Network Pharmacology and Gene Expression Signature-Based Drug Discovery

High-throughput screening (HTS) of small molecule compound libraries represents one of the biggest advances in drug discovery in recent years. Hundreds or thousands of small molecules can be assayed simultaneously for *in vitro* biological or chemical activity in a multi-well format. Compounds that elicit the desired response are considered "hits," and undergo further testing and validation in order to identify lead compounds whose potency, selectivity, and pharmacokinetics can be iteratively improved. Lead compounds are optimized to be specific for a single disease-related target so as to maximize efficacy of treatment and limit toxicity due to off-target effects.

Despite the potential for lead compound identification, there are major limitations to this screening paradigm. First, the desired target must be defined prior to the screen. This is problematic because the targets for many diseases have not been exhaustively characterized. Furthermore, not every protein involved in disease is druggable; less than 15% of human proteins are predicted to be targetable by small molecules[1]. Additionally, the identified hits do not necessarily provide information on the mechanism of action of the compound or on the underlying biology of the disease[2]. The vast majority of identified compounds also are ultimately found to demonstrate unacceptably low efficacy or high toxicity *in vivo*, resulting in high attrition rates in phase 2 and 3 clinical trials and a stagnating success rate for new single-target drugs[3-5]. This is likely due to arguably the biggest limitation of the linear "one drug, one target" model of drug discovery: failure to account for the high degree of complexity of biological systems[2].

IQPC GmbH | Friedrichstr. 94 | D-10117 Berlin, Germany
t: +49 (0) 30 2091 3330 | f: +49 (0) 30 2091 3263 | e: eq@iqpc.de | w: www.iqpc.de

Genes and proteins do not function in linear, isolated pathways, but in interconnected, overlapping networks. It is thus extremely difficult to predict the downstream effects of a poorly characterized novel compound, which may have multiple targets affecting unexpected pathways, resulting in side effects and toxicity. Focusing therapeutic efforts on a single disease target also ignores other proteins in the same biological pathway that may contribute to the disease as well as the possibility of compensatory activities of proteins in the same or unrelated pathways that lead to low efficacy. A network-based approach to drug discovery instead views disease as an altered network topology state with exploitable connections[2].

The impact of a perturbation of single gene is clearly not limited to that gene product, but to many of the proteins it is linked to by either genetic or physical interactions. Proteins involved in same disease have a high likelihood of interacting with each other, meaning that close network connections of disease-related proteins are likely also involved in the same disease. Gene products that are more highly interconnected with each other than any other gene products are referred to as a module, and their expression is often correlated. Several hundred genes can thus act together in a module to drive disease. Each disease has its own unique module, and a gene, protein, or metabolite can function in multiple modules. Identification of disease module for the pathological phenotype of interest is thus a critical step in the design of safer, more effective drug compounds[6].

The need for compound screening approaches that consider such biological complexity is the driving force behind a new concept in drug discovery: network pharmacology. Network pharmacology utilizes principles in systems biology and network analysis to advance drug discovery through the identification of connectivity, redundancy, and pleiotropy in biological pathways[4]. The network underlying a pathological state is characterized based on gene

expression profiling, metabolomic analysis, and/or protein-protein interactions. A key advantage of this approach is that disease modules can be identified without a need for prior knowledge of which genes cause disease or how genes within the module are connected[2]. Counterintuitive, paradoxical, and unexpected networks can be identified[4], likely revealing sources of redundancy and multifunctionality that impact efficacy and toxicity[7]. Because perturbation of individual components is predicted to affect entire modules[2], small molecule compound libraries can then be systematically tested to determine if and how they affect the disease network, limiting the search to compounds that induce detectable change in disease module activity without affecting networks related to other biological processes[6]. The overarching goal of network pharmacology is then to identify compounds that perturb disease networks and recreate normal network topology[5].

