

# Luca Nanni

PhD candidate, 3<sup>rd</sup> year. Adv: Stefano Ceri

Thesis submission: October 2020



## Computational methods for the inference of DNA folding mechanisms: from data management to machine learning

### Relevant publications/submissions

**Luca Nanni**, Pietro Pinoli, Arif Canakoglu, and Stefano Ceri. 2018. *Exploring Genomic Datasets: from Batch to Interactive and Back*. In Proceedings of the 5th International Workshop on **Exploratory Search in Databases and the Web** (ExploreDB 2018); collocated with SIGMOD PODS 2018. ACM, New York, NY, USA, Article 3, 6 pages.

**Luca Nanni**, Pietro Pinoli, Arif Canakoglu, Stefano Ceri, *PyGMQL: scalable data extraction and analysis for heterogeneous genomic datasets*. **BMC Bioinformatics** 20,560 (2019).

Cheng Wang, **Luca Nanni**, Boris Novakovic, Wout Megchelenbrink, Tatyana Kuznetsova, Hendrik G. Stunnenberg, Stefano Ceri & Colin Logie, *Extensive epigenomic integration of the glucocorticoid response in primary human monocytes and in vitro derived macrophages*, **Scientific Reports**, vol. 9, no. 1, p. 2772, 2019.

**Luca Nanni**, Stefano Ceri, Colin Logie, *Spatial patterns of CTCF sites define the anatomy of TADs and their boundaries*. **Under review @ Genome Biology**

Marco Varrone, **Luca Nanni (co-first)**, Giovanni Ciriello, Stefano Ceri, *Co-Expression Network Inference from Chromatin Conformation Data through Graph Embedding*. **Under review @ Intelligent Systems for Molecular Biology Conference**

# Additional research interests

## Chromatin Conformation in cancer genomes

Collaborating with the **Computational Systems Oncology Lab in Lausanne** on the study of Whole Genome Doubling and its effect on cancer evolution and prognosis.

## Deep-learning for cancer genomics

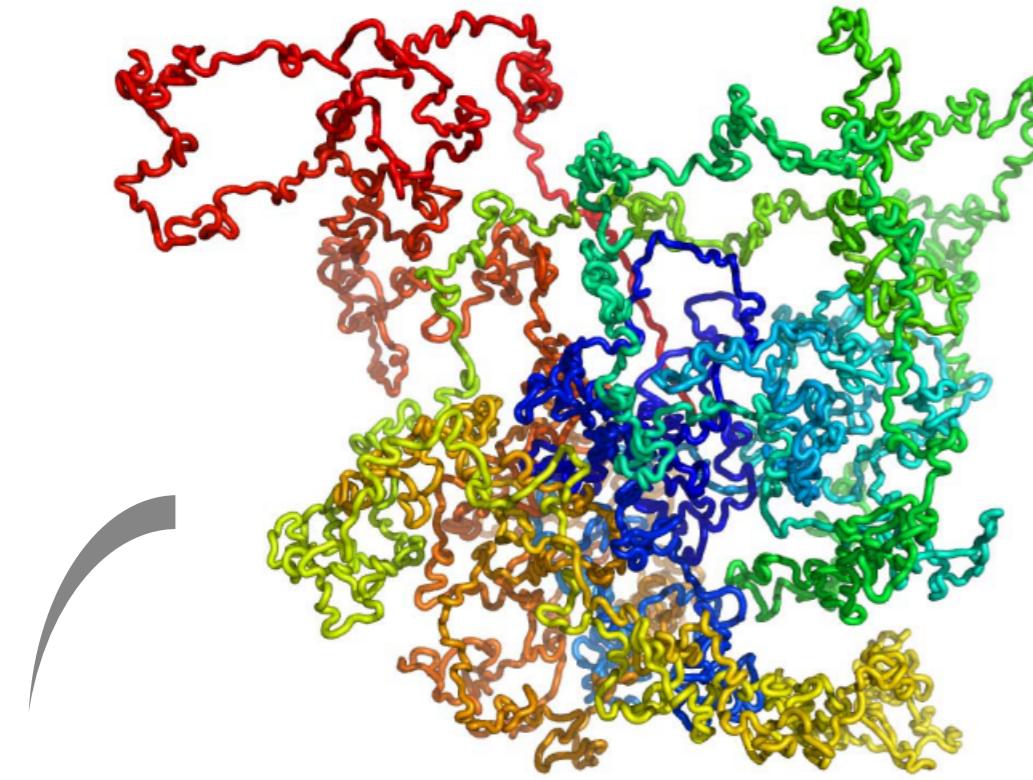
Arif Canakoglu, **Luca Nanni** (co-first), Artur Sakalouski, Ceri: *Designing and Evaluating Deep Learning Models for Cancer Detection on Gene Expression Data*. Springer **Lecture Notes in Bioinformatics** 2018.

Francisco Cristovao, Arif Canakoglu, Mark Carman, Silvia Cascianelli, **Luca Nanni**, Pietro Pinoli, Marco Masseroli; *Comparing classic, deep and semi-supervised learning for whole-transcriptome breast cancer subtyping*. Presented at the 16th **International Conference on Computational Intelligence methods for Bioinformatics and Biostatistics** (CIBB2019).

