



# Guizhi Fuling Decoction inhibiting the PI3K and MAPK pathways in breast cancer cells revealed by HTS<sup>2</sup> technology and systems pharmacology



Yifei Dai<sup>a</sup>, Weijie Qiang<sup>b</sup>, Xiankuo Yu<sup>c</sup>, Siwei Cai<sup>d</sup>, Kequan Lin<sup>e</sup>, Lan Xie<sup>f,g</sup>, Xun Lan<sup>a,\*</sup>, Dong Wang<sup>c,\*</sup>

<sup>a</sup> Department of Basic Medical Sciences, School of Medicine, Tsinghua University, Beijing 100084, China

<sup>b</sup> Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100193, China

<sup>c</sup> School of Basic Medical Sciences, Chengdu University of Traditional Chinese Medicine, Chengdu 611137, China

<sup>d</sup> Department of Electronic and Computer Engineering, College of Engineering, Drexel University, Philadelphia 19104, USA

<sup>e</sup> School of Life Sciences, Tsinghua University, Beijing 100084, China

<sup>f</sup> Medical Systems Biology Research Center, School of Medicine, Tsinghua University, Beijing 100084, China

<sup>g</sup> National Engineering Research Center for Beijing Biochip Technology, Beijing 102206, China

## ARTICLE INFO

### Article history:

Received 23 February 2020

Received in revised form 25 April 2020

Accepted 6 May 2020

Available online 18 May 2020

### Keywords:

Guizhi Fuling Decoction

Breast cancer

Systems pharmacology

HTS<sup>2</sup> assay

Bioinformatics

## ABSTRACT

As one of the classical traditional Chinese medicine (TCM) prescriptions in treating gynecological tumors, Guizhi Fuling Decoction (GFD) has been used to treat breast cancer (BRCA). Nonetheless, the potential molecular mechanism remains unclear so far. Therefore, systems pharmacology was used in combination with high throughput sequencing-based high throughput screening (HTS<sup>2</sup>) assay and bioinformatic technologies in this study to investigate the molecular mechanisms of GFD in treating BRCA. By computationally analyzing 76 active ingredients in GFD, 38 potential therapeutic targets were predicted and significantly enriched in the “pathways in cancer”. Meanwhile, experimental analysis was carried out to examine changes in the expression levels of 308 genes involved in the “pathways in cancer” in BRCA cells treated by five herbs of GFD utilizing HTS<sup>2</sup> platform, and 5 key therapeutic targets, including HRAS, EGFR, PTK2, SOS1, and ITGB1, were identified. The binding mode of active compounds to these five targets was analyzed by molecular docking and molecular dynamics simulation. It was found after integrating the computational and experimental data that, GFD possessed the anti-proliferation, proapoptosis, and anti-angiogenesis activities mainly through regulating the PI3K and the MAPK signaling pathways to inhibit BRCA. Besides, consistent with the TCM theory about the synergy of *Cinnamomi Ramulus* (Guizhi) by *Cortex Moutan* (Mudanpi) in GFD, both of these two herbs acted on the same targets and pathways. Taken together, the combined application of computational systems pharmacology techniques and experimental HTS<sup>2</sup> platform provides a practical research strategy to investigate the functional and biological mechanisms of the complicated TCM prescriptions.

© 2020 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

According to the Global Cancer Report released at the end of 2018, breast cancer (BRCA) is the most common malignancy in female patients, and it ranks second place in terms of morbidity among all malignancies [1]. Apart from chemotherapy and radiotherapy, complementary and alternative therapy (CAM) has gradually become a new therapeutic option. As an indispensable

component of CAM, traditional Chinese medicine (TCM) has been increasingly applied in cancer prevention and treatment over the last few decades [2]. Compared with chemotherapy which has a series of adverse reactions, TCM therapy has certain advantages in treating breast cancer, including less adverse reactions, safer property and higher patient compliance [3].

Guizhi Fuling Decoction (GFD) is originated from a classic medical book named the *Synopsis of the Golden Chamber* (*Jingui Yaolue*) written by the famous Chinese physician *Zhang Zhongjing* at about 1800 years ago. It is composed of 5 herbs, including *Cinnamomi Ramulus* (Guizhi), *Poria* (Fuling), *Radix Paeoniae Rubra* (Chishao), *Cortex Moutan* (Mudanpi), and *Persicae Semen* (Taoren). In ancient

\* Corresponding authors.

