Global view of a drug-sensitivity gene network

Haixiu Yang[1,*], Yunpeng Zhang[1,*], Jiasheng Wang[1], Tan Wu[1], Siyao Liu[1], Yanjun Xu[1] and Desi Shang[1]

[1]College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150081, China

[*]These authors contributed equally to this work

Correspondence to: Desi Shang, **email**: sds_46@163.com
Yanjun Xu, **email**: tonghua605@163.com

ABSTRACT

An important challenge in drug development is to gain insight into the mechanism of drug sensitivity. Looking for insights into the global relationships between drugs and their sensitivity genes would be expected to reveal mechanism of drug sensitivity. Here we constructed a drug-sensitivity gene network (DSGN) based on the relationships between drugs and their sensitivity genes, using drug screened genomic data from the NCI-60 cell line panel, including 181 drugs and 1057 sensitivity genes, and 1646 associations between them. Through network analysis, we found that two drugs that share the same sensitivity genes tend to share the same Anatomical Therapeutic Chemical classification and side effects. We then found that the sensitivity genes of same drugs tend to cluster together in the human interactome and participate in the same biological function modules (pathways). Finally, we noticed that the sensitivity genes and target genes of the same drug have a significant dense distance in the human interactome network and they were functionally related. For example, target genes such as epidermal growth factor receptor gene can activate downstream sensitivity genes of the same drug in the PI3K/Akt pathway. Thus, the DSGN would provide great insights into the mechanism of drug sensitivity.

INTRODUCTION

There is compelling evidence that predictions of anticancer drug response can be refined by identifying and applying molecular biomarkers [1, 2]. For example, the use of drugs to target the protein product of the *BCR–ABL* translocation in chronic myeloid leukemia, or the *BRAF* gene in malignant melanoma, helped transform the treatment of these diseases and substantially improve survival rates [3, 4]. In recent years, enormous efforts have been made to identify predictive biomarkers of drug response. For example, Lindsay *et al.* provided non-linear machine learning techniques, and generated biomarkers that predict drug response [5]. David *et al.* developed a novel approach named Multivariate Organization of Combinatorial Alterations (MOCA), combining many genomic alterations into biomarkers of drug response,

and found that multi-gene features have substantially higher correlation with drug response than do single-gene features [6]. It follows that methods considering the cumulative effect of many markers would make the prediction of complex phenotypes (such as drug response) more accurate [2, 7].

Based on the fact that many genes may be regarded as genomic biomarkers for drug response and one genomic biomarker may be correlated to sensitivity toward many drugs [6], a large-scale network correlating drugs and their sensitivity genes should be constructed, as it would give global clues to possible biomarker-related treatments of drug response, and network analysis would be helpful in elucidating the action of drug sensitivity.

However, it would be difficult to construct such a global drug sensitivity gene network via low-throughput biological experimental studies. One of the major concerns

is that gene expression estimates, generated on different microarray platforms or even in different batches, are not always consistent, leading to irreproducible data [8]. Another factor is that publicly available data on gene expression related to drug response are relatively limited. Fortunately, these limitations could be alleviated to a great extent by the development of high-throughput experimental and bioinformatics technologies. The NCI-60 cell line panel and associated drug screens were used to pioneer the approach of linking drug sensitivity to genomic data [9]. Meanwhile, with the development of CellMiner, rapid data retrieval of genomic data along with activity reports for ~20,000 chemical compounds across the NCI-60 was allowed [10]. Hence, we can acquire genomic data related to drug sensitivity, and build the relationship between the sensitivity genes and drug response.

In this study, we constructed a global drug-sensitivity gene network (DSGN) in which nodes represent drugs or sensitivity genes, and these are connected if the genes are related to anticancer sensitivity of the corresponding drug. We then did a series analysis of the global relationships between drugs and sensitivity genes, including the basic properties of the DSGN, shortest path analysis, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of the sensitivity genes, Anatomical Therapeutic Chemical (ATC) codes and side effects of the drugs. Through these analyses, our findings offered insight into the interplay between drugs and sensitivity genes.

RESULTS

Construction of the drug-sensitivity gene network (DSGN)

We constructed a bipartite network consisting of two disjoint kinds of nodes. One kind of node corresponded to drugs tested in the National Cancer Institute (NCI) Developmental Therapeutics Program (DTP), and the other kind of node comprised the sensitivity genes from the profiles for the 60 cell lines of the NCI DTP drug screen [11]. A drug and a gene were connected if the gene was related to the anticancer sensitivity of the corresponding drug. To obtain the relationships between them, we used CellMiner Analysis tools (https://discover.nci.nih.gov/cellminer/) to retrieve potential associations between a drug and its sensitivity genes by calculating the Pearson correlation coefficient (PCC). Setting $|PCC| \geq 0.5$ and p-value < 0.01, we obtained 16,694 significant drug-gene correlations, including 6477 genes and 234 drugs. Each drug was associated with up to 27.6 sensitivity genes on average. To reduce the false positive results and obtain the more significant drug-sensitivity gene relationships, we ranked the sensitivity genes for each drug according to the absolute PCC value, and retained the top 10% of sensitivity genes (see Materials and Methods). Finally, the

DSGN was composed of 1646 drug-gene pairs, including 1057 genes and 181 drugs that were grouped into 12 drug classes using the ATC classification system in the DSGN (see Figure 1).

Properties and functional mapping of the DSGN

The DSGN was composed of 1238 nodes (1057 sensitivity genes and 181 drugs) and 1646 edges (Figure 1; Supplementary Dataset 1). We examined the DSGN, and its network topological characteristics are displayed in Supplementary Figure 1. The degree distributions of the drug and gene nodes both followed power law distributions approximately with slopes of −5.45 and −6.24, respectively, and R^2 = 0.27 and 0.35, respectively (Supplementary Figure 1A, 1B). Thus, the DSGN was scale-free. The degree of genes was distributed from 1 to 19. Schlafen family member 11 (*SLFN11*), with the highest degree of 19, was recently discovered as a dominant response factor of cancer cells to topoisomerase I inhibitors [12, 13]. Knockdown of SLFN11 increases chemoresistance of cancer cells to a broad range of DNA damaging agents [12, 14], and ectopic expression of SLFN11 sensitizes colon cancer cells to topoisomerase I inhibitors [15], consistent with the involvement of SLFN11 in the DNA damage response [12]. The gene with the second highest degree (12) was Src-like-adaptor (SLA), which is expressed in a variety of cell types, and it can both inhibit and activate signaling downstream of various cell surface receptors including the B cell receptor, the T cell receptor, cytokine receptors and receptor tyrosine kinases, which are important regulators of immune and cancer cell signaling [16]. SLA protein appears to be an important component in regulating signal transduction required by immune and malignant cells [16], so it was also relevant for many anticancer drugs. Meanwhile, the degree of drugs was distributed from 1 to 69, and the drug Vemurafenib, which is a B-Raf enzyme inhibitor developed for the treatment of late-stage melanoma [17], had the highest degree. Other drugs with high degrees included Zalcitabine with a degree of 68, which is a potent inhibitor of HIV replication at low concentrations, acting as a chain-terminator of viral DNA by binding to reverse transcriptase [18], and Artemether with a degree of 63, which is an antimalarial agent used to treat acute uncomplicated malaria [19].

