Joel Dudley, PhD

Submitted for publication December 30, 2012. Accepted in revised form April 7, 2013.

*Correspondence: Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Pl., Box 1498, New York, NY 10029-6574 (e-mail: joel.dudley@mssm.edu).

Abbreviations and Acronyms

ALK = anaplastic lymphoma receptor tyrosine kinase

CSS = cultured skin substitute

DNA = deoxyribonucleic acid

eQTL = expression quantitative trait loci

GEO = Gene Expression Omnibus

GOBP = gene ontology biological process

ICD = International Classification of Disease

MEMN = macrophage enriched metabolic network

MET = mesenchymal epithelial transition factor

mGlu = metabotropic glutamate (receptor)

(continued)

Translational Bioinformatics Approaches to Drug Development

Ben Readhead^{1,2} and Joel Dudley^{1-3,*}

¹Center for Biomedical Informatics, ²Institute for Genomics and Multiscale Biology, and ³Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York.

Significance: A majority of therapeutic interventions occur late in the pathological process, when treatment outcome can be less predictable and effective, highlighting the need for new precise and preventive therapeutic development strategies that consider genomic and environmental context. Translational bioinformatics is well positioned to contribute to the many challenges inherent in bridging this gap between our current reactive methods of healthcare delivery and the intent of *precision medicine*, particularly in the areas of drug development, which forms the focus of this review.

Recent Advances: A variety of powerful informatics methods for organizing and leveraging the vast wealth of available molecular measurements available for a broad range of disease contexts have recently emerged. These include methods for data driven disease classification, drug repositioning, identification of disease biomarkers, and the creation of disease network models, each with significant impacts on drug development approaches.

Critical Issues: An important bottleneck in the application of bioinformatics methods in translational research is the lack of investigators who are versed in both biomedical domains and informatics. Efforts to nurture both sets of competencies within individuals and to increase interfield visibility will help to accelerate the adoption and increased application of bioinformatics in translational research.

Future Directions: It is possible to construct predictive, multiscale network models of disease by integrating genotype, gene expression, clinical traits, and other multiscale measures using causal network inference methods. This can enable the identification of the “key drivers” of pathology, which may represent novel therapeutic targets or biomarker candidates that play a more direct role in the etiology of disease.

SCOPE AND SIGNIFICANCE

A MAJORITY OF medical interventions occur late in the pathological process, when treatment outcome can be less predictable and effective, highlighting the need for new precise and preventive therapeutic development strategies that consider genomic and environmental context. Translational bioinformatics is well positioned to contribute to the many challenges inherent in bridging this

gap between our current reactive methods of healthcare delivery and the intent of *precision medicine*, particularly in the areas of drug development, which forms the focus of this review.

TRANSLATIONAL RELEVANCE

A range of powerful informatics methods have been developed to organize and capitalize on the deluge of molecular level information avail-

able in a range of disease contexts, in areas as diverse as disease classification, drug repositioning, identification of disease biomarkers and the creation of disease network models, each with significant impacts on drug development approaches.

CLINICAL RELEVANCE

The approaches discussed are relevant to multiple domains of clinical practice, and exist as a diverse array of potential and realized translational endeavors. While some methods (*e.g.*, the use of disease biomarkers to drive treatment choices) already reflect a standard of care within certain areas of medicine, other areas (*e.g.*, the modeling of causal disease networks) which do not yet routinely impact clinical care, are expected to add insight to the identification of high-quality drug targets to inform drug development, as well as guiding the interpretation of individual genetic variation in creating tailored therapeutic strategies.

BACKGROUND

Healthcare in the 21st century faces a unique set of challenges in facing rising healthcare costs and declining research and development productivity in therapeutic discovery and development. Until 2010, costs have been rising faster than economic growth in many Organization for Economic Cooperation and Development countries,¹ which has positioned healthcare expenditure as a key economic focus. As innovation of new treatments, rising provider costs, and an increased prevalence of expensive, chronic conditions press treatment costs upwards, the expanding, ageing population in many developed nations is reflected in an increased demand for medical care.

An important driver of this cost growth is the partial success of many existing treatments; medical inter-

ventions which lead to so-called “chronification” of complex, previously fatal diseases which can now be modified or at least managed symptomatically for many years (*e.g.*, ischemic heart disease, type 2 diabetes, cerebrovascular disease and chronic obstructive pulmonary disease), with the incidence of such diseases increasing with age.² Despite the benefits of being able to offer such treatments to patients, the intervention and maintenance of treatment at this level of disease progression is inefficient and expensive. Of course, historically there has been little alternative but for a patient become symptomatic before seeking clinical review, as the combination of clinically observable signs and reported symptoms have been our primary view into a patient’s physiology, and it is usually only once a significant system perturbation has occurred that an individual will notice anything unusual. However, in many pathologies, predisease conditions can now be identified (*e.g.*, prehypertension, prediabetes, mild cognitive impairment), states which imply increased risk of progression to full-blown disease for a patient, and in some cases, with an accompanying opportunity to engage in early therapeutic strategies (*e.g.*, exercise prescriptions, dietary changes, medications) which can reduce the risk of progression. Some of the key challenges in this paradigm of disease screening are in identifying predisease states which represent significantly increased risk of future adverse health outcomes, identifying reliable markers of these predisease states, and developing useful therapeutic strategies to offer individuals should they receive a predisease diagnosis. With healthcare costs amounting to US\$2.6 trillion in 2010, and three out of four treatment dollars being spent on the management of chronic conditions, the need for new approaches

Abbreviations and Acronyms (*continued*)

NSCLC = non small cell lung cancer

PAH = pulmonary artery hypertension

R&D = research and development

RNA = ribonucleic acid

SNOMED-CT = Systematized Nomenclature of Medical Terms-Clinical Terms

SNP = single nucleotide polymorphism

TTC = tissue-tissue coexpression

to address these issues has never been more pressing.³

The concept of precision medicine, an approach where a patient's health traits are interpreted alongside state-of-the-art molecular profiles, to develop accurate diagnostic, prognostic and therapeutic strategies which are tailored to reflect individual physiological context, has been put forth as a model for transitioning towards a more safe, effective, and efficient healthcare paradigm.⁴ Precision medicine offers an expectation of an improved power to diagnose disease at earlier stages, facilitating treatment initiation at more stable phases of pathogenesis. Recent advances in methods for gathering and analyzing genomic and other molecular level data across many technology platforms, and different biological contexts (*e.g.*, different diseases, populations, tissues, or time periods⁵) serves as the groundwork for enabling precision medicine.

The development of safe and effective targeted therapeutics is a core component of precision medicine; however, the current industry development pipeline is not presently well poised to meet this need. The number of Food and Drug Administration applications and subsequent approvals of new therapies has been stable for the past decade (Fig. 1) reflecting the growing costs and increasing failure rates of clinical trials.⁶ Explanations for this trend are not universally agreed upon, but

Figure 1. Development of new pharmaceuticals. Despite a general trend towards increased R&D spending by U.S. pharmaceutical companies,⁶⁷ the number of new molecular entity and new biologic entities by the U.S. Food and Drug Administration (FDA)⁶⁸ has remained steady over the past decade. As demonstrated, the number of new applications has also remained steady, suggesting an increased difficulty developing new compounds beyond the early phases of drug development. R&D costs converted to 2012 U.S. dollars.⁶⁹ R&D, research and development; NME, new molecular entity; NBE, new biological entity. To see this illustration in color, the reader is referred to the web version of this article at www.liebertpub.com/wound

some cite an exhaustion of figurative “low-hanging fruit” in drug development opportunities, forcing the pursuit of riskier development choices, and the requirement to deepen our understanding of the basic biological pathways which underlie disease processes.⁷ While target based screens, cell-based assays, and genetic studies are capable of generating many potential drug target leads, in many cases we aren’t yet capable of interpreting the greater biological and physiological context of the target modulatory effects, and thus, are limited in the ability to exploit such leads. How then, can we link the expectations of precision medicine to current paradigms in drug discovery? In this review, we discuss approaches in which bioinformatics can be used to assist these challenges, with a particular focus on *translational bioinformatics*, a field which focuses on “the development of storage, analytic and interpretive methods to optimize the transformation of increasingly voluminous biomedical data into proactive, predictive, preventative, and participatory health.”⁸

First, we outline the manner in which integrating high-dimensional data drawn from multiple levels of biological hierarchy can be used to create new classification systems for disease, and in turn allow the inference of new insights on the basis of shared features between diseases. Next, we discuss the ways in which translational bioinformatics has helped to move beyond population-level evaluation of new therapies, to identify subpopulations of patients which may derive benefit from a therapy while saving others from needless, costly or even harmful interventions. We also describe the ways in which translational bioinformatics can leverage available genomic, molecular, clinical, and other data sources to create integrated frameworks, which can be used to identify new indications for existing therapies. Furthermore, we also review some of the tools that have been developed to integrate matched DNA sequence data, gene expression data, and disease trait data to construct probabilistic causal models of disease, which allow the identification of the true drivers of the molecular networks that underlie disease.

Although the main focus of this review is the applicability of translational bioinformatics to drug development in general, we will conclude by reviewing some of the existing bioinformatic approaches which have been applied to the complex field of wound healing, and discuss some of the ways in which the methods described in this article could be utilized for novel approaches, such as transcriptomic profiling to identify wound subtypes and guide treatment decisions, or to

identify new therapeutic leads to inform drug development. The methods described here are also relevant to the extent that they can be used to improve diagnostic and therapeutic strategies for the range of systemic diseases which have an impact on wound healing, such as peripheral vascular disease, type 2 diabetes and venous hypertension.

