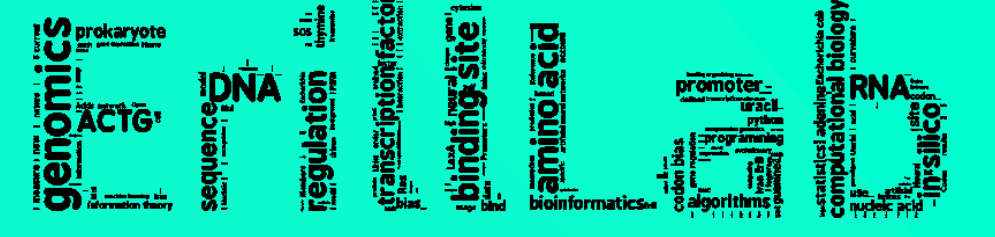


# Uncovering host-MGE cross-regulation

Elia Mascolo<sup>1</sup>, Ryan Burdick<sup>1,2</sup>, Emmanuel Mekasha<sup>1</sup>, Ivan Erill<sup>1</sup>

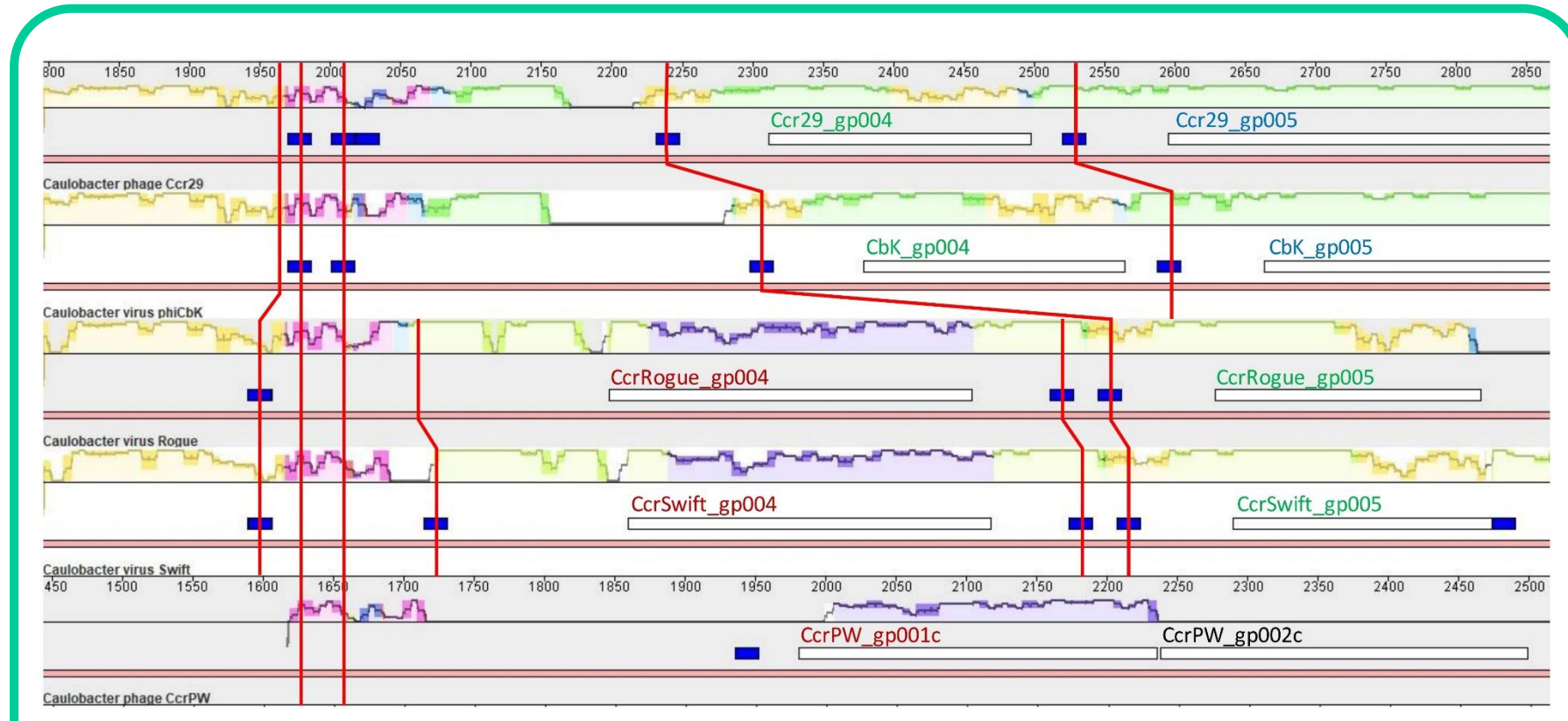
<sup>1</sup> Department of Biological Sciences, University of Maryland Baltimore County, Baltimore, MD 21250, USA

<sup>2</sup> HIV Dynamics and Replication Program, National Cancer Institute at Frederick, Frederick, MD 21702, USA

