

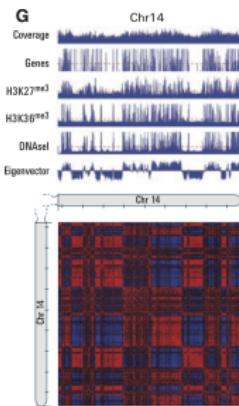
# Hi-C Tutorial: Topological Domains

Emre Sefer    Geet Duggal



# What are Topological Domains?

# Chromatin Classification by PCA

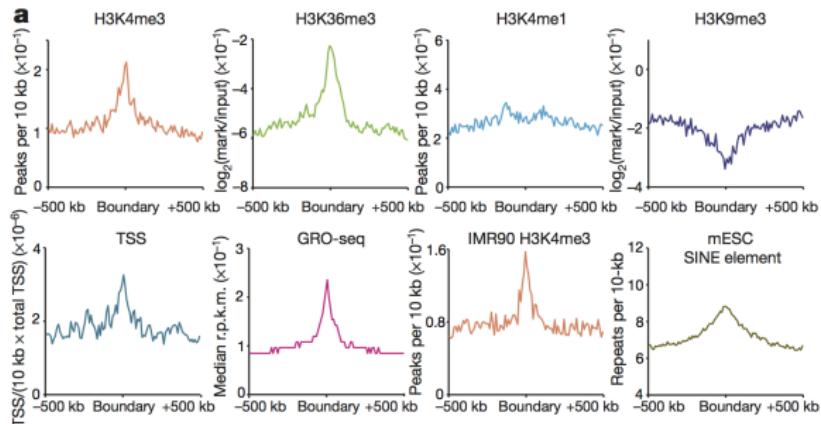


- Human chromosomes are partitioned into two types of compartments:
  - Compartment A -> open chromatin -> active gene-dense regions
  - Compartment B -> closed chromatin -> repressive gene-poor regions

---

<sup>0</sup>Lieberman-Aiden et al., "Comprehensive mapping of long-range interactions reveals folding principles of the human genome".

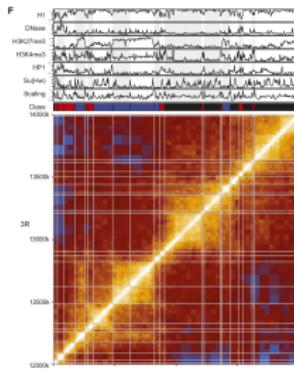
# TADs Are Enriched For Genomic Features



- H3k4me3 is enriched at TAD boundaries.
- Repressive markers such as H3k9me3 are depleted at TAD boundaries.
- CTCF binding sites are also enriched.

<sup>o</sup>Dixon et al., "Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions".

# Domain Finder: Sexton et al.



- Domain partition based on likelihood optimization

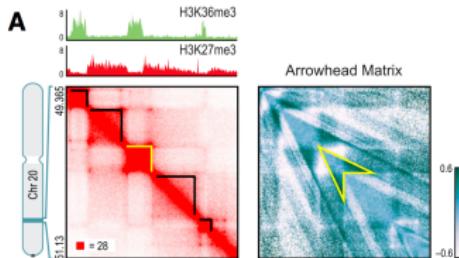
$$P(X_{a,b}) = \delta_{cl(a), cl(b)}^{domain(a,b)} \left( \sum_{a < k < b} \gamma_k l(k) \right) \cdot F_{len}(a_{len}, b_{len}) \cdot F_{gc}(a_{gc}, b_{gc}),$$

- TADs create unexpectedly low numbers of contacts crossing the boundary regions
- Peaks correspond to domain boundaries

---

<sup>0</sup>Sexton et al., "Three-dimensional folding and functional organization principles of the Drosophila genome".

# Domain Finder: Arrowhead

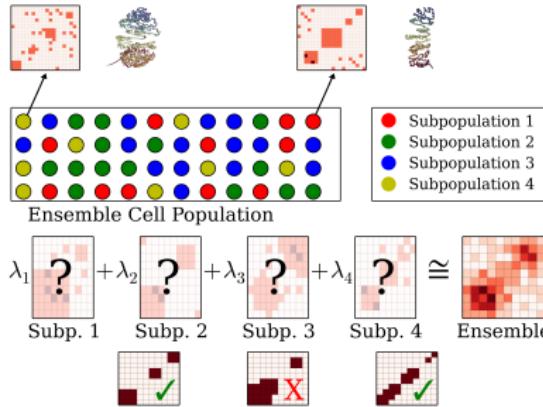


- Builds modified difference matrix corresponding to directionality biases
  - Search arrowhead pattern over this matrix heuristically
- Find corners of domains by arrowhead search
- Make use of very high resolution contact maps

---

<sup>o</sup>Rao et al., "A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping".

# Domain Finding via Deconvolution

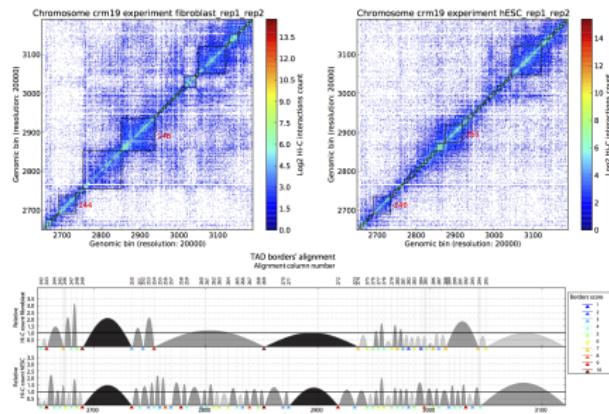


- Deconvolution: Given an ensemble Hi-C matrix, can we extract the mixing components?
  - Identify TADs indirectly by deconvolution
  - Domains in the context of structural classes
- It is based on combinatorial optimization

---

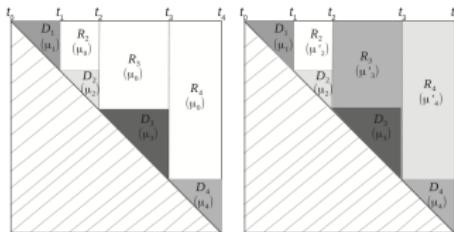
<sup>o</sup>Sefer, Duggal, and Kingsford, "Deconvolution Of Ensemble Chromatin Interaction Data Reveals The Latent Mixing Structures In Cell Subpopulations".

# Domain Finder: TADBIT



- TADBit uses breakpoint detection to detect TAD border positions
  - It uses BIC penalized estimation of likelihood
  - Poisson assumption of Hi-C counts
- Dynamic programming based algorithm

# Domain Finder: HiCseq



- Partitions into TAD segments by optimizing maximum likelihood via dynamic programming
  - Related to 1D and 2D segmentation problems
- Does not allow identification of multiscale or hierarchical domains

---

<sup>0</sup>Lévy-Leduc et al., "Two-dimensional segmentation for analyzing Hi-C data".

# Weaknesses of the existing methods

- Dixon et al. :
  - It is slow and it has many parameters
  - It is limited to a single scale
- Armatus:
  - Scaling parameter is not linear with respect to avg domain size
- HICseq:
  - Each segment must belong to a domain
- Sefer et al. :
  - Not scalable to recent high resolution datasets

## Open Problems

- How to find TADs jointly over multiple species?
- How to find true TAD boundaries efficiently over mixture cell populations?

# Conclusion

- 
- 
-

# Acknowledgements

- Thanks to Kingsford Group Members

