Importantly, the findings of Derksen et al. constitute the first demonstration in mice of a causal link between E-cadherin loss and tumor formation. Given the wealth of data from cell culture studies and human patient samples implicating E-cadherin in tumor progression, this is a central finding that establishes E-cadherin as a bona fide tumor suppressor. Notably, while E-cadherin loss is thought to be a critical step in the process of epithelial-to-mesenchymal transition (EMT), the extent to which EMT per se is involved in tumor progression remains unclear. For example, the transcriptional repressor Snail has been shown to downregulate *E-cadherin* as well as promote both EMT and tumor progression in a conditional mouse model for HER2/neu-induced breast cancer (Moody et al., 2005). In contrast, E-cadherin loss in the model described by Derksen et al. does not by itself lead to a frank mesenchymal phenotype. This suggests that, at least in this context, E-cadherin loss can promote tumorigenesis independently of its contribution to EMT.

Another puzzling feature of this disease has been the uncertain relationship between ILC and lobular carcinoma in situ (LCIS). Women with LCIS are at substantially increased risk for developing breast cancer both in the same breast in which the LCIS was identified and in the opposite breast. As such, whether LCIS represents a precursor lesion to ILC, or merely a marker of increased risk, continues to be a matter of debate. It is notable, therefore, that despite the markedly increased risk of mILC in mice in which E-cadherin and p53 have been deleted in the mammary epithelium, classic LCIS lesions are not found. This raises the important possibility that LCIS is not a precursor lesion for ILC, or that mILC models a form of ILC that does not pass through an LCIS phase.

Finally, beyond its important mechanistic implications, this study by Derksen et al. is equally significant for its establishment of a faithful model for human ILC where none existed before. This accomplishment represents a significant step forward in the effort to accurately model human cancers in mice. As they constitute only 10%–15% of all breast cancers, lobular carcinomas are likely to differ from ductal carcinomas with respect to their etiology, biology, and response to therapy. As such, the availability of this model holds significant promise for improving our ability to understand and treat this type of breast cancer. Undoubtedly, it will be important to continue refining this model to more precisely incorporate the molecular alterations found in human ILC, as these become elucidated. Ultimately, such models should prove useful for testing therapeutic regimens targeted specifically against ILC and in that manner improve the clinical management of this disease. This hope, if realized, would constitute the most meaningful validation of all for this—or any—mouse model.

James V. Alvarez,[1] Denise Perez,[1] and Lewis A. Chodosh[1,*]

[1]Department of Cancer Biology and Abramson Family Cancer Research Institute, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104
*E-mail: chodosh@mail.med.upenn.edu

Selected reading

Arpino, G., Bardou, V.J., Clark, G.M., and Elledge, R.M. (2004). Breast Cancer Res. *6*, R149–R156.

Derksen, P.W.B., Liu, X., Saridin, F., van der Gulden, H., Zevenhoven, J., Evers, B., van Beijnum, J.R., Griffioen, A.W., Vink, J., and Krimpenfort, P. (2006). Cancer Cell, this issue.

Holland, E.C. (2004). Mouse Models of Human Cancer (Hoboken, N.J.: Wiley-Liss).

Mohsin, S.K., O'Connell, P., Allred, D.C., and Libby, A.L. (2005). Breast Cancer Res. Treat. *90*, 249–256.

Moody, S.E., Perez, D., Pan, T.C., Sarkisian, C.J., Portocarrero, C.P., Sterner, C.J., Notorfrancesco, K.L., Cardiff, R.D., and Chodosh, L.A. (2005). Cancer Cell *8*, 197–209.

Papadimitriou, C.S., Costopoulos, J.S., Christoforidou, B.P., Kotsianti, A.J., Karkavelas, G.S., Hytiroglou, P.M., Koufogiannis, D.J., and Nenopoulou, H.E. (1997). Eur. J. Cancer *33*, 1275–1280.

Sharpless, N.E., and Depinho, R.A. (2006). Nat. Rev. Drug Discov. *5*, 741–754.

Stange, D.E., Radlwimmer, B., Schubert, F., Traub, F., Pich, A., Toedt, G., Mendrzyk, F., Lehmann, U., Eils, R., Kreipe, H., and Lichter, P. (2006). Clin. Cancer Res. *12*, 345–352.

