

NGS analysis for gene regulation and epigenomics

Jacques Serizay, Cyril Matthey-Doret

January 11, 2020

Day 2

Applications of chromosome conformation technologies

Biological insights

Several motivations to investigate chromosome conformation:

- Mechanics of regulation: Promoter-enhancer, DNA repair, insulation, ...
- Fundamental physics of chromatin: phase separation, compaction, ...
- 3D model of chromosomes, their organization and segregation

What to look for in a contact map ?

Extract signals for more quantitative analyses:

- Contact probability vs genomic distance
- Insulation score
- Compartment eigenvectors
- Feature detection: loops, domains, hairpins, ...

Distance-dependent contacts

The diagonal of Hi-C map contains useful information:

- Distant-dependent contact decay follows a power-law
- The slope of the curve gives compaction information
- Can be used to compare polymer models

Insulating boundaries

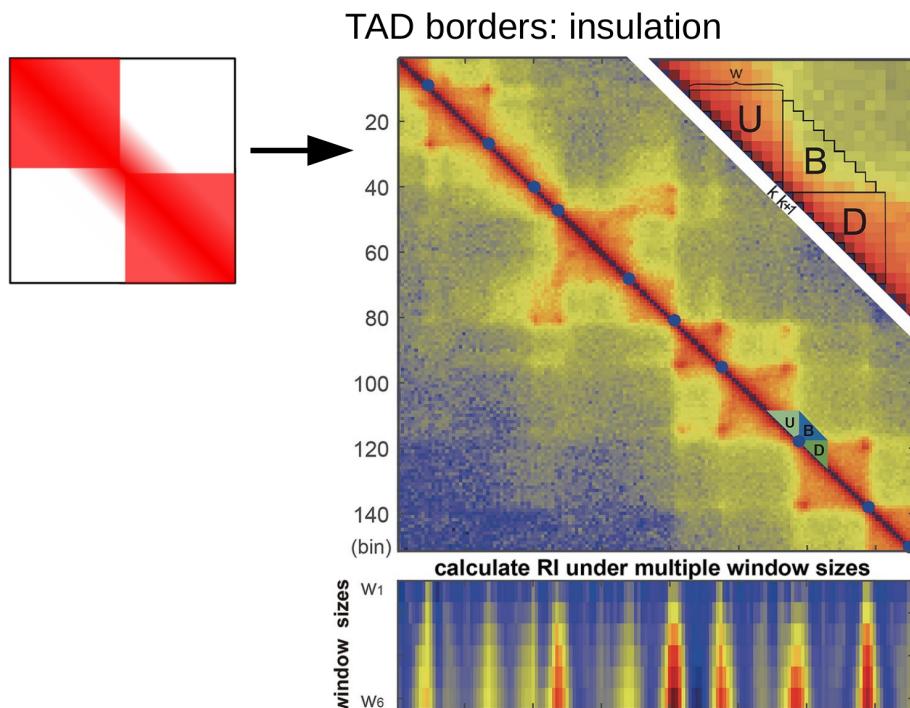
Chromatin interaction domains are important for gene regulation:

- form compact neighbourhoods of co-regulated genes
- Domain borders prevent interactions with elements outside
- TAD boundary disruption results in gene deregulation ([Lupianez et al., 2015](#))

Insulation score

Insulation: Contact depletion between domains

Insulation can be quantified with a numeric score.

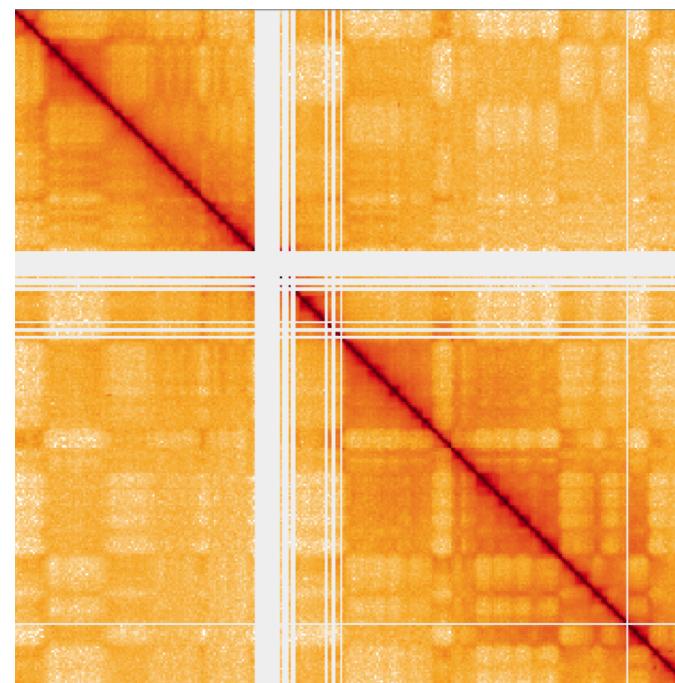


Chen et al., 2018

Chromatin compartments

Active and inactive chromatin is usually classified into A/B compartments

In Hi-C those compartments appear as a plaid-like pattern.

