Biomedical technology and the clinic of the future

Technology pioneers trade views with a clinician and an entrepreneur on the likely impact of large-scale systems technology in healthcare.

To date, large-scale 'omics data sets and systems approaches in biology have had a relatively minor impact on the practice of medicine. As new technology brings individual genome sequencing closer to reality and large-scale biology continues to progress, opportunities are likely to open up in disease prediction, prevention, diagnosis and treatment. Here the views of two researchers on the potential of disruptive biomedical technologies in clinical practice are contrasted with the perspectives of a clinician and an entrepreneur in commercial clinical information technology.

POINT: Are we prepared for the future doctor visit?

Stephen H Friend & Trey Ideker

Imagine the following visit to the doctor's office, which, although fictitious, is based on technologies that are emerging or already available. A patient, Jane Doe, enters the clinic for a routine physical exam. Today, at least seven parameters would be registered upon her admittance: sex, age, height, weight, temperature, pulse rate and blood pressure (itself a pair of values). But in the future when Jane registers, this set of routine measurements will have expanded enormously (**Table 1**).

Tomorrow's routine checkup

Either on this visit or a previous one, Jane's full genome has been sequenced, noninvasively, using a buccal swab. At the same time, and optionally on every visit, the nurse has sampled and sequenced the metagenome of the microbiome pool resident in the patient's mucosal and gastrointestinal cavities, providing a detailed characterization of the population of microbes commensal with the human host. Messenger RNA, microRNA, proteome and metabolome profiles may be gathered from urine and, if necessary, whole blood and other tissues. Finally, in addition to height and weight, a large panel of physiological parameters and images is monitored, capturing detailed information about respiration, endocrine function, cardiac and brain activity, and so on.

Another key development that will transform Jane's visit to the clinic is deeper data integration. All of the newly gathered information are banked in a unified electronic medical record, which uses a relational database to establish cross-references among the different data types. The new information augments the history of data gathered on previous visits, including all medical treatments and outcomes accumulated over the patient's lifetime.

Crucially, the new data are then integrated with a library of biological network models spanning multiple levels and scales (**Fig. 1**). First is the network of functional and molecular interactions—a.k.a. the molecular wiring diagram—providing a modular, hierarchical and executable view[1] of the cellular processes underlying human health and disease. Such networks are being assembled from diverse large- and small-scale experiments performed over decades of systems biology and biomedical research, providing an up-to-date representation of current knowledge in the field[2,3]. A second type of network model will represent the relevant nosology, which maps relationships between diseases based on their similarities in etiology, pathogenesis and symptoms. Related to this will be another network—that of pharmacologic treatments, which provides rich information about the different protocols and drugs that are available along with their quantitative inter-relationships. One more important network will be the patient's extended social network and pedigree, which will be available along with references to the integrated medical records of friends and relatives. This social network documents significant personal relationships in Jane's life, weighted by importance and, subject to privacy concerns, gathered from social networking websites, personal address books, geographical co-location data, as well as cell phone and e-mail usage. The pedigree provides a complementary set of social relationships that have a genetic basis.

The benefits of these network models to Jane are severalfold. First, they integrate an array of different lines of evidence for health or disease, enabling the formulation of compound biomarkers that are combinations or functions of many simultaneous readouts. Such compound biomarkers can be more robust than biomarkers based on individual genes, proteins or metabolites[4]. Second, the networks provide a natural interpretation of the mechanisms behind Jane's present and future conditions, in contrast to current biomarkers that often have little relation to the actual cause of disease. Third, Jane's data and outcomes can be dynamically analyzed and reintegrated to

Stephen H. Friend is at Sage Bionetworks, Seattle, Washington, USA. Trey Ideker is in the Departments of Medicine and Bioengineering, University of California, La Jolla, California, USA, and at The Institute for Genomic Medicine, University of California, La Jolla, California, USA.
e-mail: friend@sagebase.org or tideker@ucsd.edu