DISCUSSION OF FINDINGS AND RELEVANT LITERATURE

Disease taxonomy and drug development

The value of creating a framework by which to classify diseases was apparent in the time of Hippocrates, though wasn't developed in a systematic way until the Enlightenment, initially by Sauvages⁹ and then built upon by Linnaeus.¹⁰ The complexity of modern disease taxonomies have increased significantly since then, though the underlying principles of classification on the basis of observable attributes remains similar. Our current disease classification systems reflect the diverse requirements of stakeholders, including classification for processing of reimbursement (e.g., International Classification of Disease [ICD]), research purposes (e.g., Medical Subject Headings), specialist reporting (e.g., Systematized Nomenclature of Medical Terms-Clinical Terms [SNOMED-CT]), and clinical classification (ICD, Diagnostic and Statistical Manual IV). These taxonomies tend to reflect the attributes of disease that are routinely measured, which is itself a function of convenience and utility. Therefore, current systems classify diseases according to constellations of clinical, anatomical or investigative findings. Such methods have undoubtedly been beneficial, but can also be misleading, by clustering essentially dissimilar pathologies (which may share organs systems or clinical manifestations) or separating similar ones (which may appear distinct at a clinical level). Loscalzo *et al.*¹¹ discuss the examples of diseases which form clinically distinct syndromes yet emerge from the same molecular basis, and conversely, diseases that have multiple genetic bases, yet converge into a common phenotype. Sickle cell anemia, a monogenic disease, with a well-characterized, single genetic basis, can clinically manifest along a spectrum of clinical severity ranging from mild anemia to veno-occlusive stroke, dependent on additional genetic and environmental risks.¹¹ This is contrasted with diseases, such as familial pulmonary artery hypertension (PAH), which are associated with many different mutations affecting members of the transforming growth fac-

tor beta receptor superfamily, yet tending towards a common phenotype.¹²

There have been multiple efforts to classify systems of diseases according to their molecular level correlates, including investigation by the National Academy of Science into creating a knowledge network of disease to underpin a new taxonomy of disease.¹³ At a disease class level, Sirota *et al.*¹⁴ examined Genome Wide Association Study results to determine genetic variation profiles associated with a range of autoimmune disease, identifying sets of diseases which appeared to cluster together, conferring some protective effects against diseases in other classes. They reported individual single nucleotide polymorphism (SNP) and gene associations that appear to drive the similarities and differences between different clusters of diseases. Interestingly, they found that some disease pairs (e.g., Crohn's disease and rheumatoid arthritis), which share some treatments, and clinical features, possess different underlying genetic profiles suggesting a differing molecular basis. Suthram *et al.*¹⁵ used protein interaction networks to identify functional modules, which were then assessed in the context of gene expression microarray data for 54 different diseases. Each disease was comprised of a signature of "module response scores", which were used to cluster diseases. Interestingly, sets of modules relating to key biological functions were found to be perturbed in many diseases. This same set of pan-disease modules was also enriched for proteins identified as pluripotent drug targets (*i.e.*, drugs targeting these proteins tended to have many indications), suggesting the relevance of targets within these modules for a wide range of therapeutic uses.

The highlighted works demonstrate an important principle—the potential for data-driven molecular taxonomy approaches to challenge and improve existing models of disease taxonomy. Such studies offer the possibility to share insights between "similar" diseases with the potential of identifying common disease pathways (with associated drug repositioning opportunities), as well as by looking at dissimilar pairs, which can lead to the identification of exclusive pathways, with associated opportunities for further hypothesis-led studies to identify novel treatment targets (Fig. 2).

Despite an increasing ability to compare and contrast different disease entities at a range of biological levels of organization, an important challenge lies in translating this into clinical decision-making tools. While new disease descriptions and

Figure 2. Carving disease at the joints. With an increase in our collective ability to measure the different aspects of disease traits, comes an opportunity to rethink the ways in which we describe diseases, and classify them into communities, which reflect their underlying biology. **(a)** By incorporating available sources of sufficiently accurate, high-dimensional data which is entangled with a pathological process, **(b)** we will be able to produce disease signatures, which will allow comparison between instances of a disease, and between different diseases. **(c)** This will enable a classification system, which encodes the difference and similarities between diseases, according to the dynamic molecular networks distributed throughout key tissues/organ systems, and at useful multi-omic levels. **(d)** Such differences and similarities are useful to the extent that they catalyse new understandings of disease, including the transfer of knowledge between similar diseases, and the identification of key underlying principles of difference between dissimilar diseases. To see this illustration in color, the reader is referred to the web version of this article at www.liebertpub.com/wound

definitions may be scientifically informative, their utility is only maximized if they can be used to guide expectation of clinically relevant outcomes (such as disease prognosis or therapeutic response). With current methods, a lower bound on the time to identify such associations is formed by the natural time course of the disease process under investigation.

Beyond population-level evaluation of therapies

Along with our ability to measure huge amounts of genetic and other molecular data has come the realization that in many cases, certain molecular

traits can be used to predict important variability between patients in clinically relevant areas, such as treatment response, side effect profiles or toxicity risk. The translation of such approaches into panels of diagnostic markers to guide therapeutic strategies is known as companion diagnostics and forms a key component in the growing toolkit of personalized and precision medicine.¹⁶ This carries a clear set of advantages to the patients who are predicted to benefit from the therapy, and in the context of oncology, where its application is most commonly seen, tangible benefits to those who won't benefit, by saving the time, financial and physical costs of needless therapy (Fig. 3).

Figure 3. Beyond population-level evaluation of therapies. Traditionally, assessment of pharmaceutical safety and efficacy has been constrained by the application of a new therapy to a single class of patients, usually sharing a broad diagnosis. **(a)** This level of assessment necessitates combining heterogeneous patient subtypes, which often possess different treatment responses or side-effect and toxicity risks. **(b)** By combining trial data with a range of rich biological data on each patient, investigators can now try to find predictive features, which will segregate with important aspects of treatment response. Such data can consist of single gene genotyping, through to protein biomarker concentrations, to identifying disease module gene activity signatures. Identifying these kinds of predictive markers empowers the tailoring of specific therapeutic strategies on the basis of a patient's biology. Overall, this approach can allow the retention of otherwise useful medications that might be predictably toxic, or perhaps only useful in a subset of patients. To see this illustration in color, the reader is referred to the web version of this article at www.liebertpub.com/wound

The number of companion diagnostics on offer is growing rapidly, occurring in regulated and unregulated contexts. Common examples of companion diagnostics include the use of human epidermal growth factor receptor 2 immunohistochemistry and gene-amplification tests to guide the use of trastuzumab in the treatment of metastatic breast cancer, and the c-kit immunohistochemistry test to guide the use of imatinib in the treatment of gastrointestinal stromal tumors.¹⁶

Lung cancer remains the commonest cause of cancer related death worldwide.¹⁷ Since 2004,

treatment options for non small cell lung cancer (NSCLC, the most common histological type) have grown rapidly, largely as a result of increased knowledge of driver mutations which can then be used to guide appropriate treatment. Such approaches have demonstrated the value of evaluating the efficacy of medications in light of molecular subtypes, rather than broad disease class. An early drug trial of kinase inhibitors gefitinib and erlotinib initially demonstrated no clear overall benefit when applied to a cohort of patients with the broad diagnosis of NSCLC.

However, it was recognized at the time that a subpopulation of $\sim 10\%$ of patients actually experienced dramatic antitumor results, which was a sufficient basis for Food and Drug Administration approval in 2003, despite the fact that the specific underlying mechanism conferring sensitivity wasn't yet known.¹⁸ Later, it was found that positive treatment response correlated with activating epidermal growth factor receptor mutations, which could then be used as a distinct biomarker to guide initiation of treatment.^{19–22} This biomarker enabled a more effective trial design, with optimization of therapeutic indications; thus, reducing unnecessary side effects and costs in patients that wouldn't benefit from the therapy.

A more recent example, centers on the story of crizotinib, another kinase inhibitor (particularly *MET* [mesenchymal epithelial transition factor] and *ALK* [anaplastic lymphoma kinase] activity) used in the treatment of NSCLC. Part way through a clinical trial assessing its efficacy on a broad population of lung cancer patients,²³ an independent team published findings of a particular chromosomal translocation affecting the *ALK* gene with an association to tumor growth in a subset of NSCLC patients.²⁴ This allowed modification of the clinical trial to assess treatment efficacy within the frame of this particular variant, with dramatically different results, showing an increase in median survival time to about a year, as opposed to a few months with the standard of care. Such targeted approaches also offer the potential benefit of identifying useful therapeutic responses on smaller cohorts of patients.

It is clear that historically, a level of serendipity has contributed to these discoveries, suggesting that many additional opportunities exist, and have perhaps been missed. The increased use of systematic, bioinformatics approaches to correlate molecular and genomic level patient data with treatment response and adverse outcomes and drug toxicity is likely to enable more routine and accelerated discovery in relevant patient subpopulations. Such approaches could potentially leverage any accessible, meaningful biological elements, from single gene genotyping,²⁵ to non-coding RNA expression,²⁶ to multi-module disease gene expression signature.²⁷

Computational drug repositioning

The practice of identifying additional therapeutic indications for existing drug compounds is referred to as drug repositioning, and has some key benefits over traditional methods of drug

development. Estimating the average costs of bringing a drug to market is a complicated undertaking, though estimates placed capitalized costs between approximately US\$800 million²⁸ and US\$1.8 billion²⁹ with an accompanying timeline of 15 years.³⁰ A large portion of these costs are associated with early stages of development and toxicity testing, with over 90% of tested compounds failing to progress beyond this stage.³¹ The identification of additional indications for existing medications offers clear time and cost benefits, thereby avoiding many of these early hurdles.

Traditional approaches to drug development usually focus on identification of a novel treatment target, followed by a search for a compound capable of appropriately modulating the target. A lengthy and costly target validation process then follows target identification. For a given compound, additional targets are not usually investigated, and additional clinical applications, are not routinely explored.³² This represents a huge opportunity for the systematic identification of new indications for existing therapeutics—some of the most commercially successful medications are administered for different indications than envisioned in their initial development. Prominent examples amongst many include sildenafil (an antiangina medication now used for treatment of erectile dysfunction and PAH), minoxidil (anti-hypertensive, now used for alopecia) and thalidomide (antinausea, now used for multiple myeloma and erythema nodosum leprosum).