E-mail addresses: [xlan@tsinghua.edu.cn](mailto:xlan@tsinghua.edu.cn) (X. Lan), [dwang@cducm.edu.cn](mailto:dwang@cducm.edu.cn) (D. Wang).

clinical practice, Chinese physicians have used GFD to prevent and treat gynecological malignant tumors included mammary tumors, which provide theoretical support for the treatment of breast cancer [4]. In addition, numerous experimental researches have fully proved GFD could inhibit the proliferation and induce the apoptosis of breast cancer cells [5,6]. However, the underlying molecular mechanisms remain largely unknown for the time being.

Systems pharmacology, an emerging discipline that emphasizes integrity and systematization, is a more appropriate approach to investigate the complicated mechanisms of TCM, which commonly functions in a multi-component and multi-target manner [7]. Systems pharmacology has been successfully adopted in previous studies to predict the primary bioactive substances and to elucidate the potential therapeutic mechanisms of TCM prescriptions, such as Huanglian Jiedu Decoction [8]. Nonetheless, the suitable verification method is lacking to verify the results predicted by systems pharmacology, which leads to the uncertainty of research. Therefore, this study applied a novel and powerful research technique called high throughput sequencing-based high throughput screening (HTS<sup>2</sup>) [9,10].

HTS<sup>2</sup>, a high-throughput screening platform based on gene expression signature, is mainly composed of the RASL strategy and the next-generation sequencing technology. This platform allows us to simultaneously examine the expression of thousands of genes in human cells treated with thousands of herbs; importantly, it provides the large-scale herb-cell-gene datasets for effectively validating the predicted results of systems pharmacology, and brings new research inspiration. Moreover, it can quantitatively analyze gene matrices related to a specific phenotype or focusing on a specific pathway. Compared with genome sequencing, the HTS<sup>2</sup> technique is more targeted and cost-effective.

In this study, systems pharmacology was employed to construct an active ingredient-BRCA target network, so as to screen the most significantly enriched anti-BRCA signaling pathway of GFD. Thereafter, HTS<sup>2</sup> assay was carried out to verify the effect of drug intervention on gene expression involved in this pathway. Finally,

bioinformatics analysis was performed to identify the key targets of GFD and to explain its anti-BRCA mechanism. Fig. 1 presents the research flowchart in this study.

## 2. Materials and methods

### 2.1. Construction of the chemical component database

All the chemical components of GFD were collected from the Traditional Chinese Medicine Systems Pharmacology Database (TCMSP) (<http://lsp.nwu.edu.cn/>, version 2.3, updated on May 31st, 2014) [11] and the NCBI PubChem Database (<https://pubchem.ncbi.nlm.nih.gov/>, updated on May 8th, 2019) [12].

### 2.2. Screening the active ingredients

The chemical components of GFD were filtered by Oral bioavailability (OB) and Drug-likeness (DL). Of them, OB in vivo (%F), which represents the unchanged fraction of the orally administered dose that achieves systemic circulation, is one of the most commonly pharmacokinetic parameters used in drug screening. In this study, a robust calculative system, OBioavail 1.1, was employed to predict the compound OB [13]. Then, compounds with OB  $\geq 30\%$  were selected as the active ingredients in this study. Besides, DL is defined as a delicate balance among the various molecular properties that affect pharmacodynamics and pharmacokinetics, which ultimately affects its absorption, distribution, metabolism, and excretion (ADME) in the human body like a drug. In this study, the DL index of a new compound was calculated according to the Tanimoto similarity [14].

$$f(A, B) = \frac{A \cdot B}{|A|^2 + |B|^2 - A \cdot B}$$

Where A represents the new compound and B stands for the average DL index of all the 6511 molecules in the DrugBank database based on the Dragon soft descriptors.