Table 1 Current and emerging genomic technologies for network medicine

Data space	Technologies	Information	Feasibility for patient testing
Genome	Next nth-generation instruments (e.g., reversible dye terminators, sequencing by ligation and pyrosequencing)	Whole genome, including single nucleotide polymorphisms and copy number variants Characterization of patient microbiome	<$1,000 per patient within 2 years Noninvasive (buccal swab)
Epigenome	Chromatin immunoprecipitation sequencing (ChIP-seq), methyl-seq, genome-wide DNase hypersensitivity assays	Chromatin modifications and structure	Distant future: currently used for basic research only
Transcriptome	RNA sequencing, DNA microarrays and bead arrays	Whole genome transcript abundances and translation rates microRNA abundances	<$500 per patient Available now Noninvasive (urine, blood) or invasive (biopsy)
Proteome	Mass spectrometry, multiparameter fluorescence-activated cell sorting (FACS)	Protein abundances and modifications	Predominantly used in basic research
Metabolome	Mass spectrometry (electrospray ionization/ triple quadrupole) NMR, isotope labeling	Metabolic abundances and fluxes	<$200 per patient Well established for neonatal screening Noninvasive (urine, blood)
Protein binding and signaling networks	Immunoprecipitation, co-affinity purification and protein arrays	Protein-protein physical binding interactions, kinase-substrate targeting	Distant future: currently used for basic network assembly
Transcriptional networks	Genome-wide ChIP-seq and protein binding arrays	Protein-DNA, protein-RNA interactions	Distant future: currently used for basic network assembly
Forward and reverse genetic networks	Forward: gene linkage and association studies Reverse genetics: RNA interference screening and combinations, synthetic genetic analysis	Phenotypic profiling, epistatic interactions	Not applicable: networks inferred from populations of individuals

improve the network models themselves. Thus, the impact of a network can increase over time along with the coverage and accuracy of the information it captures. For this reason, all of these network models have been developed using an online public 'commons', which is open-access, crowd-sourced and hosted by a neutral party. The commons serves as a platform for sharing biomedical data, models and tools, including results from extensive clinical trials, ample proteomic and genomic information, proper curation with standard annotations and full assurance that all of the information will remain in the public domain without the constraints of intellectual property (IP). The commons is also a portal by which federal regulators monitor drugs, since, in this future world, therapies are evaluated predominantly by patient-driven trials after their initial approval as safe compounds.

On the basis of Jane's integrated data, multiple indicators are triggered that she is at moderate risk factor 12.7 for breast cancer. The molecular network model indicates both common and rare variants in genes within module 3b.AF8001D, a tumor suppressor module involved in DNA repair and cell cycle checkpoints, resulting in a quantitative decrease in its simulated functional output, which is corroborated by the mRNA and protein expression profiling data. In addition, the system predicts greater than average activation of a key onco-module involved in cell proliferation, which triggers a warning on the nurse's information management console. The entire pattern of network module activity is cross-referenced to the nosology, highlighting a web of diseases for which Jane is at risk and with tubular carcinoma type IIa3 as the most likely outcome. Type IIa3 is a tumor substratification of the future, which can only be identified using molecular profiling data in conjunction with a network model.

Jane's integrated pedigree shows that, although no immediate family members have been diagnosed with similar diseases, two family members at network distances 2 and 3, respectively, have had breast and ovarian cancer. The history for these individuals shows that both were initially placed on preventative treatment with the compound 'aleamed A' but switched to 'aleamed B' after experiencing deleterious side effects, including severe depression. Although Jane's genome sequence places her only at moderate risk for depression, this trait is strongly enriched among the social network of her immediate friends—a finding that raises Jane's own depression risk factor[5]. Thus, aleamed B is recommended as the initial course of action for Jane, or related protocols as indicated by the network of treatments.

Technological possibility or political and social pipe dream?

What are the barriers to making this scenario a reality? Technologies, such as genome sequencing and molecular profiling, are here now (**Table 1**). The required network models—representing connections at the molecular, social, chemical and disease levels—are also available in various forms, although their coverage is far from complete. Clearly, using network maps to develop therapies will require representations of disease that go far beyond the classic biopathway maps so vaunted today. It will require pathophysiological maps that highlight the protein targets lacking in redundancy, such that when altered by drugs these targets modify disease. In turn, these maps will need to highlight unforeseen secondary effects of modifying each potential target.

Assembling and interpreting such integrative network maps will also require that we populate patient records with genotypic and phenotypic changes at scales far beyond our capabilities today. It will require a new class of primary care physician who is proficient in biostatistics, the various data types, networks and modes of integration, and the contribution of each of these components to the overall disease risk and treatment plan. Presently, some of the most forward-looking tests are provided by direct-to-consumer personalized genetics companies[6], but a key challenge faced by such companies is how to provide suitable education to the patient without physician guidance.

