Gene 518 (2013) 201–208

Contents lists available at SciVerse ScienceDirect

Gene

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

In-silico drug screening and potential target identification for hepatocellular carcinoma using Support Vector Machines based on drug screening result

Wu-Lung R. Yang [a], Yu-En Lee [b], Ming-Huang Chen [c,d], Kun-Mao Chao [a,e,*], Chi-Ying F. Huang [c,f,**]

[a] Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan
[b] Institute of Biotechnology in Medicine, National Yang-Ming University, Taipei, Taiwan
[c] Institute of Clinical Medicine, National Yang-Ming University, Taipei, Taiwan
[d] Division of Hematology and Oncology, Department of Medicine, Taipei Veterans General Hospital, Taipei, Taiwan
[e] Graduate Institute of Biomedical Electronics and Bioinformatics, National Taiwan University, Taipei, Taiwan
[f] Institute of Biopharmaceutical Sciences, National Yang-Ming University, Taipei, Taiwan

ARTICLE INFO

Article history:
Accepted 27 November 2012
Available online 6 December 2012

Keywords:
Drug screening
Support Vector Machines
Hepatocellular carcinoma

ABSTRACT

Hepatocellular carcinoma (HCC) is a severe liver malignancy with few drug treatment options. In finding an effective treatment for HCC, screening drugs that are already FDA-approved will fast track the clinical trial and drug approval process. Connectivity Map (CMap), a large repository of chemical-induced gene expression profiles, provides the opportunity to analyze drug properties on the basis of gene expression. Support Vector Machines (SVM) were utilized to classify the effectiveness of drugs against HCC using gene expression profiles in CMap. The results of this classification will help us (1) identify genes that are chemically sensitive, and (2) predict the effectiveness of remaining chemicals in CMap in the treatment of HCC and provide a prioritized list of possible HCC drugs for biological verification.

Four HCC cell lines were treated with 146 distinct chemicals, and cell viability was examined. SVM successfully classified the effectiveness of the chemicals with an average Area Under ROC Curve (AUROC) of 0.9. Using reported HCC patient samples, we identified chemically sensitive genes that may be possible HCC therapeutic targets, including MT1E, MYC, and GADD45B. Using SVM, several known HCC inhibitors, such as geldanamycin, alvespimycin (HSP90 inhibitors), and doxorubicin (chemotherapy drug), were predicted. Seven out of the 23 predicted drugs were cardiac glycosides, suggesting a link between this drug category and HCC inhibition. The study demonstrates a strategy of *in silico* drug screening with SVM using a large repository of microarrays based on initial *in vitro* drug screening. Verifying these results biologically would help develop a more accurate chemical sensitivity model.

1. Introduction

Hepatocellular carcinoma (HCC) is the most common liver malignancy and one of the leading causes of cancer death worldwide. It is an aggressive tumor with poor prognosis, and its 5-year survival rate is below 12% (El-Serag, 2011). Surgical resection is the main form of therapy; however, only few patients are resectable due to difficulties of early diagnosis and poor liver preservation (Chen et al., 2003; El-Serag et al., 2008). Currently, only one drug, sorafenib (Nexavar), has been approved for HCC treatment by the FDA. However, in a Phase III, double-blind, placebo-controlled clinical trial, the median overall survival in the sorafenib group was prolonged by only 2.3 months compared with the placebo group (Cheng et al., 2008). It is therefore of great urgency to identify additional drug candidates for the treatment of HCC.

To address the issue, we have built and maintained a HCC-associated gene database, EHCO2 (Hsu et al., 2007), integrating several HCC transcriptomic and proteomic studies, as well as literature from text-mining. With the HCC-associated gene signatures, we utilized Connectivity Map (CMap) (Lamb, 2006), a chemical-induced gene expression repository and software to calculate gene expression similarities. With CMap, we were able to conduct pattern matching between the expression profiles of chemicals or drugs and our HCC gene signatures (Chen et al., 2011). One of the greatest advantages of using CMap is the possibility for drug repurposing, the practice of using drugs for treating diseases other than their original indications. Drug repurposing will help cut costs and the time spent in clinical trials because the drug has already been found to be safe for human consumption (Chong and Sullivan, 2007). Out of 1309 chemicals represented in CMap, 965 are

Abbreviations: HCC, hepatocellular carcinoma; CMap, Connectivity Map; MTT, 3-(4,5-cimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide; IC_{50}, concentration required to inhibit cell growth by 50%.