To investigate the biological functions of drug sensitivity genes in the DSGN, we employed functional analysis of the corresponding genes. We implemented KEGG pathway enrichment by using DAVID [20], and found 24 pathways were significantly enriched (p-value < 0.01) (Table 1, Figure 2A). The hematopoietic cell lineage (hsa04640) pathway was the most significantly enriched pathway, and it is significantly associated with pediatric acute lymphoblastic leukemia [21]. The primary immunodeficiency (hsa05340) pathway was also a
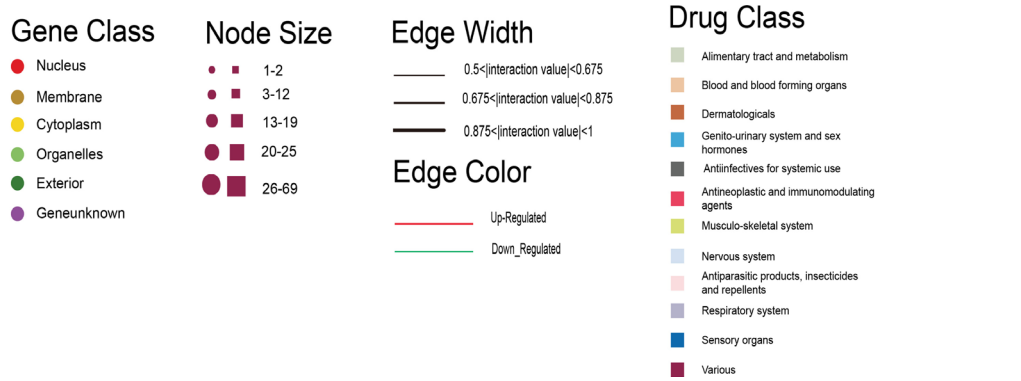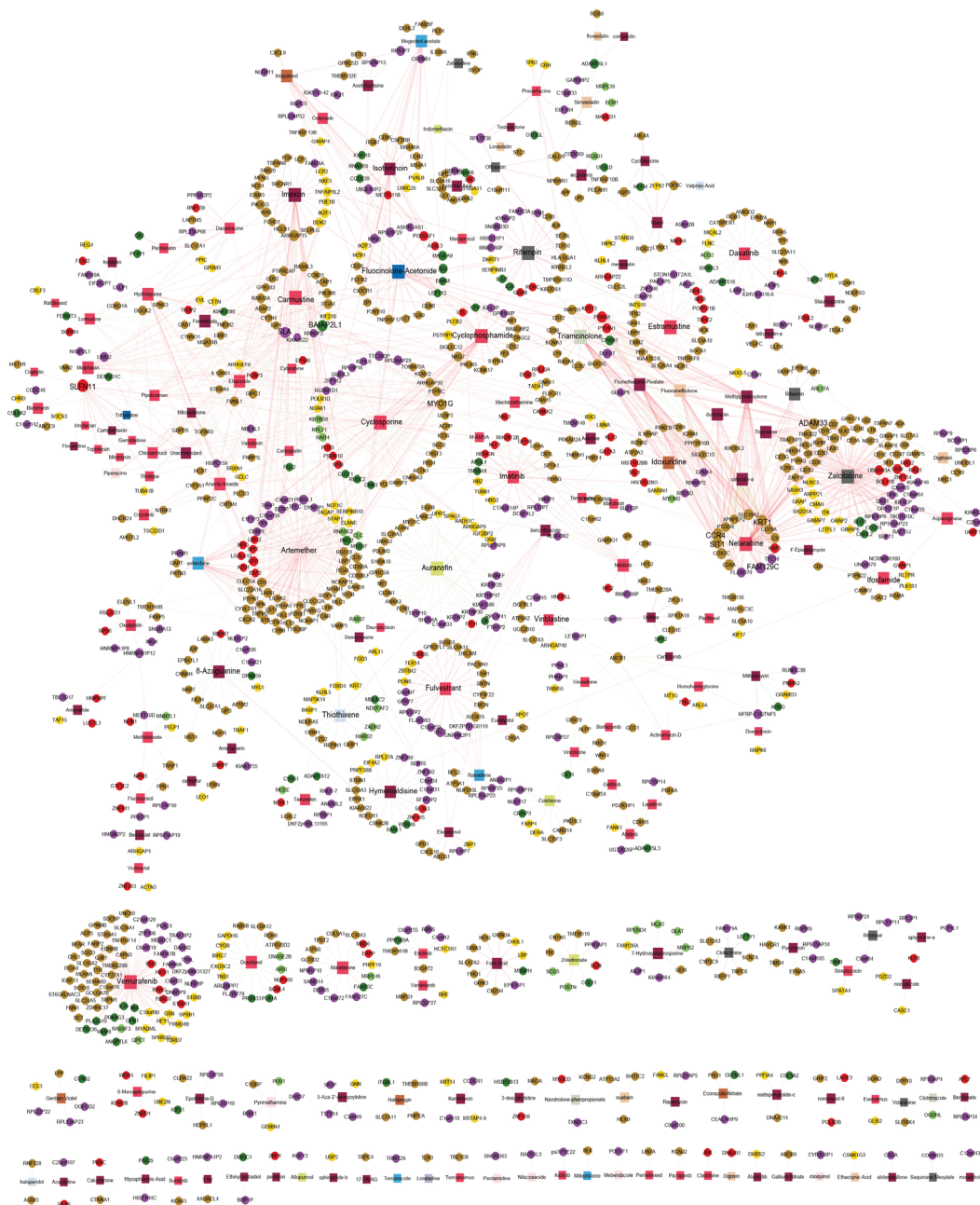
Figure 1: The DSGN network. The rectangles and circles in the network correspond to drugs and sensitivity genes, respectively. A drug and a gene are connected by an edge if the gene is related to the anticancer sensitivity of the corresponding drug. Different colors of gene and drug nodes represent subcellular localization class of genes and ATC classification of drugs, respectively. The node size represents the degree of node.

Table 1: Significantly enriched pathways

	Term	*P* Value	FDR
1	hsa04640:Hematopoietic cell lineage	2.37E-10	2.86E-07
2	hsa05340:Primary immunodeficiency	7.48E-10	9.03E-07
3	hsa04514:Cell adhesion molecules (CAMs)	6.22E-06	0.007512
4	hsa04670:Leukocyte transendothelial migration	3.85E-05	0.046522
5	hsa04060:Cytokine-cytokine receptor interaction	8.34E-05	0.100689
6	hsa04510:Focal adhesion	1.28E-04	0.154968
7	hsa04672:Intestinal immune network for IgA production	1.78E-04	0.214605
8	hsa04660:T cell receptor signaling pathway	3.70E-04	0.445435
9	hsa04512:ECM-receptor interaction	0.002096	2.501833
10	hsa04650:Natural killer cell mediated cytotoxicity	0.003898	4.606564
11	hsa04062:Chemokine signaling pathway	0.00668	7.773949
12	hsa05332:Graft-versus-host disease	0.009976	11.40227
13	hsa04810:Regulation of actin cytoskeleton	0.015899	17.59498
14	hsa05412:Arrhythmogenic right ventricular cardiomyopathy (ARVC)	0.020163	21.80382
15	hsa04530:Tight junction	0.02086	22.47301
16	hsa05322:Systemic lupus erythematosus	0.021074	22.67742
17	hsa04520:Adherens junction	0.021923	23.48353
18	hsa05330:Allograft rejection	0.023549	25.00512
19	hsa05222:Small cell lung cancer	0.037487	36.95679
20	hsa04940:Type I diabetes mellitus	0.046316	43.59542
21	hsa04666:Fc gamma R-mediated phagocytosis	0.075082	61.03248
22	hsa05416:Viral myocarditis	0.076974	61.98434
23	hsa03010:Ribosome	0.096177	70.50742
24	hsa05320:Autoimmune thyroid disease	0.099817	71.91028

significantly enriched pathway, which might be associated with acute myeloid leukemia (AML) development [22].

We then characterized the sensitivity genes in the DSGN by categorizing them into six groups according to subcellular localization: membrane, cytoplasm, organelles, nucleus, exterior and genes that did not belong to any group above (unknown). Figure 2B shows the distribution of these sensitivity genes among the six groups. We found that up to 472 (44.65%) genes belonged to the membrane group, which indicates that these genes are likely to participate in the processes of cell membrane function, such as the transport process and membrane receptor recognition process.