Multiple approaches have been used in drug repositioning efforts, ranging from blind screening of libraries of drug compounds against model systems, to data-driven computational approaches that integrate and search across links between drugs and diseases using various forms of biomolecular and clinical data. Here we focus on computational approaches, which offer substantial advantages in cost and speed of discovery compared to experimental screening based approaches.

In general, computational approaches exploit the known links between diseases and drugs, with the accompanying possibility that shared attributes between drugs used to treat the same disease, or diseases which share treatments, implies a degree of meaningful similarity between some aspects of the linked objects, which can be used to generalize existing treatments into new clinical contexts. Dudley *et al.*³² describe a means to classify computational approaches to drug repositioning according to whether the mode of inference is

Figure 4. Drug repositioning strategies. Many computational approaches to drug repositioning capitalize on known relationships between diseases and drugs, with the accompanying possibility that shared features between drugs used to treat the same disease, or diseases that share treatments, implies a degree of meaningful similarity between some aspects of the linked entities, which can then be used to generalize existing treatments into new clinical contexts. A number of emerging methods integrate multiple sources of disease and drug based knowledge, which can enable even more sophisticated inferences about new therapeutic opportunities. (Figure reprinted with permission from Dudley *et al.*³²)

drug-based or disease-based (Fig. 4). We will briefly outline these approaches, giving examples of each.

Drug-based approaches to drug repositioning

Approaches using drug similarity. Drug similarity approaches leverage known properties of existing drug compounds to compare and classify on the basis of similarity in their chemical properties. Such approaches use a variety of chemical descriptors, such as QSAR models (quantitative structure-activity models) and pharmacophores to allow comparison and clustering of drugs which share similar properties.³³ Such clustering can then allow the inference of similarity between drugs, and identification of potential novel therapeutic applications.

Noeske *et al.*³⁴ used this conceptual approach to identify new treatment targets for a group of known metabotropic glutamate (mGlu) receptor agonists by preparing pharmacophore descriptors for the antagonists as well as a range of additional known compounds, and used self-organizing maps to create a compound map. They reported subclusters containing mGlu antagonists, with proximity to additional compounds known to bind to dopamine and histamine receptors. This was then used as the basis for further experiments, which confirmed the predicted interactions between mGlu receptor antagonists and Histamine Receptor 1 and dopamine receptor 2 targets.

Keiser *et al.*³⁵ developed a quantitative approach for clustering protein targets on the basis of

similarity amongst their associated ligands. Their “Similarity Ensemble Approach” used 65,000 ligands grouped into sets on the basis of known interactions with hundreds of protein targets. A pairwise ligand similarity score was calculated between every member of each set and every other member of every other set, to derive the inter-set similarity, which reflected the similarity of the protein targets at the core of each set. In this way, protein targets can be linked even when they don’t share any common ligands. Despite utilizing a protein similarity score derived entirely from chemical similarity of ligands, biologically related protein clusters nevertheless emerged, alongside unexpected links, such as methadone antagonizing muscarinic M3 receptors and loperamide antagonizing neurokinin N2 receptors (these predictions were then confirmed experimentally). With its potential to link proteins which are functionally related, yet share nothing in the way of structural or sequence similarity, this ligand-centric approach represents a valuable tool for deriving knowledge about biological function which can circumvent some of the limitations of approaches which rely solely on protein similarity metrics.

Drug similarity approaches are limited by a number of factors, including the requirement for accurate surveys of chemical properties for any drug under consideration and closed access to these types of information, often as part of the intellectual property strategy of a pharmaceutical company which has invested heavily in the development of its compounds. In addition, the effect of a

drug is much more complex than its initial chemical properties. Many drugs undergo extensive metabolic transformations into their active components, and such processes aren't always well characterized, or well accounted for in pharmacophore descriptions.³⁶

Approaches using molecular activity similarity. Another exciting approach to drug repositioning arises by characterizing drug compounds according to their impact on molecular activity. Such an approach frames a compound as a perturbation to the system it is introduced to, which can be used to identify a characteristic signature for that compound. This can then be used to compare many medications, with many opportunities to generalize therapeutic indications between "similar" drugs.

One of the most widely used resources in this area is the Connectivity Map,^{37,38} a repository for the whole genome transcriptional response of 1,309 small bioactive molecules applied to a range of cultured human cells. Iorio *et al.*³⁹ used the connectivity map to perform pairwise comparisons between all compounds, followed by clustering to identify drug communities with cluster co-location, implying possible repositioning opportunities. By leveraging these associations, they predicted previously unknown autophagy activity for fasudil, a Rho-kinase inhibitor, which was later validated experimentally.

Sirota *et al.*⁴⁰ used an integrated approach which compared drug signatures, with analogous disease signatures to identify repositioning opportunities. Leveraging publicly available microarray gene expression data, the authors looked at 100 diseases, and 164 drug compounds, hypothesizing that if a given disease is characterized by a gene expression signature, then drugs which induce an inverse signature may have therapeutic value, while drugs with similar signatures may exacerbate the disease. Using this approach, the authors rederived many known therapeutic applications, such as the efficacy of prednisolone in treatment of Crohn's disease and ulcerative colitis. In addition, they predicted therapeutic benefit for the over-the-counter H₂-receptor antagonist cimetidine (commonly used in gastric reflux and peptic ulcer disease to reduce gastric acid secretion) in the treatment of lung adenocarcinoma. They then validated this prediction *in-vitro* using cell lines, and *in-vivo* using xenograft mouse models.

It is clear that such approaches aren't constrained to the use of gene expression data. Any method which provides meaningful, high-dimensional description of a drug compound can be considered.

Chen *et al.*⁴¹ used PubChem⁴² to create signatures on the basis of bioassay results for PubChem compounds. They then constructed a bipartite network that mapped PubChem profiles onto biological networks (protein interaction networks and metabolic pathways) on the basis of sequence similarity between compound protein targets and the proteins represented in the biological networks. This allowed interpretation of compound efficacy or adverse effects to occur in a rich biological context, which can form the basis of further hypothesis led experiments.

The primary limitations of molecular similarity approaches are their reliance on the quality of data that are used to form the disease and drug signatures, as well as the simplifications which must be made in forming the disease signatures. For example, Connectivity Map reference signatures are derived from specific cell lines, which may be irrelevant to the actions of certain compounds, as well as neglecting the often complex metabolism which compounds undergo from the time of their administration to their arrival in distant cells. Additional simplifications occur in the transcriptional classification of many diseases, which are known to affect multiple tissue types and organ systems (such as peripheral vascular disease), or diseases where the relevant tissue types may not be obvious (hypertension) or amenable to easy sampling (Alzheimer's disease).

Approaches using molecular docking. Molecular docking is a simulation-based approach that uses 3D structural protein information to characterize the interaction interface between drug compounds and targets. This can then be used as the basis for making predictions about whether drug compounds will interact with novel treatment targets. If a compound is predicted to interact with a novel treatment target, then an inference can be made that the compound may be useful in any diseases in which the treatment target is implicated.

In an interesting molecular docking approach, Kinnings *et al.*⁴² started with approved drugs, and identified binding sites from the 3D structural data of their known targets, using these sites as the basis for a similarity search with other proteins. Additional proteins that were found to possess similar sites were highlighted as predicted additional targets for the drug compound, allowing the repositioning of entacapone (most frequently used in Parkinson's disease to inhibit the metabolism of levodopa) as a treatment for multi-drug resistant tuberculosis.

Such approaches rely heavily on having access to 3D structural information of drugs and their ligands. Although many such structures are available and growing rapidly, some very significant omissions exist, including many members of the G-protein coupled receptor superfamily.

Disease-based approaches to drug repositioning

Approaches using associative indication transfer. Another approach to drug repositioning which has been recently explored is one which leverages known associations between drugs and diseases, on the assumption that drugs which share many disease indications may share mechanisms of action; thus, allowing the generalization of putatively similar drugs into additional clinical contexts.⁴³ As has been noted previously,³² this interesting approach is currently constrained by lack of systematic knowledge of the therapeutic intent which links a particular drug to a particular disease. For example, the treatment of an infectious disease may require concomitant administration of paracetamol for relief of fever and an antibiotic for the underlying infection—two treatments which clearly engage with the underlying pathology at different levels.

Approaches using shared molecular pathology. Some repositioning strategies proceed on the assumption that diseases that share drugs may do so because of common aspects in their molecular pathophysiology, implying a shared responsiveness to the therapeutic functions of the drug. On the basis of this, drugs may be repositioned into additional diseases that look quite different at the phenotypic level. The approach is analogous in many ways to methods which make use of shared molecular similarity, in fact a number of interesting studies have emerged which combine the two approaches to cluster diseases and drugs into communities to identify repositioning opportunities, including some already discussed.^{15,40} Hu and Agarwal⁴⁴ used Gene Expression Omnibus (GEO) data to formulate disease signatures, and then clustered these using a correlation based approach. These signatures were then integrated with Connectivity Map data to cluster disease molecular pathology and drug mechanisms.

Understandably, such approaches are limited by the technologies that are used to characterize a pathophysiological state (*e.g.*, gene expression microarray). Additional current limitations reside in capturing and meaningfully comparing the multi-

tissue, multi-organ and dynamic nature of many diseases over time.

Approaches using side effect similarity. Another novel repositioning approach consists of linking drugs together on the basis of their side effect similarity. On the assumption that such side effect profiles provide a read-out of the physiological effect of a compound on a host system, Campillos *et al.*⁴⁵ mapped medication side-effects from package leaflets, and mapped the listed terms to the Unified Medical Language System, creating weighted associations between side effects and drugs on the basis of side-effect frequency. Inter-drug similarity was then calculated by the sum of common weights, allowing formation of drug communities, from which repositioning opportunities could be inferred.

