

Team SB1 Paris

Maxime Queudre, Mikal Daou, Oktawia Scibior, Victor Molodij

Detecting chromatin compartments.

Input: Hi-C matrices from different cell lines data from all chromosomes at 25 kb and 100 kb res.

Methodology:

1 step: Data preparation protocol (binning, filtering, SCN, division O/E, correlation matrix)

2 step: Signal extraction: 1st eigenvector or HMM

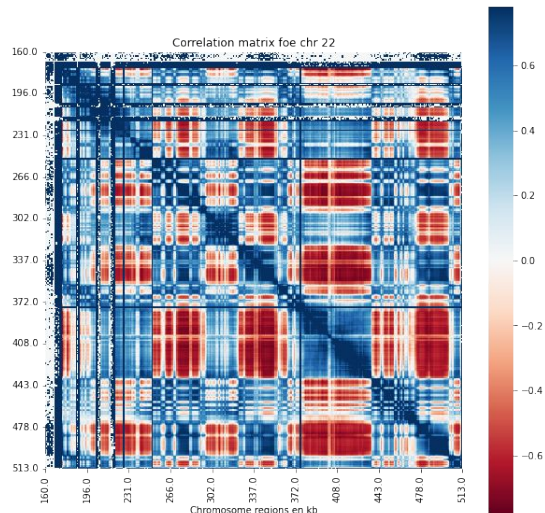
Indication of 2 compartments: **active (euchromatin)** , **inactive(heterochromatin)**

Output: Vector of chromosomal regions with corresponding compartment type.

Questions about:

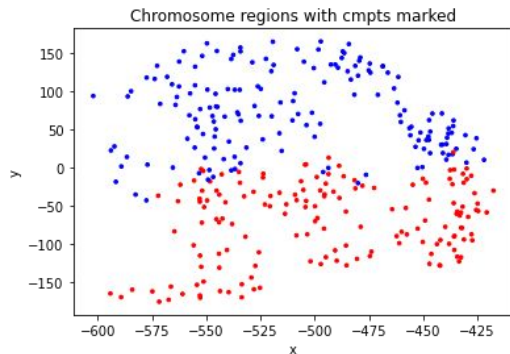
- increasing number of compartment
(HMM with more than 2 hidden states, 3D visualisation , epigenetic marks analysis ?)
- increasing of resolution ?
- analysis of inter chromosomal contacts ?

Results - Chromosome 22

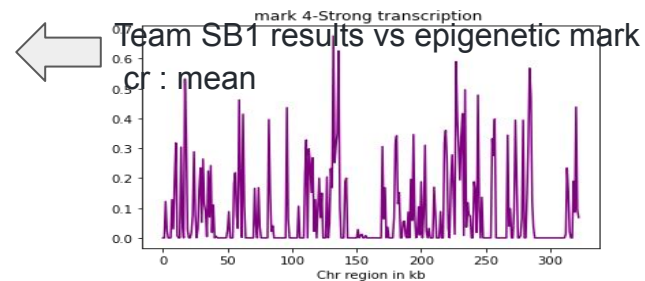
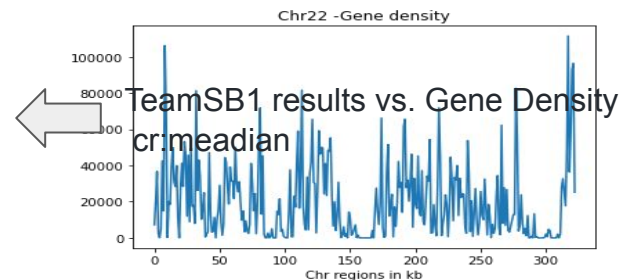
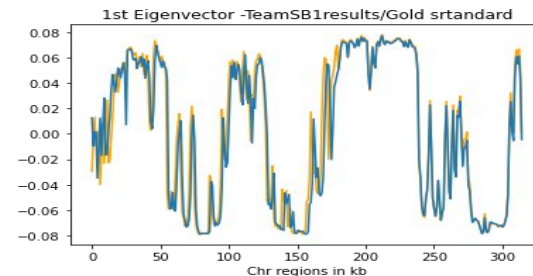


| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| Active | 0.93 | 0.83 | 0.88 | 150 |
| Inactive | 0.86 | 0.95 | 0.90 | 165 |
| weighted avg | 0.89 | 0.89 | 0.89 | 315 |

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| Active | 0.70 | 0.71 | 0.70 | 158 |
| Inanactive | 0.72 | 0.70 | 0.71 | 165 |
| weighted avg | 0.71 | 0.71 | 0.71 | 323 |

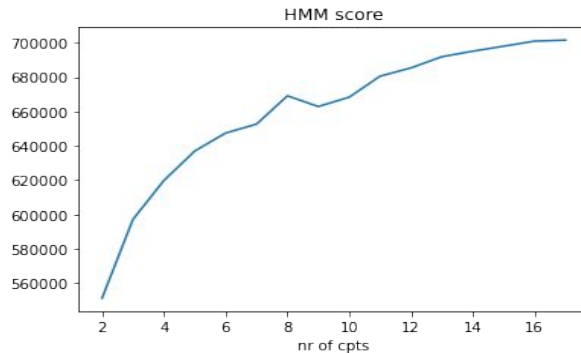


| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| Active | 0.59 | 0.83 | 0.69 | 158 |
| Inactive | 0.74 | 0.45 | 0.56 | 165 |
| weighted avg | 0.67 | 0.64 | 0.63 | 323 |



Challenges:

- **Technical** : memory issues, code implementation issues
- **Theoretical**:
 - understanding of data preparation and signal extraction pipeline especially the role of correct filtration
 - correct interpretation of gene density and epigenetic marks (setting correct thresholds)
 - Still open questions about: increasing the resolution and defining the best number of compartments ? HMM?



?