We also compared the 1057 sensitivity genes with each of the following gene sets: 571 Cancer Gene Census genes (CGC), 2712 essential genes (Essential), 547 adverse drug reaction-associated proteins (ADRPs), 11,700 Mendelian disease genes (MDGs), 2123 orphan disease-causing mutant genes (ODMGs) and 2354 drug target genes (Target). Figure 2C and 2D show the overlaps between the sensitivity genes and the other six gene sets. Although the number of overlaps is small, we still found that sensitivity genes significantly overlapped with CGC (p = 0.00546), Essential (p = 7.9e-10), ADRPs (p = 0.002763) and ODMGs (p = 9.792813e-10) (hyper geometric distribution test), indicating that many sensitivity genes are possibly cancer genes, essential genes, adverse drug reaction-associated proteins and orphan disease-causing mutant genes.

To further investigate the drugs and sensitivity genes in the DSGN, we generated two biologically relevant network projections, a "drugs–drug network" (DDN) and a "sensitivity gene network" (SGN) from the DSGN bipartite network.

Functional characteristics of the DDN

In the DDN (composed of 125 nodes and 564 edges), the nodes represented the drugs, and two drugs were connected to each other if they were connected to at least one common gene in the DSGN (Supplementary Figure 2). Some researchers have indicated that drugs that bind to similar proteins tend to have similar effects and pharmacological properties [23–25]. Thus arose the question whether, if two drugs share the same sensitivity

genes, they tend to cause the same therapeutic or side effects.

To investigate this, we obtained the ATC classifications of 125 drugs and found that 205 drug-drug pairs out of 564 edges in the DDN belonged to the same ATC classification. We then generated 564 randomized drug pairs 1000 times. We found that there were no times when the number of randomized drug pairs that shared the same ATC classification was more than 205, suggesting that connected pairs tend to share the same ATC classification (p-value < 0.001, Figure 3A). Taking Chlorambucil and Carmustine for example, Chlorambucil is a nitrogen mustard alkylating agent used as an antineoplastic agent for the treatment of various malignant and nonmalignant diseases, while Carmustine is a cell-cycle phase nonspecific alkylating antineoplastic agent used for the treatment of brain tumors, multiple myeloma, Hodgkin's disease and non-Hodgkin's lymphomas. The two drugs were connected by some sensitivity genes, and they belong to the same ATC code L (antineoplastic and immunomodulating agents).

Then we downloaded the public and accurate side effect records from the SIDER database that includes 997 drugs corresponding to 4492 side effects. In the DDN, there were 71 drugs that were also recorded in the SIDER database and these drugs formed 110 unique connected drug pairs that share the same side effect (Figure 3B). In the SIDER database, some side effects, such as dizziness and nausea, were caused by most drugs. To improve the specificity of the similarity of drug pairs, we calculated the number of side effects shared by drug pairs rather than the number of drug pairs that shared the same side effects. We
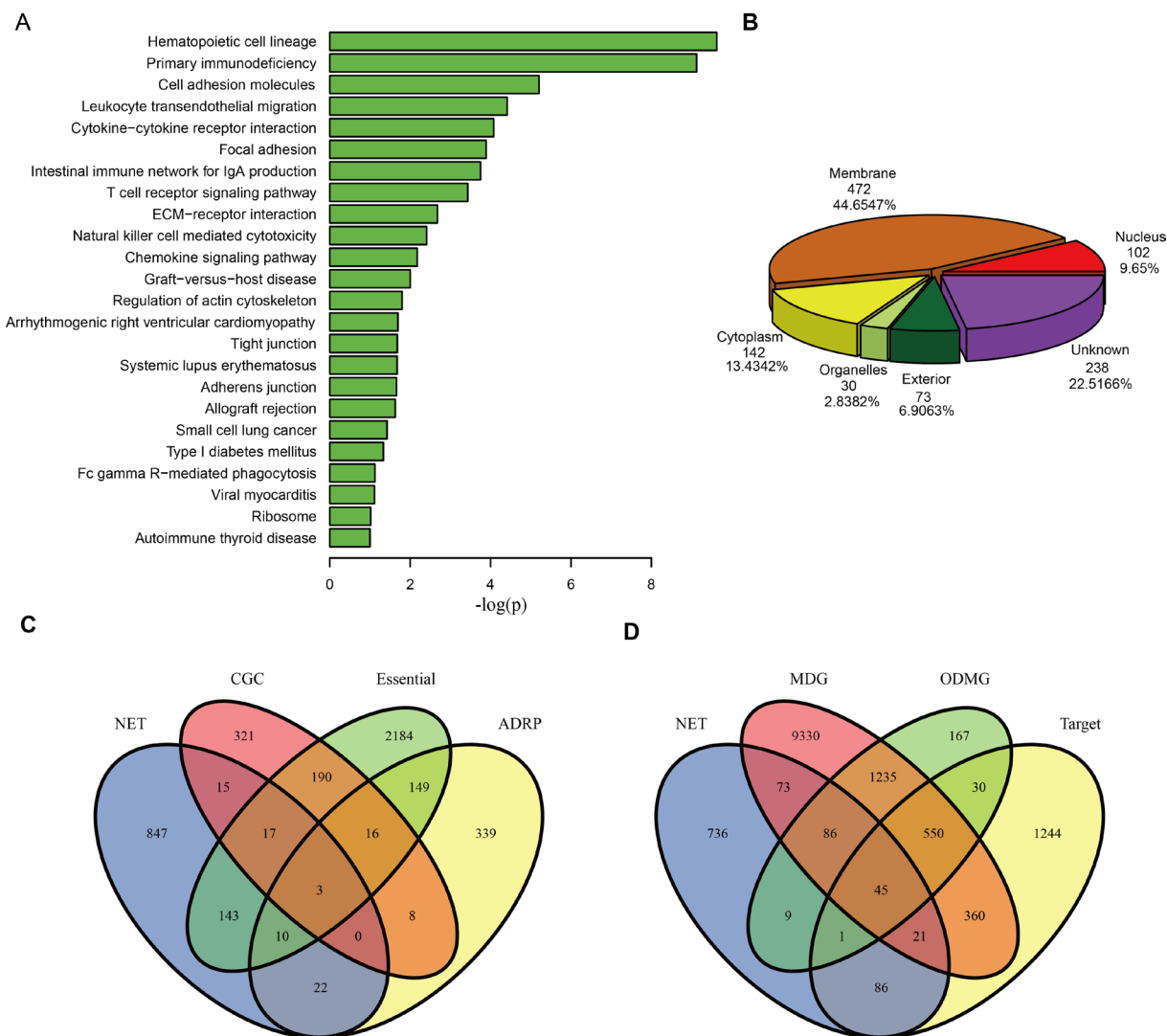