Approaches that rely on the integration of data represented as a structured language are subject to the constraints of codifying such a complex, heterogeneous, and ambiguous domain as healthcare delivery into a formalized system. Some of the more common issues include discrepancies in categorization of concepts and the development of multiple representations of the same concept (*e.g.*, SNOMED-CT possesses a concept for “Apgar score at 1 minute” as an *observable* entity, as well as a separate concept for “Apgar score at 1 minute” as a *finding* entity)⁴⁶ which can lead to ambiguities in interpreting the semantic context of a given term.

Such limitations are also compounded by the precision with which side effect profiles are recorded. Many clinical signs and symptoms are quite non-specific, and the level of granularity with which side-effects are elicited could have significant effects on characterizing a drug. For example, a reported side-effect of abdominal pain, if not characterized further may imply mechanisms as diverse as gastric reflux, hepatotoxicity or drug-induced diarrhea.

Many promising computational drug repositioning approaches exhibiting quantifiable success have emerged in the past decade. It's also clear from the described approaches that techniques can be devised which exploit any sufficiently high-dimensional description of a disease or its known efficacious treatments. Some of the most promising approaches have been ones which integrate multiple strategies to various degrees: utilizing high throughput data describing therapies, as well as diseases to create communities of drug-disease interactions, with the possibility of linking these to other forms of rich biological data, such as

protein interaction networks, DNA and protein sequence data, transcriptional and metabolic networks. Future approaches may also see the use of much higher-level phenotypic data, extracting features from imaging studies, electroencephalogram, or wireless monitoring devices as the basis for comparing disease and drug similarity.

Causal network models of disease

An emerging perspective in systems medicine is that to understand biology, and identify the true drivers of pathology in individual patients will require the construction of high-dimensional networks representing multiple levels of biological hierarchy.⁴⁷ In this view, core biological processes are mediated by networks of interacting molecules (DNA, RNA, proteins, metabolites, *etc.*) in a manner which is dynamically coordinated within and between tissues, and in a highly context dependent manner. This systems medicine view frames these molecular networks as key sensors to environmental and genetic perturbation,⁴⁸ shifting their state in response to a myriad of context dependent cues, and in the frame of pathogenesis, representing intermediate molecular phenotypes which mediate the flow of risk between disease associated DNA variants and disease states.⁴⁹

One of the great challenges implicit in developing such an understanding is the extraction of the relevant networks from an immensely noisy molecular background, to identify the molecular underpinnings of normal and pathogenic processes. Some exciting computational approaches have been developed, which integrate matched DNA variation, gene expression, and clinical data to construct probabilistic causal networks⁵⁰—models which capture functional relationships between key biological objects (such as gene regulatory networks) and phenotypic data (such as disease traits), to approach a mechanistic understanding of disease and allow the differentiation of reactive elements from the true drivers of disease (Fig. 5). Bayesian network approaches are commonly employed to learn the conditional probability distributions of molecular and clinical variables, and then used to solve conditional probability equations to infer probable causal networks.⁵¹ Bayesian networks often infer many equally possible network graphs as solutions.⁵¹ This quickly leads to nontrivial computational and methodological challenges in exploring the complete space of all probable network graphs. Other popular approaches for network reconstruction are gene coexpression networks, or

methods that first group genes and loci into functionally coherent groups called “modules”.⁵¹ Causal inference is then performed to infer networks of functional gene modules.

The Macrophage Enriched Metabolic Network (MEMN) is an example of a disease network derived by integrating gene expression, genotype and clinical data from adipose and liver tissue, in mice and humans, which was found to be causally associated with phenotypic traits of metabolic disease, such as diabetes, heart disease and obesity.⁴⁹ The authors combined co-expression network analysis, with a statistical procedure called “likelihood-based causality model selection”⁵⁰ to identify whether patterns of gene expression were causal for, or reactive to the phenotypic disease traits. The MEMN was shown to be strongly causal for the disease traits, facilitating the identification and validation of three previously unknown type 2 diabetes associated genes.

A key approach in constructing such networks, builds on weighted gene co-expression network analysis⁵² to identify sets of highly co-expressed genes in a biological context of interest. These gene modules can then be converted into a network through consideration of pair-wise gene co-expression thresholds. Schadt *et al.*⁵⁰ demonstrated that these correlation-derived association networks could be converted into causal networks (where the direction of an interaction can be known) by integrating them with a systematic form of biologically meaningful perturbation—DNA sequence variation. By utilizing genetic loci that are correlated with mRNA transcript abundance (expression quantitative trait loci [eQTL]), the authors developed a statistical approach that considers an eQTL alongside a pair of genes with transcript levels that correlate with sequence variation of eQTL. By analyzing conditional correlations between these three variables, it is possible to infer the most likely relationship between the gene pair (either gene A regulates gene B, gene B is regulated by gene A, or gene A and B do not regulate each other).

Once such directed networks of gene regulatory relationships are available, it becomes possible to identify the key drivers of that network, and thus, the disease trait it has been constructed to reflect.⁵⁰ By functionally annotating the network members, and identifying the enriched functional pathways which it is associated with, it becomes possible to work upstream into the network to identify genes which regulate many downstream members implicated in this particular function (local drivers) and ultimately to identify the most upstream genes

Figure 5. Building causal models of disease. One of the exciting possibilities in systems medicine is the capacity to build models of disease which reflect the causal interactions underlying disease at the molecular level. These models can then be used to identify new diagnostic, prognostic and therapeutic approaches. One approach utilizes **(a, b)** gene expression microarray data from disease tissue to **(c)** identify modules of genes which are coexpressed, **(d)** and for each of these modules, derive an undirected network which best accounts for the pair-wise gene expression correlations. **(e)** By layering in matched DNA variation data for each patient, it becomes possible to transform the undirected network, into a causal network, which encodes the direction of functional regulatory relationships between its genes. By considering expression quantitative trait loci (genomic locations where sequence variation correlates with mRNA transcript abundance of one or more genes) alongside a pair of its associated genes, it is possible to look at conditional correlations to find the most likely causal model connecting the eQTL and the gene pair. **(f)** By iterating over all relevant gene pairs (with their associated eQTLs) in all the modules, it becomes possible to create a probabilistic causal network, and identify “local drivers” (blue)—genes which regulate many relevant downstream genes and even the most upstream, so-called global drivers (dark blue). This allows the identification of potentially high leverage molecular targets, facilitating the prioritisation of high-value validation experiments. eQTL, expression quantitative trait locus. To see this illustration in color, the reader is referred to the web version of this article at www.liebertpub.com/wound

(global drivers).⁵³ This set of drivers represents a very high leverage set of targets that can then be investigated for disease diagnosis, monitoring and drug targeting purposes.

Such networks also offer the opportunity for *in-silico* screening of drug compounds by simulating the extended effects a given drug may impose on the network. In this way, unfavorable side effect profiles, or worthwhile therapeutic potential can be anticipated. Utilizing the network comprising the MEMN, Chen *et al.*⁴⁹ predicted that Ppm1l (at the

time, a poorly characterized protein phosphatase) had the potential to improve insulin resistance (highlighting its potential as a treatment target for Type 2 diabetes); however, due to its network position, it was also likely to increase weight gain, and increase blood pressure (therefore worsening overall cardiovascular risk). These predictions were then verified in Ppm1l knock-down mice, which demonstrated an improved diabetic profile, but with increased weight gain and blood pressure. This method of prioritizing leads for streamlined

Figure 6. Intertissue molecular networks. Dobrin *et al.*⁵⁴ uncovered a TTC network between hypothalamus, liver, and adipose tissue in the context of obesity, deriving biologically relevant interactions which weren't present within any single tissue type. Center: Depiction of the *network backbone*, comprised of a limited number of highly correlated genes, where variation in one gene is likely to induce expression variation in a connected gene in a complementary tissue. Node coloring represents membership to different TTC subnetworks, with enrichment for GOBP terms of relevance to obesity traits. (Figure reprinted from Dobrin *et al.*⁵⁴) GOBP, Gene Ontology Biological Process; TTC, tissue-tissue coexpression. To see this illustration in color, the reader is referred to the web version of this article at www.liebertpub.com/wound

experimental validation represents a very promising advance in early drug development.

The importance of recognizing and accounting for inter-tissue interactions within such disease networks is also a critical factor, particularly for complex diseases that span multiple tissues. The majority of investigations into the relationships between DNA variation and the intermediate phenotypes associated with complex disease have traditionally focused on single tissue analysis, which doesn't account for the richly interrelated networks which occur between different tissue types. Dobrin *et al.*⁵⁴ investigated multi-tissue co-expression networks in obesity, with a focus on inter-tissue networks between the hypothalamus, liver and adipose tissue (Fig. 6). Their work revealed biologically sensible networks of relevance to obesity, which weren't present in any single specific tissue, yet clearly emerged through a systems approach. The insights that can be built upon

these types of methods are expected to prove useful in better capturing disease complexity, as a dynamic entity occurring within a richly interconnected network of organs and tissue, with consequences for our ability to understand pathology and ultimately guide drug development.

SUMMARY

As treatment costs and demand for healthcare rise, it has become imperative to investigate new means of providing high quality care to patients. It's increasingly clear that we cannot continue to manage disease at the level that we have in the past. Translational bioinformatics is well poised to make key contributions to this complex set of challenges, in particular to the identification of novel, high-leverage therapeutic targets, optimizing the allocation of existing treatments to disease, and in the formation of biologically meaningful models of physiology and disease.

