Title

Network Medicine in Disease Analysis and Therapeutics

Authors

Chen, B
Butte, AJ

Network Medicine in Disease Analysis and Therapeutics

Bin Chen[1] and **Atul J. Butte**[1,*]

[1]Division of Systems Medicine, Department of Pediatrics, Stanford University School of Medicine, 1265 Welch Road, Room X-163 MS-5415, Stanford, CA, 94305, USA

Abstract

Two parallel trends are occurring in drug discovery. The first is that we are moving away from a symptom-based disease classification system to a system based on molecules and molecular states. The second is that we are shifting away from targeting a single molecule towards targeting multiple molecules, pathways or networks. Network medicine is a network-based approach to understanding disease and discovering therapeutics, and it may play a critical role in the adoption of both trends.

The contemporary classification of diseases based on signs and symptoms relies on expert observational skill and cognitive pattern-matching. However, this approach lacks sensitivity with respect to identifying early stage preclinical disease and lacks specificity in defining disease unambiguously (1). Symptom-based classification may miss opportunities for prevention especially for many diseases that are asymptomatic in early stages. For example, women who carry mutations in *BRCA1* or *BRCA2* are now well known to have a higher risk of breast cancer, and while early breast cancer can be found through mammography, these mutations obviously cannot be detected by simple observation and the use of risk-reducing surgery would otherwise be missed.

As the generation of various types of molecular data in the genomic era continues to increase, a molecular characterization of diseases is now required to build a new taxonomy of disease (or nosology) to devoid the problems of symptom-based classification. Among its benefits, the new disease classification will facilitate more precise diagnosis and personalized treatment. Again, breast cancer, for instance, can now be categorized into major classes by analyzing data from genomic DNA copy number arrays, DNA methylation, exome sequencing, messenger RNA arrays, microRNA sequencing and reverse-phase protein arrays. These classes are encouraging for the discovery of effective individualized treatments (2).

In addition to a molecular-based taxonomy of disease, the ideal level of understanding of disease would include all levels of molecular changes, from DNA to RNA to microRNA to proteins, and even include disease determinants such as environmental factors. With so many complex and interacting factors that may contribute to disease, networks of these interacting factors are a more optimal way to represent and model our understanding of disease. For

---

*Correspondence: Atul J. Butte, abutte@stanford.edu.

example, analysis of a disease-gene network would reveal diseases that are closely related genetically (3), and an integrated analysis of disease-related mRNA expression data and the human protein interaction network can identify common functional modules enriched for pluripotent drug targets as well as some surprising new relationships among diseases (4).

It is evident that many diseases are complex, and caused not by a single factor or genetic variation. Instead, they are consequences of defects in a complex network comprising a multitude of environmental factors, genetic mutations and polymorphisms, whose effects operate from conception through adulthood. Nonetheless, contemporary drug discovery efforts have largely relied on a "magic bullet" to perturb the network by targeting a key component. This approach has led to a number of drugs, such as trastuzumab (HER2 in breast cancer), crizotinib (ALK in non-small cell lung carcinoma), and dabrafenib (BRAF in melanoma). However, as many diseases are driven by complex molecular and environmental interactions, targeting a single component may not be sufficient to disrupt them. For example, EGFR inhibitors are used to treat lung tumors, but the tumors typically develop resistance and relapse occurs due to EGFR-T790M gate-keeper mutations, MET amplification, or induction of FGFR1 and FGF2. This process suggests that targeting other molecules in the tumors may be necessary (5). Furthermore, functional genomics studies have also revealed that many single gene knockouts have no or little effect on phenotype (6).

Given the inherent shortcomings of the "magic bullet" approach, a more appealing strategy in the drug discovery stage is to modulate the disease network by targeting multiple components using a designed polypharmacological ligand or a combination of drugs (6). Examples include trastuzumab in combination with paclitaxel in breast cancer, cetuximab in combination with irinotecan in metastatic colorectal cancer and lapatinib (a dual inhibitor of the EGFR and HER2 tyrosine kinases) in combination with capecitabine in advanced breast cancer. Additionally, combination antiretroviral therapy is known as an effective treatment for HIV.

