POLICY

Complexity in Common Diseases: Big Biology for All

John R. Lamb* and Neil W. Gibson

Published 31 March 2010; Volume 2 Issue 25 25cm11

The Commentary of Mills and Sykes in *Science Translational Medicine* presented their thesis on the advent of high-throughput technologies and the dangers they may represent for the future of biomedical research. In response, we argue that true progress on the diagnosis and treatment of common human diseases will require the advent of big biology and its deep integration with focused research as practiced in both academic and industrial institutions.

We read with interest the recent Commentary of Mills and Sykes (*1*) in *Science Translational Medicine* on the advent of high-throughput technologies and the negative influence they may have on the future of biomedical research. The thesis they present is that "focused research" (deep investigations into the function of a single gene product or the regulatory mechanisms of a specific biological process) and "high-throughput research" (genome-scale screens of various kinds) are in competition for limited resources and that there will be undesirable outcomes if the latter is promoted at the expense of the former. We would like to strike a rather different tone by arguing that true progress on the diagnosis and treatment of common human diseases will require the advent of big biology and its deep integration with focused research as practiced in both academic and industrial institutions. This metamorphosis will require changes in funding mechanisms and career reward structures in order to promote collaborative research (Fig. 1).

THE DANGERS OF NEW TECHNOLOGIES

With the availability and increasing affordability of high-throughput genomic technologies, what stresses will be placed on scientific research as it has been practiced to date? Mills and Sykes (*1*) argue that focused research must be protected from the consequences of the easy sell (to funders and the public) and instant gratification (for scientists and investors) of high-throughput work. The underlying idea is that because the successful practitioners of high-throughput approaches will necessarily be skilled in political maneuvering, they will have a higher success rate in acquiring funding, generating high-profile results, and publishing in high-impact journals than will scientists trained in the more traditional single–principal investigator mode. As such, this politicization may lead to a degradation of the quality of science and, in their words, the loss of a generation of appropriately trained scientists. So, to summarize, Mills and Sykes believe that high-throughput science (i) is practiced by politically savvy scientists, (ii) is inappropriately attractive to funding agencies and journal editors, and (iii) will not yield real gains in the diagnosis and treatment of diseases. Furthermore, they fear that if the scientific community as a whole succumbs to the allure of these vast data–generating approaches, public funding for other kinds of (academic) science will dry up.

Fig. 1. Finding focus. Big biology can serve as a roadmap for guiding researchers to their ultimate destination: a more complete understanding of common diseases.

STEPPING STONES OF ADVANCEMENT: OPTIMISM, REALITY, INNOVATION

It is worth stepping back and thinking about the life cycles of new technologies. All technological advancements experience a cycle of initial excitement, which can become overblown and is then followed by a discovery of the limits of the new technology. Only through this process can realistic application of the technology and further innovation occur. From this perspective, continued funding for high-throughput methods will come only with the tangible success of increased biological knowledge, but this will be achieved only by risk taking, the development of novel approaches and analyses, and long-term commitment—in other words, all the components that Mills and Sykes think are missing from high-throughput science.

DeCODE Genetics in fact provides quite a good example of these issues. Their great expertise was to positionally clone human disease genes (a far cry from high-throughput technology), something that they did better than anyone else. The business argument for funding the company was that such genomic information would be valuable to other parties for use in drug development or as predictive biomarkers. The failure came largely from the realization that common diseases arise from many biological changes and that cloning single disease genes describes only a tiny fraction of the genetic variance present in the population. This emerging perspective has also been discussed extensively in the context of genome-wide association studies (GWASs) (*2*). In other words, the biological reality, and not the technology used, meant that what deCODE had to offer was not seen as sufficiently valuable.

COMPLEXITY OF DISEASE MEANS STATISTICS IS IN OUR FUTURE

As scientists have gained the ability to simultaneously measure more and more biological variables, we have come to the realization that common diseases do not generally arise from simple monogenic variants, but rather from

Oncology Research Unit, Pfizer, San Diego, CA 92121, USA.
*Corresponding author. E-mail: John.Lamb@pfizer.com

the combined effects of up to thousands of individual ones (*2*–*4*). Furthermore, mRNA and protein variation in populations of individuals do not occur in isolation but rather as part of an integrated biological system (*3*, *4*). Taken together, this more holistic view of disease biology means that scientists simply will never understand common diseases by considering one variable at a time in isolation. The challenge before us, therefore, is to grasp that complexity in its entirety and try to make sense of it. As a practical example, we currently have only partial knowledge of what the drivers of most common diseases are in the general population (for example, cancer, diabetes, obesity, and Alzheimer's disease). Without a deeper understanding of the drivers of disease progression and the complex interactions within and between them, picking targets for intervention and matching them to patient subpopulations is difficult.