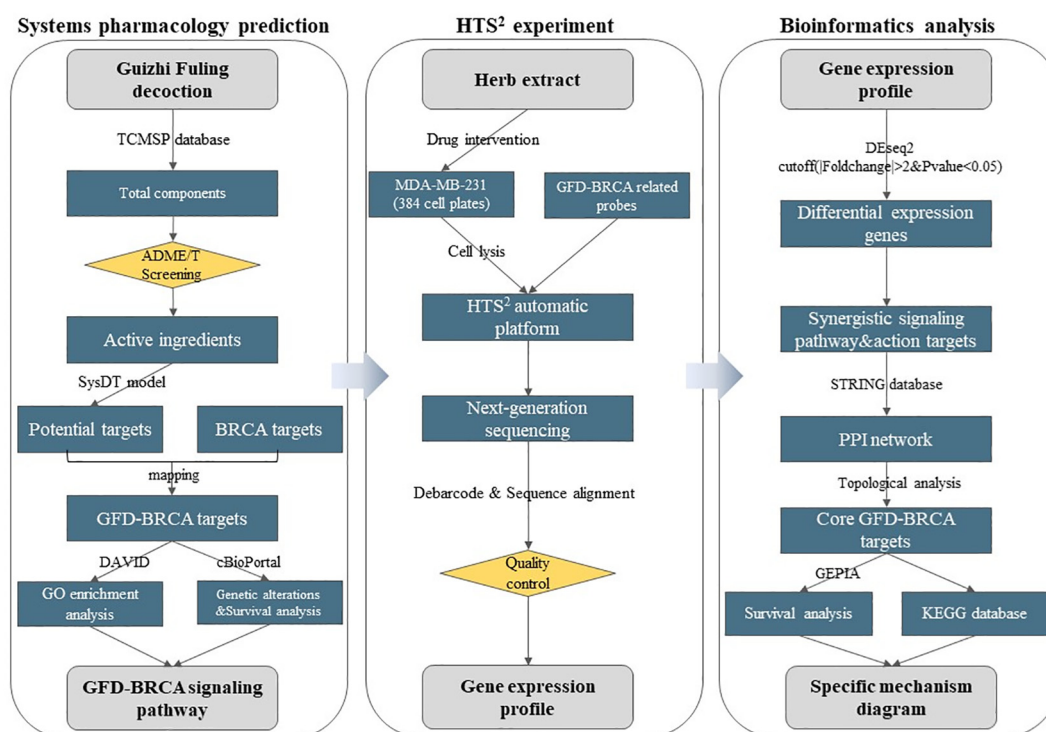


Fig. 1. Experimental technical roadmap.

Afterwards, compounds with DL <0.18 were removed. Finally, compounds with both OB  $\geq$ 30% and DL  $\geq$ 0.18 were considered as the active ingredients, as suggested by the TCMSP database.

### 2.3. Prediction of the potential targets of active ingredients

SysDT, the drug-target prediction model, was adopted in this study to predict the targets of active ingredients. Notably, SysDT is based on 6511 drugs and 3987 targets of the DrugBank database, as well as the mutual correlation degree [15]. It is established according to the support vector machine (SVM) algorithm, with the consistency, sensitivity, and specificity of 82.83%, 81.33% and 93.62%, respectively. Using this model, targets with SVM >0.7 were predicted as the potential targets of active ingredients. Additionally, the target information from the STITCH (<http://stitch.embl.de/>, version 5.0) [16], the Therapeutic Target Database (TTD) (<http://bidd.nus.edu.sg/group/cjttd/>, updated on September 15th, 2017) [17], and the Uniprot (<https://www.uniprot.org/>, updated on July 31st, 2019) [18] databases, was integrated to supplement this prediction model.

### 2.4. Collection of BRCA therapeutic targets and network mapping

All BRCA therapeutic targets were collected from the OMIM (<https://omim.org/>, updated on June 28th, 2019) [19], the DrugBank (<https://www.drugbank.ca/>, version 5.1.4, updated on July 2nd, 2019) [20], and the TTD (<http://bidd.nus.edu.sg/group/cjttd/>, updated on September 15th, 2017) [17] databases. Thereafter, the BRCA target network was established using Cytoscape v3.7.1 [21], which was then mapped with the compound-target network to obtain the active ingredient-BRCA target network of GFD, including all the GFD-related targets for BRCA treatment. Moreover, the Venn diagram was acquired through the Venn diagram web tool (<http://bioinformatics.psb.ugent.be/webtools/Venn/>) to visualize the overlapping targets.

