

Analysis of differential transcription factor occupancy in leukaemic cells

Background

Each cell type in the human body expresses a different set of transcription factors. These are proteins that can read the sequences of the DNA and then cooperate to switch genes on or off. This process is highly regulated and lies at the heart of all developmental processes. In cancer due to specific genetic changes, normal transcription factor activity is disturbed, leading to the mis-expression of multiple genes and aberrant growth. Currently, we know very little about how different genetic changes cause different types of cancer. This project will address this question.

DNase-seq is a state-of-the-art approach to analyse the global pattern of protein-DNA binding in regulatory regions of DNA. In this assay, DNA is digested by DNaseI and high-throughput sequencing is applied to reveal the genome-wide pattern of DNA digestion in hypersensitive regions. Regions of DNA bound by regulatory proteins are protected from digestion and leave behind “footprints”, regions of reduced cut frequencies. A single DNaseI-seq data set provides a “snapshot” of the global occupancy pattern of regulatory proteins on DNA, revealing binding of a large number of different proteins at once which is highly informative. Our collaborators at Birmingham, Constanze Bonifer and Peter Cockerill, are using this type of data to study aberrant regulatory events in leukaemic cells.

Objectives

To use DNaseI data from leukaemia patients carrying different genetic changes, obtain footprinting data and compare these with normal cells and with each other. If time permits, this will be correlated with gene expression profiles.

Approach in mini-project

In this project we are going to make use of data sets where both DNase-seq and gene expression data are available. We are going to detect DNase hypersensitive sites (DHSs), footprints inside DHSs, and differences in footprints between different samples. We are going to check differential footprints for statistical links with differential expression profiles and over-representation of TF-binding motifs. Time permitting, we will analyse the combinatorial patterns of motif presence in these regions. All computations will be carried out using publicly available tools as well as tested software tools in our labs.

Prospects for follow-on PhD projects

In a PhD project we will expand this work into method development and study statistical approaches to infer a more refined picture of re-arrangements of proteins and protein complexes binding to DNA.

References

The following paper is describing a recent method for the analysis of single DNase-seq data sets that we published recently. It has become very popular and we know that it is being used by a number of leading labs in the field. We are keen to build on the current momentum to establish this tool and related tools we are developing in the future as a widely used means for DNase-seq data analysis.

<http://nar.oxfordjournals.org/content/41/21/e201.full?keytype=ref&%2520ijkey=C5ERIRQ7eNX7whC>

This paper explains how such a method can be used in leukaemic research:

<http://www.sciencedirect.com/science/article/pii/S2211124714006871>