A BOND OF NECESSITY

If we accept this perspective, what is the path forward to a better understanding of human disease? The underlying complexity of disease, the integrative nature of biological systems, and our current collective lack of biological knowledge mean that we must employ high-throughput technologies. However, as Mills and Sykes point out, the simple collection of data and correlation to endpoints will not in and of itself be sufficient to define the best points of intervention for treating common diseases. Although methods for inferring causality from high-throughput data have been and continue to be developed (*3*, *5*, *6*), the validation of these methods has required and will continue to require focused mechanistic research (*7*, *8*). These examples are perhaps instructive: High-throughput technology and data integration led to the identification of genes that appeared to be key drivers of disease. This hypothesis was then tested in a focused manner, by individually knocking out all of these genes in mice and asking whether the predicted alteration in the disease state occurred. Another very well-known example of this series of events is the sequencing of the human genome, which was a hugely expensive exercise when it was done for the first time, but one that is destined to become routine [indeed, the $1000 genome is not far away (*9*)] and has spawned innumerable subsequent studies, from deep population sequencing of clinical samples (*10*) to genome-wide interactome maps (*11*), which in turn have spurred cascades of more focused research. A third, very topical example is the recent description of somatic copy number variation in a broad array of tumor types (*12*), which will probably open the way for many directed studies in academic and industrial laboratories across the world to define the drivers of most tumors. The lesson from these and other examples is that focused and high-throughput approaches in combination advance our knowledge of disease biology, and we argue that it is inappropriate to consider them otherwise.

THE ADVENT OF BIG BIOLOGY AS A RESOURCE FOR ALL

It is all too common that science and scientists are binned into silos, seeing each other as irrelevant or competitive: focused versus high-throughput, academic versus industrial, nonhuman versus human biology, gene versus protein expression, both versus post-translational modification. The realization of the true integrated complexity of biology is a fitting moment, perhaps, to turn the tide on this view and try to bring the diversity of science together. The focus should be on the structures, training, and resources necessary to address the monumental task of making personalized medicine a reality, and this will require the close collaboration of many disciplines and skills sets. As mentioned above, the need for combining many disciplines may require the invention of new operating models in the biological sciences. In the spirit of this, we would like to end here by noting that we, Pfizer Oncology, in collaboration with Merck and Eli Lilly, have established the Asian Cancer Research Group (ACRG), an independent, not-for-profit company formed to accelerate research and ultimately improve treatments for patients affected with the most commonly diagnosed cancers in Asia. Under the agreement, the data will be made available to all, and it is our hope that this movement will form the foundation of a common understanding of the complexity of those diseases from which targeted therapies will emerge. Further, we hope that formation of ACRG will serve as an example of the kinds of innovative organizational structures that are needed to tackle common diseases.

