Molecular Systems Biology 3; Article number 94; doi:10.1038/msb4100132
Citation: *Molecular Systems Biology* 3:94

www.molecularsystemsbiology.com

molecular systems biology

NEWS AND VIEWS

Higher order structure in the cancer transcriptome and systems medicine

Edison T Liu[1] and Thomas Lemberger[2]

[1] Genome Institute of Singapore, Singapore and [2] European Molecular Biology Organization, Heidelberg, Germany

Molecular Systems Biology 13 March 2007; doi:10.1038/msb4100132

A formal goal of gene expression studies in cancer is to reconstruct the pathways that characterize tumor behavior through its transcriptional signature (Liu, 2005; Liu *et al*, 2006). The ultimate utility is not only to discover molecular components of these cancers, but also to exploit this knowledge to improve therapeutics (Bild *et al*, 2006a). To accomplish this, a number of approaches have been used to achieve higher level order in the expression of gene sets (Mootha *et al*, 2003; Segal *et al*, 2004; Subramanian *et al*, 2005; Bild *et al*, 2006b). Recently, Tomlins *et al* (2007) made a further contribution to this effort by devising an approach to construct a molecular concept map (MCM) in the analysis of human prostate cancers. In this work, they microdissected 101 prostate cancers that ranged from benign epithelium to metastatic disease and assessed the expression of transcripts on a genome-wide scale using printed cDNA microarrays. Approximately 14 000 molecular 'concepts' representing biologically connected genes were used to analyze the expression profiles and identify a number of relevant pathways that might drive prostate cancer biology.

As in many good scientific articles, several layers of importance that are significant to different research communities can be found. Pertinent to this journal, this paper by Tomlins *et al* (2007) provides an instructive focal point for the discussion of fundamental problems that bedeviled systems biologists working on primary human cancers. For example, given the complexity of mammalian systems and the further genetic scrambling that takes place with cancer, is it possible to develop systems approaches that can solidly map, in cancer cells, interactions relevant to clinical cancer biology? Does MCM provide definitive answers to these problems? The honest answer must be no, but Tomlins *et al* (2007) do lay out solid ideas as to the direction of future work in systems medicine.

The primary contribution is the application of the MCM approach. 'Molecular concepts' are sets of genes related in a biologically meaningful way, but these concepts may reside in different planes of association. The molecular concepts that populate an MCM come in three forms that are heterogeneous in character: (1) gene and protein annotations from external databases (including from cell lines as well as primary cancers), (2) computationally derived regulatory networks and (3) microarray gene expression profiles. The external annotation included chromosomal locations, protein domains and families, molecular functions, cellular localizations, biological processes, signaling and metabolic pathways, protein–protein interaction networks and protein complexes. Presumably any elements that are consistently associated with each other would be candidates. The strength of such an approach is that it takes into account all available knowledge of genetic relationships pertinent to prostate cancer, and knowledge that is not restricted to only one technical platform.

Expression signatures resulting from the profiling of prostate cancer samples are then analyzed using this large collection of molecular concepts. More precisely, pairs of concepts are tested for significant association as indicated by enrichment and disproportionate overlap of the respective gene sets within the top-ranking genes in the expression signature. In this way, a map of connected concepts, the MCM, is constructed, which delineates the landscape of interlinked biological events and processes that underlie the expression signature analyzed.