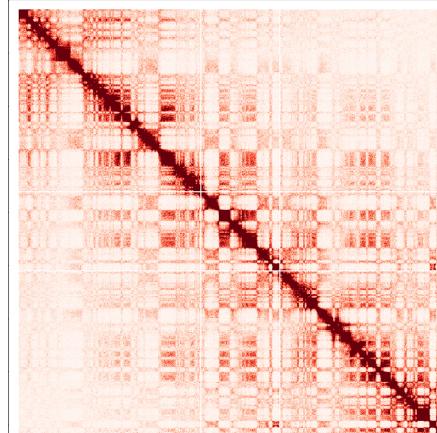


Dataset [4DNESYTUBW2E](#) from Oksuz, Yang et al., bioRxiv 2020.

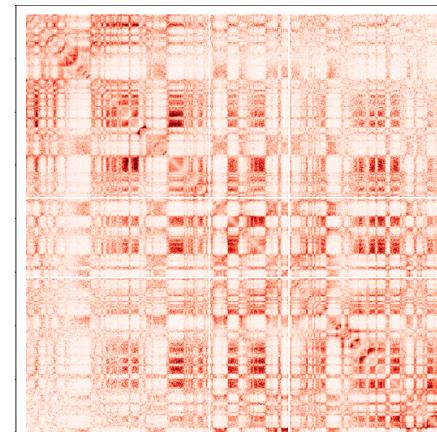
Compartments eigenvector

- Most common method to identify compartments: PCA on the Hi-C matrix
- Eigenvectors explaining the most variance will contain compartments.
- Must be validated with an external correlated signal (e.g. GC%)

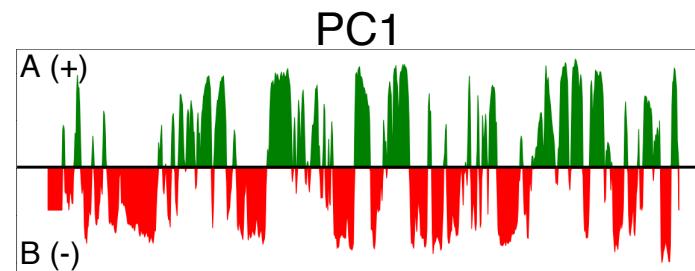
Compartments eigenvector



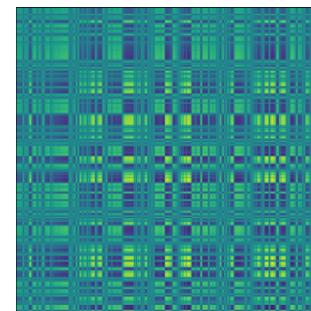
Obs / Exp



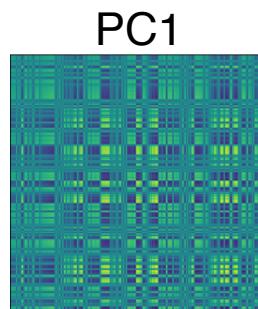
PCA



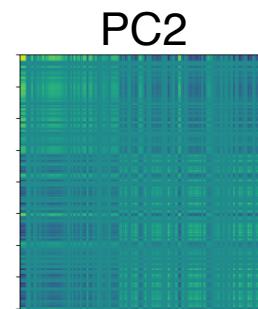
Phase (orient) PC



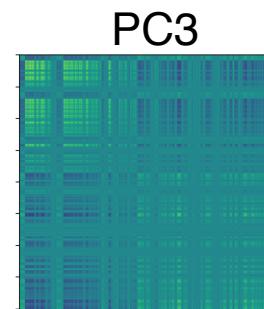
Select PC



PC1



PC2

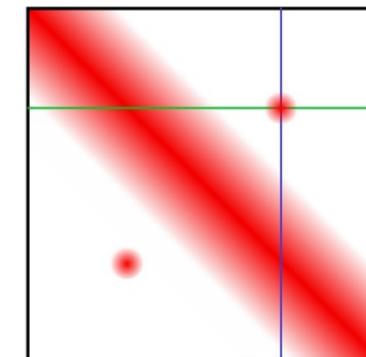


PC3

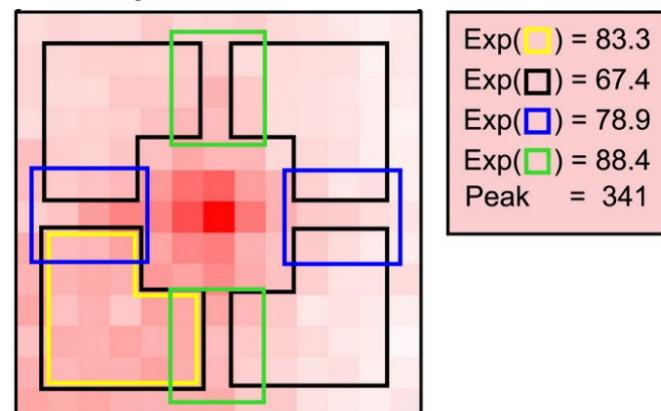
Feature detection

Often we want to automatically find *where* changes are happening in the genome, such as:

