

# Gene position scoring within transcription regulation networks

Ivan Junier, Joan Hérisson, Mohamed Elati, François Képès

Programme d'Épigénomique, Évry, France



# Outline

- \* Why positions?

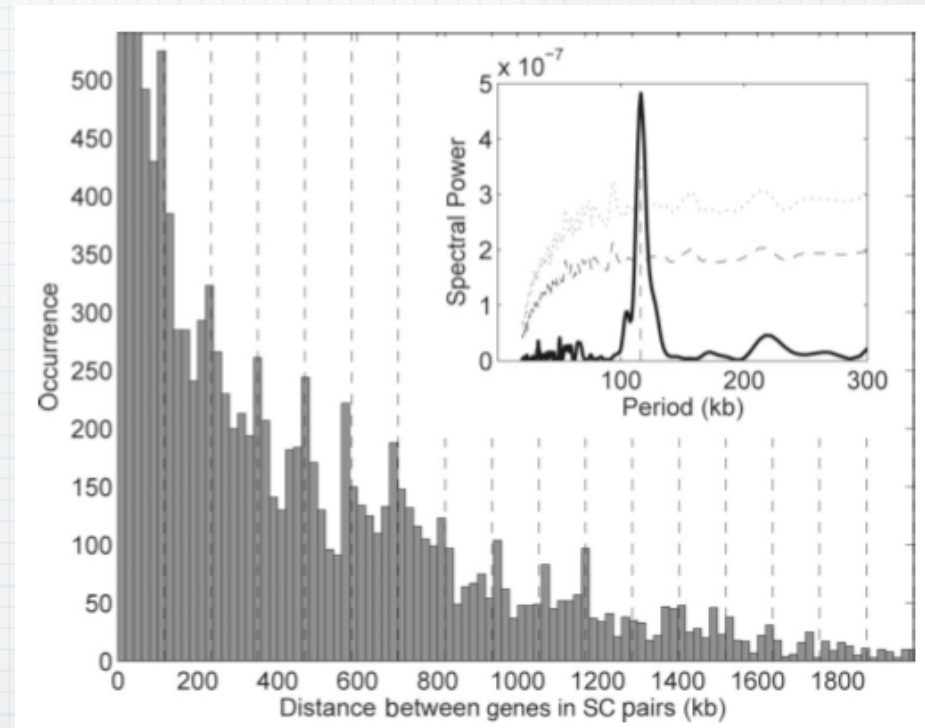
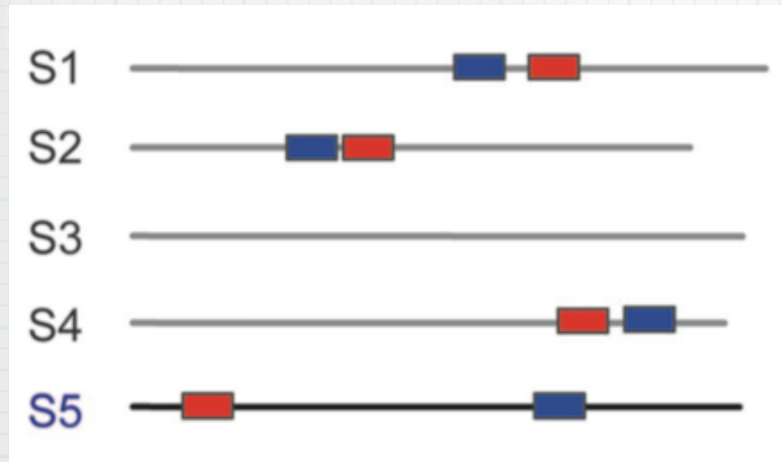
- \* How to score?

- \* Which outcome?