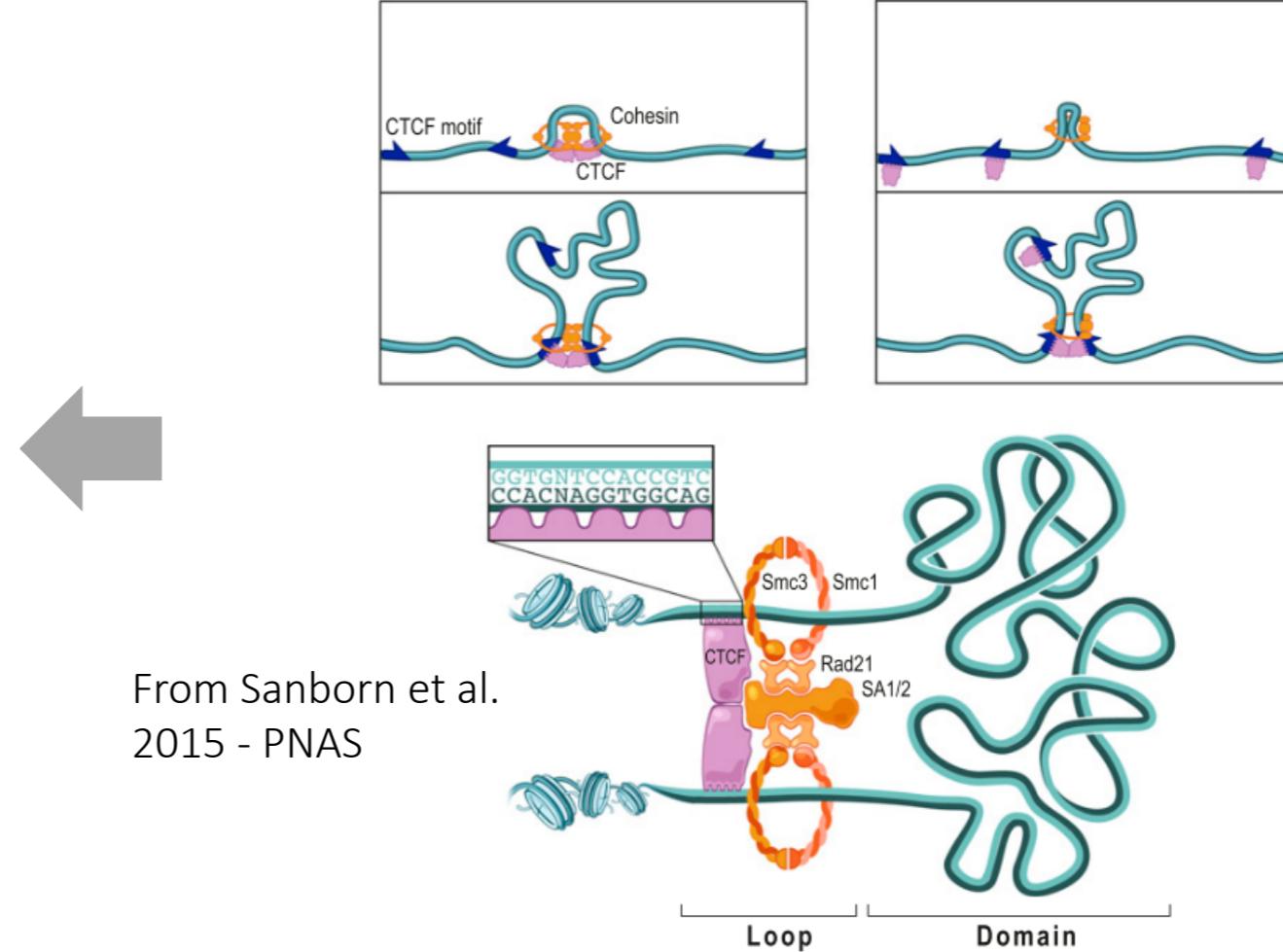
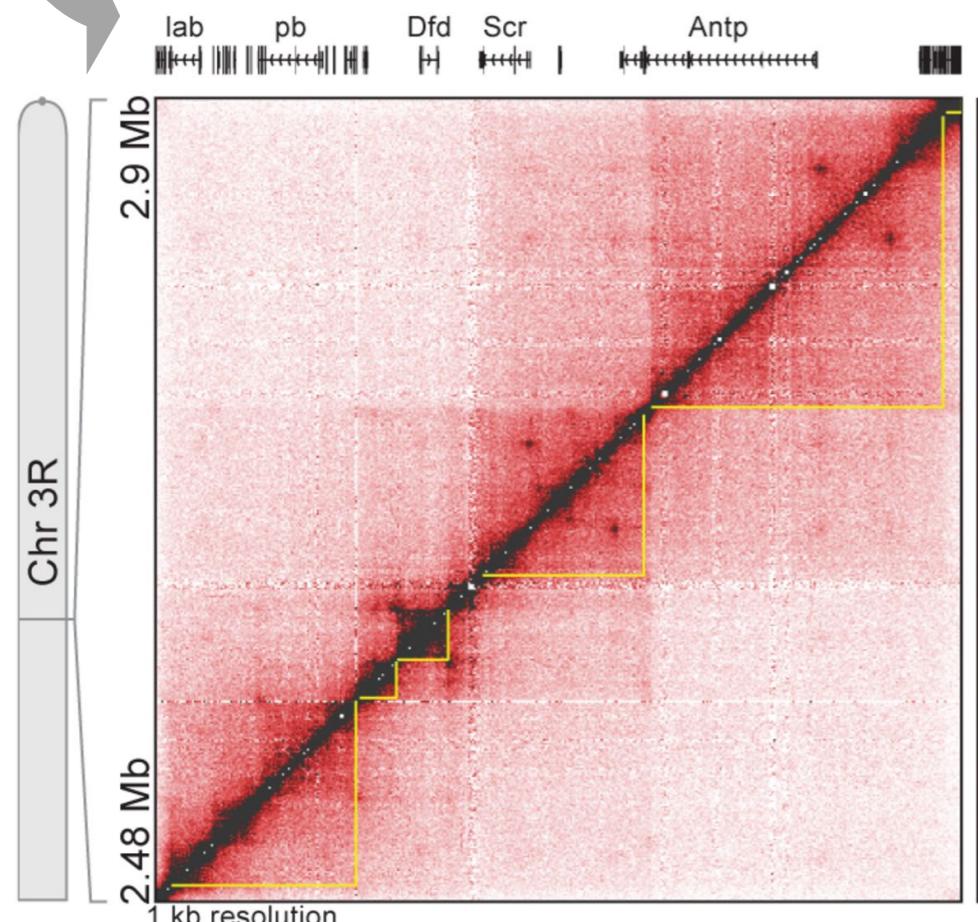
## High-performance computing for genomic research

