Predicting therapeutic and aggravating drugs for hepatocellular carcinoma based on tissue-specific pathways

Liang Yu1*, Fengdan Xu1, Lin Gao1

1School of Computer Science and Technology, Xidian University, Xi’an, PR China

Abstract
Motivation: Hepatocellular carcinoma (HCC) is a significant health problem worldwide and annual number of cases are nearly more than 700,000. However, there are few safe and effective therapeutic options for HCC patients. Here, we propose a new approach for predicting therapeutic and aggravating drugs for HCC based on tissue-specific pathways, which considers not only liver tissue and functional information of pathways, but also the changes of single gene in pathways.

Results: Firstly, we map genes related to HCC to the liver-specific protein interaction network and get an extended tissue-specific gene set of HCC. Then, based on the extended gene set, 12 enriched KEGG functional pathways are extracted. Using Kolmogorov–Smirnov statistic, we calculate the therapeutic scores of drugs based on the 12 tissue-specific pathways. Finally, after filtering by Comparative Toxicogenomics Database (CTD) benchmark, we get 3 therapeutic drugs and 3 aggravating drugs for HCC. Furthermore, we validate the 6 potentially related drugs of HCC by analyzing their overlaps with drug indications reported in PubMed literatures, and also making CM approach similarity analysis and KEGG enrichment analysis based on their targets. All analysis results suggest that our approach is effective and accurate for discovering novel therapeutic options for HCC and it can be easily extended to other diseases. More importantly, our method can clearly distinguish therapeutic and aggravating drugs for HCC, which also can be used to indicate unmarked drug-disease associations in CTD as positive or negative.

Contact: lyu@xidian.edu.cn (LY)

1 Introduction
Hepatocellular carcinoma (HCC), also called malignant hepatoma, is a major health problem being one of the leading causes of death worldwide with annual number of cases exceeding 700,000. It mostly affects patients with liver cirrhosis and currently is their most common reason of death. However, there are few safe and effective therapeutic options for HCC patients. Drug development takes too long and costs too much before taking new drugs to market (Booth and Zemmel, 2004). By rough estimating, it takes about 15 years (DiMasi, 2001) and $800 million to bring a single drug to market (Booth and Zemmel, 2004). Drug repositioning has been used as a profitable and successful strategy for decades to save time and money for drug discovery and development. Because the existing drugs are usually vetted in terms of safety, dosage and toxicity, they can be applied in clinical phases more rapidly than newly developed drugs (Ashburn and Thor, 2004). Well-known examples include Minoxidil (originally designed for hypertension; now indicated for hair loss) (Lucky et al., 2004), Sildenafil (originally designed for erectile dysfunction) (Boolell et al., 1996), Amphotericin B (originally designed for serious systemic fungal infections; now indicated for acute promyelocytic leukemia)(Kleiotriene et al., 2013) and Lipitor (originally designed for cardiovascular disease; now indicated for Alzheimer's) (Feldman et al., 2010).

As the development of technology over the past decade, researchers have proposed many computational methods for new drug indications through various strategies. Human genetics studies offer the strongest evidence to connect genes to human diseases. Zhang et al. (2016) identified 22 liver cancer-related enhancer SNPs through integrating GWAS data and histone modification CHIP-seq data in 2016. Sanseau et al. (2012) found that disease genes extracted from GWAS data were enriched 2.7-fold in drug targets. Besides, they uncovered 123 drug targets which were associated with GWAS traits when not to consider their original indications. These mismatches opened the door for drug repositioning. Bamshad et al. (2011) dissected the genetic basis of diseases and traits by exom sequencing which is the targeted sequencing of the subset of the human genome that encodes proteins. Their strategies have focused on identifying individual genes that exhibit differences between normal and disease
states. In fact, genes within a cell do not function alone. They interact with each other to carry out specific biological functions. Complex diseases are generally caused by the dysregulation of a set of genes that share common biological function.

Some pathway-based computational approaches were proposed recently to overcome this problem. Ye et al. (2012) integrated known drug target information and proposed a disease-oriented strategy for evaluating the relationship between drugs and disease based on their pathway profiles. Pathway profile is a vector with 185 dimensions and each element in the vector is a p-value of the hypergeometric test corresponding to 185 KEGG pathways (Kanehisa et al., 2010). In the next year, Li and Lu (2013) constructed a multi-layer causal network (CauseNet) consisting of chains from drug to target, target to pathway, pathway to downstream gene, and gene to disease. The transition likelihood of each causal chain in the network was measured by a statistical method based on known drug-disease treatment relationships. Zhao et al. (2013) extended the concept of cancer signaling bridges (CSB) and developed a computational model to derive specific downstream signaling pathways that revealed previously unknown target-disease connections and new mechanisms for specific cancer subtypes. Pathway analysis can effectively uncover the perturbation of biological systems by diseases. However, a common problem of these methods is that they treat pathways simply as gene sets and ignore the changes of single gene in pathways and tissue-specific information of pathways. Moreover, these methods cannot clearly distinguish therapeutic and aggravating drugs for diseases.

In this study, we propose an approach to identify related drugs for treatments of HCC, which considers the functional information of pathways, the changes of single gene in pathways and the special relationship between HCC and liver tissue. More importantly, our method can clearly distinguish the potential drugs are therapeutic or aggravating for HCC, which can be used to indicate unmarked drug-disease associations in CTD as positive or negative. Fig 1 gives the flowchart of our proposed method. First, we map all the genes related to HCC to the tissue-specific protein interaction network of liver and get an extended tissue-specific gene set of HCC. Then, based on the extended gene set, we extract 12 liver-specific pathways by KEGG pathway enrichment analysis. Using Kolmogorov–Smirnov statistic (Hollander et al., 2013 ), we calculate the therapeutic scores of drugs based on the 12 pathways. Finally, after filtering by Comparative Toxicogenomics Database (CTD) benchmark (Davis et al., 2015), we get top-6 potential drugs by analyzing their overlaps with drug indications reported in PubMed literatures. In addition, we also make Connectivity Map (CMap)(Lamb et al., 2006) profiles similarity analysis and KEGG enrichment analysis on their related genes. All these findings suggest that our approach is effective for accurate discovering novel therapeutic options for HCC and easily to be extended to other diseases.