# E. coli : conserved paired genes and their relative position

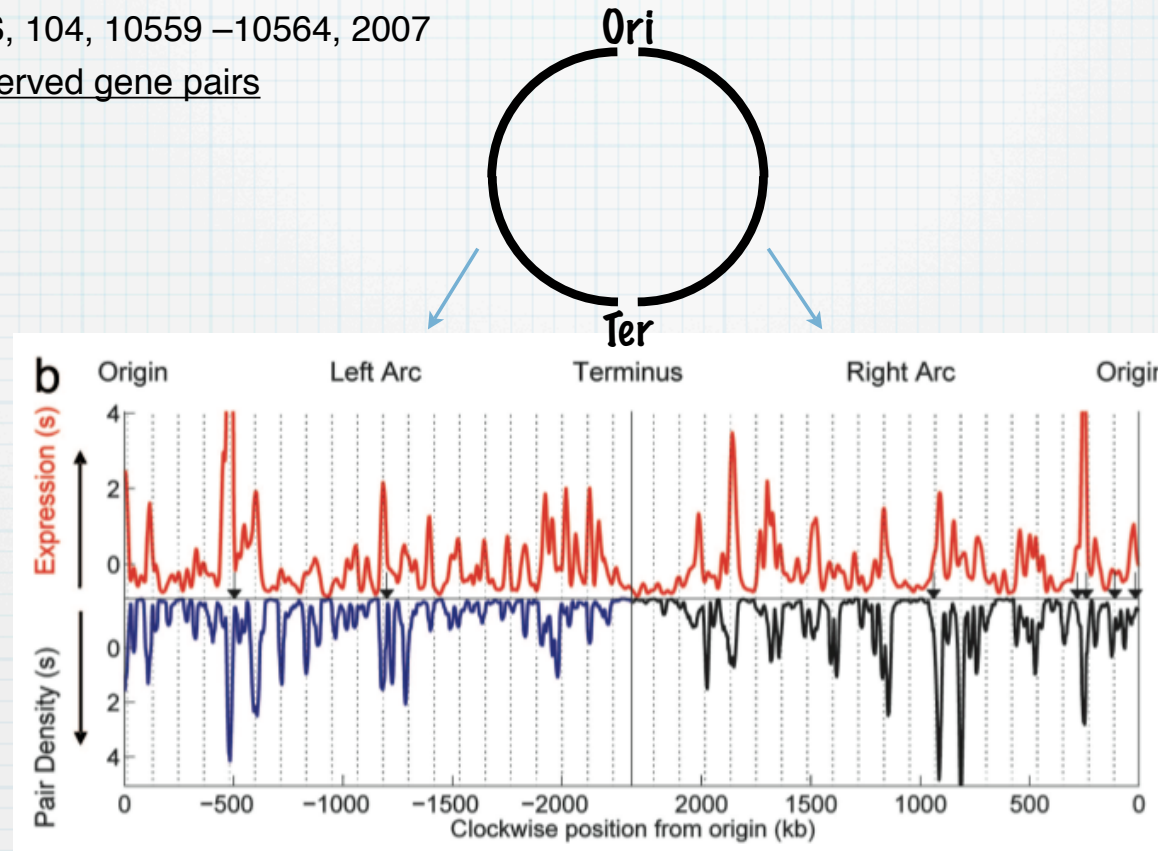
Wright et al., PNAS, 104, 10559 –10564, 2007

Evolutionarily conserved gene pairs



# E. coli : conserved paired genes and their relative position

Wright et al., PNAS, 104, 10559 –10564, 2007  
Evolutionarily conserved gene pairs

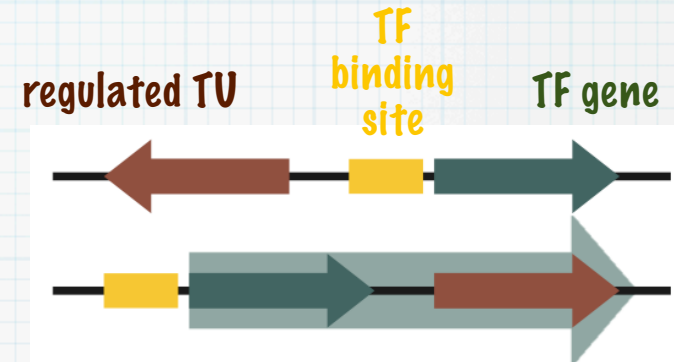


Gene position  $\leftrightarrow$  Transcription regulation

## Co-regulation and gene position

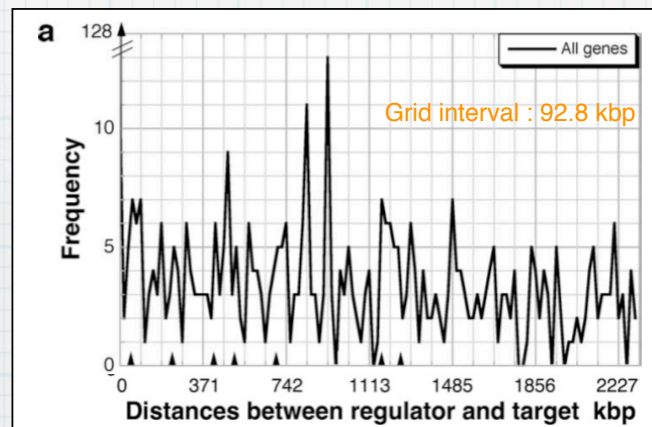
- \* 1D co-regulation is frequent in prokaryotes