Figure 2: Functional annotations of the drug sensitivity genes in DSGN. (**A**) KEGG pathways enriched by sensitivity genes. (**B**) Pie chart of 1057 sensitivity genes categorized into six groups according to subcellular localization: membrane, cytoplasm, organelles, nucleus, exterior and genes that did not belong to any group above (unknown). (**C**) The Venn diagram of overlaps among the 1057 sensitivity genes (NET), 571 Cancer Gene Census genes (CGC), 2712 essential genes (Essential) and 547 adverse drug reaction-associated proteins (ADRP). (**D**) The Venn diagram of overlaps among the 1057 sensitivity genes (NET), 11,700 Mendelian disease genes (MDG), 2123 orphan disease-causing mutant genes (ODMG) and 2354 drug target genes (Target).
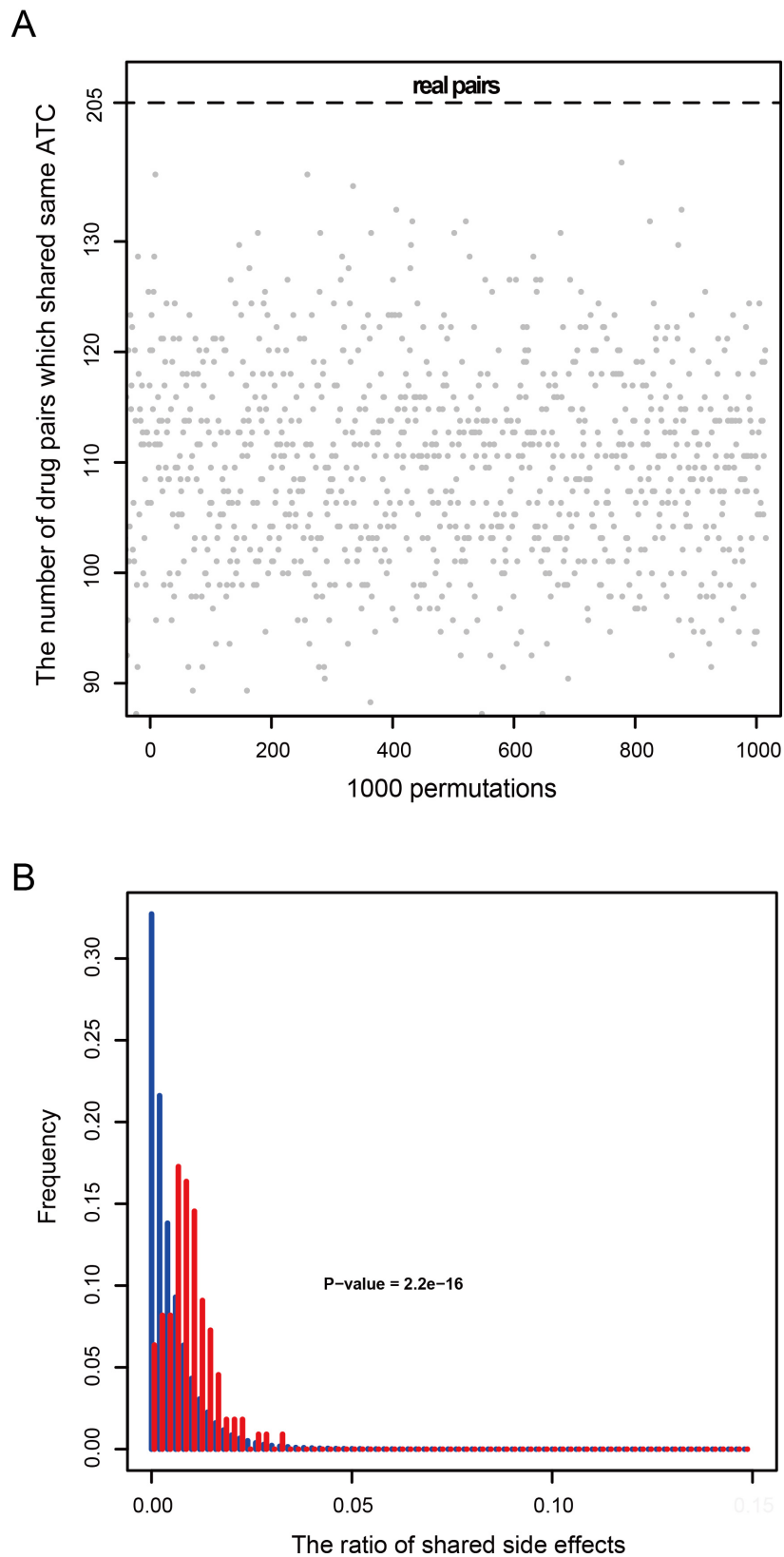
Figure 3: Characteristics of drug pairs that have the same sensitivity genes. (**A**) 205 drug pairs that have the same sensitivity genes share the same ATC classification, compared to 1000 times permutations. (**B**) The proportion of shared side effects by drug pairs that have the same sensitivity genes (red), compared to the proportion of shared side effects among the total drug pairs in the SIDER database (blue).

found that the number of side effects shared by connected drug pairs of the DDN was significantly higher than the number of side effects shared by total drug pairs in the SIDER database (p-value $= 2.2^*$e-16, Wilcoxon rank sum test). These results suggested that two drugs sharing the same sensitivity genes in the DDN tend to cause the same side effects.

Functional characteristics of the SGN

We also constructed the (SGN), in which two genes are related by anticancer sensitivity toward the same drug in the DSGN. SGN was composed of 1033 nodes and 15,224 edges (Supplementary Figure 3).

Yildirim *et al*. indicated that targets of a single drug tend to be in the same module of a network of physical protein-protein interactions (PPIs) and participate in the same molecular complex or cellular pathway [26]. Thus arose the problem whether the sensitivity genes of the same drug also display this feature. First, we investigated the distribution of sensitivity genes belonging to the same drug by calculating the shortest path between sensitivity genes in the Human Protein Reference Database (HPRD) [27] and STRING [28] network using the classical Dijkstra algorithm. We observed a strong enrichment in the regions of first, second and third neighbors compared with the distribution of the shortest path between all nodes in the HPRD and STRING PPI network (Figure 4A, 4B), showing a bias toward clustering of the sensitivity genes of the same drug in the SGN network.

We then investigated whether the sensitivity gene pairs of the same drug tend to participate in the same pathway. We found that there were 253 connected gene pairs in the SGN engaging in 39 pathways. We then randomized the gene pairs 1000 times, and we found that there were no times when the number of randomized gene pairs that engaged in the same pathway was more than 253, suggesting that connected pairs tend to participate in the same pathway (p-value < 0.001, Figure 4C).

Correlation between a sensitivity gene and drug target gene in the context of the cellular network

Tumor-targeted delivery of compounds to the site of malignancy allows for enhanced cellular uptake and increased therapeutic effects, so the development of targeted delivery greatly improved anticancer therapeutic research [29]. Meanwhile, molecular biomarkers have been widely used to predict anticancer drug response. Thus arose the question whether the sensitivity genes and drug target genes are functionally related, and we explored the mechanism of action of sensitivity genes by investigating the relationships between sensitivity genes and target genes. To further investigate the functions of sensitivity genes, we also analyzed the relationships between the sensitivity genes and the target genes of the

same drug based on the PPI network. We obtained the drug targets from the DrugBank database (http://www.drugbank.ca), which is a richly annotated database of drug and drug target information [30]. All of the sensitivity genes and target genes were mapped into the PPI network (HPRD and STRING). In the PPI network, we calculated the shortest path between a sensitivity gene and target gene for each drug, and we found that the shortest paths followed a normal distribution, the shortest paths being between 0 and 10 (0 represented the sensitivity gene and the target gene being the same) and the average of the shortest paths being 4.04 for HPRD (4.07 for STRING). Then we compared the distribution of the above shortest paths with that of all nodes in the HRPD (or STRING). We found that the distribution of the shortest paths between all nodes in the PPI network was significantly larger than the distribution of the shortest paths between sensitivity genes and target genes for the same drug (Figure 5A, 5B). Thus we concluded that the sensitivity gene and the target gene of the same drug are densely connected in the PPI network, which indicates that they are functionally related. Besides, we found 103 drugs for which the shortest path between the sensitivity gene and the target gene was less than 4 in the HPRD network and 91 drugs in the STRING network, of which 78 drugs overlapped between the two PPI networks (Figure 5C), which further indicated the robustness of our results.

