Project description – 21/2/2018

Title: Sample specific enhancer-promoter network inference

Background: Deciphering the regulatory role of the noncoding part of the human genome is a major challenge. With the completion of the sequencing of the genome, efforts have shifted over the last decade towards understating the epigenome. These efforts aim at understanding regulatory mechanisms outside the protein-coding sequences that allow the production of thousands of different cell types from the same DNA blueprint. Consequently, large-scale epigenomic projects set out to identify all the cis-regulatory elements (e.g., enhancers and promoters) that are encoded in the genome. Prominent among them are the ENCODE and Roadmap consortia. Given the plethora of candidate enhancer regions called by these projects, the next pressing challenge is to identify which of them is actually functional and map them to the genes they regulate. We developed FOCS, a machine learning based method that infers global enhancer-promoter (E-P) links across many cell types. Our next goal is to use these global E-P links to identify cell-type specific E-P maps. Such identification could help developing treatments targeting sample-specific enhancers instead of their regulated genes.

Goal: A high quality predictor of cell-type specific enhancer-promoter interactions.

Details: In this project we will apply machine learning (ML) methods for inferring sample-specific enhancer-promoter (E-P) links based on large scale high-throughput sequencing data. The data covers diverse cell types, including cancer ones, across many samples. We will combine many different data sources such as RNA-seq, DNase-seq, ChIP-seq, and 3D chromatin capture (ChIA-PET, HiChIP) techniques. Project steps will include (1) download of the data and quality analysis and filtering; (2) application of state of the art ML methods to infer global E-P links across all cell types; (3) inference of cell-type specific E-P links using more sophisticated ML methods; (4) extensive testing and validation of the performance on multiple datasets. This project will provide a deep understanding of the regulatory networks far beyond promoters and on the genome structure.

Prerequisites: R programming, passion to statistics and familiarity with genomics and molecular biology.

Project leader: Tom Hait

Reading material: Review on enhancers and promoters [1], FANTOM5 enhancer atlas [2], FOCS method for inferring global E-P links [3], and JEME method for inferring specific E-P links [4].

1. Shlyueva D, Stampfel G, Stark A. Transcriptional enhancers: from properties to genome-wide predictions. Nat. Rev. Genet. 2014;15:272–86.

2. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, et al. An atlas of active enhancers across human cell types and tissues. Nature. 2014;507:455–61.

3. Hait TA, Amar D, Shamir R, Elkon R. An extensive enhancer-promoter map generated by genome-scale analysis of enhancer and gene activity patterns. doi.org [Internet]. Cold Spring Harbor Laboratory; 2017 [cited 2017 Sep 24];190231. Available from: https://www.biorxiv.org/content/early/2017/09/18/190231.1

4. Cao Q, Anyansi C, Hu X, Xu L, Xiong L, Tang W, et al. Reconstruction of enhancer – target networks in 935 samples of human primary cells , tissues and cell lines. 2017;