- \* Rapid search hypothesis



- \* Periodic positioning of genes regulated by the same TFs

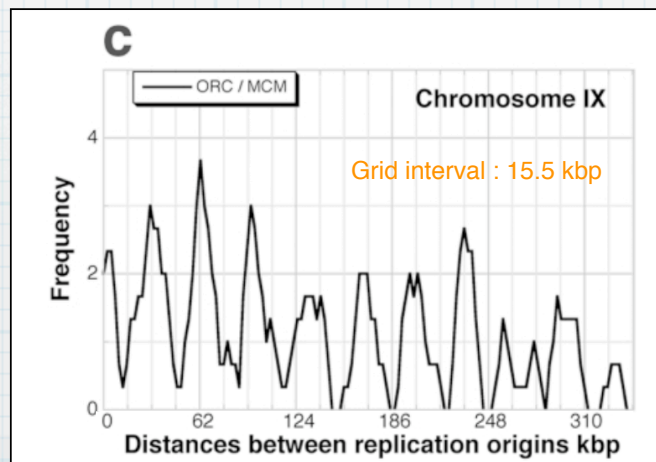
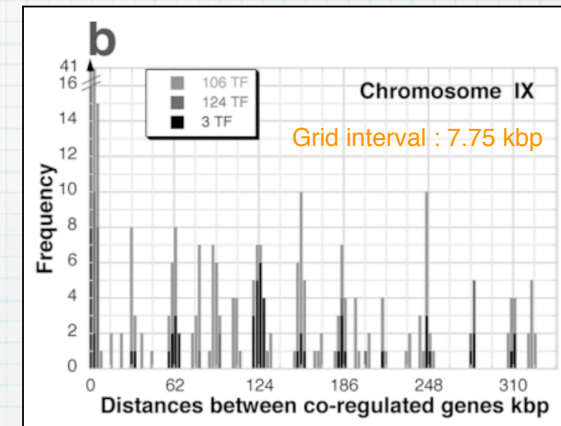
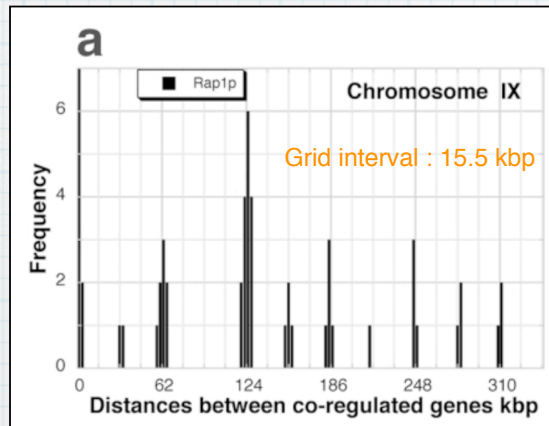
- \* 3D co-localization --> rapid search hypothesis