Integration of disease analysis and drug discovery

Integrating disease analysis into drug discovery is critical to find effective therapeutics for complex diseases. For example, gefitinib was discovered to treat lung cancer patients, but a subsequent study showed that it was only effective in a subset of patients. Analysis of the patient samples identified that EGFR mutations account for the efficacy. This information guides the application of gefitinib only to those patients with EGFR mutations. However, many of those patients tend to develop resistance later on due to compensatory pathways, secondary mutations or other issues. Analysis of the resistant samples may enable elucidation of drug resistance networks or pathways and further sub-classify patients based on resistance mechanisms. This could guide discovery of new drugs to treat each specific subgroup. For instance, an FGFR inhibitor (AZD4547) may have the potential to treat patients who relapse because of the bypass pathway induced by FGFR1 and FGF2 (5). Likewise, in breast cancer, hormone therapy was first discovered for ER+/PR+ patients; subsequently, trastuzumab was discovered to treat patients with HER2 amplification or overexpression. Patients with the triple-negative breast cancer (TNBC) still have a poor prognosis. However, prognosis improves if complete pathologic response is achieved

following administration of genotoxic drugs. Lee et al. recently found that EGFR inhibitors can sensitize TNBC cells to these genotoxic drugs. Further transcriptional, proteomic and computational analyses of signaling networks and phenotypes in drug-treated cells has revealed that the rewired oncogenic signaling pathway is involved in the efficacy increase (7).

The role of public data

But if our goal is thus to study diseases and drugs through layers of molecular measurements made at various levels of resolution, connected by networks of across layers and disease progression and time, it quickly becomes apparent that no one researcher, lab, or institute will be able to collect enough samples to reach this goal. This is why public datasets will continue to be important elements in the analysis of disease and therapeutic discovery using network-based approaches. Table 1 lists public data sources that can be harnessed to create and study networks linking drugs, diseases and other biological factors in specific contexts (e.g., cells, tissues, organisms, populations). For example, biclustering of drug-induced gene expression profiles taken from the Connectivity Map (CMap) revealed transcription modules that can predict novel gene functions, give insights into the mechanisms of drug actions, and provide leads for drug repositioning projects (8). Combining gene expression profiles of diseases from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) and drug-induced gene expression profiles from CMap can help identify novel drug-disease relations (9).

In particular, the recent release of several new flagship datasets in chemical genomics, functional genomics and clinical trials will aid efforts to cross-link multiple data sources. For example, a large-scale drug response network could be created from the NIH Library of Integrated Network-Based Cellular Signatures (LINCS) project. LINCS aims to create over 1 million gene expression profiles of transcriptional response data for off-patent drugs and thousands of genetic reagents, across 15 cell lines including primary and cancer cells. Another source of data with high potential comes from Chip-Seq experiments from ENCODE. Their goal is to create the first complete regulatory networks of the relationships between vital transcription factors and their target genes. Links between clinical outcomes and genomics and cellular measurements could be established using data from the NIAID Immunology Database and Analysis Portal (ImmPort).

There are challenges to realizing the potential of network medicine using public sources, however. For example, the structure of the data is complex and the lack of standards for data representation may hinder integration. Clinical trial datasets, for instance, have different data schemas and ontologies. Semantic web technologies might help to formalize data representation. In addition, current network measurements (e.g., scale-free property) and modeling methods (e.g., random walk algorithm) developed for homogeneous networks need to be adjusted for heterogeneous networks, where a semantic type is assigned to each node (e.g., disease, gene, protein, drug) and each edge (e.g., protein-protein interaction, DNA-protein interaction). Moreover, biological networks are context-dependent. Two transcriptional response networks of the same drug in two cell lines can be completely different. Thus, integrating a drug-response network from one cell line into a network from

another cell line would be problematic. Sometimes, the findings from different studies may be neither reproducible nor robust due to reasons such as improper analysis or validation, small sample size, insufficient control of false positives, etc. Meta-analysis has the potential to increase statistical power and generalizability of single-study analysis (10). Finally, neither the models nor the findings should be complicated from a clinical perspective. For example, there are significant hurdles in executing, interpreting, and paying for multi-marker tests. Nevertheless, these issues are solvable, and as they continue to be resolved, network medicine will continue to increase our understanding of diseases and aid discovery of new therapeutics.