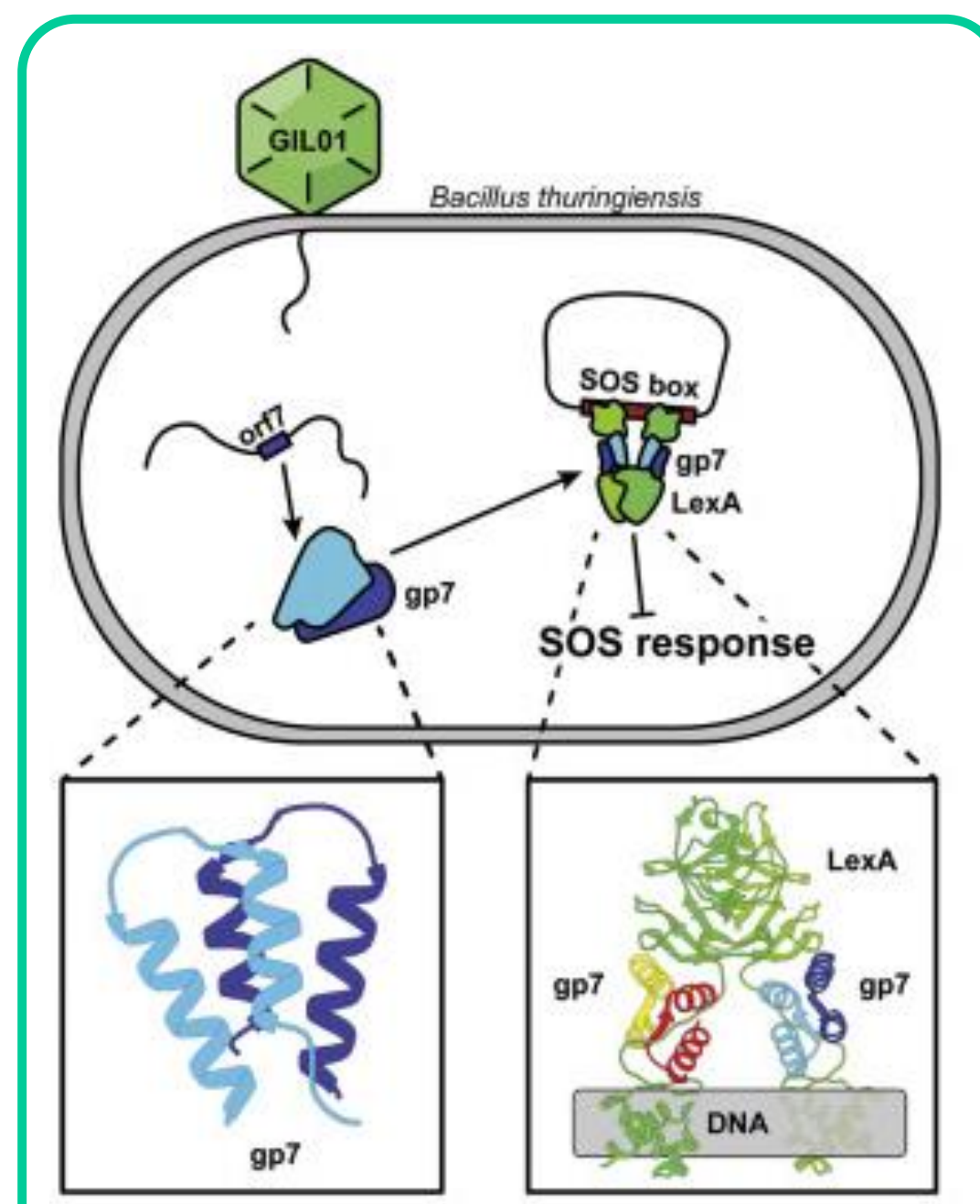


## Background and aim

Some **mobile genetic elements (MGE)** have genes regulated by **transcription factors (TFs)** of the host cell.