**2 Methods**

### 2.1 Extract tissue-specific subnetwork of HCC.

HCC is a cancer that starts in the liver. In order to obtain the tissue-specific subnetwork of HCC, we first get nine related genes of HCC from Online Mendelian Inheritance in Man (OMIM) (Hamosh et al., 2005), shown in Table 1. OMIM is a comprehensive, authoritative compendium of human genes and genetic disorders to support research and education in human genomics and the practice of clinical genetics. Then, the nine genes related to HCC are extended through obtaining the direct neighbors of nine genes and genetic disorders to support research and education in human genomics and the practice of clinical genetics. Then, the nine genes related to HCC are extended through obtaining the direct neighbors of nine genes.

![Fig 1. The framework of our method to predict related drugs for HCC.](image-url)

(1) We compare control samples with tumor samples by Limma to generate a list of up-regulated and down-regulated genes for HCC disease. Genes are ranked by their expression values. The red circles represent up-regulated genes, while the blue circles represent down-regulated genes. (2) We integrate disease related genes and liver tissue-specific network to construct a tissue-specific genes signature of HCC. Based on the tissue-specific gene sets, we extract 12 related pathways by KEGG pathway enrichment analysis. The purple circles represent the related genes of HCC, the other green circles represent other genes in the network. (3) Genes in each pathway are ranked based on their expressions obtained from (1). (4) 6100 gene expression instances of drugs from the Connectivity Map. (5) Different orders of the same gene in pathway \( P(i = 1,2,...,12) \) and instance \( I(j = 1,2,...,6100) \). Based on the different orders of genes in pathway \( P(i = 1,2,...,12) \) and instance \( I(j = 1,2,...,6100) \), we calculate the therapeutic score of each drug using Kolmogorov–Smirnov statistic. (6) Drugs are ranked according to their therapeutic scores. The orange bars (therapeutic scores > 0) represent the therapeutic drugs and the green bars (therapeutic scores < 0) represent the aggravating drugs.
2.2 Construct HCC tissue-specific pathway signatures.

Because HCC is a complex disease and complex diseases generally involve several gene mutations and pathway dysregulations (Chen et al., 2009; Chen et al., 2010), we further analyze the biological functions of 58 genes got from the tissue-specific PPI network of HCC. By using DAVID tool (Huang et al., 2009), we perform pathway enrichment analysis on these genes. The parameters of DAVID are set as: p-value = 0.001 and count = 5. Finally, we obtain 12 KEGG pathways (Kanehisa et al., 2010) related to the 58 genes (see Table 3). There are totally 28 of 58 (48.3%) tissue-specific genes are enriched in these 12 KEGG pathways. Therefore, based on 12 HCC-specific pathways and 6100 drug-related gene expression instances, we use a nonparametric, rank-based pattern matching strategy originally introduced by Lamb et al. (2006) to evaluate the relationship between drugs and HCC (shown in Fig 1 (5)).

We take each ranked drug expression profile as reference signature and assess their similarity to each HCC pathway. For a HCC pathway $i$ and a drug expression profile $j$, we compute a connectivity score separately for the set of up- or down-regulated genes in the pathway $i$: $CS^{up}_i$ or $CS^{down}_i$.

For the HCC pathway $i$, it needs up- ($G_{up}$) or down-regulated genes ($G_{down}$) as inputs. The computational formulas as follows (Lamb et al., 2006):

\[
\begin{align*}
    a &= \max_{p \in \mathbb{L}} \left\{ \frac{p}{m} \cdot \frac{V\{p\}}{n} \right\} \\
    b &= \max_{p \in \mathbb{L}} \left\{ \frac{V\{p\}}{n} \cdot \frac{p-1}{m} \right\} \\
    CS &= \begin{cases} 
    a & \text{(if } a > b) \\
    -b & \text{(if } a \leq b) 
    \end{cases}
\end{align*}
\]

Where $n$ represents the total number of genes in the reference drug expression profile $j$; $m$ represents the number of up-regulated genes $G_{up}$, or the number of down-regulated genes $G_{down}$; $p$ represents the po-

---

**Table 1.** HCC related genes extracted from OMIM

| Gene Names | Gene Entrez IDs |
|------------|-----------------|
| IGF2R      | 3482            |
| CASP8      | 841             |
| MET        | 4233            |
| PDGFRL     | 5157            |
| TP53       | 7157            |
| PIK3CA     | 5290            |
| LCO        | 3935            |
| CTSNB1     | 1499            |
| AXIN1      | 8312            |

**Table 2.** Network topological attributes of the tissue-specific sub-network of HCC

| Attribute names       | Values |
|-----------------------|--------|
| Number of nodes       | 58     |
| Number of edges       | 838    |
| Average node degree   | 28.9   |
| Average clustering coefficient | 0.69   |
| Average the shortest path length | 1.5    |
| Maximal diameter      | 3      |

2.3 Calculate the therapeutic scores of drugs.

We get 6100 gene expression instances covered 1,309 drugs from Cmap database. In other words, a drug may correspond to multiple instances. We rank the genes in each instance by taking the differential expression values between drug-treated and drug-un treated cell lines (shown in Fig 1(4)). Finally, we get 6100 drug-related gene lists. If the gene expression profiles related to a drug positively correlate to that of HCC-specific pathways, the drug may worsen HCC. On the contrary, if their relationship is negative, the drug may treat HCC. The reason is that the positive correlations show that drugs may strengthen the expressions of disease genes and the negative correlations represent that the drugs may inhibit the expressions of disease genes. Therefore, based on 12 HCC-specific pathways and 6100 drug-related gene expression instances, we use a nonparametric, rank-based pattern matching strategy originally introduced by Lamb et al. (2006) to evaluate the relationship between drugs and HCC (shown in Fig 1 (5)).