Marco Masseroli, Arif Canakoglu, Pietro Pinoli, Abdulrahman Kaitoua, Andrea Gulino, Olha Horlova, **Luca Nanni**, Anna Bernasconi, Stefano Perna, Eirini Stamoulakatou, Stefano Ceri, *Processing of big heterogeneous genomic datasets for tertiary analysis of Next Generation Sequencing data*, **Bioinformatics**, Volume 35, Issue 5, 01 March 2019, Pages 729–736

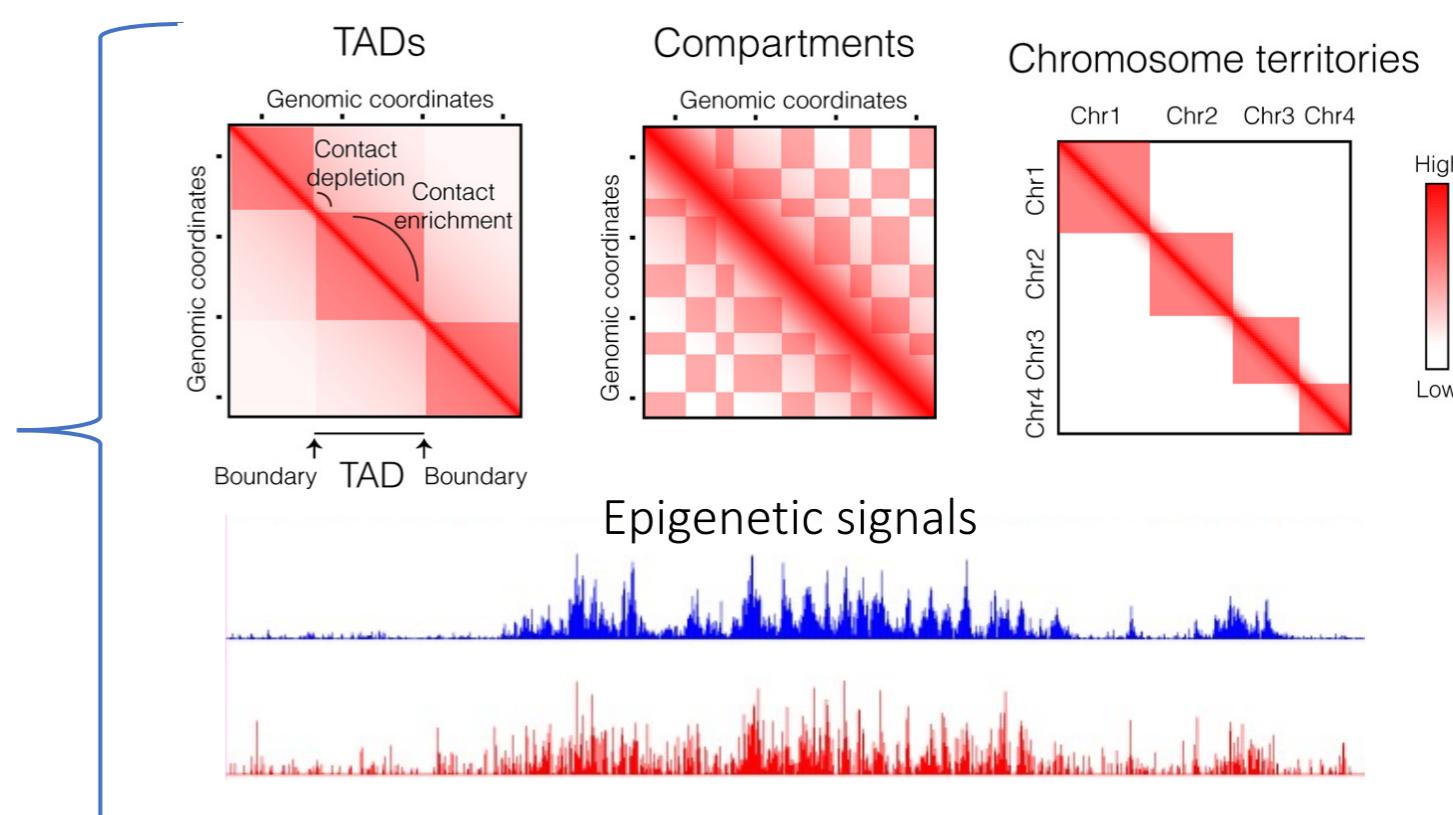
Pietro Pinoli, Stefano Ceri, Davide Martinenghi, **Luca Nanni**, *Metadata management for scientific databases*. **Information Systems** 81 (2019): 1-20.



