Molecular Cell
Previews

A FIRE-y PAGE in the Computational Analysis of Cancer Profiles

John R. Prensner[2] and Arul M. Chinnaiyan[1,2,*]
[1]Howard Hughes Medical Institute
[2]Michigan Center for Translational Pathology
Department of Pathology and Urology, Comprehensive Cancer Center, University of Michigan Medical School, Ann Arbor, MI 48109, USA
*Correspondence: arul@umich.edu
DOI 10.1016/j.molcel.2009.11.019

In this issue of *Molecular Cell*, Goodarzi et al. (2009) employ novel computational approaches to demonstrate the power of global systems biology analyses in elucidating cancer biology.

It seems contradictory that the more we study genome sequences, the less confidently we seem to understand fundamental biological processes. In the world of regulatory factors, such as transcription factors, the emergence of global profiling technologies such as ChIP-on-chip or ChIP-Seq has led to the perplexing discovery that many of these factors show binding at many more genomic positions than previously thought (Fullwood et al., 2009). Whether these findings represent functionally dismissible nonspecific interactions, technical or procedural artifacts, or novel biology is yet to be fully determined.

To this end, much emphasis has been placed on the computational identification of functionally important DNA sequence motifs or *cis*-regulatory elements that may indicate genomic regions with coordinately regulated functions (Das and Dai, 2007). These efforts become even more powerful when combined with a vast amount of publicly available gene expression and protein interactome data that enable a global picture of biological processes and their dysregulation in diseases such as cancer. Indeed, a systems biology approach to the global interpretation of these data may now replace a decades-old pathway-based approach that overemphasizes conventional, although undoubtedly important, signal transduction pathways dominated by familiar names and behaviors. Such systems biology analyses of oncogenic and wild-type *Ras* have shown that there is much to learn even for such a well-characterized small GTPase (Bluthgen et al., 2009; Sachs et al., 2005). Additionally, the surprising finding that 4% of protein kinases also bind DNA—and, in the case of *Erk2*, have direct functionality in gene repression—further underscores the broad horizons enabled by global systems-biology-based methods (Hu et al., 2009).

In this issue of *Molecular Cell*, Tavazoie and colleagues (Goodarzi et al., 2009) offer a new approach to the formidable task of making biological sense from a collection of genome-wide profiling studies. Previous work from this lab developed FIRE (finding informative regulatory elements), a computational method to predict and identify high-confidence DNA motifs whose presence or absence in a regulatory region correlates with other informative traits, including the motif's position or orientation relative to a gene, its relationship to other motifs, and, importantly, its correlation with gene expression (Elemento et al., 2007). To accomplish this, FIRE employs the concept of mutual information, which provides a general measure of the dependency between the motif and its functional constraints. In the current study, Goodarzi et al. (2009) develop a counterpart to FIRE, termed iPAGE (information-theoretic pathway-level analysis of gene expression) (Figure 1).

iPAGE uses gene expression data to compile known pathways, cellular processes, and gene ontologies to quantify biological processes that are dysregulated in cancer (Goodarzi et al., 2009). Importantly, iPAGE is able to cluster these data to reveal intracancer heterogeneities specific to cancer subtypes as well as perform broad cancer versus normal analyses. In this respect, iPAGE enters a field already populated by a variety of methodologies also looking for statistically significant nonrandom patterns in pathway clusters and gene expression (Sinha et al., 2008; Subramanian et al., 2005).

The key insight with iPAGE comes from its joint utility with FIRE. Together, the two identify biological phenomena (gene ontologies, pathways, etc.) dysregulated in cancer from gene expression data and simultaneously nominate putative motifs that are commonly shared in specific gene expression clusters (Goodarzi et al., 2009). Ultimately, the idea is to intuit key regulatory factors (microRNAs, transcription factors, etc.) that recognize these specific motifs that contribute to a variety of dysregulated processes in cancer (Figure 1). These regulatory factors then become candidate drivers of oncogenesis.

To demonstrate the efficacy of the iPAGE-FIRE system, Goodarzi et al. (2009) apply their algorithms to several case examples, including bladder cancer (representing a cancer versus normal situation), Burkitt's lymphoma and diffuse large B cell lymphoma (DLBCL) (representing two cancer subtypes), and finally a global meta-analysis of cancers. In each case, they can recapitulate known observations as well as offer new ones. In bladder cancer, for example, they accurately identify gene ontologies pertaining to increased mitosis and DNA replication and offer intriguing evidence that decreased *Elk1*, a member of the *Ets* family of transcription factors, may account for several of these gene expression patterns. Knockdown of *Elk1* in an in vitro model system generated similar patterns of expression, further suggesting that this gene deserves to be studied

