3D transmap. A package to analyse associations between transcriptomics and 3D genomics

Projet long Proposition – M2BI 2018

Costas Bouyioukos

Team PTER, CNRS UMR7216 Epigenetics and Cell Fate

Goal

To develop a full python application able to analyse the spatial correlations and associations between 3D genome organisation and gene transcription. Provide and interactive 3D visualisation of the genome.

Introduction

Genome conformation capturing techniques (from 3C to Hi-C) have lighted the way genomes are organised in the three-dimensional space (3D). At the same time, advances in next generation sequencing (NGS) and its application to transcriptomics (RNA-seq) allow to access the activity of whole transcriptomes with elevated accuracy.

Genome architecture both affects and is affected by transcriptional activity, long distance regulatory elements (enhancers) are reported to regulate, in a fine tuning manner, the transcription of a large amount of genes in eukaryotic genomes.

On that epigenomic level of gene regulation the proposed concept of transcription factories, that is the local concentration of transcriptional activity in 3D hotspots, is playing a key role.

We have introduced a method, relatively easy to its use and interpretation, that allows the genome wide study of the associations between genome architecture, transcriptional activity and genome regulation and with this project I want to develop a proper python application.

Project

We will obtain 3D genome positions and the gene expression tables of *Plasmodium falciparum* and *Saccharomyces serevisiae* from [1] and [2]. We then use that as the only required input of our 3D transcriptome mapping python package. The calculations begin by the computation of the transcription correlation matrices (Pearson, Spearman and geometric) of each gene in these genomes and its n (user specified) neighbours. The computation of the same matrices for the 1D genome coordinates follows. The analyses continue with the detection of "positional sensitive" genes using a clustering algorithm and

the generation of interactive plots of the 3d transcription atlas of the whole genome using the `matplotlib` python package.

The software will report the 3D interactive visualisation of the three-dimensional map of transcriptional activity based on the spatial correlations of each gene in the genomes of *P. falciparum* and *S cerevisiae*. These visualisations can directly lead to discoveries of clusters of transcriptionally connected genes in the 3D space. We further investigate the "positional sensitivity" of each gene's transcription activity and separate and cluster genes of the studied organisms into three distinct categories based on their positional sensitivity.

References

[1] Ay, F.; Bunnik, E. M.; Varoquaux, N.; Bol, S. M.; Prudhomme, J.; Vert, J.-P.; Noble, W. S. & Le Roch, K. G.
Three-dimensional modeling of the P. falciparum genome during the erythrocytic cycle reveals a strong connection between genome architecture and gene expression.
Genome Research, 2014, 24, 974-988

[2] Duan, Z.; Andronescu, M.; Schutz, K.; McIlwain, S.; Kim, Y. J.; Lee, C.; Shendure, J.; Fields, S.; Blau, C. A. & Noble, W. S.
A three-dimensional model of the yeast genome.
Nature, 2010, 465, 363-367

[3] Hunter, J. D.

 Matplotlib: A 2D graphics environment

Computing In Science & Engineering, IEEE COMPUTER SOC, 2007, 9, 90-95