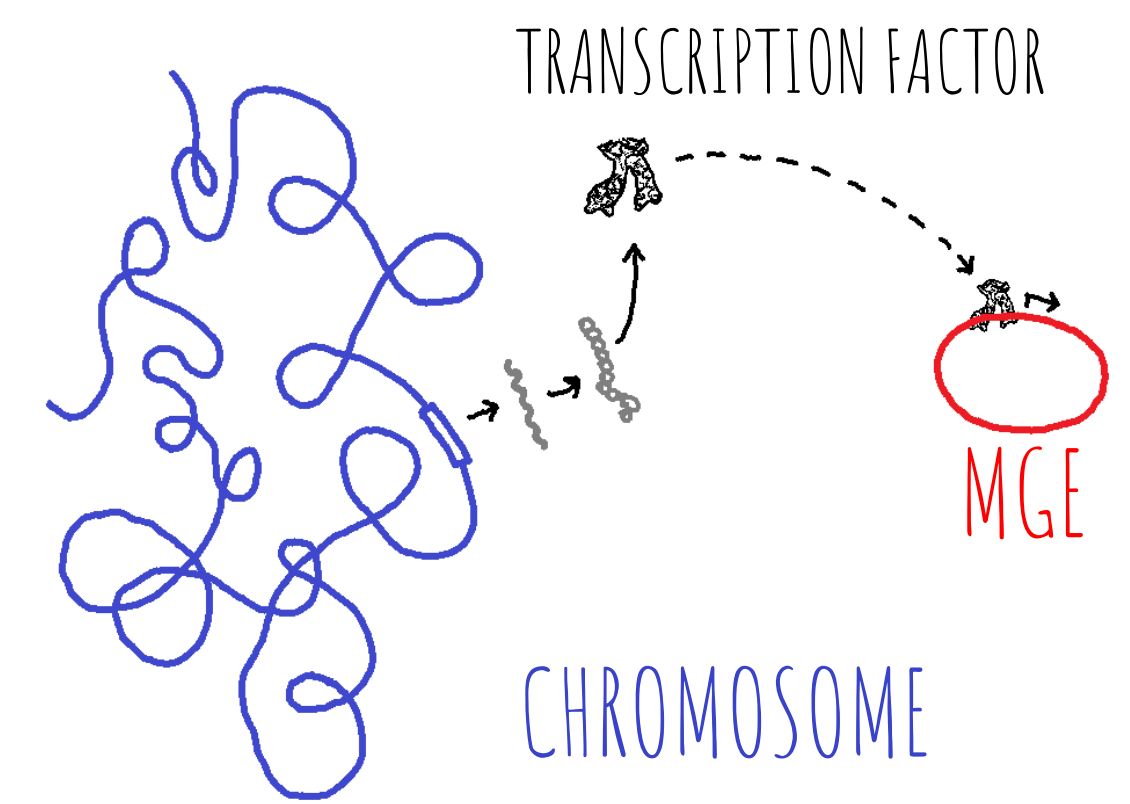
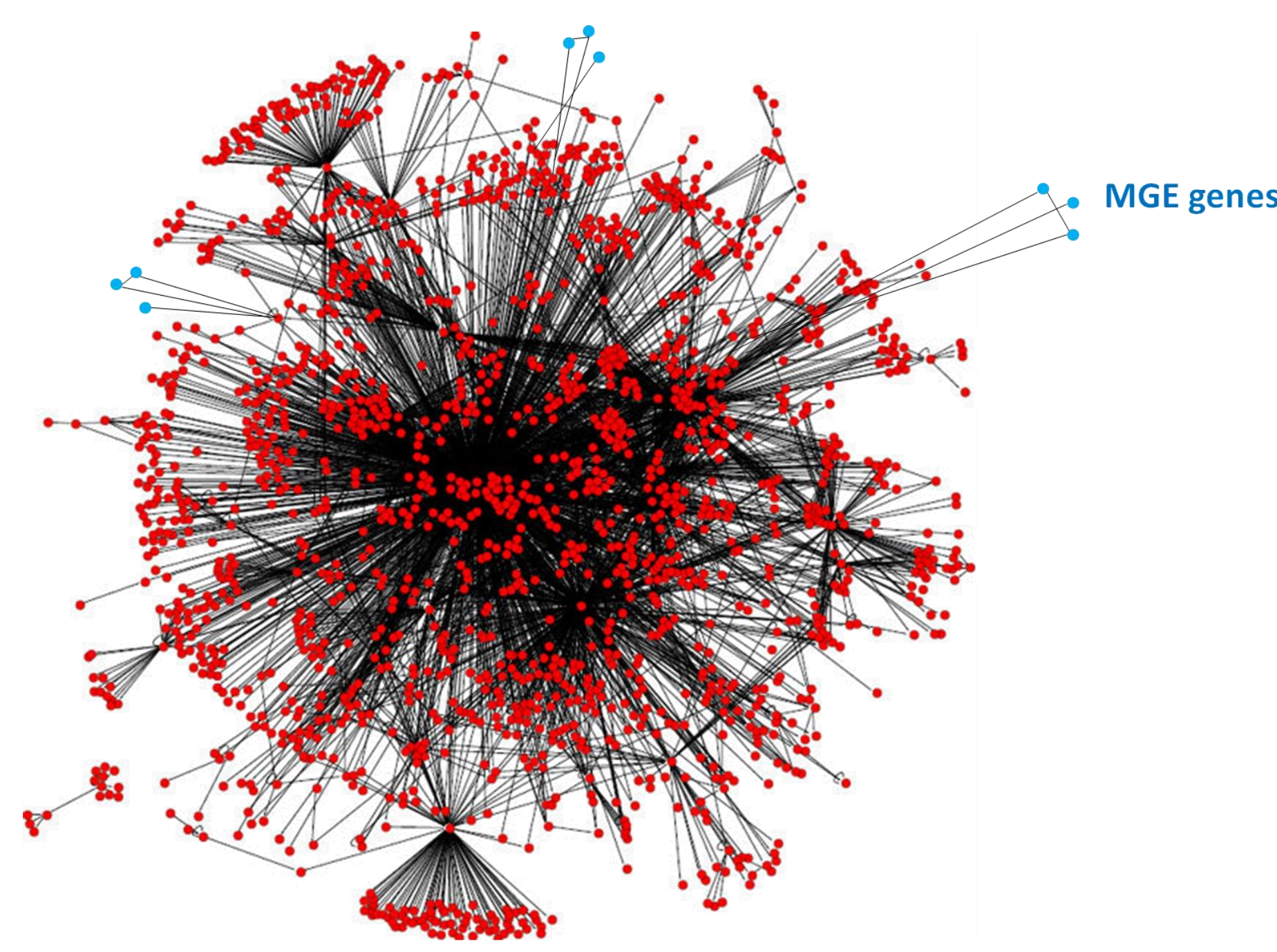


CtrA, the master cell cycle regulator in Alpha-proteobacteria, regulates gene expression in several Alpha-infecting phages (Mascolo et al., 2022).



Regulation of the lytic switch in Bacteriophage GIL01 by the host's LexA (Fornelos et al., 2011).

An extended model of transcriptional **gene regulatory networks (GRNs)**



**Aim:** to develop and benchmark a computational platform to systematically assess the extent of this novel GRN paradigm.

