Drug Discovery in a Multidimensional World: Systems, Patterns, and Networks

Joel T. Dudley · Eric Schadt · Marina Sirota · Atul J. Butte · Euan Ashley

Abstract Despite great strides in revealing and understanding the physiological and molecular bases of cardiovascular disease, efforts to translate this understanding into needed therapeutic interventions continue to lag far behind the initial discoveries. Although pharmaceutical companies continue to increase investments into research and development, the number of drugs gaining federal approval is in decline. Many factors underlie these trends, and a vast number of technological and scientific innovations are being sought through efforts to reinvigorate drug discovery pipelines. Recent advances in molecular profiling technologies and development of sophisticated computational approaches for analyzing these data are providing new, systems-oriented approaches towards drug discovery. Unlike the traditional approach to drug discovery which is typified by a one-drug-one-target mindset, systems-oriented approaches to drug discovery leverage the parallelism and high-dimensionality of the molecular data to construct more comprehensive molecular models that aim to model broader bimolecular systems. These models offer a means to explore complex molecular states (e.g., disease) where thousands to millions of molecular entities comprising multiple molecular data types (e.g., proteomics and gene expression) can be evaluated simultaneously as components of a cohesive biomolecular system. In this paper, we discuss emerging approaches towards systems-oriented drug discovery and contrast these efforts with the traditional, unidimensional approach to drug discovery. We also highlight several applications of these system-oriented approaches across various aspects of drug discovery, including target discovery, drug repositioning and drug toxicity. When available, specific applications to cardiovascular drug discovery are highlighted and discussed.

Keywords Genomics · Drug Discovery · Systems Biology

J. T. Dudley · M. Sirota
Program in Biomedical Informatics, Stanford University School of Medicine,
Stanford, CA, USA

J. T. Dudley · M. Sirota · A. J. Butte
Department of Pediatrics, Stanford University School of Medicine,
Stanford, CA, USA

J. T. Dudley · M. Sirota · A. J. Butte
Lucile Packard Children's Hospital,
Palo Alto, CA, USA

E. Schadt
Pacific Biosciences,
1505 Adams Drive,
Menlo Park, CA, USA

E. Schadt
Sage Bionetworks,
Palo Alto, CA, USA

E. Ashley (✉)
Division of Cardiovascular Medicine, Stanford University School of Medicine,
Falk CVRC 300 Pasteur Drive; MC5406,
Stanford 94305-5406 CA, USA
e-mail: euan@stanford.edu

Introduction

Traditional approaches to drug discovery are generally regarded as protracted and costly, with data showing that it takes approximately 15 years and over $1 billion to develop and bring a novel drug to market [1]. Yet as drug development costs continue to rise, the number of drugs gaining FDA approval is in decline [2]. A substantial portion of drug development costs are incurred during early development and

toxicity testing, with more than 90% of drugs failing to move beyond these early stages [3, 4]. These facts highlight a declining trend in research and development (R&D) productivity across the pharmaceutical industry, which is cited as a principal factor underlying diminishing drug discovery pipelines [5, 6]. Increased costs and regulatory burdens associated with conducting clinical trials are driving recent shifts towards the outsourcing of early-stage R&D in drug development, and there is a trend across the industry to shift funding strategies from early- to late-stage projects [7]. A consequence of this strategy is increased collaborations and early-stage research engagements both across the industry and also between industry and academia through the sharing of data in the public domain [8]. Consequently, an increasing proportion of pre-competitive research data, once considered to be a strategic asset proprietary to the company, is finding its way into the public domain. Much of this data is coming into the public domain by way of pharmaceutical research consortia or partnerships with non-profit or academic research organizations. Prominent examples include the ChEMBL database (http://www.ebi.ac.uk/chembldb/), Sage Commons (http://www.sagebase.org), the Asian Cancer Research Group, and the CTSA Pharmaceutical Assets Portal (http://www.ctsapharmaportal.org). The public availability of this once-proprietary data offers novel opportunities to explore and develop new approaches to drug discovery.

A portion of the decline in pharmaceutical R&D might be explained by the prevailing unidimensional research strategy that has guided drug discovery for the past 70 years. The unidimensional strategy operates under the one-target-one-disease philosophy, where research activities are typically constrained to focus on a specific therapeutic indication, and the goal is to find the single druggable target (e.g., enzyme or receptor) that is able to facilitate a positive change in disease biology though pharmaceutical modulation [9]. Once a drug target is identified, the bulk of R&D goes into the discovery and development of compounds that will safely and effectively modulate the target in the desired manner. Presently, this is achieved through use of multiplex, high-throughput screening approaches that enable automated screening of large chemical libraries against targets for biological activity. Another approach is to use high-throughput screening technology to screen large chemical libraries against multiple biological targets. However, the combinatorial space of drug compounds and targets is so vast, that some prior selection of targets is necessary [10]. Nonetheless, the unidimensional strategy ultimately prevails because the goal of these R&D efforts is to hone in on the single drug–target combination that can be further evaluated towards the development of a new drug product. While the reductionism of the unidimensional approach has demonstrated historical success in bringing safe and effective drugs to market, there is mounting evidence that a large number of approved drugs interact promiscuously with biological elements outside of their canonical targets [11], and that multi-target strategies can sometimes be more effective in treating complex disease conditions [12]. Simultaneous consideration for multiple drugs and biological targets does not lend itself easily to reductionist approaches, and therefore non-reductionist approaches must be developed to serve as the basis for departure from unidimensionality.