We take each ranked drug expression profile as reference signature and assess their similarity to each HCC pathway. For a HCC pathway $i$ and a drug expression profile $j$, we compute a connectivity score separately for the set of up- or down-regulated genes in the pathway $i$: $CS^{up}_i$ or $CS^{down}_i$. For the HCC pathway $i$, it needs up- ($G_{up}$) or down-regulated genes ($G_{down}$) as inputs. The computational formulas as follows (Lamb et al., 2006):

\[
\begin{align*}
    a &= \max_{p \in \mathbb{L}} \left\{ \frac{p}{m} \cdot \frac{V\{p\}}{n} \right\} \\
    b &= \max_{p \in \mathbb{L}} \left\{ \frac{V\{p\}}{n} \cdot \frac{p-1}{m} \right\} \\
    CS &= \begin{cases} 
    a & \text{(if } a > b) \\
    -b & \text{(if } a \leq b) 
    \end{cases}
\end{align*}
\]

Where $n$ represents the total number of genes in the reference drug expression profile $j$; $m$ represents the number of up-regulated genes $G_{up}$, or the number of down-regulated genes $G_{down}$; $p$ represents the po-
sition of the input set \( p = 1 \ldots m \); \( V(p) \) is the position of the \( p \)-th input gene in the gene list of drug expression profile \( j \); \( CS \) represents the connectivity score \( CS_{up} \) or \( CS_{down} \) for the up- or down-regulated gene set.

The therapeutic score \((TS)\) between a drug with \( k \) instances in CMap and HCC with 12 functional pathways is calculated as follows:

\[
TS = \frac{1}{k} \sum_{j=1}^{k} CS_{ij}
\]

(4)

where the drug has \( k \) instances in CMap and HCC has 12 functional pathways; \( CS_{ij} \) represents the connectivity score between HCC pathway \( i \) and instance \( j \), which is defined as:

\[
CS_{ij} = CS_{up_{ij}} - CS_{down_{ij}}
\]

(5)

where \( CS_{up_{ij}} \) and \( CS_{down_{ij}} \) represent the connectivity scores for the up-regulated gene set and the down-regulated gene of pathway \( i \) respectively, and their definitions are given in formula (3) to (5).

If the up-regulated pathway genes are near the top (up-regulated) of the rank-ordered drug gene lists and the down-regulated pathway genes are near the bottom (down-regulated) of the rank-ordered drug gene lists, we can get high positive therapeutic scores (TS), which indicate the drugs and the HCC pathways have similar expression profiles and the drugs might aggravate HCC. On the other hand, if the up-regulated pathway genes are near the bottom of the rank-ordered drug gene lists and the down-regulated pathway genes are near the top of the rank-ordered drug gene lists, we can get negative therapeutic scores(TS), which imply the given drugs and the HCC pathways have adverse expression profiles and the drugs could be treatment candidates for HCC. Both cases are significant for HCC treatment.

### 2.4 Datasets

HCC gene expression data. We download the gene expression profiles of HCC from the NCBI Gene Expression Omnibus (GEO)(Edgar et al., 2002). The gene expression dataset (GSE45436) contains 134 samples, including 95 tumor samples and 39 control samples. Each probe is mapped to a gene using GPL570 platform obtained from GEO, while the probes are removed if they do not match any gene. If multiple microarray probes are mapped to a same gene, we assign the average of expression values to the gene.

Here, the expression values of all genes are standardized as follows:

\[
z_{ij} = \frac{g_{ij} - \text{mean}(g_{ij})}{\text{std}(g_{ij})}
\]

(6)

where \( g_{ij} \) represents the expression value of gene \( i \) in sample \( j \), and \( \text{mean}(g_{ij}) \) and \( \text{std}(g_{ij}) \) respectively represent mean and standard deviation of the expression vector for gene \( i \) across all samples.

Gene expression data related to drugs. The gene expression data related to drugs is downloaded from the CMap database (Lamb et al., 2006). It contains 6,100 instances, which cover 1,309 drugs. These instances are measured on five types of human cancer cell lines, including the breast cancer epithelial cell lines MCF7 and soMCF7, the prostate cancer epithelial cell line PC3, the nonepithelial lines HL60 (leukemia) and SKMEL5 (melanoma).

### 3 Results

#### 3.1 Choose Potential HCC Drugs through CTD Benchmark

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

For a drug in CMap, if it cannot find corresponding chemical name in CTD, we will not calculate its therapeutic score (defined in section “Methods”). In this way, we finally get 1168 scored drugs. Because most drug-disease associations in CTD are not marked as positive or negative, we rank the 1168 drugs in descending order by the absolute values of their therapeutic scores (S1 Table). We know the top drugs imply stronger connections with HCC. And then we calculate the precisions of our approach at different top-x drugs, which are shown in Fig 2. The precision is calculated as follows:

\[
\text{precision} = \frac{P_{CTD}}{P_{top}}
\]

(7)

where \( P \) represents the number of top-x drugs; \( P_{CTD} \) represents the number of drugs in the top-x drugs can be found related with HCC in CTD database.