### References

Fornelos, N., Bamford, J. K. H., & Mahillon, J. (2011). Phage-borne factors and host LexA regulate the lytic switch in phage GIL01. *Journal of Bacteriology*. <https://doi.org/10.1128/JB.05618-11>  
 Mascolo, E., Adhikari, S., Caruso, S. M., deCarvalho, T., Folch Salvador, A., Serra-Sagrà, J., Young, R., Erill, I., & Curtis, P. D. (2022). The transcriptional regulator CtrA controls gene expression in Alphaproteobacteria phages: Evidence for a lytic deferment pathway. *Frontiers in Microbiology*, 0, 2999. <https://doi.org/10.3389/FMICB.2022.918015>

Frontiers Microbiology  
 "2022 outstanding article"

## Methods

PSSM: a model of transcription factor binding specificity

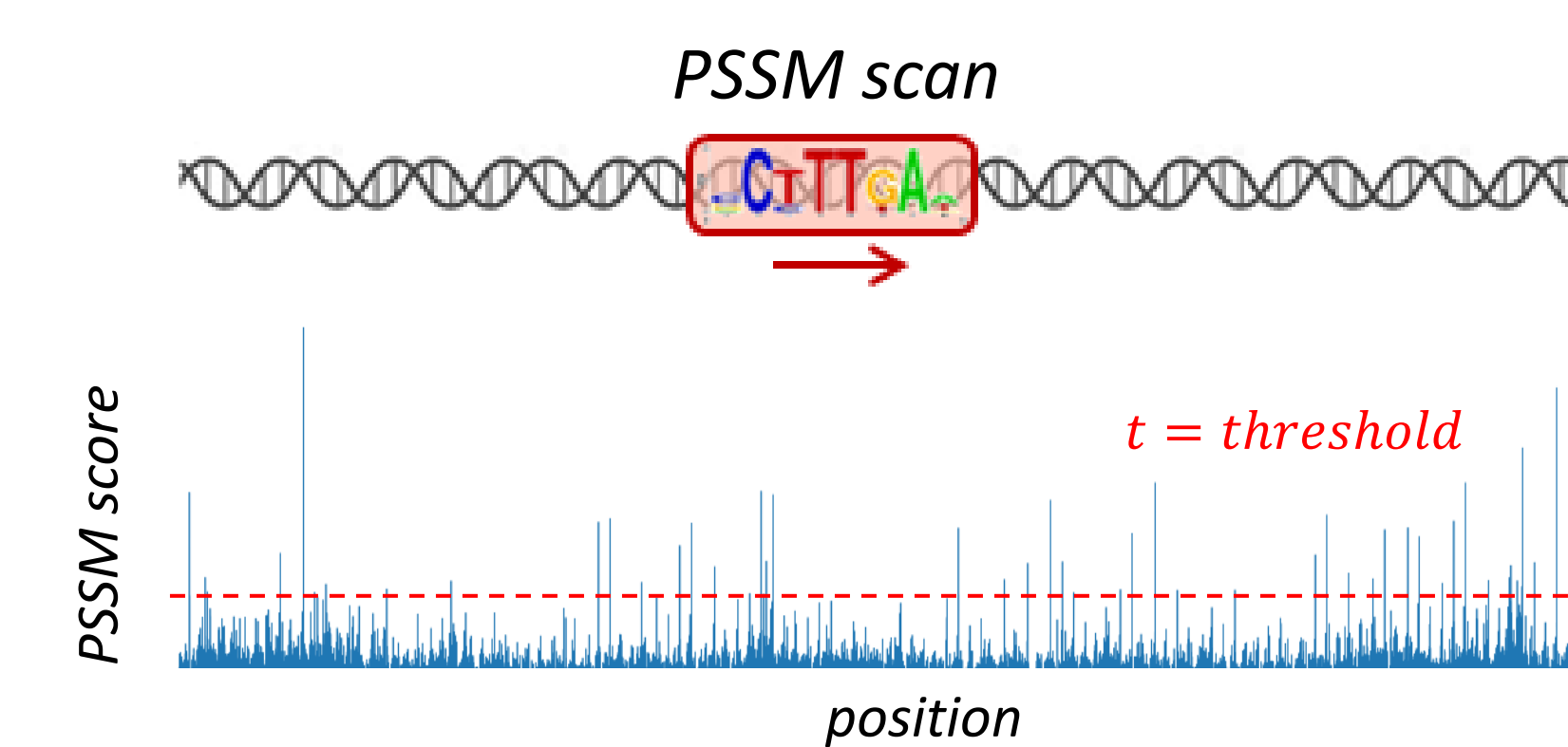
Position-specific probability matrix	A	0.33	0.17	0.00	0.00	0.00	0.17	0.83	0.50
	C	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.17
	G	0.17	0.67	0.00	1.00	1.00	0.83	0.17	0.00
	T	0.50	0.17	1.00	0.00	0.00	0.00	0.00	0.33

Elements are transformed using a background model

Position-specific scoring matrix (PSSM)	A	0.40	-0.56	-∞	-∞	-∞	-0.56	1.73	1.00
	C	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-0.56
	G	-0.56	1.42	-∞	2.00	2.00	1.73	-0.56	-∞
	T	1.00	-0.56	2.00	-∞	-∞	-∞	-∞	0.40