From Lieberman et al. 2009 - Science



From Sanborn et al.  
2015 - PNAS



Bringing scalable computation to 3D genomics research through interactive computing and key data structures



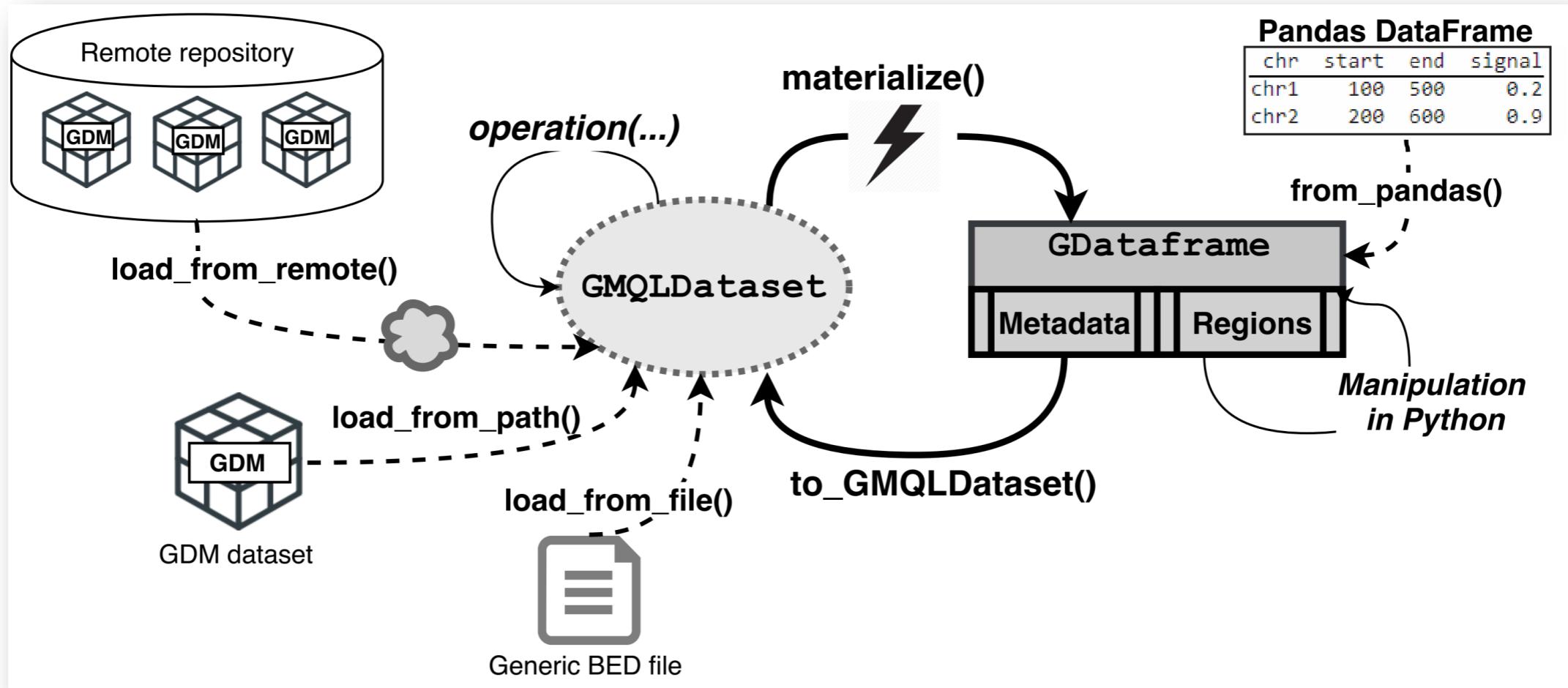
What are the epigenetic drivers of chromosome folding?

Can we infer the rules of genome folding from an essential set regulators?

How does chromatin folding affect gene expression?

