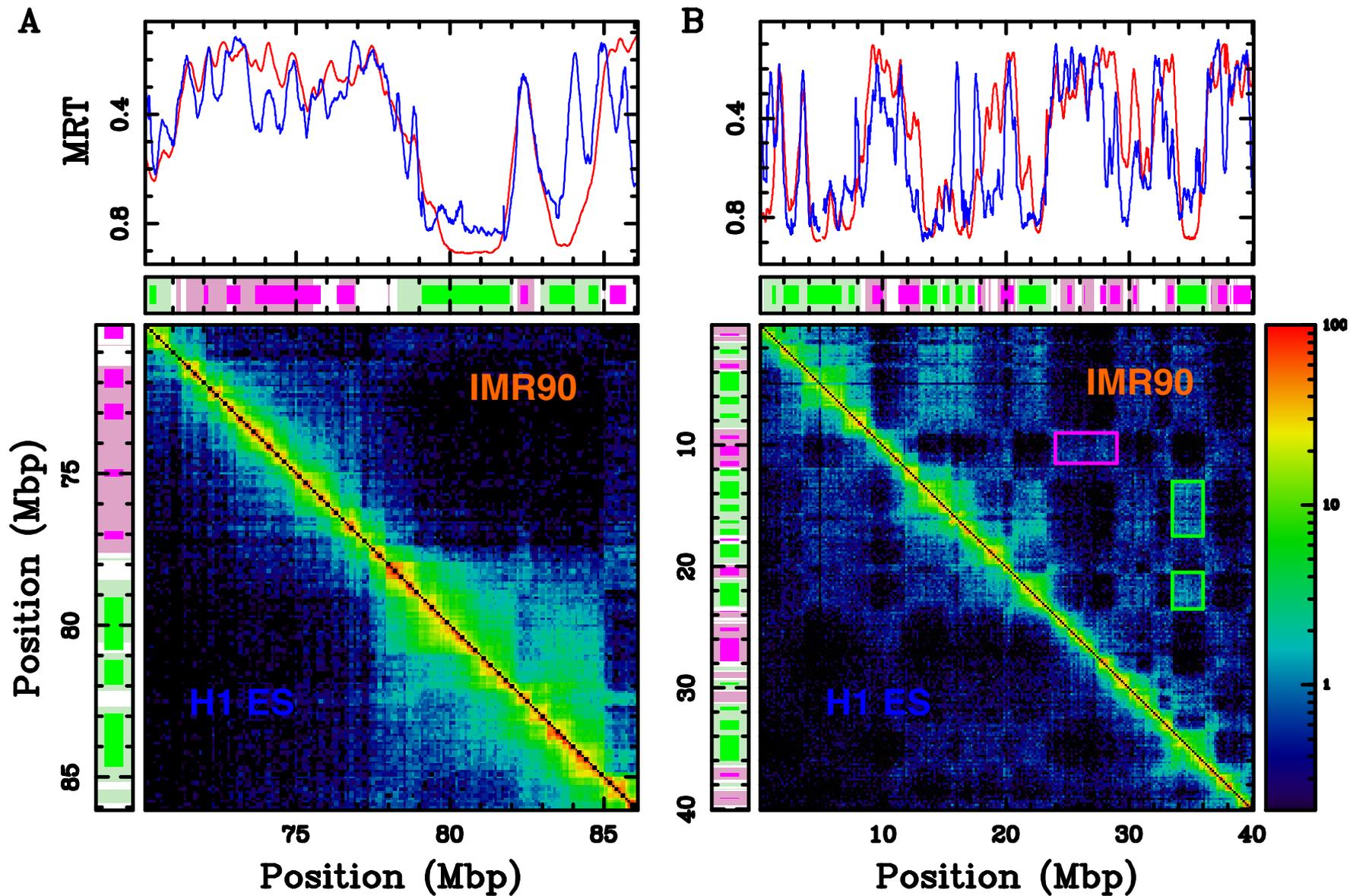
Structural organization of chromatin domains in K562

1+2 vs 1+2 interactions

1+2 vs 3+4 interactions

3+4 vs 3+4 interactions

✓ **Exponent α = -1 is specific to 3+4 domains**

✓ **1+2 domains compatible with α = -3/2**

✓ **Segregation between 1+2 and 3+4 domains**

Structural organization of chromatin domains in IMR90
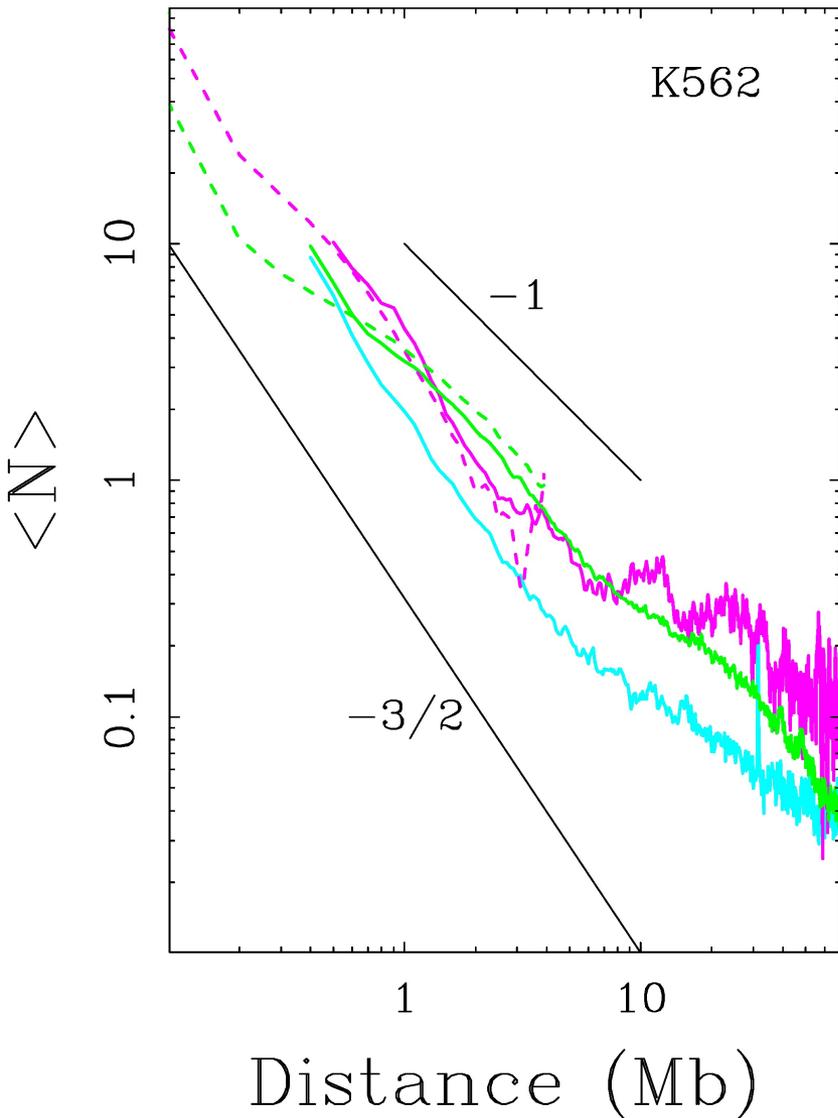
1+2 vs **1+2** interactions

1+2 vs **3+4** interactions

3+4 vs **3+4** interactions

✓ **Exponent α = -1 is specific to 3+4 domains**

✓ **1+2 domains compatible with α = -3/2**

✓ **Segregation between 1+2 and 3+4 domains**

Structural organization of chromatin domains in IMR90

Boulos, FEBS Letters (2015)
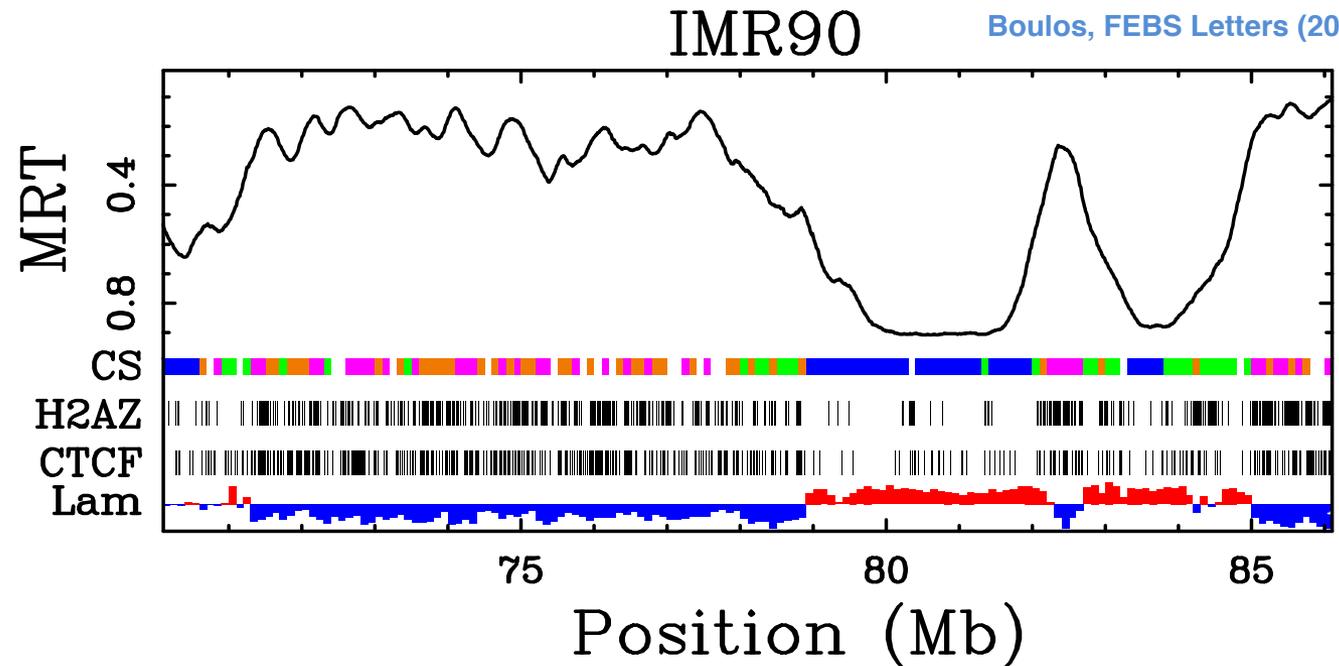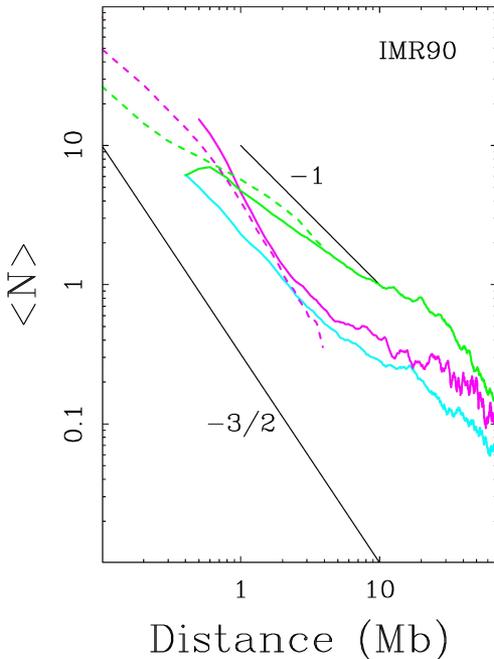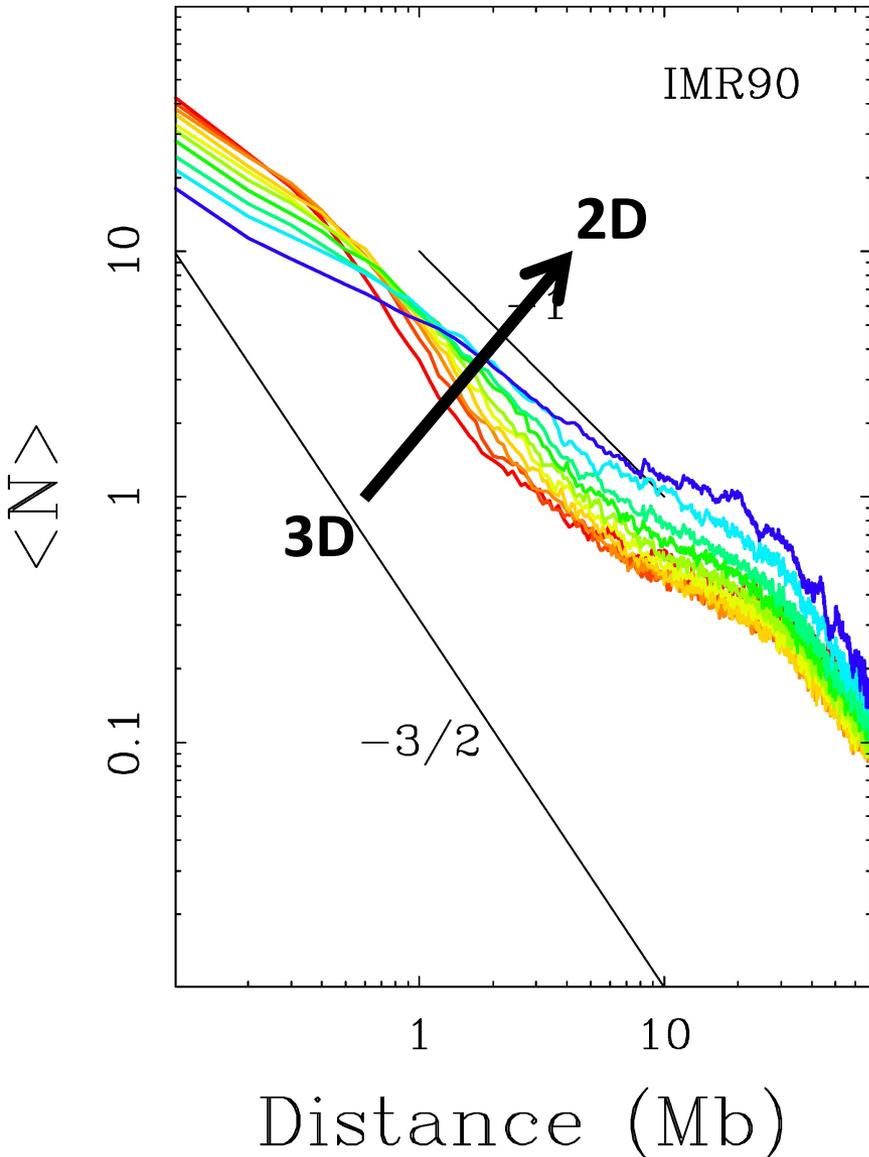
3+4 domains associated to lamina at nuclear envelop ➜ d_f=2: plane-filling state