### From PSSM scores to genome-wide metrics

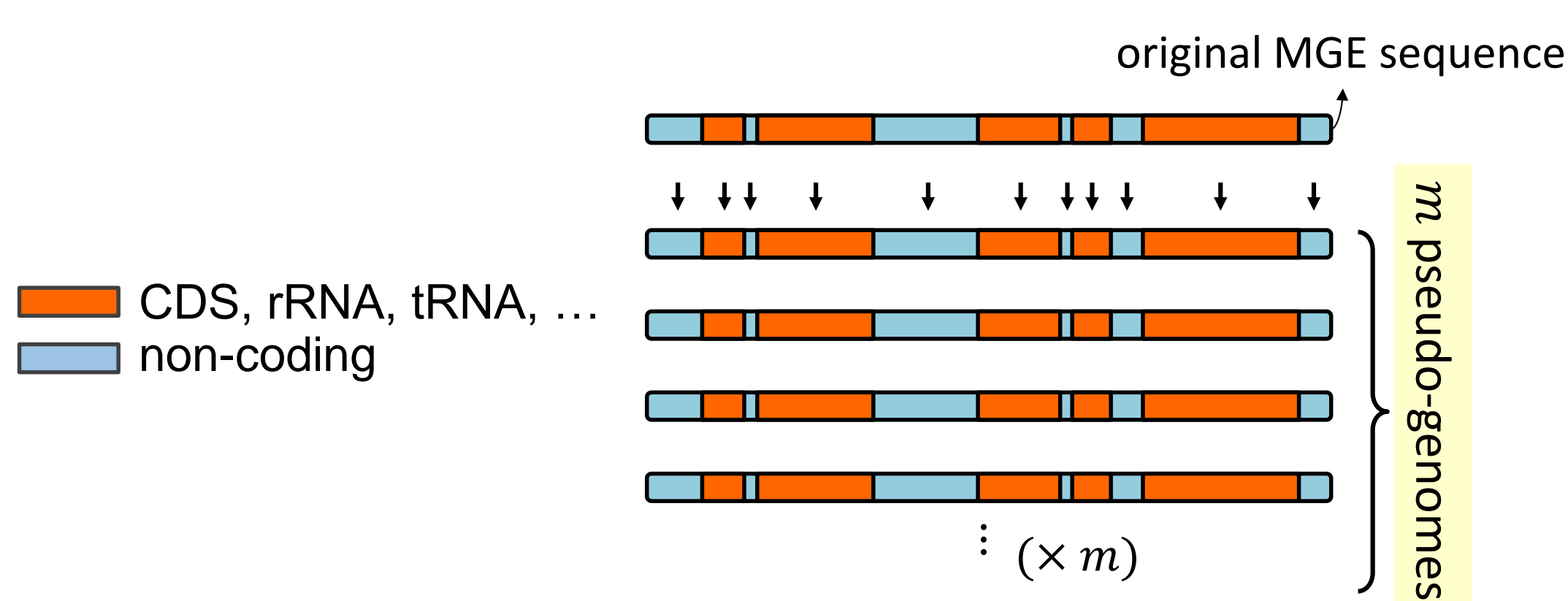
The problem of false positives



Three genomic metrics to catch signatures of cross-regulation

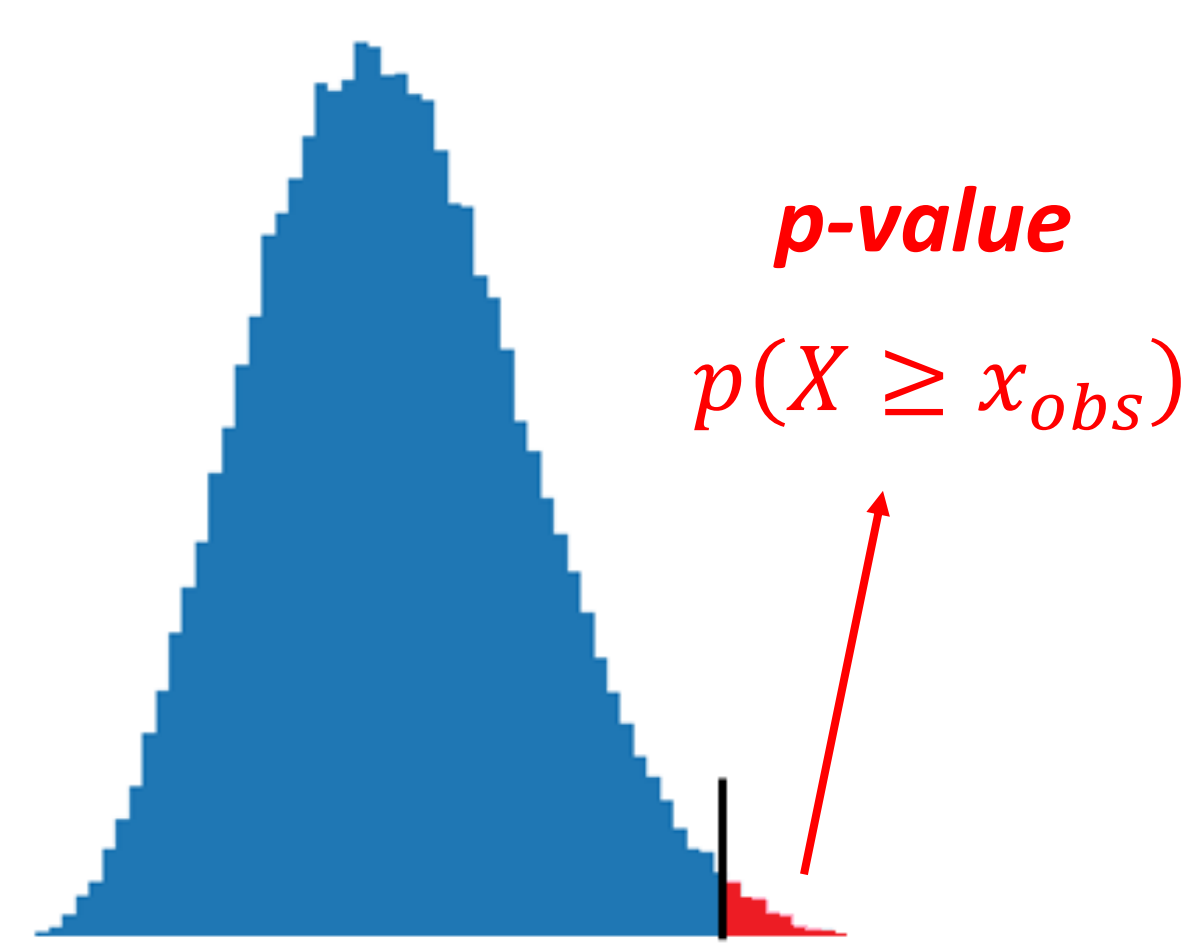
- scores extremeness**  $e = \sum_i (s_i - t)$  Accounts for the quantity and value of scores above the threshold
- colocalization**  $l = L(200)$  Detects if predicted sites tend to be less than 200 bp from each other more often than expected by chance  
Ripley's L function
- intergenicity**  $i = \frac{\# S_{inter}}{\# S_{inter} + \# S_{intra}}$  Proportion of predicted sites that are intergenic