* Correspondence to: K.-M. Chao, Department of Computer Science and Information Engineering, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei 10617, Taiwan. Tel.: +886 2 33664888; fax: +886 2 23628167.
** Correspondence to: C.-Y.F. Huang, Institute of Biopharmaceutical Sciences, National Yang-Ming University, No.155, Sec.2, Linong Street, Taipei 112, Taiwan. Tel.: +886 2 28267000; fax: +886 2 28745074.
E-mail addresses: f95922170@ntu.edu.tw (W.-L.R. Yang), youann.lee@gmail.com (Y.-E. Lee), mhchen9@vghtpe.gov.tw (M.-H. Chen), kmchao@csie.ntu.edu.tw (K.-M. Chao), cyhuang5@ym.edu.tw (C.-Y.F. Huang).

0378-1119/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.gene.2012.11.030

approved drugs. The focus of our study is to find an FDA-approved drug that can be used for the treatment of HCC.

In our previous study (Chen et al., 2011), we employed several techniques, including clique analysis and majority votes, to extract key genes from 4020 HCC-associated genes in EHCO2 and utilized CMap to hypothesize possible drug candidates for treating HCC. Overall, 28 out of the 50 predicted chemicals successfully killed the HCC cell lines (using cell-proliferation assays) or inhibited their colony formation ability (using cell-colony formation assays), confirming the usefulness of CMap. This study provided us with several HCC chemical candidates, and the next immediate challenges are (1) to understand the mechanism of action (MOA) of these chemicals, and (2) to broaden the search for other chemicals in CMap.

In this study, we continued the search for HCC drug candidates. A total of 146 chemicals in CMap have been tested for effectiveness against HCC using several HCC cell lines. Several of these chemicals, predicted ineffective from CMap, were ultimately shown to be effective, indicating the need for a more accurate prediction model. To this end, we employed Support Vector Machines (SVM) to classify the effectiveness of the chemicals, in the hopes of building a signature that can predict the effectiveness of chemicals from their gene expression profiles. To determine the effectiveness of a given chemical against a cell line, one could instead conduct a cell viability test without any computational predictions. However, with thousands of chemicals undergoing testing, computational predictions will provide us with a priority list of drugs to screen biologically. Furthermore, high throughput drug expression profiling of 4000 small-molecule compounds will soon be available (http://www.broadinstitute.org/lincs/), making the prioritization process crucial. We hope to demonstrate the success of establishing the pipeline of drug screening using CMap as an initial prediction model and subsequently SVM for a more precise prediction. Finally, the reversed gene signatures will provide us with a hypothesis on the mechanism of action of these drugs.

2. Material and methods

A flowchart of the study is illustrated in Fig. 1.

2.1. Cell lines

Four hepatocellular carcinoma cell lines, Mahlavu (Human), Huh7 (Human), PLC5 (Human), and HepG2 (Human), were used in this study. The Huh7 and PLC5 cell lines were obtained from Dr. Zhong-Zhe Lin, National Taiwan University Hospital, Taiwan. The HepG2 cell line was purchased from Bioresource Collection and Research Center (BCRC, http://www.bcrc.firdi.org.tw/). The Mahlavu cell line was provided by Dr. Muh-Hwa Yang, National Yang-Ming University, Taiwan. The 4 HCC cell lines were maintained in Dulbecco's Modified Eagles Medium (DMEM, GIBCO) with 10% fetal bovine serum (FBS, GIBCO), 1× non-essential amino acids (NEAA, GIBCO), 1× L-glutamine (GlutaMAXTM-I Supplement, GIBCO) and penicillin–streptomycin (PSA, GIBCO).