To further demonstrate the mechanism underlying the interaction between the sensitivity gene and the target gene, we selected epidermal growth factor receptor (EGFR) as an example. ErbB2/EGFR inhibitor is an important type of anticancer drug (e.g. Lapatinib and Varlitinib) and its gene is also a sensitivity gene in our DSGN. Tetsu *et al*. reported that inhibiting EGFR could evoke innate drug resistance in lung cancer cells by preventing Akt activity and thus inactivating Ets-1 function [31]. The expression of EPH receptor A2 (EPHA2) is activated by EGFR and EGFRvIII in the human cancer cell lines [32]. EPHA2 is also the target of Dasatinib. Interestingly, the combination of Dasatinib and Gefitinib (an EGFR inhibitor) presents anti-tumor properties that are superior to those of platinum-based combinations, indicating that this combination may be a promising new treatment modality to be tested in the clinic [33]. We extracted a subnetwork of epidermal growth factor receptor (EGFR) gene and its neighbor nodes from the HPRD network (Figure 5D). Then we found that some genes in this subnetwork (for example, PIK3R1, PIK3R3 and MET) are involved in the PI3K/Akt signaling pathway and some genes (JAK1, JAK2 and STAT5B) are involved in the JAK/STAT signaling pathway. PI3K/Akt was overexpressed and activated in cancer cells and was found to induce chemoresistance in various cancers [34–38]. For example, the PI3K/Akt pathway was found to be related to multidrug resistance in gastric cancer cells [39]. In addition, O'Gorman *et al*. found PI3-kinase inhibition
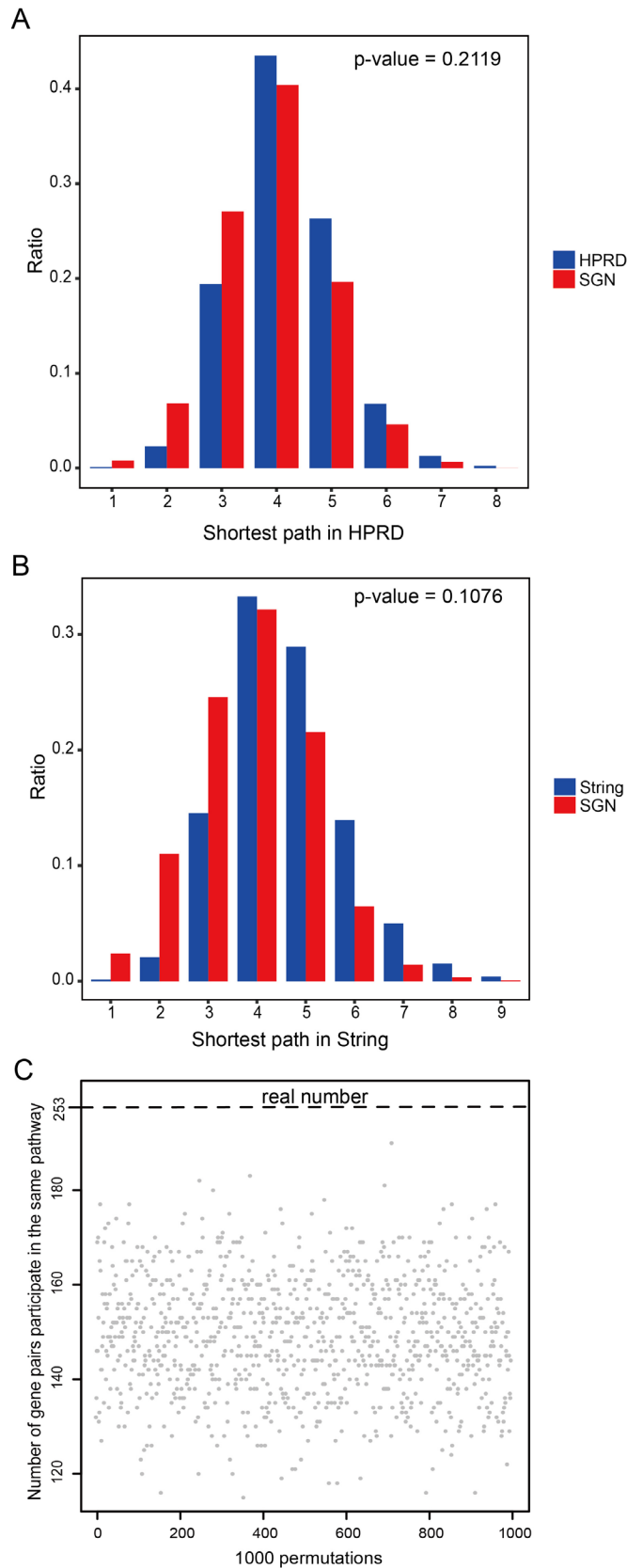
Figure 4: Functional characterizations of sensitivity genes belonging to the same drug. (**A**) Distribution of the shortest paths of sensitivity genes belonging to the same drug in the HPRD network (red), compared to the shortest paths among all nodes in the HPRD (blue). (**B**) Distribution of the shortest paths of sensitivity genes belonging to the same drug in the STRING network (red), compared to the shortest paths among all nodes in the STRING (blue). (**C**) 253 gene pairs participated in the same pathway, compared to 1000 permutations.

significantly increased sensitivity in HL60 human leukemia cells [40]. Meanwhile, the increased JAK/STAT signaling and enhanced interference with aerobic glycolysis and autophagy are associated with resistance to Afatinib [41]. Also, the crosstalk involving PI3K/Akt and JAK/STAT pathways was related to resistance to Sorafenib, an oral multikinase inhibitor [42].

DISCUSSION AND CONCLUSIONS

One of the biggest challenges associated with cancer chemotherapy resistance is discovering the unknown underlying mechanisms of drug sensitivity. Insight into the mechanisms of drug sensitivity is critical for effective treatment strategies in drug development. With the development of high-throughput technology, some high-quality data made it possible to reveal the potential mechanism underlying drug sensitivity. NCI-60 data assessed gene expression profiles in 60 human cancer cell lines and characterized drug activities by treatment with more than 70,000 different compounds.

CellMiner provided a rapid data retrieval of genomic data along with activity reports for ~20,000 chemical compounds across the NCI-60. They pioneered the approach of linking drug sensitivity to genomic data by the development of high-throughput experimental and bioinformatics technologies.

We constructed a DSGN using CellMiner Analysis tools to retrieve potential associations between a drug and its sensitivity genes in the NCI-60 cell line panel.