![Fig 2. The precision of our approach at different top-x drugs.](image)

We find in the top-10 drugs (x=10), there are 9 drugs associated with HCC in CTD. That is to say, the precision is 0.9. For the top-20 drugs (x=20), the precision is 0.85 and there are three potentially HCC-related drugs. When x is 30, its precision is 0.8 and we get six potential drugs with HCC. From the Fig 2, we notice that with the increase of x, the precision declines and the number of potential drugs increases. We hope we can predict relatively more HCC-related drugs with high precision. Therefore, we choose top-30 (x = 30) drugs for further analysis.
3.2 Validate potentially HCC-related Drugs through PubMed Literature

In the above section, we choose the top-30 drugs (precision = 0.8) for further analysis. There are 12 therapeutic drugs with negative TS values in the top-30 drugs, shown in Table 4. Nine of them can be found having connections with HCC in CTD. But only one of the 9 drugs is marked as therapeutic drug (Rank=3 and Evidence = "T" in Table 4) for HCC and the other 8 inferred drugs are unmarked in CTD. Here, we can indicate these 8 unmarked drugs are possibly therapeutic drugs for HCC. The rest three drugs (Securinine, Isocarboxazid and Ajmaline) are newly predicted ones by our method, which are marked as bold in Table 4. Based on PubMed, we analyze the three drugs further. PubMed (pubmed.gov) is a free resource developed and maintained by the National Center for Biotechnology Information (NCBI) at the National Library of Medicine (NLM). PubMed comprises more than 26 million references and abstracts on life sciences and biomedical topics.

Table 4. Twelve Therapeutic Drugs for HCC in the Top-30 Drugs.

| Rank | Drug Name          | Evidence | Ref Count |
|------|--------------------|----------|-----------|
| 1    | Geldanamycin       | inferred | 33        |
| 2    | Lomustine          | inferred | 1         |
| 3    | Securinine         | NULL     | NULL      |
| 4    | Clozapine          | inferred | 35        |
| 5    | Suramin            | inferred | 14        |
| 6    | Rosiglitazone      | inferred | 79        |
| 7    | Furazolidone       | inferred | 7         |
| 8    | Hydrocortisone     | inferred | 30        |
| 9    | Isocarboxazid      | NULL     | NULL      |
| 10   | Ajmaline           | NULL     | NULL      |
| 11   | 1-(5-Isoquinolinesulfonfyl)-2-Methylpiperaizine | inferred | 15 |
| 12   | Troglitazone       | T        | 86        |

Evidence represents a drug-disease association is curated, inferred or not existed in CTD database. Curated associations include three types: marker/mechanism (Evidence = "M"), therapeutic (Evidence = "T"), marker/mechanism & therapeutic (Evidence = "M&T"). If an association is inferred by CTD, Evidence = "inferred", and if it is not existed in CTD, Evidence = "NULL"; Ref Count represents the number of reference(s) for the curated and inferred associations. If an association is not existed in CTD, Ref Count = "NULL".

Securinine (Rank=3 in Table 4), a quinolizine pseudoalkaloid (not from amino acid) from securinega suffrutiosa or securinini nitras, is one of central nervous stimulants and clinically applied to treat amyotrophic lateral sclerosis (ALS), poliomyelitis and multiple sclerosis (Buravtseva, 1958; Copperman et al., 1973; Copperman et al. 1974). It has been found to be active as a γ-amino butyric acid (GABA) receptor antagonist. GABA stimulates HCC cell line HepG2 growth. Consequently, it means that securinine is a promising agent with therapeutic effect on HCC.

Isocarboxazid (Rank=9 in Table 4) is a non-selective, irreversible monoamine oxidase inhibitor (MAOI) used as an antidepressant (Fagervall et al., 1986). It has been investigated in the treatment of Parkinson’s disease and other dementia-related disorders. MAOIs inhibit the activity of the monoamine oxidase enzyme family and decrease dopamine production (Goldstein et al., 2016). Dopamine is an organic chemical belong to catecholamine and phenethylamine families and plays several important roles in the brain and body. It is synthesized from L-DOPA, which is produced in the brain and kidneys. Studies have revealed that dopamine receptors (DRs) are over expressed in the HCC (Villa et al., 2005; Lu et al., 2015), and the overexpression of DRs improves the chance combining with dopamine. However, isocarboxazid can decrease this chance by reducing dopamine. Therefore, isocarboxazid has the potential role in the treatment of HCC.

Ajmaline (Rank=10 in Table 4) is a sodium ion channel blocker. It has served as a treatment for Wolff-Parkinson-White syndrome (WPW) which is a disorder of the electrical system of the heart referred to arrhythmia with the ventricles contracting prematurely. Sodium channel blockers are drugs which impair conduction of sodium ions (Na+) through sodium channels. The sustained low extracellular sodium ion concentration decrease glutamate uptake (Fujisawa et al., 2015). Glutamine is an essential compound in cellular metabolism and a fuel for cell proliferation, so the decrease of glutamine will inhibit tumor proliferation (Hensley et al., 1958; Copperman, 1973; Copperman et al., 1974).
Pridinol (Rank=7 in Table 5) is a muscle relaxant that is able to alleviate pain (Merck, 1989). In PubMed, there are only eight articles related to pridinol. Due to lack of information about pridinol, it is hard for us to find evidences to infer that pridinol may aggravate HCC. We are confident that our prediction for pridinol aggravating HCC will be verified in the future.

