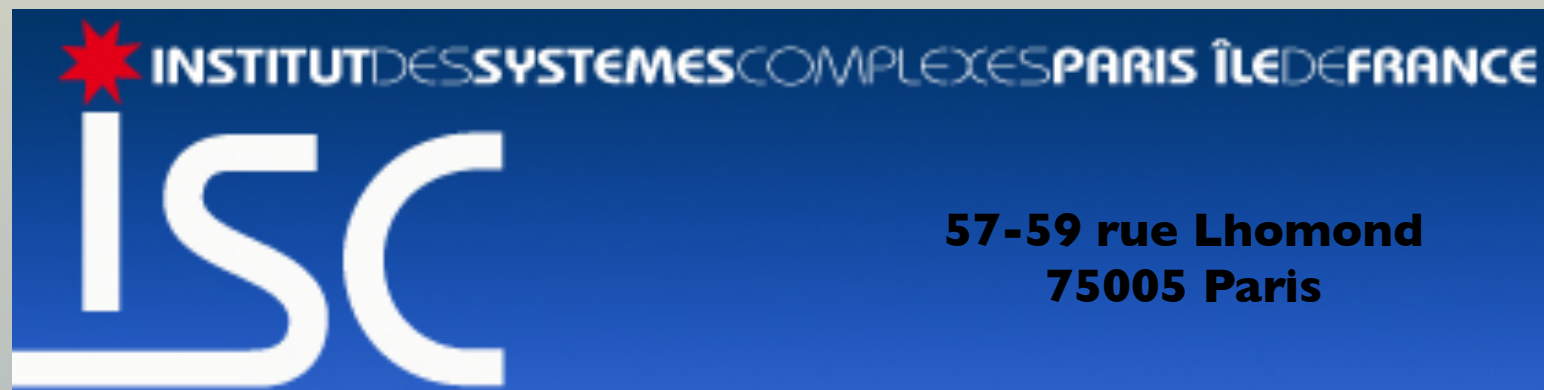


Insights into the functional organization of genomes: confronting **polymer physics** & **comparative genomics**

Ivan Junier



<http://iscpif.fr/~junier>

Joint work with :

François Képès

Epigenomics project
Genopole, Evry, France



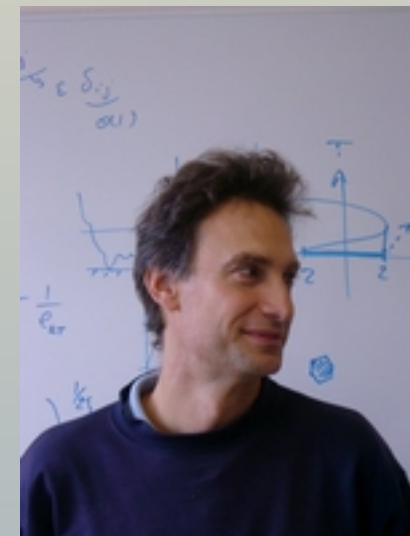
Joan Hérissou

Epigenomics project
Genopole, Evry, France



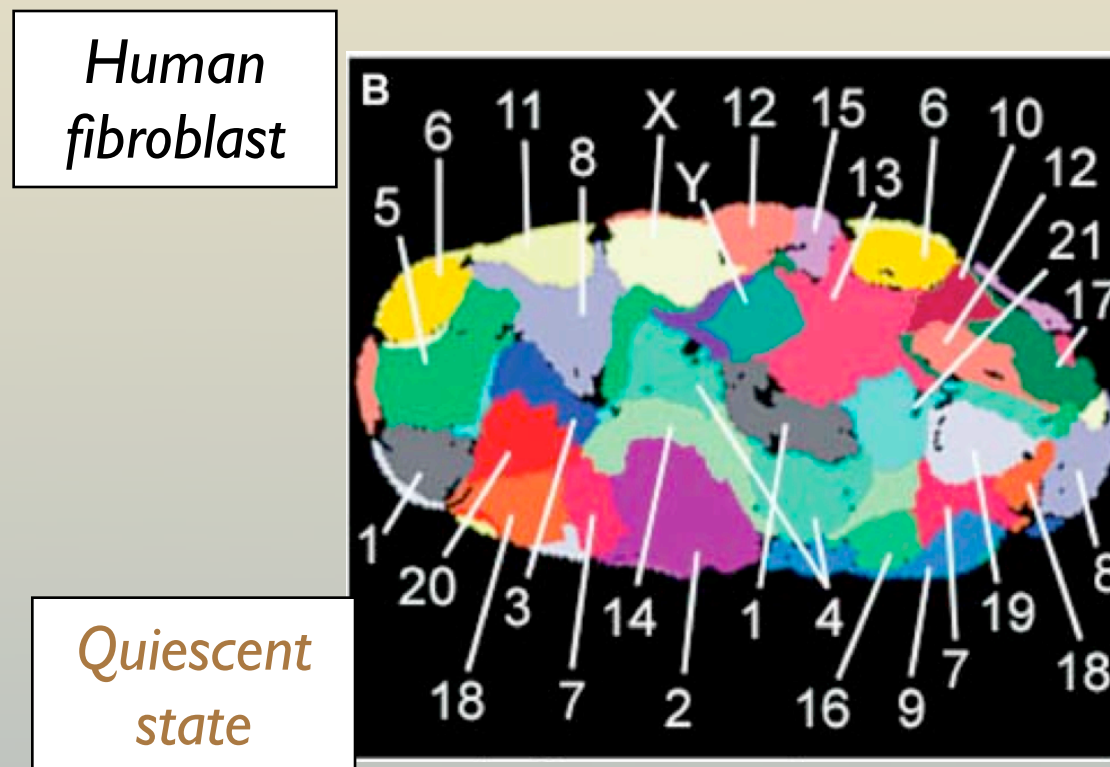
Olivier Martin

LPTMS, Orsay, France



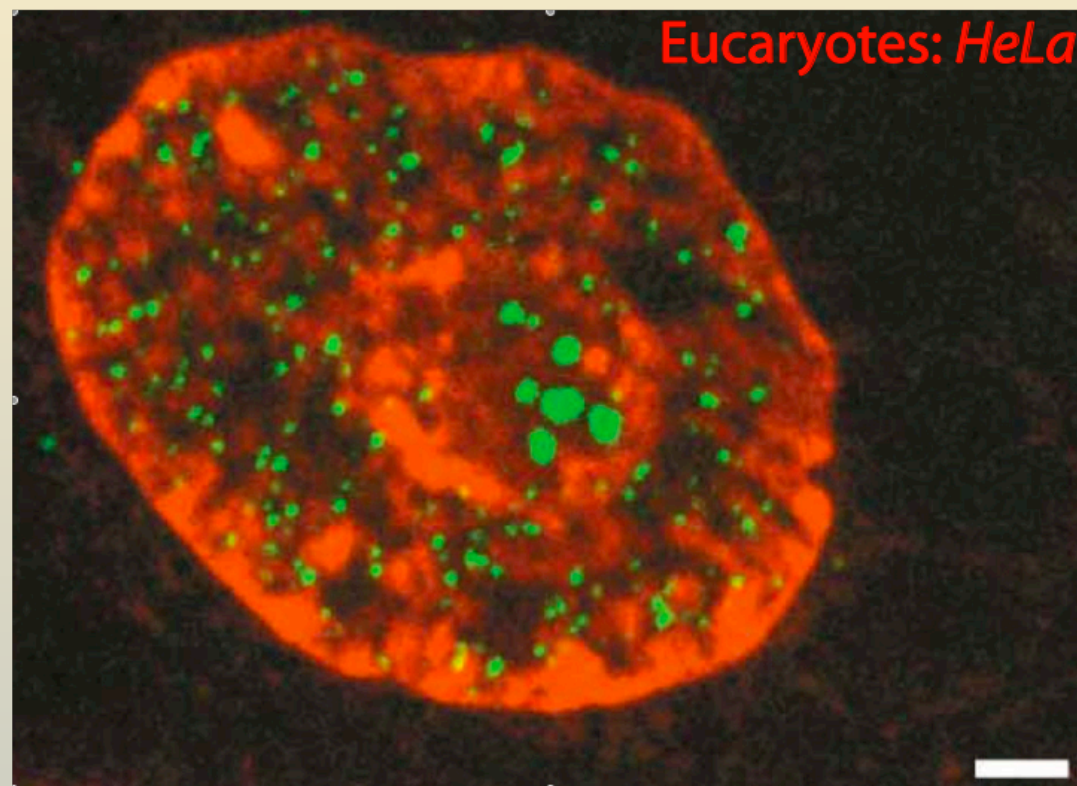
Chromosomal organization is not random

Example: chromosome territories



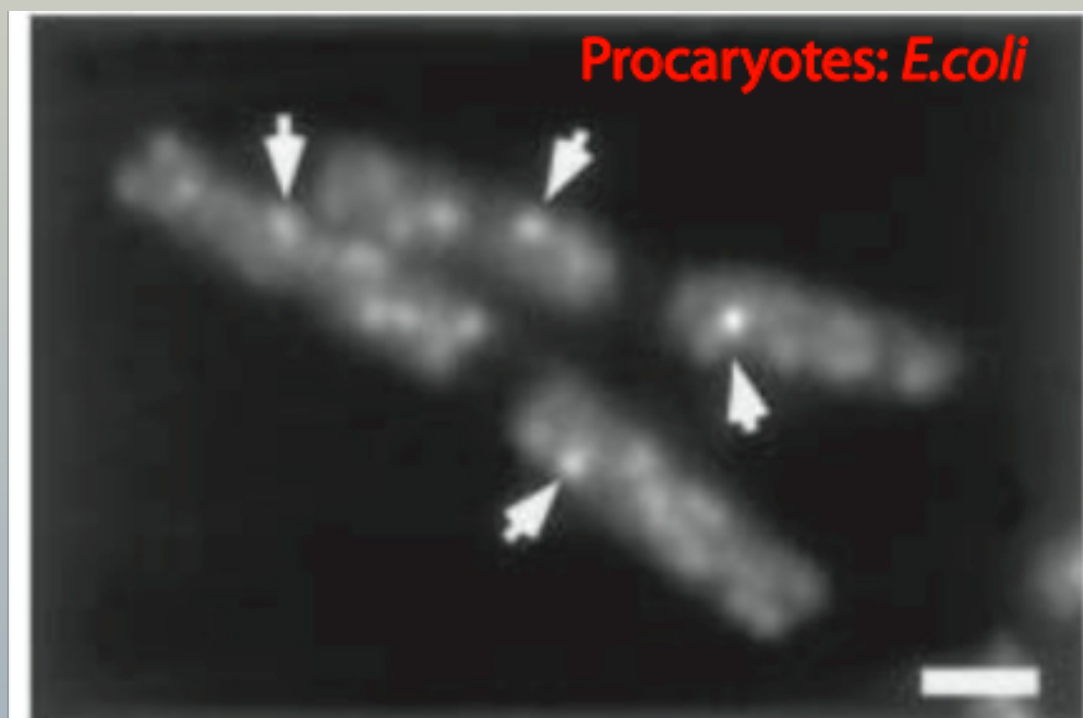
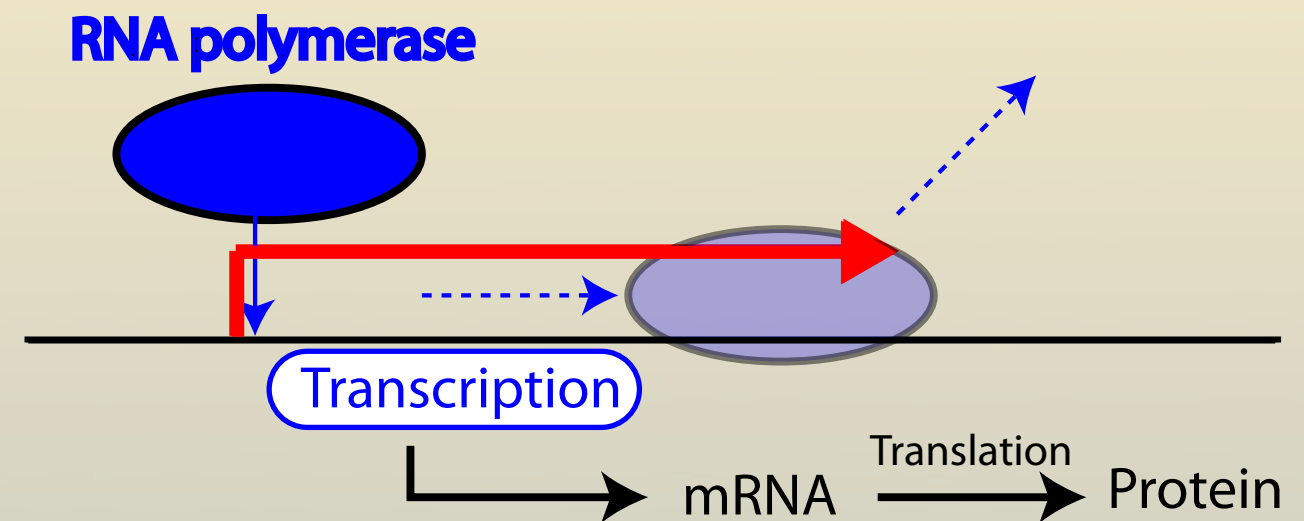
Bolzer et al, *PLoS Biology*, 3, e157, 2005