DOI 10.1016/j.ccr.2006.10.016

Accelerating drug discovery: Open source cancer cell biology?

The possibility that experimental data from diverse cell biology experiments might shed light on other experiments has been generally outside the realm of cancer biologists. Recent experiments suggest that core RNA expression profiles distilled from experiments using a set of known members with related attributes may be used as query tools to probe expression profiles from other unrelated experiments. The potential benefit arises from the possibility to share findings without fully reconstructing the exact initial conditions. The limitations will be framed by the robustness of the hypotheses so generated.

As one looks back over past decades, or for that matter, past centuries, scientific progress, though remarkable, has been limited by tools and the ways insights have been shared. Three recent papers by the Golub and Armstrong groups (Lamb et al., 2006; Hieronymus et al., 2006; Wei et al., 2006) describe a process, a tool, and even better, examples of how that tool functions that suggest an opportunity which, if fully implemented, has the potential to truly change the speed and efficiency of drug discovery.

One of the rather hidden inadequacies of cancer biology and drug discovery is the full extent to which we fly blind. From the size of textbooks and the complexity of pathway diagrams it is hard to recognize that for virtually all cancer processes we do not have confidence in the complete "parts list" of components

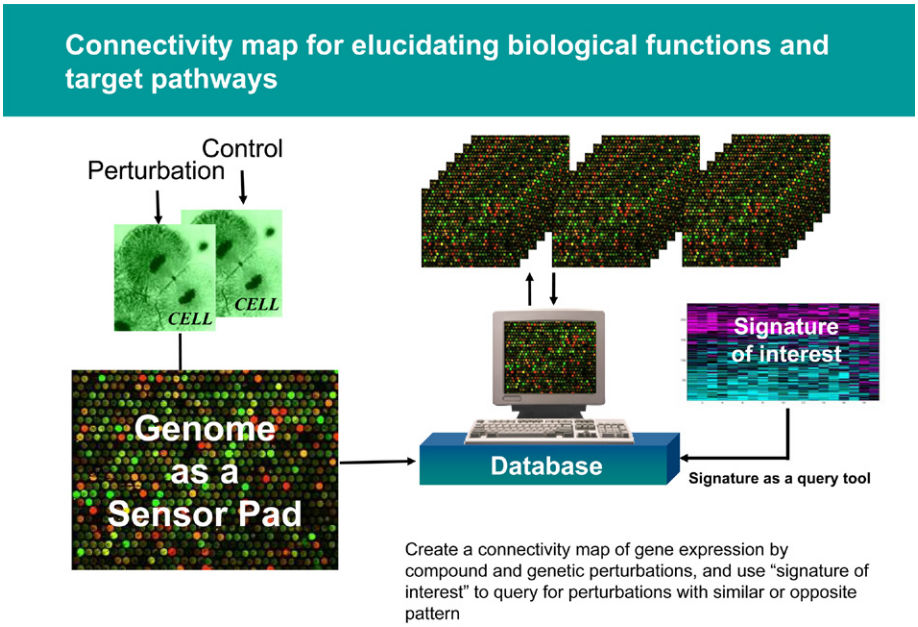

Figure 1. Using a gene expression pattern for the "Connectivity Map"

Cells are treated with diverse molecular compounds (or genetic perturbations), and gene expression patterns are stored in the database. A "signature of interest" derived from a disease type (for example, obese versus normal, or drug resistant versus sensitive) can be used to query the database to find which compound or drug treatment could reverse the pattern. A "signature of interest" could also be from a novel compound. The target pathways of this novel compound can be elucidated by pattern matching to the compounds or drugs in the database with known targets or functions.

for these pathways, let alone their crosswiring diagrams. There is, however, a more fundamental and critical problem, and that is the way cancer biologists have until now shared their data.