2D equilibrium globule: $\alpha = -d_f/d_w$

    geometrical fractal dimension: d_f=2 plane-filling state

    dynamical fractal dimension: d_w=2: normal diffusion

    ➜ $\alpha = -1$

Measurement in mouse embryonic fibroblast:

d_f = 2.2, d_w = 2.6 (subdiffusion) ➜ $\alpha = -0.85$

Bancaud, EMBO Journal (2009); Nucleic Acids Research (2012)

Structural organization and the DNA replication program in IMR90

Segmentation of the genome in replication timing deciles

Early (D1) vs early (D1) interactions

Mid-early (D4) vs mid-early (D4) interactions

Mid-late (D7) vs mid-late (D7) interactions

Late (D10) vs late (D10) interactions

Timing as a measure of the radial positioning within the nucleus

Consistently with the change of spatial distribution of replication foci during S-phase from central to peripheral positioning

Structural organization of chromatin domains in ES cells
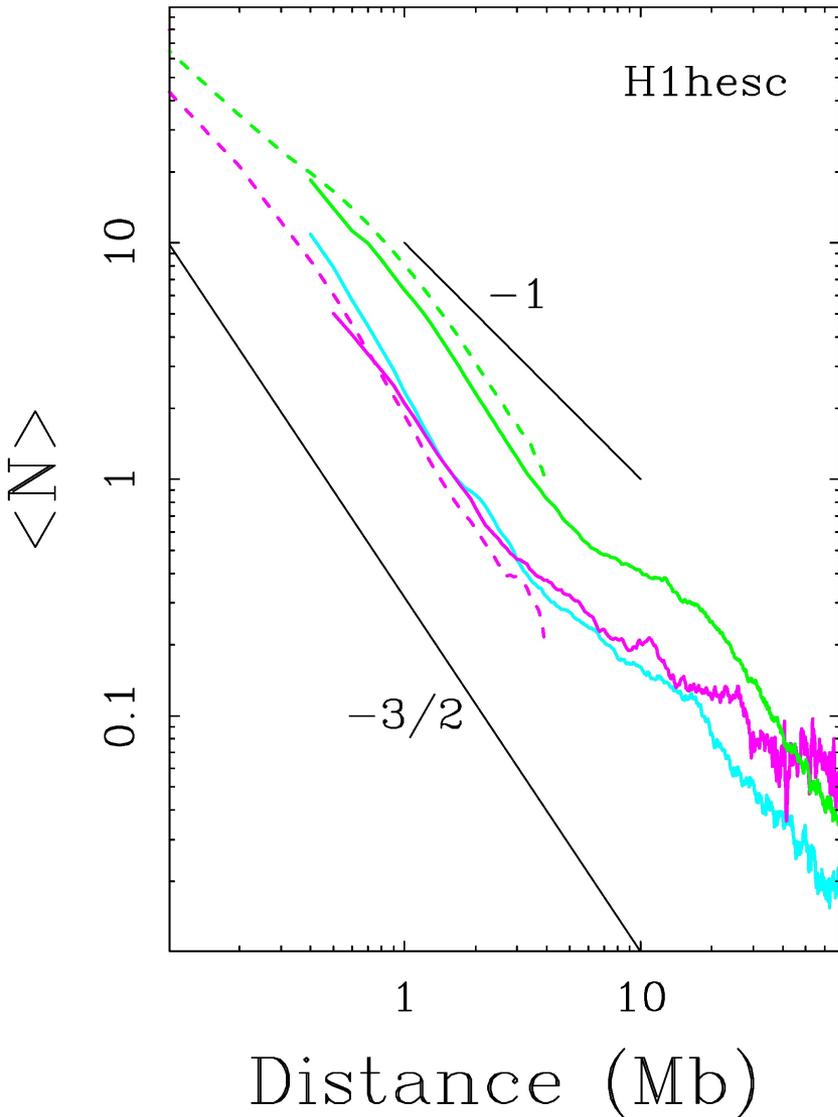
1+2 vs **1+2** interactions

1+2 vs **3+4** interactions

3+4 vs **3+4** interactions

✓ **All domains compatible with α = -3/2**

✓ **Weak segregation between 1+2 and 3+4 domains**

Structural organization and the DNA replication program in ES cells
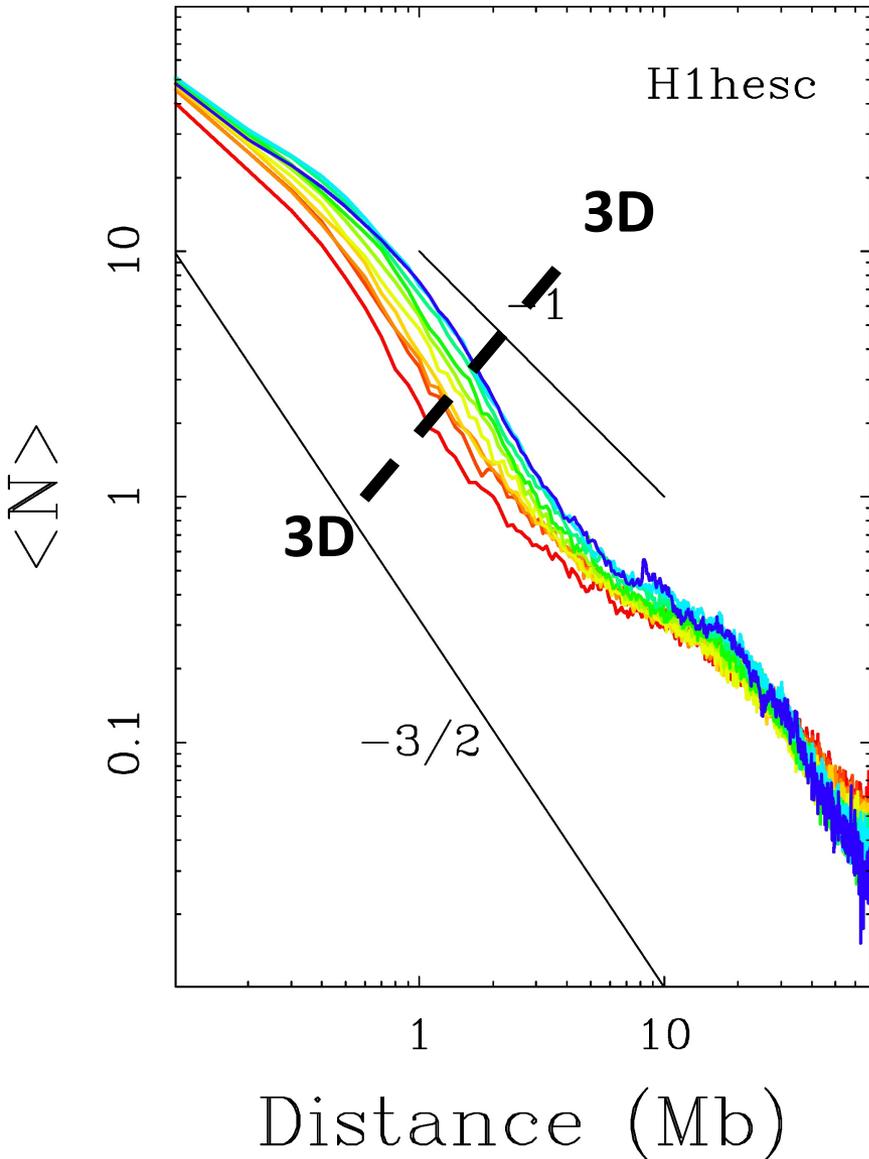
Segmentation of the genome in replication timing deciles

Early (D1) vs early (D1) interactions

Mid-early (D4) vs mid-early (D4) interactions

Mid-late (D7) vs mid-late (D7) interactions

Late (D10) vs late (D10) interactions

All timing domains in 3D space

Spatial organization of chromatin domains related to cell fate decision ?

Dynamics of replication foci during S-phase in differentiated cells

Quantitative Live Imaging of Endogenous DNA Replication in Mammalian Cells
Burgess, PLoS One (2012)