Network analysis of the DSGN offered insight into the interplay between drugs and sensitivity genes. We noticed that two drugs sharing the same sensitivity genes tended to share the same ATC classification and tended to cause the same side effects. Interestingly, we investigated the biological functions of the drug sensitivity genes, and found significantly enriched pathways were associated with cancer occurrence and development. Furthermore, compared to other functional gene sets, the sensitivity genes tended to be cancer genes, essential genes, adverse drug reaction-associated protein genes and orphan disease-causing mutant genes.
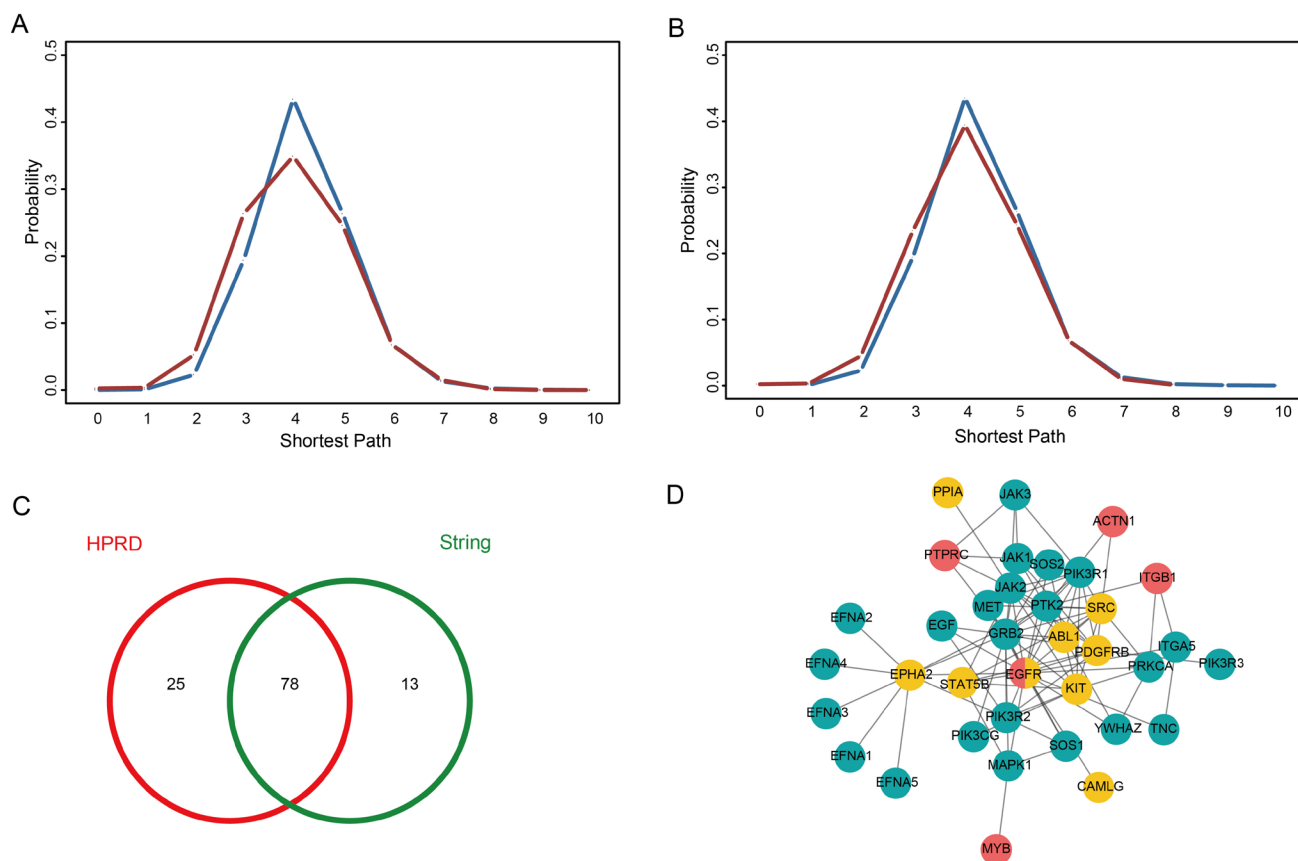

Figure 5: The relationships between a sensitivity gene and a target gene for the same drug. (**A**) Distribution of the shortest paths between a sensitivity gene and a target gene for the same drug in the HPRD network (red), compared to the shortest path distribution for all nodes in the HPRD (blue). (**B**) Distribution of the shortest paths between a sensitivity gene and a target gene for the same drug in the STRING network (red), compared to the shortest path distribution for all nodes in the STRING network (blue). (**C**) Among the shortest paths between sensitivity genes and target genes that are less than 4, 103 drugs are in the HPRD network and 91 drugs are in the STRING network, of which 78 drugs overlap between the two PPI networks. (**D**) Subnetwork of epidermal growth factor receptor (EGFR) and its neighbor nodes from the HPRD network; red corresponds to sensitivity genes, yellow corresponds to target genes, and green corresponds to enzymes in the PI3K/Akt signaling pathway.

We also investigated the different sensitivity genes of the same drug. We found the sensitivity genes tended to cluster together in the PPI network when we calculated the shortest path. Besides, by implementing pathway enrichment analysis, we found sensitivity gene pairs that belonged to the same drug tended to participate in the same pathway. That indicated that sensitivity genes of the same drug tended to engage in the same module of the PPI network and participate in the same biological functions. In view of the fact that anticancer therapeutic research is significantly associated with targeted delivery and molecular biomarkers, we investigated the relationships between the sensitivity genes and target genes based on the PPI network by calculating the shortest path between them, and we found that the sensitivity gene and target gene of the same drug were densely connected, indicating that they are functionally related. Interestingly, we found that the drug combination Dasatinib and Gefitinib could be inferred from the interaction of their targets, *EPHA2* and *EGFR*, suggesting that targeting sensitivity genes/proteins may be a potential strategy against drug resistance.

We also noticed that there were some limitations of our current study. Firstly, compared to tens of thousands of drugs, DSGN only include 181 FDA-approved anticancer small molecular drugs. It would be improved by the development of high-throughput experiments and pharmacogenomics. Another limitation of our study is the data source of DSGN was relatively simple. Besides NCI-60 dataset, integrating more drug-affected gene expression profiles of other resources and literatures will alleviate this limitation and complement DSGN. Although these data sets and the methodology are far from complete, our network analyses still provide statistically significant characteristics of the relationships between drugs and sensitivity genes.

In conclusion, we constructed a drug-sensitivity gene network (DSGN) based on the potential relationships between drugs and their sensitivity genes in the NCI-60 cell line panel, which contained 181 drugs, 1057 sensitivity genes and 1646 associations. Then we did a series of analysis to look for insights into the global relationships between drugs and their sensitivity genes, such as network analysis of DSGN, shortest path analysis between sensitivity gene and target gene, KEGG pathway enrichment analysis of sensitivity genes, ATC and side effects analysis of drugs, and so on. At last, our analyses provide statically significant characteristics of the relationships between drugs and sensitivity genes and help to gain insight into the mechanism of drug sensitivity.

MATERIALS AND METHODS

Generating the DSGN

We collected drugs and sensitivity genes from the NCI-60 cancer cell line database. The NCI-60 cancer cell line database is a large-scale information set with multiple genomic and drug response platforms. We retrieved potential associations between the drug activity and the expression levels of mRNAs by CellMiner, which is a powerful platform that allows rapid data retrieval of transcripts for genes along with activity reports for chemical compounds. The CellMiner provides 'NCI-60 Analysis Tools' to study the relationships between the mRNA expression and the 50% growth inhibitory concentration (GI50) values of drugs by calculating the Pearson correlation coefficient (PCC) between them. Firstly, we selected U.S. food and drug administration (FDA)-approved drugs and clinical trials drugs, and filtered out drugs not in the DrugBank database. Then we followed the steps to retrieve the correlations: (i) in the 'NCI-60 Analysis Tools' page, click 'Pattern comparison' and 'Drug NSC#' option in Step 1 section; (ii) input the drug NSC ID in Step 2 section; (iii) enter e-mail address and CellMiner would send the result documents of Pearson correlations between all genes and each input drug; (iv) integrated all drug files together. Assigning 0.5 as the PCC threshold, we obtained 16,694 correlations of drug-gene pairs, encompassing 6477 genes and 234 drugs, each drug was associated with up to 27.6 sensitivity genes on average. Then we did some dealing steps to reduce the false positive results and obtain the more significant drug-sensitivity gene relationships. For each drug: (i) we ranked the drug-gene pairs according to the absolute PCC value in descending order, and counted the number of sensitivity genes, and rounded this number up to 10-fold value rd (e.g. if the drug was related with 32 sensitivity genes, then rd = 40); (ii) we retained the top 10%*rd of the ranked drug-gene pairs (e.g. if rd = 40, then we retained top 4 drug-sensitivity gene pairs); (iii) we mapped NSC ID to the drug names. Finally, we obtained 1646 drug-gene pairs, encompassing 1057 genes and 181 drugs.

SIDER database

We obtained drug side effects data from a public computer-readable side effect resource, the side effect resource (SIDER) [43], which is freely available for academic research on the website http://sideeffects.embl.de. We collected 997 drugs corresponding to 4492 side effect terms. In the DSGN, 71 drugs were recorded in the SIDER database.

Different gene sets

CGC genes

The cancer genes are genes for which mutations have been causally implicated in cancer. We downloaded 571 cancer genes from the CGC (http://cancer.sanger.ac.uk/cancergenome/projects/census/), which is an ongoing effort to catalogue these cancer genes.