Table 1. *Upregulated genes*

<i>Gene Symbol</i>	<i>Adjusted p-Value</i>	<i>Fold Change</i>	<i>Biological Function</i>
<i>PTX3</i>	6.73×10^{-7}	31	Zymosan binding
<i>IGF2BP3</i>	1.79×10^{-6}	26	RNA binding; mRNA 3'-UTR binding; mRNA 5'-UTR binding; nucleotide binding; protein binding; translation regulator activity
<i>S100A12</i>	4.31×10^{-6}	22	RAGE receptor binding; calcium ion binding; protein binding; zinc ion binding
<i>MMP1</i>	7.90×10^{-6}	20	Calcium ion binding; metalloendopeptidase activity; peptidase activity; zinc ion binding
<i>PI3</i>	7.90×10^{-6}	20	Endopeptidase inhibitor activity; serine-type endopeptidase inhibitor activity
<i>MME</i>	8.03×10^{-6}	20	Metal ion binding; metalloendopeptidase activity; metallopeptidase activity; peptidase activity; peptide binding; protein binding; zinc ion binding
<i>MMP3</i>	1.33×10^{-5}	19	Calcium ion binding; metalloendopeptidase activity; peptidase activity; zinc ion binding
<i>PI3</i>	1.33×10^{-5}	18	Endopeptidase inhibitor activity; serine-type endopeptidase inhibitor activity
<i>WISP1</i>	1.61×10^{-5}	18	Insulin-like growth factor binding; protein binding
<i>SLC16A1</i>	1.64×10^{-5}	18	Mevalonate transmembrane transporter activity; monocarboxylic acid transmembrane transporter activity; protein binding; secondary active monocarboxylate transmembrane transporter activity; symporter activity
<i>TCN1</i>	1.74×10^{-5}	18	Cobalamin binding
<i>ADAM12</i>	1.84×10^{-5}	17	SH3 domain binding; metal ion binding; metalloendopeptidase activity; metallopeptidase activity; peptidase activity; protein binding; zinc ion binding
<i>KCNJ15</i>	1.89×10^{-5}	17	Inward rectifier potassium channel activity; potassium channel activity; protein binding; voltage-gated ion channel activity
<i>HMGA2</i>	2.56×10^{-5}	16	AT DNA binding; DNA binding; protein binding
<i>KRT6B</i>	2.79×10^{-5}	16	Structural constituent of cytoskeleton
<i>KRT16</i>	4.92×10^{-5}	15	Protein binding; structural constituent of cytoskeleton
<i>IL6</i>	5.08×10^{-5}	14	Cytokine activity; cytokine activity; growth factor activity; interleukin-6 receptor binding; contributes_to interleukin-6 receptor binding; interleukin-6 receptor binding
<i>TMEM158</i>	5.23×10^{-5}	14	Peptide binding
<i>SLC6A14</i>	5.73×10^{-5}	14	Amino acid transmembrane transporter activity; neurotransmitter:sodium symporter activity; symporter activity
<i>MME</i>	6.42×10^{-5}	14	Metal ion binding; metalloendopeptidase activity; metallopeptidase activity; peptidase activity; peptide binding; protein binding; zinc ion binding
<i>SERPINB3</i>	7.22×10^{-5}	14	Peptidase inhibitor activity; protein binding; serine-type endopeptidase inhibitor activity
<i>DSC2</i>	7.22×10^{-5}	14	calcium ion binding
<i>SLC16A1</i>	7.31×10^{-5}	13	Mevalonate transmembrane transporter activity; monocarboxylic acid transmembrane transporter activity; protein binding; secondary active monocarboxylate transmembrane transporter activity; symporter activity
<i>DSC2</i>	7.31×10^{-5}	13	Calcium ion binding
<i>ANPEP</i>	7.40×10^{-5}	13	Aminopeptidase activity; metal ion binding; metallopeptidase activity; peptidase activity; peptide binding; receptor activity; zinc ion binding

Upregulated genes arranged according to fold change, with adjusted *p*-values (Benjamini-Hochberg corrected⁷⁰) and associated Gene Ontology biological functions.⁷¹

We have discussed a range of translational bioinformatics approaches which can contribute to these challenges at multiple levels, including the appropriate characterization and classification of diseases on the basis of their underlying molecular features, enabling the transfer of insights between pathologies deemed to be similar, and aiding the search for underlying principles which may separate pathologies deemed to be distant. One implication of this type of disease map is the possibility to identify specific biomarkers that correlate with disease subtypes, with relevance for choice of therapy, and allowing prediction and prevention of unfavorable side-effect profiles or toxicity effects.

We have reviewed some of the translational bioinformatics approaches for computational drug repositioning, which has the potential to powerfully leverage from pre-existing drug development efforts to expand and reshape the catalog of therapies which are currently available. We can see

that current data-driven methods can be readily adapted to utilize increasingly voluminous, high-quality biological data to identify ever more precise attributes of disease and therapy, which can be used to generalize between therapeutic contexts.

We have also provided a brief overview of the evolving area of probabilistic causal network modeling of disease, which integrates genotype, expression and clinical data to provide deep insights into disease physiology and facilitating the development of powerful new therapeutic options. We expect that causal network modeling of disease will play a key role in building truly predictive models of natural disease progression in the near future. As the cost of molecular profiling technologies continues to decrease, it will become tractable to score multiscale molecular traits from individuals over time, and thus, become possible to characterize the dynamic changes driving network states from health to disease.⁵⁵

Table 2. Downregulated genes

<i>Gene Symbol</i>	<i>Adjusted p-Value</i>	<i>Fold Change</i>	<i>Biological Function</i>
<i>P2RY14</i>	5.41×10^{-7}	-38	G-protein coupled receptor activity; UDP-activated nucleotide receptor activity; G-protein coupled purinergic nucleotide receptor activity; receptor activity
<i>DCT</i>	5.41×10^{-7}	-36	Copper ion binding; dopachrome isomerase activity; isomerase activity; metal ion binding; oxidoreductase activity
<i>CD74</i>	5.41×10^{-7}	-35	MHC class II protein binding; beta-amyloid binding; cytokine binding; cytokine receptor activity; identical protein binding
<i>C1QA</i>	5.41×10^{-7}	-34	Complement activation; cell-cell signaling; innate immune response
<i>MRC1L1</i>	6.03×10^{-7}	-33	Binding; receptor activity; sugar binding; mannose binding
<i>HLA-DRA</i>	6.63×10^{-7}	-32	MHC class II receptor activity
<i>FCER1A</i>	6.73×10^{-7}	-31	IgE receptor activity; receptor activity
<i>TYROBP</i>	1.58×10^{-6}	-28	Protein binding; receptor signaling protein activity
<i>HLA-DRA</i>	1.62×10^{-6}	-27	MHC class II receptor activity
<i>TYRP1</i>	1.79×10^{-6}	-26	Copper ion binding; metal ion binding; oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, another compound as one donor, and incorporation of one atom of oxygen; protein binding; protein heterodimerization activity
<i>F13A1</i>	1.79×10^{-6}	-26	Acyltransferase activity; metal ion binding; protein-glutamine gamma-glutamyltransferase activity; transferase activity
<i>LYZ</i>	1.79×10^{-6}	-26	Hydrolase activity, acting on glycosyl bonds; lysozyme activity; protein binding
<i>CDHR1</i>	1.92×10^{-6}	-25	Calcium ion binding; receptor activity
<i>LOC100133811</i>	1.92×10^{-6}	-25	DNA binding; metal ion binding; zinc ion binding; endonuclease activity; hydrolase activity; nucleic acid binding; pancreatic ribonuclease activity; ribonuclease activity; MHC class II receptor activity
<i>RGS5</i>	2.54×10^{-6}	-24	GTPase activator activity; signal transducer activity
<i>MLANA</i>	2.71×10^{-6}	-24	Protein binding
<i>PECAM1</i>	2.71×10^{-6}	-24	Protein binding
<i>CCL27</i>	3.09×10^{-6}	-23	Chemokine activity
<i>PER3</i>	5.64×10^{-6}	-22	signal transducer activity
<i>NTRK2</i>	7.90×10^{-6}	-20	ATP binding; neurotrophin binding; neurotrophin receptor activity; nucleotide binding; protein binding; transferase activity; transmembrane receptor protein tyrosine kinase activity
<i>FBXW7</i>	7.90×10^{-6}	-20	Protein binding
<i>HLA-DRB1</i>	7.90×10^{-6}	-20	MHC class II receptor activity
<i>N4BP2L1</i>	7.90×10^{-6}	-20	ATP binding
<i>LAPTM5</i>	8.03×10^{-6}	-20	Transport
<i>DCT</i>	8.03×10^{-6}	-20	Copper ion binding; dopachrome isomerase activity; isomerase activity; metal ion binding; oxidoreductase activity
<i>ANO1</i>	1.23×10^{-5}	-19	Intracellular calcium activated chloride channel activity; ion channel activity
<i>CD93</i>	1.27×10^{-5}	-19	Calcium ion binding; complement component C1q binding; protein binding; receptor activity; sugar binding
<i>HLA-DPB1</i>	1.74×10^{-5}	-18	Protein binding
<i>HCLS1</i>	1.74×10^{-5}	-18	DNA binding; SH3 domain binding; protein complex binding; transcription factor activity
<i>DCT</i>	1.88×10^{-5}	-17	Copper ion binding; dopachrome isomerase activity; isomerase activity; metal ion binding; oxidoreductase activity
<i>DARC</i>	1.91×10^{-5}	-17	G-protein coupled receptor activity; chemokine receptor activity; receptor activity; transmembrane receptor activity
<i>TNXB;TNXA</i>	2.08×10^{-5}	-17	Heparin binding; integrin binding; molecular function
<i>RGS5</i>	2.22×10^{-5}	-17	GTPase activator activity; signal transducer activity
<i>PTPRC</i>	2.22×10^{-5}	-17	Hydrolase activity; protein binding; protein kinase binding; protein tyrosine phosphatase activity; transmembrane receptor protein tyrosine phosphatase activity
<i>LOC100133811</i>	2.22×10^{-5}	-17	DNA binding; metal ion binding; zinc ion binding; endonuclease activity; hydrolase activity; nucleic acid binding; pancreatic ribonuclease activity; ribonuclease activity; MHC class II receptor activity
<i>CBX7</i>	2.41×10^{-5}	-17	Chromatin binding; protein binding
<i>NFIB</i>	2.41×10^{-5}	-16	Transcription factor activity
<i>FUCA1</i>	2.41×10^{-5}	-16	Alpha-L-fucosidase activity; hydrolase activity, acting on glycosyl bonds; sugar binding
<i>DBP</i>	2.67×10^{-5}	-16	RNA polymerase II transcription factor activity; protein dimerization activity; sequence-specific DNA binding; transcription factor activity
<i>HLA-DMA</i>	2.79×10^{-5}	-16	Protein binding
<i>TPSAB1</i>	2.91×10^{-5}	-16	Peptidase activity; protein binding; serine-type endopeptidase activity; serine-type peptidase activity
<i>LOC100133811</i>	3.23×10^{-5}	-16	DNA binding; metal ion binding; zinc ion binding; endonuclease activity; hydrolase activity; nucleic acid binding; pancreatic ribonuclease activity; ribonuclease activity; MHC class II receptor activity
<i>CRIP1</i>	3.69×10^{-5}	-15	Metal ion binding; zinc ion binding
<i>RGS5</i>	3.73×10^{-5}	-15	GTPase activator activity; signal transducer activity
<i>HBA2;HBA1</i>	3.84×10^{-5}	-15	Heme binding; metal ion binding; oxygen binding; oxygen transporter activity; protein binding
<i>MS4A4A</i>	3.87×10^{-5}	-15	Receptor activity
<i>HLF</i>	3.90×10^{-5}	-15	DNA binding; double-stranded DNA binding; protein dimerization activity; sequence-specific DNA binding; transcription factor activity
<i>CHP2</i>	4.40×10^{-5}	-15	Calcium ion binding
<i>CD14</i>	4.45×10^{-5}	-15	Lipopolysaccharide binding; lipoteichoic acid binding; opsonin receptor activity; peptidoglycan receptor activity; protein binding

