Team SB2 – Digest

SORBONNE UNIVERSITÉ

Chromosomes's Compartments Detection
A Contact, Epigenetic and Consensus Approach

Team Members : Cédric Cornede, Hamid Hachemi, Damien Legros, Rouquaya Mouss, Arnaud Quelin

Method

In this work, we present our approach to predict genomic compartments from two different approaches, based on HMM, which we then confront and combine to obtain the most fine analysis possible. These two approaches are HMM Contact Prediction and HMM Expression-Repression Prediction as described in the floowing worflow:

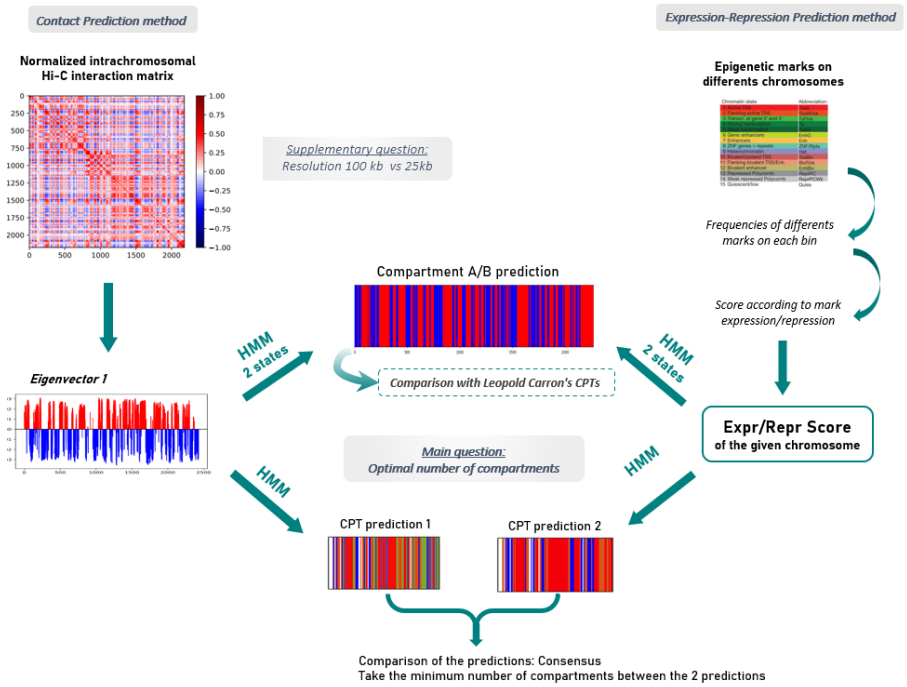

Fig. 1 : Worflow analysis

For the HMM Contact Prediction method, we use the Hidden Markov model in the aim to determine a number of compartments ranging from 2 to 16. Following the process that brought us to the intrachromosomal correlation matrices for each chromosome, we extract the first eigenvector. The optimal number of compartments was then determined from the HMM score.

The Expr-Rep Prediction method is also based on the Hidden Markov model but this time using epigenetic marks. From these frequencies of the different epigenetic marks we then determine an expression/repression score for a set of bins along the chromosomes. We then train our HMM in order to determine, as for the previous method, a number of compartments ranging from 2 to 16 from the HMM score.

At this stage we obtained a prediction of the compartments with two different methods. We then consider a consensus of our methods. The goal is to see how similar our methods are. We realized a function allowing to reach a consensus of our predicted compartments. The consensus is considered to be the association of labels (compartment number) between the two methods which gives the best similarity score.

Results

First, for the two prediction methods applied to all chromosomes for all cell lines, we produced a similarity score with the compartments predicted by L. Carron. We first noticed that the similarity score is higher for the HMM

Contact method than for the HMM Expr-Repr method (Fig. 1B). This is expected because this method is built on the model of the one used for target predictions. We represent a median case of similarity (Fig. 1A). Nevertheless, the Expr-Repr Prediction method achieves good scores in some cases and brings an interesting approach for our analysis.

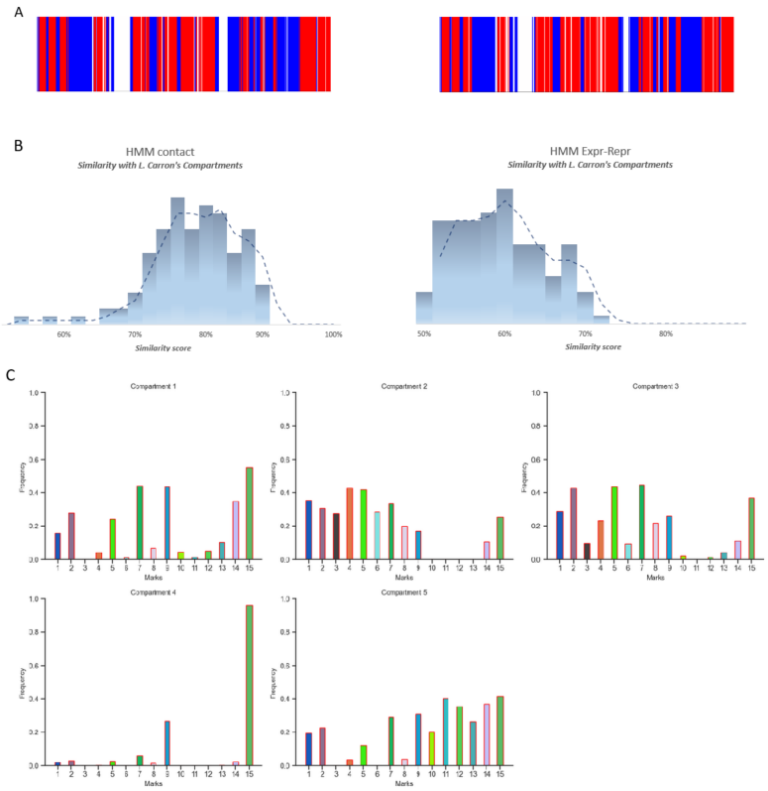

Fig. 2 : Main results : A - Example for a median case of similarity with L. Carron. B- Distribution of the simalrity for the two methods. C - Frequency of epigenetic marks with the Warsaw Team 1 method

We then consider the optimal number of clusters. This number is higher with the HMM contact method compared to the HMM Exp/Repr method. The HMM Contact Prediction method tells us about the spatial compartmentalization of the chromosome in the nucleus. The Expr-Repr Prediction method provides information on the expression-repression compartmentalization of the chromosome, and therefore indirectly on the spatial compartmentalization. With the HMM Expr/Repr method, it can be observed that in most cases, chromatins have two distinct states: heterochromatin and euchromatin. But when this is not the case, we notice that the compartmentalization is carried out in five states that we can connect with different known states of chromatin. We were then able to perform a visualization of our predictions with PYMOL which allowed us to observe the consensus with regard to the prediction methods.

We then compared our results to the approach of our partner team Warsaw CPT Team 1 who also used an HMM to predict the compartments on the distance matrix for 100 kb. Warsaw Team 1 determined and plotted the frequency of epigenetic marks of their predicted compartments obtained with their methods.
In Fig. 1C, we can observe that the compartments have different mark profiles, so in this case epigenetic marks can help us to determine a significant number and significant limits of compartments.

We also made our predictions with a resolution of 25kb. When we consider this resolution, we realize that our Expr-Repr prediction method and the consensus no longer have added value because we have the same input of the epigenetic marks and with and with the preliminary results we have considered, it seems that the Contact prediction method allows us to identify TADs.