The growing wealth of molecular profiling data and development of sophisticated informatics methods have enabled completely new approaches to drug discovery that attempt to understand drug activity from a data-driven, systems perspective. Methods are available to integrate genome-wide proteomics, gene expression, and RNA interference data to construct in silico systems models of molecular pathology. Such models can be inspected computationally to generate predictions about the biomolecular network effects of pharmacologically perturbing one or more biological targets. Molecular profiling data can also be used towards more non-reductionist approaches, where molecular "signatures" can be used as patterns to match diseases and drugs considering only their gross molecular characteristics (e.g., genome-wide transcript abundance). Initial developments in these new approaches to drug discovery are already yielding novel therapeutic relationships unlikely to be discovered by traditional means. Continued research and development in these directions is likely to herald in a new generation in drug discovery, with promise to further systematize a discipline traditionally reliant on some serendipity [13], and to reinvigorate pharmaceutical pipelines with much-needed pharmaceutical innovations.

Here, we put forward a view that drug discovery is now undergoing a fundamental shift towards systems-based research and development. We suggest that this shift is engendered by a confluence of advances in molecular profiling technologies and developments in informatics that enable integrative analysis of diverse molecular data towards systems-based analysis, and we highlight some of the specific approaches driving this change. This perspective is organized into several major areas of drug discovery: target identification, drug repositioning, and side effect discovery. We aim to illustrate how systems-based approaches are impacting each of these areas individually, and also how systems-based approaches unify these areas under the emerging paradigm of systems-based drug discovery (Fig. 1, Table 1).

Reductionist Biology Tends Towards Linearity and Unidimensionality

In the early part of the twentieth century, physiologists were among our most prominent biological scientists. Animal systems were developed where an input was varied and a variety of physiological parameters were measured (e.g.,

Fig. 1 Schematic overview of systems-based approaches for drug discovery. Molecular profiles measured from disease states or drug effect can be leveraged in many ways to enable systems-oriented approaches towards therapeutic discovery. Multiple types of molecular data and characteristics can be integrated using network paradigms to build more comprehensive models of bimolecular systems relevant to disease conditions and drug effect
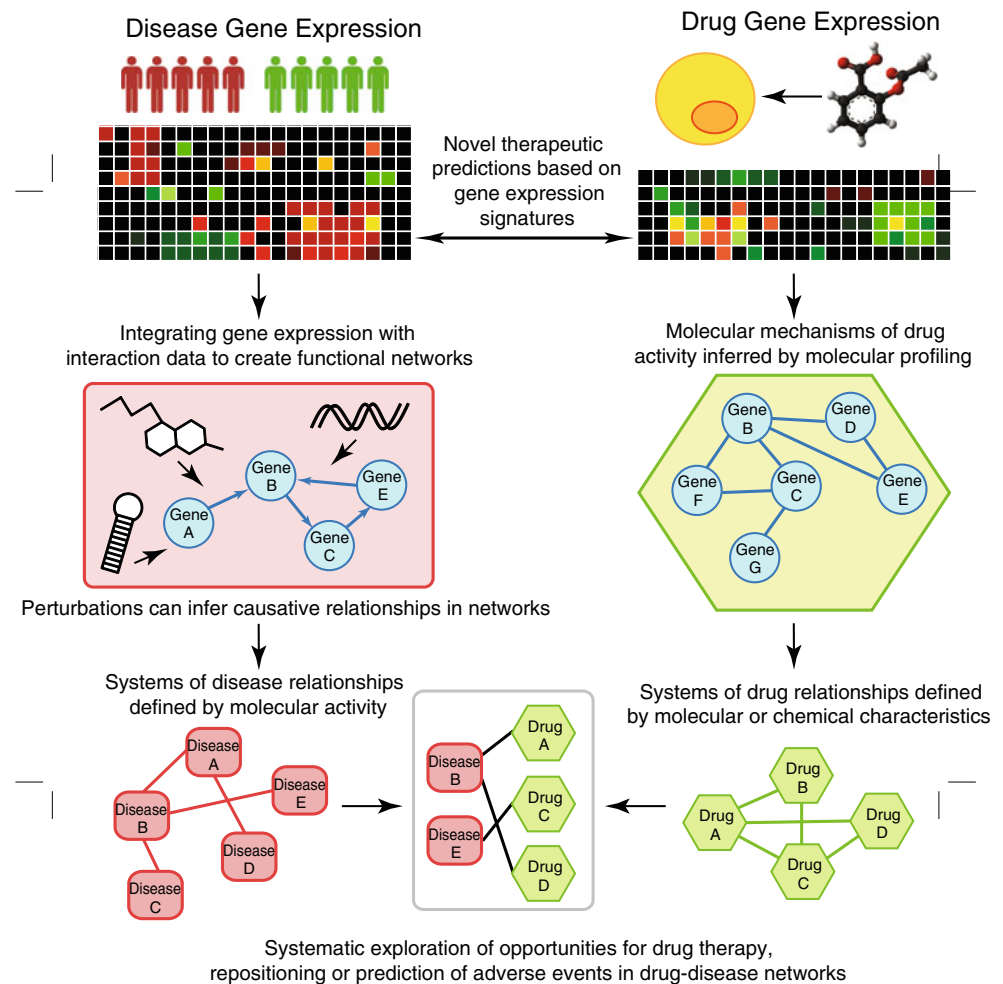
Starling's dog model of cardiac mechanics); the emphasis was on describing the physiological system response to a perturbation. With the emergence of the molecular biology era, the testing of scientific hypotheses has tended to focus on the manipulation of one molecular parameter and the measurement of another, usually a continuous molecular variable. For example, a gene is knocked down and a signaling reporter is detected. Multiple versions of this experiment are often repeated in different wells or with cells under different conditions and a plot of the data might reveal a linear dependency between the gene and the reporter. For in vivo experiments, a gene can be deleted by homologous recombination in embryonic stem cells, which can then lead to colonies of animals that lack the specified gene product from the moment of conception. Commonly, differences in phenotype from littermate control animals are ascribed to a direct effect of the genetic ablation with little attempt to analyze the effect of the perturbation on the underlying biological system. Despite the fact that most common diseases implicate multiple organs, tissues or physiological systems in their underlying pathology, unidimensional molecular approaches continue to dominate the research

landscape. The tendency to invoke linearity is understandable given the benefits of simplicity and the wide range of linear statistical methods that can be readily applied for analysis. Nonetheless, sophisticated informatics methods now exist that enable study of the non-linear dynamics among molecular entities perturbed by drug or disease for early discovery stages drug discovery.