Piperacillin (Rank=9 in Table 5) is a well-known β-lactam antibiotic (Holten, 2000). β-lactam antibiotics work by inhibiting the biosynthesis of the peptidoglycan layer of bacterial cell walls. The peptidoglycan layer is the outermost and fundamental component of cell walls, and it is crucial for cell wall structural integrity. Blocking the synthesis of the peptidoglycan layer destroys the bacterial cell wall and induces bacterial cell rupture. Lipopolysaccharides (LPS) are released because of bacterial cell rupture (Parija, 2014). Studies demonstrate that LPS promotes proliferation, adhesion and invasion of hepatoma cells (Liu et al., 2014). Therefore, piperacillin probably aggravates HCC indirectly by disturbing the bacterial cell wall.

Ethisterone (Rank=13 in Table 5) is a synthesized progestin. Studies revealed that natural and synthetic progesterin increased the expression of vascular permeability factor (VEGF). VEGF is a signal protein involved in vasculosogenesis and angiogenesis and its principal function is to produce new blood vessels during embryonic development after blood vessels injury. VEGF also is a potent angiogenesis growth factor. The overexpression of VEGF and its receptors has been associated with poor prognosis in HCC (Dhar et al., 2001; Tseng et al., 2008). Moreover, the overexpression of VEGF can contribute to the proliferation of HCC tumor cells. If HCC patients orally take ethisterone, the level of VEGF will increase. The situation of those patients will be more terrible. Hence, ethisterone gets some evil influence over HCC patients (The mechanism depicts in Fig 3B).

### Table 5. Eighteen Aggravating Drugs for HCC in the Top-30 Drugs.

| Rank | Drug Name                | Evidence | References_count |
|------|--------------------------|----------|------------------|
| 1    | Scriptaid                | inferred | 7                |
| 2    | Entinostat               | inferred | 37               |
| 3    | Trichostatin A           | inferred | 89               |
| 4    | Harman                   | inferred | 1                |
| 5    | Estradiol                | inferred | 116              |
| 6    | Riboflavin               | inferred | 18               |
| 7    | Pridinol                 | NULL     | NULL             |
| 8    | Diethylstilbestrol       | inferred | 69               |
| 9    | Pipercillin              | NULL     | NULL             |
| 10   | Nilutamide               | inferred | 2                |
| 11   | N-(2-cyclohexoxy-4-nitrophenyl)-methyleneisufonamide | inferred | 32 |
| 12   | Trioxsalen               | inferred | 1                |
| 13   | Ethisterone              | NULL     | NULL             |
| 14   | Fulvestrant              | inferred | 59               |
| 15   | Acacetin                 | inferred | 1                |
| 16   | Cetirizine               | inferred | 2                |
| 17   | Dinoprost                | inferred | 18               |
| 18   | Haloperidol              | inferred | 33               |

Evidence represents a drug-disease association is curated, inferred or not existed in CTD database. Curated associations include three types: marker/mechanism (Evidence = "CT"), therapeutic (Evidence = "T"), marker/mechanism & therapeutic (Evidence = "M&T"). If an association is inferred by CTD, Evidence = "inferred", and if it is not existed in CTD, Evidence = "NULL". Ref Count represents the number of reference(s) for the curated and inferred associations. If an association is not existed in CTD, Ref Count = "NULL".

### 3.3 Analyze potentially HCC-related Drugs through CMap database

The CMap database can not only be applied to calculate drug-disease correlations, but also can be used to identify connections between two drugs. In particular, for a same disease, if two drugs have strongly positive relationship, they may have similar effects on this disease. On the contrary, if their relationship is negative, they may have opposite effects. In this section, we further analyze the six predicted drugs (three therapeutic drugs shown in Table 4: Securinine, Isocarboxazid and Ajmaline; three aggravating drugs shown in Table 5: Pridinol, Piperacillin and Ethisterone) based on CMap and estimate their correlations (evaluated by formula (5)) with known HCC drugs marked as “therapeutic” in CTD database. The results are shown in Table 6.

For the three potentially therapeutic drugs (Securinine, Isocarboxazid and Ajmaline) marked as bold in Table 6, we find that securinine yields highly positive connectivity score (calculated by formula (5)) for daunorubicin, troglitazone and paclitaxel. Daunorubicin is a chemotherapy medication used to treat cancer. The liposomal formulation of the anthracycline daunorubicin has low systemic toxicity and is taken up strongly by the liver. The researchers found that daunorubicin was an active agent against HCC (Daniele et al., 1999). Troglitazone is a thiazolidinedione PPARγ agonist that exhibits anti-diabetic, anticancer, anti-fibrotic, and anti-inflammatory activities. Studies (Yu et al., 2006) found a significant decrease in expression of PPARγ mRNA and protein in human liver cancer cells compared with surrounding nontumorous liver. Troglitazone induces PPARγ expression and inhibits cell growth both in vitro and in vivo (Yu et al., 2006). Paclitaxel is a cancer medication that interferes with the growth and spread of cancer cells in the body (Huang et al., 2015). Studies indicate that paclitaxel is cytotoxic to cultured hepatocellular carcinoma cells (Gigandeep et al., 1999) and is a common chemodrug for therapy of HCC patients (Yan et al., 2013; Chen et al., 2015; Chao et al., 1998).

