Journal of Biomedical Informatics xxx (2013) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Network-based target ranking for polypharmacological therapies

Francesca Vitali [a,*], Francesca Mulas [a], Pietro Marini [b], Riccardo Bellazzi [a]

[a] Dipartimento di Ingegneria Industriale e dell'Informazione, University of Pavia, Pavia, Italy
[b] Demetra Pharmaceutical, Piacenza, Italy

ARTICLE INFO

Article history:
Received 1 February 2013
Accepted 29 June 2013
Available online xxxx

Keywords:
Polypharmacology
PPI network
Network-based bioinformatics
Drug discovery
Target ranking

ABSTRACT

With the growing understanding of complex diseases, the focus of drug discovery has shifted from the well-accepted "one target, one drug" model, to a new "multi-target, multi-drug" model, aimed at systemically modulating multiple targets. In this context, polypharmacology has emerged as a new paradigm to overcome the recent decline in productivity of pharmaceutical research. However, finding methods to evaluate multicomponent therapeutics and ranking synergistic agent combinations is still a demanding task.

At the same time, the data gathered on complex diseases has been progressively collected in public data and knowledge repositories, such as protein–protein interaction (PPI) databases. The PPI networks are increasingly used as universal platforms for data integration and analysis. A novel computational network-based approach for feasible and efficient identification of multicomponent synergistic agents is proposed in this paper. Given a complex disease, the method exploits the topological features of the related PPI network to identify possible combinations of hit targets. The best ranked combinations are subsequently computed on the basis of a synergistic score. We illustrate the potential of the method through a study on Type 2 Diabetes Mellitus. The results highlight its ability to retrieve novel target candidates, which role is also confirmed by the analysis of the related literature.

1. Introduction

Over the past decade, there has been a significant decrease in the rate of translation of new drug candidates into effective clinical therapeutic agents. Many reasons have been argued for such decline of productivity in pharmaceutical research. The assumption of the current approach is that safer, more effective drugs will result from designing very selective ligands for a good target. Ideally, a good target would regulate the pathway of interest: thus, blocking the target would result in effective medical treatment. The rationale for this strategy is that the specificity to the selected target leads to reduced side effects that may be caused by undesirable, non-therapeutic off-target binding. Recently it has been appreciated that many effective drugs (in therapeutic areas as diverse as oncology, psychiatry and antiinfectives) act on multiple rather than single targets, a phenomenon known as polypharmacology [1].

Searching for multi-target drugs is a novel and emerging drug discovery paradigm based on the idea that improved therapeutic efficacy and safety can be achieved by designing new individual chemical entities that can simultaneously target different molecular aspect of a given disease [2]. A multi-target approach to discover innovative medicines is necessary given the multifaceted nature of several complex diseases. In fact, the multidimensional view of diseases is replacing the linear causality model based on the "one disease, one gene, one target" and the "one single-target drug" paradigms. A first step toward this change is the multi-target strategy, where different targets at different key points within the same or concurrent pathogenic pathways are carefully chosen for their potential additive or synergistic effects.

However, the rational design of multi-target drugs faces considerable challenges. These mainly arise from the need for new methods to validate target combinations and to identify preliminary hit compounds.

One of the fundamental advantages of multicomponent therapeutics is the production of "synergy", that is, the combinational effect must be greater than the sum of the individual effects. Although some experimental methods have been proposed to screen favourable drug combinations by disease-relevant phenotypic assays, the high-throughput identification of synergistic agent combinations arising from numerous agents remains an unresolved issue. In this context, computational approaches that take advantage of the rapid accumulation of massive data may provide a more promising and desirable strategy for multicomponent drug studies.

* Corresponding author. Fax: +39 0382 525638.
E-mail addresses: francesca.vitali03@atenopv.it (F. Vitali), francesa.mulas@unipv.it (F. Mulas), scientifico@demetrapharmaceutical.it (P. Marini), riccardo.bellazzi@unipv.it (R. Bellazzi).

1532-0464/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jbi.2013.06.015

Based on the theory that complex disease arise from a change of the equilibrium of a system, we need multi-targeted drugs capable of modulating the balance of the system in an attempt to minimise this change.

In order to develop drugs that meet those standards it is essential to design the system (i.e. network) best describing the complex "omics" interactions (i.e. nodes) of the underlying disease. In particular, a network-based representation and analysis seems the elective strategy to deal with multicomponent therapeutics in complex diseases, as it "naturally" offers new therapeutic views and recommendations for drug repositioning [3,4]. This approach is very powerful when networks are coupled with data collected with -omics technologies (proteomics and genomics) for uncovering dynamic correlations within targets and drug actions. Moreover, it can provide efficient tools to better define the global picture of disease status and dynamic interactions of pathological targets at the molecular network level; likewise, all of these informations can be used for drug design based on network targeting [4].