2.2. Chemicals

Supplementary File 1 lists the vendors where the 146 chemicals were purchased. Geldanamycin and withaferin A were dissolved as 1 μM stocks, while all of the other drugs were dissolved as 10 μM stocks and stored at −20 °C. The stocks (other than geldanamycin and withaferin A) were diluted to 10 μM, 3.3 μM, 1 μM and 0.33 μM.

2.3. 3-(4,5-Cimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide (MTT) assay

Cell survival rates were assessed by MTT colorimetric assay. MTT is taken up by living cells by endocytosis to yield a blue formazan product. Four cell lines were seeded in a 96-well plate (1500–4000 cells/well) overnight and treated with each of the 146 drugs selected from the CMap. After 72 h of treatment, the cells were incubated with 0.5 μg/ml MTT for 2 h. The medium was removed, and the cells were dissolved in deep-blue crystal with 100% DMSO at room temperature for 10 min. The OD values were then measured at 570 nm by an ELISA reader.

2.4. Drug response profiles

The chemical response profiles were downloaded from the CMap (Lamb, 2006), specifically, the file 'ratioMatrix.txt'. CMap hosted the gene expression profiles of 1309 chemicals in 4 cell lines (MCF7, PC3, HL60, and SKMEL5), resulting in 6100 profiles, as well as 959 control profiles. The 'ratioMatrix.txt' file contains the fold changes of the genes for all of the drug treatment profiles. To fairly weigh the effects of up-regulated and down-regulated expressions, all fold change values were log-2 transformed.

Because SVM is a binary classifier, two binary classes are required. With the cell survival rate for each drug being a continuous value, we

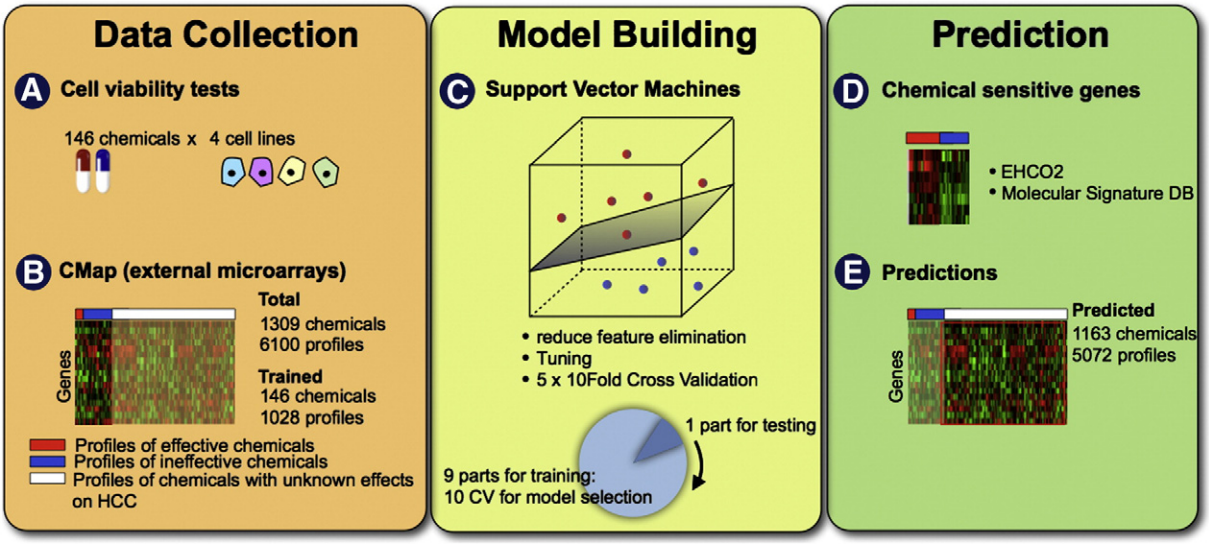

Fig. 1. Flowchart. A) The cell viability of 4 HCC cell lines treated with 146 chemicals was determined. B) The corresponding gene expression profiles were collected and processed. C) Support Vector Machines were used to classify the effectiveness of the drugs. D) The gene features chosen were compared against EHCO2 data and Molecular Signature Database for pathway enrichment study. E) Predictions were made to the remaining CMap chemicals.