

Hi-C Topological Domains Tutorial

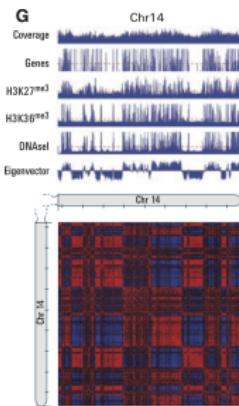
Emre Sefer

Geet Duggal



What are Topological Domains?

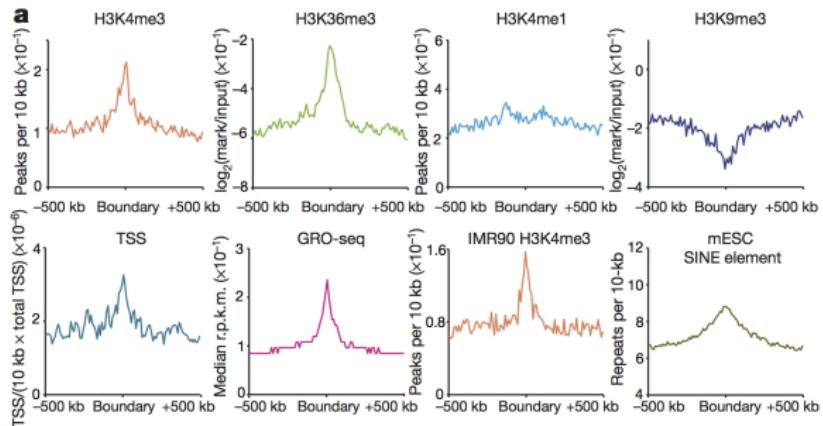
Chromatin Classification by PCA



- 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

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

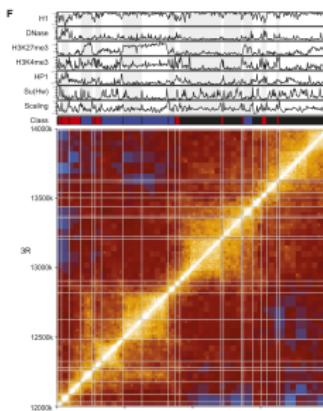
TADs Are Enriched For Genomic Features



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

^oDixon et al., "Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions".

Domain Finder: Sexton et al.

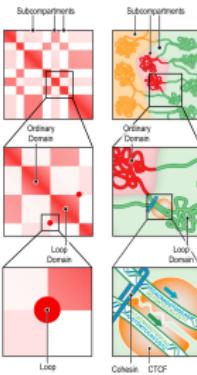


- Domain partition based on likelihood optimization
- Model parameters are estimated independently for each chromosome

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

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

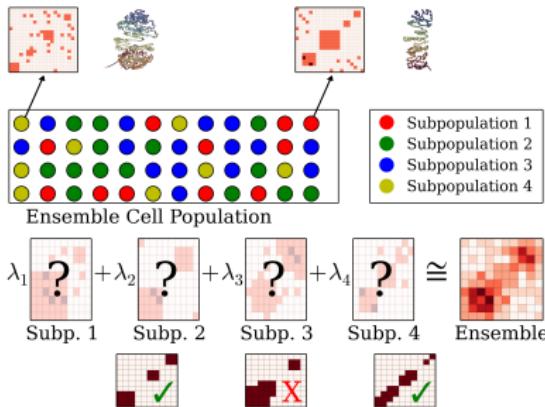
Domain Finder: Arrowhead



- Arrowhead estimates a modified difference matrix A varying between positive and negative values
 - $A_{i,i+d}$ is strongly positive if locus $i - d$ is inside a domain and locus $i + d$ is not.
 - $A_{i,i+d}$ is close to zero if both are inside a domain.
- Runs dynamic programming to find domains

^oRao et al., "A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping".

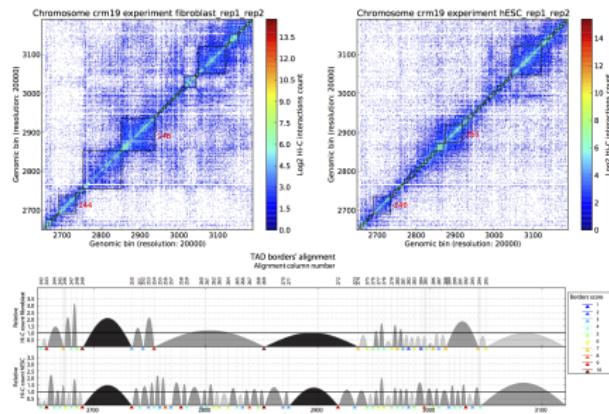
Domain Finding via Deconvolution



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

⁰Sefer, 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

Disadvantages of the existing methods

Acknowledgements

- Thanks to Kingsford Group Members