### 2.5. Functional enrichment analysis

The functional enrichment analysis of the overlapping targets was performed by DAVID (<https://david.ncifcrf.gov/>, version 6.8) [22]. Later, the gene ontology (GO) biological process (BP) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were enriched based on the adjusted *P*-value of <0.05. Thereafter, the top 10 signaling pathways were drawn into a bubble chart using the R software [23].

### 2.6. Genetic alteration and survival analysis

cBioPortal (<https://www.cbioportal.org/>, updated on August 13th, 2019) [24], a web-based integrated data mining system, was utilized to examine the genetic alterations and to perform survival analysis on GFD-related targets for BRCA treatment.

### 2.7. The HTS<sup>2</sup> assay

HTS<sup>2</sup> is a high-throughput screening platform based on the gene expression signature that quantitatively analyzes cell transcriptional profiles at a large scale [9,25]. In this study, the HTS<sup>2</sup> assay was carried out to detect the regulation of GFD on the most significantly enriched signaling pathway in BRCA cells.

#### 2.7.1. Cell culture and materials

MDA-MB-231 cells were obtained from the China Infrastructure of Cell Line Resources (Beijing, China), and cultured in DMEM medium (Gibco) supplemented with 10% fetal bovine serum (FBS, Gemini), 100 U/mL streptomycin and 100 U/mL penicillin (Gibco) under

the humidified atmosphere of 95% air and 5% CO<sub>2</sub> at 37°C. Meanwhile, the extracts of five herbs contained in GFD were provided by CapitalBio Corporation (Beijing, China). All probes were purchased from Invitrogen (Shanghai, China).

#### 2.7.2. Preparation of herb extracts

Firstly, the five herbs, namely, *Cinnamomi Ramulus*, *Poria*, *Radix Paeoniae Rubra*, *Cortex Moutan*, and *Persicae Semen*, were pulverized into powders using a grinder. Secondly, the powders were extracted with the 90% ethanol solvent in a Soxhlet extractor. Then, the solvent was dried to extractum by the drying oven at 45 °C. Finally, a certain amount of extractum was weighed and diluted with DMSO to 100 µg/mL, and preserved in a refrigerator at –80 °C for further studies.

#### 2.7.3. Probe design

Afterwards, genes involved in the most significant KEGG pathway in GFD were selected as the probes [26]. Typically, three pairs of probes were designed for each gene according to exon-exon junctions near the 3'-terminal of target mRNA. Moreover, the T<sub>m</sub> range of those designed probes was required to limit G/C content in probe design, and all probes must be tested to prevent the high-affinity probes from excessively occupying the sequencing space. Then, one pair of probes was selected to represent a transcript.

#### 2.7.4. HTS<sup>2</sup> screening

Gene signature represents a phenotype of interests, which is used to search for herbs that affect phenotype in the presence or absence of a validated drug target. This study attempted to explore the synergetic mechanisms of the five herbs in GFD in MDA-MB-231 cells. To this end, about 5000 MDA-MB-231 cells were grown in the 384-well plates for 24 h, treated with herb extracts for 24 h, lysed with lysis buffer, and incubated at room temperature for 10 min. Then, the lysis was preserved at –80 °C for subsequent automated HTS<sup>2</sup> assay. Later, the sample obtained from the HTS<sup>2</sup> assay was sequenced using the Illumina HiSeq X Ten sequencer.

### 2.8. Data processing

First of all, all reads were mapped to the probe sequences, which allowed for three mismatches and were normalized relative to the expression of 18 stable housekeeping genes. Subsequently, to evaluate the reliability and repeatability of the transcriptional profile, the Pearson correlation coefficients among normalized transcriptional data were calculated by R software after treatments with 16 replicates of DMSO and 3 replicates of the five herbs in GFD. Notably, the correlation coefficients of >0.9 indicated that the HTS<sup>2</sup> assay results were reliable and repeatable. Then, the FC and *P*-value of gene expression were calculated via R package DESeq2 [27], and genes with |FC| > 2 and *P*-value <0.05 were considered as the differentially expressed genes (DEGs). Later, the heatmap and volcano plots of DEGs were drawn using the R package pheatmap and ggplot2, respectively [28,29].

