Computers in Biology and Medicine 61 (2015) 163–171

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Integration of a prognostic gene module with a drug sensitivity module to identify drugs that could be repurposed for breast cancer therapy

Lida Zhu, Juan Liu *

Key State Laboratory of Software Engineering, School of Computer, Wuhan University, Wuhan 430072, PR China

ARTICLE INFO

Article history:
Received 25 February 2014
Accepted 22 December 2014

Keywords:
Prognostic associated module
Drug sensitivity module
Drug reposition
Breast cancer

ABSTRACT

Background: Efficiently discovering low risk drugs is important for drug development. However, the heterogeneity in patient population complicates the prediction of the therapeutic efficiency. Drug repositioning aiming to discover new indications of known drugs provides a possible gateway.
Method: We introduce a novel computational method to identify suitable drugs by using prognosis information of patients. First, we identify prognostic related gene modules, Prognostic Gene Ontology Module (PGOMs), by incorporating multiple functional annotations. Then, we build the drug sensitivity modules based on gene expressions and drug activity patterns. Finally, we analyze the potential effects of drugs on prognostic gene modules and establish the links between PGOMs and drugs.
Result and discussion: With PGOMs generated based on the patient outcome, FDA approved drugs for breast cancer treatment have been successfully identified on one hand; several drugs that have not been approved by FDA, such as Etoposide, have found to strongly associate with the outcome on the other hand. With PGOMs generated based on the patient ER status, Tamoxifen and Exemestane rank at the top of the drug list, suggesting that they may be more specific to ER status of breast cancer. Especially, the rank difference of Exemestane in ER+ group and ER− group is very large, demonstrating that Exemestane may be more specific to ER+ breast cancer and would cause side-effect to ER− breast cancer patients. Our method can not only identify the drugs that could be repurposed for breast cancer therapy, but also can reveal their effective pharmacological mechanisms.

1. Introduction

Due to the expensive cost and time consuming processes in traditional drug discovery, the number of new drugs approved by FDA has declined recently [1]. Under the circumstances, one possible strategy to speed up the drug discovery and lower the developing cost is drug repositioning which aims to discover new indications of known drugs based on their comprehensive profiles. Some of the most successful examples of drug repositioning include Pfizer's Viagra (for erectile dysfunction) and Celgene's thalidomide (for severe erythema nodosum leprosum) [2]. However, the successful examples of drug repositioning are just from the experimental or clinical stage, for the underlying molecular mechanisms of complex diseases are still unclear.

On one hand, human diseases are generally caused by the abnormalities of multiple genes. These dysfunctional genes may further perturb other biological functional pathways and rewire the whole cellular network [3,4], which may lead to some other diseases. On the other hand, with the rapid development of high-throughput technologies, the study of drug action has been characterized within the context of regulatory networks that might include additional undiscovered drug targets. Therefore, it is increasingly important to find the disease–drug associations, which are then used for repositioning drugs, by examining both drug actions and disease outcomes via biological functional modules.

Up to now, a great number of computational methods have been proposed to reposition drugs. Some methods characterize the similar components of drugs by first building “disease–disease” and “drug–drug” associations, and then discovering new “drug–disease” associations by imposing the known “drug–disease” connections. The similar components of drugs vary from the overlap of the target protein sequences to the integrated features of multiple sources such as compound chemistry structures, phenotypes and side effect indications [5,6]. Incorporating microarray technology, some researches focus on using the gene expression data to build the connections between drugs and diseases for drug repositioning. For example, the “Connectivity Map” (CMAP) project has produced a reference collection of gene expression profiles by treating cultured human cells with bioactive small molecules; and compounds that have the most negative correlation with the gene expression profiles are identified as

* Corresponding author. Tel./fax: +86 27 68775711.
E-mail address: liujuan@whu.edu.cn (J. Liu).

