# Revealing mechanisms of cancer progression by pan-cancer deconvolution of tumoral transcriptomes

A. Zinovyev, A. Biton, I. Bernard-Pierrot, F. Radvanyi

*Institut Curie, 26 rue d'Ulm, Paris,* `andrei.zinovyev@curie.fr`

Transcriptomal profiling is one of the most used methods for characterizing a tumor sample in order to detect important biological differences between the functioning of tumor and normal cells, to precise the diagnosis and to suggest personalized treatment [1]. Recently, large-scale efforts (such as The Cancer Genome Atlas, TCGA, http://cancergenome.nih.gov) were undertaken to produce transcriptome profiling for a large number of tumor samples for various cancer types and to make these data available for analysis by statistical methods. Currently, we have access to several tens of thousands of tumoral transcriptomes, most of which are collected for the most prevailing cancers such as breast, colon, lung and ovarian cancers.

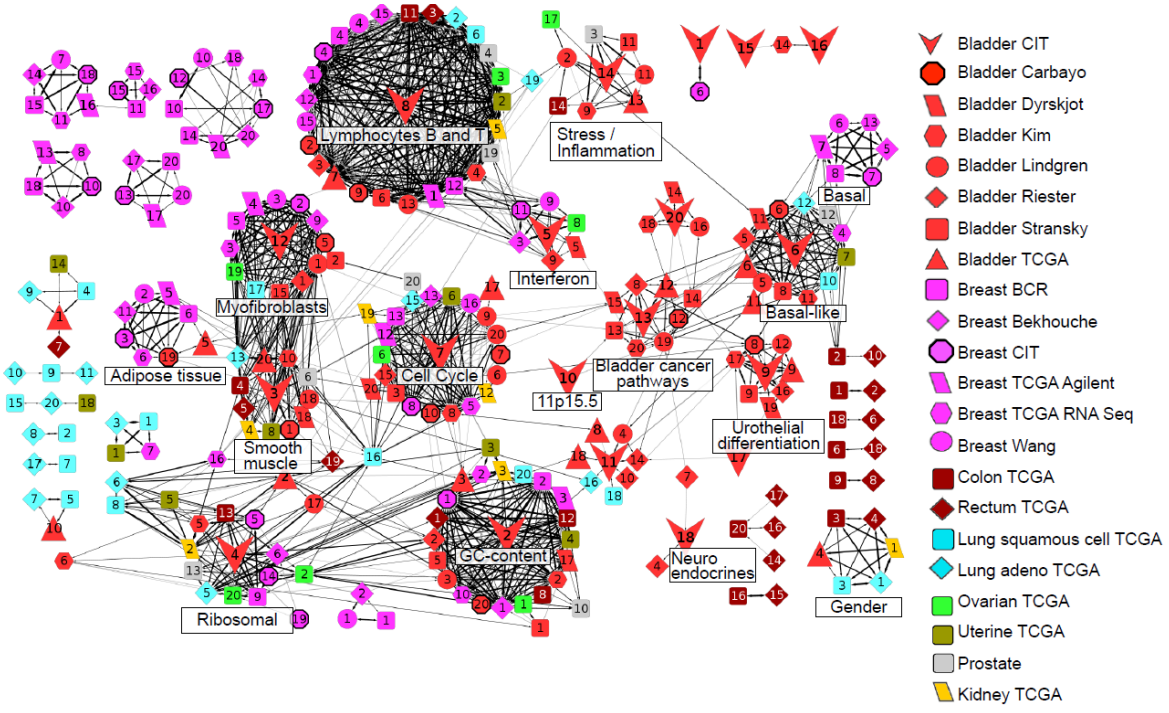


**Figure 1.** Correlation graph showing connections between independent components computed for 22 datasets collecting tumor sample transcriptomes in various types of cancer (6671 tumor samples are

considered altogether in all datasets). Each node is a component, the color denotes the type of cancer and the shape distinguishes a particular dataset. An edge connects two components if their gene projections are correlated with Pearson correlation larger than 0.35 by absolute value. A particular focus is given to bladder (red) and breast (pink) cancers. The detected tightly connected node communities show the cancer type-specific signals or the signals common for many cancer types. Reproduced from [3].

