In the pursuit of complexity: Systems medicine in cancer biology

Adler et al., in a paper appearing in *Nature Genetics*, exploited the intersect of genetic information from expression profiles with that from array comparative genomic hybridization in human breast cancers to identify genes that may induce the transcription of the prognostic "wound response" expression signature. The amplification of two genes, *MYC* and *CSN5*, appeared to be correlated with the wound response cassette. In vitro validation showed that the wound signature could be induced in MCF10A cells only when *MYC* and *CSN5* were coexpressed. This work shows that the intersect analysis of gene amplification and transcriptional expression on a genome-wide scale can uncover complex conditional interactions embedded in the systems map of transcriptional regulation.

Since its inception, the promise of microarray technologies has been its potential to lay bare the molecular circuitry controlling cellular pathways and disease states. Indeed, a "holy grail" of the expression array field has been the elucidation of pathway maps and the mechanisms underlying biological systems and pathological states from the output of a genome-wide transcriptome analysis. However, with few exceptions, the microarray has fallen short of this promise—bona fide mechanisms have remained elusive, while "candidate" genes have been identified by the thousands. Some of the reasons for this shortcoming relate to the maturity of the technical platform (e.g., probe design and reproducibility), and to the state of knowledge of the transcriptome (from ESTs to full-length transcripts, to more advanced builds of the human genome sequence, to progressively better transcriptional representation of pathways). However, one of the problems is an underappreciation of array data dimensionality, which has limited the imagination in experimental design.

Array dimensionality can be defined by the complexity of the units of analysis. In its most basic form, arrays can be viewed as simple collections of individual probes. The analysis would therefore be focused on identifying an individual probe (or gene) associated with a phenotype. This analysis can be done with basic tools but has been troubled by irreproducibility across experimental systems, perhaps because of the inherent imprecision of the technology, and from multiple sampling error. Higher-dimensional analysis assesses the behavior of clusters of genes using approaches such as principle component analysis (PCA), and class prediction and feature selection procedures such as prediction analysis for microarrays (PAM) (Tibshirani et al., 2002), statistically weighted sysndromes (SWS) (Kuznetsov et al., 1996; Broet et al., 2006), and significance analysis of microarrays (SAM) (Tusher et al., 2001). Such approaches allow for the identification of coordinately expressed genes with biological or clinical associations. Linkages with other databases such as protein-protein interaction databases or Gene Ontology expand the dimensionality of the array data set to encompass biochemistry (protein interaction) and biological functional (Gene Ontology). The successes from this approach have relied on raising hypotheses from these pathway analyses and confirmation by direct experimentation. Thus, deciphering biological mechanisms requires looking beyond a single biological context and the unidimensional view of gene expression. Moreover, the intersection of heterogeneous databases including diverse biological and clinical information provides higher-dimensional views of gene function and their clinical significance.

Adler and coworkers, in their work published in a recent issue of *Nature Genetics* (Adler et al., 2006), combined multiple genomics approaches with both in vitro and in vivo measurements, to elucidate the causative transcriptional mechanisms responsible for a wound response signature that has prognostic significance in breast cancer. Building on previous work that identified a group of "wound response" genes coordinately induced in fibroblasts upon serum stimulation (Iyer et al., 1999) and subsequently found to be independently correlated with poor prognosis in breast cancer (Chang et al., 2005), Adler and colleagues correlated the wound response signature (defined by expression microarrays) to gene copy number changes (defined by array-CGH assessing chromosomal amplification) that might explain the transcriptional origin of the signature in aggressive breast cancer. Uniquely, they used the wound response expression cassette in a set of tumors as the dependent variable and asked what regions of gene amplification appeared to be correlated with this expression profile. The goal was to uncover the genetic components in amplicons that might transcriptionally control this wound response cassette. Adler and colleagues found that several genes on chromosome 8q were amplified in those tumors with the activated wound signature. In a larger set of 85 tumors for which they had wound scores derived from gene expression data, Adler et al. observed a number of genes on 8q24 and 8q13 that showed a high correlation between expression and wound score. Looking at candidates with biological plausibility, they hypothesized that the oncogene *MYC* and *CSN5*, a general activator of the cullen ubiquitin ligase complexes, both amplified in those tumors with the activated wound expression signature may functionally interact to regulate the transcription of the wound response cassette. Triangulation of these multiple nodes of genetic, genomic, and clinical information, together with some literature-based reasoning, allowed the pinpointing of two candidate genes that the authors hypothesized might interact functionally to modulate the wound signature genes.