However, the proposal we make here is that the most challenging hurdles that will keep this reality from occurring may not be related to technology or education but will be social and political in nature. We acknowledge that the complex technology and informatics meth-

ods that will need to be developed will require massive efforts extending over more than just a few years. At the same time, we anticipate that overcoming the accompanying social and political hurdles will be the more vexing problems, as they will involve addressing issues such as how we will need to work together, how we will need to reward individuals and what we will value.

First and foremost, the future of biomedicine will require that the data are generated and used in a sustainable way. Currently, we fund researchers to perform large clinical studies as if they were indigenous hunter-gatherers. The assumption is that these individuals must not only generate large data sets but should also zealously defend their right to use the data to deliver conclusions that develop the careers of themselves and their laboratories. The data, when finally made available, are often not formatted in a way that is accessible for other investigators to use further, other than as a conclusion. It is as if the patient, who is the actual donor and owner of their data, is sidelined by the biomedical institutions that take on a paternalistic ownership role. Should it be a surprise that this situation typically places the institution's interests and incentives in control of how the data are distributed?

Another driver of current behavior within our medical-industrial complex is the publisher, who wishes to charge for access to the results wrapped within the paper, because this paper is the main scientific currency with which authors are recognized. How can we expect researchers to share their insights before they have written papers, if there are no means to provide them recognition for the actual work itself, including their models and representations of disease? Because the models will require massive amounts of data, building these models will require data sharing in ways that issues of privacy and IP typically obstruct. Dealing with these issues effectively will require that the patients with disease be highly visible. If patients come to better understand the Byzantine cloistering of data that is prevalent today, they will likely demand a shift in culture to one that places the impact squarely on patients, not on the careers of academic investigators.

In a more positive frame, there is enormous potential for the coming tsunami of clinical-genomic data to fundamentally improve the process of developing therapies, which has been atrociously ineffective[7]. Most necessary, we posit, will be to establish a shared infrastructure for the data, tools and models needed to evolve our understanding of disease and its treatment (that is, the online public commons featured in Jane Doe's visit to the doctor

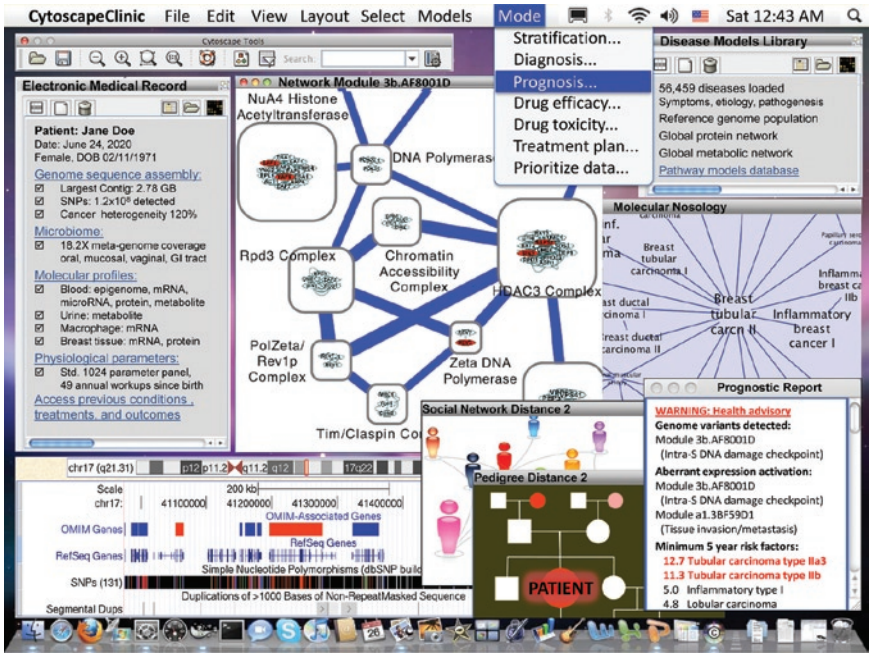

Figure 1 Layers of genomic and network-based information in integrative healthcare. The future primary care physician may need to cope with a staggering array of integrated patient data including genome sequences and biological networks. Access to the full electronic medical record (far left) will provide data at the level of genome sequence (lower left), pedigree and social network (lower center), nosology of disease (far right) and molecular network modules (center). The module 3b.AF8001D is represented as a map of functional interactions among protein complexes, with red nodes indicating proteins for which significant genetic variants were identified. Integrative analysis of these data and model simulation yields a patient prognostic report (lower right). Sequence view is adapted from the UC Santa Clara Genome Browser (http://genome.ucsc.edu/). Network views are from Cytoscape (http://www.cytoscape.org/).

described above). Such a platform must grant unrestricted use of data to develop therapies, and it will benefit greatly from public-private partnerships.