**Pseudogenome:** k-mer sampling of original MGE sequence maintaining its genomic structure

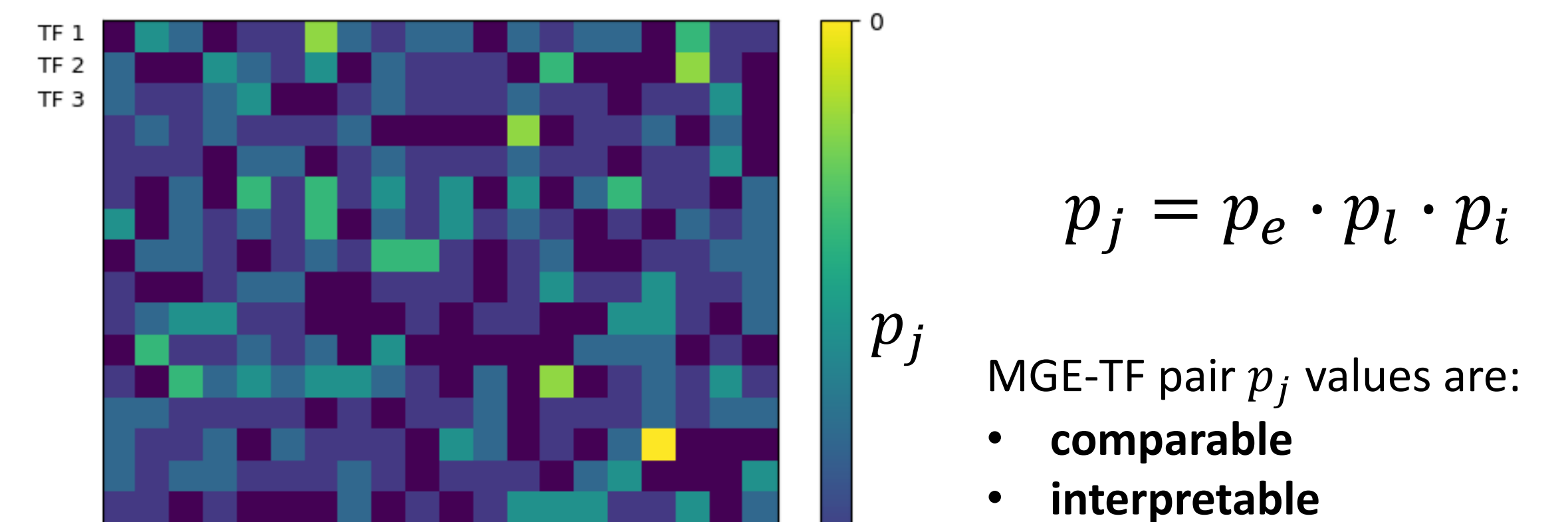


### Assessing statistical significance

Metric calculation on **pseudogenomes**. Inference of statistical significance



**Joint p-value,  $p_j$ :** integration of genomic metrics

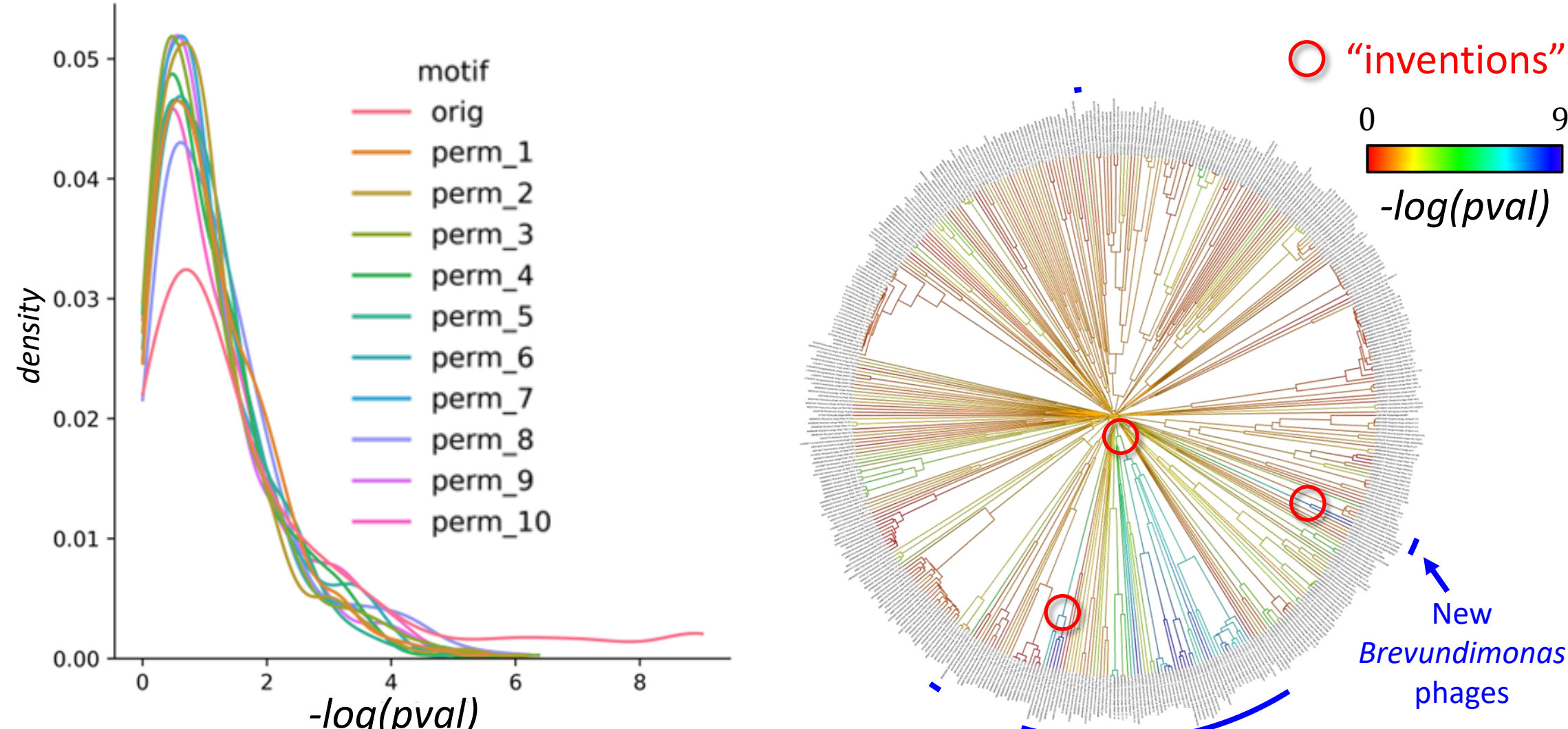


## Results

### CtrA (life cycle)

Recapitulation of results from (Mascolo et al., 2022)

476 complete genomes of phages infecting the Alphaproteobacteria, scanned with the CtrA motif