As physiologists have known for some time, the efficacy of many drugs is rooted in their ability to perturb physiological systems away from pathological states to attenuate manifest disease. Examples include reduction of blood sugar in diabetes, modulation of RAAS by ACE inhibitors, or the clearance of fluid build-up by diuretics in heart failure. In this way, the drugs are not necessarily addressing the root molecular cause of pathology, but they still offer tremendous clinical value to physicians and patients. It is becoming increasingly difficult to discover such drugs by means of a reductionist, unidimensional approach. It is, of course, possible to start with a biologically well-characterized canonical target, such as an adrenergic receptor, and simply look for compounds that bind to this receptor to produce the desired physiological

Table 1 Systems-based approaches for drug discovery

Name	Summary	References
Connectivity Maps	Cell lines are exposed to large drug libraries and genome-wide mRNA gene expression is measured. Genes are ordered into a ranked list according to expression changes to create a "signature" of drug effect. This signature can be queried against transcription changes in other biological conditions, such as disease, to see if genes perturbed by the drug are statistically enriched in the biological condition. Enrichment suggests a therapeutic connection	[65, 78–83]
Chemical systems biology	Data on chemical properties of drugs and biological molecules are integrated into systems models to model the effect of drug compounds in a biological system. This approach typically incorporates quantitative structure activity relationship (QSAR) data with physiochemical and structural properties of biological targets. Computational predictions of drug–target interactions are sometimes inferred using sophisticated docking simulation algorithms	[74, 76]
Chemical similarity networks	Drug compounds are organized into a network of relationships based primarily on chemical and structural properties, and also based on their target binding profiles. Drug compounds are connected to targets and diseases based on their chemical similarity to other compounds in the network. This approach differs from chemical systems biology because there is no attempt to model the biology of drug effect	[11, 15, 72]
Causal network inference	Molecular variability caused by directed perturbation (e.g. RNAi) or natural genetic variability is integrated with profiling data on typically molecular quantitative traits such as gene expression levels. Informatics approaches are applied to link molecular perturbation to variability in the quantitative trait, and probabilistic methods are used to model these links as causal networks of relationships. Once established, these causal networks can be explored to predict the effects of pharmacological perturbation, or to predict molecular drivers or observed changes in a quantitative trait (e.g. changes in gene expression) caused by exposure to a pharmacological perturbagen	[23, 26, 37, 38, 42, 44, 45, 47]

response in vivo. However, even if a compound is found by molecular or cellular assay to bind to the target of interest, there is little recourse to ensure that the compound will demonstrate efficacy when administered to a complete, live organism. One major source of this uncertainty is the problem of polypharmacology, which is the tendency of drugs to bind to more than one biological target. The efficacy of a particular drug might not come entirely through interaction with its canonical target, but rather though complex physiological modulation that results from promiscuous binding with multiple targets. The molecular data is supporting an emerging view that polypharmacology is a widespread phenomenon among currently approved drugs [14]. This might be expected to a degree, given that many drug targets share membership with closely related genes in multi-gene families that have shared structural features and evolutionary histories. However recent studies reveal that a number of approved drugs exhibit significant affinities for targets found in classes completely distinct from their canonical targets [11, 15]. Examples include the demonstrated affinity of Vadilex (NMDA ion channel antagonist) for 5-HTT (transporter) and μ-opioid receptors, and Dorelese (α_1 blocker) binding affinity for dopamine D4 receptor [15]. It is still possible to take a reductionist approach to dealing with the problem of polypharmacology through consideration for pharmacophore or other physiochemical similarities among receptors. However, the three-dimensional structures for many pharmacologically important receptors remain undetermined, severely

limiting this approach. An alternative approach is to use the available molecular profiling data to attempt to model the dynamics of the entire system, given data indicating how the system will respond to perturbation.

Systems Molecular Biology Attempts to Account for Network Dynamics

In the mid-1990s, a number of researchers began to multiplex molecular biology experiments using microarrays. Microarrays were revolutionary because they offered an efficient and high-throughput means to assay a complete biological system (the transcriptome) under perturbation, instead of just a single variable. Technological advancements have expanded upon the multiplex paradigm introduced by microarrays towards the development of SNP genotyping, protein, cytokine and exon microarrays; and more recently high-throughput sequencing technologies capable of generating high-coverage sequence profiles of complete genomes. A salient feature of these technologies is that they generate tremendous amounts of data. A single gene expression microarray might generate more than 50,000 data points from a biological sample whereas full-genome profiling might generate many billions of data points. These assays are typically used to measure multiple samples across a range of biological conditions, producing data sets containing millions to trillions of data points. Such data streams have served to reinvigorate biostatistics and a new generation of informatics specialists has grown up around the need to

understand such data. Even here, the initial impulse was to repeat the prior tendencies towards linearity. One of the most successful early tools for microarray data was the significance analysis of microarrays (SAM) [16]. The SAM approach was notable because until that point, the idea of fold change, carried over from molecular biology, was prevalent and the idea of accounting for variance and false discovery, rather than simply magnitude of effect, was reborn.