### 2.9. Interaction network construction and analysis

Based on the KEGG enrichment analysis results on DEGs, the herb-target-pathway interaction network was constructed [26]. Besides, information related to the phenotypes associated with the KEGG pathway was also collected. Subsequently, the herb-pathway-phenotype interaction network was constructed based on the links, so as to clarify the synergistic effects of herbs.

### 2.10. Construction of the protein–protein interaction (PPI) network and screening of key targets

To screen the key targets among the DEGs verified by the HTS<sup>2</sup> assay, all DEGs were mapped to the STRING database (<https://string-db.org/>, version 11.0, updated on January 19th, 2019), a database of known and predicted PPIs [30]. Afterwards, the PPI network with a combined score of >0.9 was constructed. Subsequently, topological analysis was carried out using the Network Analyzer plug-in contained in Cytoscape, and the main topological parameters of the PPI network were obtained. In this study, the degree centrality was adopted as the major parameter, which was used in combination with the closeness centrality and the betweenness centrality to identify the key anti-BRCA targets in GFD.

### 2.11. Survival analysis

The Kaplan-Meier Plotter (<http://kmplot.com/>) was adopted for analyzing the association of key anti-BRCA target expression in GFD with BRCA patient survival rate (n = 3951) [31].

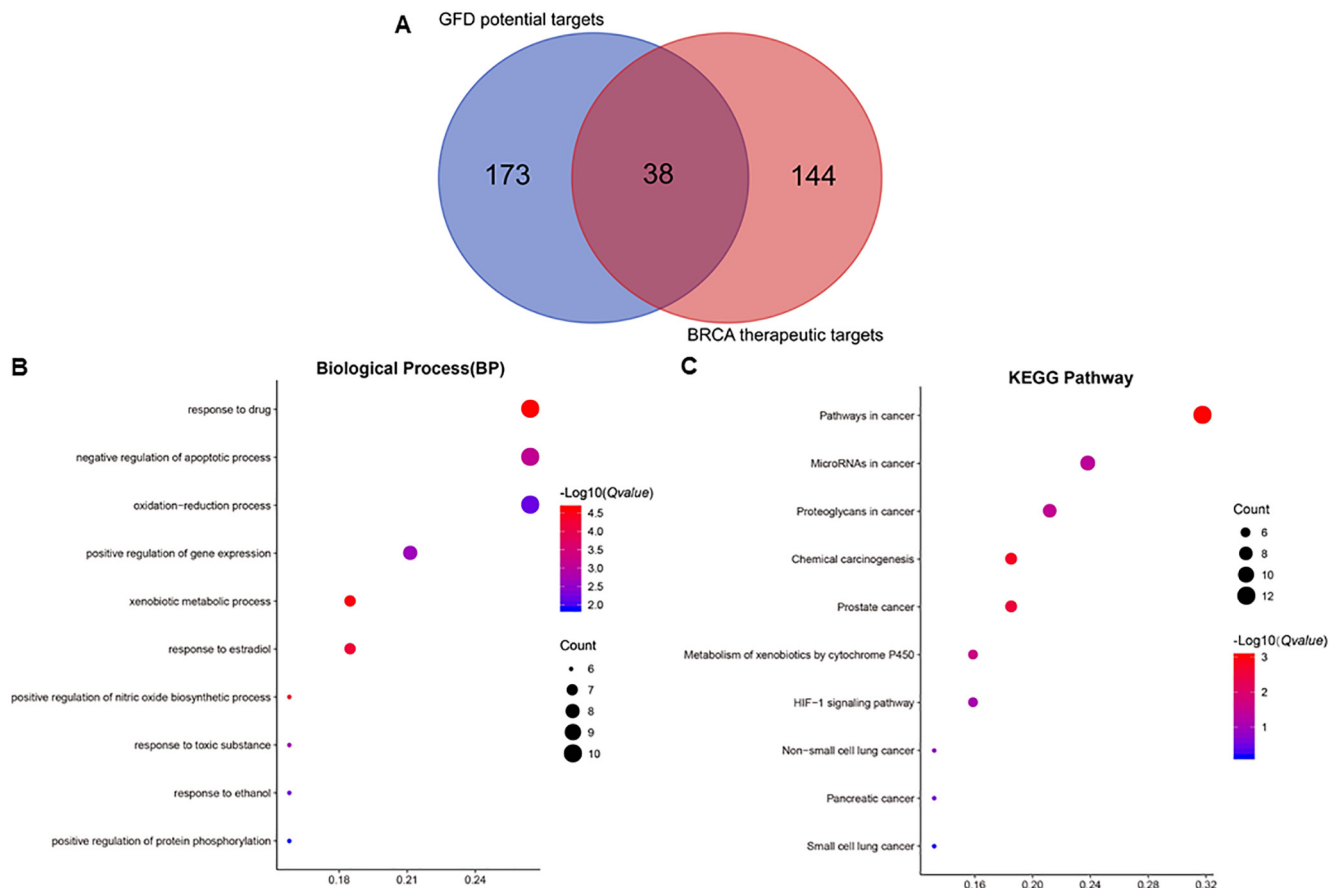
### 2.12. Molecular docking

The crystal structures of key anti-BRCA targets in GFD were obtained from the RCSB Protein Data Bank (PDB) [32]. Then, the crystal structure of each protein was selected based on the optimal available resolution. Moreover, the protein preparation wizard

module of Schrodinger's Maestro Molecular modeling suit (Schrodinger Release, 2019–1) was utilized to prepare the protein crystallographic structures [33]. In addition, the LigPrep module of Schrodinger's Maestro Molecular modeling suit was employed to obtain the 3D structures and energy minimization of the active ingredients in GFD. Based on the specific known active sites of the protein targets, Glide was adopted for all molecular docking simulations and calculations [34]. Moreover, the ligand–target interactions were visualized by the ligand interaction diagram module.