The conceptual approach of deconvolving microarray expression profiles using predefined gene sets that integrate heterogeneous biological knowledge is not completely new and has been described by other groups and applied to human disease in approaches bearing related names such as 'gene set enrichment analysis (GSEA)' (Mootha *et al*, 2003; Subramanian *et al*, 2005), 'gene module maps' (Segal *et al*, 2004) and 'pathway signatures' (Bild *et al*, 2006b). The GSEA technique was initially applied to identify genes deregulated in human diabetic muscle (Mootha *et al*, 2003). Screening of 150 gene sets, among which are most of the classical metabolic pathways, led to the identification of a set of PGC1α-regulated genes involved in oxidative phosphorylation. In their 'module map' approach, Segal *et al* (2004) used almost 3000 gene sets to analyze a compendium of cancer-related microarray data sets (Segal *et al*, 2004). After simplifying the overlaps between gene sets, 400 modules were extracted, and combinatorial activation or deactivation of these can be followed in clinical samples to identify, for example, core processes common to heterogeneous tumor types. Building as this concept of reconstructing pathway-based systems maps, Bild *et al* (2006b) mapped multiple oncogenic pathway expression signatures from human cell line models using transgenic mouse mammary tumor models for validation and regressed the data into pathway activity scores. When applied to human tumors of different types, the pathway activity scores were able to classify tumors into groups associated with disease recurrence. Moreover, the pathway activity signatures could be used to predict the response of cell lines to drugs known to target specific components of the given oncogenic pathways.

This group went on to apply this strategy in the identification of transcriptional signatures predictive of therapeutic outcome in patients (Potti *et al*, 2006).

Common to these various approaches are the detection of coordinate changes of groups of genes, pre-defined as biologically related either experimentally or by data and literature mining. These techniques therefore go beyond a gene-by-gene analysis and are usually more sensitive in detecting modest changes in expression levels, which makes them particularly powerful for the analysis of human clinical samples. In addition, the fact that biologically and functionally related genes are tested collectively provides a direct biological interpretation to the analysis. Compared to the techniques described above, the main aspect of novelty of the MCM approach are the impressive size of the gene set collection (more than 14 000), which enables the exploration of a myriad of biological functions, and the fact that pairwise association of concepts are identified rather than simply ranking individual candidate pathways. Here, the MCM provides, in principle, a more integrated view of the network of biological processes that are active under a given physiological setting and, therefore, facilitate the interpretation of microarray expression signatures.

Although the above methodologies are efficient in uncovering groups of coordinately regulated genes, follow-up studies still face the challenge of identifying the molecular mechanisms responsible for the co-regulation of the identified pathways/concepts/modules. Chang *et al* (2005) attacked the problem in mammalian cancers in a series of publications. First, they extracted gene expression modules from expression arrays of primary breast cancers and derived the expression signatures of specific physiologic mechanisms pertinent to the disease: wound healing and serum response of cultured fibroblasts. By removing the cell-cycle-associated genes, they settled on 512 core serum response (CSR) genes that were considered representative of a 'wound' signature and found a strong correlation with survival. Integrating gene amplification and gene expression information, they later discovered that this wound signature is activated by coordinate amplification of *CSN5*, an activator of cullin-based E3 ubiquitin ligases, and the oncogenic transcription factor *MYC* (Adler *et al*, 2006). Taken together, their map of the transcriptional control network for the wound response in a complex system like human breast cancer revealed that the induction of a transcriptional wound response (CSR) requires not one but two genetic elements, MYC and CSN5, to be activated. Using a systems approach, these investigators identified components that provide a synthetic or conditional effect.

The work by Tomlins *et al* (2007) and works by the others mentioned in this article show that such systems maps can be applied in complex mammalian systems and yield results that can provide insight into human relevant diseases. Tomlins *et al* (2007) included a few more incremental innovations in experimentation design for studies focused on network maps in cancer. First, the authors heroically microdissected 101 prostate cancers for analysis to eliminate variable and confounding signals from non-malignant stromal tissues. In most expression signature studies of primary tumors, whole tissues rather than dissected tissues are used and have generated clinically important data. They correctly point out that if prognosis is the sole goal, then the stromal signals contribute to patient outcome predictions. In this case, the stromal signals are greater in localized disease and less in the disease with metastatic potential. However, if a map of the intracellular network of a prostate cancer cell is the goal, then elimination of the stromal signal is important.