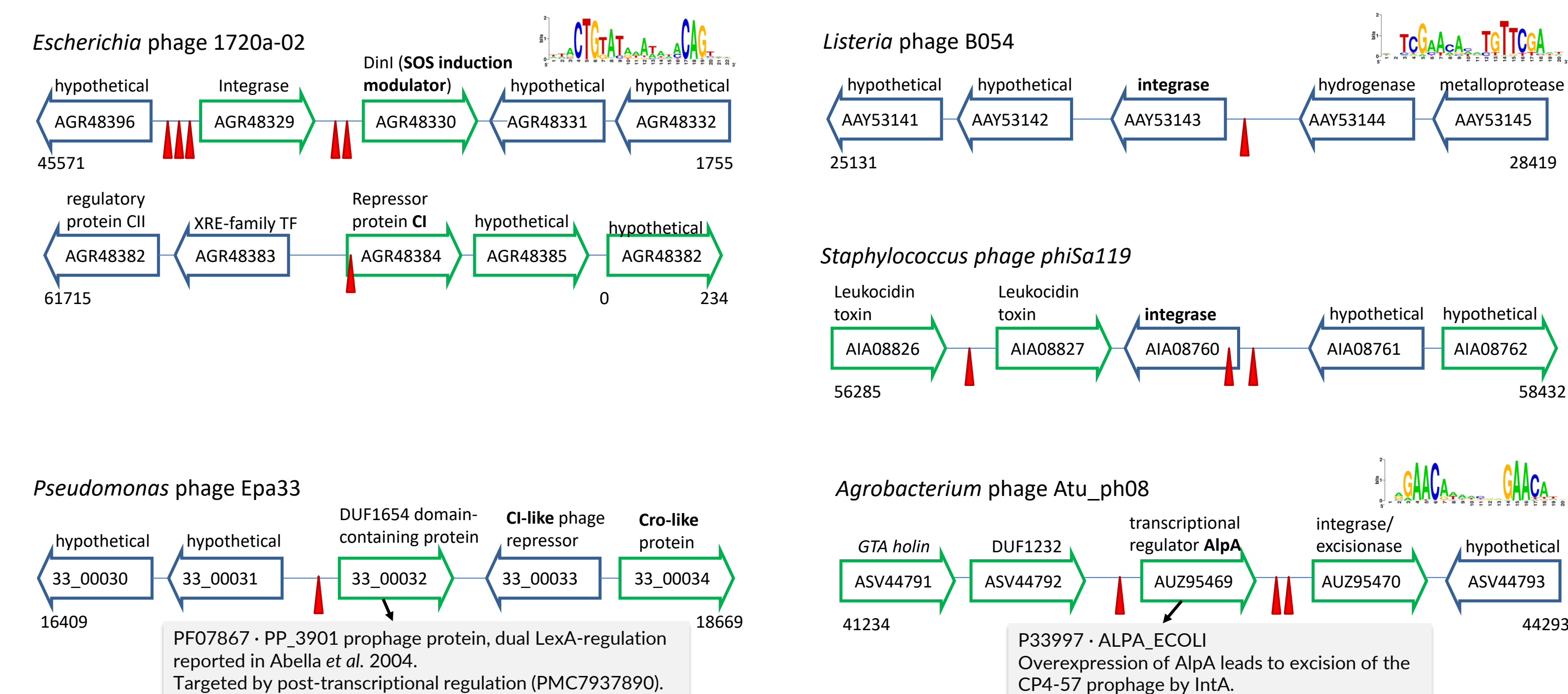


The joint p-value distribution stands out compared to 10 control experiments using permuted CtrA motifs.

The program identifies the known cases and provides a reconstruction of the evolution of cross-regulation.

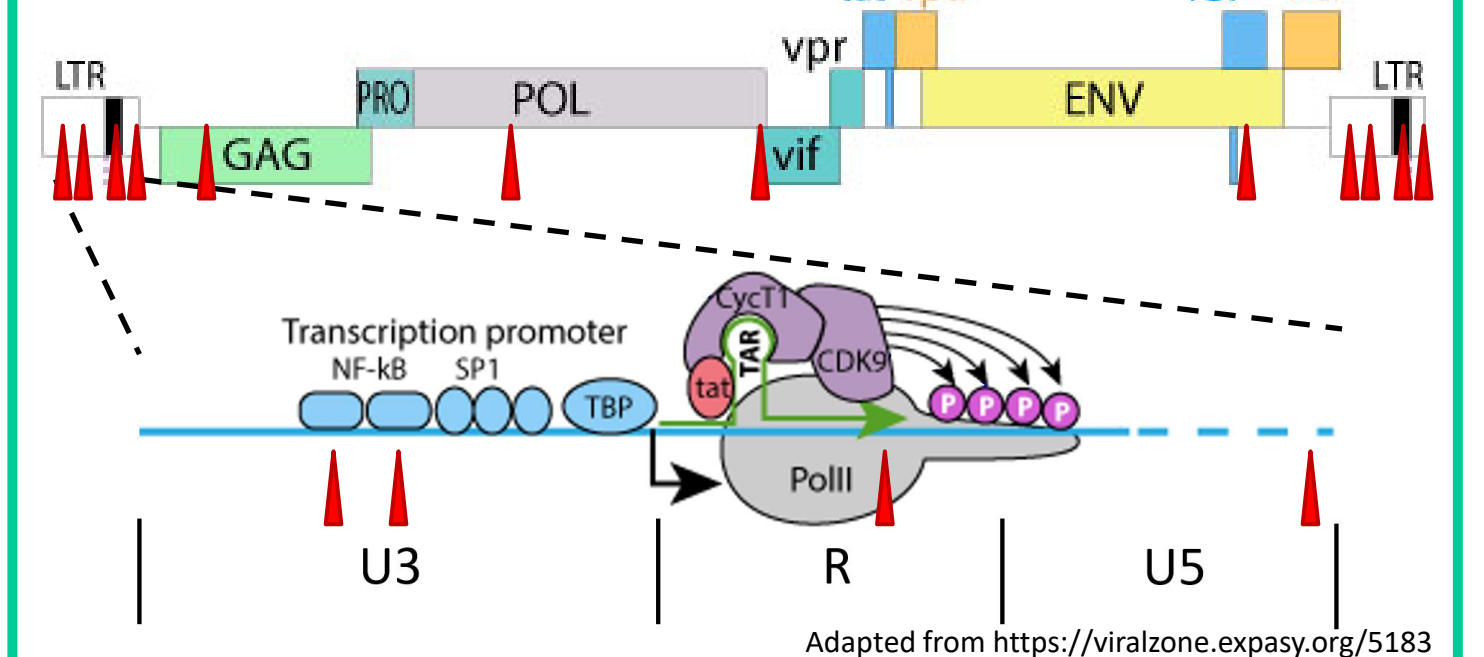
### LexA (SOS response to DNA damage)

LexA regulation of MGE genes across many bacterial clades



### HIV

Analysis of HIV-1/2 with 680 human TF PSSM models identifies known instances of regulation



	Rank based on joint p-value	
	RelA	NFKB1
HIV-1	1/680	2/680
HIV-2	2/680	11/680

## Acknowledgements

The UMBC High Performance Computing Facility (HPCF) is funded by the NSF CNS/SCREMS MRI program. Elia Mascolo and Emmanuel Mekasha were funded by the CNMS - Merck Academic Fellowship Program.