Isocarboxazid is found having strongly positive relationships with doxorubicin, sirolimus and troglitazone. Doxorubicin has been regarded as one of the effective chemotherapy medication used to treat cancer (Weiss, 1992). One way that doxorubicin works is by blocking an enzyme called topo isomerase 2 that cancer cells need to divide and grow. Sorafenib, a multikinase inhibitor, has been shown to be effective and safe monotherapy in patients with advanced HCC. A recent phase II study revealed the safety and efficacy of the combined use of sorafenib and doxorubicin-TACE in patients from the Asia-Pacific region with intermediate HCC (Chung et al., 2013). The disease control rate and overall response rate were reported to be 91.2% and 52.4%, respectively. Also, doxorubicin is used to improve the survival rate of unresectable HCC patients (Shang et al., 2016). Sirolimus is a macrolide compound and is indicated for the prevention of organ transplant rejection. The data suggested a beneficial effect of sirolimus immunosuppression on the recurrence of hepatocellular carcinoma after liver transplantation (Hanish and Knechtle, 2011; Chinnakotla et al., 2009). Moreover, researchers found sirolimus could inhibit the growth of human hepatoma cells (Schumacher et al., 2002).
Ajmaline gets high positive connectivity scores with nimesulide, dexamethasone, doxorubicin and paclitaxel. Nimesulide is a cyclooxygenase-2 (COX-2) inhibitor and is used to inhibit the proliferation and promote apoptosis of Hep G2 (a human liver cancer cell line) by up-regulating Smad4 (Yang et al., 2012). Dexamethasone, a type of corticosteroid medication, can restore gluconeogenesis in malignant hepatocytes by bypassing the abnormal regulation of 11β-HSD enzymes, which indicates that targeting altered metabolism in liver cancer could prove useful as a therapeutic strategy for HCC (Ma et al., 2013).

For the three potentially aggravating drugs (Pridinol, Piperacillin and Ethisterone) in Table 3, they all have negative relationship with known HCC drugs. Pridinol has high negative connectivity scores with genistein, nimesulide, troglitazone and sirolimus. Piperacillin have clear negative connection scores with dexamethasone, genistein, paclitaxel and sirolimus. Ethisterone have clear negative connection scores with dexamethasone, genistein, sirolimus and troglitazone. Genisten is one of the most abundant isoflavones in soy that is described as an angiogenesis inhibitor and a phytoestrogen (Walter, 1941). Researcher found genistein and its inhibitor isoflavones in soy that is described as an angiogenesis inhibitor and a phytoestrogen (Walter, 1941). Researchers found genistein and its inhibition of multiple signal transduction pathways could control the invasiveness and metastatic potential of HCC (Wang et al., 2014) and the combination of ATO with genistein would present a promising therapeutic strategy for HCC (Ma et al., 2013). It has been shown that targeting multiple signal transduction pathways could control the invasiveness and metastatic potential of HCC (Wang et al., 2014) and the combination of ATO with genistein would present a promising therapeutic strategy for HCC (Ma et al., 2013).

All in all, these results indicate that our method can find both therapeutic drugs and aggravating drugs for diseases. Moreover, our method can be applied to mark the known drug-disease associations as positive or negative, such as some drug-disease associations in CTD database, to provide a better basis for the treatment of diseases.

### 3.4 Pathway Enrichment Analysis

Furthermore, we try to apply DAVID functional annotation tool (Huang et al., 2009) to identify enriched KEGG pathways of the above six drugs based on their targets. If the targets of a drug have close correlations with HCC-related pathways, it is very likely that the drug is helpful for the treatment of HCC. The pathway details are shown in Table 7. Because securinine, pridinol and ethisterone have no target information at present in DrugBank, they have no related KEGG pathways.

The potentially therapeutic drugs of HCC are marked as bold. The other three drugs are potentially aggravating drugs of HCC: “NULL” represents the drug has no targets in DrugBank at present. Thus, its corresponding KEGG pathway is “NULL” too.

Isocarboxazid has two targets “MAOA and MAOB” and they relate with 13 KEGG pathways. Some of them have been reported having close relationships with liver cirrhosis or HCC. For example, for the alcoholism pathway, the researchers found chronic alcohol consumption had long been associated with progressive liver disease toward the development of hepatic cirrhosis and the subsequent increased risk for developing HCC (McKillop and Schrum, 2009). Metabolic pathways are linked series of chemical reactions occurring within a cell. The results showed that the metabolism in the HCC tumor was modified to promote cellular proliferation or escape from apoptosis (Huang et al., 2013). For the drug metabolism-cytochrome P450 pathway, cytochrome P450 1A1 is a major enzyme in the bioactivation of exogenous procarcinogens of hepatocellular carcinoma (HCC) (Li et al., 2009). For the drug ajmaline, it has one enriched pathway, adrenergic signaling in cardiomyocytes. The adrenergic signaling pathway plays an important role in cancer progression by regulating multiple cellular processes (Cole and Sood, 2012; Fitzgerald, 2012; Fitzgerald, 2009).

Table 6. The relationships of six predicted drugs with known HCC therapeutic drugs in CTD.

| Predicted Drugs | Known HCC Drugs in CTD | Connectivity Scores |
|-----------------|------------------------|---------------------|
| Securinine      | Doxorubicin            | 0.397               |
|                 | Troglitazone           | 0.504               |
|                 | Paclitaxel             | 0.321               |
| Isocarboxazid   | Doxorubicin            | 0.31                |
|                 | Sirolimus              | 0.432               |
|                 | Troglitazone           | 0.436               |
|                 | Nimesulide             | 0.542               |
| Ajmaline        | Dexamethasone          | 0.41                |
|                 | Doxorubicin            | 0.36                |
|                 | Paclitaxel             | 0.424               |
| Pridinol        | Genistein              | -0.377              |
|                 | Nimesulide             | -0.269              |
|                 | Troglitazone           | -0.442              |
|                 | Sirolimus              | -0.382              |
| Piperacillin    | Dexamethasone          | -0.436              |
|                 | Genistein              | -0.336              |
|                 | Paclitaxel             | -0.299              |
|                 | Sirolimus              | -0.408              |
| Ethisterone     | Dexamethasone          | -0.369              |
|                 | Genistein              | -0.223              |
|                 | Sirolimus              | -0.236              |
|                 | Troglitazone           | -0.267              |

The potentially therapeutic drugs of HCC are marked as bold. The other three drugs are potentially aggravating drugs of HCC.