Transcriptional **activity** in the cellular space



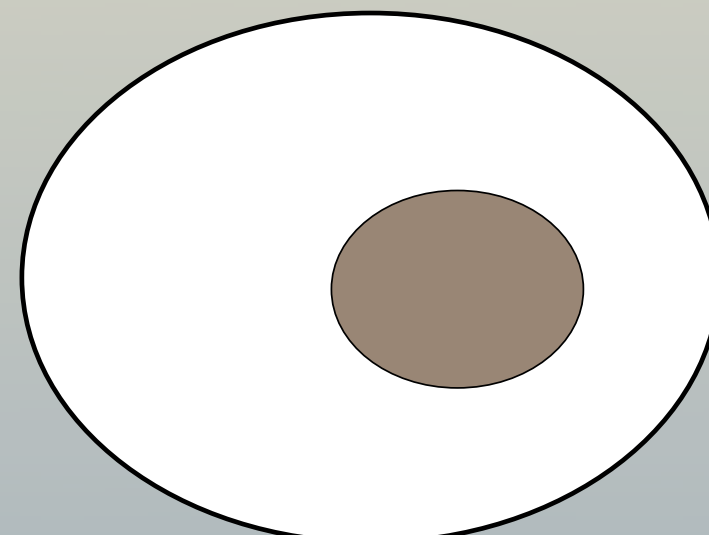
Cook, Nature Genetics, 2002

2 μm



Cabrera & Jin, Mol. Microbiol., 2003

1 μm

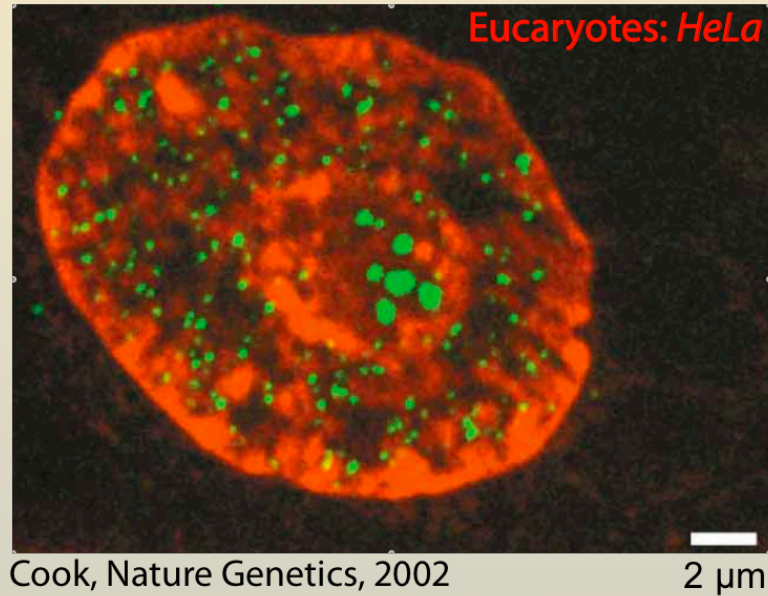


Eukaryote

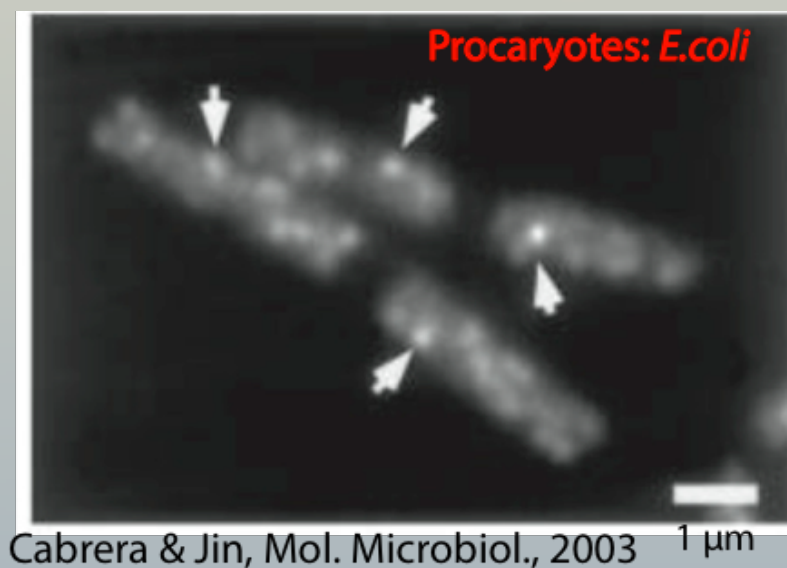


Prokaryote

Part I



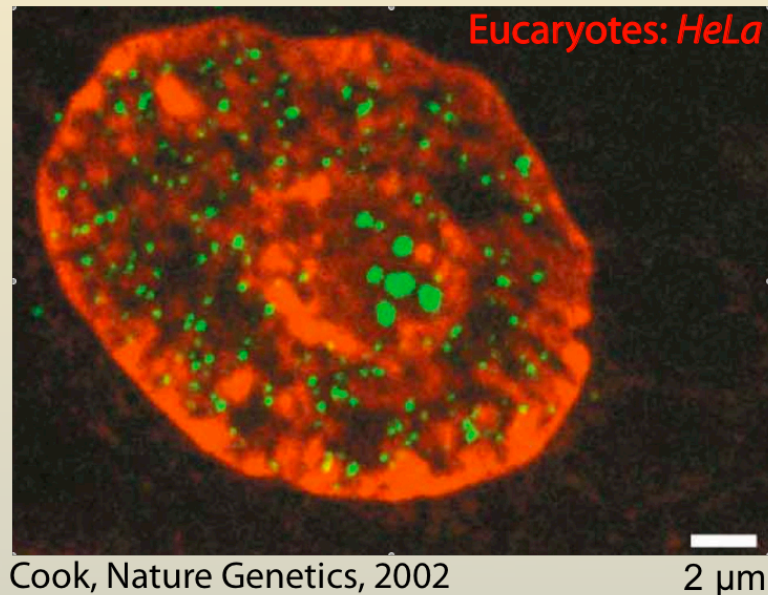
Active or passive?
DNA conformation?



Transcription factories

Self-organized process where
genome organization plays a crucial role (Part II)

Transcriptional regulation in the cellular space



“Associations between *distal genes* are transcription-dependent and require coassociation with a shared transcription factory.”

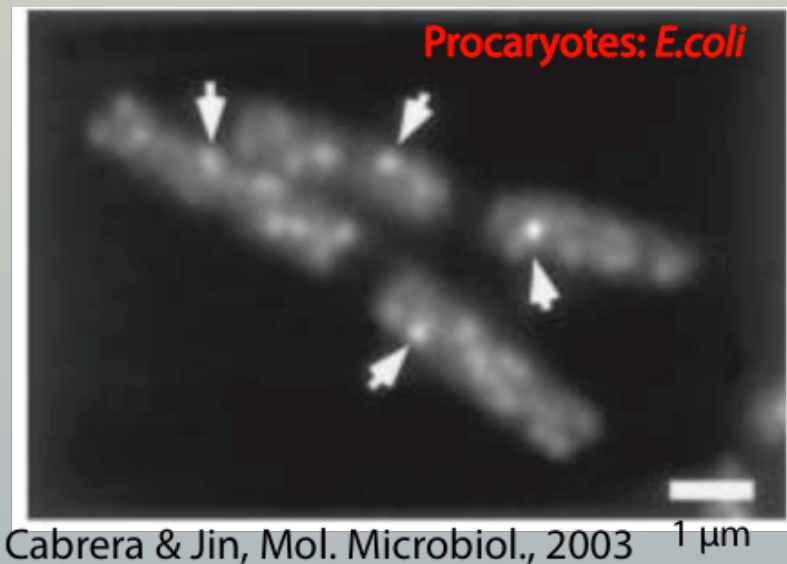
Mitchell & Fraser, *Genes Dev.*, 22, 20-5, 2008

“Similar active genes cluster in *specialized* shared transcription factory.”

Xu & Cook, *J. Cell Biolohy*, 181, 615, 2008

“active *co-regulated* genes and their *regulatory factors* cooperate to create specialized nuclear hot spots optimized for efficient and coordinated transcriptional control.”

Schoenfelder et al., *Nature Genetics*, 42, 53, 2010



“Active transcription of *rRNA operons* condenses the nucleoid in *Escherichia coli*.”