A number of recent studies have been performed to analyse multi-target drug discovery with a network-based approach [4,5]. In particular, many works revealed that protein and ligand promiscuity is a phenomenon much more common then previously hypothesised. Yildirim et al. applied network analysis to drugs and their targets by integrating publicly available drug data [6]. The resulting network of polypharmacological interactions was dense in nature and revealed not only that a single target can often bind multiple drugs, but also that it is far more common than expected that single drugs modulate several different molecular targets, which may be involved in multiple diseases.

Li et al. have proposed an integrated network model, which is aimed to transfer correlations between drugs to the interactions among their molecular target [7]. The authors computed a score based on two elements, a Topology Score (TS) and an Agent Score (AS), which were used for assessing agents interactions with the biological targets. The TS is derived from the topological features of the background network related to certain disease condition, while the AS is used to quantify the effect of two agents on disease phenotype, as derived by applying text mining on the OMIM disease descriptions. The score was calculated multiplying TS and AS. The higher the score, the greater the probability of synergy of the drug combination.

Li et al. used a biological network-based multi-target computational estimation scheme to screen anticoagulant activity of a series of argatroban intermediates and eight natural products based on affinity predictions from their multi-target docking scores and on a network efficiency analysis [8]. This scheme has been derived from the traditional single agent virtual screening method, which relies on evaluating binding affinity targets. Li et al. built a network by using the Reactome repository and designed a method for selecting targets that can be applied only to the proteins whose virtual screening is known.

We hereby propose a novel computational approach for a efficient identification of multicomponent synergies that relies on a network-based representation of the complex disease.

Our work differs from the papers described above. In particular, instead of performing a drug–drug and a drug–target network as described in [6], our approach focuses on disease-specific protein networks to find new target candidates.

The network is created by integrating different data sources to represent the disease efficiently. In this way, the prediction of the targets also includes the proteins for which a virtual screening is unknown or infeasible, differently from [8]. Although a number of computational approaches have been developed to integrate data from multiple sources to propose new drug candidates, relatively few of them focus on identifying and ranking potential targets, as our investigation suggests. Our work was inspired by the paper of Zhang et al. [9], which ranked potential targets for a specific drug, thus limiting the target discovery to a drug, instead of focusing on a specific disease. The authors also built a network whose nodes were genes. In our case, we build a protein–protein interaction network in order to increase the number of possible associations retrieved, i.e. the physical interaction between proteins and their binding type.

A detailed description of the proposed approach and the results obtained for the selected disease is reported in the following of the paper. Given the social impact, the scientific interest and the high number of available data sets, we chose to test our methods on type 2 Diabetes Mellitus (T2DM), a noteworthy example of multi-factorial and complex disease [10].

2. Material and methods

As mentioned in the previous section, we developed a network-based method that aims at extracting the core disease causative pathways and then at proposing possible combinations of targets suitable for a multicomponent therapy.

The main steps of the methods are described in the following sections.

2.1. Network design

The first step involved the creation of a protein–protein interaction (PPI) network for the disease under study, by integrating different databases and high-throughput datasets. In the resulting biological network, the nodes are proteins and the edges represent various biological associations among them.

The PPI network was built by first retrieving the proteins involved in the Reactome disease pathways [11] and then, on the basis of the proteins retrieved so far, by retrieving the human PPI data in the STRING repository [12]. We selected only the relationships derived from experimental data and with a *confidence score* > 0.7, corresponding to high confidence for the association predicted by STRING [13]; this allowed us building a network with weights on the edges: the weights are equal to the confidence of the association.

2.1.1. Disease network

Starting from the PPI network, the next step was to find out which proteins are differentially expressed in the disease state. Considering that such proteins correspond to the nodes of the network, we called them *disease proteins* (DP). To identify the DP, we used human microarray data downloaded from Stanford Microarray Database (SMD) [14] and Gene Expression Omnibus (GEO) [15]. This selection was computed on the basis of the fold change (FC) obtained from case-versus-control comparisons with a cutoff of 2. Then, we matched these genes with the genes that encode the network's proteins in order to identify the over- or under-expressed nodes. This step clearly has limits, as it does not take into account post-translational proteins modification and regulation.

2.1.2. Target features

Once DPs had been identified, the next step consisted in network targeting. Instead of selecting all the network nodes as possible targets, we decided to introduce some constraints that allowed restricting the nodes space. We called these proteins $Source_T$ S_T, because they are the sources of a potential synergistic pharmacological effect.

First, we decided to discard hub nodes, i.e. highly connected nodes. Besides the special topological and functional significance in a network, hubs have also special biological properties: they