Simultaneously, statistical approaches to understanding networks were beginning to emerge from social science and were applied to problems in biological science. Classic experiments attempting to explain the "small world" characteristics (e.g., cliques and sub-graphs) of network systems were carried out. The challenge was to explain why the maximum path length (the number of hops required to get from one node to another in a network) was often very small. Early solutions [17] were passed over in favor of the scale free hypothesis [18]. In this model, the probability P that a node will connect to k other nodes declines as a power function $P(k) \sim k^{-x}$. This was shown to hold for large networks such as the World Wide Web, and it was not long before key experiments replicated these network properties in biological systems. Such networks are self-organizing and promote preferential connectivity that gives rise to "hub" nodes that tend to be more critical. In yeast, it was shown that the likelihood that ablation of a protein would lead to lethality directly correlates with the number of connections in a protein network [19]. In other work, gene hubs derived from co-expression networks were shown to be more predictive of cancer survival [20, 21], emphasizing the centrality and biological significance of these nodes. We and others have argued for the use of network analysis techniques to identify such critical biological signaling players to enhance drug discovery [22–24]. Networks can be generated in different ways. Commonly, a co-expression matrix is generated from pair-wise Pearson correlations among gene expression values derived from microarray profiling [25]. A network topology is then generated according to different approaches to quantifying shared gene expression, such as a correlation threshold. Other approaches to network generation incorporate different data sources such as protein binding data [24], gene ontology [26], drug activity [25], or transcription factor binding data [27]. It is even possible to incorporate phenotypic traits, such as those derived from patient clinical records or high-throughput phenotypic animal screens, along with molecular data within a single, coherent network [28].

Molecular Profiling of Human Cardiovascular Tissue can Reveal Targets for Drug Discovery

We recently used transcriptional profiling of samples of human heart to identify the most significantly differentially regulated genes in a reverse heart failure model. Samples were collected at the time of implantation of a left ventricular assist device and then again from the same patient at the time of transplant. Gene expression was compared in these paired samples and through a series of analyses; we identified the recently de-orphanized apelin receptor as the most significantly up-regulated gene after offloading [29]. An angiotensin-like receptor, it became a lead candidate for heart failure therapeutics in our laboratory and others. Through a series of experiments, a fundamental role for this relatively unknown signaling system was demonstrated in cardiovascular biology. Disease models followed with potential efficacy shown in multiple small animal models. In a rat model of isoproterenol-induced heart failure, apelin improved function and outcome [30]. In a mouse model of aortic aneurysm formation, apelin delivered by minipump almost completely abrogated the aneurysm and provided a mortality benefit [31, 32]. Apelin also seemed to be able to enhance insulin sensitivity in models of insulin resistance [33]. Most recently a series of human experiments have established a role for this peptide in enhancing cardiac hemodynamics [34] thus bringing investigators full circle back around to human therapeutics. At this time, several privately funded and well established biotechnology and pharmaceutical companies are exploring potential exploitation of the therapeutic efficacy of this system.

We have also used transcriptional profiling along with text mining of the published literature for identifying candidate systems for atherosclerosis and in-stent restenosis [22]. In the latter case, we carried out transcriptional profiling on human atherectomy samples and derived a gene expression signature of in-stent restenosis by subtracting the profile of de novo atherosclerosis. Separately, we used a text-mining algorithm to generate self-assembling networks from the published literature. Gene names, aliases, and symbols were used as nodes and a lookup table of modifier verbs used to generate edge relationships. Subnetworks generated in this way were then given a significance score for the extent to which they reflected the signature gene expression of in-stent restenosis. Top candidates for drug development were then identified and are currently being tested in mouse models of stenting [35, 36].

Since in vivo experiments are time consuming, we performed an in silico test of this approach by measuring the mean score of subnetworks that feature the current drugs used in drug eluting stents [22]. We carried out such an analysis for paclitaxel and sirolimus, two of the most common agents used in coronary artery stents. When comparing the mean subnetwork significance scores for subnetworks generated for these two agents, we found a clear signal in favor of sirolimus recapitulating the direct head-to-head comparison of these agents in clinical trials [22].

Connecting Genetics to Molecular Phenotypes Through Causal Inference and Network Perturbation

Another powerful perspective afforded by systems-based approaches comes through the incorporation of system dynamics to infer the downstream effects of pharmaceutical perturbation. Networks enable representation of the topology of relationships in a system, and patterns (e.g., genome-wide gene expression) provide snapshots of the molecular state of a system sampled from a particular physiological context (e.g., diseased tissue). While these representations enable powerful characterizations and quantifications of a system, the incorporation of a dynamic component that leverages these data to infer cause–effect relationships between perturbations of the network and changes in molecular states enables new modes of predictive systems analysis not afforded by static representations alone. A primary challenge in this area is first learning the rules and parameters that govern dynamic relationships between elements in a biomolecular system. A formal set of baseline rules and parameters must be learned with an appreciable degree of accuracy before it is possible to infer how a perturbation might alter a system. Although it is not yet possible to perform real-time profiling of all aspects of molecular systems, statistical approaches make it possible to infer causal dynamics through integration of the available molecular profiling data. Given molecular profiling data measured from a system under various states of perturbation, these methods attempt to infer the causal relationships between members of a system that identify the most probable cause–effect relationships explaining the observed response.

Both natural and directed genetic variation can serve as rich sources of variation for the inference of causal relationships between genetic causes and phenotypic effects in biological systems. In this case, perturbations in the form of variation in genotype are linked to variation in molecular phenotypes (e.g., gene transcript abundance) to identify quantitative trait loci (QTLs) affecting phenotypes. This variability can be profiled in natural breeding populations, or more precisely in carefully controlled crosses of inbred animal strains. Even under controlled breeding conditions, the elucidation of QTLs presents substantial computational and statistical challenges; such as limiting false discovery, avoiding magnification of data artifacts, and disentangling proxy-associated QTLs from true causal QTLs [37]. Computational and statistical approaches have been developed to address these concerns [38], however the construct of QTLs themselves do not offer any notion of causal ordering, therefore additional steps are needed to infer the conditional relationships between QTLs to establish a framework for propagating perturbation dynamics across a system model.