Dynamics of replication foci during S-phase in differentiated cells

Quantitative Live Imaging of Endogenous DNA Replication in Mammalian Cells
Burgess, PLoS One (2012)
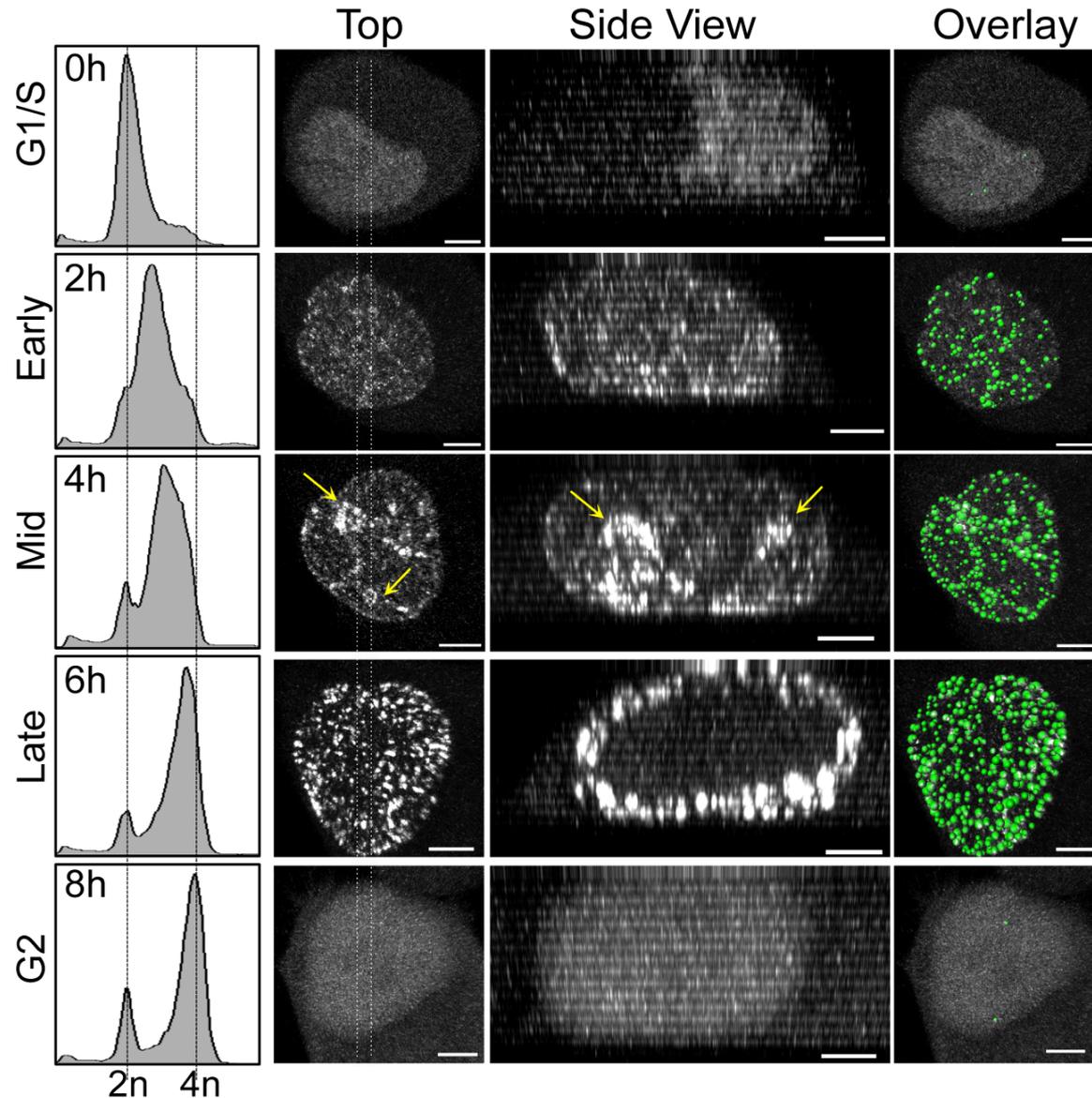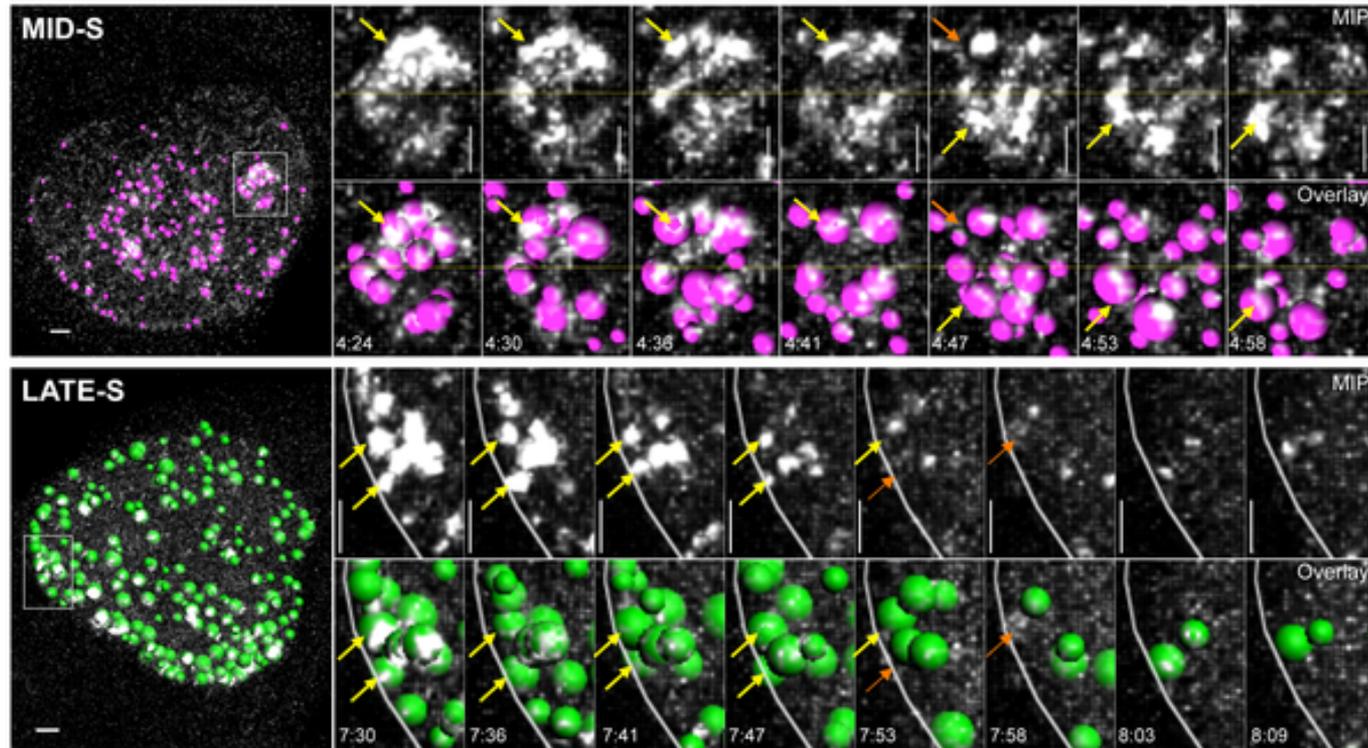
Epigenetically controlled functional and structural organization of the human genome

"Master" replication zones

3D

IMR90

H1 ES

Position (Mbp)

2D/3D

Early

Time

Late

Chromatin openness

Open   Closed   Open

1   2   3   4   3   2   1
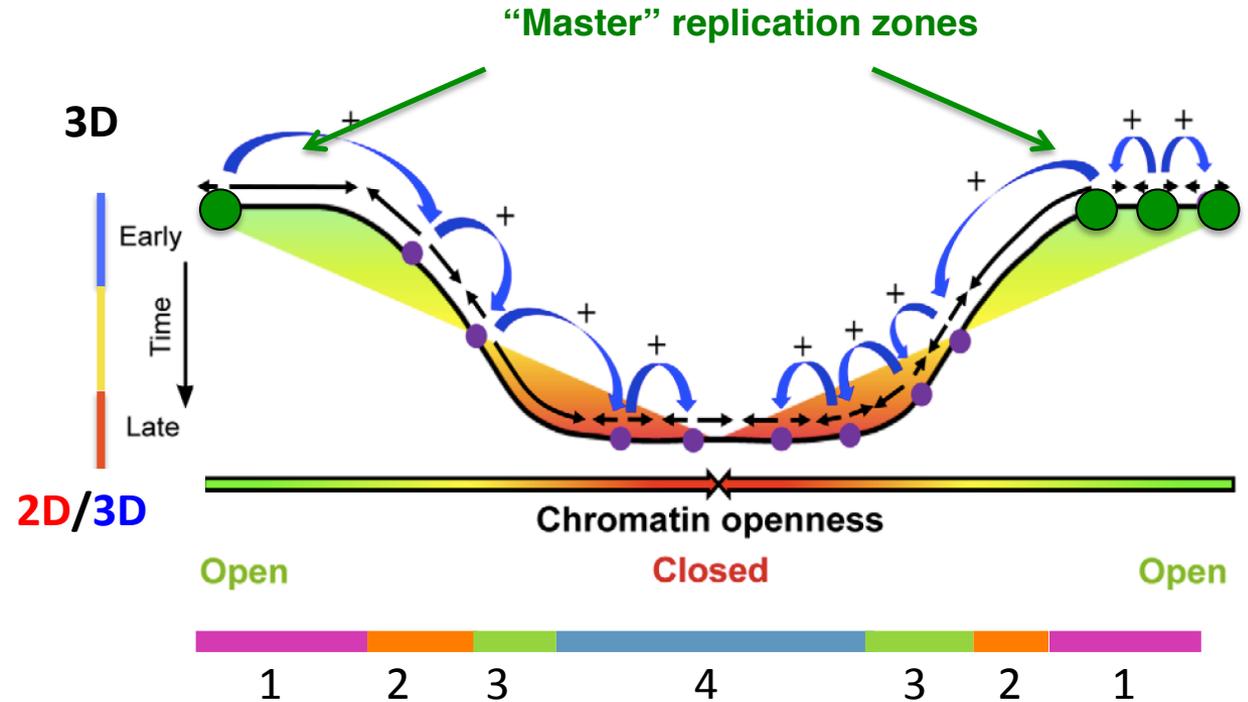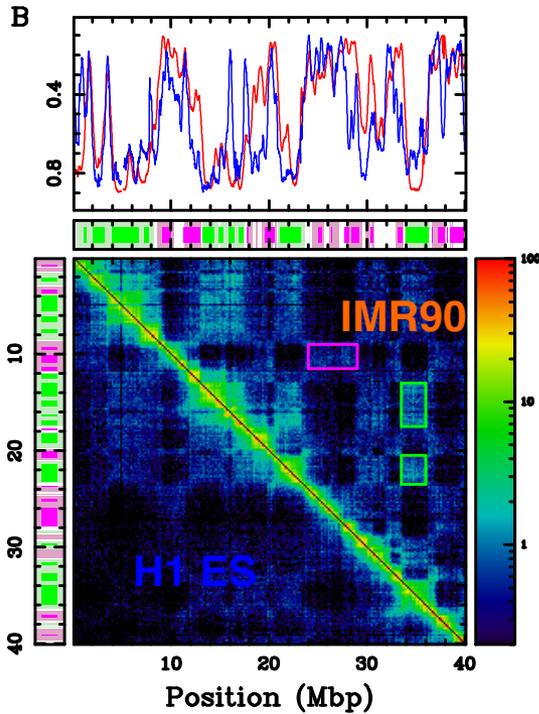
A cascade model for replication origin activation through 4 chromatin states along U-domains

Hyrien, Journal of Molecular Biology (2013)

Replication domains and chromatin conformation domains

Replication timing U-domains appear as large scale structural units

Baker, PLoS Computational Biology (2012)
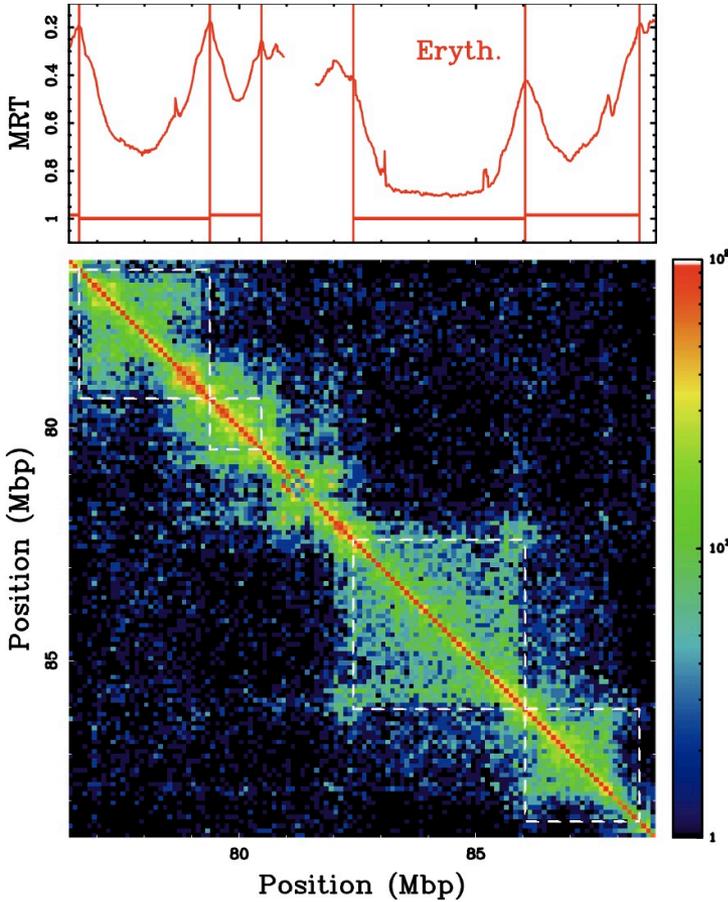
Replication domain boundaries share a near one-to-one correlation with topologically associating domains (TAD) boundaries