### 2.13. Molecular dynamics (MD) simulation

To further optimize the docking results, MD simulation was performed by GROMACS 2019 with all-atomic force field and SPC water model [35]. The simulation temperature was set to 300 K, and simulation system formed by the water molecules added around the protein was served as the periodic boundary. In the simulation process, the particle-mesh Ewald (PME) algorithm was used to calculate the long-range electrostatic interactions, and 2 fs integral time step was applied. Under the NVT (keeping the atomic number, volume and temperature constant) ensemble, the system was balanced and the water was optimized for 500 ps. Afterwards, another 500 ps equilibration under NPT ensemble conditions (keeping atomic number, pressure and temperature constant) was conducted, which was followed by a final production run of 100 ns. The binding energy of the compound and the protein was calculated using MM-PBSA method. And the images were drawn by pymol 2.3 and Maestro 11.9 [33,36].



**Fig. 2.** Functional enrichment analysis. (A) The Venn diagram of the potential targets in GFD and the therapeutic targets for BRCA. (B) GO enrichment analysis on 38 anti-BRCA targets in GFD. (C) KEGG pathway enrichment analysis on 38 anti-BRCA targets in GFD.

### 2.14. Plotting of the GFD-related mechanism diagram

According to the results obtained from HTS<sup>2</sup> assay, KEGG pathway enrichment analysis, topological analysis, and molecular docking, a specific mechanism diagram related to those key anti-BRCA targets in GFD was plotted to visualize the significant molecular mechanisms of GFD in treating BRCA.

## 3. Results and discussion

### 3.1. Screening of active ingredients

All chemical components of the five herbs contained in GFD, including Guizhi (2 2 0), Fuling (34), Chishao (1 1 9), Mudanpi (55), and Taoren (66), were collected through TCMSP and PubChem. Afterwards, altogether 76 compounds with OB  $\geq$ 30% and DL  $\geq$ 0.18 were identified as the active ingredients (Table S1). Among all active ingredients contained in Guizhi (7), Fuling (15), Chishao (29), Mudanpi (11), and Taoren (23), 6 were shared by 2 or 3 herbs. For example, beta-sitosterol was shared by Guizhi, Chishao, and Taoren.

### 3.2. Potential targets of GFD

The SysDT model was utilized to predict the potential targets for all active ingredients in GFD, and a total of 211 potential protein

targets were finally acquired. The detailed information about these potential targets is provided in Table S2.