(continued)

Table 2. (Continued)

Gene Symbol	Adjusted p-Value	Fold Change	Biological Function
CLDN5	4.45×10^{-5}	-15	Identical protein binding; structural molecule activity
RABGAP1L	4.61×10^{-5}	-15	GTPase activator activity; Rab GTPase activator activity
CYP39A1	4.61×10^{-5}	-15	24-Hydroxycholesterol 7 α -hydroxylase activity; electron carrier activity; heme binding; metal ion binding; monooxygenase activity; oxysterol 7- α -hydroxylase activity; oxysterol 7- α -hydroxylase activity; steroid 7- α -hydroxylase activity
TNXB	4.92×10^{-5}	-15	Heparin binding; integrin binding; molecular function
TSPAN7	4.92×10^{-5}	-15	Interspecies interaction between organisms
NFIB	4.92×10^{-5}	-15	Transcription factor activity
TPSB2	4.96×10^{-5}	-15	Peptidase activity; protein binding; serine-type endopeptidase activity; serine-type peptidase activity; peptidase activity; protein binding
NMU	5.08×10^{-5}	-14	Neuromedin U receptor binding; receptor binding
PTPRC	5.08×10^{-5}	-14	Hydrolase activity; protein binding; protein kinase binding; protein tyrosine phosphatase activity; transmembrane receptor protein tyrosine phosphatase activity
GIMAP6	5.42×10^{-5}	-14	GTP binding; nucleotide binding
GATA3	5.62×10^{-5}	-14	DNA binding; metal ion binding; sequence-specific DNA binding; transcription factor activity; transcription factor binding; zinc ion binding
ZFPM2	5.65×10^{-5}	-14	DNA binding; RNA polymerase II transcription factor activity; metal ion binding; transcription corepressor activity; transcription factor binding; zinc ion binding
LDB2	5.88×10^{-5}	-14	LIM domain binding; enzyme binding; transcription cofactor activity
CXCR4	5.91×10^{-5}	-14	C-C chemokine receptor activity; C-X-C chemokine receptor activity; G-protein coupled receptor activity; actin binding; coreceptor activity; myosin light chain binding; protein binding; ubiquitin protein ligase binding
TPSB2	6.36×10^{-5}	-14	Peptidase activity; protein binding; serine-type endopeptidase activity; serine-type peptidase activity
CEP68	6.42×10^{-5}	-14	Centrosome organisation
BCL2	6.73×10^{-5}	-14	BH3 domain binding; identical protein binding; protease binding; protein binding; protein heterodimerization activity; protein homodimerization activity; sequence-specific DNA binding; transcription activator activity; transcription factor binding
HLF	7.22×10^{-5}	-14	DNA binding; double-stranded DNA binding; protein dimerization activity; sequence-specific DNA binding; transcription factor activity
HBA2	7.22×10^{-5}	-14	Heme binding; metal ion binding; oxygen binding; oxygen transporter activity; protein binding
DKK2	7.31×10^{-5}	-14	Multicellular organismal development
TNXB	7.31×10^{-5}	-13	Heparin binding; integrin binding; molecular_function
A2M	7.31×10^{-5}	-13	Enzyme binding; growth factor binding; interleukin-1 binding; interleukin-8 binding; protein binding; serine-type endopeptidase inhibitor activity; tumor necrosis factor binding
GPR116	7.36×10^{-5}	-13	G-protein coupled receptor activity
SYNE2	7.40×10^{-5}	-13	Actin binding; protein binding

Downregulated genes arranged according to fold change, with adjusted *p*-values (Benjamini-Hochberg corrected⁷⁰) and associated Gene Ontology biological functions.⁷¹

Table 3. Connectivity map compound hits

Compound	Connectivity	p-Value	Specificity
0198306-0000	-0.88	0	0
Cromoglicic acid	-0.88	0.03	0.04
Mimosine	-0.81	0.01	0
Prestwick-682	-0.77	0.01	0
Benzocaine	-0.76	0.01	0.03
Disulfiram	-0.74	0	0
Ondansetron	-0.74	0.01	0.01
Dimenhydrinate	-0.73	0.01	0.01
Levcycloserine	-0.72	0.01	0.04
5194442	-0.71	0.01	0.03
Glycocholic acid	-0.71	0.02	0.01
Moracizine	-0.67	0.03	0.03
Ketorolac	-0.67	0.03	0.05
Dropropizine	-0.66	0.03	0.02
Cefadroxil	-0.64	0.04	0
Sulfamethoxazole	-0.63	0.02	0.02

Using query signature based on Tables 1 and 2, the top 16 "hit" compounds ordered according to connectivity score, with negative values denoting an opposite transcriptional profile to our query signature. Low specificity values refer to a unique connectivity between the query signature and the listed compounds.

PERSPECTIVES FOR WOUND HEALING

A variety of studies have looked at the complex field of wounds and wound healing using bioinformatics approaches, including transcriptional and proteomic profiling of different wound areas,⁵⁶ different wound types in animal models⁵⁷ and human subjects.⁵⁸ These kinds of approaches are valuable steps in elucidating the underlying molecular pathways, which may be aberrant, particularly in chronic wounds. Such studies could also lay a foundation for further works that place such molecular wound signatures within the landscape of other diseases, to allow for comparisons between different wound subtypes and other phenotypically different diseases, to identify opportunities for adapting existing therapies. An example of a hypothesis led approach was described which explored similarities between chronic wound healing and stroma tissue in cancers;⁵⁹ however, to our

knowledge an approach integrating profiles from a wider range of disease has not yet been performed. Other interesting bioinformatic wound studies have included the use of transcriptomic profiling to identify wound edge biomarkers which could then be used to guide surgical debridement.⁶⁰ Such approaches could be adapted to identify topical therapeutics that could help guide enzymatic debridement, or to identify therapeutics with complementary transcriptional profiles. Additional interesting directions may include the use of transcriptomic or proteomic profiling to identify different molecular subtypes of wound, to guide therapy choice, or a data-driven comparison between wound signatures and compound molecular signatures to identify new topical or systemic treatments.

Tables 1–3 provide an example of a bioinformatics approach to wound care. We used publicly available microarray data, collected by Smiley *et al.*,⁶¹ (stored in the GEO repository,⁶² GSE3204) in an investigation of the transcriptome of cultured skin substitute (CSS), an autologous culture of patient keratinocytes and fibroblasts which can facilitate healing in large burns. Although useful, such CSS only contain two cell types, and are thus, limited in their capacity to emulate normal skin. We hypothesized that identifying the difference signature between CSS and normal skin may highlight an opportunity for an existing drug compound to act as an adjunctive therapy alongside CSS, to augment its transcriptional response in the direction of normal skin.