Bayesian network approaches are commonly employed to learn the conditional probability distributions of variables in the network, which are then used to solve conditional probability equations to infer probable causal networks [39]. However, Bayesian networks incorporating many thousands of QTL elements carry several limitations. In addition to being acyclic, and therefore not allowing for modeling of feedback regulation, Bayesian networks often infer many equally possible network graphs as solutions [40]. This quickly leads to non-trivial computational and methodological challenges in exploring the complete space of all probable network graphs. Some of these limitations can be overcome by application of dynamic Bayesian networks, which incorporates temporal relationships between variables [41]. Another popular approach to reducing the complexity of the network is to apply co-expression correlation, or more sophisticated statistical methods to first group genes and loci into functionally coherent groups called "modules" [42, 43]. Causal inference is then performed to infer networks of functional gene modules with reduced network complexity and added statistical power.

Once established, causal molecular networks can be explored bi-directionally to infer causal events explaining an observed phenotype, or to propagate the effects of a molecular perturbation to infer a probably effect on phenotype. Schadt et al. were among the first to demonstrate the feasibility of generating causal networks of gene expression phenotypes across multiple species [37]. They later demonstrated that the inferred causal networks could be integrated across-species and across-tissues to build a system of causal networks that identify molecular drivers of complex disease operating in a broader physiological context [44, 45]. Ferrara et al. integrated QTLs with metabolomics data to infer causal networks underlying metabolic processes in the liver and used these findings to link causal molecular networks to metabolic disease [46]. Wang et al. used causal network inference to enable a reverse-genetics approach that used genomic analysis of crosses between strains of atherosclerotic mice to infer upstream, cross-tissue molecular drivers of atherosclerosis [47]. Causal network inference methods can be extended to incorporate many different data types and multiple sources of perturbation [48], and therefore the incorporation of pharmaceutical perturbation is a natural extension of these methods. Causal networks can serve as the basis for multidimensional, network-based modalities of drug discovery that provide frameworks for linking pharmaceutical perturbations to system-wide phenotypic effects, or to enable reverse-genetics approaches to drug discovery that begin with a clinical phenotype of interest and infer molecular drivers of the phenotype that might serve as promising targets for novel drug therapies. Instead of using

444

J. of Cardiovasc. Trans. Res. (2010) 3:438–447

high-throughput drug screens to single out drug–target pairs for lead discovery, the standard approach can be retooled to identify relationships between drugs and causal subnetworks, whose members are identified as targets for therapy [23].

Drug Repositioning

Drug repositioning (also known as drug repurposing) is the identification of novel disease indications for approved drugs, and it offers several advantages over traditional drug development [49]. The repositioning of drugs already approved for human use mitigates the costs and risks associated with early stages of drug development, because the drugs have already proven to be generally safe for use in humans. Drug repositioning also offers potentially shorter routes to approval for novel therapeutic indications because, with a drug's toxicity characteristics already known, the repositioned drug indication has the potential to skip ahead to later phases of the drug approval process. Successful examples of drug repositioning include the indication of sildenafil for erectile dysfunction and pulmonary hypertension, thalidomide for severe erythema nodosum leprosum, and retinoic acid for acute promyelocytic leukemia [50]. Several approaches to drug repositioning are being taken, ranging from purely experimental to purely computational and hybrids thereof. The prevailing approach to drug repositioning inherits from the established tools and techniques developed for screening libraries of lead compounds against biological targets of interest in early stage drug discovery. In the case of drug repositioning, libraries of approved drugs are typically screened across a broad set of putative biological targets using high-throughput screening approaches [49, 50]. Computational approaches to the discovery of novel biological targets for approved drugs have been developed to complement HTS approaches, allowing for greatly expanded searches of the possible drug–target relationships space for novel repositioning opportunities [15, 23, 31, 51–53].

Gene expression microarrays are regularly and broadly applied in clinical studies of human diseases, providing genome-wide characterization of a disease state. Comparative gene expression analysis of primary affected, peripheral and secondary organs, tissues, and biofluids are used to study the molecular pathophysiology of a disease condition, or to identify expression patterns that serve as prognostic or diagnostic indicators. By examining the sets of genes that are up-regulated and down regulated in a disease state as compared to a normal state, it is possible to create a gene expression profile, or signature, of a disease [54–58]. When derived from clinical samples, these signatures represent accurate and consistent gene expression states imparted by

immunological, metabolic, and other complex factors comprising the broad physiological manifestation of a disease [44, 59]. Therefore, gene expression signatures derived from microarrays can be used to generate high-quality signatures of disease that are representative of the molecular disease pathology in vivo [59].

Microarrays are also used to discover gene expression patterns that signify pharmacologic perturbation, allowing for the development of high-quality signatures of drug effect [60–64]. Lamb and colleagues published a collection of genome-wide transcriptional expression data from cultured human cells treated with a broad range of bioactive small molecules to create a drug signature library called the Connectivity Map [65]. The initial study consisted of several hundred experiments with differing dosages of 164 chemical compounds and corresponding vehicle controls. The set of drugs used to build the Connectivity Map included a number of FDA approved chemical drugs and biologics. The drug expression signatures from the Connectivity Map can be used to connect the molecular basis of drug effects to disease, and several have used data from the Connectivity Map to study particular diseases of interest. Garman et al. queried the Connectivity Map using a gene expression signature derived from colon cancer samples signature to identify novel therapeutic opportunities [66] and Setlur et al. queried a gene expression signature from a prostate cancer subtype against the Connectivity Map to show that the molecular phenotype of the disease was associated with estrogen receptor signaling [67]. The evident weakness in the Connectivity Map approach is the reliance on cultured cell lines as a proxy for the true biological context of human physiology. For diseases in which broad physiological context is important, such as immunological and metabolic disorders, signatures derived under the Connectivity Map approach may be weak surrogates for the in vivo drug effect. Even so, the Connectivity Map approach has uncovered significant new opportunities for therapeutic intervention in cancers and some aspects of metabolic disorders [68, 69]. Given that the precise molecular mechanisms underlying a substantial proportion of drug and disease activity remain uncharacterized, signature-based approaches offer an effective and expedient non-reductionist approach to discovery of drug repositioning candidates that is amenable to our current understanding of human disease pathology and pharmacology.