Dixon, Nature (2012)
Pope, Nature (2014)
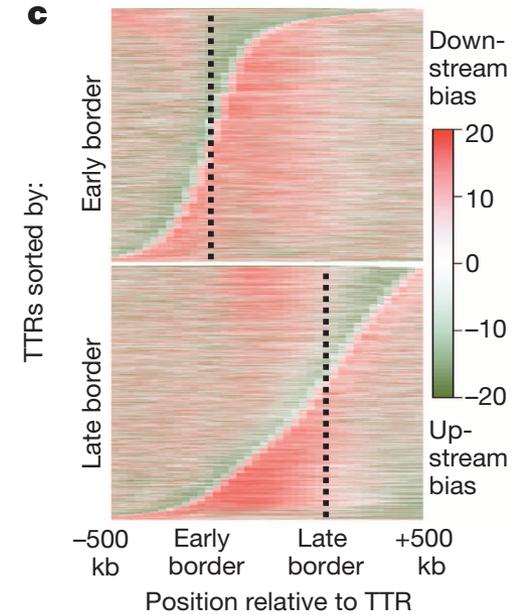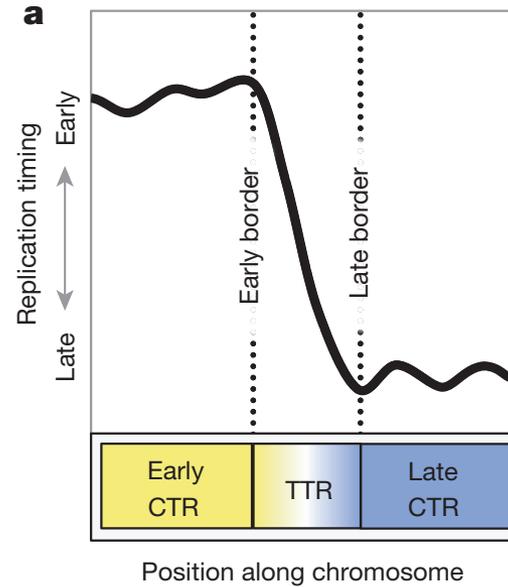
Replication domain borders are giant hubs in the chromatin conformation graph in K562

Co-localisation matrix as the adjacency matrix of the chromatin interaction graph

Long- (>4Mb) and **short- (<4Mb) distance interactions**
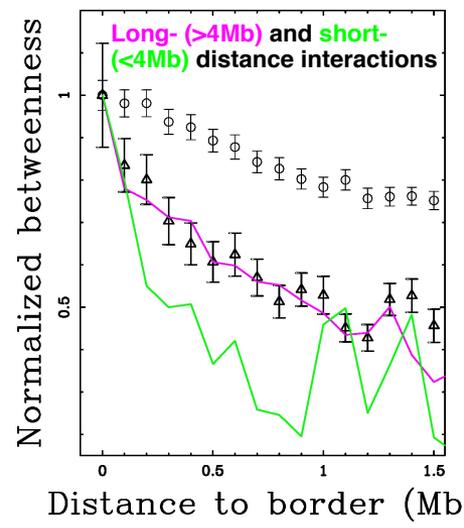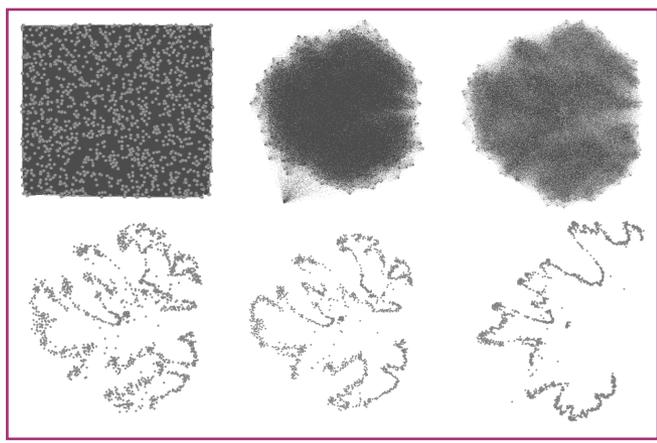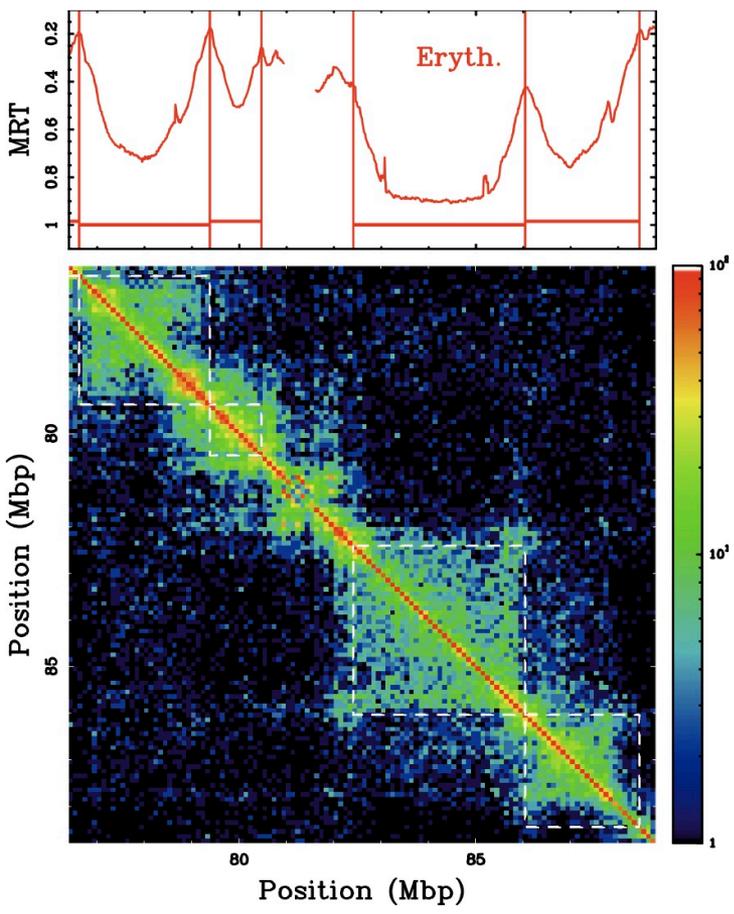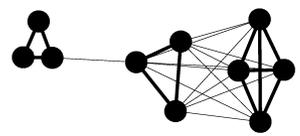
Intra-chromosomal betweenness centrality along replication domains

Replication domain borders and centers
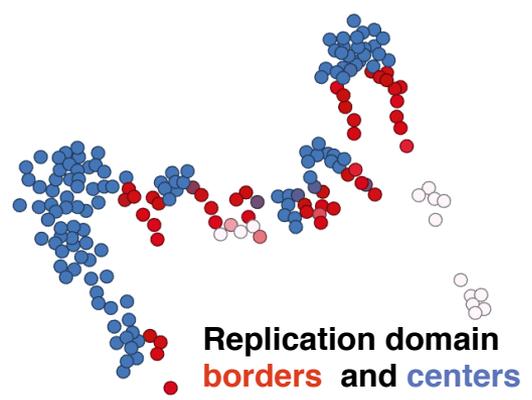
Delineating Structural Domains in Hi-C Data

Directionality index
Dixon, Nature (2012)
Pope, Nature (2014)

Interactions downstream
B

A
Interactions upstream

Putative boundary

⇐ A | B ⇒

Biased
downstream

Biased
upstream

Degree of bias
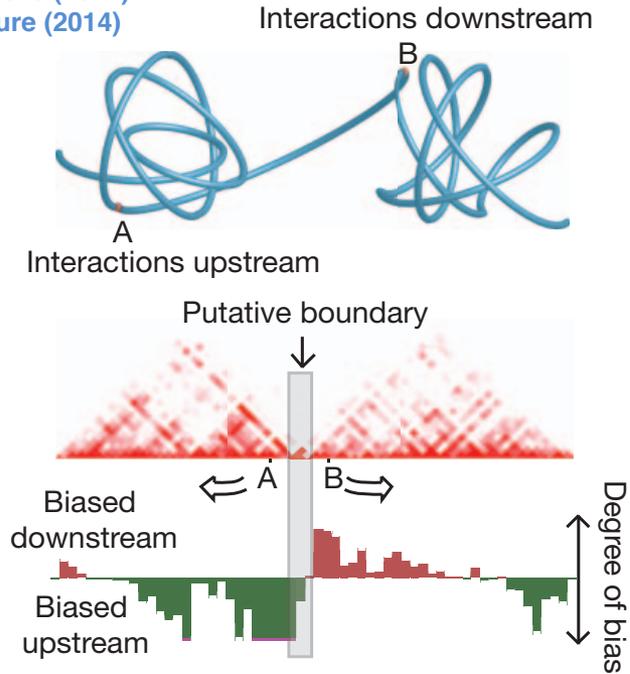
Wavelet change point detection
Shavit, Molecular BioSystems (2014)

e   1D Contact Profile
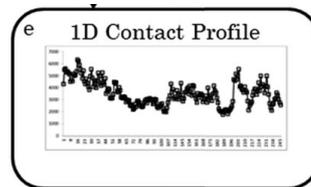
Dynamical programming
Filippova, Algorithms for Molecular Biology (2014)
Lévy-Leduc, Bioinformatics (ECCB 2014)

t_0 t_1 t_2 t_3 t_4

D_1 (μ_1)
R_2 (μ_0)
D_2 (μ_2)
R_3 (μ_0)
R_4 (μ_0)
D_3 (μ_3)
D_4 (μ_4)
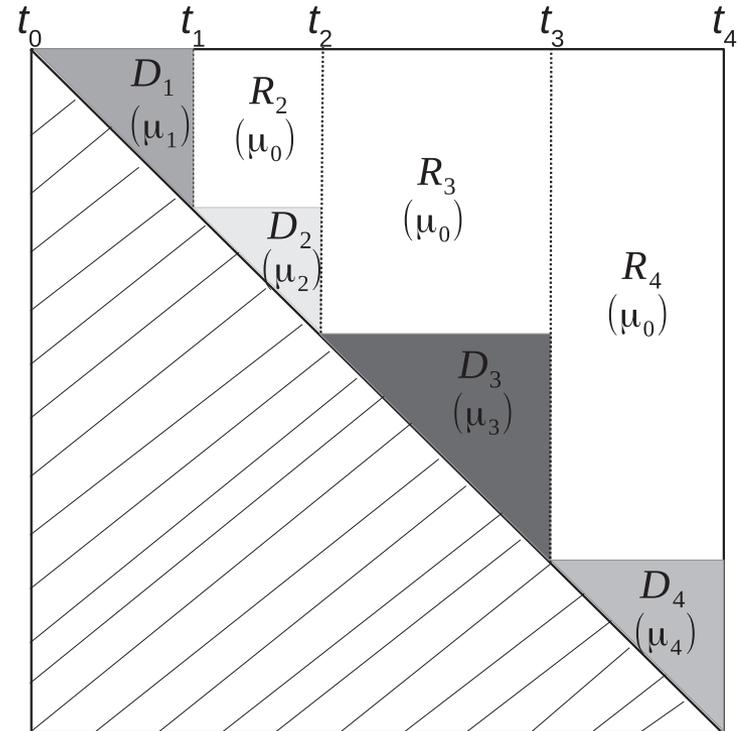
Most methods suppose that structural domains are chromosome intervals and/or do not allow for nested structures

Delineating Structural Domains in Hi-C Data
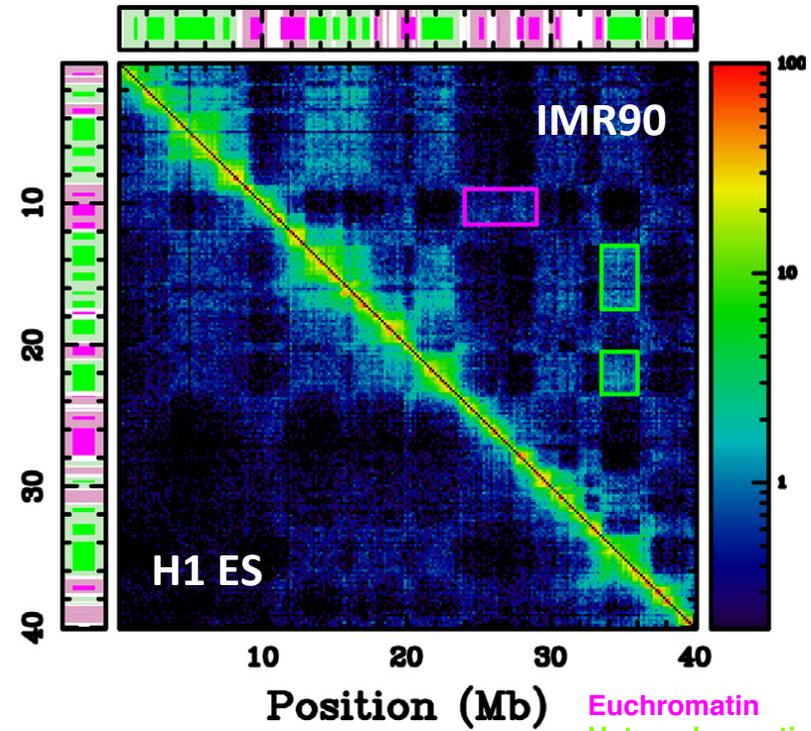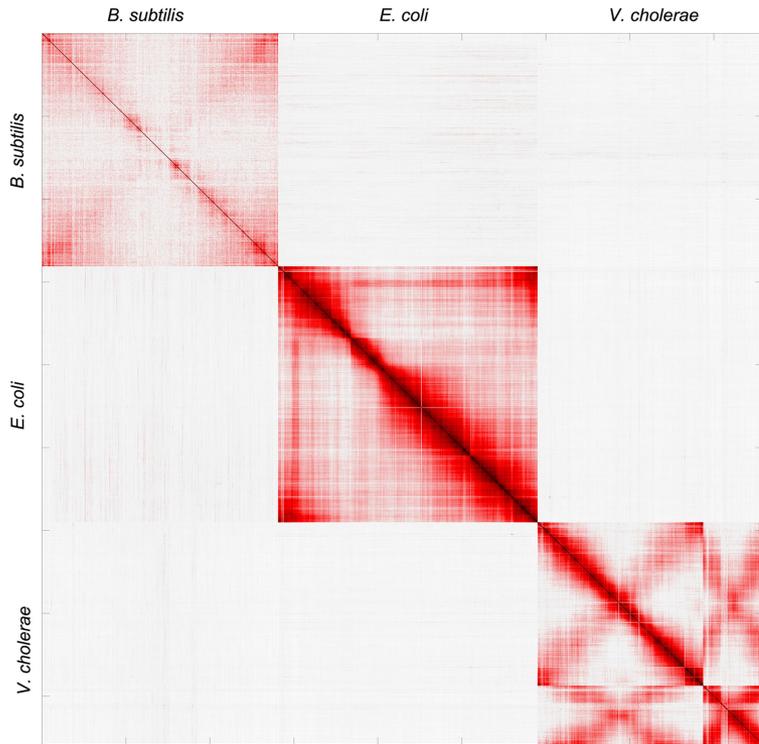
Genome assembly using chromosomal contact data