The outcome of their analysis of the clustering of the expression signatures is also interesting. In terms of expression architecture, the expression concepts show similarity between PIN (a very early form of prostate cancer) and prostate cancers both of which differ from benign prostatic tissue. This is also what has been described for breast cancer, where the expression signature of DCIS (ductal carcinoma *in situ*, a preinvasive lesion) is similar to invasive cancer. This suggests that at least in some common epithelial cancers, the majority of the molecular changes in a cancer cell has already taken place in the earliest forms of the cancer. More specifically, the transition from benign to localized prostate cancer to metastatic disease is accompanied by an increase in expression of protein synthesis and proliferation concepts, by a decrease in androgen effects and by an increase in concepts related to the activity of the ETS family of transcription factors. This last result was important in view of the recent discovery of *TMRPSS2:ETS* gene fusions in the majority of prostate cancers (Tomlins *et al*, 2005), suggesting that ETS transcription factors may play a key role in prostate cancer progression.

Certainly, the most elegant systems network studies have been conducted in lower organisms such as bacteria, yeast or the sea urchin for obvious reasons. The simplicity of the systems, the ability to control experimental conditions and the knowledge base of how these organisms behave to specific stimuli all provide the precise high content data needed to render good dynamic interaction maps. However, the work reviewed herein, all focused on human cancers, show that some of these lessons learned in lower organisms can be applied to human tissues. Importantly, the information generated in this fashion are clinically relevant. Perhaps systems medicine is finally coming of age.

References

Adler AS, Lin M, Horlings H, Nuyten DS, van de Vijver MJ, Chang HY (2006) Genetic regulators of large-scale transcriptional signatures in cancer. *Nat Genet* **38:** 421–430

Bild AH, Potti A, Nevins JR (2006a) Linking oncogenic pathways with therapeutic opportunities. *Nat Rev Cancer* **6:** 735–741

Bild AH, Yao G, Chang JT, Wang Q, Potti A, Chasse D, Joshi MB, Harpole D, Lancaster JM, Berchuck A, Olson Jr JA, Marks JR, Dressman HK, West M, Nevins JR (2006b) Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* **439:** 353–357

Chang HY, Nuyten DS, Sneddon JB, Hastie T, Tibshirani R, Sorlie T, Dai H, He YD, van't Veer LJ, Bartelink H, van de Rijn M, Brown PO, van de Vijver MJ (2005) Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. *Proc Natl Acad Sci USA* **102:** 3738–3743

Liu ET (2005) Mechanism-derived gene expression signatures and predictive biomarkers in clinical oncology. *Proc Natl Acad Sci USA* **102:** 3531–3532

Liu ET, Kuznetsov VA, Miller LD (2006) In the pursuit of complexity: systems medicine in cancer biology. *Cancer Cell* **9:** 245–247

Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstrale M, Laurila E, Houstis N, Daly MJ, Patterson N, Mesirov JP, Golub TR, Tamayo P, Spiegelman B, Lander ES, Hirschhorn JN, Altshuler D, Groop LC (2003) PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* **34:** 267–273

Potti A, Dressman HK, Bild A, Riedel RF, Chan G, Sayer R, Cragun J, Cottrill H, Kelley MJ, Petersen R, Harpole D, Marks J, Berchuck A, Ginsburg GS, Febbo P, Lancaster J, Nevins JR (2006) Genomic signatures to guide the use of chemotherapeutics. *Nat Med* **12:** 1294–1300

Segal E, Friedman N, Koller D, Regev A (2004) A module map showing conditional activity of expression modules in cancer. *Nat Genet* **36:** 1090–1098

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* **102:** 15545–15550

Tomlins SA, Mehra R, Rhodes DR, Cao X, Wang L, Dhanasekaran SM, Kalyana-Sundaram S, Wei JT, Rubin MA, Pienta KJ, Shah RB, Chinnaiyan AM (2007) Integrative molecular concept modeling of prostate cancer progression. *Nat Genet* **39:** 41–51

Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao X, Tchinda J, Kuefer R, Lee C, Montie JE, Shah RB, Pienta KJ, Rubin MA, Chinnaiyan AM (2005) Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* **310:** 644–648