Here, we do not refer to the prevalent data sets in the silos of intellectual property-driven universities or even the more striking company silos. Instead, we are referring to what happens to the data sets at the end of any biological experiment. Most often, it is simply thrown away. In a fashion that would be deplored by environmentalists, cancer biologists do not readily recycle the data sets because we can't. For today's cellular biology experiment, the primary light of day is in the hallowed realm of publications, where the sections appear as follows: title, data, methods, and conclusions. Our concern is that the ability to leverage this information is primarily limited by the need to reconstruct that specific set of conditions to truly extend the finding. To point out this inefficiency, we need only refer to software engineers. One could argue that the most exciting development in the software field is not the benefit properly attributed to the engineers and their creations that drive Moore's law (Moore, 1965), but the open source strategies driven originally by Richard Stallman (see, for example, http://www.opensource.org/). The excitement regarding open source approaches arises from the power of having computer science's currency of discovery, lines of code, increasingly made accessible and, more importantly, evolved by many people. As biologists, we have marveled with sheer jealousy at how efficient this sharing of discovery can be. To understand how the papers by Golub, Armstrong, and colleagues may potentially help release this data set segmentation problem in biology, it is necessary to describe what they have done. These papers suggest a novel way to share the data sets generated within experiments.

In a recent issue of *Science*, Golub and coworkers published a striking paper about a "Connectivity Map" for a mammalian system (Lamb et al., 2006). They treated the cells with a collection of diverse small-molecule compounds, recorded gene expression alterations caused by each of the compounds, stored these expression patterns in the database, and then used the pattern matching method to link up the compounds of unknown function with compounds with known biological targets and functions (see Figure 1). In addition, they show that this same database can be used to find new therapeutics that can reverse a disease pattern. Two applications for this Connectivity Map are fully illustrated by the papers in *Cancer Cell* (Hieronymus et al., 2006; Wei et al., 2006).

In the first paper, Wei et al. tackle the phenomenon of drug resistance, a major obstacle to successful cancer treatment. Armstrong and coworkers first identified a gene expression pattern associated with resistance to dexamethasone treatment in primary acute lymphoblastic leukemia (ALL) cells and, by the use of a Connectivity Map, they identified that a marketed drug, rapamycin, can reverse the resistance pattern (Wei et al., 2006). They thus hypothesized that rapamycin would induce dexamethasone sensitivity in lymphoid malignancy cells and found that it indeed sensitized the resistant lymphoid cell line to dexamethasone-induced apoptosis via a modulation of antiapoptotic protein MCL-1. This result suggests that dexamethasone in combination with rapamycin could be an effective therapy for childhood ALL and should be tested immediately in clinical trials.

In the second example, Golub and coworkers demonstrated the power of gene expression-based high-throughput screening for identifying novel compounds and used a Connectivity Map to elucidate the target pathways (Hieronymus et al., 2006). The androgen receptor (AR)-mediated pathway that they explored is associated with prostate cancer progression and eventual hormonal therapy resistance, but the ability to modulate this pathway is limited. Using a defined AR activation gene signature in a prostate cell line and bead-based expression platform, they identified celestrol and gedunin from 2500 compounds as novel inhibitors for the AR signaling pathway. Then, using the "Connectivity Map," they showed that the modulation of AR pathway acted through an upstream HSP90 pathway.

Considering that the first publication of a yeast compendium of genetically altered organisms, conditions, and compounds describing the benefits of using knowns to query unknowns was published in 2000 (Hughes et al., 2000), why did it take more than 5 years to generate a counterpart in mammalian

cells? First, there has been a common fear that a mammalian system would be intrinsically complex compared to the yeast. Apart from shared problems like dose and time of treatment, there are many extra hurdles in the mammalian system, such as the complexity resulting from many cell types, many tissues, in vitro and in vivo, tumor and normal, individual variability, etc. Furthermore, many groups became stuck at the stage of discussing which cell to choose to construct a compound expression compendium. Current results by Golub and Armstrong et al. suggest that a big fraction of connectivity appears to be to a large extent cell line independent. They even include examples that suggest connectivity across species, and from in vitro to in vivo (obesity example from Lamb et al., 2006). The second reason could be the inadequacies of existing analytical methods. Previous attempts were more or less limited to the genome-wide comparison, like hierarchical clustering, which has limited sensitivity and specificity. The approach used by Golub and coworkers focuses on the "signature of interest," like a gene set related to disease, or a gene set related to drug resistance. Knowing how to define a gene signature set is also a key factor to success. The rank-based method used by the authors also helps them query the patterns across multiple platforms.