Burton, Genes Genomes Genetics (2014)
Marbouty, Elife (2014)
Marie-Nelly, Nature Communications (2014)

Boulos, FEBS letters (2015)

Detect multi-scale communities in the chromatin interaction graph

Graph clustering using spectral graph wavelets

Classical wavelets
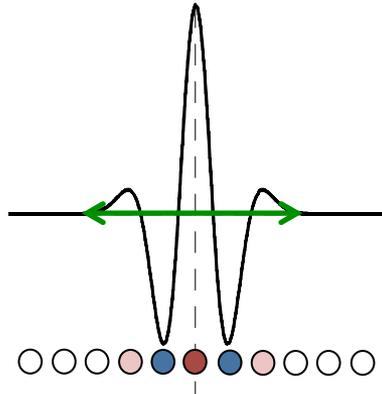
Spectral graph wavelets

small scale
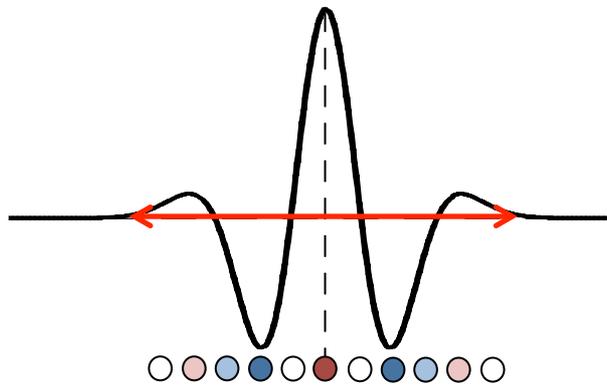
large scale
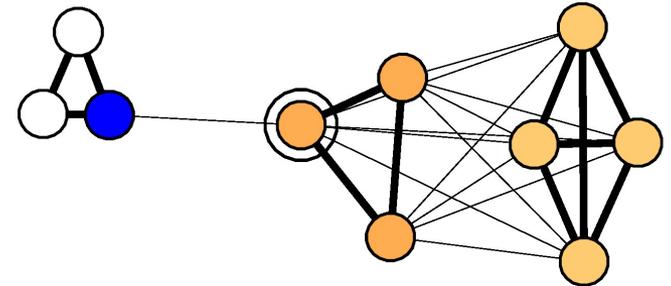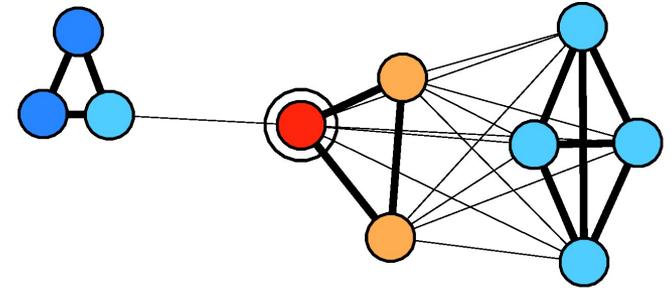
$$\psi_{b,a}(x) = \psi\left(\frac{x-b}{a}\right)$$

$$\widehat{\psi_{b,a}}(\omega) = e^{i\omega b}\widehat{\psi}(a\omega)$$

➡

$$\widehat{\Psi_{i,a}}(k) = \chi_i(k)g(a\lambda_k)$$

Graph clustering using spectral graph wavelets
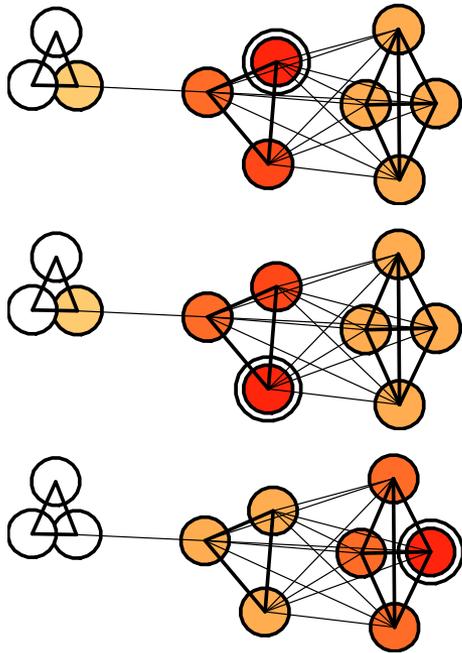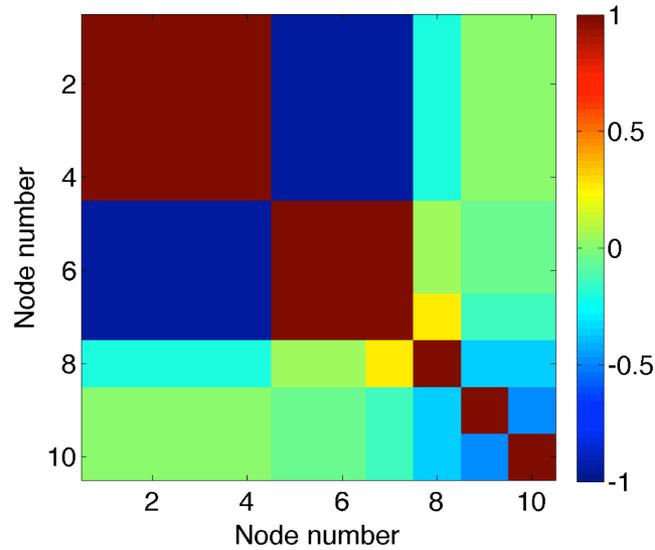
Wavelet coefficients **Wavelets correlation matrix** **Community detection**
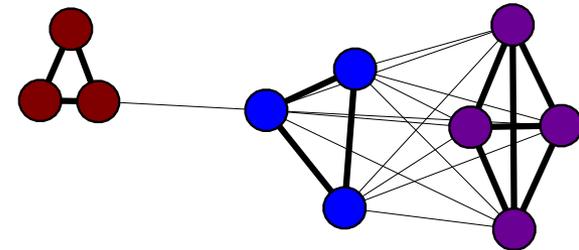
A partition in 3 communities at small scales

Graph clustering using spectral graph wavelets

Wavelet coefficients
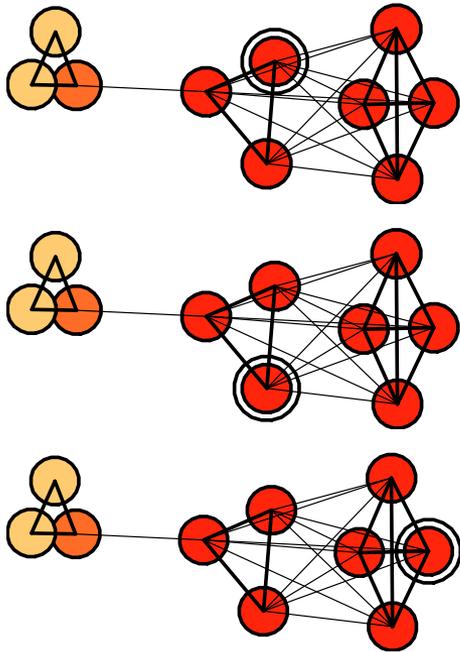
Wavelets correlation matrix
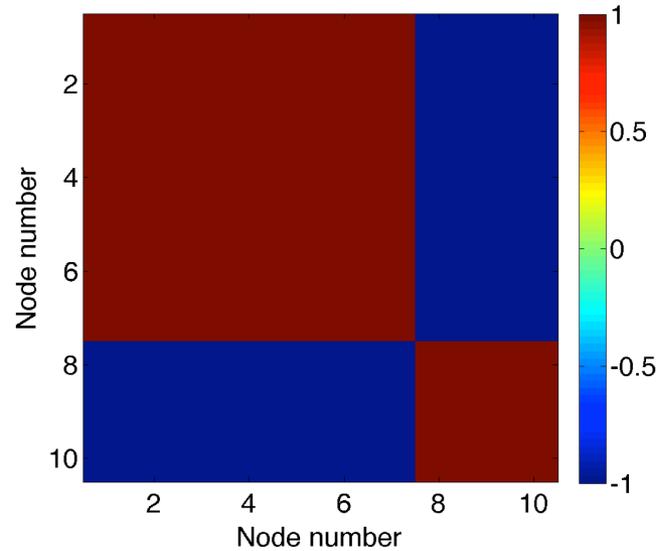
Community detection
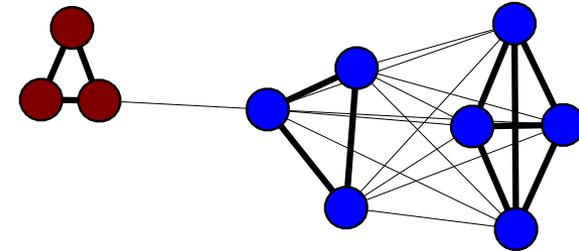
A partition in 2 communities at large scales