The expression status of normal or disease gene modules can easily be determined by microarray- or transcriptome sequencing (RNA-seq)-based profiling. A gene expression signature can then serve as a surrogate for cellular states that can be interrogated by high throughput screening methods. Gene expression-based high-throughput screening (GE-HTS) can be used in cases where the mechanism underlying a disease state is unknown and it is therefore impossible to perform a small molecule screen against a validated target protein[8]. In this cell-based method, genomewide expression data is gathered from normal and disease samples and a select group of genes whose expression levels highly correlate with the normal phenotype are chosen as a signature. A relevant cell type recapitulating the disease state is then grown in multi-well (384-well or higher) plates and treated with the contents of a small molecule compound library. Multiplexed real time-PCR of the signature genes is then performed and the products quantitatively detected by mass spectrometry or bead-based

technology. Top-scoring compounds induce the desired signature without inducing non-specific cell death. GE-HTS has successfully identified compounds that induce a differentiation signature in cancer[8], suppress the effects of a non-druggable, oncogenic fusion protein[9,10], and inhibit the PDGFR signaling cascade[11].

Major advantages of GE-HTS include the lack of a requirement for prior target validation and its adaptability to any cellular phenotype that affects gene expression. GE-HTS also does not require any specialized assays or reagents, unlike traditional assays that require antibodies or reporter constructs[8]. However, a main drawback of GE-HTS is that cell culture-based systems likely are not completely accurate representations of *in vivo* disease. Additionally, causal gene signatures may be obscured by experimental artifacts and sources of variation[2]. Regardless, compounds identified by this approach are now being used in clinical trials for diseases such as acute myeloid leukemia and Ewing sarcoma[8,10]. GE-HTS has also generated valuable data about disease states and has provided insight into novel mechanisms of action of compounds that were thought to be well characterized.

In addition to genomewide expression data, huge amounts of protein-protein interaction and metabolomic data from normal, diseased, and small molecule-treated states have rapidly been generated in recent years. It has been estimated that the amount of information on molecular, cellular, and physiological states has skyrocketed from ~10 gigabytes (GB) in 2000 to ~1 terabyte (1000 GB) in 2004 to ~1 *petabyte* (1,000,000 GB) in 2008[2]. Processing and integrating this existing data will undoubtedly provide valuable information for drug discovery. Virtual screening has long been used to assess large compound libraries *in silico* to identify compounds likely to bind to a given target based on their structure or chemical properties, but the failure of this approach to account for biological relationships has resulted

in limited success in driving drug discovery[12]. Virtual gene expression signature screening can avoid this limitation through the incorporation of information from normal physiological processes, disease states, and responses to small molecules[13].

An example of a gene expression-based virtual small molecule compound screening platform is the Connectivity Map (http://www.broadinstitute.org/cmap/). The Connectivity Map uses pattern-matching tools to detect similarities between gene expression signatures induced by small molecules in cell lines representing multiple physiological and disease states[14]. Users can query the Connectivity Map database, which currently includes more than 7,000 expression profiles representing treatment with varying concentrations of 1,309 compounds, with a gene expression signature from their biological state of interest. This approach can be used to gain insight into the mechanism of an uncharacterized compound[15], or identify compounds that induce signatures that are similar to a desired state[16]. For example, the Connectivity Map successfully predicted that the uncharacterized natural products celastrol and gedunin were heat shock protein 90 (HSP90) inhibitors based on similarities to the expression signatures of HSP90-inhibitor treated cells[15]. The Connectivity Map also correctly predicted that sirolimus, an inhibitor of the mammalian target of rapamycin (mTOR), would be effective in the treatment of dexamethasone-resistant acute lymphoblastic leukemia (ALL) based on the connection to signatures observed in dexamethasone-sensitive cell lines[16]. A main drawback of *in silico* gene expression signature screening is that compound effects that require multiple cell types or *in vivo* microenvironments cannot be queried[17]. However, applications of biological network-based virtual screening are not only limited to gene expression signatures. Screening approaches can exploit other biological relationships between small molecules to identify potential targets

and mechanisms such as similarities in side effects[18] and chemical similarities between known drug and ligand pairs[19].