One powerful example of such a partnership within the realm of drug discovery is the Structural Genomics Consortium (SGC) led by Aled Edwards and Chas Bountra[8]. Now 6 years old, this consortium has stimulated sharing of data and models to the extent that the majority of crystal structures solved today no longer have IP attached to them. This is an important example of how a domain of scientific discovery has been transformed—from the traditional assumption that solving structures of targets is a competitive proprietary benefit, to the modern realization that such competitive activities end up crippling all parties because each effort is only a small piece of the whole and has access to only a fraction of the data. Since its inception with a focus on crystal structures, SGC has diversified to tackle other components of basic drug discovery, such as the generation of chemical probes, guided by the same open-access, IP-free philosophy.

A second example of real data sharing is the Coalition Against Major Diseases[9], which has worked to provide open access to clinical trial data from Alzheimer's and other neurological disorders. Patient-led clinical trials, such as those facilitated by PatientsLikeMe or the Life Raft Group, are also a promising direction, provided certain challenges can be met, such as the establishment of appropriate controls. Beyond these needs, it will be essential that information technology companies be shown what a key role they will have in hosting massive amounts of biomedical data and resources in 'the cloud'.

An additional interconnected hurdle relates to the legal friction that the integration of clinical and genomic data will spawn. The desire to capture economic benefits from potential discoveries associated with the data and resulting integrative network models will, if not kept in check, lead to layered legal ownership constraints that could cripple sharing. Avoiding this paralysis will require cooperation among academic institutions, nonprofit foundations, government funders and journals, which set many of the current research rules and reward structures.

If we are going to be able to guide the future care of Jane Doe, we will need to engage in

"institutional analysis" akin to that described by Elinor Ostrum, who won the 2009 Nobel Prize in Economic Sciences[10]. Within the institution of academic research, the most important cultural issues are recognition and reward. We will need to develop robust ways to recognize scientists for their work before, and independent of, publication of journal articles. For example, if we were able to publish models of disease that could be cited by others, then academic institutions might be willing to grant tenure based on the citation impact associated with the models themselves. Similarly, funding agencies might judge potential grantees by the impact of their disease models and, in parallel, set standards for how grantees should share data and models in publicly accessible ways. Such mechanisms could speed the transition to a world in which public access to data and models, as ingredients for future experiments, is not the exception but the rule. It also would greatly help if others were to follow the example set by the Wellcome Trust (London), which has opened discussions about standard legal tools that enable disease-to-therapy projects within an IP-free zone[11]. Here, too, patients as advocates will need to harness their energy and visibility as we navigate the delicate path to robust public clinical-genomic data access while protecting key issues of patient privacy.

Conclusions

In summary, the technologies are here that will entirely transform healthcare. For that reason, it is vitally important that we now focus on realigning the cultural and institutional incentives driving researchers, academic institutions and publishers. The way forward is at least threefold. First, to engage the patients, who must demand methods for data sharing that move past current privacy issues; second, to promote open-access platforms for sharing of data, models and tools; and third, to reward scientists for publication of models, not papers. If these challenges can be met, the future promises to be a world of healthcare honed by data collected from a vast majority of patients being treated in real time.

At the same time, the world of drug discovery will no longer be filled by the top ten pharmaceutical giants of the present day. Instead, these titans will be complemented by a distributed chain of groups who each build a given tool, reagent or product—much closer to the archipelago of software engineers that currently provide applications for iPhones.

It is indeed possible that certain forces—in pharma, in insurance or in hospital administrations—will be aligned against this view. Nonetheless, the tasks described are not impossible, especially if we the people—as citizens, as scientists and as patients—are willing to experiment with how we work together. Don't doubt that the technology will be powerful enough to provide deep understandings. Do doubt whether we are willing to take the cultural and institutional steps to fundamentally change how we work together, and how we share the data and models that will be needed to take advantage of the upcoming opportunities.