Graph clustering using spectral graph wavelets

small scale

large scale

Interaction matrix

Wavelet correlations

Wavelet correlations
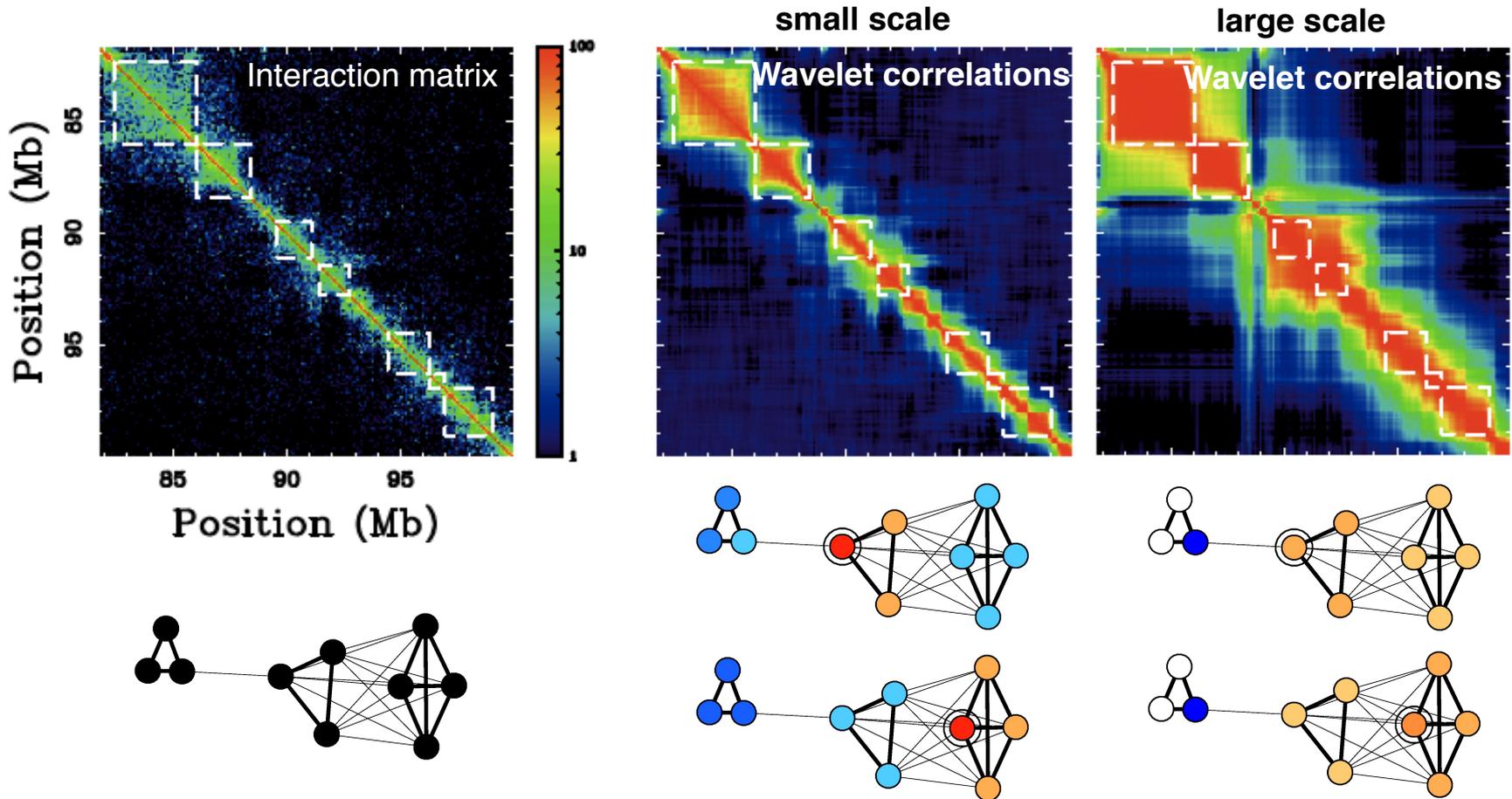
Obtaining an objective multi-scale segmentation of the human genome into structural communities

Multi-scale structural communities in IMR90 and H1 ES

Proportion of interval communities in groups of 100 communities

Communities size (Mb)
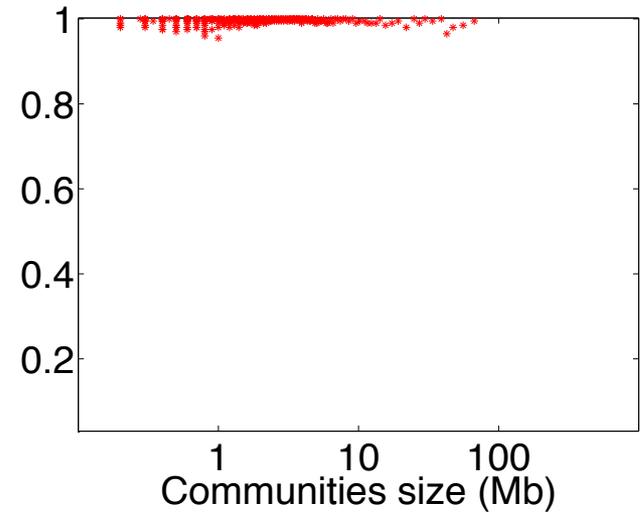
> 99% of structural communities are chromosomal intervals in both cell lines

Multi-scale structural communities in IMR90 and H1 ES

Multi-scale structural communities in IMR90 and H1 ES

Boulos, GRETSI (2015)

IMR90

H1 ES
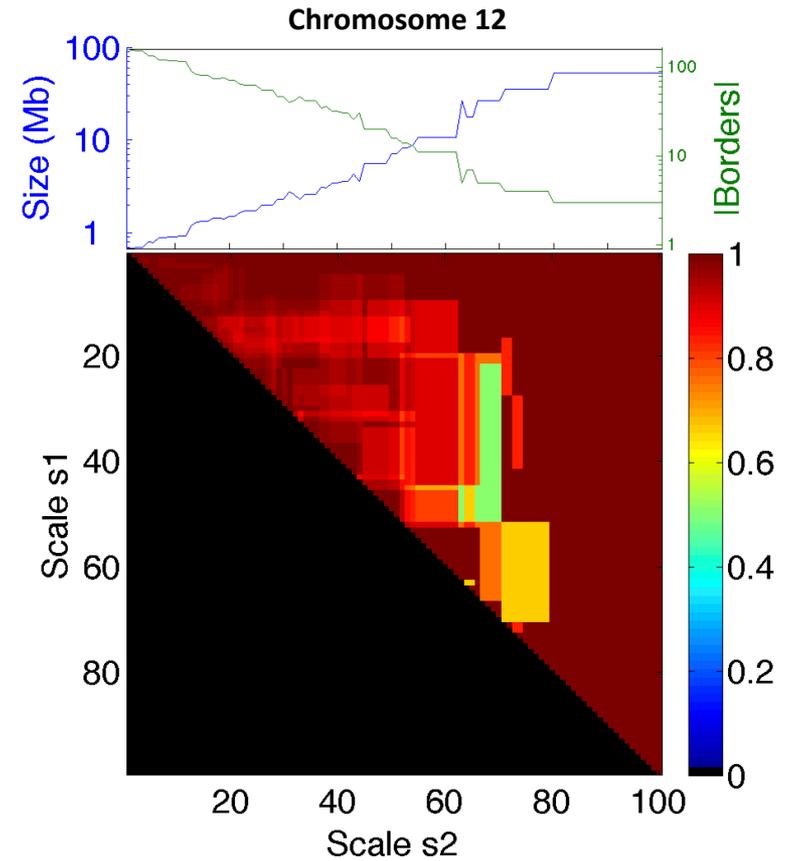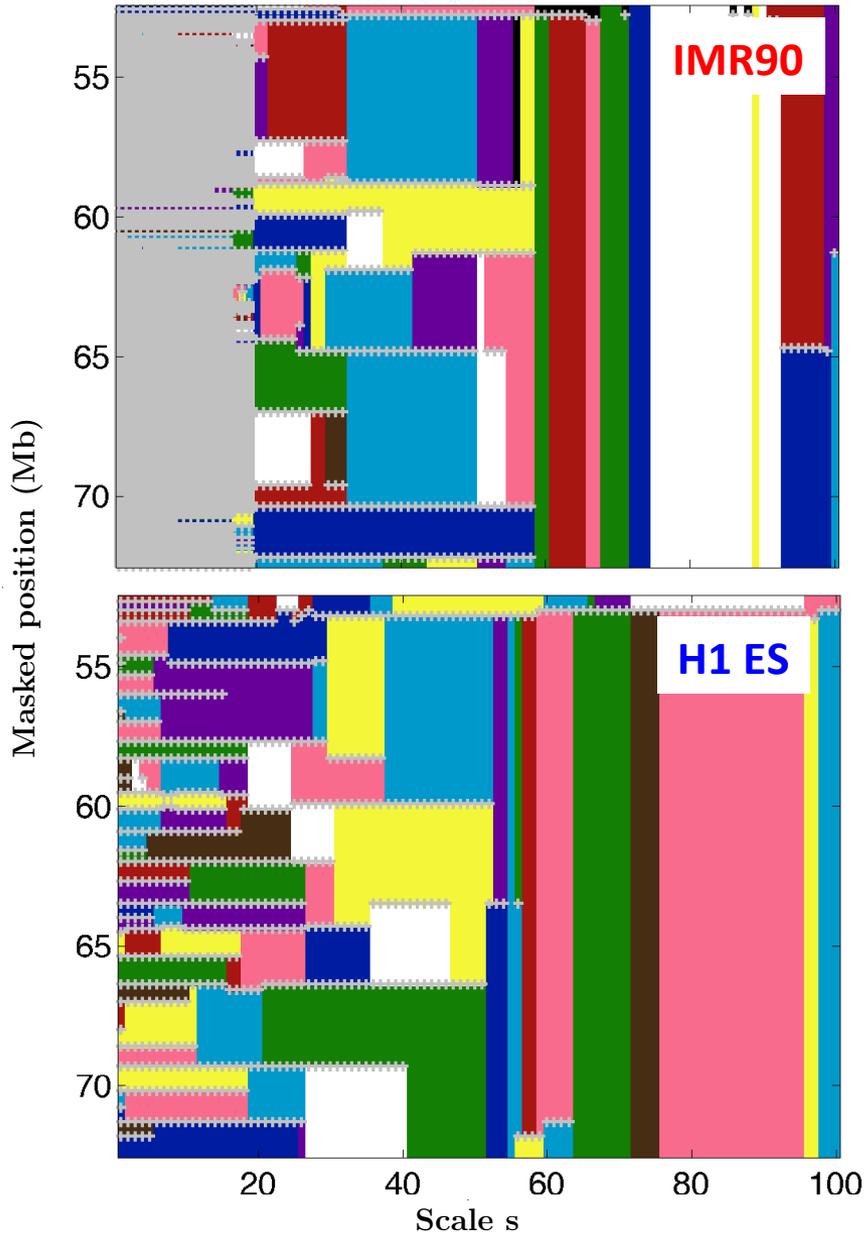
Chromosome 12

Structural communities form a hierarchy of chromosome intervals

Structural community borders have 'insulator' like properties

Boulos, PhD thesis (2015)
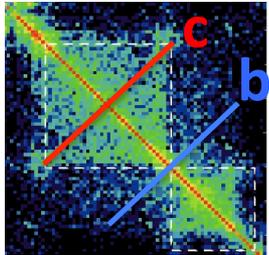
Interval communities
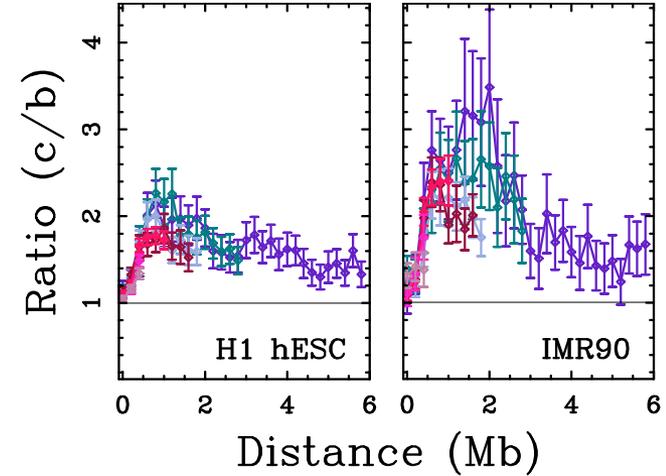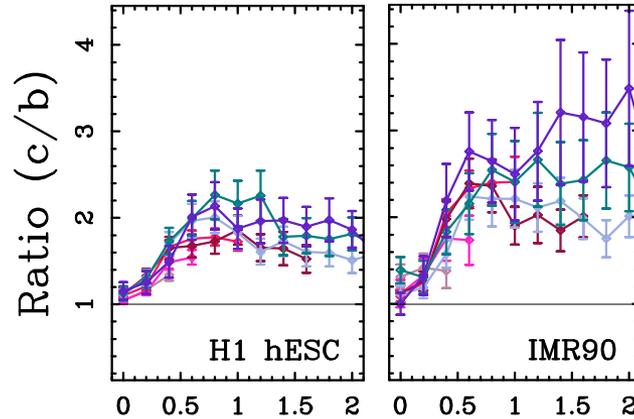
TADs
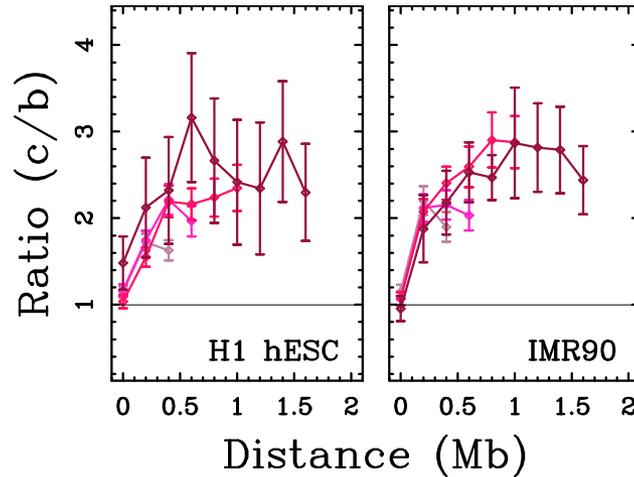
0.6 ≤ L < 1 Mb
1 ≤ L < 2 Mb
2 ≤ L < 3 Mb
3 ≤ L < 5 Mb
5 ≤ L < 10 Mb
10 ≤ L < 100 Mb