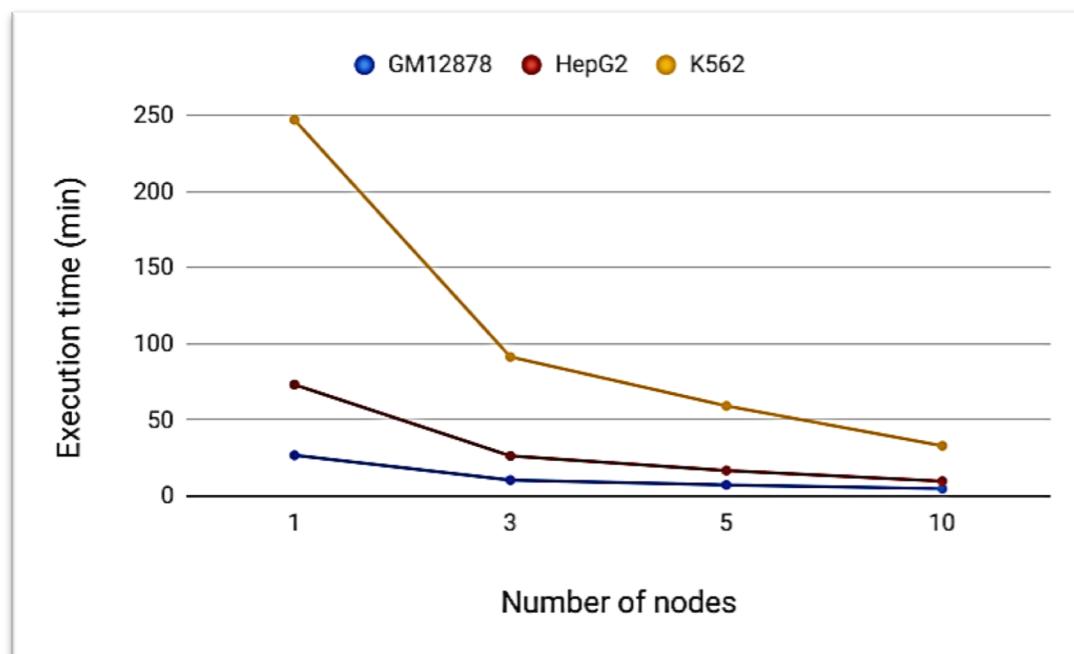
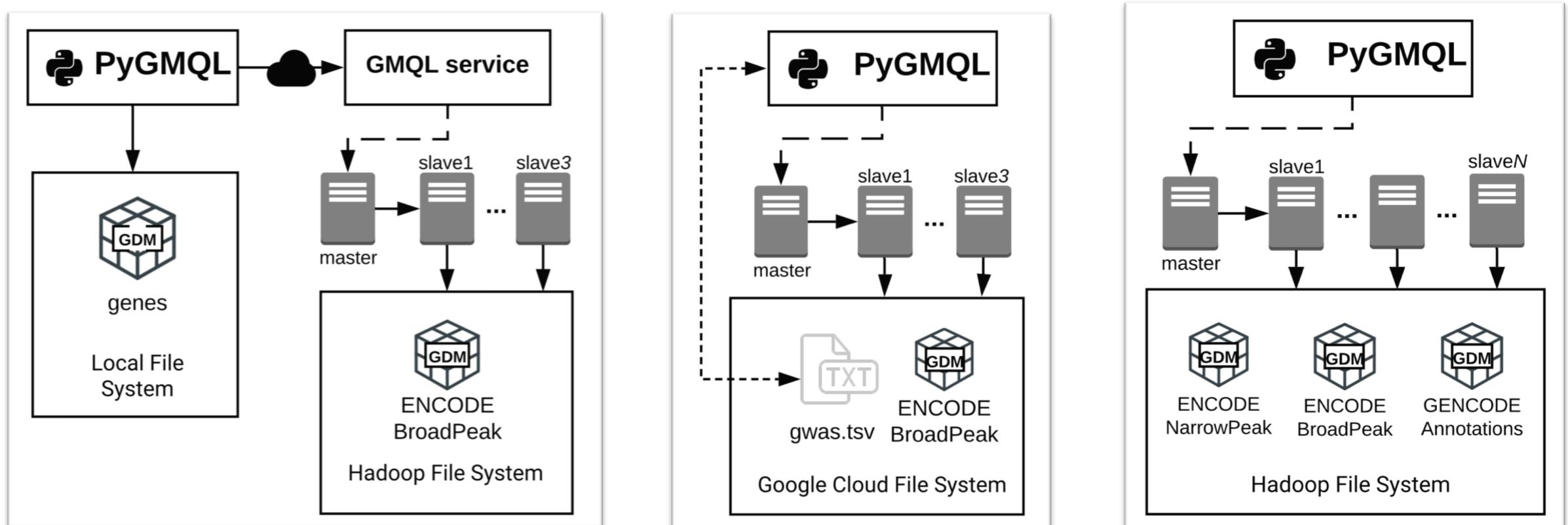
Envisioning a unified model of chromatin folding and epigenetic regulation

# Data structure design for genomic data exploration



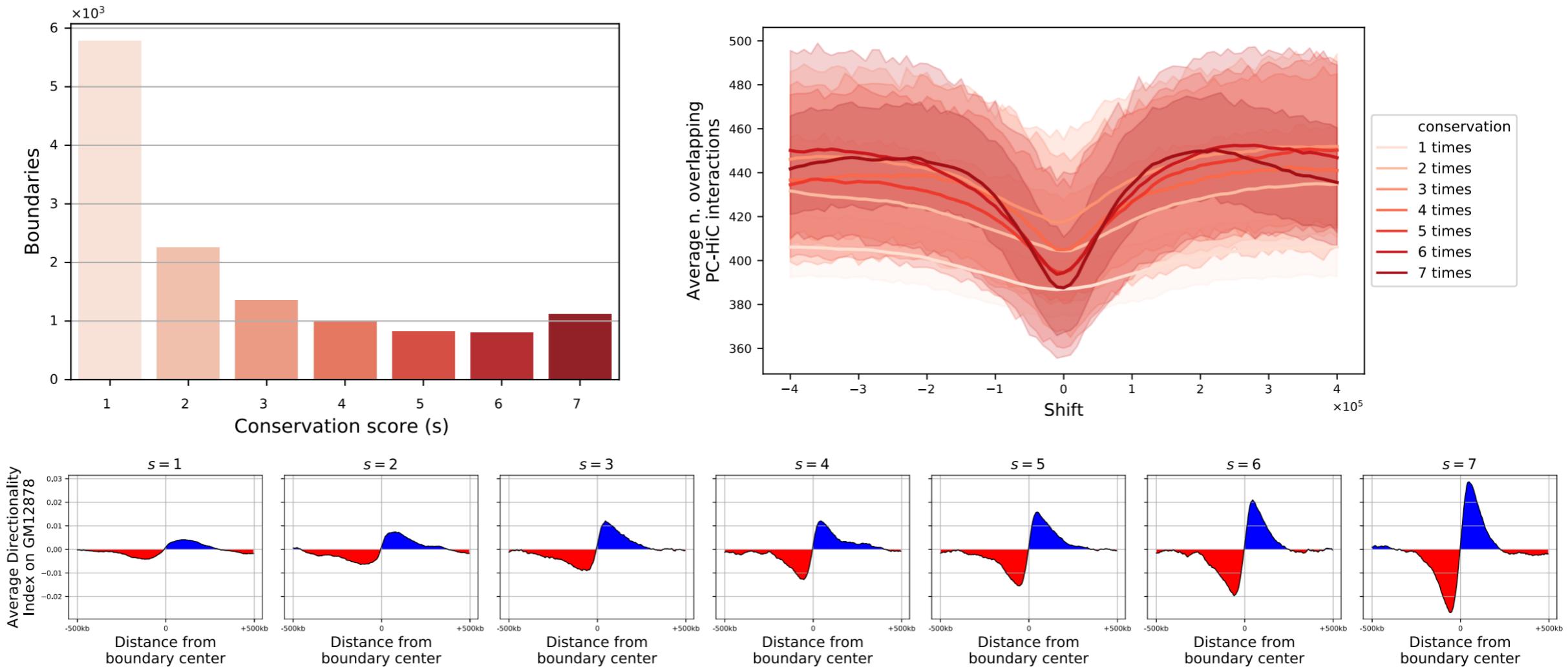
Luca Nanni, Pietro Pinoli, Arif Canakoglu, and Stefano Ceri. 2018. *Exploring Genomic Datasets: from Batch to Interactive and Back*. In Proceedings of the 5th International Workshop on Exploratory Search in Databases and the Web. ACM, New York, NY, USA.