ACKNOWLEDGMENTS

We are grateful to G. Siuzdak and S. Choi for helpful comments on the manuscript. T.I. is a David and Lucille Packard Fellow and was funded by a grant from the US National Institutes of Health (RR031228). S.H.F. is funded in part by the National Cancer Institute Centers for Cancer Systems Biology (CCSB) Program and the State of Washington Life Sciences Discovery Fund.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

COUNTERPOINT: Do not opine before it's time

Isaac S Kohane & David M Margulies

Ms. Jane Janus stumbled into the office of Dr. Jill Askepulus pale and sweating. Before the administrative assistant could intercept the unfortunate woman, Dr. Askepulus took her friend by the arm and guided her to a soft landing on her office couch. When Jane had sufficiently recovered, Dr. Askepulus gently asked her what had happened. After a few quavering aborted attempts, she managed to whisper, "I know you warned me, but I went to the Network Integromics Clinic [NIC]." Jane was alternately glum and anxious. She explained to Jill that, of course, she knew she already had a risk of cancer because of her family history of ovarian and breast cancer, but then the NIC had shown her these complicated diagrams, which their physicians informed her demonstrated a high risk that required very close attention. They also had suggested a drug based on the genomic measurements taken at the NIC, which their models suggested could reduce her risk.

Jill paused for a moment, then brought her electronic tablet over to the chair next to Jane and went over with her what appeared to be a prognostication of a track similar to those of hurricanes often seen on the video news. "Jane," she started, "given that you are a professor of mathematics, I figured you could appreciate this. Here," she said pointing to a 95% confidence interval, shaded in red, growing and broadening with age, "is the risk that we know you have and that increases with age for these various cancers. And here are the trajectories that are peeling away from the main risk trajectory under the influence of lifestyle choices, which you and I have already discussed. This broad trajectory in green is the estimated effect of the drug that they suggested to you, and some variations based on different predictive models of cancer based on your genetic markers. Jane stared for a minute at this display and remarked, "I see that I can change my risk somewhat by lifestyle and I do see that this drug might be able to reduce the risk. But I was expecting that all these genomic and proteomic measurements were going to give me a much more accurate and personalized perspective of my medical future. They all seem to overlap a lot." Jill nodded, "They might be much more accurate one day soon, but we have had considerable challenges integrating these various clinical and experimental databases and results from other high-throughput data types to come up with a more accurate prognosis and individualized therapeutic decision-making. We will get there eventually, but the science still has to be worked out and frankly we need more research to be sure our models are accurate. Right now, let's make sure you understand the certainty or lack of it that comes from these various new data types. And let's weigh, with common sense, the preponderance of evidence to date. I could bore you with an accounting of untold suffering that occurred as a result of an insufficiently informed use of tests such as prostate specific antigen, mammograms or urinary screening for neuroblastoma. But I won't. Let's talk about how we are going to make the right decision for you, with you."

Isaac S. Kohane is at the Harvard Medical School Center for Biomedical Informatics and Children's Hospital Informatics Program, Boston, Massachusetts, USA. David M. Margulies is at Correlagen Diagnostics Inc., Waltham, Massachusetts, USA.
e-mail: isaac_kohane@harvard.edu

An alternative view

The above slightly tongue-in-cheek sequel to the scenario proposed by Friend and Ideker is provided to emphasize where we believe the

current challenges lie. To be sure, increased openness, transparency, data sharing and academic rewards for team and multidisciplinary behavior are important ingredients in developing a vibrant and productive biomedical discovery establishment. However, they do not constitute structural impediments to the translation of genome-scale measurements into safe clinical practice.

Moreover, although we have a long way to go, historical trends point to steady progress towards openness and collaboration. This includes an ever-widening fraction of open-access publications with steadily rising impact, the opening to a world of researchers of cohort studies (e.g., the Framingham Study and the Gene Expression Omnibus storing the data of over half a million microarrays), each measuring tens of thousands of genes. It includes the evidence of the increased impact and frequency of large multinational studies with hundreds of authors; historic achievements such as trial registries like clinicaltrials.gov, which even now are being upgraded to include more primary data; multiple consumer-driven, data-sharing efforts, from the corporate, such as PatientsLikeMe, to purely voluntary and extensive social network support groups. Already, biomedical research groups are discussing publication formats that follow the lead of our colleagues in astronomy that include the full data within the publication document itself[12]. We can cheer on these efforts, but the translation of existing and future 'massively parallel' measurements to clinical-grade decision support and therapeutics remains a methodological and scientific challenge for which there has been far less progress than the sociological trends appropriately lauded by Friend and Ideker.