There exists an important question of what can be learned from these "big data": in particular, what kind of signals shape the transcriptomes in many cancer types or are specific to a particular cancer type. We've approached this question by deconvoluting into independent components 22 different datasets collecting transcriptomal profiles for various types of cancer, focusing at comparing bladder and breast cancer in more details. For each dataset, we've computed 20 components which seemed to be sufficient to capture the most important biological or technical signals. We've made particular effort to understand the meaning of the components for one particular dataset for bladder cancer, produced at Institut Curie (CIT bladder cancer dataset, where CIT stands for Identitity Card of a Tumor in French) [3].

We systematically compared all 440 components by computing the Pearson correlation between the values of gene projections onto each component, between the genes common for any pair of independent components. The resulting correlation graph is shown in Figure 1. We analysed this graph for existence of communities of tightly connected nodes, which we interpreted as highly reproducible signals. Those communities that contained the nodes of only one cancer type were considered as corresponding to the cancer type-specific signals, while those communities collecting components from cancers of different types were considered as intrinsically associated with cancer progression in general. Among such generic signals we've identified the changes in the transcriptome connected to infiltration of B- and T-lymphocytes, to the presence of myofibroblasts, to the cell cycle or to the functioning of mitochondria or translation. Among such generic signals we also identified a community associated with GC-content of the gene sequences, which can indicate to specific technological biases induced by preparation of RNA for quantification by PCR-based amplification. Among cancer type-specific signals, we observed a clear signal connected to

existence of the basal-like breast cancer subtype (one of the most aggressive). An adjacent pseudo-clique, which is composed from mostly bladder cancer components from different datasets, corresponds to the recently identified basal-like subtype of bladder cancer [4]. Overall, the correlation graph presented at Figure 1 shows the global landscape of the biological and technical signals shaping cancer transcriptome variability.

The urothelial differentiation component (CIT-9), found in all bladder cancer data sets studied, was specifically associated with bladder luminal tumors. We studied this component in more detail, both computationally and experimentally. First, we did not find any relationship between molecular differentiation measured on this component and morphological grade assessed by architectural organization and nuclear atypia. To identify protumorigenic genes potentially involved in the carcinogenesis of luminal tumors we looked for altered genomic regions associated with the urothelial differentiation component. Among these regions, we selected the regions of gains that were found in tumors associated with the differentiated side of the component and that contained a contributing gene associated with differentiation. We found that genomic alterations of PPARG were correlated with its expression in both tumors and bladder tumor-derived cell lines. To demonstrate the functional involvement of PPARG in bladder tumors, we studied the effect of siRNA-mediated PPARG knockdown on the growth of nine bladder cancer-derived cell lines with different levels of PPARG mRNA. The experimental results indicated that PPARG was involved in tumor cell growth in PPARG expressing bladder cancer cells. The presence of PPAR target genes among the contributing genes of the urothelial differentiation component suggested that PPARG could control the expression of several contributing genes of this component. We tested this hypothesis by comparing the transcriptome of the SD48 cell line treated with three different siRNAs targeting PPARG or with the transfection reagent (Lipofectamine).

**References:**

1. E. Barillot, L. Calzone, P. Hupe, J.-P. Vert, A. Zinovyev (2012) *Computational Systems Biology of Cancer*. Chapman & Hall, CRC Mathematical & Computational Biology, 452 p.

2. A. Zinovyev, U. Kairov, T. Karpenyuk, E. Ramanculov (2013) Blind Source Separation

Methods For Deconvolution Of Complex Signals In Cancer Biology. *Biochemical and Biophysical Research Communications* **430**(3) 1182–1187.

3. A. Biton, I. Bernard-Pierrot, Y. Lou, C. Krucker, E. Chapeaublanc, C. Rubio Perez, N. Lopez Bigas, A. Kamoun, Y. Neuzillet, P. Gestraud, G. Grieco, S. Rebouissou, A. de Reynies, S. Benhamou, T. Lebret, J. Southgate, E. Barillot, Y. Allory, A. Zinovyev, F. Radvanyi (2014) Independent component analysis uncovers the landscape of the bladder tumor transcriptome and reveals insights into luminal and basal subtypes. *Cell Reports* **9**(4), 1235–1245.

4. The Cancer Genome Atlas Research Network (2014) Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* **20** (507), 315–322.