Essential genes

Essential genes are crucial in the study of the robustness of biological systems and effective drug target identification, while knockouts of them would result in a lack of cell viability or embryonic lethality. We collected 2721 essential genes from the Online GEne Essentiality (OGEE) database [44].

ADRPs

ADRPs are proteins that mediate adverse drug reactions or toxicity by binding to drugs or their reactive metabolites. We collected 547 ADRPs from Chen's previously published work [45].

MDGs

We collected 11,700 MDGs from the Online Mendelian Inheritance in Man database [46], which is a comprehensive, authoritative and timely knowledge base of human genes and genetic disorders compiled to support human genetics research and education and the practice of clinical genetics.

ODMGs

An orphan disease is a disease that affects fewer than 200,000 inhabitants, which is equivalent to approximately 6.5 patients per 10,000 inhabitants, and ODMGs are orphan disease-causing mutant genes. We collected 2123 ODMGs from Zhang's previously published work [47].

Drug target data source

The drug-target associations were downloaded from the DrugBank database [48]. We obtained target genes of 6905 drugs. Duplicated target genes without Entrez ID were excluded, and we collected 2354 drug target genes.

Extraction of subnetwork related to sensitivity gene and drug target gene

We selected epidermal growth factor receptor (EGFR) as example to study the mechanism underlying the interaction between the sensitivity gene and the target gene. EGFR gene and its first neighbor nodes were extracted from the HPRD to construct a subnetwork.

Author contributions

D.S.S. and Y.J.X. conceived and designed the study, H.X.Y., Y.P.Z. and D.S.S. collected and processed the data, D.S.S., H.X.Y. and J.S.W. performed the experiments. H.X.Y. and D.S.S. wrote the manuscript. All (T.W. and S.Y.L.) authors analyzed the data. Y.J.X. and D.S.S. supervised the research and revised the manuscript. All authors reviewed the manuscript.

CONFLICTS OF INTEREST

No potential conflicts of interest.

FUNDING

REFERENCES

1. Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, Bindal N, Beare D, Smith JA, Thompson IR, Ramaswamy S, Futreal PA, Haber DA, et al. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res. 2013; 41:D955–61. https://doi.org/10.1093/nar/gks1111.

2. Geeleher P, Cox NJ, Huang RS. Clinical drug response can be predicted using baseline gene expression levels and *in vitro* drug sensitivity in cell lines. Genome Biol. 2014; 15:R47. https://doi.org/10.1186/gb-2014-15-3-r47.

3. Druker BJ, Guilhot F, O'Brien SG, Gathmann I, Kantarjian H, Gattermann N, Deininger MW, Silver RT, Goldman JM, Stone RM, Cervantes F, Hochhaus A, Powell BL, et al, and IRIS Investigators. Five-year follow-up of patients receiving imatinib for chronic myeloid leukemia. N Engl J Med. 2006; 355:2408–17. https://doi.org/10.1056/NEJMoa062867.

4. Chapman PB, Hauschild A, Robert C, Haanen JB, Ascierto P, Larkin J, Dummer R, Garbe C, Testori A, Maio M, Hogg D, Lorigan P, Lebbe C, et al, and BRIM-3 Study Group. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. N Engl J Med. 2011; 364:2507–16. https://doi.org/10.1056/NEJMoa1103782.

5. Stetson LC, Pearl T, Chen Y, Barnholtz-Sloan JS. Computational identification of multi-omic correlates of anticancer therapeutic response. BMC Genomics. 2014 (Suppl 7); 15:S2. https://doi.org/10.1186/1471-2164-15-S7-S2. Erratum in: Erratum: Computational identification of multi-omic correlates of anticancer therapeutic response. [BMC Genomics. 2015].

6. Masica DL, Karchin R. Collections of simultaneously altered genes as biomarkers of cancer cell drug response. Cancer Res. 2013; 73:1699–708. https://doi.org/10.1158/0008-5472.CAN-12-3122.

7. van den Brink WJ, Wong YC, Gulave B, van der Graaf PH, de Lange EC. Revealing the Neuroendocrine Response After Remoxipride Treatment Using Multi-Biomarker

Discovery and Quantifying It by PK/PD Modeling. AAPS J. 2017; 19:274–285. https://doi.org/10.1208/s12248-016-0002-3.

8. Shi L, Reid LH, Jones WD, Shippy R, Warrington JA, Baker SC, Collins PJ, de Longueville F, Kawasaki ES, Lee KY, Luo Y, Sun YA, Willey JC, et al, and MAQC Consortium. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. Nat Biotechnol. 2006; 24:1151–61. https://doi.org/10.1038/nbt1239.

9. Liu H, D'Andrade P, Fulmer-Smentek S, Lorenzi P, Kohn KW, Weinstein JN, Pommier Y, Reinhold WC. mRNA and microRNA expression profiles of the NCI-60 integrated with drug activities. Mol Cancer Ther. 2010; 9:1080–91. https://doi.org/10.1158/1535-7163.MCT-09-0965.

10. Reinhold WC, Sunshine M, Liu H, Varma S, Kohn KW, Morris J, Doroshow J, Pommier Y. CellMiner: a web-based suite of genomic and pharmacologic tools to explore transcript and drug patterns in the NCI-60 cell line set. Cancer Res. 2012; 72:3499–511. https://doi.org/10.1158/0008-5472.CAN-12-1370.

11. Bethesda. NCI-60 DTP Human Tumor Cell Line Screen. Available from: https://dtp.cancer.gov/databases_tools/bulk_data.htm.

12. Zoppoli G, Regairaz M, Leo E, Reinhold WC, Varma S, Ballestrero A, Doroshow JH, Pommier Y. Putative DNA/RNA helicase Schlafen-11 (SLFN11) sensitizes cancer cells to DNA-damaging agents. Proc Natl Acad Sci USA. 2012; 109:15030–35. https://doi.org/10.1073/pnas.1205943109.

13. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, Wilson CJ, Lehár J, Kryukov GV, Sonkin D, Reddy A, Liu M, Murray L, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature. 2012; 483:603–07. https://doi.org/10.1038/nature11003.

14. Kiianitsa K, Maizels N. A rapid and sensitive assay for DNA-protein covalent complexes in living cells. Nucleic Acids Res. 2013; 41:e104. https://doi.org/10.1093/nar/gkt171.

15. Tian L, Song S, Liu X, Wang Y, Xu X, Hu Y, Xu J. Schlafen-11 sensitizes colorectal carcinoma cells to irinotecan. Anticancer Drugs. 2014; 25:1175–81. https://doi.org/10.1097/CAD.0000000000000151.

16. Kazi JU, Kabir NN, Rönnstrand L. Role of SRC-like adaptor protein (SLAP) in immune and malignant cell signaling. Cell Mol Life Sci. 2015; 72:2535–44. https://doi.org/10.1007/s00018-015-1882-6.

17. Blank CU, Larkin J, Arance AM, Hauschild A, Queirolo P, Del Vecchio M, Ascierto PA, Krajsova I, Schachter J, Neyns B, Garbe C, Chiarion Sileni V, Mandalà M, et al. Open-label, multicentre safety study of vemurafenib in 3219 patients with BRAFV600 mutation-positive metastatic melanoma: 2-year follow-up data and long-term responders'

analysis. Eur J Cancer. 2017; 79:176–84. https://doi.org/10.1016/j.ejca.2017.04.007.

18. Naeger LK, Margot NA, Miller MD. ATP-dependent removal of nucleoside reverse transcriptase inhibitors by human immunodeficiency virus type 1 reverse transcriptase. Antimicrob Agents Chemother. 2002; 46:2179–84. https://doi.org/10.1128/AAC.46.7.2179-2184.2002.

19. Phanalasy S, and ACTwatch Group. The malaria testing and treatment landscape in the southern Lao People's Democratic Republic (PDR). Malar J. 2017; 16:169. https://doi.org/10.1186/s12936-017-1769-0.