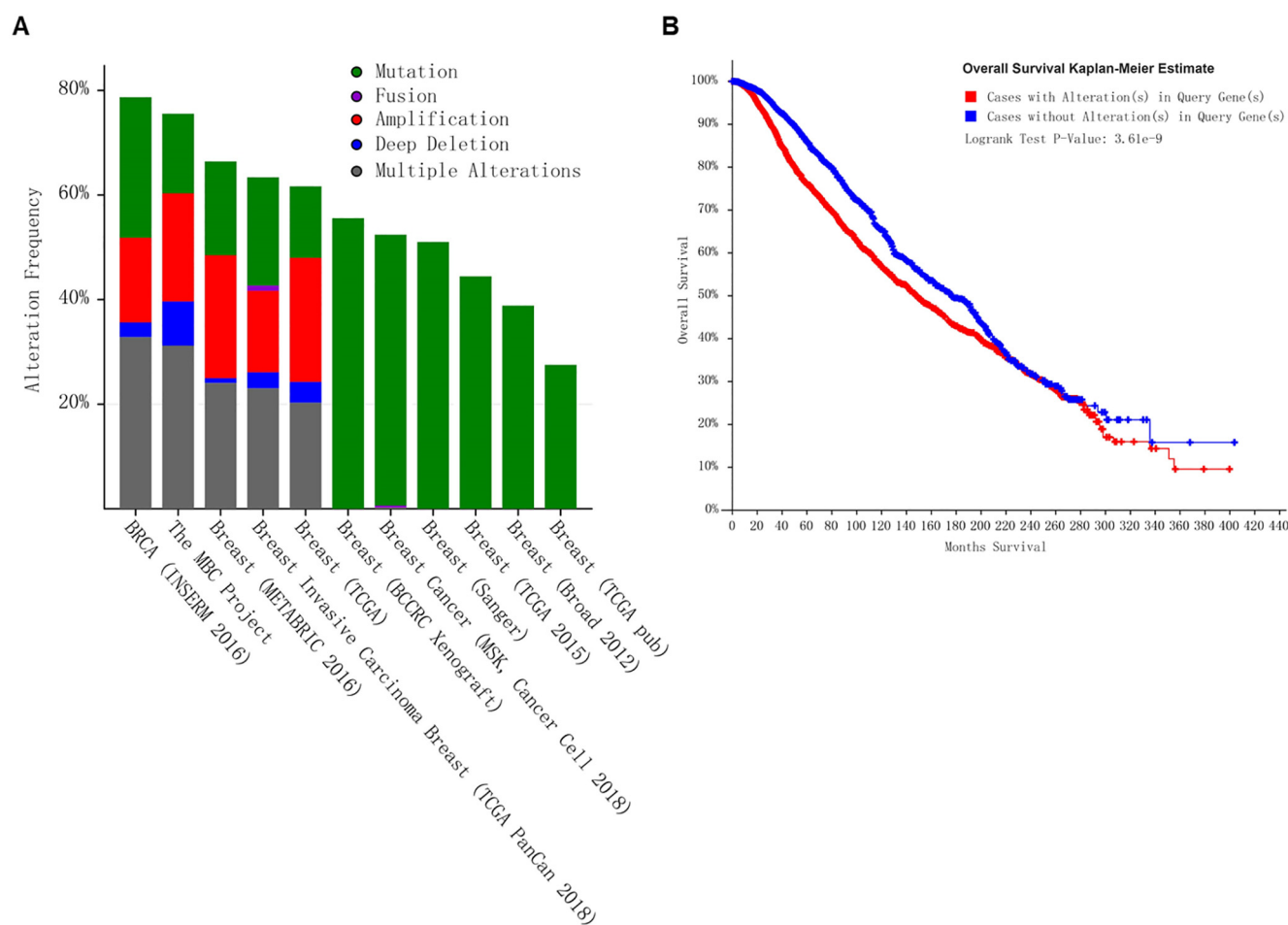
### 3.3. Known therapeutic targets for BRCA

To acquire all the therapeutic targets for BRCA, they were collected manually from the TTD, OMIM and DrugBank databases, respectively. In total, 58, 21 and 118 known therapeutic protein targets for BRCA were acquired from the TTD, OMIM and DrugBank databases, respectively. Ultimately, 182 known BRCA targets were collected after eliminating the redundancy. The detailed information is presented in Table S3.

