Modelling normal cells identifies master regulators in cancer

Rémy Nicolle[1,2], Mohamed Elati[1], Jennifer Southgate[3], and François Radvanyi[2]

[1] iSSB, Institute of Systems and Synthetic Biology, CNRS FRE3561, University of Evry-Val-d'Essonne, Genopole Campus 1, Genavenir 6, 5 rue Henri Desbruères, 91030 Evry, France,
(remy.nicolle, mohamed.elati)@issb.genopole.fr,
[2] Institut Curie, CNRS UMR 144, 26 rue d'Ulm, 75248 Paris,cedex 05, France
francois.radvanyi@curie.fr
[3] Jack Birch Unit of Molecular Carcinogenesis, Department of Biology, University of York, United Kingdom.
js35@york.ac.uk

Motivation

Tumor cells display many functions possessed by their normal counterparts. Their ability to migrate, to proliferate, to attract new vessels and to exist in various differentiation states are properties also found in normal tissue during wound healing. Following tissue injury normal cells can operate these processes in a tightly regulated and coordinated manner leading to the healing of the wound.

We propose an original systems biology approach to identify and analyse the regulatory networks found in the normal states to then assess whether networks of the normal regenerative process are specifically maintained or altered in the tumor state. This strategy was applied to bladder cancer, a cancer derived from the bladder urothelium, because normal bladder urothelium can be grown in culture at various different stages of proliferation and differentiation thereby mimicking wound healing.

Constructing the network of normal proliferation and differentiation

Gene expression data of primary Normal Human Urothelium (NHU) non cancerous primary cell cultures in various states of differentiation and proliferation was considered as an *in vitro* model of wound healing and used to infer a large regulatory network. We applied LICORN[1], a data mining algorithm introduced by our team that infers the targets of transcription factors from genome wide expression data. LICORN was shown to be suitable for cooperative regulation and to scale up to the complexity of mammalian transcriptional networks. LICORN is able to find the set of Transciption Factors that cooperatively regulate the expression of a given gene. Furthermore, LICORN was previously applied[2] in yeast to infer a large regulatory network from gene expression data. The authors showed that the clusters of genes extracted from the inferred regulatory network had a higher functional enrichment than clusters based solely on gene expression.

Additionally to expression based information, the inferred normal regulatory network, comprising approximately 5000 genes and 400 co-regulators, was enriched with systematic promoter sequence analysis of known transcription factor binding sites model, public ChIP-chip and ChIP-Seq data as well as Protein Protein interactions between co-regulators.

Note that the rest of our approach is not dependent on the network inference method. Aside from the co-regulation information inferred by LICORN, any method able to infer large-scale regulatory networks such as ARACNE[3] or GENIE3[4] could be used.

Measuring context specific regulation activity

The concept of using the knowledge over the network structure was shown to be successful at identifying key regulators of specific phenotypes and processes [5, 6]. However, we were interested in a data transformation approach in which neither a predefined gene-signature nor sample classification was needed to identify central regulators.

In order to identify key regulators and to quantify their impact on their regulatory programs, we propose to measure the influence of regulator genes on their targets in a given sample. The idea is to be able to quantify the extent to which a Master Regulator (MR) is active on it's target genes in a given sample or set of samples. The measure is based on the divergence between the expression of the set of activated and repressed target genes of a given regulator in a given sample. The basic idea is that if a set of genes is effectively activated

and another repressed by the same MR, and that this MR is active, the activated set of genes should be over expressed and the repressed set should be under expressed. Therefore the more a MR is active on given sets of targets (activated and repressed) the greater the distance between these sets will be. Measuring this divergence will give an idea on the activity of a MR in a given sample, or set of samples, and more importantly on the pertinence of the structure of the network.

Interestingly, when measured for each regulator in each sample, the measure of regulatory influence produces a data set with the same number of samples but a reduced number of features representing the master regulators activity. Therefore, we proposed[7] to use this measure in the context of classification and feature extraction. We showed that the transformation of the data through the regulatory activity greatly improves the stability and robustness of models trained in different datasets.