Képès, JMB, 2004

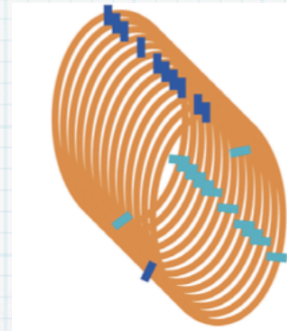
# Co-regulation and gene position

\* Periodic positioning of genes regulated by the same TFs in yeast



# Why positions are important? Spatial co-localization

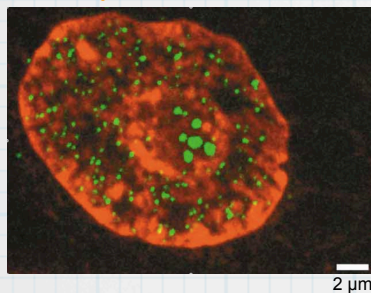
## \* Conceptual framework



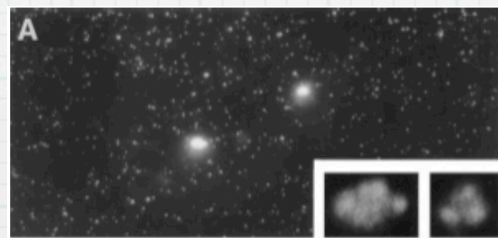
Képès, Vaillant, ComplexUS, 2003

## \* Experimental facts

### Eukaryotic nucleus

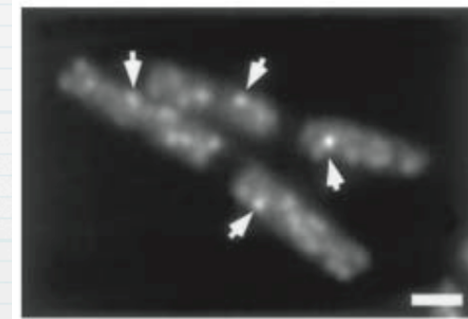


Cook et. al, Nature, 2002



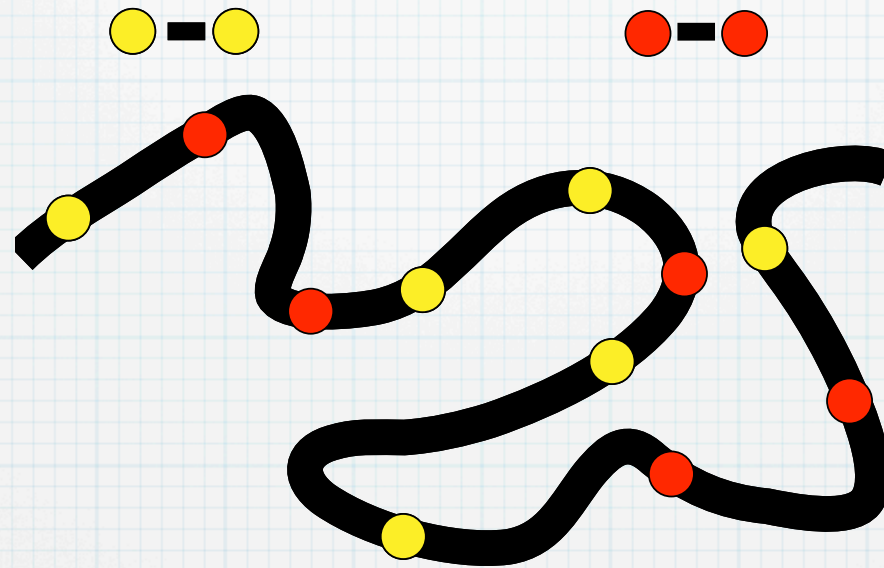
Jackson et al., Mol. Biol. Cell, 1998

### Bacterial cells



Cabrera, Jin, JMB, 2004 1 μm

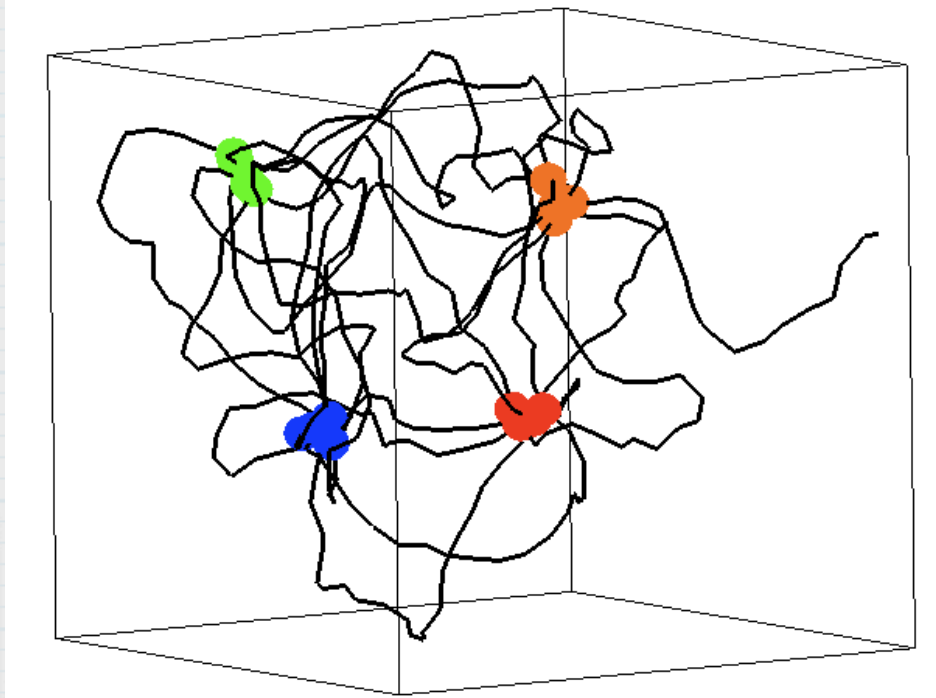
# Why positions are important? Polymer theory



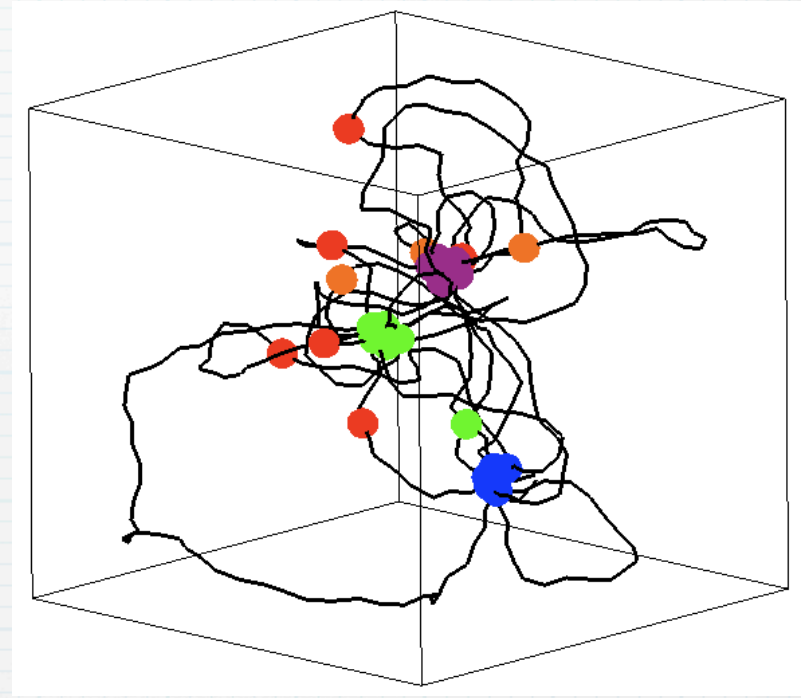
I. Junier, O. Martin, F. Képès, submitted to  
Biophys. J.