We used the web-based tool GEO2R⁵ to rederive the most differentially expressed genes, reflecting the difference signature between CSS and normal skin cells. We extracted the 100 most differentially expressed genes, which comprised 25 upregulated (Table 1) and 75 downregulated genes (Table 2), which we used as the basis of a query signature, for use with Connectivity Map.^{37,38} We present the top 16 compounds (Table 3) which were found to have an “opposite” transcriptional

TAKE-HOME MESSAGES

- A key driver of healthcare expenditure is the treatment of chronic disease, accounting for three out of four healthcare dollars spent in the United States.
- Most interventions for chronic conditions occur late in the pathological process, at a time when a patient is symptomatic, when intervention is less predictable, less effective, and thus, more costly.
- Translational bioinformatics is concerned with “the development of storage, analytic and interpretive methods to optimize the transformation of increasingly voluminous biomedical data into proactive, predictive, preventative, and participatory health.”
- Precision medicine represents an approach where a patient’s disease traits are interpreted alongside state-of-the-art molecular profiles, to develop accurate diagnostic, prognostic and therapeutic strategies that are tailored to reflect individual biological contexts.
- A key milestone in the progress of medicine toward the ideals of precision medicine will be the organization of a new classification system for disease, which reflects the set of clinically meaningful and exploitable similarities and difference between disease traits. Such a map will likely integrate high-dimensional signatures reflecting network level activity for a range of key biological networks (gene regulatory networks, RNA networks, protein networks, *etc.*), as well as higher-level data, such as clinical and investigative findings.
- Translational bioinformatics can exploit known associations between diseases and treatments, to reposition existing medications into new therapeutic contexts.
- Translational bioinformatics can use molecular level patient data (DNA sequence or structural variation, protein concentrations, *etc.*) to help identify subpopulations within groups of patients that share a broad diagnosis. These subpopulations can then be stratified and/or assessed separately when testing new therapies, to identify useful markers that are predictive for important clinical outcomes, such as treatment responsiveness, risks of toxicity, *etc.*
- By integrating genotype, gene expression and clinical data collected on groups of animals or humans with a disease trait, scientists can build causal models of the molecular underpinnings of disease, effectively separating its causal and reactive components.
- This can enable the identification of the true drivers of pathology, which may represent high-leverage treatment targets, and help prioritize time-consuming experimental validation.
- An important bottleneck in the application of bioinformatic methods to clinical domains is the presence of scientists conversant in both fields. Creative solutions are required to nurture both sets of competencies within individuals, as well as to increase interfield visibility, to facilitate cross-pollination between domain experts within each field.

signature to the CSS signature, comprising a selection of well characterized, approved medications, as well as a number of experimental compounds. We suggest that these results could form the basis of hypothesis led studies investigating the role of adjuvant therapies in the application of CSS to large healing wounds.

Additional ways in which translational bioinformatic approaches are being applied to the field of wound care, is in the contribution of better treatments of the diseases which predispose to chronic wounds, such as type 2 diabetes,⁴⁹ venous hypertension and peripheral arterial disease.^{63,64}

One of the key challenges facing the systematic integration of bioinformatic approaches with clinical medicine is the relative paucity of “bilingual” practitioners—computer scientists with a sense of front-line clinical priorities, and clinicians familiar with a broad range of computational tools and approaches. To remain current in any one field is a full-time undertaking, so dual trainees are still a limited resource. An ideal scenario would be one in which individuals possess enough biological knowledge to ask clinically useful questions, and also possess a command of existing computational and bioinformatic approaches to be able to address this question in a scientific way. Indeed, one of the major challenges presented by large-scale, integrative informatics approaches is the fact that these approaches enable discovery and evaluation of large number of hypotheses, and it can require significant clinical domain expertise to reduce or focus the set of hypotheses evaluated by informatics approaches.

Some universities have begun integrating bioinformatic content into their curriculum, including units on translational bioinformatics⁶⁵ and interpretation of whole genome sequencing,⁶⁶ for bioinformaticists, PhDs, genetic counselors and medical practitioners alike. An initial approach may also lie in broad, cross-field educational modules, for example, a series of reviews directed towards clinicians to provide outlines of the field of bioinformatics, without necessarily focusing on developing the computational skills to build specific solutions. A value inflection point will occur when clinicians possesses some fluency in the capabilities and limits of the available bioinformatics tools, and are able to recognize and articulate some broad applications to their areas of clinical expertise. Likewise, a similar approach could be envisioned where a survey of key clinical needs within medical domains are made available

to bioinformaticians. This may then help catalyze conversations and collaboration with domain experts in both fields, informing the next iteration of tool-building and biological inquiry.

AUTHOR DISCLOSURE AND GHOSTWRITING

J.D. declares that he owns an equity interest in NuMedii, Inc. and has served as a consultant to GSK and Janssen Pharmaceuticals. B.R. declares that he has no competing financial interests. The content of this article was expressly written by B.R. and J.D. No ghostwriters were used to write this article.

ABOUT THE AUTHORS

Dr. Ben Readhead, MBBS, is an Australian trained medical practitioner with a long-standing interest in innovation within medicine and the life sciences. This has included the design and exploration of novel technologies for the delivery of advanced wound care in difficult to heal wounds. More recent work has looked at network level features of disease genes to gain biological insight into the mechanisms of pathogenesis in certain disease types. He has recently joined the Institute for Genomics and Multiscale Biology (Icahn School of Medicine at Mount Sinai) as a Biomedical Informatician with a focus on translational bioinformatic approaches to drug repositioning, gene co-expression network analysis and causal disease modeling. **Dr. Joel Dudley, PhD**, is a veteran bioinformatics and genomics researcher with more than 10 years of professional experience studying the genomic basis of species evolution and human disease. He has published more than 40 peer-reviewed research articles pertaining to personal genomics, genomic medicine, pharmacogenomics, drug discovery, bioinformatics, and evolutionary genomics. Joel is Director of Informatics and Assistant Professor of Genetics and Genomics Sciences at Icahn School of Medicine at Mount Sinai in New York. He earned a BS in Microbiology from Arizona State University and a PhD in Biomedical Informatics from Stanford University.

REFERENCES

1. OECD Statistics. (GDP, unemployment, income, population, labour, education, trade, finance, prices, health, debt). <http://stats.oecd.org>
2. Manton KG, Corder L, and Stallard E: Chronic disability trends in elderly United States populations: 1982–1994. *Proc Natl Acad Sci USA* 1997; **94**: 2593.
3. National Health Expenditure Projections 2010–2020. <https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/downloads/proj2010.pdf>

4. Mirnezami R, Nicholson J, and Darzi A: Preparing for precision medicine. *N Engl J Med* 2012; **366**: 489.
5. Barrett T, Troup DB, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Muetter RN, Holko M, Ayambule O, Yefanov A, and Soboleva A: NCBI GEO: archive for functional genomics data sets—10 years on. *Nucleic Acids Res* 2010; **39**: D1005.
6. Hu M, Schultz K, Sheu J, and Tschopp D: The innovation gap in pharmaceutical drug discovery & new models for R&D success. Kellogg School of Management (2007). <https://ksmb02.kellogg.northwestern.edu/research/biotech/faculty/articles/NewRDModel.pdf>
7. Cockburn IM: The changing structure of the pharmaceutical industry. *Health Affairs* 2004; **23**: 10.
8. Butte AJ: Translational bioinformatics: coming of age. *J Am Med Inform Assoc* 2008; **15**: 709.
9. Sauvages F: *Nosologia methodica, sistens morborum classes, genera [et] species.* (de Tourne, 1763). http://books.google.com.au/books?hl=en&lr=&id=SD6AjVv4Nk4C&oi=fnd&pg=PA1&dq=Nosologia+methodica+sistens+morborum+classes&ots=3geJwbc6o9&sig=WiNUZS2kHLan8-O-Qz715vwnas4&redir_esc=y
10. Linnaeus C: *Genera morborum.* (1763). http://books.google.com.au/books?id=ZIMZAAAAYAAJ&printsec=frontcover&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false
11. Loscalzo J, Kohane I, and Barabasi A-L: Human disease classification in the postgenomic era: a complex systems approach to human pathobiology. *Mol Syst Biol* 2007; **3**: 124.
12. Farber HW and Loscalzo J: Pulmonary arterial hypertension. *N Engl J Med* 2004; **351**: 1655.
13. National Research Council Committee on a Framework for Developing a New Taxonomy of Disease: *Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease.* Washington, DC: National Academies Press (U.S.), 2011. Available online at: www.ncbi.nlm.nih.gov/books/NBK91503
14. Sirota M, Schaub MA, Batzoglou S, Robinson WH, and Butte AJ: Autoimmune disease classification by inverse association with SNP alleles. *PLoS Genet* 2009; **5**: e1000792.
15. Suthram S, Dudley JT, Chiang AP, Chen R, Hastie TJ, and Butte AJ: Network-based elucidation of human disease similarities reveals common functional modules enriched for pluripotent drug targets. *PLoS Comput Biol* 2010; **6**: e1000662.
16. Hamburg MA and Collins FS: The path to personalized medicine. *N Engl J Med* 2010; **363**: 301.
17. GLOBOCAN: Country Fast Stat. <http://globocan.iarc.fr/factsheets/populations/factsheet.asp?uno=900>
18. Memorial Sloan-Kettering Cancer Center: Why some lung cancers stop responding to tarceva and iressa. www.sciencedaily.com/releases/2005/02/050223164218.htm
19. Kris MG, Natale RB, Herbst RS, Lynch TJ Jr, Prager D, Belani CP, Schiller JH, Kelly K, Spiridonidis H, Sandler A, Albain KS, Cella D, Wolf MK, Averbuch SD, Ochs JJ, and Kay AC: Efficacy of gefitinib, an inhibitor of the epidermal growth factor receptor tyrosine kinase, in symptomatic patients with non-small cell lung cancer. *JAMA* 2003; **290**: 2149.
20. Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, Harris PL, Haserlat SM, Supko JG, Haluska FG, Louis DN, Christiani DC, Settleman J, and Haber DA: Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* 2004; **350**: 2129.
21. Paez JG, Jänne PA, Lee JC, Tracy S, Greulich H, Gabriel S, Herman P, Kaye FJ, Lindeman N, Boggon TJ, Naoki K, Sasaki H, Fujii Y, Eck MJ, Sellers WR, Johnson BE, and Meyerson M: EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Sci Signal* 2004; **304**: 1497.
22. Pao W, Miller V, Zakowski M, Doherty J, Politi K, Sarkaria I, Singh B, Heelan R, Rusch V, Fulton L, Mardis E, Kupfer D, Wilson R, Kris M, and Varmus H: EGF receptor gene mutations are common in lung cancers from 'never smokers' and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc Natl Acad Sci USA* 2004; **101**: 13306.
23. Kwak EL, Bang YJ, Camidge DR, Shaw AT, Solomon B, Maki RG, Ou SH, Dezube BJ, Jänne PA, Costa DB, Varella-Garcia M, Kim WH, Lynch TJ, Fidias P, Stubbs H, Engelman JA, Sequist LV, Tan W, Gandhi L, Mino-Kenudson M, Wei GC, Shreeve SM, Ratain MJ, Settleman J, Christensen JG, Haber DA, Wilner K, Salgia R, Shapiro GI, Clark JW, and Iafrate AJ: Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. *N Engl J Med* 2010; **363**: 1693.
24. Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, Fujiwara S, Watanabe H, Kurashina K, Hatanaka H, Bando M, Ohno S, Ishikawa Y, Aburatani H, Niki T, Sohara Y, Sugiyama Y, and Mano H: Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature* 2007; **448**: 561.
25. Druker BJ, Talpaz M, Resta DJ, Peng B, Buchdunger E, Ford JM, Lydon NB, Kantarjian H, Capdeville R, Ohno-Jones S, and Sawyers CL: Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. *N Engl J Med* 2001; **344**: 1031.
26. Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, Sweet-Cordero A, Ebert BL, Mak RH, Ferrando AA, Downing JR, Jacks T, Horvitz HR, and Golub TR: MicroRNA expression profiles classify human cancers. *Nature* 2005; **435**: 834.
27. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, and Lander ES: Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999; **286**: 531.
28. DiMasi JA, Hansen RW, and Grabowski HG: The price of innovation: new estimates of drug development costs. *J Health Econ* 2003; **22**: 151.
29. Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, and Schacht AL: How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat Rev Drug Discov* 2010; **9**: 203.
30. DiMasi JA: New drug development in the United States from 1963 to 1999. *Clin Pharmacol Therapeut* 2001; **69**: 286.
31. Chong CR and Sullivan DJ: New uses for old drugs. *Nature* 2007; **448**: 645.
32. Dudley JT, Deshpande T, and Butte AJ: Exploiting drug-disease relationships for computational drug repositioning. *Brief Bioinform* 2011; **12**: 303.
33. Lill MA: Multi-dimensional QSAR in drug discovery. *Drug Discov Today* 2007; **12**: 1013.
34. Noeske T, Sasse BC, Stark H, Parsons CG, Weil T, and Schneider G: Predicting compound selectivity by self-organizing maps: cross-activities of metabotropic glutamate receptor antagonists. *Chem Med Chem* 2006; **1**: 1066.
35. Keiser MJ, Roth BL, Armbruster BN, Ernsberger P, Irwin JJ, and Shoichet BK: Relating protein pharmacology by ligand chemistry. *Nat Biotechnol* 2007; **25**: 197.
36. Guha R and Jurs PC: Development of QSAR models to predict and interpret the biological activity of artemisinin analogues. *J Chem Inform Comput Sci* 2004; **44**: 1440.
37. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, Lerner J, Brunet JP, Subramanian A, Ross KN, Reich M, Hieronymus H, Wei G, Armstrong SA, Haggarty SJ, Clemons PA, Wei R, Carr SA, Lander ES, and Golub TR: The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. *Sci Signal* 2006; **313**: 1929.
38. Lamb J: The connectivity map: a new tool for biomedical research. *Nat Rev Cancer* 2007; **7**: 54.
39. Iorio F, Bosotti R, Scacheri E, Belcastro V, Mithbaokar P, Ferriero R, Murino L, Tagliaferri R, Brunetti-Pierri N, Isacchi A, and di Bernardo D: Discovery of drug mode of action and drug repositioning from transcriptional responses. *Proc Natl Acad Sci USA* 2010; **107**: 14621.
40. Sirota M, Dudley JT, Kim J, Chiang AP, Morgan AA, Sweet-Cordero A, Sage J, and Butte AJ: Discovery and preclinical validation of drug indications using compendia of public gene expression data. *Sci Transl Med* 2011; **3**: 96ra77.
41. Chen B, Wild D, and Guha R: PubChem as a source of polypharmacology. *J Chem Inform Model* 2009; **49**: 2044.
42. Kinnings SL, Liu N, Tonge PJ, Jackson RM, Xie L, and Bourne PE: A machine learning-based method to improve docking scoring functions and its application to drug repurposing. *J Chem Inform Model* 2011; **51**: 408.