Table 7. Pathway enrichment analysis result of six selected drugs.

| Drug Name | Drug Targets | KEGG Pathways |
|-----------|--------------|---------------|
| Securinine| NULL         | Phenylalanine metabolism; Histidine metabolism; Tyrosine metabolism; Glycine, serine and threonine metabolism; Tryptophan metabolism; Arginine and proline metabolism; Aromatic amino acid metabolism; Tyrosine, tryptophan and aromatic amino acid metabolism; Hypertrophic cardiac myocyte; Rho signaling pathway; VEGF signaling pathway; Adipocytokine signaling pathway |
| Isocarboxazid| MAOA, MAOB | Cocaine addiction; Arginine and proline metabolism; Amphetamine addiction; Drug metabolism - cytochrome P450; Serotoninergic synapse; Dopaminergic synapse; Alcoholism; Metabolic pathways |
| Ajmaline  | SCN5A        | Adrenergic signaling in cardiomyocytes |
| Pridinol  | NULL         | NULL |
| Piperacillin| PBP3, PENA, PBP2A, PBP1B | Peptidoglycan biosynthesis |
| Ethisterone| NULL         | NULL |

There are nine of thirteen pathways in which isocarboxazid drug targets exist and these nine pathways have common genes with the 12 tissue-specific KEGG pathway of HCC (Table 3). The interaction between the nine pathways and the 12 tissue-specific KEGG pathway of HCC is shown in
Fig. 4. Tyrosine metabolism has common genes with one pathways which are tissue-specific KEGG pathway of HCC. There are common genes between Cocaine addiction and seven HCC related pathways. For Arginine and proline metabolism, it exists common genes with one HCC related pathway. There are common genes between Amphetamine addiction and seven HCC related pathways. Drug metabolism-cytochrome P450 has common genes with two HCC pathways. Serotonergic synapse has common genes with ten pathways of liver cancer. Dopaminergic synapse has common genes with nine pathways of liver cancer. There are common genes between Alcoholism and five HCC related pathways. For Metabolic pathways, it exists common genes with seven HCC related pathway.

Figure 4. Pathway analysis of isocarboxazid drug targets. The purple circle represents pathways of the mercaptopurine drug targets appeared, and the green circle represents tissue-specific KEGG pathway of HCC. The gray side indicates that there is a common gene between two pathways, and the more number of public genes is, the more rough edge is. There is only one pathway in which Ajmaline drug targets exist. However, Adrenergic signaling in cardiomyocytes has common genes with ten of 12 tissue-specific KEGG pathway of HCC (Table 3). The interaction between the six pathways and the 12 tissue-specific KEGG pathways of HCC is shown in Fig. 5.

These results further show that these potential drugs have close relationships with HCC. With the improvement of data, our algorithm will perform better.

Figure 5. Pathway analysis of reserpine drug targets. The purple circle represents pathways of the reserpine drug targets appeared, and the green circle represents tissue-specific KEGG pathway of HCC. The gray side indicates that there is a common gene between two pathways, and the more number of public genes is, the more rough edge is.

4 Conclusions

Hepatocellular carcinoma (HCC), also called malignant hepatoma, is the most common type of liver cancer. In this study, we propose an approach to identify potential drugs for treatments of HCC, which takes into account the special relationships among HCC, liver tissue and KEGG pathways. First, we map genes related to HCC to the tissue-specific protein interaction network of liver and get an extended gene set of HCC. Then, based on the new tissue-specific gene set of HCC, we extract 12 enriched functional pathways by KEGG pathway enrichment analysis. The 12 KEGG pathways are used to calculate the therapeutic score (TS) of each drug using Kolmogorov–Smirnov statistic. High positive therapeutic scores (TS) indicate the drugs and the HCC pathways have similar expression profiles and the drugs might aggravate HCC. On the contrary, negative therapeutic scores (TS) imply the given drugs and the HCC pathways have adverse expression profiles and the drugs could be treatment candidates for HCC. That is to say, our method can clearly distinguish therapeutic and aggravating drugs for HCC, which make it possible for providing a more accurate reference for the treatment of HCC. Furthermore, it is helpful in marking some known associations as therapeutic or aggravating between drugs and diseases in CTD database. Finally, we choose top-six potentially related drugs of HCC for further analysis, including three therapeutic drugs and three aggravating drugs. They are validated by analyzing their overlaps with drug indications reported in CTD benchmark and PubMed literature. Furthermore, CMap profiles similarity analysis and KEGG enrichment analysis are also made on their related genes. These findings suggest that our approach is more useful and effective in predicting novel therapeutic options for HCC, and it can be easily extended to other diseases. The performance of our proposed method can be further improved by integrating more pathway information and tissue data, which will help predicting novel drugs for diseases in a more accurate way.

References

K.Takahashi et al.

Fig.4. Tyrosine metabolism has common genes with one pathways which are tissue-specific KEGG pathway of HCC. There are common genes between Cocaine addiction and seven HCC related pathways. For Arginine and proline metabolism, it exists common genes with one HCC related pathway. There are common genes between Amphetamine addiction and seven HCC related pathways. Drug metabolism-cytochrome P450 has common genes with two HCC pathways. Serotonergic synapse has common genes with ten pathways of liver cancer. Dopaminergic synapse has common genes with nine pathways of liver cancer. There are common genes between Alcoholism and five HCC related pathways. For Metabolic pathways, it exists common genes with seven HCC related pathway.

Figure 4. Pathway analysis of isocarboxazid drug targets. The purple circle represents pathways of the mercaptopurine drug targets appeared, and the green circle represents tissue-specific KEGG pathway of HCC. The gray side indicates that there is a common gene between two pathways, and the more number of public genes is, the more rough edge is. There is only one pathway in which Ajmaline drug targets exist. However, Adrenergic signaling in cardiomyocytes has common genes with ten of 12 tissue-specific KEGG pathway of HCC (Table 3). The interaction between the six pathways and the 12 tissue-specific KEGG pathways of HCC is shown in Fig. 5.