More pressing challenges

What are the components of this most pressing and thorny challenge in achieving meaningful, clinical-grade, integrative medicine that leverages the various data types enumerated by Friend and Ideker? First, we have to develop suitable technical methods and user interaction models to integrate the diverse data sources. Although there are isolated instances of integration of, for example, expression data with underlying pathways, or expression data in the context of specific somatic genome variation, there is no general purpose architecture or model for integrating the complexity of data types with physiology and anatomy over time.

Second, we have to ensure that what we know is accurate. That is, we have to clean up our existing evidentiary knowledge base. For example, of the at least 150,000 genomic variants documented to have some import to disease, a substantial minority have not been reproduced or have been contradicted by subsequent reports.

Third, we have to ensure that we know what is known. In the context of a medical education system that is already straining to keep physicians informed of best practices using only a few thousand clinical variables, the challenge of supporting sound and efficient decision-making in the context of millions of variants will require substantial progress in data reduction, user interfaces and automated support.

> Increased openness, transparency, data sharing and academic rewards for team and multidisciplinary behavior do not constitute the primary structural impediments to the translation of genomic measurements into safe clinical practice.

Fourth, we have to know whether we can safely proceed to clinical decision-making from computer models that are not completely based on human clinical trials, randomized or observational. That is, can our models achieve the same mechanistic and predictive qualities as the Henderson-Hasselbalch equation for acid-base equilibrium, the Frank-Starling Curve for cardiac contractility or at least the Framingham cardiovascular risk scores? If not, are they only useful for hypothesis exploration rather than clinical care? Breakthroughs in both measurement and modeling technology may be required to achieve clinical-grade soundness of our models.

Fifth, there will need to emerge regulatory clarity around the use of data displays of this complexity. Who will decide whether what we think we know is safe? What are the boundaries of the US Food and Drug Administration's (FDA) so-called 'IVDMIA' (*in vitro* diagnostic multivariate assay) threshold? How will the regulatory framework of the FDA and its international analogs cope with models as complex as those of *in silico* airplane design?

Finally, how much better is our new knowledge than older knowledge? When is the incremental benefit of a genomic variant(s) or gene expression profile relative to a family history or classic histopathology insufficient and when does it add rather than subtract variance? If we are able to rationalize the selection of cancer chemotherapeutic agents by integrating information about responsiveness of cells with specific cell expression profiles, that would be an important 'emergent' benefit of deep integration. But it is important that we identify potential transformative benefits to focus and prioritize data integration efforts.

The clinical perspective exemplified by these questions poses substantial challenges. We do not doubt that our biomedical research community is up to successfully addressing them, some even in the very near term.

Like our colleagues, we are excited to be able to collaborate in integromic research that we are convinced will benefit many who suffer from disease. And like Friend and Ideker, we are optimistic that the trends to collaboration and transparency, already underway, can only help.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

1. Fisher, J. & Henzinger, T.A. *Nat. Biotechnol.* **25**, 1239–1249 (2007).
2. Friend, S.H. *Clin. Pharmacol. Ther.* **87**, 536–539 (2010).
3. Chuang, H.Y., Hofree, M. & Ideker, T. *Annu. Rev. Cell Dev. Biol.* **26**, 721–744 (2010).
4. Ideker, T. & Sharan, R. *Genome Res.* **18**, 644–652 (2008).
5. Rosenquist, J.N., Fowler, J.H. & Christakis, N.A. *Mol. Psych.* published online, doi: 10.1038/mp.2010.13 (16 March 2010).
6. Wagner, J.K. *Am. J. Hum. Genet.* **87**, 451–456 (2010).
7. David, E., Tramontin, T. & Zemmel, R. *Nat. Rev. Drug Discov.* **8**, 609–610 (2009).
8. Weigelt, J. *EMBO Rep.* **10**, 941–945 (2009).
9. Romero, K. *et al. Clin. Pharmacol. Ther.* **86**, 365–367 (2009).
10. Ostrum, E. *Understanding Institutional Diversity* (Princeton University Press, Princeton, NJ, USA, 2005).
11. Friend, S.H. *Scientist* **24**, 22–32 (2010).
12. Goodman, A.A. *et al. Nature* **457**, 63–66 (2009).