# Why positions are important? Polymer theory



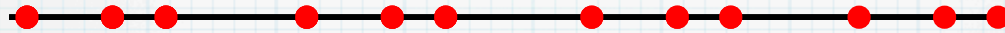
**Periodic**



**Random**

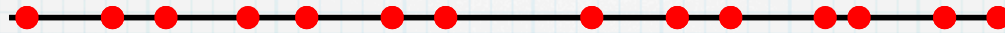
# How to detect periodicity?

\* 1) What is periodic?



\* 2) Noise

\* Genes out of the periodicity / False positives

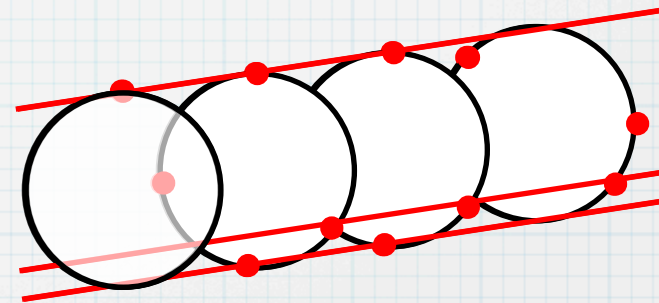
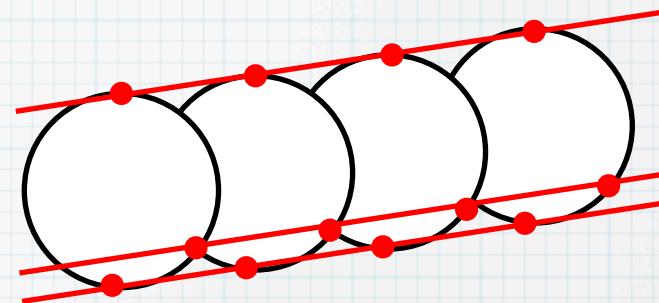
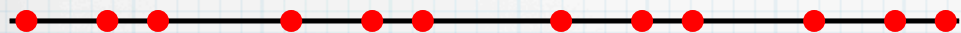


\* Blank sites / False negatives

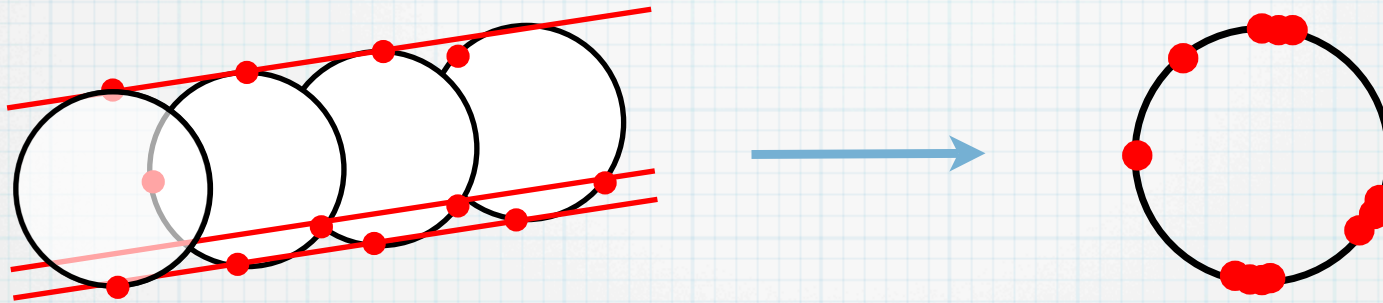


\* + fluctuations around the original sites

# Solenoidal framework



# Periodicity detection = clustering detection

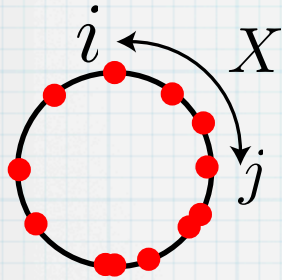


Principle :



Play with the period : **spectrum** (score vs. period)

# Statistics of circular distributions



**Binomial:**

$$\rho_{N,|j-i|}(X = x) = \mathcal{C}_{N-1}^{|j-i|} x^{|j-i|-1} (1-x)^{N-|j-i|-2}$$