Structural communities vs TAD

Proportion of TADs that have a structural community counter part (80% mutual overlap)
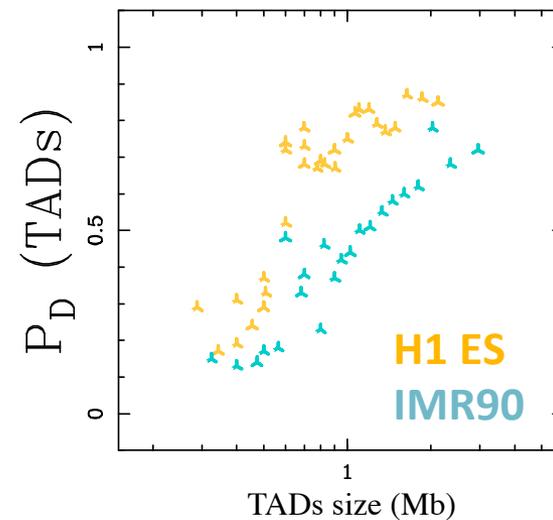
Proportion of TADs' borders that have a structural community border counter part (±100kp)
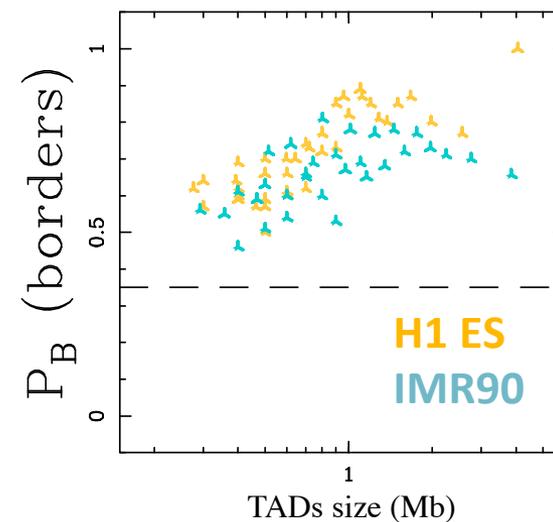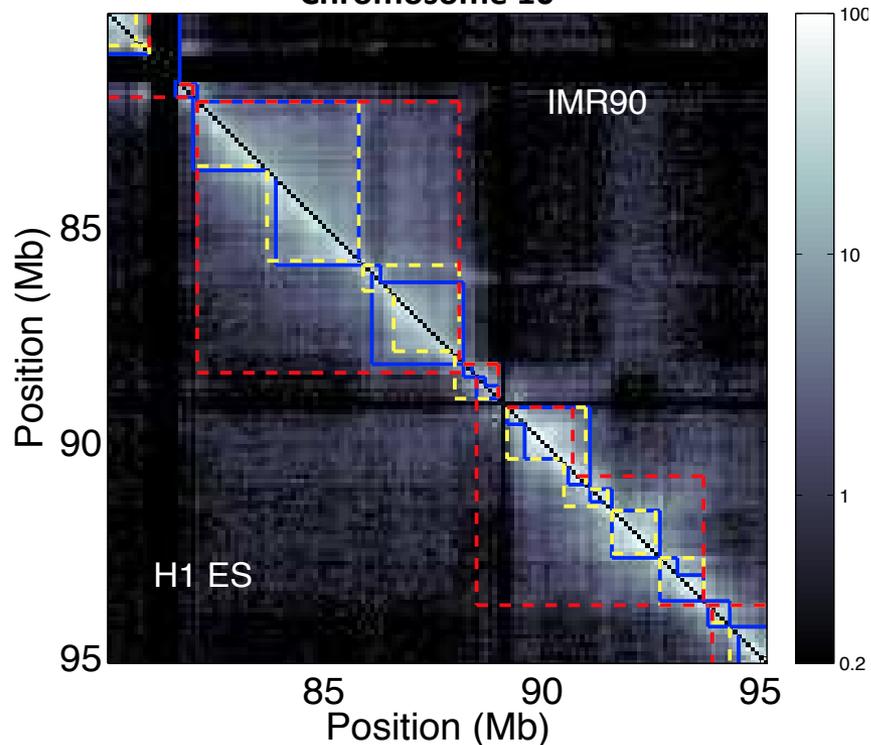
Large and small scale structural communities
TADS

Conservation of structural communities between cell lines
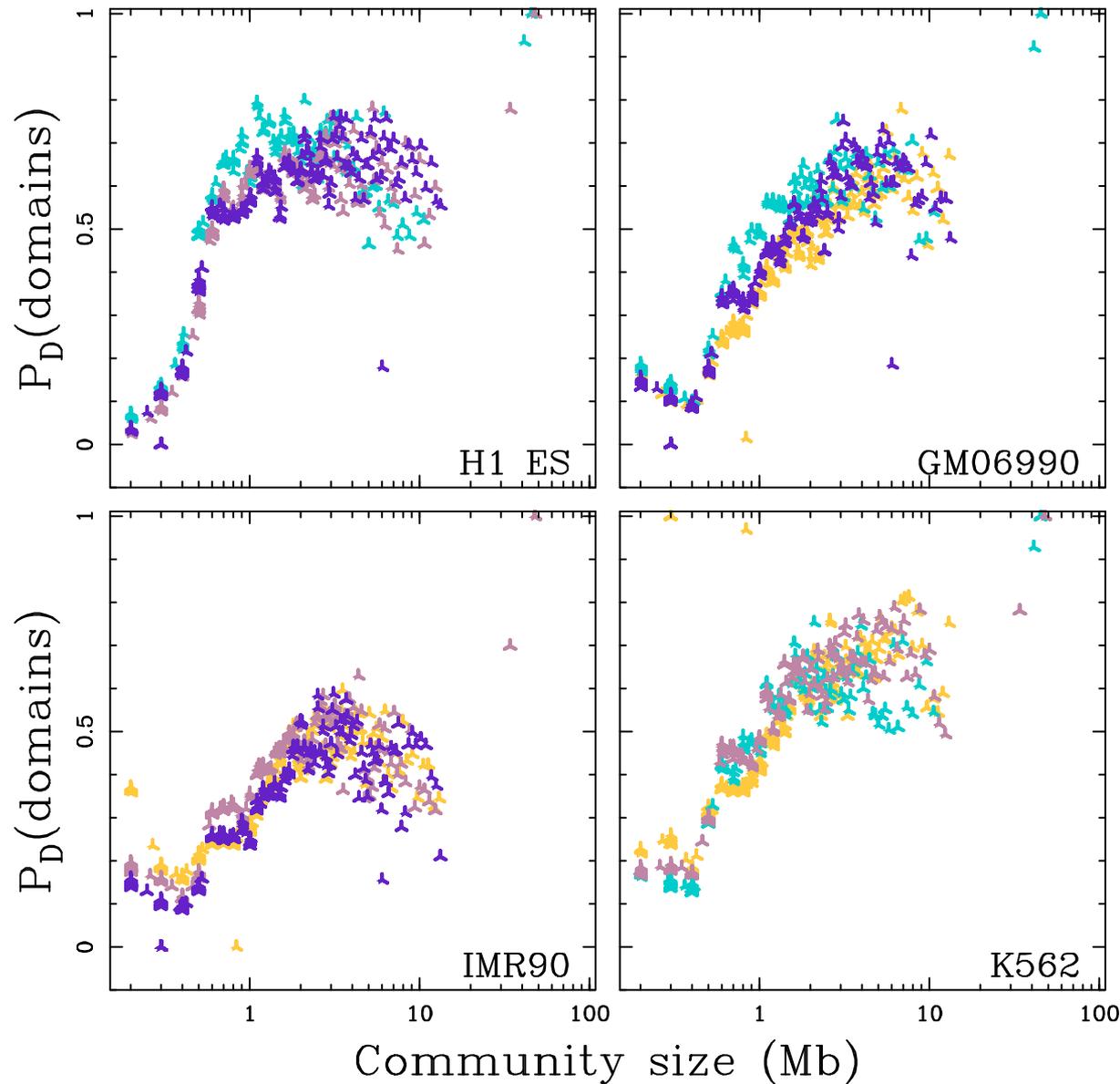
Proportion of structural communities in one cell line that have a counter part in an other cell line (80% mutual overlap)

H1 ES
IMR90
GM06990
K562

Structural communities vs replication U-domains

Boulos, PhD thesis (2015)

Large and small scale structural communities
Replication U-domains in IMR90 and H1 ES

Structural community borders are encoded in the DNA sequence via a local enrichment in nucleosome excluding energy barriers

Boulos, PhD thesis (2015)

E
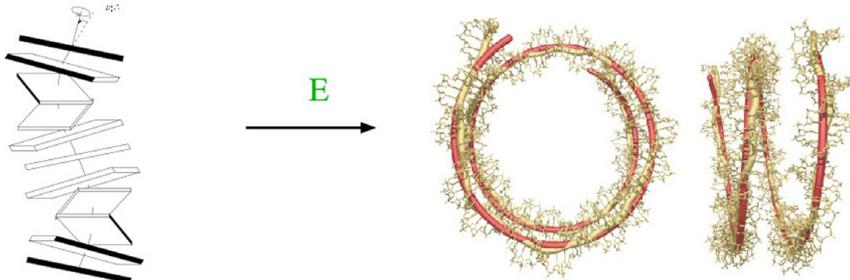
Vaillant, Physical Review Letters (2007)
Chevereau, Physical Review Letters (2009)

Drillon, Journal of Physics: Condensed Matter (2015)
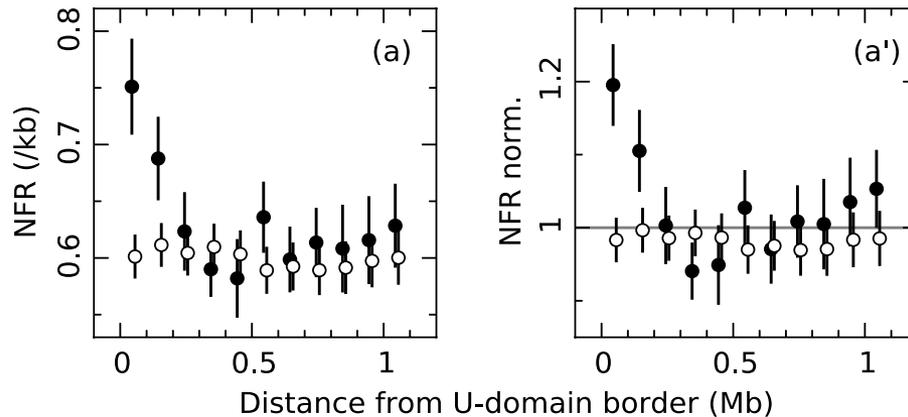
● : Ubiquitous U-domain borders
○ : GM specific U-domain borders

$0.6 \leq L < 1$ Mb $3 \leq L < 5$ Mb
$1 \leq L < 2$ Mb $5 \leq L < 10$ Mb
$2 \leq L < 3$ Mb $10 \leq L < 100$ Mb