# HPC interactive deployment



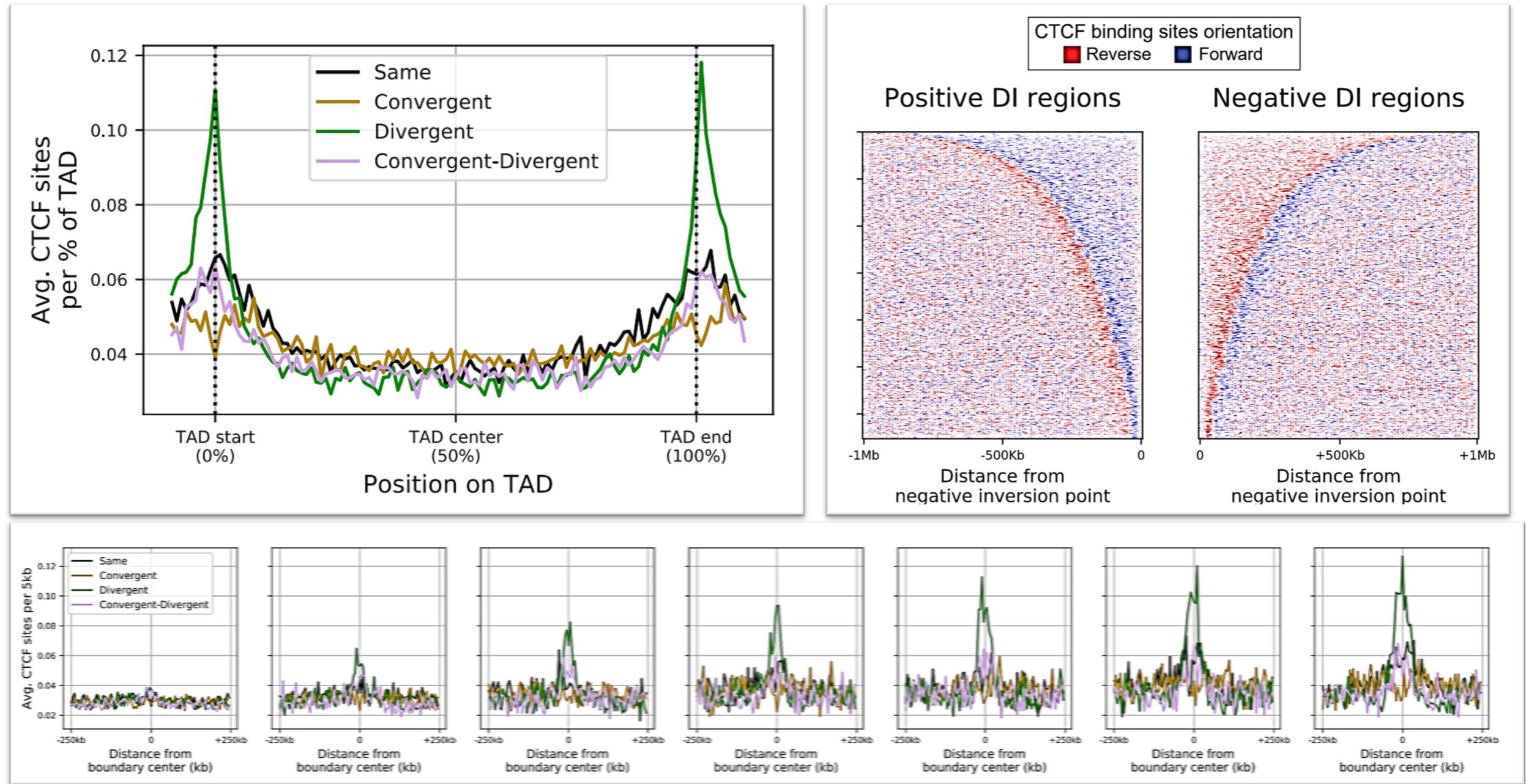
Luca Nanni, Pietro Pinoli, Arif Canakoglu,  
Stefano Ceri, *PyGMQL: scalable data extraction  
and analysis for heterogeneous genomic  
datasets*. BMC Bioinformatics 20,560 (2019).