Acknowledgments

References

1. Barabasi AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nature reviews Genetics. 2011; 12:56–68.

2. Comprehensive molecular portraits of human breast tumours. Nature. 2012; 490:61–70. [PubMed: 23000897]

3. Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL. The human disease network. Proceedings of the National Academy of Sciences of the United States of America. 2007; 104:8685–90. [PubMed: 17502601]

4. Suthram S, Dudley JT, Chiang AP, Chen R, Hastie TJ, Butte AJ. Network-based elucidation of human disease similarities reveals common functional modules enriched for pluripotent drug targets. PLoS computational biology. 2010; 6:e1000662. [PubMed: 20140234]

5. Ware KE, et al. A mechanism of resistance to gefitinib mediated by cellular reprogramming and the acquisition of an FGF2-FGFR1 autocrine growth loop. Oncogenesis. 2013; 2:e39. [PubMed: 23552882]

6. Hopkins AL. Network pharmacology: the next paradigm in drug discovery. Nat Chem Biol. 2008; 4:682–90. [PubMed: 18936753]

7. Lee MJ, et al. Sequential application of anticancer drugs enhances cell death by rewiring apoptotic signaling networks. Cell. 2012; 149:780–94. [PubMed: 22579283]

8. Iskar M, et al. Characterization of drug-induced transcriptional modules: towards drug repositioning and functional understanding. Molecular systems biology. 2013; 9:662. [PubMed: 23632384]

9. Sirota M, et al. Discovery and preclinical validation of drug indications using compendia of public gene expression data. Science translational medicine. 2011; 3:96ra77.

10. Ramasamy A, Mondry A, Holmes CC, Altman DG. Key issues in conducting a meta-analysis of gene expression microarray datasets. PLoS medicine. 2008; 5:e184. [PubMed: 18767902]

Table 1

Public data sources for network medicine. Websites were accessed in July 2013.

Data source	URL	Contents
PubChem	http://pubchem.ncbi.nlm.nih.gov	Chemical compounds and bioassay experiments, including >47M unique chemical compounds and >700K bioassays.
ChEMBL	https://www.ebi.ac.uk/chembl	Bioactivities for drug-like small molecules, including 9,844 targets, >1.2M distinct compounds, and >11M activities.
Connectivity Map (CMap)	http://www.broadinstitute.org/cmap	>7,000 gene expression profiles using 1,309 compounds in 3 main cell lines.
Library of Integrated Network-based Cellular Signatures (LINCS)	http://lincscloud.org	Plans to create >1M gene expression profiles representing >5,000 compounds and >3,500 genes, in >15 cell types.
Cancer Cell Line Encyclopedia (CCLE)	http://www.broadinstitute.org/ccle	Genetic and pharmacologic characterization of >1000 cancer cell lines.
Gene Expression Omnibus (GEO)	http://www.ncbi.nlm.nih.gov/geo	Functional genomics data repository with >957,000 samples.
Encyclopedia of DNA Elements (ENCODE)	http://genome.ucsc.edu/ENCODE	Functional elements in the human genome (genes, transcripts, transcriptional regulatory regions, etc.); >800 ChIP-Seq experiments in >90 cell lines; various tissues, treatments, and conditions.
Immunology Database and Analysis Portal (ImmPort)	https://immport.niaid.nih.gov	Flow cytometry, protein and gene expression data, and clinical assessments in immunology: 55 clinical studies and 566 experiments on >13,000 subjects.