Network-based approaches to drug repositioning seek to relate approved drugs to new indications through network relationships. In many ways, network-based approaches to drug repositioning will resemble the network-based approaches used to discover drug–target relationships for new chemical entities. However, a wealth of information exists for approved compounds that are not available for

NCEs, such as canonical sets of known drug targets captured in repositories such as DrugBank [70], and information regarding both the on- and off-label prescribing information that provides connections between drugs and clinical phenotypes, such as diseases and side effects [71]. These additional data sources afford novel methodological opportunities to reposition existing drugs for novel indications using network properties. Keiser et al. developed a network-based cheminformatics approach that utilized knowledge of the chemical structures of approved drugs and their canonical biological targets to discover novel connections between ligands and targets based on chemical similarities of ligand binding sets [15]. Campillos et al. constructed a drug–drug network using a phenotypic similarity metric that was computed from drug side effect information extracted from FDA drug labels [72]. In similar fashion, we created a drug–drug network in which drugs were associated in the network based on shared on- and off-label disease indications, and discovered novel drug indications discovered using a "guilt by association" approach [73]. In other work, we constructed network modules from gene expression profiles of disease and integrated these with protein–protein interaction data to build functional protein network modules associated with various disease states [24]. This work uncovered a number of common network modules shared across many diseases, and revealed that these modules were enriched for drug targets; and therefore present rich opportunities for the reposition of drugs within the same functional modules. The clear benefit of the network paradigm in all of these cases is the ability to build relationships between multiple types of relevant data, and then to discovery novel relationships between drugs and targets through statistical analysis of these complex relationships.

Predicting Drug Side Effects

Systems approaches to the study of pharmacological perturbation enable the discovery of efficacious connections between drugs and disease, yet these methods are just as easily applied towards the discovery of relationships of toxicity between drugs and clinical phenotypes of adverse drug effect [74]. Berger et al. proposed a systems pharmacology approach to uncover a pharmacological basis for drug-induced long-QT syndrome (LQTS) [75]. In their analysis, they used a seed of genes associated with LQTS and expanded this set using protein interaction data to build a LQTS disease "neighborhood" network. They compared the LQTS network with the networks of other diseases known to exhibit cardiac complications to uncover a shared molecular basis for cardiac complications affecting QT interval. Finally, they incorporated drug information and

found that FDA approved drugs having LQTS as side effects targeted genes enriched in the LQTS network. This enabled the development of a statistical classifier to predict drugs likely to cause LQTS, which was applied to implicate several additional drugs likely to cause LQTS. Xie et al. developed a chemical systems biology approach to uncover a molecular basis for the serious hypertension side effect of cholesteryl ester transfer protein (CETP) inhibitors, which caused them to be withdrawn from development in phase III clinical trials [76]. Through integration of a ligand-target binding network and pathway analysis using a simple gene regulation model, they were able to predict off-target binding targets in the RAAS explaining the putative molecular basis by which the CETP inhibitor torcetrapib caused severe hypertension. Scheiber et al. constructed a network linking adverse drug reactions (ADRs) based on the chemical similarity of their associated drug compounds to create a global network map of linking ADRs based on chemical properties [77]. Whereas characterization of drug toxicity and drug efficacy are pursued disjointly under traditional approaches to drug discovery, systems-based approaches enable simultaneous consideration for both drug efficacy and toxicity along many possible dimensions.

Conclusion

We have discussed the limitations of reductionism and unidimensionality from the perspective of drug discovery, and highlighted a number of systems-oriented approaches that provide means to overcome these limitations by enabling multidimensional perspectives on complex clinical phenomena. Of course, recognition is due to the long tradition of unidimensional approaches that have brought forth innumerable biological and clinical insights, as well as a majority of effective therapeutics used regularly in the course of modern clinical care. It is mostly due to the recent explosion in the volume and diversity of available high-throughput molecular measurements that unidimensional methods stand out to be so ostensibly limited. Going forward, the most successful efforts in multidimensional, systems-oriented approaches to drug discovery will likely build from the long and successful tradition of unidimensionality, to create opportunities that synergize the strengths of both perspectives; enabling completely new directions in drug discovery that address the new opportunities and challenges facing cardiovascular drug discovery the post-genomic era.

