Detecting and comparing genomic compartments

Cyril KURYLO, Sylvain FOISSAC and Matthias ZYTNICKI

Genomic compartmentalization is a biological factor affecting cell functionality. Analysis of data produced by the Hi-C protocol reveals compartmentalization of chromatin in the nucleus, which can vary as a tissue develops. Today, existing methods to detect genomic compartmentalization are limited in at least one of the following ways: detecting compartments qualitatively with no confidence measure, ignoring experimental biases, and/or dismissing replicate variability.

We propose an improvement over existing methodology to detect compartments and compare compartmentalization between conditions. First, we properly correct the diverse technological and biological biases inherent to Hi-C data. Then, we use an unsupervised learning method, constrained k-means, to detect compartments from normalized data. This method enables us to produce quantitative “concordance” values for each genomic region in each replicate, supporting our compartment predictions. Finally, we use these concordance values for differential analysis of compartmentalization between conditions. From their distributions, we obtain p-values revealing the significance of each predicted compartment change.

The method was implemented in an R package available on github.com/mzytnicki /HiCDOC, and was validated with Hi-C data originating from muscles of fetal pigs. Our data consists of three biological replicates at 90 days of pregnancy and three biological replicates at 110 days of pregnancy. The detected compartment changes open a way towards a better understanding of neonatal mortality affecting piglets.