To validate this, the authors overexpressed these two genes, *MYC* and *CSN5*, in MCF10A cells (derived from noncancerous human breast epithelial cells). Whereas overexpression of *MYC* and *CSN5*, individually, yielded only partial modulation of the wound signature genes, the simultaneous overexpression of both genes induced the synergistic transcriptional response of a large fraction of the wound signature genes. This confirmed the functional role of this gene pair in regulating the prognostic wound signature phenotype.

This work by Adler and coworkers, as well as that of several others recently (Lamb et al., 2003; Wei et al., 2006; Miller et al., 2005; Bild et al., 2006; Garraway et al., 2005; Sweet-Cordero et al., 2005) demonstrates that mechanistic discoveries are possible with microarray-based investigations if one searches the intersects of multiple data dimensions: transcription/genomic, bio-

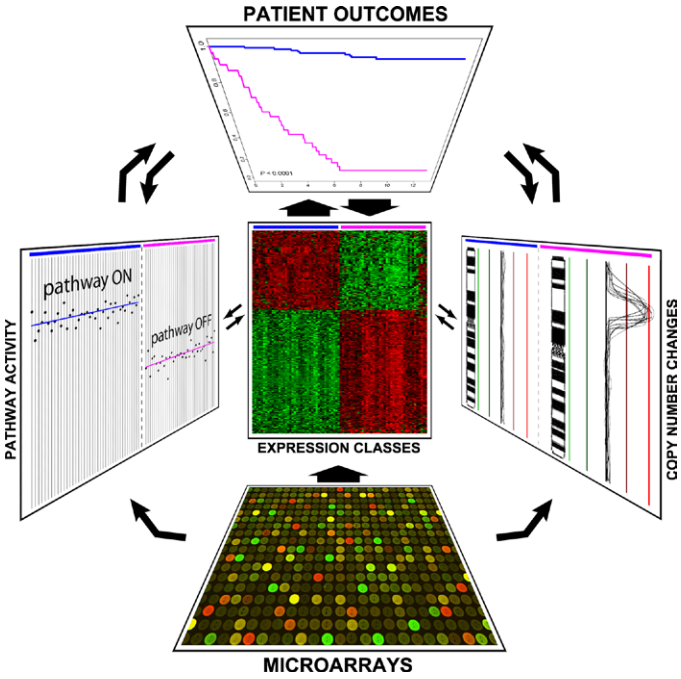

Figure 1. Concept schematic of intersecting planes of genomic, biological, and clinical data that inform mechanistic discoveries.

logical/clinical, etc. (see Figure 1). Adler and coworkers focused on the interface between copy number and gene expression, and gene expression and patient survival, to discover transcriptional mechanisms in cancer with clinical significance. Other important discoveries have recently been made at the boundaries of gene expression, pathway activation, and genome-wide transcription factor binding (Lamb et al., 2003; Wei et al., 2006; Miller et al., 2005; Bild et al., 2006). These studies also demonstrate the power of hypothesis building and testing between in vitro and in vivo systems, and even across species (Sweet-Cordero et al., 2005). To identify modifiers of *Cyclin D1* activity, Lamb and coworkers (Lamb et al., 2003) identified a reproducible cassette of *Cyclin D1*-responsive genes in vitro and determined that a large fraction of these genes were significantly correlated with *Cyclin D1* transcript levels in a panel of 190 tumors of mixed type (assessed by expression array analysis). The authors discovered that several transcription factors were both induced by *Cyclin D1* and correlated with *Cyclin D1* overexpression in vivo. Further analysis of >380 tumors and cell lines from publicly available data sets revealed that one transcription factor in particular, *C/EBP*β, remained significantly correlated with *Cyclin D1* expression in different biological contexts. To test the hypothesis that *C/EBP*β might modify *cyclin D1* activity, the authors performed expression and promoter assays in cultured cells. Their results indicate that endogenous *C/EBP*β acts as a constitutive repressor of *Cyclin D1* target gene promoters, thereby providing functional confirmation that *C/EBP*β is a mechanism of *Cyclin D1* action.