446

J. of Cardiovasc. Trans. Res. (2010) 3:438–447

References

1. DiMasi, J. A., Hansen, R. W., & Grabowski, H. G. (2003). *Journal of Health Economics, 22*, 151.
2. Hughes, B. (2009). *Nature Reviews. Drug Discovery, 8*, 93.
3. Chong, C. R., & Sullivan, D. J., Jr. (2007). *Nature, 448*, 645.
4. Dickson, M., & Gagnon, J. P. (2004). *Nature Reviews. Drug Discovery, 3*, 417.
5. Garnier, J. P. (2008). *Harvard Business Review, 86*, 68.
6. Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, Schacht AL at Rev Drug Discov 9: 203
7. Barnes, M. R., Harland, L., Foord, S. M., Hall, M. D., Dix, I., Thomas, S., et al. (2009). *Nature Reviews. Drug Discovery, 8*, 701.
8. Woodcock J Clin Pharmacol Ther 87: 521
9. Drews, J. (2000). *Science, 287*, 1960.
10. Hüser, J. (2006). *High-throughput screening in drug discovery.* Weinheim: Wiley.
11. Keiser, M. J., Roth, B. L., Armbruster, B. N., Ernsberger, P., Irwin, J. J., & Shoichet, B. K. (2007). *Nature Biotechnology, 25*, 197.
12. Morphy, R., Kay, C., & Rankovic, Z. (2004). *Drug Discovery Today, 9*, 641.
13. Schlueter, P. J., & Peterson, R. T. (2009). *Circulation, 120*, 255.
14. Overington, J. P., Al-Lazikani, B., & Hopkins, A. L. (2006). *Nature Reviews. Drug Discovery, 5*, 993.
15. Keiser, M. J., Setola, V., Irwin, J. J., Laggner, C., Abbas, A. I., Hufeisen, S. J., et al. (2009). *Nature, 462*, 175.
16. Tusher, V. G., Tibshirani, R., & Chu, G. (2001). *Proceedings of the National Academy of Sciences of the United States of America, 98*, 5116.
17. Watts, D. J., & Strogatz, S. H. (1998). *Nature, 393*, 440.
18. Barabasi, A. L., & Albert, R. (1999). *Science, 286*, 509.
19. Jeong, H., Mason, S. P., Barabasi, A. L., & Oltvai, Z. N. (2001). *Nature, 411*, 41.
20. Horvath, S., Zhang, B., Carlson, M., Lu, K. V., Zhu, S., Felciano, R. M., et al. (2006). *Proceedings of the National Academy of Sciences of the United States of America, 103*, 17402.
21. Taylor, I. W., Linding, R., Warde-Farley, D., Liu, Y., Pesquita, C., Faria, D., et al. (2009). *Nature Biotechnology, 27*, 199.
22. Ashley, E. A., Ferrara, R., King, J. Y., Vailaya, A., Kuchinsky, A., He, X., et al. (2006). *Circulation, 114*, 2644.
23. Schadt, E. E., Friend, S. H., & Shaywitz, D. A. (2009). *Nature Reviews. Drug Discovery, 8*, 286.
24. Suthram S, Dudley JT, Chiang AP, Chen R, Hastie TJ, Butte AJ PLoS Comput Biol 6: e1000662
25. Butte, A. J., Tamayo, P., Slonim, D., Golub, T. R., & Kohane, I. S. (2000). *Proceedings of the National Academy of Sciences of the United States of America, 97*, 12182.
26. Yang, X., Deignan, J. L., Qi, H., Zhu, J., Qian, S., Zhong, J., et al. (2009). *Nature Genetics, 41*, 415.
27. Dewy F, Perez M, Wheeler M, Watt C, Spin J, Langfelder P, Horvath S, Hannenhalli S, Cappola T, Ashley E (2010) Circ Cardiov Genetics (in revision)
28. Segal, E., Sirlin, C. B., Ooi, C., Adler, A. S., Gollub, J., Chen, X., et al. (2007). *Nature Biotechnology, 25*, 675.
29. Chen, M. M., Ashley, E. A., Deng, D. X., Tsalenko, A., Deng, A., Tabibiazar, R., et al. (2003). *Circulation, 108*, 1432.
30. Jia, Y. X., Pan, C. S., Zhang, J., Geng, B., Zhao, J., Gerns, H., et al. (2006). *Regulatory Peptides, 133*, 147.
31. Leeper, N. J., Tedesco, M. M., Kojima, Y., Schultz, G. M., Kundu, R. K., Ashley, E. A., et al. (2009). *American Journal of Physiology. Heart and Circulatory Physiology, 296*, H1329.
32. Chun, H. J., Ali, Z. A., Kojima, Y., Kundu, R. K., Sheikh, A. Y., Agrawal, R., et al. (2008). *Journal of Clinical Investigation, 118*, 3343.
33. Perez-Echarri, N., Perez-Matute, P., Marcos-Gomez, B., Martinez, J. A., & Moreno-Aliaga, M. J. (2009). *The British Journal of Nutrition, 101*, 1059.
34. Japp AG, Cruden NL, Barnes G, van Gemeren N, Mathews J, Adamson J, Johnston NR, Denvir MA, Megson IL, Flapan AD, Newby DE Circulation 121: 1818
35. Ali, Z. A., Alp, N. J., Lupton, H., Arnold, N., Bannister, T., Hu, Y., et al. (2007). *Arteriosclerosis, Thrombosis, and Vascular Biology, 27*, 833.
36. Ali, Z., Raiesdana, A., Qu, X., Ali, A., Pavlovic, A., Wheeler, M., et al. (2009). *Circulation, 120*, S1139.
37. Schadt, E. E., Monks, S. A., Drake, T. A., Lusis, A. J., Che, N., Colinayo, V., et al. (2003). *Nature, 422*, 297.
38. Lee, S. I., Dudley, A. M., Drubin, D., Silver, P. A., Krogan, N. J., Pe'er, D., et al. (2009). *PLoS Genetics, 5*, e1000358.
39. Friedman, N., Linial, M., Nachman, I., & Pe'er, D. (2000). *Journal of Computational Biology, 7*, 601.
40. Rockman, M. V. (2008). *Nature, 456*, 738.
41. Dojer, N., Gambin, A., Mizera, A., Wilczynski, B., & Tiuryn, J. (2006). *BMC Bioinformatics, 7*, 249.
42. Segal, E., Shapira, M., Regev, A., Pe'er, D., Botstein, D., Koller, D., et al. (2003). *Nature Genetics, 34*, 166.
43. Litvin, O., Causton, H. C., Chen, B. J., & Pe'er, D. (2009). *Proceedings of the National Academy of Sciences of the United States of America, 106*, 6441.
44. Dobrin, R., Zhu, J., Molony, C., Argman, C., Parrish, M. L., Carlson, S., et al. (2009). *Genome Biology, 10*, R55.
45. Chen, Y., Zhu, J., Lum, P. Y., Yang, X., Pinto, S., MacNeil, D. J., et al. (2008). *Nature, 452*, 429.
46. Ferrara, C. T., Wang, P., Neto, E. C., Stevens, R. D., Bain, J. R., Wenner, B. R., et al. (2008). *PLoS Genetics, 4*, e1000034.
47. Wang, S. S., Schadt, E. E., Wang, H., Wang, X., Ingram-Drake, L., Shi, W., et al. (2007). *Circulation Research, 101*, e11.
48. Zhu, J., Zhang, B., Smith, E. N., Drees, B., Brem, R. B., Kruglyak, L., et al. (2008). *Nature Genetics, 40*, 854.
49. Ashburn, T. T., & Thor, K. B. (2004). *Nature Reviews. Drug Discovery, 3*, 673.
50. Aronson, J. K. (2007). *British Journal of Clinical Pharmacology, 64*, 563.
51. Lum, P. Y., Derry, J. M., & Schadt, E. E. (2009). *Pharmacogenomics, 10*, 203.
52. Xie, L., Li, J., Xie, L., Bourne, P. E., & Nussinov, R. (2009). *PLoS Computational Biology, 5*, e1000387.
53. Kinnings, S. L., Liu, N., Buchmeier, N., Tonge, P. J., Xie, L., & Bourne, P. E. (2009). *PLoS Computational Biology, 5*, e1000423.
54. Yap, Y. L., Zhang, X. W., Smith, D., Soong, R., & Hill, J. (2007). *Computational Biology and Chemistry, 31*, 275.
55. Milano, A., Pendergrass, S. A., Sargent, J. L., George, L. K., McCalmont, T. H., Connolly, M. K., et al. (2008). *PLoS ONE, 3*, e2696.
56. Wang WZ, Guo X, Duan C, Ma WJ, Zhang YG, Xu P, Gao ZQ, Wang ZF, Yan H, Zhang YF, Yu YX, Chen JC, Lammi MJ (2008) Osteoarthritis Cartilage
57. Walsh, C. S., Ogawa, S., Karahashi, H., Scoles, D. R., Pavelka, J. C., Tran, H., et al. (2008). *Journal of Clinical Oncology, 26*, 2952.
58. Antonacopoulou, A. G., Grivas, P. D., Skarlas, L., Kalofonos, M., Scopa, C. D., & Kalofonos, H. P. (2008). *Anticancer Research, 28*, 1221.
59. Dudley, J. T., Tibshirani, R., Deshpande, T., & Butte, A. J. (2009). *Molecular Systems Biology, 5*, 307.
60. Del Rio, M., Molina, F., Bascoul-Mollevi, C., Copois, V., Bibeau, F., Chalbos, P., et al. (2007). *Journal of Clinical Oncology, 25*, 773.
61. Li, J., Wood, W. H., 3rd, Becker, K. G., Weeraratna, A. T., & Morin, P. J. (2007). *Oncogene, 26*, 2860.
62. Fichtner, I., Slisow, W., Gill, J., Becker, M., Elbe, B., Hillebrand, T., et al. (2004). *European Journal of Cancer, 40*, 298.