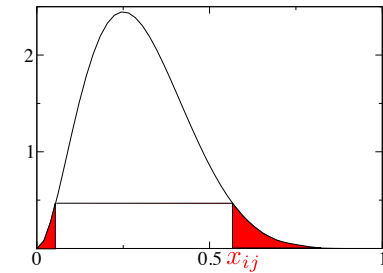
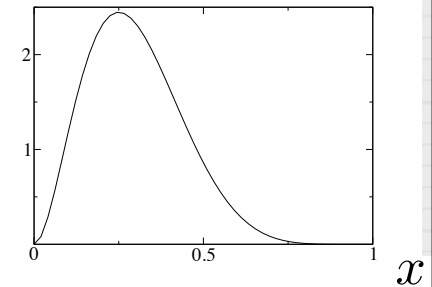
**Pair score:**

$$s(x_{ij}) = -\log[p_v(x_{ij} | \rho_{N,|j-i|})]$$

**Gene score (sum up the n first neighbors):**

$$S_i(\{x_{ij}\}) = \frac{1}{n} \sum_{j=(i+1)\%N}^{(i+n)\%N} s(x_{ij})$$

$\rho_{11,3}(x)$

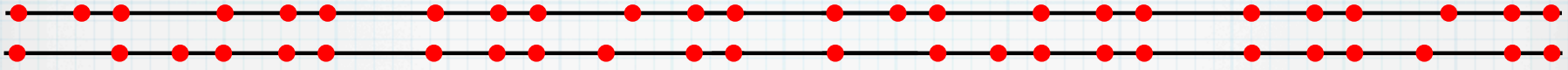


## Final score

$$\mathcal{S}(\{x\}) = \frac{1}{N} \sum_i S_i(\{x_{ij}\})$$

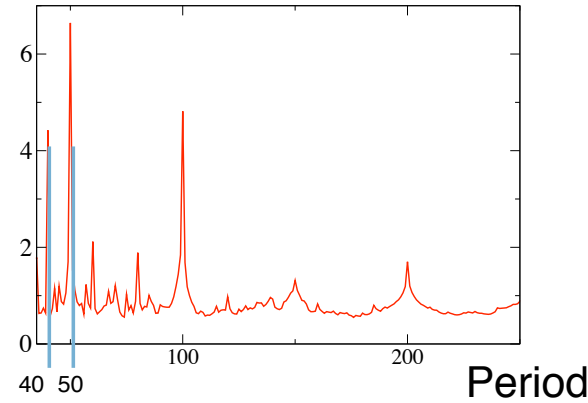
# Exemple

0 40 55 100 ...