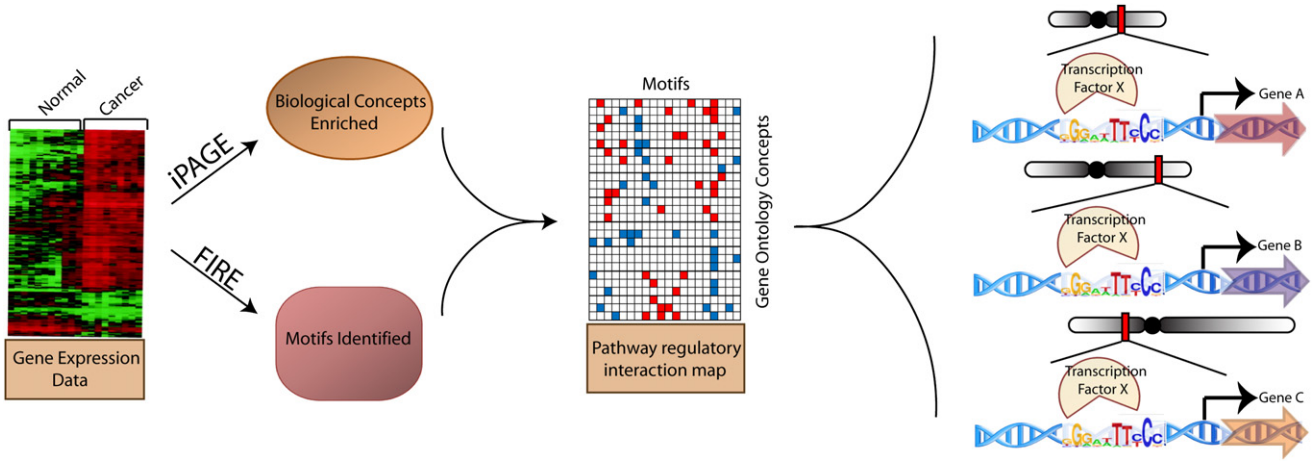

Figure 1. A Schematic of the iPAGE-FIRE System

Gene expression data are processed independently by FIRE and iPAGE in order to nominate motifs and gene ontologies, respectively. Integration of the subsequent nominations enables global mapping of candidate driver motifs (and therefore their respective regulatory factors, if known) and the putative ontologies and pathways that they control.

more closely in bladder cancer (Goodarzi et al., 2009). Similar results were observed in lymphoma, where another transcription factor, *NF-Y*, was nominated in Burkitt's lymphoma based on its profile in that disease when compared to DLBCL. Knockdown of this gene also confirmed the profile observed by iPAGE analysis of Burkitt's lymphoma (Goodarzi et al., 2009).

To apply their algorithms globally, Goodarzi et al. (2009) experimentally test an upstream sequence motif (AAAA [AGT]TT) that they observe in multiple cancer iPAGE-FIRE data sets. By performing a titration assay in which an exogenously administered synthetic AAAA [AGT]TT double-stranded DNA oligonucleotide competitively binds its transcription factor, they show that this motif—and whatever regulatory factors bind to it—regulates the expression of many of the predicted gene clusters identified by iPAGE (Goodarzi et al., 2009).

Still, there are many areas for systems biology to explore. Frequent reliance on literature mining coupled with little direct experimental confirmation leaves many hypotheses untested. Indeed, it is encouraging that Goodarzi et al. (2009) provide experimental evidence for their findings, albeit in a tangential model system—they use a breast cancer cell line to validate findings in bladder cancer and lymphoma. In light of this, a major question remains as to whether different cellular contexts can be modeled accurately to reflect cell-type-specific biological functions. On a tumor level, such heterogeneity can expand to tumors that have varying degrees of overlap between their molecular lesions—e.g., what interactions distinguish a *Kras* and *Egfr* mutant lung adenocarcinoma from a *Kras* mutant, *Egfr* wild-type one? Systems biology may also play a key role in understanding the biology of putative tumor-initiating cells that express certain cell surface markers.

With these results, Goodarzi et al. (2009) provide intriguing evidence that a systems biology approach can identify important aspects of cancer on several levels: (1) by clustering gene expression profiles and related ontology signatures, (2) by nominating *cis*-regulatory motifs consistent within and across these signatures, and (3) nominating candidate mediators of specific cancer phenotypes. Such a schema promises to elucidate fundamental questions in tumor biology, and future studies ought to utilize this emerging technology.

REFERENCES

Bluthgen, N., Legewie, S., Kielbasa, S.M., Schramme, A., Tchernitsa, O., Keil, J., Solf, A., Vingron, M., Schafer, R., Herzel, H., and Sers, C. (2009). FEBS J. *276*, 1024–1035.

Das, M.K., and Dai, H.K. (2007). BMC Bioinformatics *8* (*Suppl 7*), S21.

Elemento, O., Slonim, N., and Tavazoie, S. (2007). Mol. Cell *28*, 337–350.

Fullwood, M.J., Liu, M.H., Pan, Y.F., Liu, J., Xu, H., Mohamed, Y.B., Orlov, Y.L., Velkov, S., Ho, A., Mei, P.H., et al. (2009). Nature *462*, 58–64.

Goodarzi, H., Elemento, O., and Tavazoie, S. (2009). Mol. Cell *36*, this issue, 900–911.

Hu, S., Xie, Z., Onishi, A., Yu, X., Jiang, L., Lin, J., Rho, H.S., Woodard, C., Wang, H., Jeong, J.S., et al. (2009). Cell *139*, 610–622.

Sachs, K., Perez, O., Pe'er, D., Lauffenburger, D.A., and Nolan, G.P. (2005). Science *308*, 523–529.

Sinha, S., Adler, A.S., Field, Y., Chang, H.Y., and Segal, E. (2008). Genome Res. *18*, 477–488.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Proc. Natl. Acad. Sci. USA *102*, 15545–15550.