# Conserved TAD boundaries across cell lines show higher level of insulation



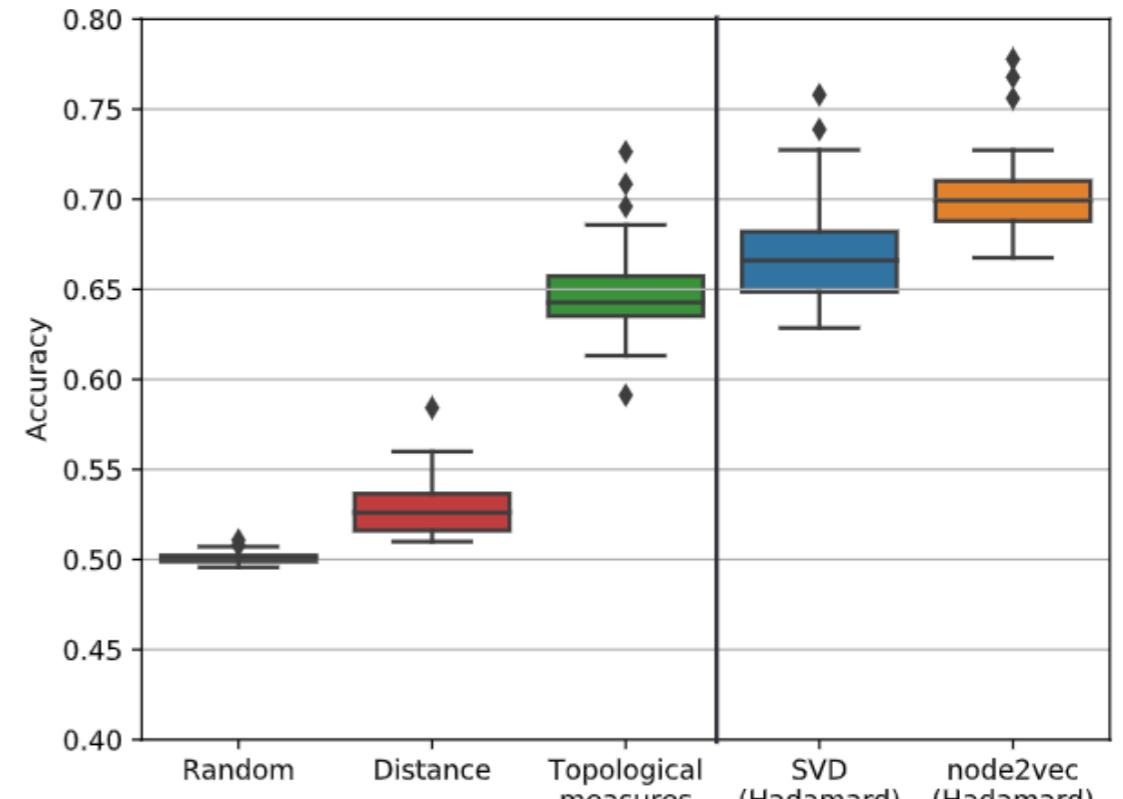
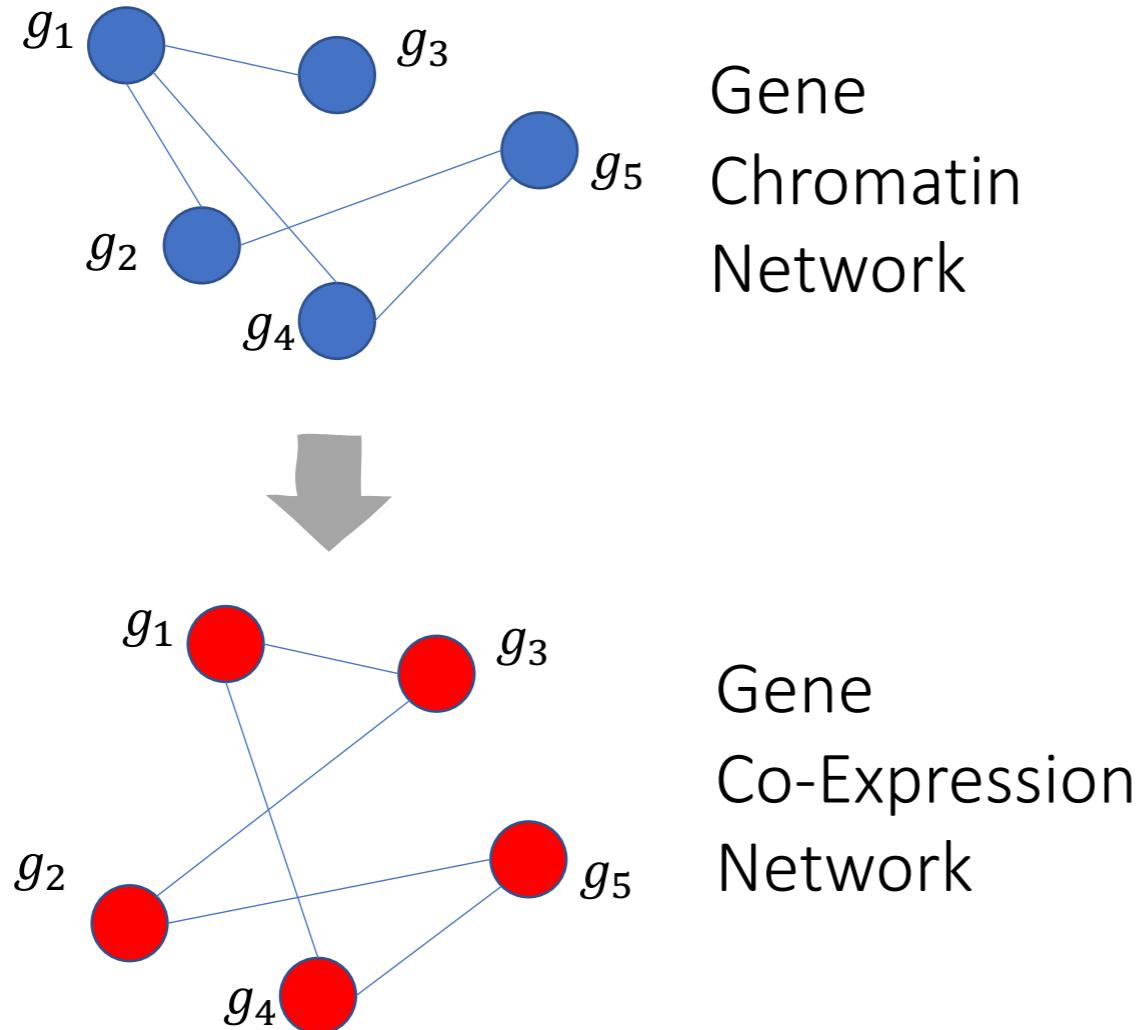
Luca Nanni, Stefano Ceri, Colin Logie, *Spatial patterns of CTCF sites define the anatomy of TADs and their boundaries. Under review @ Genome Biology*