Regulatory influences underline function of Master Regulators in normal and cancer cells

The regulatory network inferred from the NHU data pointed out several previously described regulator of normal urothelial differentiation and their validated gene targets. Alongside to these known MR, the computation of the influence characterized new MR as well as their involvement in normal urothelial differentiation, proliferation and growth arrest.

In an identical way the influence of the normal regulators was measured in 3 cohorts of 60, 120 and 180 bladder cancer transcriptomes (respectively from Stransky et. al. [8], the TCGA consortium and a private unpublished data set). In order to estimate to what extent the normal regulation of growth is conserved in tumor cells, the global influence (defined as the sum of squared influence in all samples for all MR) of the normal network was compared to the influence of 1000 randomly generated networks with similar topology (each regulator with the same number of target genes). The global influences were compared and shows that the normal network is significantly conserved, more influent, than any random network (150 times the standard deviation above the mean).

An analysis of the influence of normal regulators in bladder cancers pointed out a major loss of function of Master Regulators of urothelial differentiation. This loss of differentiation was also observed in the co-regulatory network in which known and novel regulators of normal differentiation form a dense network of cooperative regulators and are virtually all lost in most bladder tumors. Additionally to these results, several MR show the same activity profile in some bladder tumors than in the proliferating NHUs suggesting that the regulation driving normal proliferation is maintained during tumorigenesis.

References

1. M. Elati, P. Neuvial, M. Bolotin-Fukuhara, E. Barillot, F. Radvanyi, and C. Rouveirol. LICORN: learning cooperative regulation networks from gene expression data. *Bioinformatics*, 23(18):2407–2414, Sept. 2007.
2. E. Birmelé, M. Elati, C. Rouveirol, and C. Ambroise. Identification of functional modules based on transcriptional regulation structure. *BMC Proc*, 2 Suppl 4:S4, 2008.
3. A. A. Margolin, I. Nemenman, K. Basso, C. Wiggins, G. Stolovitzky, R. Favera, and A. Califano. ARACNE: An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context. *BMC Bioinformatics*, 7(Suppl 1):S7, 2006.
4. V. A. Huynh-Thu, A. Irrthum, L. Wehenkel, and P. Geurts. Inferring Regulatory Networks from Expression Data Using Tree-Based Methods. *PLoS ONE*, 5(9):e12776, Sept. 2010.
5. M. S. Carro, W. K. Lim, M. J. Alvarez, R. J. Bollo, X. Zhao, E. Y. Snyder, E. P. Sulman, S. L. Anne, F. Doetsch, H. Colman, A. Lasorella, K. Aldape, A. Califano, and A. Iavarone. The transcriptional network for mesenchymal transformation of brain tumours. *Nature*, 463(7279):318–325, Dec. 2009.
6. C. Lefebvre, P. Rajbhandari, M. J. Alvarez, P. Bandaru, W. K. Lim, M. Sato, K. Wang, P. Sumazin, M. Kustagi, B. C. Bisikirska, K. Basso, P. Beltrao, N. Krogan, J. Gautier, R. Dalla-Favera, and A. Califano. A human B-cell interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. *Molecular Systems Biology*, 6:1–10, June 2010.
7. R. Nicolle, M. Elati, and F. Radvanyi. Network transformation of gene expression for feature extraction. *11th International Conference on Machine Learning and Applications (ICMLA)*, pages 1–6, 2012.
8. N. Stransky, C. Vallot, F. Reyal, I. Bernard-Pierrot, S. G. D. de Medina, R. Segraves, Y. de Rycke, P. Elvin, A. Cassidy, C. Spraggon, A. Graham, J. Southgate, B. Asselain, Y. Allory, C. C. Abbou, D. G. Albertson, J. P. Thiery, D. K. Chopin, D. Pinkel, and F. Radvanyi. Regional copy number–independent deregulation of transcription in cancer. *Nat Genet*, 38(12):1386–1396, Nov. 2006.