43. Chiang AP and Butte AJ: Systematic evaluation of drug-disease relationships to identify leads for novel drug uses. *Clin Pharmacol Ther* 2009; **86**: 507.
44. Hu G and Agarwal P: Human disease-drug network based on genomic expression profiles. *PLoS ONE* 2009; **4**: e6536.
45. Campillos M, Kuhn M, Gavin AC, Jensen LJ, and Bork P: Drug target identification using side-effect similarity. *Science* 2008; **321**: 263.
46. Qamar R and Rector A: Semantic issues in integrating data from different models to achieve data interoperability. *Stud Health Technol Inform* 2007; **129**: 674.
47. Schadt EE and Björkregren JLM: NEW: network-enabled wisdom in biology, medicine, and health care. *Sci Transl Med* 2012; **4**: 115v1.
48. Schadt EE: Molecular networks as sensors and drivers of common human diseases. *Nature* 2009; **461**: 218.
49. Chen Y, Zhu J, Lum PY, Yang X, Pinto S, MacNeil DJ, Zhang C, Lamb J, Edwards S, Sieberts SK, Leonardson A, Castellini LW, Wang S, Champy MF, Zhang B, Emilsson V, Doss S, Ghazalpour A, Horvath S, Drake TA, Lusis AJ, and Schadt EE: Variations in DNA elucidate molecular networks that cause disease. *Nature* 2008; **452**: 429.
50. Schadt EE, Lamb J, Yang X, Zhu J, Edwards S, Guhathakurta D, Sieberts SK, Monks S, Reitman M, Zhang C, Lum PY, Leonardson A, Thieringer R, Metzger JM, Yang L, Castle J, Zhu H, Kash SF, Drake TA, Sachs A, and Lusis AJ: An integrative genomics approach to infer causal associations between gene expression and disease. *Nat Genet* 2005; **37**: 710.
51. Dudley JT, Schadt E, Sirota M, Butte AJ, and Ashley E: Drug discovery in a multidimensional world: systems, patterns, and networks. *J Cardiovasc Transl Res* 2010; **3**: 438.
52. Zhang B and Horvath S: A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* 2005; **4**: 1128.
53. Tran L, Zhang B, Zhang Z, Zhang C, Xie T, Lamb J, Dai H, Schadt E, and Zhu J: Inferring causal genomic alterations in breast cancer using gene expression data. *BMC Syst Biol* 2011; **5**: 121.
54. Dobrin R, Zhu J, Molony C, Argman C, Parrish ML, Carlson S, Allan MF, Pomp D, and Schadt EE: Multi-tissue coexpression networks reveal unexpected subnetworks associated with disease. *Genome Biol* 2009; **10**: R55.
55. Zhu J, Chen Y, Leonardson AS, Wang K, Lamb JR, Emilsson V, and Schadt EE: Characterizing dynamic changes in the human blood transcriptional network. *PLoS Comput Biol* 2010; **6**: e1000671.
56. Roy S, Patel D, Khanna S, Gordillo GM, Biswas S, Friedman A, and Sen CK: Transcriptome-wide analysis of blood vessels laser captured from human skin and chronic wound-edge tissue. *Proc Natl Acad Sci USA* 2007; **104**: 14472.
57. Soulet F, Kilarski W, Antczak P, Herbert J, Bicknell R, Falciani F, and Bikfalvi A: Gene signatures in wound tissue as evidenced by molecular profiling in the chick embryo model. *BMC Genomics* 2010; **11**: 495.
58. Eming SA, Koch M, Krieger A, Brachvogel B, Kreft S, Bruckner-Tuderman L, Krieg T, Shannon JD, and Fox JW: Differential proteomic analysis distinguishes tissue repair biomarker signatures in wound exudates obtained from normal healing and chronic wounds. *J Proteome Res* 2010; **9**: 4758.
59. Chang HY, Sneddon JB, Alizadeh AA, Sood R, West RB, Montgomery K, Chi J-T, van de Rijn M, Botstein D, and Brown PO: Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. *PLoS Biol* 2004; **2**: e7.
60. Brem H, Stojadinovic O, Diegelmann RF, Entero H, Lee B, Pastar I, Golinko M, Rosenberg H, and Tomic-Canic M: Molecular markers in patients with chronic wounds to guide surgical debridement. *Mol Med* 2007; **13**: 30.
61. Smiley AK, Klingenberg JM, Aronow BJ, Boyce ST, Kitzmiller WJ, and Supp DM: Microarray analysis of gene expression in cultured skin substitutes compared with native human skin. *J Invest Dermatol* 2005; **125**: 1286.
62. Edgar R, Domrachev M, and Lash AE: Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 2002; **30**: 207.
63. Cooke JP and Wilson AM: Biomarkers of peripheral arterial disease. *J Am Coll Cardiol* 2010; **55**: 2017.
64. Masud R, Shameer K, Dhar A, Ding K, and Kullo IJ: Gene expression profiling of peripheral blood mononuclear cells in the setting of peripheral arterial disease. *J Clin Bioinform* 2012; **2**: 6.
65. BIOMEDIN217 Translational Bioinformatics: Stanford University Online. at <http://scpd.stanford.edu/search/publicCourseSearchDetails.do?method=load&courseId=1165711>
66. Mount Sinai School of Medicine Offers First-Ever Course with Whole Genome Sequencing: Mount Sinai School of Medicine. www.mssm.edu/departments-and-institutes/genetics-and-genomic-sciences/news/mount-sinai-school-of-medicine-offers-first-ever-course-with-whole-genome-sequencing
67. PhRMA 2012 Profile. www.phrma.org/research/publications/profiles-reports
68. Center for Drug Evaluation and Research and FDA Basics—Is it true FDA is approving fewer new drugs lately? www.fda.gov/AboutFDA/Transparency/Basics/ucm247348.htm
69. CPI Inflation Calculator. <http://data.bls.gov/cgi-bin/cpicalc.pl>
70. Benjamini Y and Hochberg Y: Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B* 1995; 289.
71. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G; The Gene Ontology Consortium: Gene ontology: tool for the unification of biology. *Nat Genet* 2000; **25**: 25.