In another study, using a novel genomic technology called chIP-PET, Wei and coworkers (Wei et al., 2006) recently identified hundreds of novel bona fide *p53* binding sites in the human genome. Genes adjacent to these binding sites were assembled as candidate direct target genes. They showed that the expression pattern in primary breast cancers of these new putative direct targets predicted for the mutational status of the *p53* gene in these tumors and were prognostic of patient outcomes in human breast cancers. Similarly, we have described a 32-gene expression profile that is closely associated with *p53* status in human breast cancers and found that this profile can predict survival in breast cancer patients regardless of therapeutic intervention (Miller et al., 2005). Moreover, this *p53* mutational cassette is associated with *p53* mutations in hepatocellular carcinoma as well (Miller et al., 2005). In both examples, candidate cancer biomarkers were developed around a specific gene pathway and only afterwards associated with clinical outcome.

Forwarding this concept of pathway-specific expression signatures, Bild and coworkers (Bild et al., 2006) used expression arrays to map multiple oncogenic pathway signatures in human cell line models. The signatures were then regressed into pathway activity scores and validated in mouse mammary tumors derived from the activation of corresponding oncogenic pathways. In human tumors of different type, pathway activity scores were able to cluster tumors into groups associated with disease recurrence. Additionally, 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. Thus, pathway signatures that gauge the activity of cancer-related pathways in vitro can serve as biological, prognostic, and therapeutic readouts in vivo.

This work by Adler et al. highlights the fact that the intersection of two genome-wide scans (expression and chromosomal amplification) can yield interesting higher-order mechanistic explanations for biological processes such as cancer. The weakness still is that the process required investigator intervention to arrive at the candidate genes and is therefore subject to operator bias. For example, the 8q24 and 8q13 amplicons contain many other genes, but the selection of *MYC* and *CSN5* was because of "expert" input. These were most likely selected because of the knowledge that *MYC* is an important oncogene in breast cancer behavior and that *CSN5* is linked biochemically to post-translational modification mechanisms that control *MYC* function. The investigators were fortunate that the two molecules when tested in vitro in model systems gave confirmatory results. Still, the possibility remains that other partners in the amplicon can give the same interactive outcome but for very different reasons. Therefore, it will be important to show that *MYC* and *CSN5* have prognostic potential in other tumor sets from larger independent cohorts, and ultimately in prospective studies.

Nevertheless, this paper by Adler et al. highlights several larger principles in genomic medicine. In effect, they mapped the transcriptional control network for the wound response gene signature cassette in a complex system like human breast tumors and found that not one but two genetic elements, *MYC* and *CSN5*, had to be amplified in order for the cassette to be manifest. They therefore used a systems approach to identify components that provide a synthetic or conditional effect. Certainly synthetic lethality in genetic mutations is a well-known phenomenon in yeast genetics (Ooi et al., 2006). This is where the disruption of two genes gives a discernable phenotype when the knockout of either one has no effect. Such a principle has been proposed for cancer therapeutics (Kaelin, 2005). In addition, conditional response where the effect of a drug is dependent on the presence of another molecule that is not the target for the drug has been well described (Tan et al., 2005; Zhao et al., 2005). The *MYC* and *CSN5* conditional interaction raises the possibility that

targeted inhibition of *CSN5*-mediated regulation of *MYC* might be a potential new therapeutic approach for breast cancer with *MYC* amplification. Thus, this systems approach to mapping transcriptional control networks can provide pharmacological leads to targeted therapeutics that take advantage of cellular complexity. That *MYC* amplification is not found in normal tissues suggests that the therapeutic index of any such treatment targeting the *CSN5-MYC* interaction may be high.

Taken together, this work demonstrates that the intersection of orthogonal but precise data sets such as gene amplification and transcriptional expression on a genome-wide scale can provide insights into complex conditional interactions embedded in the systems map of cancer cell transcription.