# TAD formation is regulated by a strict *grammar*, defined by CTCF binding site orientations



Luca Nanni, Stefano Ceri, Colin Logie, *Spatial patterns of CTCF sites define the anatomy of TADs and their boundaries. Under review @ Genome Biology*

# Chromatin Conformation can predict co-expression patterns (w. Marco Varrone)

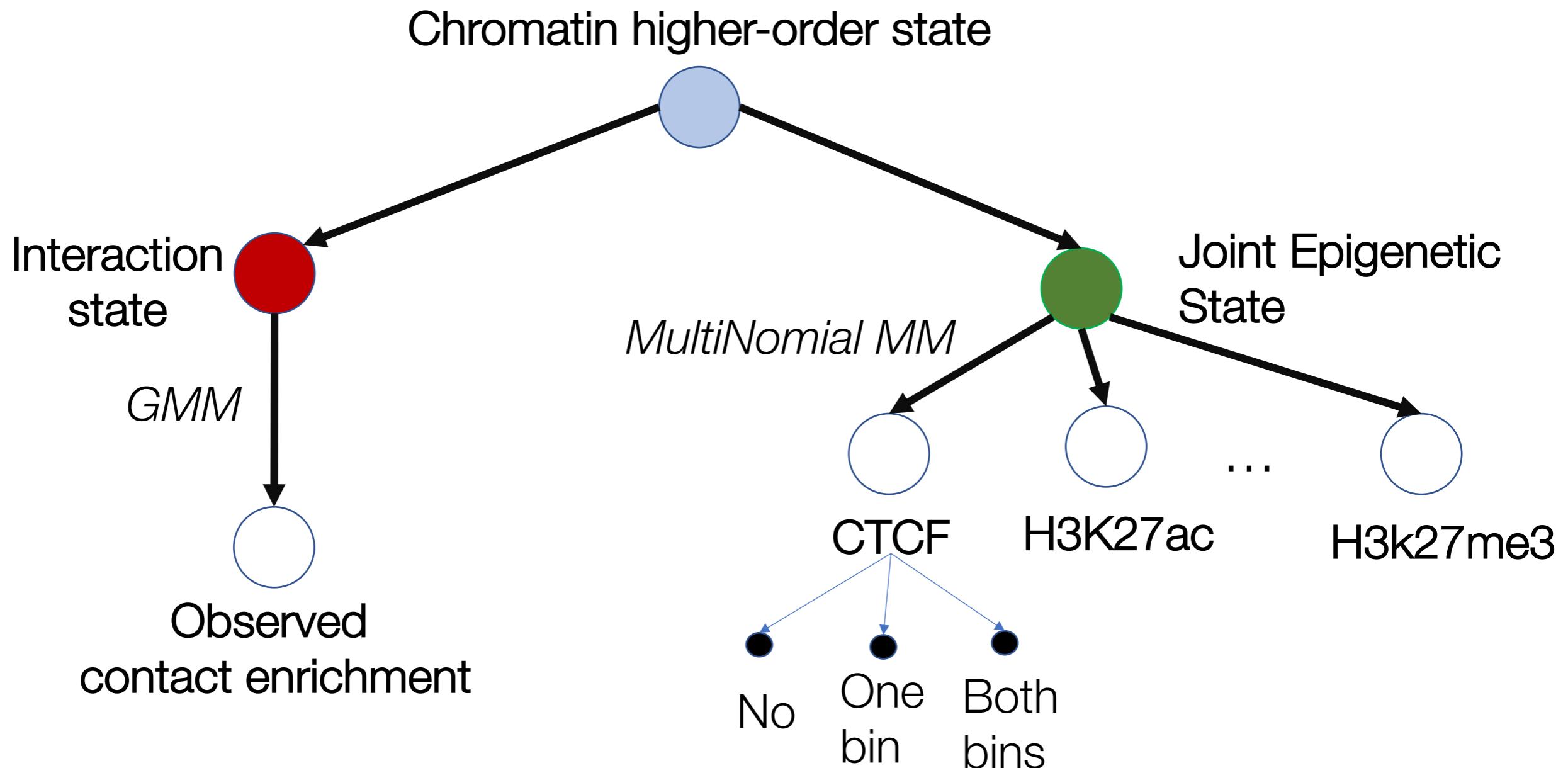


## Graph Embedding:

- Matrix Factorization
- Node embedding

Marco Varrone, Luca Nanni (co-first), Giovanni Ciriello, Stefano Ceri, *Co-Expression Network Inference from Chromatin Conformation Data through Graph Embedding*.  
Under review @ Intelligent Systems for Molecular Biology Conference

# Towards a unified model of chromatin conformation and epigenetic regulation



Currently under development under the supervision of Giovanni Ciriello from the Computational Systems Oncology Lab in University of Lausanne/EPFL