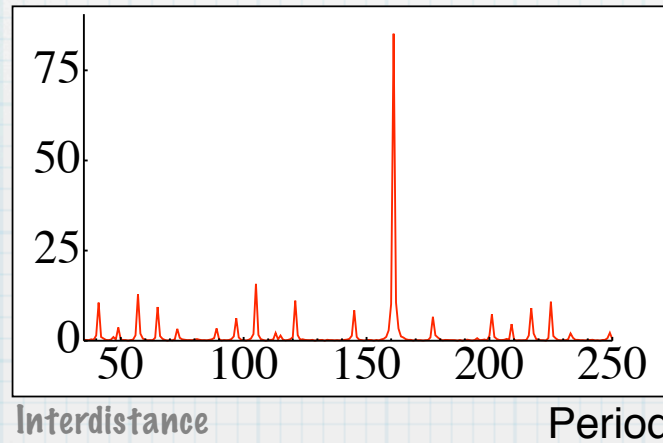
Clustering spectrum

Score

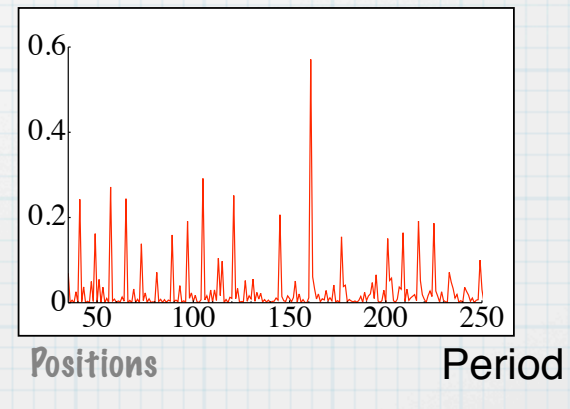


Discrete Fourier spectrum

Amplitude

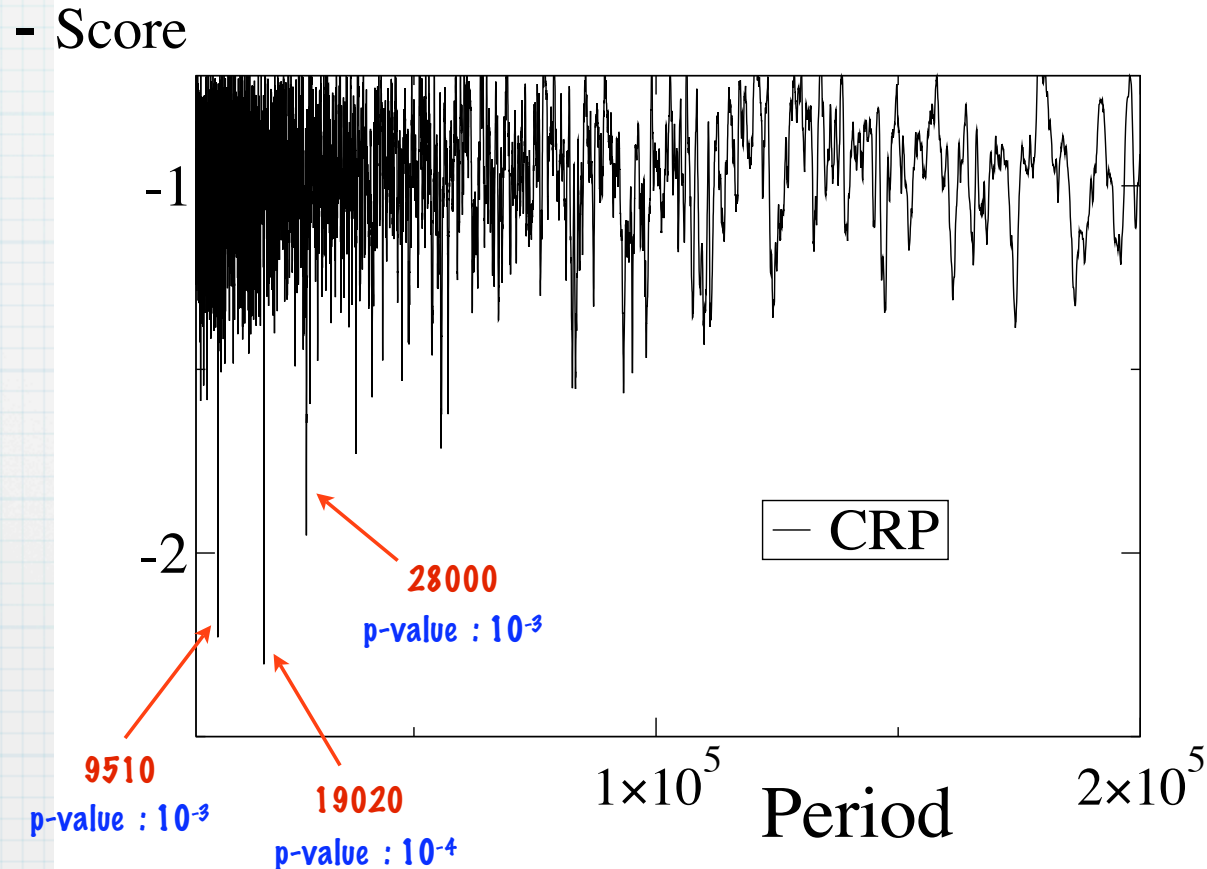


Amplitude



# Some results in E. coli

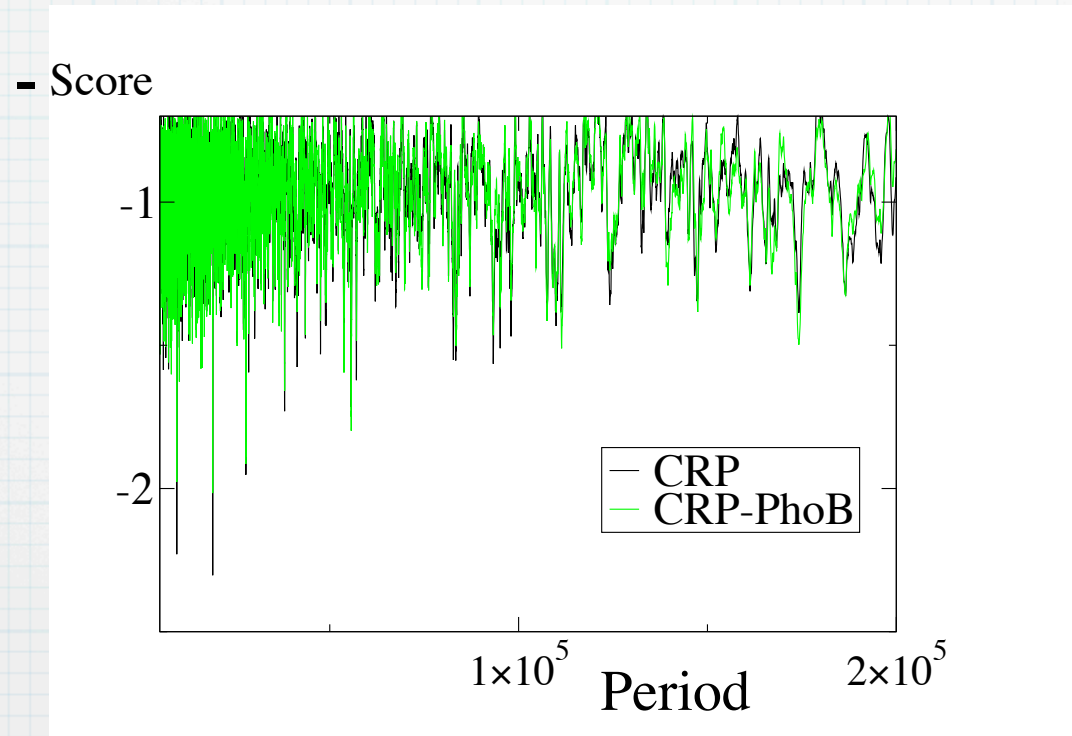
CRP binding sites (RegulonDB 2008) : 160 targets (~450 genes) --> 90 strong evidence



