is mediated by the FGFR2b receptor isoform in vivo and only signals weakly through the FGFR1b isoform (Ornitz and Itoh, 2001; Zhang et al., 2006). One possible explanation is that high levels of FGF10 may be sufficient to activate FGFR1 in vivo—further analyses of FGF10 signaling in the prostate will undoubtedly shed additional light on this issue.

In summary, the two new manuscripts in this issue of *Cancer Cell* indicate that paracrine actions of FGF signaling between stromal and epithelial compartments may be critical for prostate tumorigenesis, as has long been proposed in the literature. These provocative studies open new avenues for further research, particularly to determine how FGFR pathway activation can drive EMT and promote metastasis. Notably, the interactions between FGFR signaling and the AR and Akt pathways during prostate tumorigenesis suggest that potentially valuable therapeutic strategies may exist for targeting FGF receptors in conjunction with Akt or related pathway inhibitors.

REFERENCES

Acevedo, V.D., Gangula, R.D., Freeman, K.W., Li, R., Zhang, Y., Wang, F., Ayala, G.E., Peterson, L.E., Ittmann, M., and Spencer, D.M. (2007). Cancer Cell, this issue.

Cunha, G.R., Donjacour, A.A., Cooke, P.S., Mee, S., Bigsby, R.M., Higgins, S.J., and Sugimura, Y. (1987). Endocr. Rev. *8*, 338–362.

Freeman, K.W., Welm, B.E., Gangula, R.D., Rosen, J.M., Ittmann, M., Greenberg, N.M., and Spencer, D.M. (2003). Cancer Res. *63*, 8256–8263.

Hayward, S.W., Wang, Y., Cao, M., Hom, Y.K., Zhang, B., Grossfeld, G.D., Sudilovsky, D., and Cunha, G.R. (2001). Cancer Res. *61*, 8135–8142.

Kwabi-Addo, B., Ozen, M., and Ittmann, M. (2004). Endocr. Relat. Cancer *11*, 709–724.

Lin, Y., Liu, G., Zhang, Y., Hu, Y.P., Yu, K., Lin, C., McKeehan, K., Xuan, J.W., Ornitz, D.M., Shen, M.M., et al. (2007). Development *134*, 723–734.

Memarzadeh, S., Xin, L., Mulholland, D.J., Mansukhani, A., Wu, H., Teitell, M.A., and Witte, O.N. (2007). Cancer Cell, this issue.

Ornitz, D.M., and Itoh, N. (2001). Genome Biol. *2*, REVIEWS3005.

Ropiquet, F., Giri, D., Kwabi-Addo, B., Schmidt, K., and Ittmann, M. (2000). Prostate *44*, 334–338.

Thomson, A.A., and Cunha, G.R. (1999). Development *126*, 3693–3701.

Xin, L., Teitell, M.A., Lawson, D.A., Kwon, A., Mellinghoff, I.K., and Witte, O.N. (2006). Proc. Natl. Acad. Sci. USA *103*, 7789–7794.

Zhang, X., Ibrahimi, O.A., Olsen, S.K., Umemori, H., Mohammadi, M., and Ornitz, D.M. (2006). J. Biol. Chem. *281*, 15694–15700.

Zhong, C., Saribekyan, G., Liao, C.P., Cohen, M.B., and Roy-Burman, P. (2006). Cancer Res. *66*, 2188–2194.

New Breast Cancer Genes—Discovery at the Intersection of Complex Data Sets

Joseph R. Nevins[1,*]
[1]Duke Institute for Genome Sciences & Policy, Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, NC 27708, USA
*Correspondence: nevin001@mc.duke.edu
DOI 10.1016/j.ccr.2007.11.019

The identification of genes that contribute to the oncogenic process, including those that determine risk of cancer onset, holds the key not only in understanding mechanisms of oncogenesis but also in the identification of new targets for therapeutic development. Traditional methods of genetics and molecular biology have been successful but are slow and laborious. The advent of genome technologies, leading to the generation of large data sets describing various properties of genes and proteins relevant to cancer phenotypes, has afforded a new opportunity for discovery. M. Vidal and colleagues have made use of this data, and in particular the integration of various forms of genome-scale data, to identify new genes involved in breast cancer.