20. Huang W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009; 4:44–57. https://doi.org/10.1038/nprot.2008.211.

21. Sui S, Wang X, Zheng H, Guo H, Chen T, Ji DM. Gene set enrichment and topological analyses based on interaction networks in pediatric acute lymphoblastic leukemia. Oncol Lett. 2015; 10:3354–62.

22. Zhao Y, Zhang X, Zhao Y, Kong D, Qin F, Sun J, Dong Y. Identification of potential therapeutic target genes, key miRNAs and mechanisms in acute myeloid leukemia based on bioinformatics analysis. Med Oncol. 2015; 32:152. https://doi.org/10.1007/s12032-015-0572-4.

23. Campillos M, Kuhn M, Gavin AC, Jensen LJ, Bork P. Drug target identification using side-effect similarity. Science. 2008; 321:263–66. https://doi.org/10.1126/science.1158140.

24. Emig D, Ivliev A, Pustovalova O, Lancashire L, Bureeva S, Nikolsky Y, Bessarabova M. Drug target prediction and repositioning using an integrated network-based approach. PLoS One. 2013; 8:e60618. https://doi.org/10.1371/journal.pone.0060618.

25. Torres NB, Altafini C. Drug combinatorics and side effect estimation on the signed human drug-target network. BMC Syst Biol. 2016; 10:74. https://doi.org/10.1186/s12918-016-0326-8.

26. Yildirim MA, Goh KI, Cusick ME, Barabási AL, Vidal M. Drug-target network. Nat Biotechnol. 2007; 25:1119–26. https://doi.org/10.1038/nbt1338.

27. Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen B, Venugopal A, Balakrishnan L, Marimuthu A, Banerjee S, et al. Human Protein Reference Database—2009 update. Nucleic Acids Res. 2009; 37:D767–72. https://doi.org/10.1093/nar/gkn892.

28. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, von Mering C. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 2015; 43:D447–52. https://doi.org/10.1093/nar/gku1003.

29. Tortorella S, Karagiannis TC. The significance of transferrin receptors in oncology: the development of functional

nano-based drug delivery systems. Curr Drug Deliv. 2014; 11:427–43. https://doi.org/10.2174/1567201810666140106115436.

30. Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, Maciejewski A, Arndt D, Wilson M, Neveu V, Tang A, Gabriel G, Ly C, et al. DrugBank 4.0: shedding new light on drug metabolism. Nucleic Acids Res. 2014; 42:D1091–97. https://doi.org/10.1093/nar/gkt1068.

31. Phuchareon J, McCormick F, Eisele DW, Tetsu O. EGFR inhibition evokes innate drug resistance in lung cancer cells by preventing Akt activity and thus inactivating Ets-1 function. Proc Natl Acad Sci USA. 2015; 112:E3855–63. https://doi.org/10.1073/pnas.1510733112.

32. Larsen AB, Pedersen MW, Stockhausen MT, Grandal MV, van Deurs B, Poulsen HS. Activation of the EGFR gene target EphA2 inhibits epidermal growth factor-induced cancer cell motility. Mol Cancer Res. 2007; 5:283–93. https://doi.org/10.1158/1541-7786.MCR-06-0321.

33. Thibault B, Jean-Claude B. Dasatinib+Gefitinib, a non platinum-based combination with enhanced growth inhibitory, anti-migratory and anti-invasive potency against human ovarian cancer cells. J Ovarian Res. 2017; 10:31. https://doi.org/10.1186/s13048-017-0319-2.

34. Gao X, Deeb D, Jiang H, Liu Y, Dulchavsky SA, Gautam SC. Synthetic triterpenoids inhibit growth and induce apoptosis in human glioblastoma and neuroblastoma cells through inhibition of prosurvival Akt, NF-kappaB and Notch1 signaling. J Neurooncol. 2007; 84:147–57. https://doi.org/10.1007/s11060-007-9364-9.

35. Lane D, Robert V, Grondin R, Rancourt C, Piché A. Malignant ascites protect against TRAIL-induced apoptosis by activating the PI3K/Akt pathway in human ovarian carcinoma cells. Int J Cancer. 2007; 121:1227–37. https://doi.org/10.1002/ijc.22840.

36. Lim WT, Zhang WH, Miller CR, Watters JW, Gao F, Viswanathan A, Govindan R, McLeod HL. PTEN and phosphorylated AKT expression and prognosis in early- and late-stage non-small cell lung cancer. Oncol Rep. 2007; 17:853–57.

37. Chen Y, Wang Z, Chang P, Xiang L, Pan F, Li J, Jiang J, Zou L, Yang L, Bian Z, Liang H. The effect of focal adhesion kinase gene silencing on 5-fluorouracil chemosensitivity involves an Akt/NF-kappaB signaling pathway in colorectal carcinomas. Int J Cancer. 2010; 127:195–206. https://doi.org/10.1002/ijc.25025.

38. Ripka S, Neesse A, Riedel J, Bug E, Aigner A, Poulsom R, Fulda S, Neoptolemos J, Greenhalf W, Barth P, Gress

TM, Michl P. CUX1: target of Akt signalling and mediator of resistance to apoptosis in pancreatic cancer. Gut. 2010; 59:1101–10. https://doi.org/10.1136/gut.2009.189720.

39. Han Z, Hong L, Han Y, Wu K, Han S, Shen H, Li C, Yao L, Qiao T, Fan D. Phospho Akt mediates multidrug resistance of gastric cancer cells through regulation of P-gp, Bcl-2 and Bax. J Exp Clin Cancer Res. 2007; 26:261–68.

40. O'Gorman DM, McKenna SL, McGahon AJ, Cotter TG. Inhibition of PI3-kinase sensitises HL60 human leukaemia cells to both chemotherapeutic drug- and Fas-induced apoptosis by a JNK independent pathway. Leuk Res. 2001; 25:801–11. https://doi.org/10.1016/S0145-2126(01)00024-8.

41. van der Wekken AJ, Saber A, Hiltermann TJ, Kok K, van den Berg A, Groen HJ. Resistance mechanisms after tyrosine kinase inhibitors afatinib and crizotinib in non-small cell lung cancer, a review of the literature. Crit Rev Oncol Hematol. 2016; 100:107–16. https://doi.org/10.1016/j.critrevonc.2016.01.024.

42. Zhu YJ, Zheng B, Wang HY, Chen L. New knowledge of the mechanisms of sorafenib resistance in liver cancer. Acta Pharmacol Sin. 2017; 38:614–22. https://doi.org/10.1038/aps.2017.5.

43. Kuhn M, Letunic I, Jensen LJ, Bork P. The SIDER database of drugs and side effects. Nucleic Acids Res. 2016; 44:D1075–79. https://doi.org/10.1093/nar/gkv1075.

44. Chen WH, Minguez P, Lercher MJ, Bork P. OGEE: an online gene essentiality database. Nucleic Acids Res. 2012; 40:D901–06. https://doi.org/10.1093/nar/gkr986.

45. Chen X, Liu X, Jia X, Tan F, Yang R, Chen S, Liu L, Wang Y, Chen Y. Network characteristic analysis of ADR-related proteins and identification of ADR-ADR associations. Sci Rep. 2013; 3:1744. https://doi.org/10.1038/srep01744.

46. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res. 2005; 33:D514–17. https://doi.org/10.1093/nar/gki033.

47. Zhang M, Zhu C, Jacomy A, Lu LJ, Jegga AG. The orphan disease networks. Am J Hum Genet. 2011; 88:755–66. https://doi.org/10.1016/j.ajhg.2011.05.006.

48. Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A, Pon A, Banco K, Mak C, Neveu V, Djoumbou Y, Eisner R, Guo AC, Wishart DS. DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. Nucleic Acids Res. 2011; 39:D1035–41. https://doi.org/10.1093/nar/gkq1126.

.