63. Holleman, A., Cheok, M. H., den Boer, M. L., Yang, W., Veerman, A. J., Kazemier, K. M., et al. (2004). *The New England Journal of Medicine, 351*, 533.

64. Robert, J., Vekris, A., Pourquier, P., & Bonnet, J. (2004). *Critical Reviews in Oncology/Hematology, 51*, 205.

65. Lamb, J., Crawford, E. D., Peck, D., Modell, J. W., Blat, I. C., Wrobel, M. J., et al. (2006). *Science, 313*, 1929.

66. Garman, K. S., Acharya, C. R., Edelman, E., Grade, M., Gaedcke, J., Sud, S., et al. (2008). *Proceedings of the National Academy of Sciences of the United States of America, 105*, 19432.

67. Setlur, S. R., Mertz, K. D., Hoshida, Y., Demichelis, F., Lupien, M., Perner, S., et al. (2008). *Journal of the National Cancer Institute, 100*, 815.

68. Hieronymus, H., Lamb, J., Ross, K. N., Peng, X. P., Clement, C., Rodina, A., et al. (2006). *Cancer Cell, 10*, 321.

69. Loboda, A., Kraft, W. K., Fine, B., Joseph, J., Nebozhyn, M., Zhang, C., et al. (2009). *BMC Medical Genomics, 2*, 7.

70. Wishart, D. S., Knox, C., Guo, A. C., Shrivastava, S., Hassanali, M., Stothard, P., et al. (2006). *Nucleic Acids Research, 34*, D668.

71. Kuhn M, Campillos M, Letunic I, Jensen LJ, Bork P Mol Syst Biol 6: 343

72. Campillos, M., Kuhn, M., Gavin, A. C., Jensen, L. J., & Bork, P. (2008). *Science, 321*, 263.

73. Chiang, A. P., & Butte, A. J. (2009). *Clinical Pharmacology and Therapeutics, 86*, 507.

74. Tatonetti, N. P., Liu, T., & Altman, R. B. (2009). *Genome Biology, 10*, 238.

75. Berger SI, Ma'ayan A, Iyengar R Science Signaling 3: ra30

76. Xie, L., Li, J., & Bourne, P. E. (2009). *PLoS Computational Biology, 5*, e1000387.

77. Scheiber, J., Jenkins, J. L., Sukuru, S. C., Bender, A., Mikhailov, D., Milik, M., et al. (2009). *Journal of Medicinal Chemistry, 52*, 3103.

78. Toyoshiba, H., Sawada, H., Naeshiro, I., & Horinouchi, A. (2009). *Toxicology Letters, 186*, 52.

79. Zimmer, M., Lamb, J., Ebert, B. L., Lynch, M., Neil, C., Schmidt, E., et al. (2010). *Cancer Research, 70*, 3071.

80. Rosenbluth, J. M., Mays, D. J., Pino, M. F., Tang, L. J., & Pietenpol, J. A. (2008). *Molecular and Cellular Biology, 28*, 5951.

81. Hahn, C. K., Ross, K. N., Warrington, I. M., Mazitschek, R., Kanegai, C. M., Wright, R. D., et al. (2008). *Proceedings of the National Academy of Sciences of the United States of America, 105*, 9751.

82. Hassane, D. C., Guzman, M. L., Corbett, C., Li, X., Abboud, R., Young, F., et al. (2008). *Blood, 111*, 5654.

83. Vilar, E., Mukherjee, B., Kuick, R., Raskin, L., Misek, D. E., Taylor, J. M., et al. (2009). *Clinical Cancer Research, 15*, 2829.