Does this mean that having a Connectivity Map is equivalent to having the final answer? No. We are still far away from that, and it probably will never be the case. The Connectivity Map will help us to generate hypotheses and to guide our next cell- or lab-based experiments to test these hypotheses, like the authors showed in their examples. Take the drug resistance case—we still do not have a "home run." The intermediate results (based on the cell models in the lab) look encouraging, but the final test will be clinical trials based on these hypotheses. At least now, with this tool, we are not as "blinded" as before.

There is also some distance between identifying target pathways and identifying the true target. The HSP90 example showed that the compounds modulated this pathway to regulate the AR-mediated signaling, but work, probably a significant amount, still remains to identify the target protein in the HSP90 pathway.

These approaches and applications are strong, but why did we chose in the introduction to use the phrase "potential to truly change the speed and efficiency of drug discovery"? This goes back to the initial descriptions of sharing data. The power of the Connectivity Map described by Golub et al. is that it has been built so that it is driven by the use of signatures as the query tool. By distilling a core signature out of an experiment involving a set of compounds or conditions, as opposed to searching the entire signature against other signatures in a critical way, the robustness of the signature apparently frequently increases to where it can be used to link experiments that have not all been done under identical conditions. The power stems from the fact that unrelated experiments might potentially be queried for their component parts, and even better, be queried in the future as new questions and data sets are generated such that the opportunity for open source cell biology can be considered. Work done by Schadt et al. suggests that using core signatures may also be a tool for exploring causality in mapping cellular pathways (Schadt et al., 2005). Both these methods, of course, apply not just to cancer biology but more broadly. The long-term benefits of the Connectivity Maps, therefore, may be that signatures be they RNA-based or protein-based become the currency that empowers biological scientists to recycle their data and to share and build on it in an open architecture context.

Is it premature to declare the utility of the end state? Here again, we biologists should take a page out of the software engineer's rule book and declare that this is just version 1.0. Is it worth the NIH investing in the next version? You bet. Is it worth involving the FDA and companies? You bet. So that expectations aren't inadvertently set too high and later crashed, here is an opportunity for groups to work together as partners. Here is a chance to define together the appropriate scope for version 2.0, be it around a compendium of all FDA drugs, or tackling the safety and efficacy of emerging treatments in a critical path initiative, or even potentially focusing on a treatment modality that could speed a specific treatment opportunity such as prostate cancer. Each of these would be large projects where the synergy would be maximized if academia, industry, and government efforts were able to be linked.

Acknowledgments

Rosetta Inpharmatics LLC is a wholly owned subsidiary of Merck & Co., Inc. S.H.F. and H.D. are both employees of Merck and Co., Inc.

Stephen H. Friend[1,2,*] and Hongyue Dai[2]

[1]Merck Research Laboratories, Merck & Co., Inc., Upper Gwynedd, Pennsylvania 19454
[2]Rosetta Inpharmatics LLC, Seattle, Washington 98109
*E-mail: stephen_friend@merck.com

Selected reading

Hieronymus, H., Lamb, J., Ross, K.N., Peng, X.P., Clement, C., Rodina, A., Nieto, M., Du, J., Stegmaier, K., Raj, S.M., et al. (2006). Cancer Cell *10*, 321–330.

Hughes, T.R., Marton, M.J., Jones, A.R., Roberts, C.J., Stoughton, R., Armour, C.D., Bennett, H.A., Coffey, E., Dai, H., He, Y.D., et al. (2000). Cell *102*, 109–126.

Lamb, J., Crawford, E.D., Peck, D., Modell, J.W., Blat, I.C., Wrobel, M.J., Lerner, J., Brunet, J.P., Subramanian, A., Ross, K.N., et al. (2006). Science *313*, 1929–1935.

Moore, G.E. (1965). Electronics, 38.

Schadt, E.E., Lamb, J., Yang, X., Zhu, J., Edwards, S., Guhathakurta, D., Sieberts, S.K., Monks, S., Reitman, M., Zhang, C., et al. (2005). Nat. Genet. *37*, 710–717.

Wei, G., Twomey, D., Lamb, J., Schlis, K., Agarwal, J., Stam, R.W., Opferman, J.T., Sallan, S.E., den Boer, M.L., Pieters, R., et al. (2006). Cancer Cell *10*, 331–342.

DOI 10.1016/j.ccr.2006.10.011