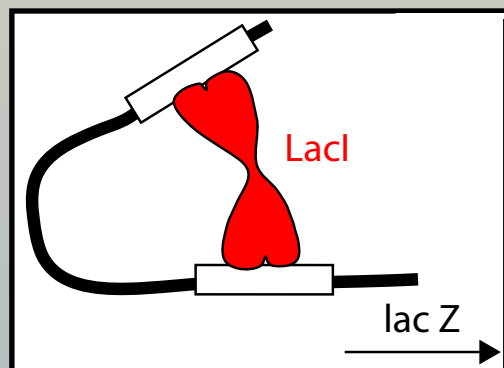
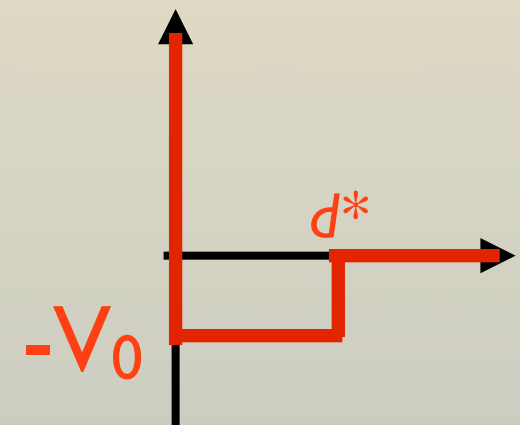
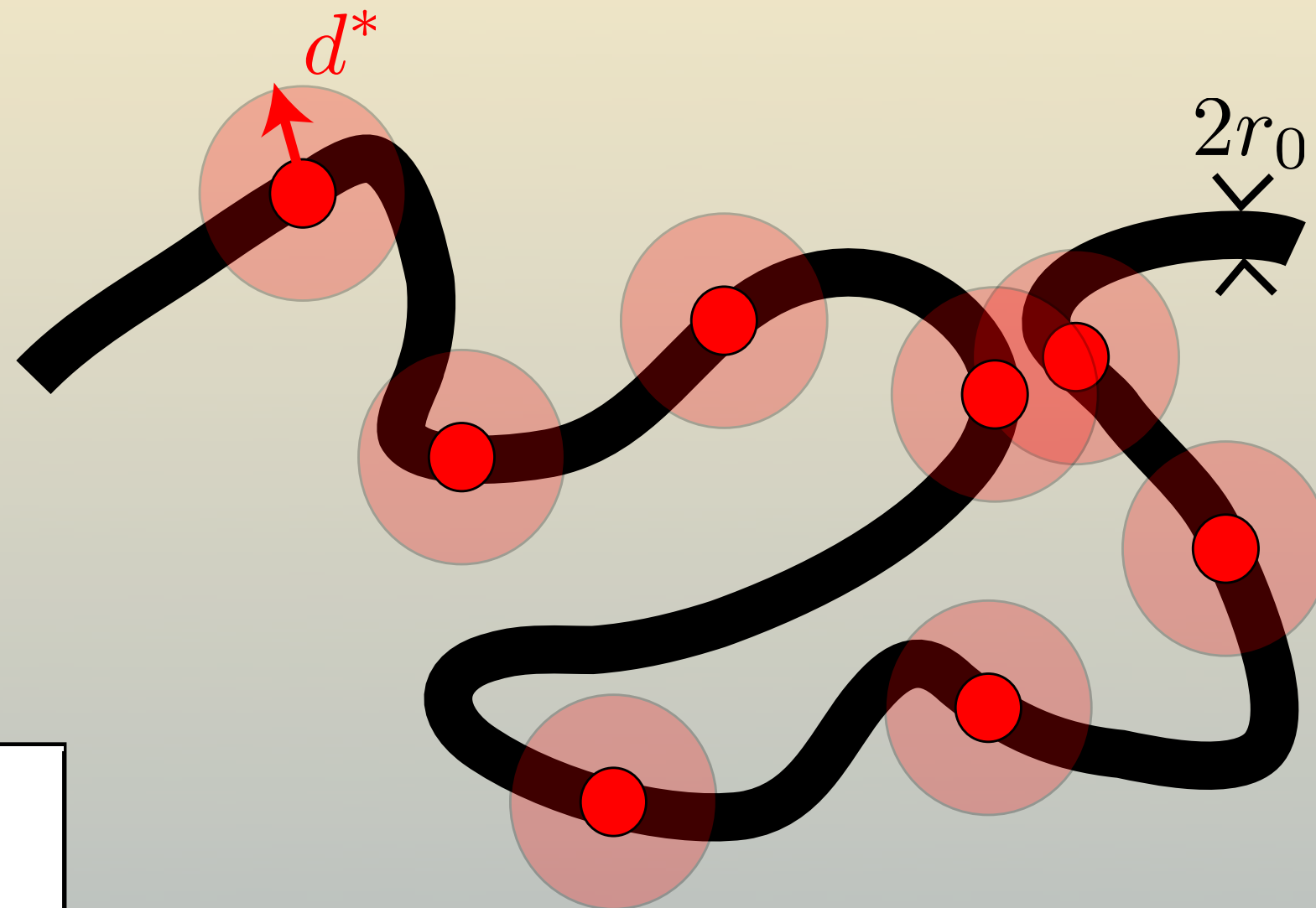
Cabrera, Cagliero, Quan, Squires, Jin, *J. Bacteriology*, 191, 4180, 2009

“Two types of localization of the *DNA-binding proteins* within the *Escherichia coli* nucleoid.”

Azam, Hiraga, Ishihama, *Genes to Cells*, 5, 613, 2000

Taniguchi et al., *Science*, 329, 533, 2010

Understanding the formation of transcription factories using single chain polymer physics



Transcription factories as a micro-phase separation

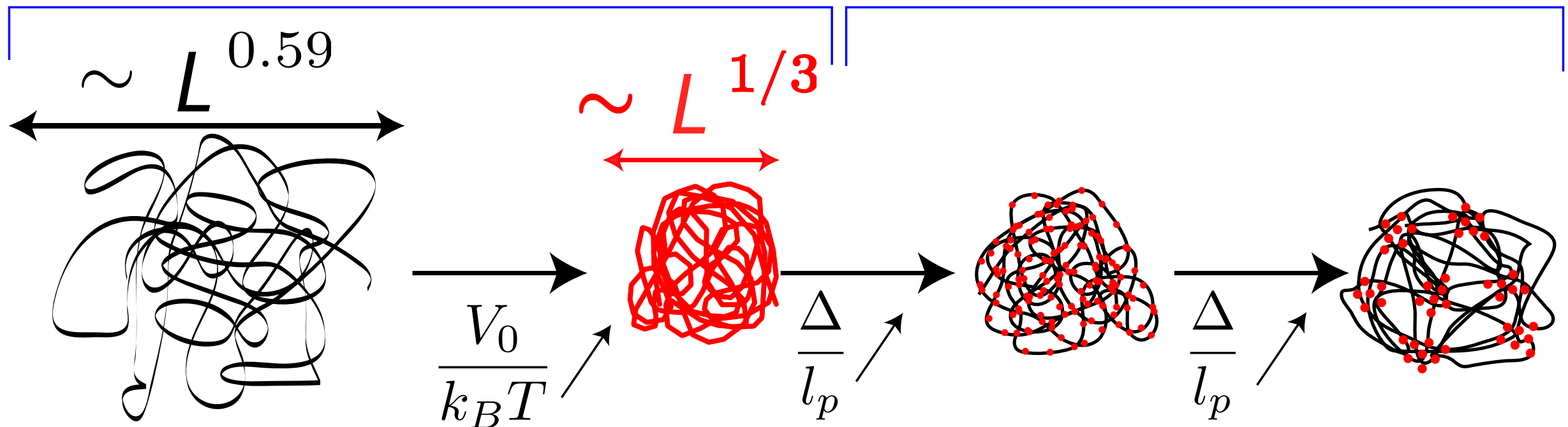
V_0 ($\sim \text{few } k_B T \sim \text{few kcal/mol}$) Protein-DNA binding free energy

Δ Mean distance between two successive interacting sites along the DNA

l_p Persistence length of DNA/chromatin

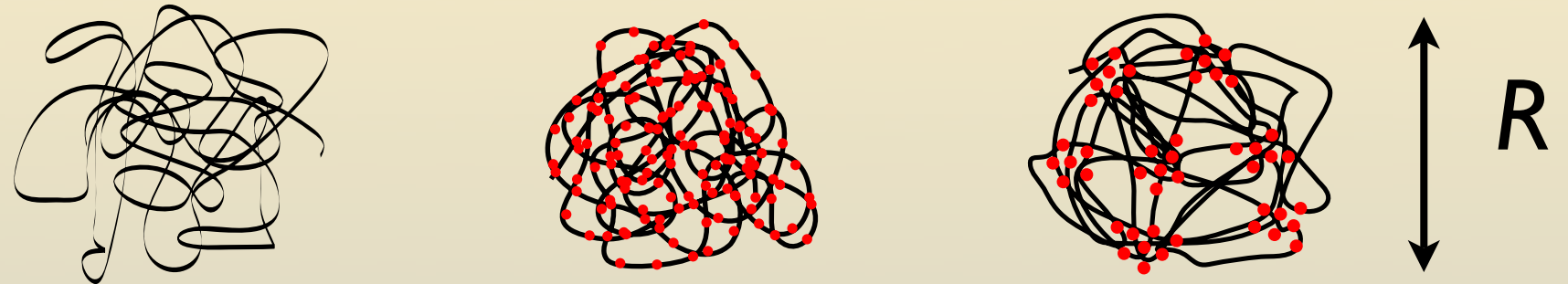
Self-attracting polymer

Sparse interacting sites



Calculation principles (I)

- Comparing free energies



- Saddle point approximation

$$F(R) \rightarrow R^* \quad \partial_R F|_{R^*} = 0$$
$$F^* = F(R^*)$$

$$F(R) = F_a(R) + F_b(R) + F_{ev}(R) + F_s(R)$$

Attractive term Bending cost Excluded volume Configurational entropy

- Flory theory : mean-field like $\langle c_i c_j \rangle = \langle c \rangle^2$

Stability results

l_p Persistence length

$\bar{n}$ Max. number of binding sites

V_0 Protein-DNA binding free energy

r_0 Polymer radius

Δ

Mean distance between two successive interacting sites along the DNA

L Contour length

R Gyration radius

Swollen vs. globular state

$$\frac{\bar{n}V_0}{k_B T} \times \frac{\Delta}{l_p} \times \left(\frac{r_0}{\Delta}\right)^3$$