- Chromatin loops
- Domains (CID, TADs, ...)
- General contact intensity changes



Loops: local enrichment



Adapted from [Rao et al., 2014](#)"

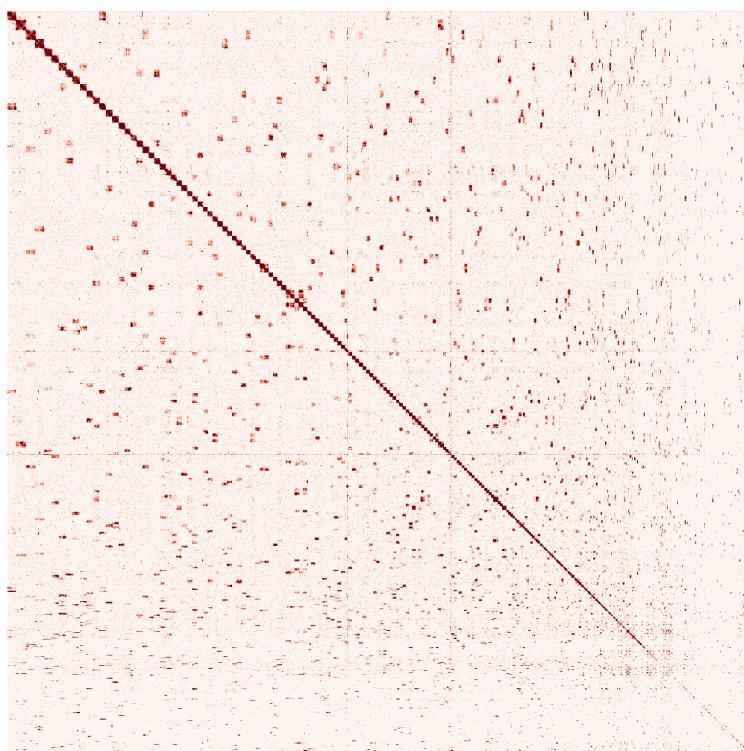
Limitations of Hi-C

Despite all its uses, Hi-C has several limitations to gain biological insights:

- Resolution limited by restriction enzyme and coverage
- Limited to 2-ways interactions. (methods in development)
- No absolute quantifications, counts are relative

Alternative uses of Hi-C

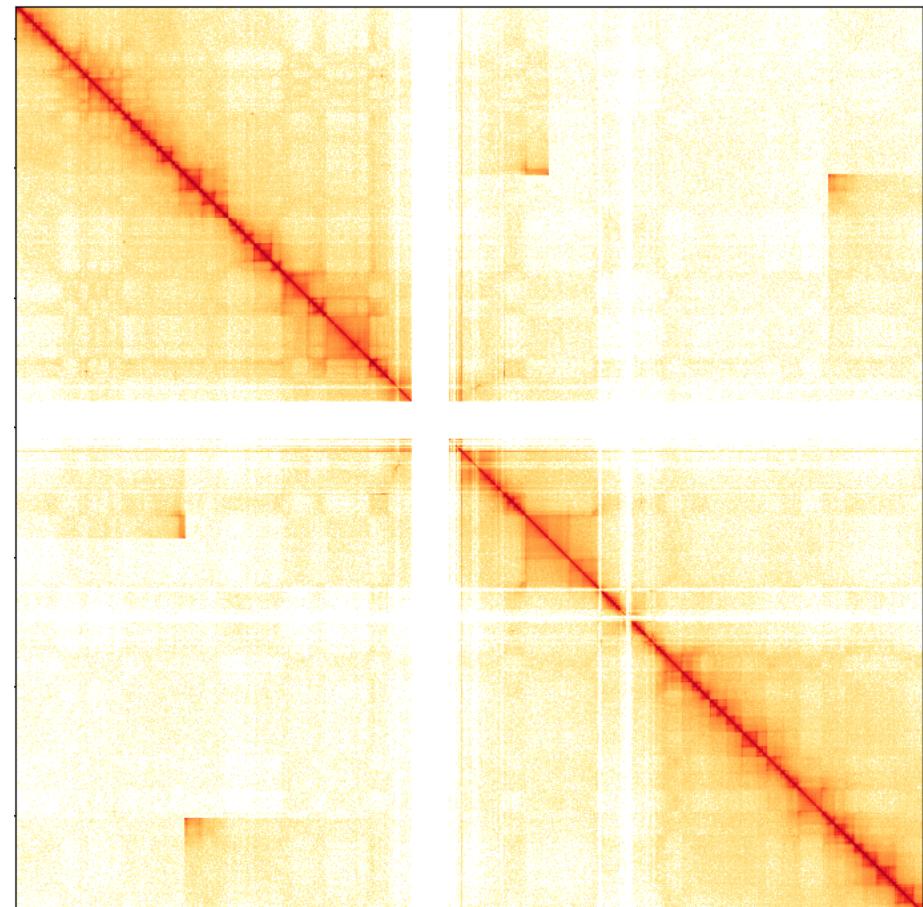
- Scaffolding genomes: Distance-dependent contacts
 - [instaGRAAL](#)
 - [3D-DNA](#)