http://dx.doi.org/10.1016/j.compbiomed.2014.12.019
0010-4825/

potential therapeutics candidates [7,8]. CMAP is used for gene expression profiling on drug perturbation. However, the mystery of the molecular mechanism of diseases still remains uncovered. Recently, some studies have shown that cellular components belonging to the same topological, functional, or disease modules have a high likelihood of being involved in similar diseases [9,10]. Thus, one can discover repurposed drugs by beginning with identifying gene modules that are associated with the same topologies, functions or diseases. A gene module is a group of genes that share a similar function or regulation mechanism and can be defined as: (1) a set of genes annotated by using the same biological process or cellular component or Gene Ontology (GO) molecular function term [11,12]; (2) a set of genes associated within a network from protein–protein interaction (PPI) [13]; (3) a set of genes involved in the same pathway[14]; or (4) a set of genes regulated by the same miRNA [15,16]. Up to now, there are some studies that reposition drugs based on the gene modules existed. For instance, Li et al. [17] defined the Gene Ontology Modules (GOMs) based on GO; and then they investigated the influence of diseases on GOMs and searched for the repositioned drugs based on the similarity of GOMs. Using GO, Xia et al. [18] investigated the functions of genes regulated by the differential miRNAs in various diseases and the functions of target genes of small molecules (drugs), and identified the links between diseases and small molecules. The existed work did not take the clinical prognosis information of patients into consideration, though some studies have suggested that the therapy of patients could be guided by related prognostic signature [19]. Since many studies on prognoses of diseases were also sampled by high-throughput gene expression profiles, it is possible to build an appropriate in silico model that could extract the intrinsic yet complex genetic interaction patterns from such prognosis data. Therefore, we propose to incorporate the clinical prognosis data with multiple functional annotations to generate the prognosis related gene modules, named as the Prognostic Gene Ontology Modules (PGOMs) hereinafter. In this work, we use associative-learning strategies based on multiple clinical trials to identify the PGOMs. To avoid overestimation of true clinical accuracy, we use two separate trial cohorts to train and evaluate the predictive signature respectively. In the meantime, by linking the chemotherapy information with bioinformatics, we define a sensitivity module for each drug to interrogate the mechanism of drug actions. The sensitivity module for each anti-cancer drug is generated by correlating the sensitivity data (e.g. GI50) from the DrugBank with the National Cancer Institute's NCI60 cell line panel profiling [20,21]. By analyzing the significant perturbation effects of drug sensitive modules on the PGOMs, we can build links between PGOMs and drugs and can then find the repurposed drugs. Since breast cancer is highly heterogeneous, and some studies have shown that estrogen receptor (ER) status in breast cancer is remarkably associated with the prognostic value [19], we further identify specific gene modules for patients with ER clinical status, trying to provide insights of underlying mechanisms into the association between drug sensitivity and disease.

2. Material and method

2.1. Main framework

Here we proposed a systematic framework to identify the potential candidate drugs for breast cancer and the associated mechanism modules. The framework consists of three parts: (1) identification of the PGOMs: the modules were selected based on the statistical differences among the prognostic status of the patients from clinical trials of breast cancer; (2) generation of the sensitivity module network of drugs: the sensitivity module network of drugs was used to interrogate the mechanism of drug action; and (3) integration of the results of (1) and (2), to search for the potential candidate drugs. In the final step, we integrated the PGOMs with the anti-cancer drug sensitivity modules to filter the most significant potential anti-cancer treatment for breast cancer as shown in Fig. 1.

2.2. Materials and data preparation

We downloaded the mRNA expression data and the corresponding clinical information (time to relapse or last follow-up and ER status score) of three breast cancer clinical trials (GSE2034, GSE7390 and GSE11121). Patients with positive or negative relapse scores were classified as those with poor or good prognosis respectively. All the data were processed by using Affymetrix HG-U133A gene-chips, and normalized by MAS5. All the probes in the mRNA expression data were mapped into Entrez Gene ID as

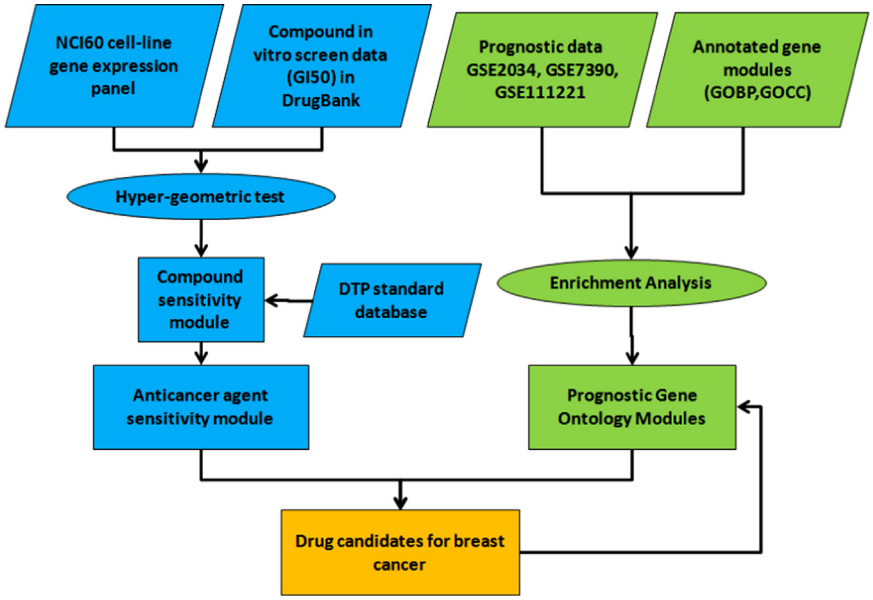

Fig. 1. The workflow of the method integrating the PGOMs with the drug sensitivity modules.