Structural communities vs chromatin state domains

Chromosome 11
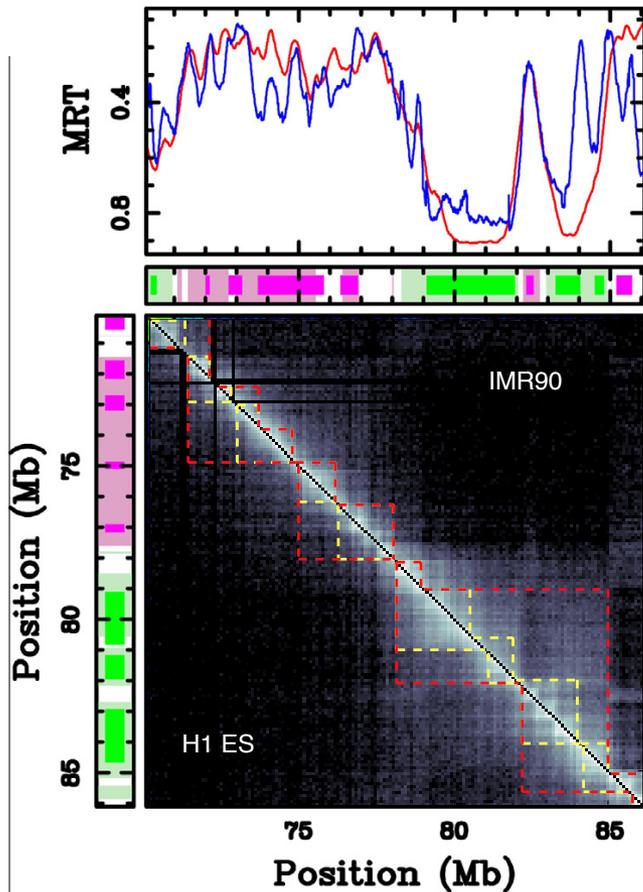
Proportion of chromatin domains that have a structural community counter part (80% mutual overlap)
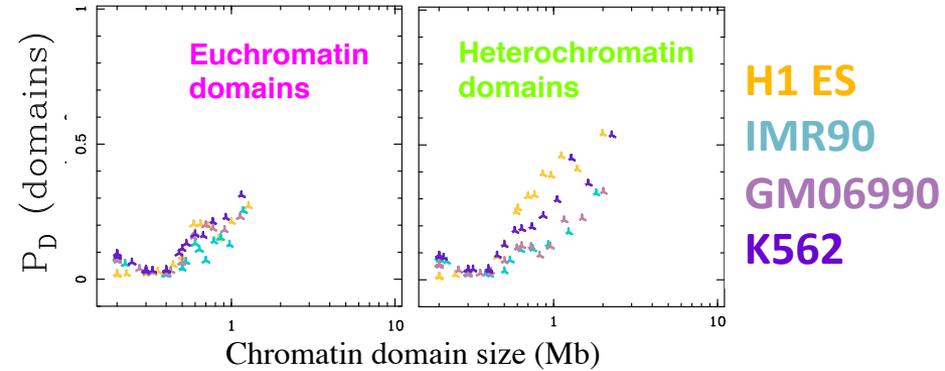
Proportion of chromatin domain borders that have a structural community border counter part (±100kp)
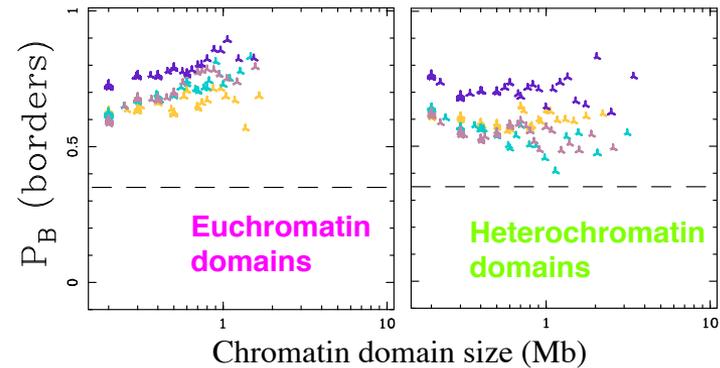
Euchromatin and **heterochromatin** domains
Large and **small** scale structural communities

Conclusion

Towards a multivariate view of genome organisation in the nucleus

Are they / what are the link between chromatin states, replication, transcription and structural domains ?

What are the implications for development, cancer progression ?

Gene transcription

DNA replication

Chromosome structure
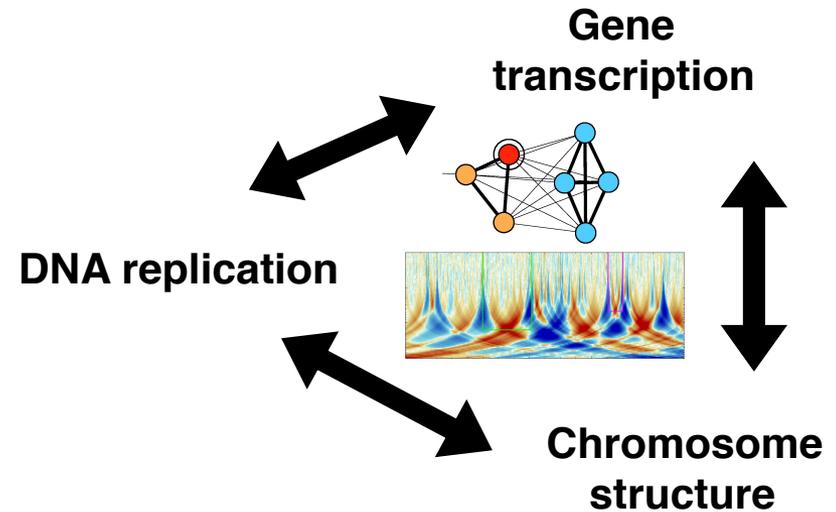
Characterizing replication stress in cancer using replication timing and fork polarity profiles

In some leukemias, replication timing changes associated to translocations precede and possibly predispose chromosomes to the translocation

Ryba, Genome Research (2012)

RUNX1 is involved in normal hematopoiesis and is one of the most frequently disrupted genes in leukemia
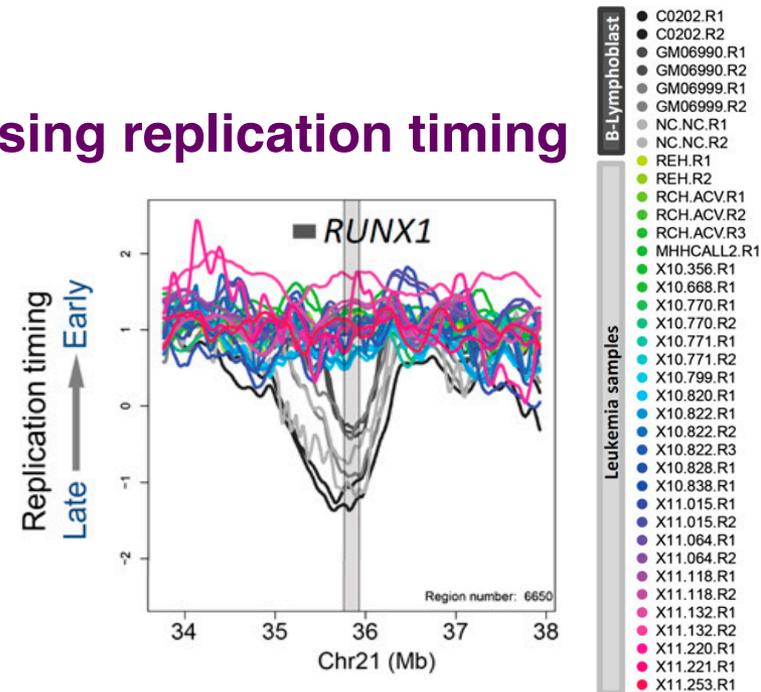
From the epigenome to the functional and structural nuclear organisation

Benjamin Audit

Equipe SIgnaux, SYstèmes et PHysiquE (SISYPHE)
Laboratoire de Physique de l'ENS-Lyon

ENS de Lyon

Françoise Argoul

Alain Arneodo

Antoine Baker

Pierre Borgnat

Rasha Boulos

Guénola Drillon

Pablo Jensen

Hanna Julienne

Nicolas Tremblay

Cédric Vaillant

Lamia Zaghloul

CGM, Gif-sur-Yvette

Yves d'Aubenton-Carafa

Chun-Long Chen

Claude Thermes

ENS, Paris

Guillaume Guilbaud

Olivier Hyrien

Malik Kahli

Nataliya Petryk

Aurélien Rappailles

CEA, Saclay

Arach Goldar

Towards graph spectral analysis

$$S(t) = \sum a(\omega)\cos(\omega t)$$
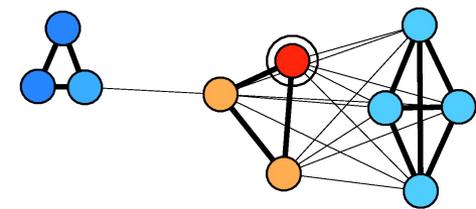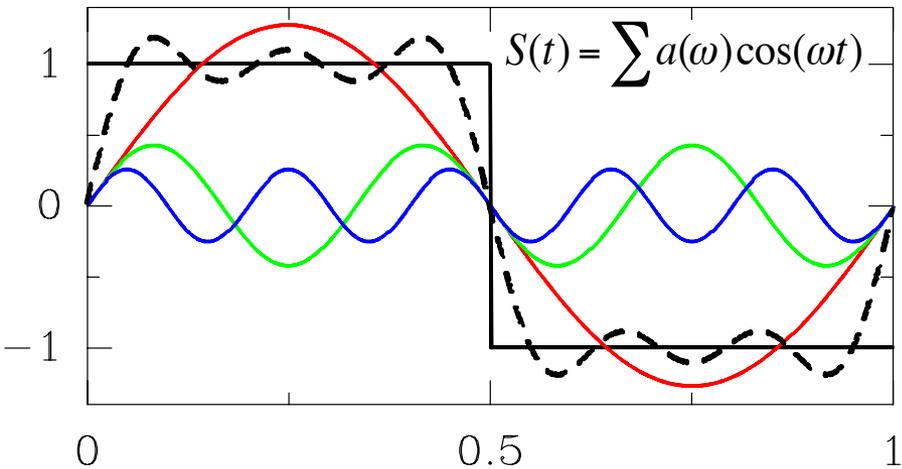
Fourier modes are eigen-functions of the Laplacian operator

$$\frac{d^2}{dt^2}\cos(\omega t) = -\omega^2 \cos(\omega t)$$

Discretized Laplacian operator *L*
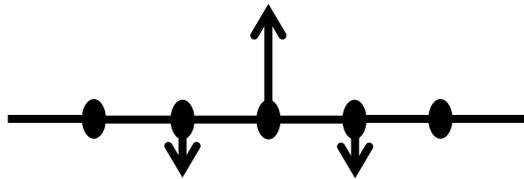
$$L = \begin{matrix} 2 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & -1 & 2 \end{matrix}$$

Discrete Fourier modes are eigen-functions of *L*

$$L\chi_i = \lambda_i \chi_i$$

L* is related to the adjacency matrix *A* and degree matrix *D* of the linear graph: *L = D - A

$$A = \begin{matrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{matrix} \qquad D = \begin{matrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{matrix}$$

Towards graph spectral analysis

Graph Fourier modes are the eigen-functions of the graph Laplacian operator $L = D - A$
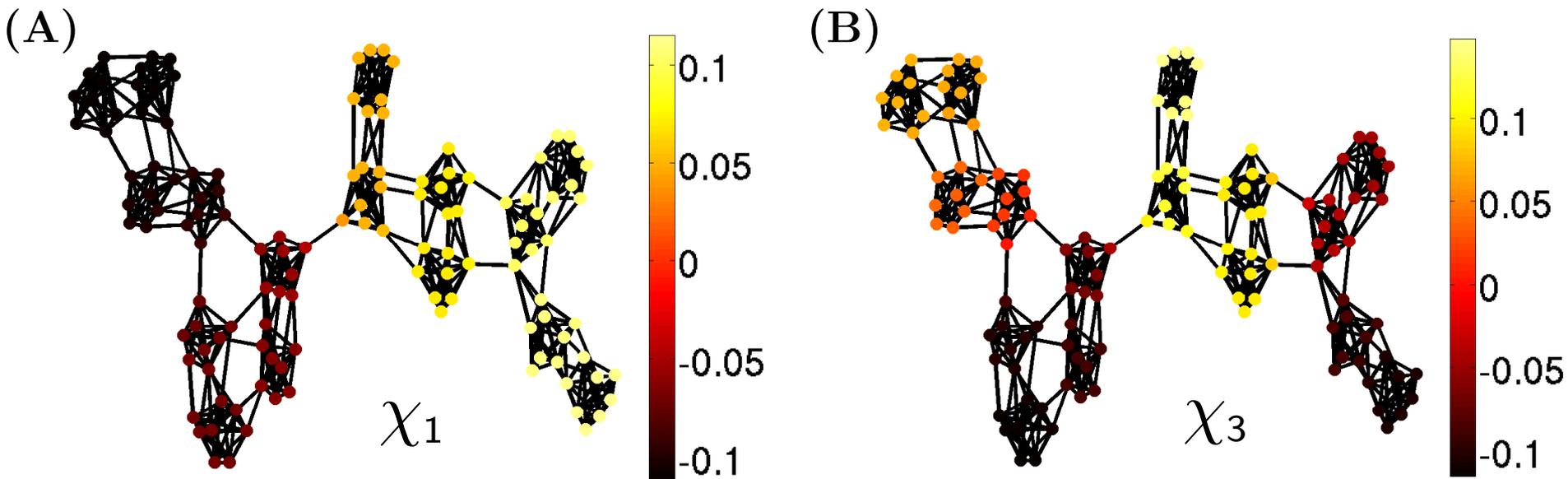
Some Fourier modes

Graph Fourier modes convey information on the graph topology
Used in graph spectral clustering

Structural communities during HeLaS3 cell cycle

G1   Mitosis