< 1 swollen

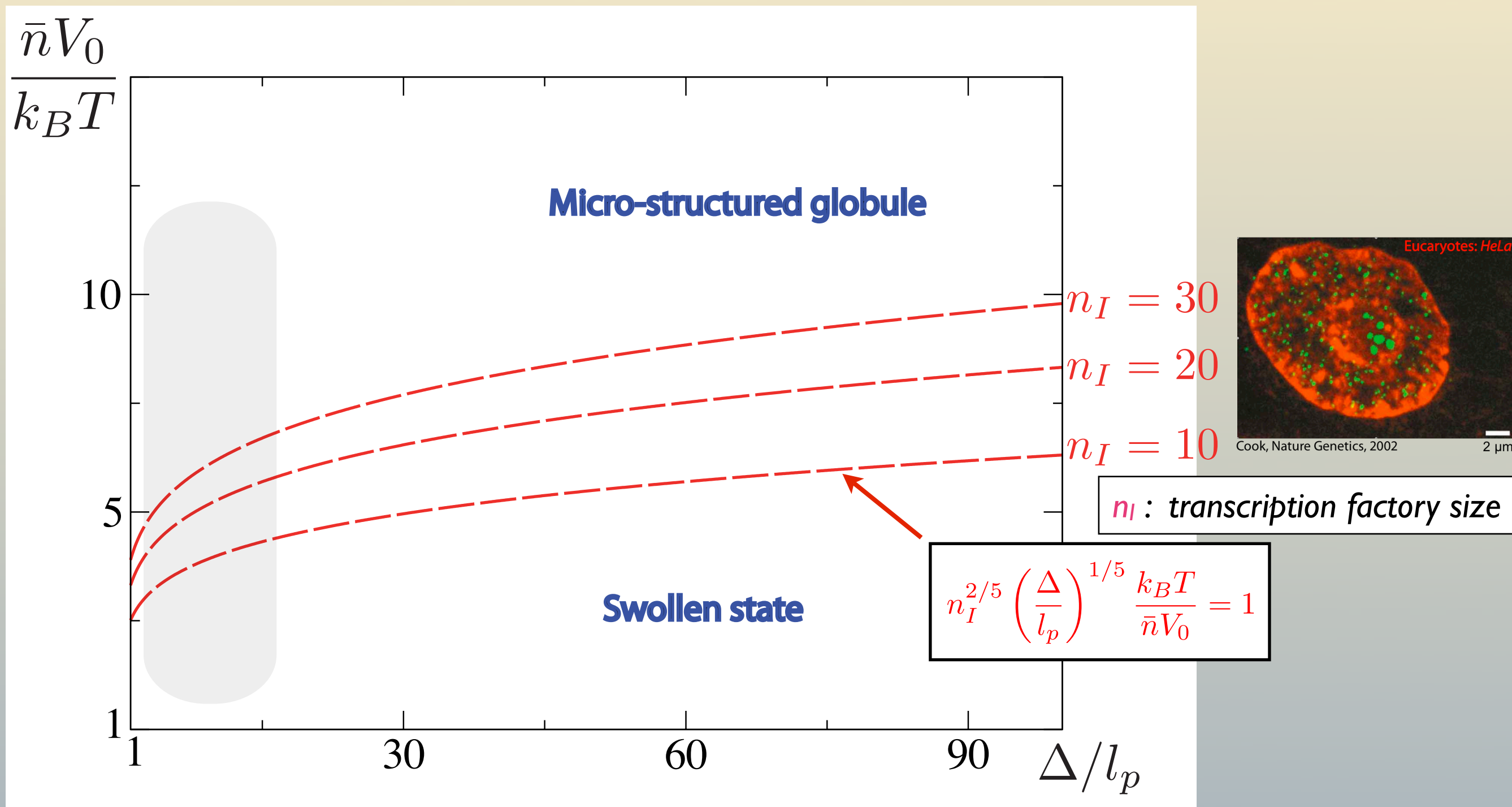
> 1 compact

Swollen vs. heterogeneous

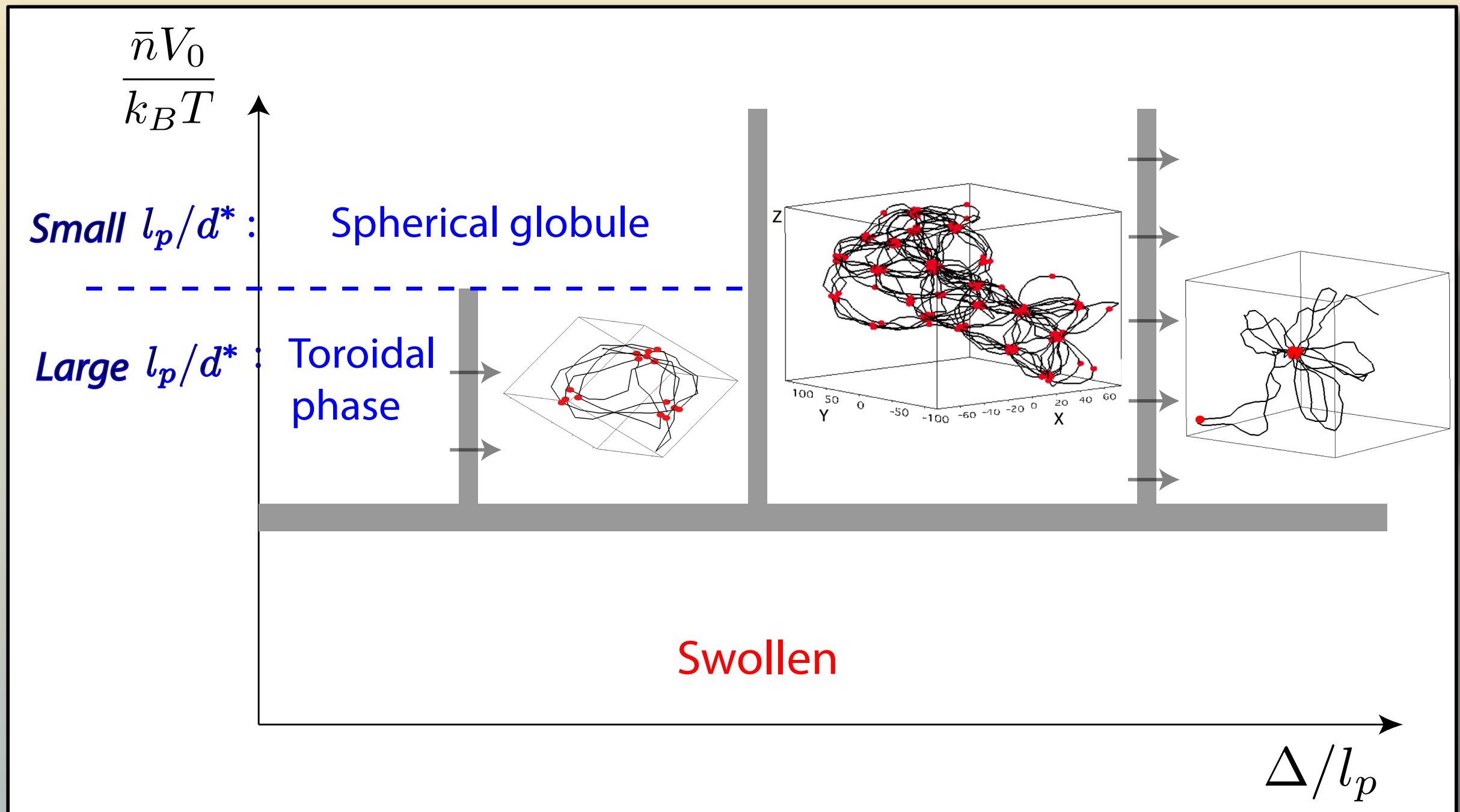
n_I Nb of sites per factory

$$n_I^{2/5} \left(\frac{\Delta}{l_p}\right)^{1/5} \frac{k_B T}{\bar{n}V_0}$$

Transcription factors are good candidates for mediating the interaction



Numerics: Topological ordering around the foci



Part II

- Functional organization of bacterial genomes is partly driven by the spatial proximity of co-functional genes
- The easiest (i.e. = most well-known) way : 1D clustering along DNA
- Long-range strategies?

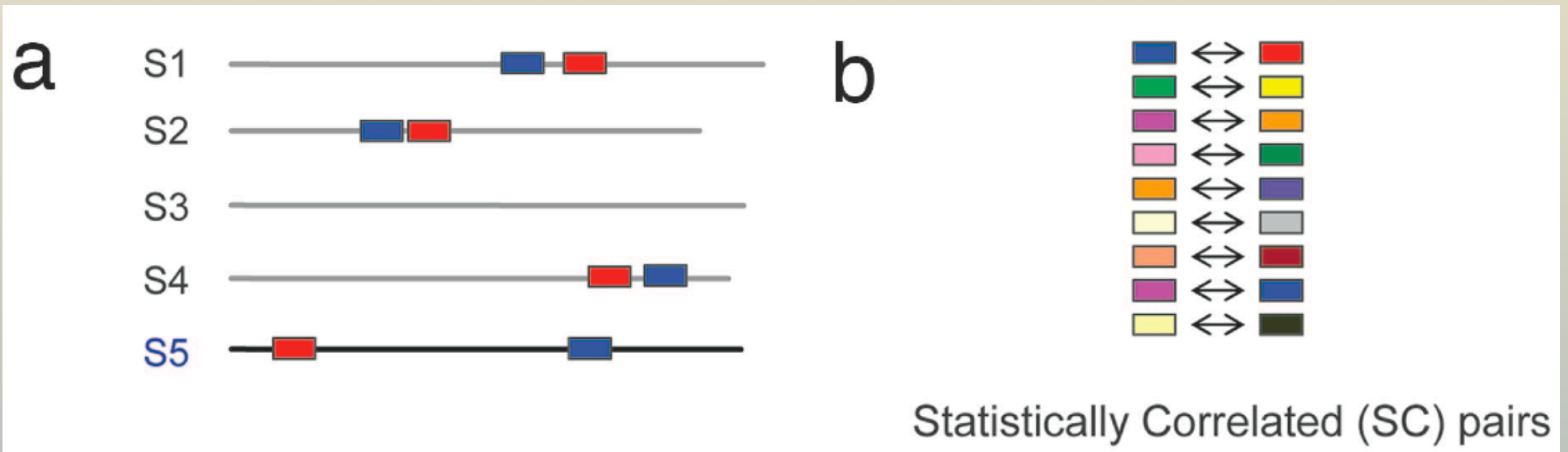
Dataset: statistically correlated genes in bacteria