Network pharmacology has the potential to accelerate drug discovery through a holistic, multidimensional approach to cellular disease states. Evaluating the effects of known and uncharacterized compounds on gene expression networks can reveal unexpected mechanisms of action, which can either be avoided or exploited. Small molecule compounds that exhibit binding to more than one molecular target, or polypharmacology[20], may be effective against multiple nodes of a disease network. A comprehensive characterization of gene expression modules in disease will also promote the development of novel therapeutic approaches that focus on compounds that affect the entire module or target neighborhood genes rather than the disease gene. Because many diseases are driven by small changes in many genes rather than large changes in few genes[2], network pharmacology will drive a "magic shotgun"[20,21] rather than a "magic bullet" approach to drug discovery.

References

[1] Hopkins AL and Groom CR. The druggable genome. *Nature Reviews Drug Discovery* 1(9): p. 727-730 (2002).

[2] Schadt EE, et al. A network view of disease and compound screening. *Nature Reviews Drug Discovery* 8(4): p. 286-95 (2009).

[3] Csermely, P et al. The efficiency of multi-target drugs: the network approach might help drug design. *Trends in Pharmacological Sciences* 26(4): p. 178-182 (2005).

[4] Hopkins AL. Network pharmacology: the next paradigm in drug discovery. *Nature Chemical Biology* 4(11): p. 682-690 (2008).

[5] Nolan, GP. What's wrong with drug screening today. *Nature Chemical Biology* 3(4): p. 187-191 (2007).

[6] Barabási A-L, et al. Network medicine: a network-based approach to human disease. *Nature Genetics* 12(1): 56-68 (2011).

[7] Fitzgerald JB, et al. Systems biology and combination therapy in the quest for clinical efficacy. *Nature Chemical Biology* 2(9): p. 458-466 (2006).

[8] Stegmaier K, et al. Gene expression–based high-throughput screening (GE-HTS) and application to leukemia differentiation. *Nature Genetics* 36(3): p. 257-263 (2004).

[9] Corsello SM, et al. Identification of AML1-ETO modulators by chemical genomics. *Blood* 113(24): p. 6193-6205 (2009).

[10] Stegmaier K, et al. Signature-based small molecule screening identifies cytosine arabinoside as an EWS/FLI modulator in Ewing sarcoma. *PLoS Medicine* 4(4): p. e122 (2007).

[11] Antipova AA, et al. Gene expression-based screening for inhibitors of PDGFR signaling. *Genome Biology* 9(3): p. R47 (2008).

[12] Schneider G. Virtual screening: an endless staircase? *Nature Reviews Drug Discovery* 9(4): p. 273-276 (2010).

[13] Hughes TR, et al. Functional discovery via a compendium of expression profiles. *Cell* 102(1): p. 109-126 (2000).

[14] Lamb, J et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313(5795): p. 1929-1935 (2006).

IQPC GmbH | Friedrichstr. 94 | D-10117 Berlin, Germany
t: +49 (0) 30 2091 3330 | f: +49 (0) 30 2091 3263 | e: eq@iqpc.de | w: www.iqpc.de

[15] Hieronymus, H et al. Gene expression signature-based chemical genomic prediction identifies a novel class of HSP90 pathway modulators. *Cancer Cell* 10(4): p. 321-30 (2006).

[16] Wei G, et al. Gene expression-based chemical genomics identifies rapamycin as a modulator of MCL1 and glucocorticoid resistance. *Cancer Cell* 10(4): p. 331-42 (2006).

[17] Lamb J. The Connectivity Map: a new tool for biomedical research. *Nature Reviews Cancer* 7(1): p. 54-60 (2007).

[18] Campillos M, et al. Drug Target Identification Using Side-Effect Similarity. *Science* 321(5886): p. 263-266 (2008).

[19] Keiser MJ, et al. Predicting new molecular targets for known drugs. *Nature* 462(7270): p. 175-181 (2009).

[20] Roth BL, et al. Magic shotguns versus magic bullets: selectively non-selective drugs for mood disorders and schizophrenia. *Nature Reviews Drug Discovery* 3(4): p. 353-359 (2004).

[21] Hopkins, AL. Drug discovery: Predicting promiscuity. *Nature* 462(7270): p. 167-168 (2009).

IQPC GmbH | Friedrichstr. 94 | D-10117 Berlin, Germany
t: +49 (0) 30 2091 3330 | f: +49 (0) 30 2091 3263 | e: eq@iqpc.de | w: www.iqpc.de