The development of technologies that facilitate genome-scale analyses of gene expression, DNA sequence variation, protein accumulation, protein interactions, DNA copy number, and more has had a transforming effect on biology and medicine. This is perhaps best seen in cancer, where complex data sets have been coupled with powerful analytical methods to extract a level of detail of the underlying biology not achievable with the once powerful methods of molecular biology. As an example, the use of large-scale gene expression data has dissected cancer into a variety of subtypes that begin to address the full complexity of the disease, recognizing that lymphoma, breast cancer, and others are not single diseases, or even two or three diseases, but rather a vast array of complex diseases defined by a variety of genetic alterations (Alizadeh et al., 2000; Golub et

al., 1999; Perou et al., 2000). Although in many instances, these complex data sets have been developed for the purpose of a specific study such as the ability to predict breast cancer recurrence or to identify lymphoma subtypes that represent distinct disease entities, or to predict the response to specific therapeutic regimens, others have been generated for the purpose of large-scale descriptions of gene expression, protein interaction, and other characteristics without regard to a specific context. Either way, they represent a resource of data that describes the biological complexity of the disease state in question. As such, they represent potentially enormously valuable resources of information that can form the basis for discovery of new disease mechanisms and ultimately new targets for cancer therapy.

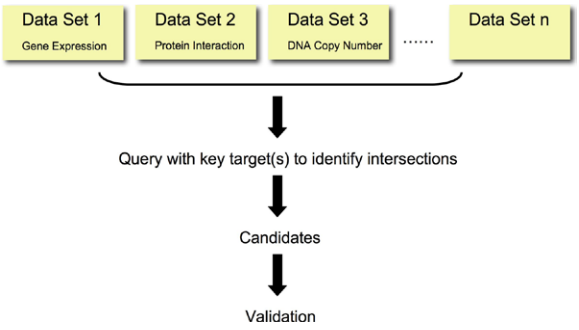

Figure 1. Integration of Multiple Genome-Scale Data Sets for Discovery
The schematic depicts the utilization of multiple data sets describing gene expression, protein interaction, and other forms of large-scale data to search for interactions that are evident in multiple states.

A recent study by Marc Vidal and colleagues, published in *Nature Genetics*, describes a strategy to take advantage of this available data and link it in a way that allows for discovery of genes not previously recognized to play a role in a particular process (Pujana et al., 2007). The context in this case is breast cancer, starting from the genes identified in genetic studies to confer increased susceptibility to breast cancer—BRCA1 and BRCA2. These were combined with two additional genes—ATM and CHEK2, genes also known to confer risk of breast cancer, albeit at much lower penetrance—as the starting point in a genomic data search for additional activities contributing to breast cancer risk. Importantly, each of these gene products is involved in a functional pathway of DNA damage recognition and repair. As such, the common function provided a basis for search for additional genes with shared properties that might then lead to additional genes involved in breast cancer susceptibility. The logic of the approach to utilizing these resources is relatively simple. If two genes are known to share expression properties across diverse circumstances, they might represent genes with shared function. But if they also are known to encode proteins that physically interact, or that genetically interact, then the likelihood that they share function increases substantially. The power of this strategy lies in the availability of diverse data sets describing various properties of genes and gene products—DNA sequence variation, chromosomal alterations, gene expression, protein interactions, and more.