E. coli : statistically correlated genes and their relative positions

Phylogenetic co-occurrence : $P < 10^{-10}$

Chromosomal proximity : $P < 10^{-4}$

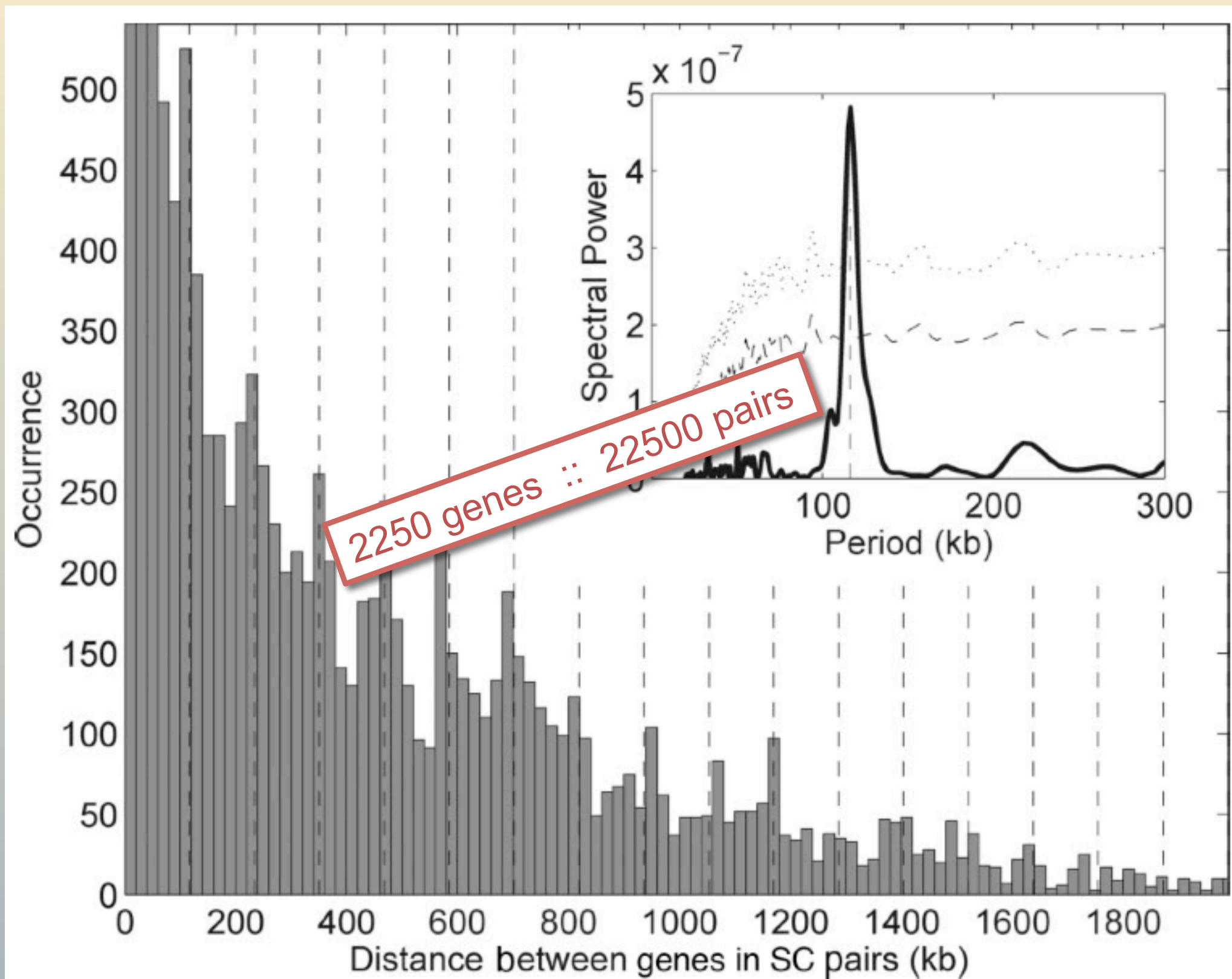
~ 100 bacterial genomes



Wright et al., PNAS, 104, 10559 –10564, 2007

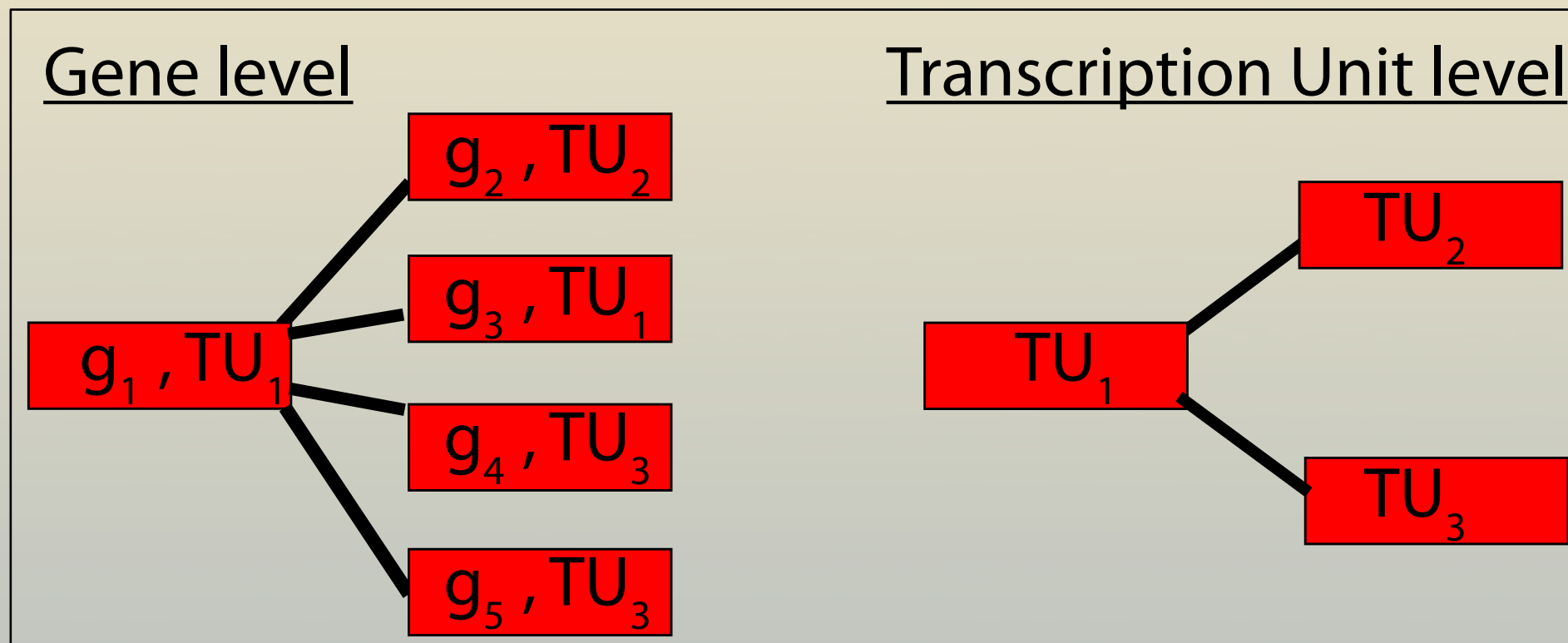
Statistically correlated pairs = functional links (likely)

Periodic spacing of distant correlated genes



Refinement of the analysis

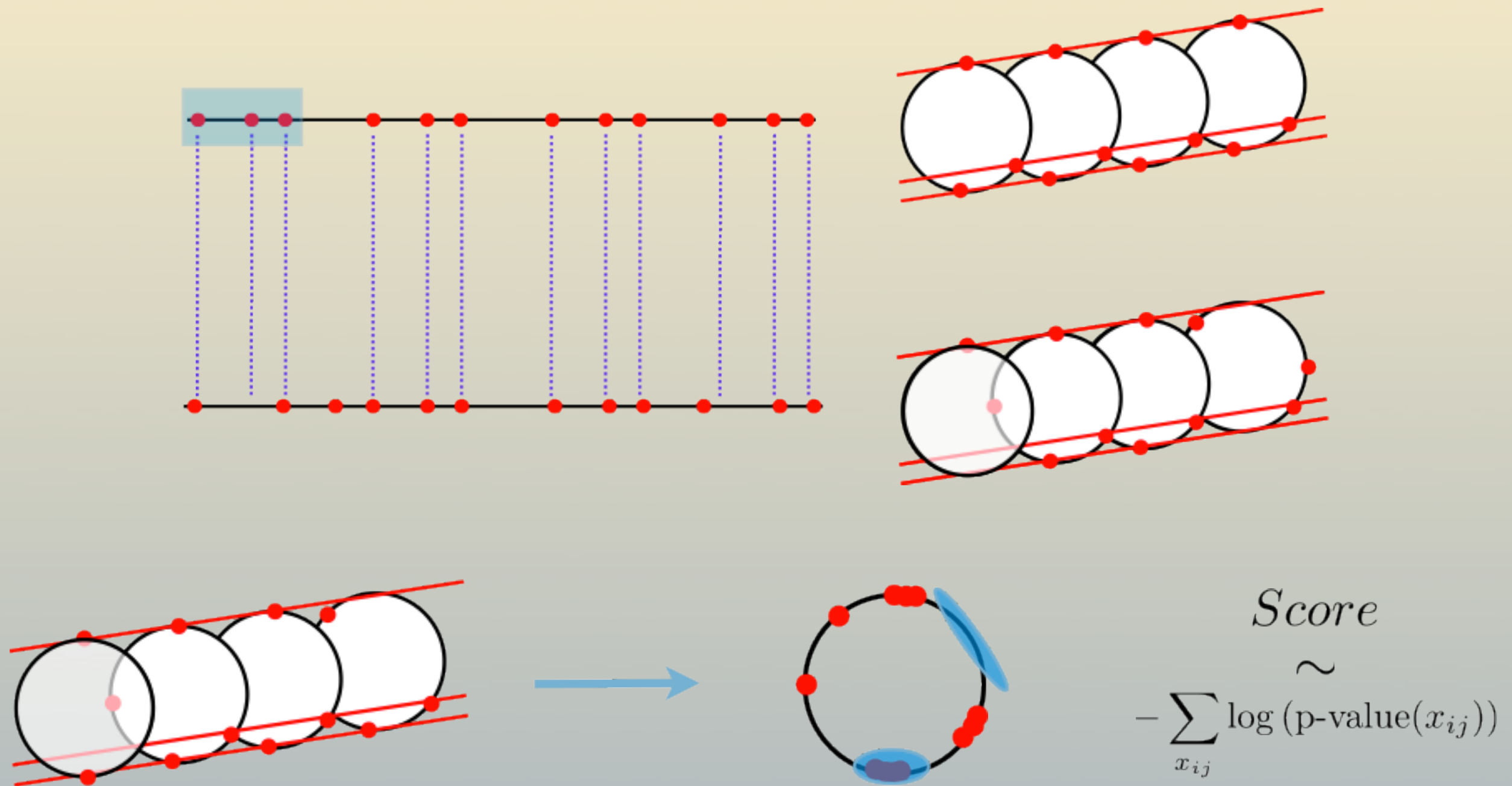
Groups of correlated **T**ranscription **U**nits



TU Group sizes: **1 to 129**

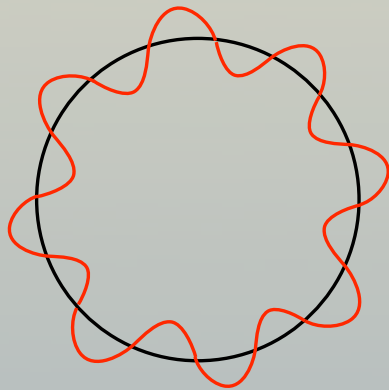
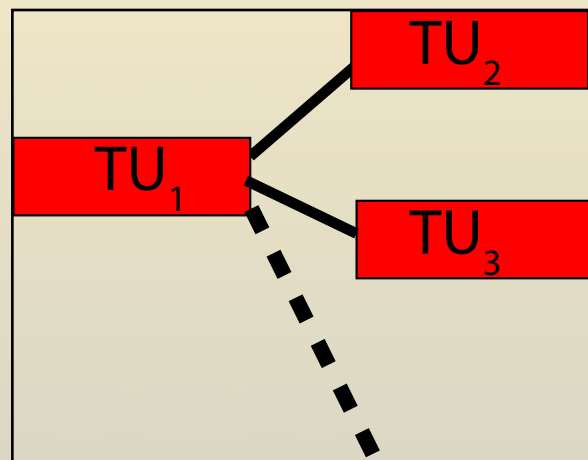
A solution to the periodicity detection problem