REFERENCES AND NOTES

1. I. G. Mills, R. B. Sykes, Taking risks with translational research. *Sci. Transl. Med.* **2**, 24cm10 (2010).
2. J. H. Moore, F. W. Asselbergs, S. M. Williams, Bioinformatics challenges for genome-wide association studies. *Bioinformatics* **26**, 445–455 (2010).
3. E. E. Schadt, Molecular networks as sensors and drivers of common human diseases. *Nature* **461**, 218–223 (2009).
4. M. Oti, H. G. Brunner, The modular nature of genetic diseases. *Clin. Genet.* **71**, 1–11 (2007).
5. J. Millstein, B. Zhang, J. Zhu, E. E. Schadt, Disentangling molecular relationships with a causal inference test. *BMC Genet.* **10**, 23 (2009).
6. E. E. Schadt, J. Lamb, X. Yang, J. Zhu, S. Edwards, D. Guhathakurta, S. K. Sieberts, S. Monks, M. Reitman, C. Zhang, P. Y. Lum, A. Leonardson, R. Thieringer, J. M. Metzger, L. Yang, J. Castle, H. Zhu, S. F. Kash, T. A. Drake, A. Sachs, A. J. Lusis, An integrative genomics approach to infer causal associations between gene expression and disease. *Nat. Genet.* **37**, 710–717 (2005).
7. Y. Chen, J. Zhu, P. Y. Lum, X. Yang, S. Pinto, D. J. MacNeil, C. Zhang, J. Lamb, S. Edwards, S. K. Sieberts, A. Leonardson, L. W. Castellini, S. Wang, M. F. Champy, B. Zhang, V. Emilsson, S. Doss, A. Ghazalpour, S. Horvath, T. A. Drake, A. J. Lusis, E. E. Schadt, Variations in DNA elucidate molecular networks that cause disease. *Nature* **452**, 429–435 (2008).
8. V. Emilsson, G. Thorleifsson, B. Zhang, A. S. Leonardson, F. Zink, J. Zhu, S. Carlson, A. Helgason, G. B. Walters, S. Gunnarsdottir, M. Mouy, V. Steinthorsdottir, G. H. Eiriksdottir, G. Bjornsdottir, I. Reynisdottir, D. Gudbjartsson, A. Helgadottir, A. Jonasdottir, A. Jonasdottir, U. Styrkarsdottir, S. Gretarsdottir, K. P. Magnusson, H. Stefansson, R. Fossdal, K. Kristjansson, H. G. Gislason, T. Stefansson, B. G. Leifsson, U. Thorsteinsdottir, J. R. Lamb, J. R. Gulcher, M. L. Reitman, A. Kong, E. E. Schadt, K. Stefansson, Genetics of gene expression and its effect on disease. *Nature* **452**, 423–428 (2008).
9. E. R. Mardis, Anticipating the 1,000 dollar genome. *Genome Biol.* **7**, 112 (2006).
10. J. Kaiser, DNA sequencing: A plan to capture human diversity in 1000 genomes. *Science* **319**, 395 (2008).
11. J. F. Rual, K. Venkatesan, T. Hao, T. Hirozane-Kishikawa, A. Dricot, N. Li, G. F. Berriz, F. D. Gibbons, M. Dreze, N. Ayivi-Guedehoussou, N. Klitgord, C. Simon, M. Boxem, S. Milstein, J. Rosenberg, D. S. Goldberg, L. V. Zhang, S. L. Wong, G. Franklin, S. Li, J. S. Albala, J. Lim, C. Fraughton, E. Llamosas, S. Cevik, C. Bex, P. Lamesch, R. S. Sikorski, J. Vandenhaute, H. Y. Zoghbi, A. Smolyar, S. Bosak, R. Sequerra, L. Doucette-Stamm, M. E. Cusick, D. E. Hill, F. P. Roth, M. Vidal, Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **437**, 1173–1178 (2005).
12. R. Beroukhim, C. H. Mermel, D. Porter, G. Wei, S. Raychaudhuri, J. Donovan, J. Barretina, J. S. Boehm, J. Dobson, M. Urashima, K. T. Mc Henry, R. M. Pinchback, A. H. Ligon, Y. J. Cho, L. Haery, H. Greulich, M. Reich, W. Winckler, M. S. Lawrence, B. A. Weir, K. E. Tanaka, D. Y. Chiang, A. J. Bass, A. Loo, C. Hoffman, J. Prensner, T. Liefeld, Q. Gao, D. Yecies, S. Signoretti, E. Maher, F. J. Kaye, H. Sasaki, J. E. Tepper, J. A. Fletcher, J. Tabernero, J. Baselga, M. S. Tsao, F. Demichelis, M. A. Rubin, P. A. Janne, M. J. Daly, C. Nucera, R. L. Levine, B. L. Ebert, S. Gabriel, A. K. Rustgi, C. R. Antonescu, M. Ladanyi, A. Letai, L. A. Garraway, M. Loda, D. G. Beer, L. D. True, A. Okamoto, S. L. Pomeroy, S. Singer, T. R. Golub, E. S. Lander, G. Getz, W. R. Sellers, M. Meyerson, The landscape of somatic copy-number alteration across human cancers. *Nature* **463**, 899–905 (2010).
13. N. W. Gibson is the Chief Scientific Officer of Pfizer Oncology. J. R. Lamb is an Associate Research Fellow at Pfizer Oncology. The authors declare that they have no competing interests.

10.1126/scitranslmed.3000882

Citation: J. R. Lamb, N. W. Gibson, Complexity in common diseases: Big biology for all. *Sci. Transl. Med.* **2**, 25cm11 (2010).

Science Translational Medicine

Complexity in Common Diseases: Big Biology for All

John R. Lamb and Neil W. Gibson

Sci Transl Med **2**, 25cm1125cm11.
DOI: 10.1126/scitranslmed.3000882