The starting point in this search was the use of gene expression data from normal human breast tissue. The concept is straightforward—genes that share expression properties across a large number of biological samples might also share function, an approach applied previously as a basis for discovery of similar function (Stuart et al., 2003). Various methods for analyzing this data, and identifying coexpression networks, can be employed; Vidal and colleagues used a simple determination of correlation coefficient. In this case, the goal is to identify expression patterns that might yield functionally related gene products. Evidence that the coexpression data did yield genes of interest was suggested by an enrichment for functional interactions evident in a literature search as well as shared Gene Ontology terms.

Obviously, this one piece of data alone would not be sufficient. Genes that are coexpressed are candidates for further study, but the number of such genes is very large. It is at this point that the real power of the strategy comes into play, by combining these patterns of gene expression covariation with other equally powerful data sets such as protein accumulation, genetic interactions, and protein interactions (Figure 1). Such an analysis of a BRCA1-based network led to the identification of the hyaluronan-mediated motility receptor gene (HMMR) as a component of the network with the highest correlation with BRCA1 expression. HMMR was found to interact with components of the centrosome, and further analysis demonstrated an interaction of BRCA1 with the centrosome, specifically as cells enter mitosis.

Of course, even with the compelling nature of the data intersections that point toward HMMR as a BRCA1-related activity, this is nevertheless only a candidate. The power in the data intersection approach is to generate such candidates with high probability of being involved in the process of interest. Nevertheless, proof of a role for the gene in breast cancer and DNA repair requires experimental validation. Vidal and colleagues took two approaches to providing such validation. First, siRNA-based knockdown of HMMR resulted in an increase in centrosome number; importantly, the same phenotype was obtained upon knockdown of BRCA1. Second, and most importantly in the context of identification of new risk-associated genes, an analysis of DNA sequence variation (SNPs) in the HMMR gene demonstrated a link with breast cancer susceptibility, independent of BRCA1 status. In particular, multiple SNPs within the HMMR locus were associated with risk in multiple independent cohorts of breast cancer patients. Moreover, further analyses indicated a link between expression

of the HMMR gene and risk—higher expression of HMMR was associated with an early age of diagnosis.

With a starting point of examining BRCA1 coexpression, and then expanding through the analysis of physical and functional intersections observed in various complex data sets, a new breast cancer gene has emerged that now extends the function of the major breast cancer susceptibility genes and provides the basis for further study. Perhaps most important is the demonstration of a strategy that can make use of the wealth of accumulating genome-scale data for in silico discovery. The opportunities for exploiting these data sets with a strategy such as that elaborated by Vidal and colleagues is only limited by the question to be addressed and a knowledge of the appropriate starting point in the search. Other recent examples can be seen in the connection of gene expression signatures with cancer genome alterations as a mechanism to identify genes linked to the relevant cancer phenotypes. This has led to the identification of Myc and Jab1 amplification in breast cancer based on a link with a wound response expression signature (Adler et al., 2006) and identification of the MITF transcription factor as a melanoma oncogene based on expression properties linked to a melanoma gene amplification profile (Garraway et al., 2005). As with the example of the identification of HMMR as a new breast cancer gene, these further studies highlight the identification of genes based on an integration of powerful data sets and analytical tools. We have every reason to believe that more is to be mined with these tools.

REFERENCES

Adler, A.S., Lin, M., Horlings, H., Nuyten, D.S., van de Vijver, M.J., and Chang, H.Y. (2006). Nat. Genet. *38*, 421–430.

Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.C., Sabet, H., Tran, T., Yu, X., et al. (2000). Nature *403*, 503–511.

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.

Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., et al. (1999). Science *286*, 531–537.

Perou, C.M., Sorlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A., et al. (2000). Nature *406*, 747–752.

Pujana, M.A., Han, J.-D.J., Starita, L.M., Stevens, K.N., Tewari, M., Ahn, J.S., Rennert, G., Morena, V., Kirchhoff, T., Gold, B., et al. (2007). Nat. Genet. *39*, 1338–1349.

Stuart, J.M., Segal, E., Koller, D., and Kim, S.K. (2003). Science *302*, 249–255.