Data driven analysis of epigenetic landscapes using Boolean logic

All organisms constitute a great variety of biological states and responses on molecular level to continuously changing intra/inter-cellular and environmental signals. Thus, despite its limitations, probing the gene regulation is one of the most visited challenges of genome biology. Analysis of gene regulation using Boolean logic by transforming gene expression levels to Boolean space and using genes as Boolean variables is one way of tackling this great biological variety. In an analogy to circuits and logic gates such analysis proved useful in investigating gene regulation networks of relatively small groups of genes. Nevertheless a system level analysis/understanding is still emerging. Once moved into Boolean space, an opportunity for a system level approach is arising: assessing the epigenetic landscape (Waddington landscape) constituted by a totality of gene expression machinery. However the number of states that can be assumed by a cell is huge even in Boolean space due to the high number of genes ($\sim 2^{10k}$). On the other hand this task is rendered possible by using clustering approaches and by using modules of genes instead of single genes in the analysis. This project aims to carry out a system level analysis of gene regulation in Boolean space and finally constructing the epigenetic landscape from gene expression data. To this end project brings forward seven main steps,

i) identifying and collecting high quality/clean rnaseq data from public repositories,
ii) analyzing rnaseq data and assigning gene expression levels,
iii) identifying significant modules of genes,
iv) identifying logical relations between these modules,
v) carrying out simulations of epigenetic landscapes using these logical relations as updating rules for the gene expression system,
vi) investigating the resulting epigenetic landscape, identifying critical states and paths,
vii) going back to data, scrutinizing results from simulations, investigating correspondence between states/path from simulations and phenotypic properties and finally biological reasoning of the results.

These steps encapsulate several basic bioinformatics tasks/methods, i.e. rnaseq/gene expression analysis, clustering, logical analysis, network/graph analysis. Furthermore a pipeline that works on unix systems is aimed, which combines all these mentioned steps.

For further information

Dr. rer. nat. Mehmet Volkan Cakir

vlknckr@gmail.com

Rudolf-Schönheimer-Institute für Biochemie
Medizinische Fakultät - Universität Leipzig
Johannisallee 30
2.OG, Raum D 202/203
04103 Leipzig