Baudry et al., 2019

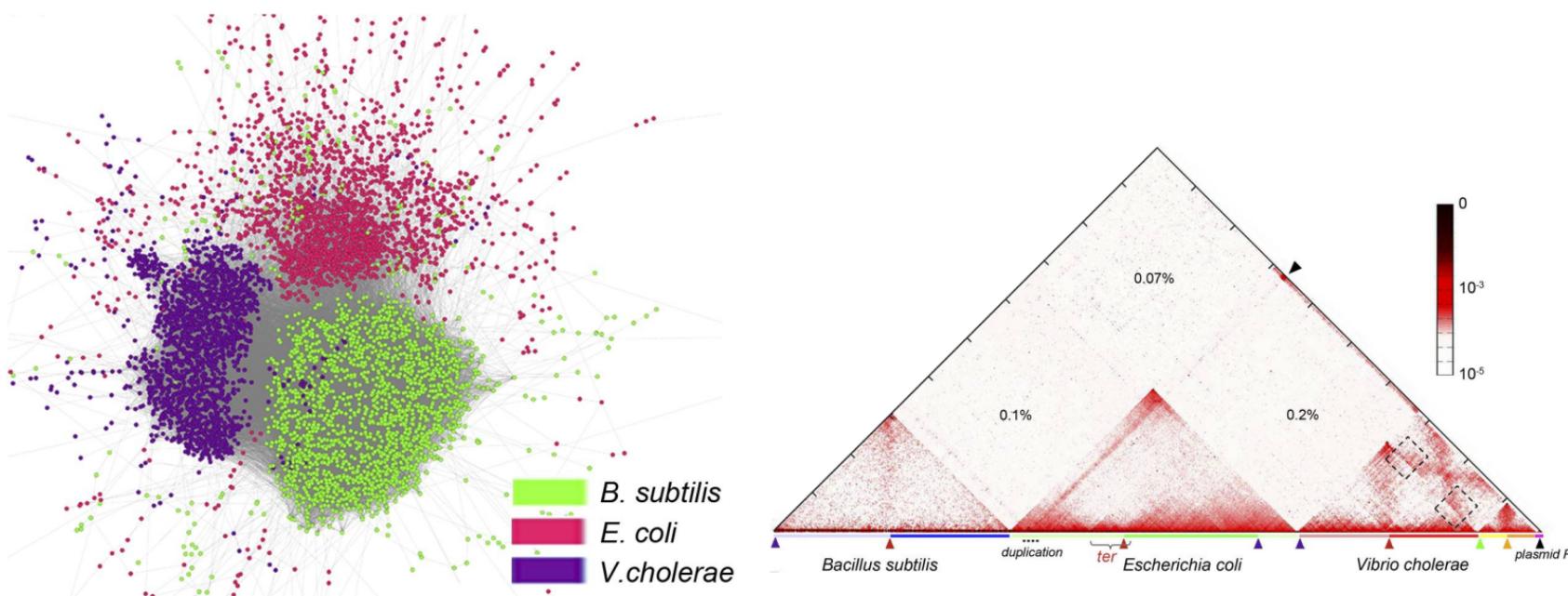
Alternative uses of Hi-C

- Structural variant detection
(deletions, inversions, ...)
 - [Hint](#)
 - [HiCnv](#)



Alternative uses of Hi-C

- Resolving species from metagenomic assemblies:
 - MetaTOR: Metagenome binning using Hi-C.
 - Strain3C: Strain-level genomes resolution using Hi-C.



Marbouty et al., 2014

Alternative uses of Hi-C

- Phasing haplotypes: Stronger cis- than trans-contacts
 - FALCON-Phase (Pacbio + Hi-C)
 - HapCUT2 (Shotgun, Nanopore, Pacbio, 10X, Hi-C)

- Use python scripting to interact with Hi-C data
- Extract quantitative signal from contact maps
- Feature detection on Hi-C

Exercises