The solenoid coordinate method

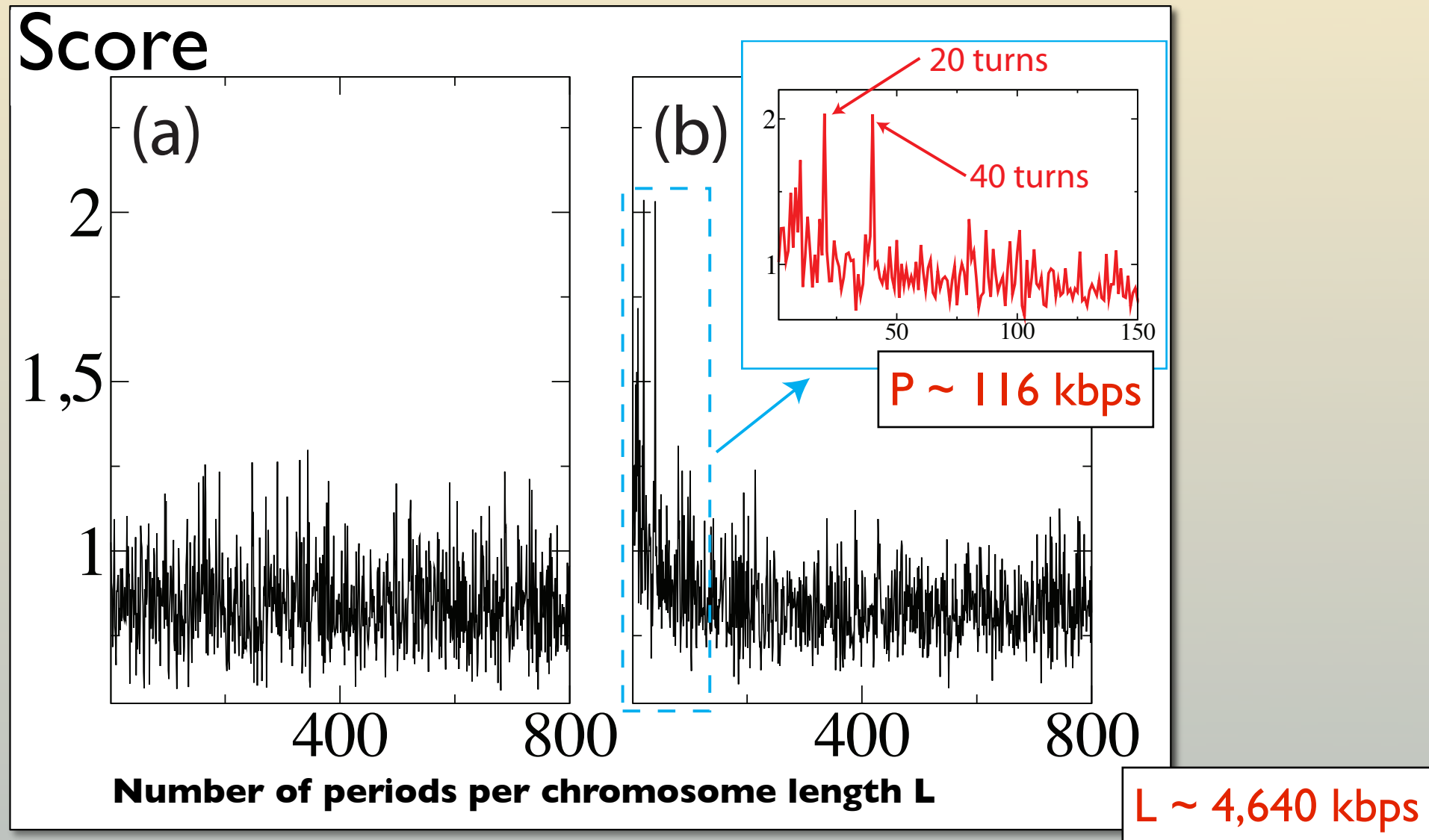


Different periods $\Rightarrow$ **spectrum**

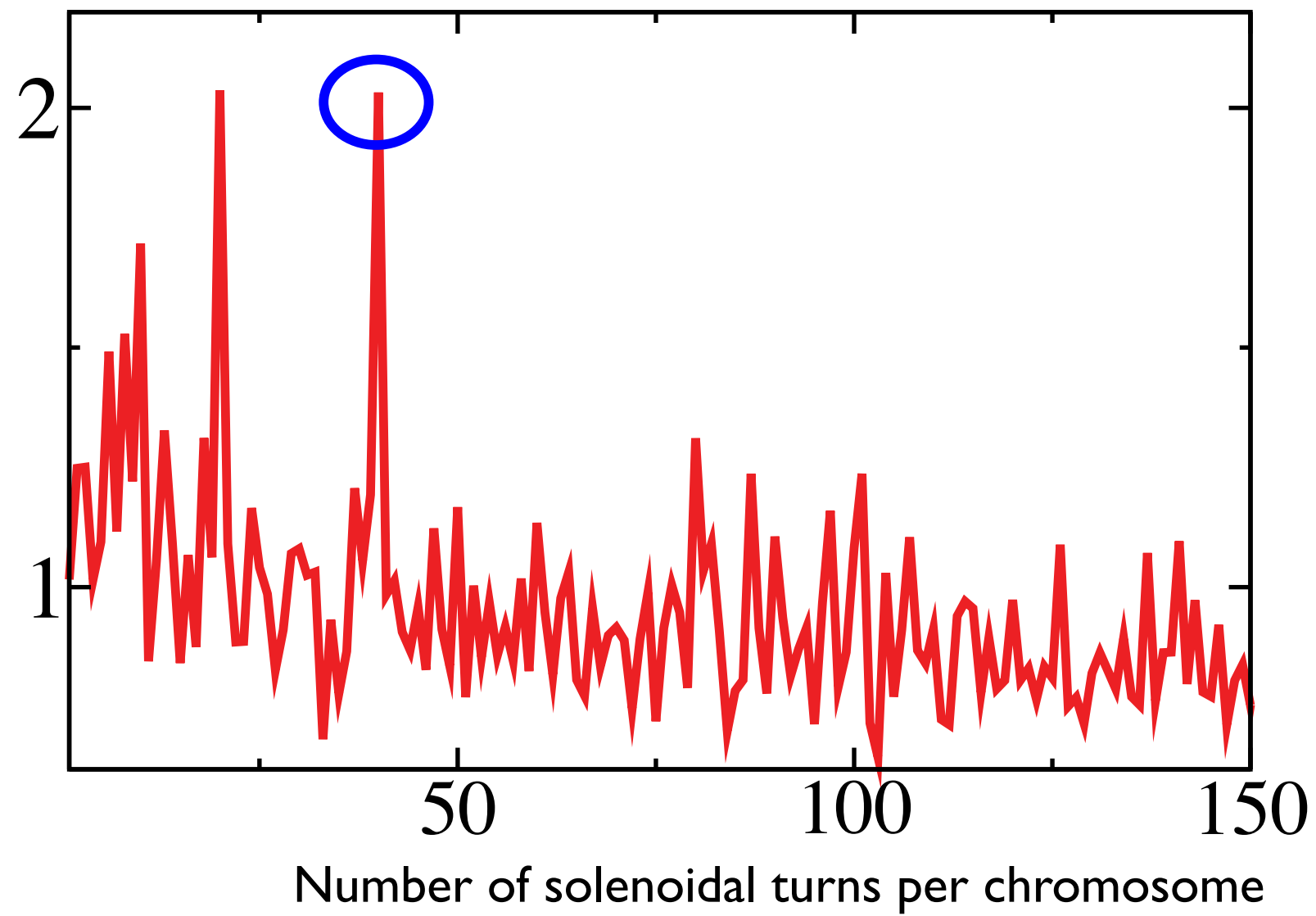
Groups of statistically correlated TUs



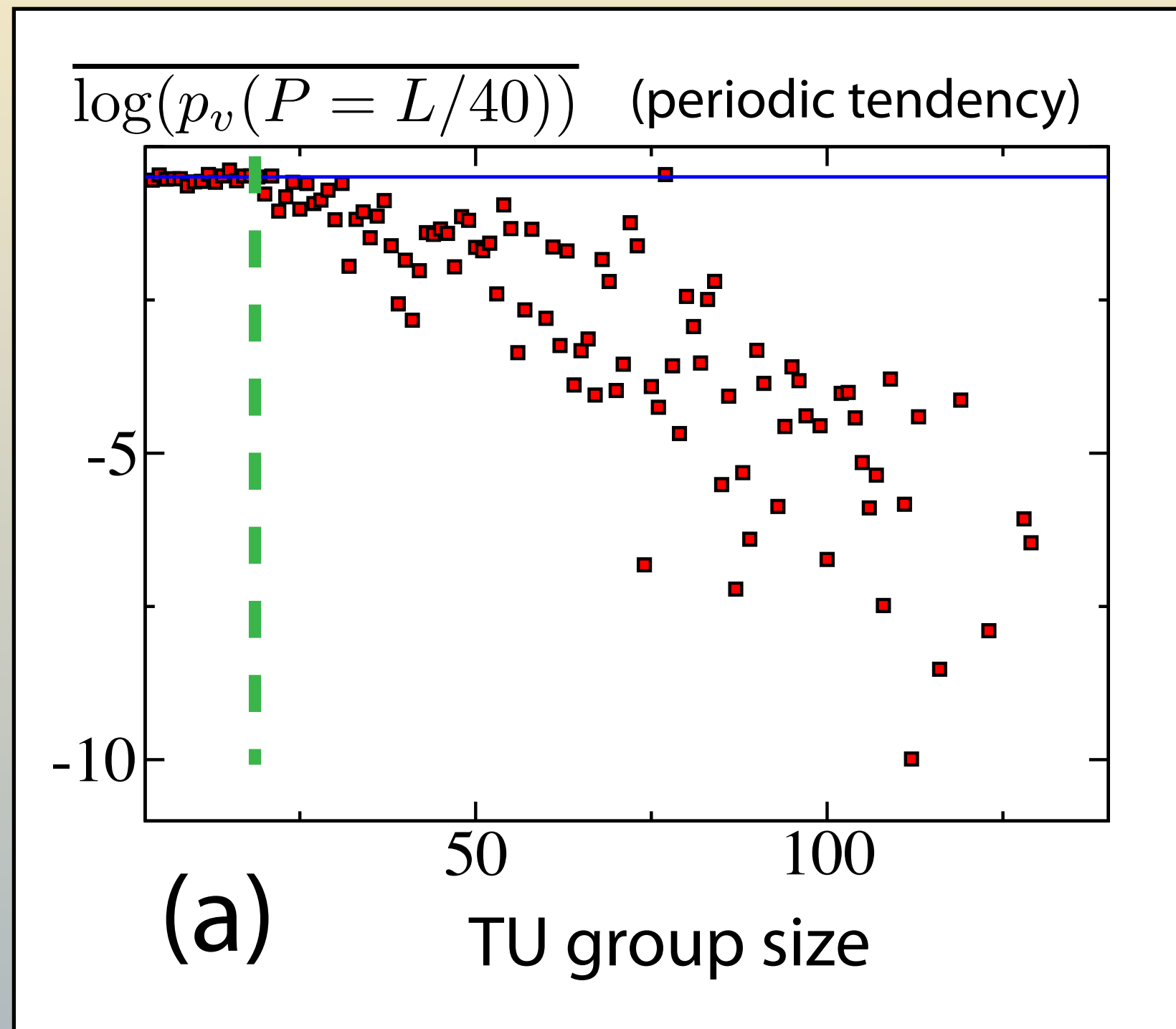
Example: 8 periods



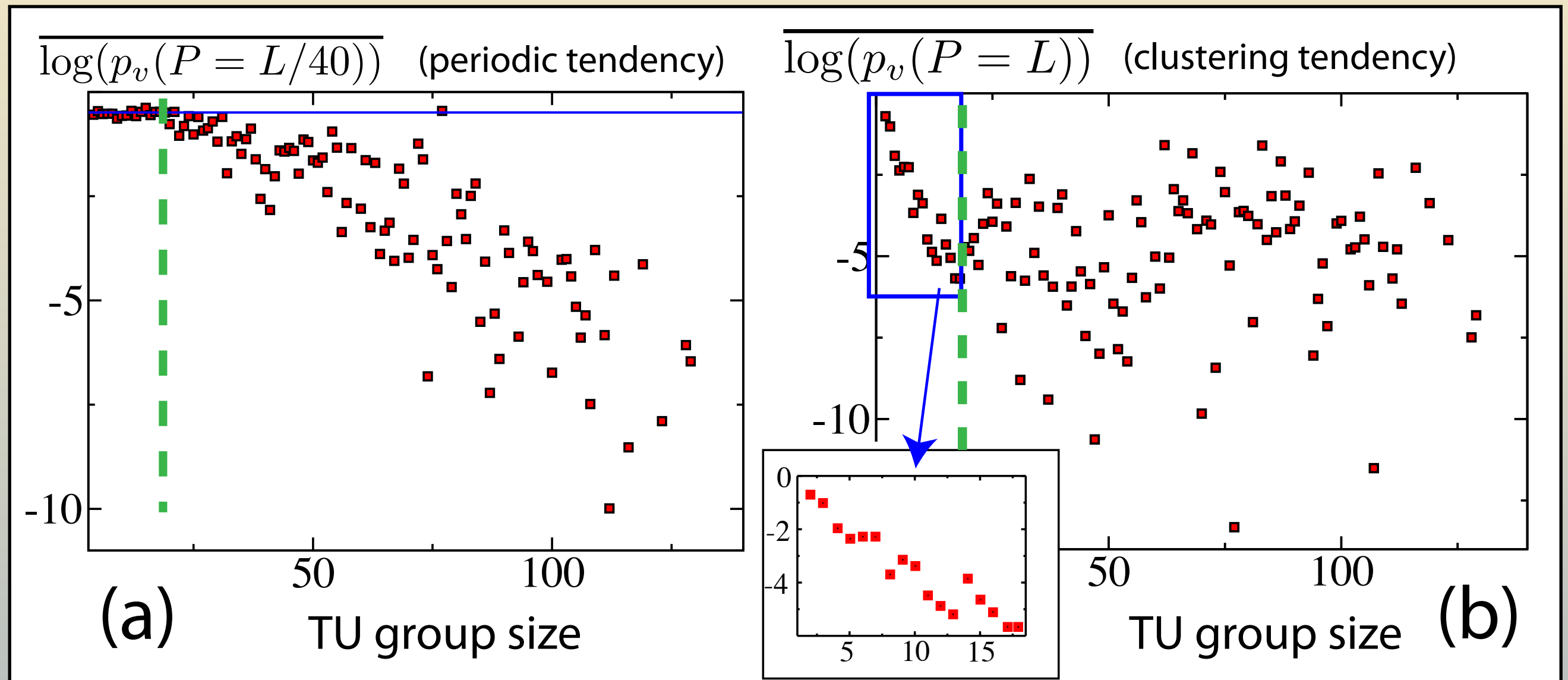
Score



Periodic tendency of correlated TUs

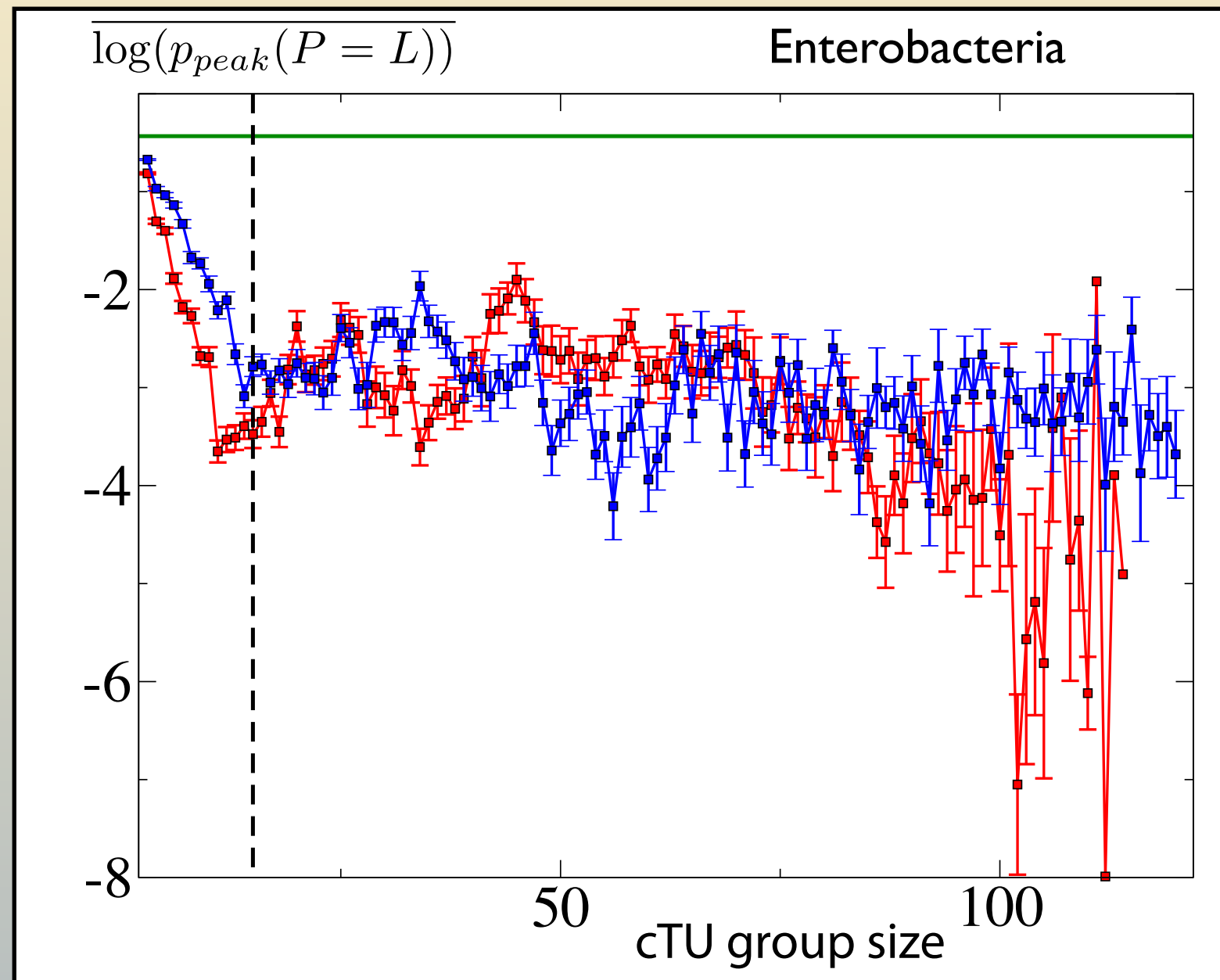


Genomic organization of correlated TUs: from 1D to 3D



