
Identification of deregulated transcription factors in bladder cancer

Magali Champion^{*1}, Julien Chiquet², Pierre Neuvial³, and Etienne Birmelé¹

¹Laboratoire de Mathématiques Appliquées (MAP5) – Université Paris V - Paris Descartes – France

²Mathématiques et Informatique Appliquées (MIA) – AgroParisTech – France

³Institut de Mathématiques de Toulouse (IMT) – CNRS : UMR5219 – France

Résumé

Transcription factors (TFs) are key genes involved in the regulation of gene expression and commonly deregulated in the pathogenesis of human cancer. Identifying deregulated TFs and understanding their causal pathways thus remain one of the major challenge for cancer therapy. Interactions between genes are usually represented through Gene Regulatory Networks (GRNs), which give an overview of the existing regulation processes between TFs and target genes.

We developed a four-steps strategy for the identification of deregulated TFs involved in specific subtypes of cancer. The true GRN being unknown, we first inferred a GRN, which could be used as a reference for regulations between genes using hLICORN (Elati et al. 2007). This method aims at learning bipartite networks in which co-regulator TFs act together to activate or repress a target gene by heuristic approaches. To make the GRN more realistic, we corrected our model by removing the effect of genomic alterations, such that copy number variations (CNV), on gene expression (Delatola et al. 2015). We then computed a deregulation score for each target gene and each sample using the methodology of Picchetti et al. (2015). In few words, each gene of the model is associated with a hidden status (ternary variable indicating the level of expression) and is allowed to be randomly deregulated. An EM-strategy is then used for parameter inference. The final step consists in finding the most significant TFs involved in the deregulation of the target genes across all samples. A multi-linear model representing the deregulation importance of each TF in each sample is thus cast and a threshold of significance is set to maximize the classification of the samples in subtypes using Linear Discriminant Analysis.

We applied our method to the bladder cancer data set (Nicolle et al. 2015), which included a set of 179 bladder cancer samples with both expression and copy number data for 18,671 genes split into a set of TFs and target genes. This leads to the identification of eleven TFs involved in specific subtypes of bladder cancer.

*Intervenant