Edison T. Liu,[1,*] Vladimir A. Kuznetsov,[1] and Lance D. Miller[1]

[1]Genome Institute of Singapore, Singapore 138672
*E-mail: liue@gis.a-star.edu.sg

Selected reading

Adler, A.S., Lin, M., Horlings, H., Nuyten, D.S., van de Vijver, M.J., Chang, H.Y. (2006). Nat. Genet., in press. Published online March 5, 2006. 10.1038/ng1752.

Bild, A.H., Yao, G., Chang, J.T., Wang, Q., Potti, A., Chasse, D., Joshi, M.B., Harpole, D., Lancaster, J.M., Berchuck, A., et al. (2006). Nature *439*, 353–357.

Broet, P., Kuznetsov, V.A., Bergh, J., Liu, E., Miller, L.D. (2006). Bioinformatics, in press. Published online March 21, 2006. 10.1093/bioinformatics/btl110.

Chang, H.Y., Nuyten, D.S., Sneddon, J.B., Hastie, T., Tibshirani, R., Sorlie, T., Dai, H., He, Y.D., van't Veer, L.J., Bartelink, H., et al. (2005). Proc. Natl. Acad. Sci. USA *102*, 3738–3743.

Garraway, L.A., Widlund, H.R., Rubin, M.A., Getz, G., Berger, A.J., Ramaswamy, S., Beroukhim, R., Milner, D.A., Granter, S.R., Du, J., et al. (2005). Nature *436*, 117–122.

Iyer, V.R., Eisen, M.B., Ross, D.T., Schuler, G., Moore, T., Lee, J.C., Trent, J.M., Staudt, L.M., Hudson, J., Jr., Boguski, M.S., et al. (1999). Science *283*, 83–87.

Kaelin, W.G., Jr. (2005). Nat. Rev. Cancer *5*, 689–698.

Kuznetsov, V.A., Ivshina, A.V., Sen'ko, O.V., and Kuznetsova, A.V. (1996). Math. Comput. Model. *23*, 92–112.

Lamb, J., Ramaswamy, S., Ford, H.L., Contreras, B., Martinez, R.V., Kittrell, F.S., Zahnow, C.A., Patterson, N., Golub, T.R., and Ewen, M.E. (2003). Cell *114*, 323–334.

Miller, L.D., Smeds, J., George, J., Vega, V.B., Vergara, L., Ploner, A., Pawitan, Y., Hall, P., Klaar, S., Liu, E.T., et al. (2005). Proc. Natl. Acad. Sci. USA *102*, 13550–13555.

Ooi, S.L., Pan, X., Peyser, B.D., Ye, P., Meluh, P.B., Yuan, D.S., Irizarry, R.A., Bader, J.S., Spencer, F.A., and Boeke, J.D. (2006). Trends Genet. *22*, 56–63.

Sweet-Cordero, A., Mukherjee, S., Subramanian, A., You, H., Roix, J.J., Ladd-Acosta, C., Mesirov, J., Golub, T.R., and Jacks, T. (2005). Nat. Genet. *37*, 48–55.

Tan, J., Zhuang, L., Leong, H.S., Iyer, N.G., Liu, E.T., and Yu, Q. (2005). Cancer Res. *65*, 9012–9020.

Tibshirani, R., Hastie, T., Narasimhan, B., and Chu, G. (2002). Proc. Natl. Acad. Sci. USA *99*, 6567–6572.

Tusher, V.G., Tibshirani, R., and Chu, G. (2001). Proc. Natl. Acad. Sci. USA *98*, 5116–5121.

Wei, C.L., Wu, Q., Vega, V.B., Chiu, K.P., Ng, P., Zhang, T., Shahab, A., Yong, H.C., Fu, Y., Weng, Z., et al. (2006). Cell *124*, 207–219.

Zhao, Y., Tan, J., Zhuang, L., Jiang, X., Liu, E.T., and Yu, Q. (2005). Proc. Natl. Acad. Sci. USA *102*, 16090–16095.

DOI 10.1016/j.ccr.2006.03.026