Transition size ~ 15 TUs ~ 18 genes

Genomic organization of correlated TUs: from 1D to 3D



1D $\rightarrow$ periodic spacing

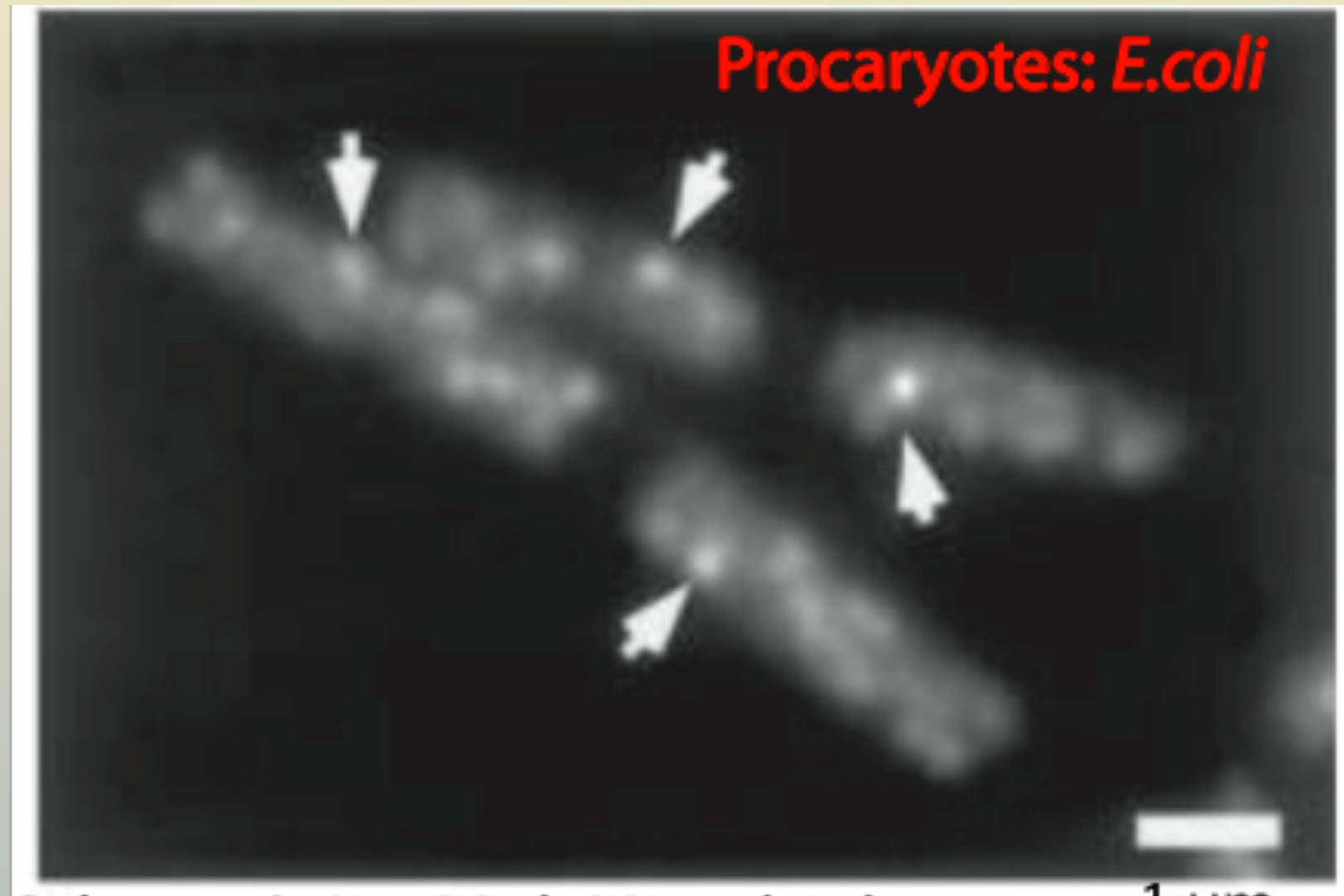
=

1D clustering (e.g. micro-domain of superhelicity)



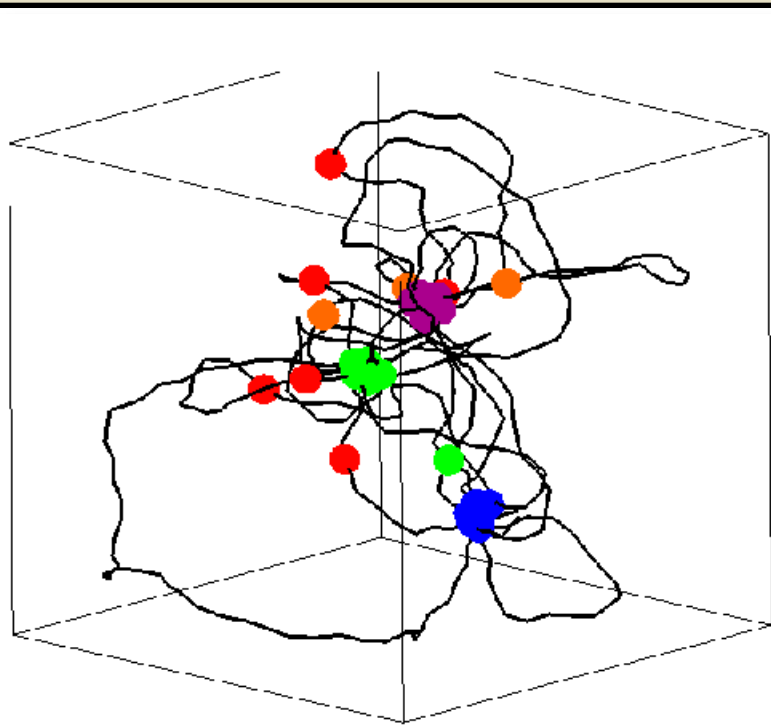
3D clustering (e.g. transcription factories)

rRNAs are transcribed
within transcription factories in *E. coli*

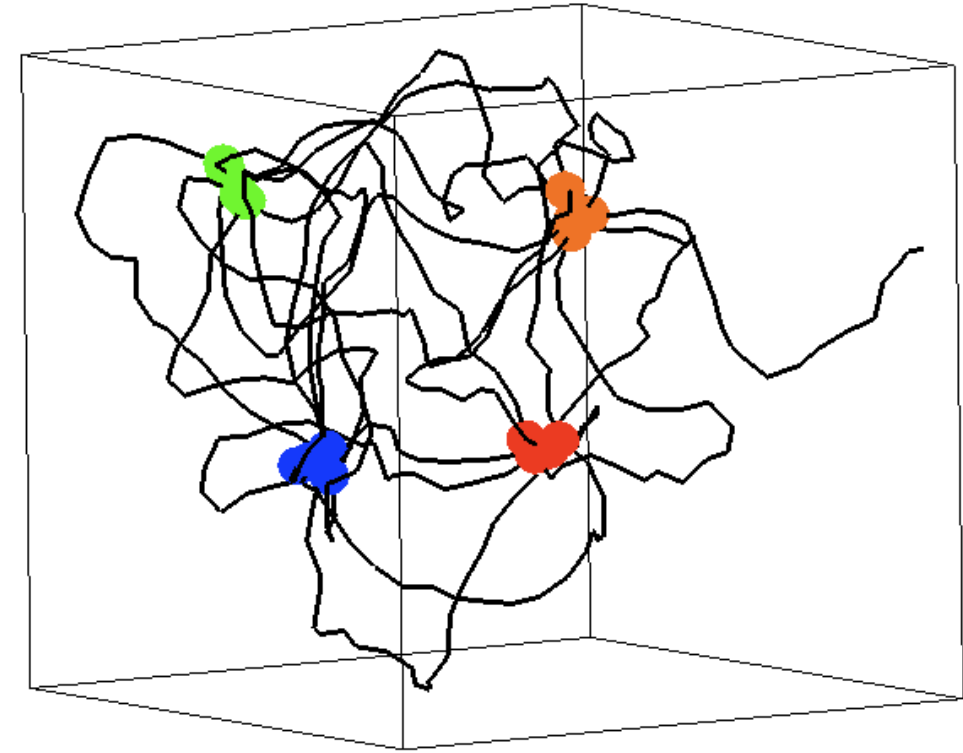
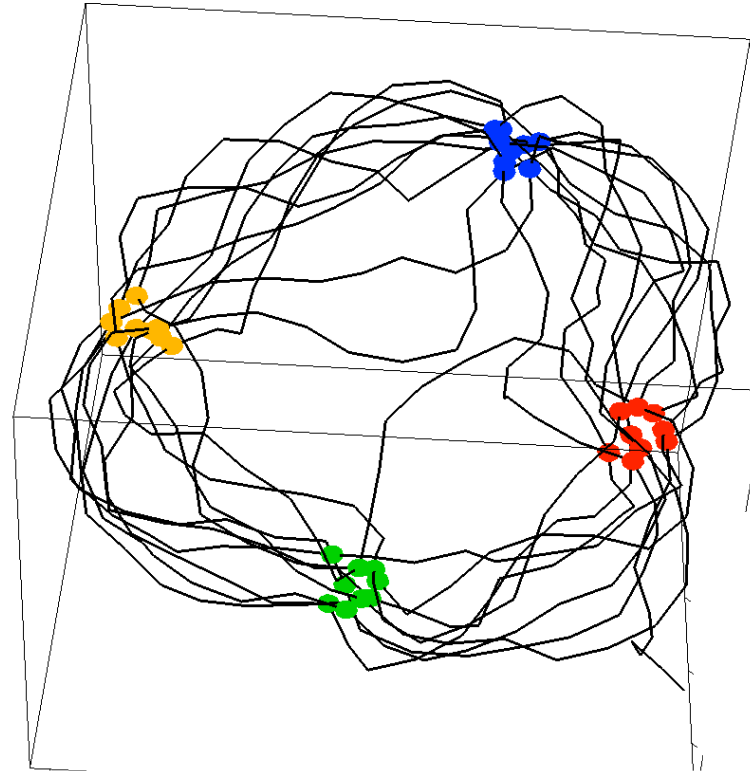