Unveiling chromosomal structures

# Some results in E. coli

CRP binding sites (RegulonDB 2008) : 160 targets (~450 genes) --> 90 strong evidence



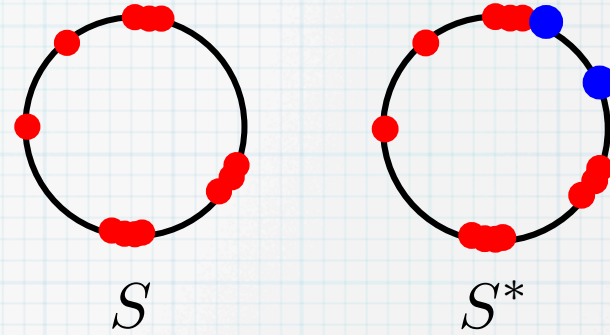
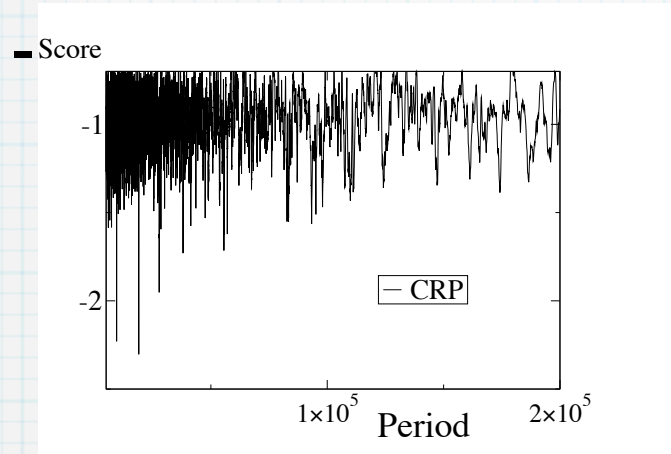
Unveiling functional relation between TFs : inference of transcription regulation

J. Hérisson, I. Junier, F. Képès, in preparation



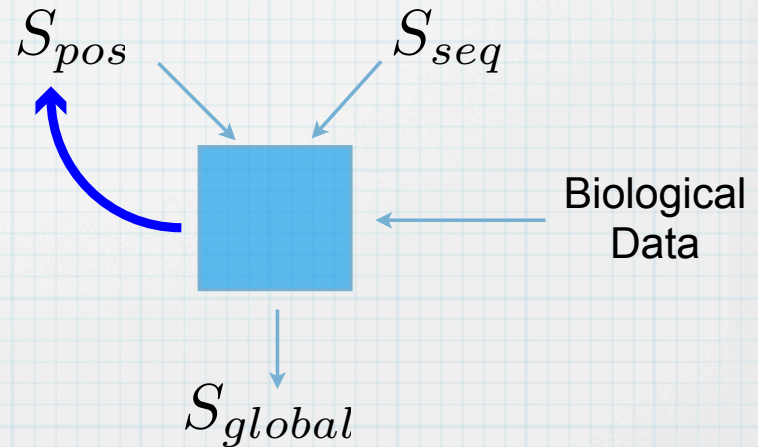
# Positional score of a site (binding sites, genes,...)

Needs to be specific with respect to which TF



$$S_{pos} = f\left(\left|\frac{S^* - S}{S}\right|\right) \times g(\max(S^*, S))$$

Combining scores : learning machine technique  
with J. hérisson, M. Elati, F. Képès



# Conclusion

## \* Why positions?

- \* Biological data show regular pattern --> space co-localization ↔ transcription regulation

## \* How to score?

- \* Method based on a solenoidal framework + clustering detection => valuable information for finding repetitive patterns

## \* Which outcome?

- \* Structural information about spatial organization of chromosomes
- \* Predicting functional relation between genes