Chapter 13
Network-Driven Analysis Methods and their Application to Drug Discovery

Daniel Ziemek
Pfizer Inc., USA

Christoph Brockel
Pfizer Inc., USA

ABSTRACT

Drug discovery and development face tremendous challenges to find promising intervention points for important diseases. Any therapeutic agent targeting such an intervention point must prove its efficacy and safety in patients. Success rates measured from first studies in human to registration average around 10% only. Over the last decade, massive knowledge on biological systems has been accumulated and genome-scale primary data are produced at an ever increasing rate. In parallel, methods to use that knowledge have matured. This chapter will present some of the problems facing the pharmaceutical industry and elaborate on the current state of network-driven analysis methods. It will focus especially on semi-quantitative methods that are applicable to large-scale data analysis and point out their potential use in many relevant drug discovery challenges.

INTRODUCTION

Drug discovery and development continues to be a high-risk endeavor. Success rates for small molecule therapeutics average to around 10% from entry into Phase 1 ('First in Human') to successful registration across therapeutic indications and companies (Kola & Landis, 2004).

While historically much of the attrition was based on safety issues identified in Phase 1, more recent data indicate that the average success rate in Phase 1 has climbed to 60% and the success rates in Phase 2 and Phase 3 have reached 30-40% and 60%, respectively. Failure of a compound in a late

DOI: 10.4018/978-1-60960-491-2.ch013

stage of development is particularly undesirable, since it has then already incurred most of its total cost. With the total cost of drug discovery and development estimated to approach, and frequently exceed, $1B per registration, reduction in late stage attrition will be critical to affordable therapeutics.

In the pharmaceutical industry, the path to a therapy starts in many cases with the selection of a target, the intervention point that promises to cure or at least treat the disease of interest. Finding the right target is both crucial and non-trivial and there are several different approaches in use. The association of a clinical phenotype with a genetic mutation is very promising, since observations are made directly in human. Unfortunately, even in the case of monogenic diseases where the cause for the clinical phenotype can be linked to a single mutation, this information does not always lead directly to selection of a target. Reasons can be that the activity of the mutated protein cannot be modulated by a small molecule compound or that the mutation doesn't result in any direct change of a protein at all. Good examples for this problem are the known MODY ('Maturity Onset Diabetes of the Young') gene mutations causing early-onset diabetes by rendering the corresponding proteins non-functional. No compounds have been found to counteract this loss-of-function directly. The 'phenotype-mutation-target' approach fails completely for multigenic, complex diseases. Methodologies are needed to mechanistically connect the genetic layer to the clinical phenotype to then enable selection of optimal intervention points.

Once the target is identified, the search for compounds or biologicals that demonstrate selective activity against the target, in in-vitro assays and in-vivo models of the disease begins. These programs are typically lengthy and produce optimized molecules that still fail about half of the time in safety studies in human - often caused by off-target effects. The ones that pass that hurdle then frequently fail to demonstrate efficacy.

The development of technologies that enable the collection of comprehensive 'genome wide', data in biological systems – frequently referred to as 'OMICs' technologies has created the promise to understand human biology and disease on a molecular and mechanistic level.

Dramatic improvements in speed, throughput and cost to sequence DNA and RNA are adding to this promise and introduce the possibility to understand genetic drivers for complex diseases and possibly differences in susceptibility to treatment in different sub-populations. Personalized medicine, the matching of the right treatment for the individual patient seems within reach.

While system wide generation of experimental data in disease models and to some degree in human is feasible today, the interpretation of these data for specific, testable hypotheses is still a significant bottleneck.

As illustrated in Figure 1, the approach to understand disease and drug action involves several stages from experiment to data generation and analysis. Integrating experimental results – possibly recorded at different levels (transcript, protein abundance, protein modification, etc) - with existing biological knowledge is a key step towards interpretation and hypothesis generation. Pathway or network analysis promises to provide the missing link between the lists of statistically significant findings from experiments and the understanding of the underlying biological mechanisms.

In the following, we describe network-based approaches that can help understand disease at a molecular level, select alternate and hopefully better targets for cases where genetic drivers of the disease themselves are not suitable for small molecule approaches, understand the mechanism of action and toxicity of candidate compounds and select mechanistic markers that can be used to confirm exposure of the target to the compound in human. We present some example applications of network analysis, provide a review of the currently available methods/algorithms and close with a discussion of gaps and future perspectives.