Cabrera & Jin, Mol. Microbiol., 2003 1 μm

Specialized transcription factories: the importance of genome organization



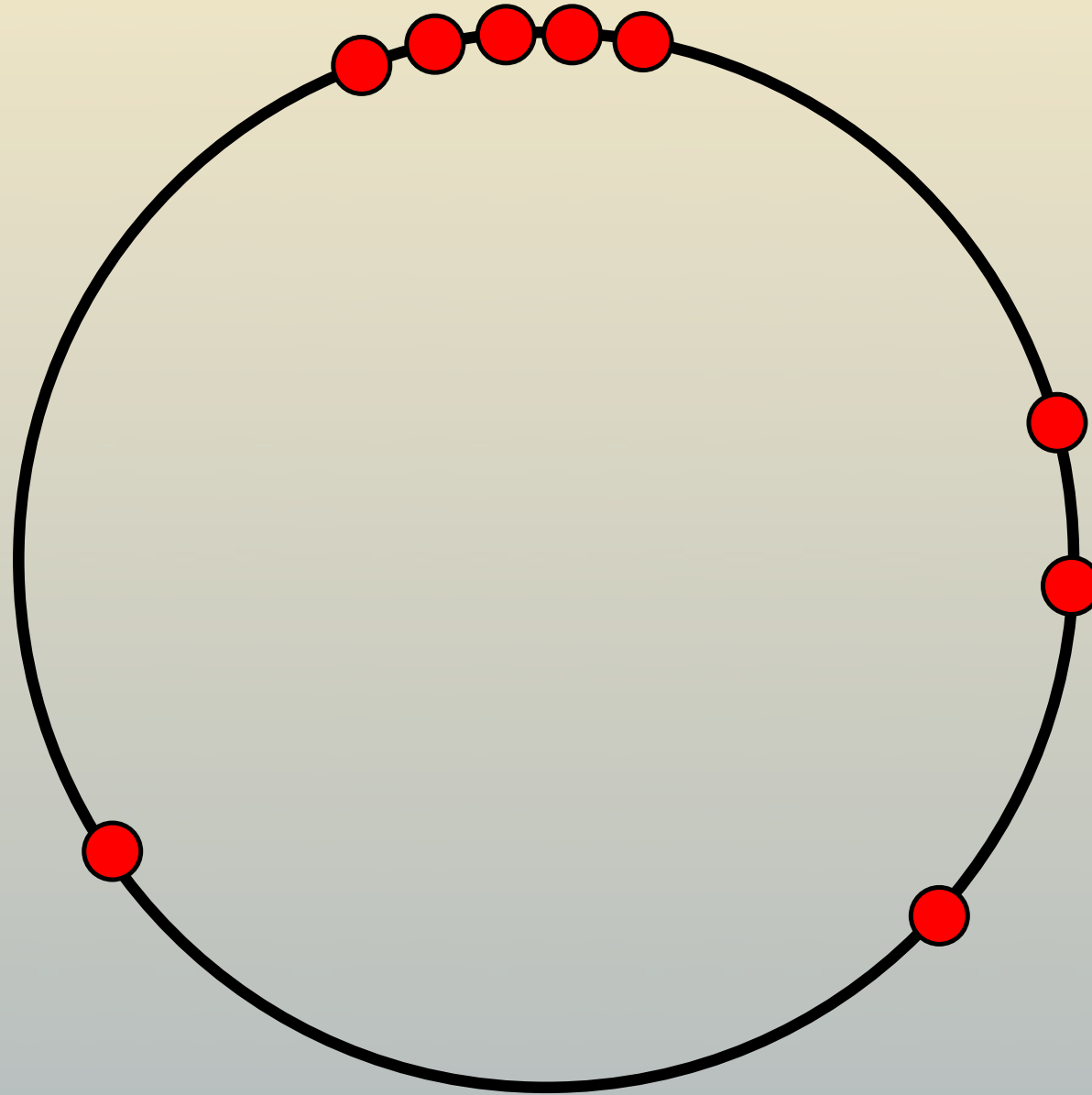
Random site positions



(Long-range) periodic site positions

Identifying periodic elements

$$\text{Score}(i) = - \min_{j \in n.n.} \{ \log(p_{<}(x_{ij})) \}$$



PTU genes are periodic in all phyla

Strain	Genome length	Period
Acidobacterium_capsulatum_ATCC_51196	4127356	687892
Acidovorax_JS42	4448856	193428
Acidovorax_avenae_citrulli_AAC00-1	5352772	223032
Acinetobacter_baumannii_AYE	3936291	34528
Aeromonas_hydrophila_ATCC_7966	4744448	474444
Aeromonas_salmonicida_A449	4702402	204452
Aggregatibacter_aphrophilus_NJ8700	2313035	192752
Agrobacterium_radiobacter_K84	4005130	174136
Alteromonas_macleodii_Deep_ecotype_	4412282	275767
Bacillus_halodurans	4202352	420235
Bacteroides_thetaiotaomicron_VPI-5482	6260361	32606
Bacteroides_vulgatus_ATCC_8482	5163189	258159
Bifidobacterium_longum_infantis_ATCC_15697	2832748	708187
Bordetella_bronchiseptica	5339179	889863
Bordetella_parapertussis	4773551	140398
Bradyrhizobium_BTAl	8264687	113214
Bradyrhizobium_OR5278	7456587	1065226
Burkholderia_383	1395069	17659
Burkholderia_ambifaria_MC40_6	2769414	197815
Burkholderia_cenocepacia_MC0_3	3213911	189053

Periodic TUs (PTU's) : functional bias

205 PTU's (of 1303)

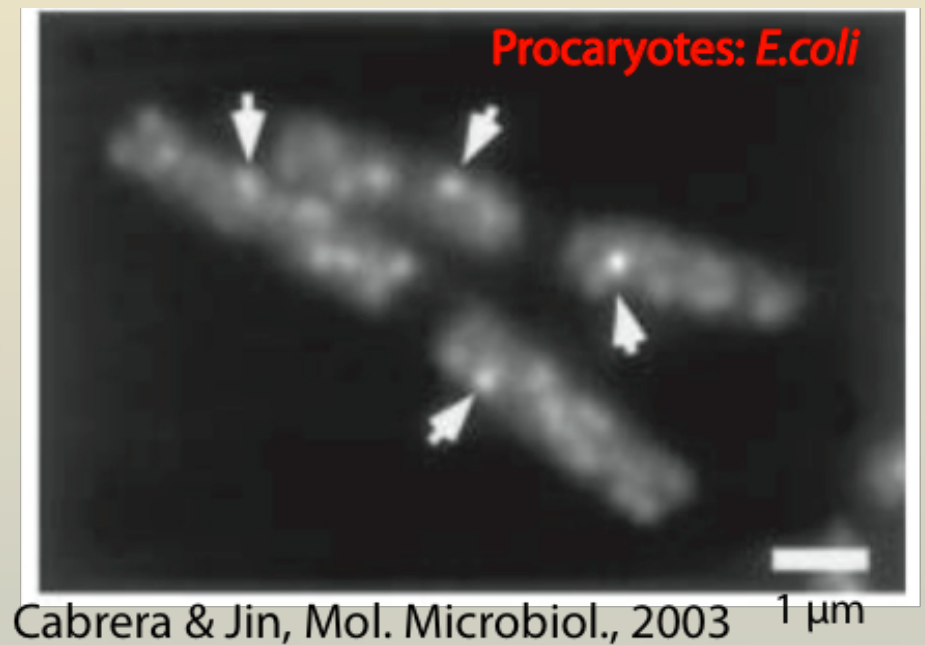
41/54 cons. **Ribosomal proteins** ($P < 10^{-18}$)

20/38 cons. **Aminoacyl tRNA synthetase** ($P < 10^{-6}$)

7/9 cons. Subunits of the **DNApol III** ($P < 2 \times 10^{-4}$)

5/7 cons. Subunits of **RNApol** ($P < 3 \times 10^{-3}$)

1/2 **elongation factors**



GO annotation:

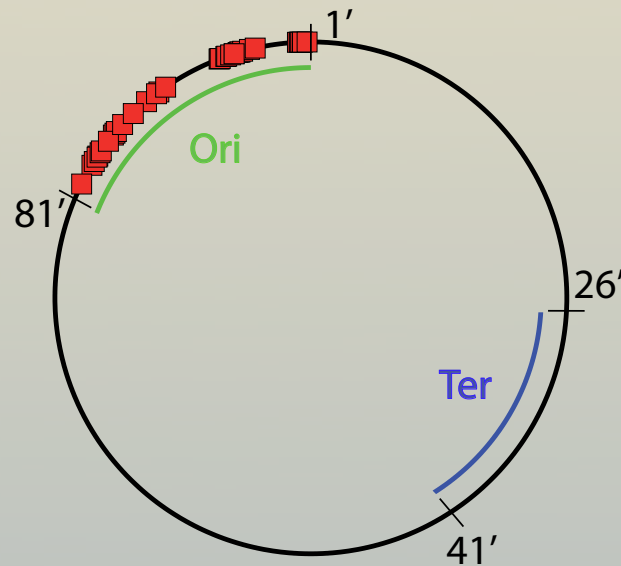
28/41 cons. "regulation and cellular component organization" ($P \approx 2 \times 10^{-7}$)

12/13 cons. "cellular protein complex disassembly" ($P \approx 2 \times 10^{-4}$)

22/39 cons. "establishment of localization in cell" ($P \approx 3 \times 10^{-3}$)

Conclusion

- Transcription factories : self-organized process where genome organization plays a crucial role
- ID clustering limited to 15 TUs (18 genes) in enterobacteria
- Above 15 TUs : different large-scale strategies related to the chromosome folding
- **Macrodomains**



Espéli, Boccard
Journal of Structural Biology (2006)

Valens et al.
EMBO J (2004)

- **Periodicity (< 10 % of the genome) ⇒ dynamically contributes to the overall organization of the cell.**

Llopis et al., Spatial organization of the flow of genetic information in bacteria.

I. Junier, J. Hérisson, F. Képès, under review