These results further show that these potential drugs have close relationships with HCC. With the improvement of data, our algorithm will perform better.

Figure 5. Pathway analysis of reserpine drug targets. The purple circle represents pathways of the reserpine drug targets appeared, and the green circle represents tissue-specific KEGG pathway of HCC. The gray side indicates that there is a common gene between two pathways, and the more number of public genes is, the more rough edge is.

4 Conclusions

Hepatocellular carcinoma (HCC), also called malignant hepatoma, is the most common type of liver cancer. In this study, we propose an approach to identify potential drugs for treatments of HCC, which takes into account the special relationships among HCC, liver tissue and KEGG pathways. First, we map genes related to HCC to the tissue-specific protein interaction network of liver and get an extended gene set of HCC. Then, based on the new tissue-specific gene set of HCC, we extract 12 enriched functional pathways by KEGG pathway enrichment analysis. The 12 KEGG pathways are used to calculate the therapeutic score (TS) of each drug using Kolmogorov–Smirnov statistic. High positive therapeutic scores (TS) indicate the drugs and the HCC pathways have similar expression profiles and the drugs might aggravate HCC. On the contrary, negative therapeutic scores (TS) imply the given drugs and the HCC pathways have adverse expression profiles and the drugs could be treatment candidates for HCC. That is to say, our method can clearly distinguish therapeutic and aggravating drugs for HCC, which make it possible for providing a more accurate reference for the treatment of HCC. Furthermore, it is helpful in marking some known associations as therapeutic or aggravating between drugs and diseases in CTD database. Finally, we choose top-six potentially related drugs of HCC for further analysis, including three therapeutic drugs and three aggravating drugs. They are validated by analyzing their overlaps with drug indications reported in CTD benchmark and PubMed literature. Furthermore, CMap profiles similarity analysis and KEGG enrichment analysis are also made on their related genes. These findings suggest that our approach is more useful and effective in predicting novel therapeutic options for HCC, and it can be easily extended to other diseases. The performance of our proposed method can be further improved by integrating more pathway information and tissue data, which will help predicting novel drugs for diseases in a more accurate way.

References

K.Takahashi et al.

Fig.4. Tyrosine metabolism has common genes with one pathways which are tissue-specific KEGG pathway of HCC. There are common genes between Cocaine addiction and seven HCC related pathways. For Arginine and proline metabolism, it exists common genes with one HCC related pathway. There are common genes between Amphetamine addiction and seven HCC related pathways. Drug metabolism-cytochrome P450 has common genes with two HCC pathways. Serotonergic synapse has common genes with ten pathways of liver cancer. Dopaminergic synapse has common genes with nine pathways of liver cancer. There are common genes between Alcoholism and five HCC related pathways. For Metabolic pathways, it exists common genes with seven HCC related pathway.

Figure 4. Pathway analysis of isocarboxazid drug targets. The purple circle represents pathways of the mercaptopurine drug targets appeared, and the green circle represents tissue-specific KEGG pathway of HCC. The gray side indicates that there is a common gene between two pathways, and the more number of public genes is, the more rough edge is. There is only one pathway in which Ajmaline drug targets exist. However, Adrenergic signaling in cardiomyocytes has common genes with ten of 12 tissue-specific KEGG pathway of HCC (Table 3). The interaction between the six pathways and the 12 tissue-specific KEGG pathways of HCC is shown in Fig. 5.

These results further show that these potential drugs have close relationships with HCC. With the improvement of data, our algorithm will perform better.

Figure 5. Pathway analysis of reserpine drug targets. The purple circle represents pathways of the reserpine drug targets appeared, and the green circle represents tissue-specific KEGG pathway of HCC. The gray side indicates that there is a common gene between two pathways, and the more number of public genes is, the more rough edge is.
K.Takahashi et al.

Weiss, R. B. (1992). The anthracyclines: will we ever find a better doxorubicin?. Seminars in Oncology, 19(6), 670.

Yan, H., Wang, S., Yu, H., Zhu, J., & Chen, C. (2013). Molecular pathways and functional analysis of mirna expression associated with paclitaxel-induced apoptosis in hepatocellular carcinoma cells. Pharmacology, 92(3-4), 167.

Yang, S., Guo, R., Huang, L., Yang, L., & Jiang, B. (2012). Nimesulide inhibits the proliferation of HepG2 by up-regulation of Smad4. Indian journal of pharmacology, 44(5), 599.

Ye, H., Yang, L., Cao, Z., Tang, K., & Li, Y. (2012). A pathway profile-based method for drug repositioning. Chinese Science Bulletin, 57(17), 2106-2112.

Yu, J., Qiao, L., Zimmermann, L., Ebert, M. P., Zhang, H., & Lin, W., et al. (2006). Troglitazone inhibits tumor growth in hepatocellular carcinoma in vitro and in vivo. Digest of the World Core Medical Journals, 43(1), 134-143.

Zhang, T., Hu, Y., Wu, X., Ma, R., Jiang, Q., & Wang, Y. (2016). Identifying Liver Cancer-Related Enhancer SNPs by Integrating GWAS and Histone Modification ChIP-seq Data. BioMed Research International, 2016.

Zhao, H., Jin, G., Cui, K., Ren, D., Liu, T., Chen, P., ... & Chang, J. (2013). Novel modeling of cancer cell signaling pathways enables systematic drug repositioning for distinct breast cancer metastases. Cancer research, 73(20), 6149-6163.