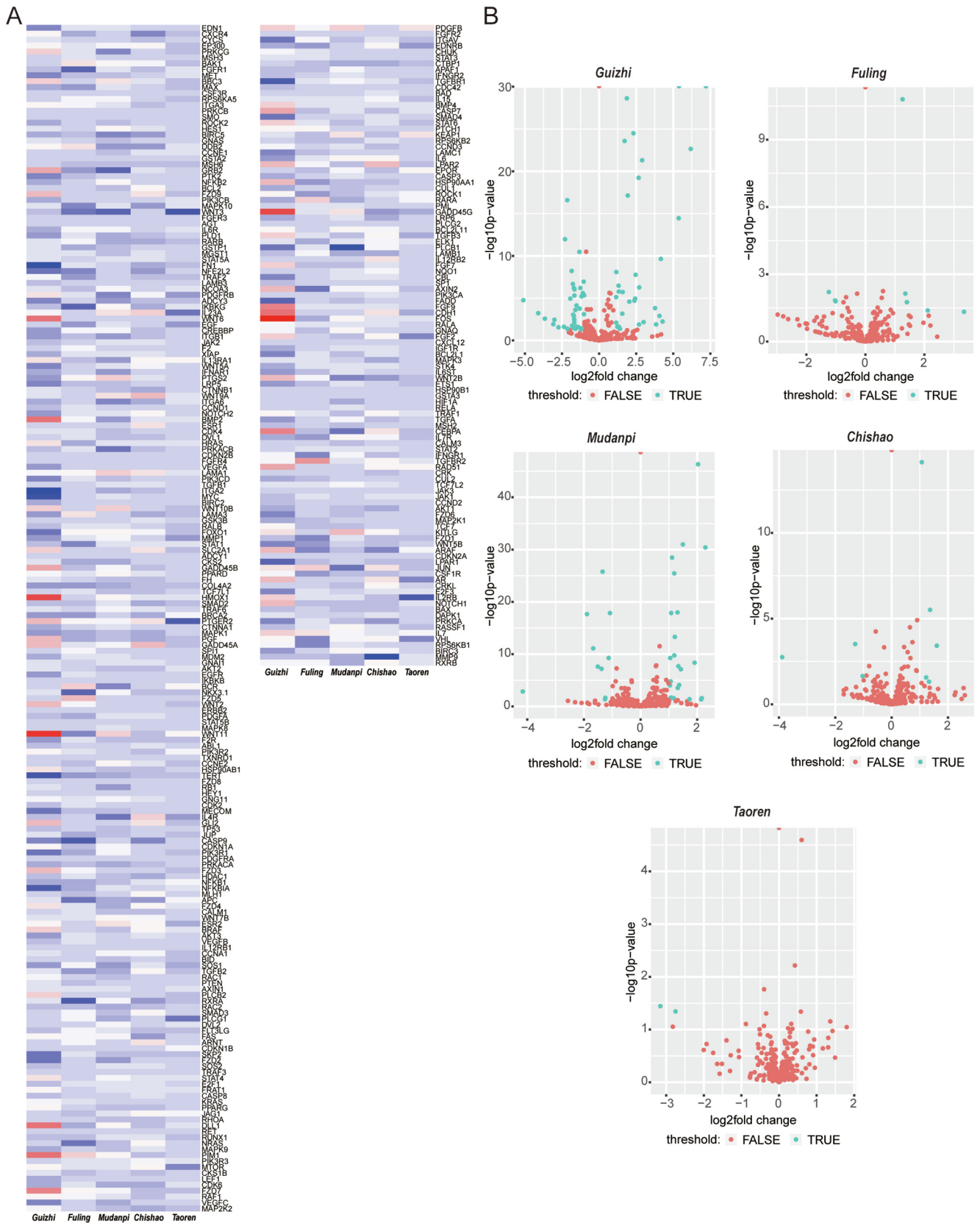
### 3.4. Mining of overlapping targets and enrichment analysis

To obtain the anti-BRCA targets in GFD, a comparative analysis was carried out for the potential targets in GFD and the therapeutic targets for BRCA. The results showed that, 38 proteins overlapped between the potential targets in GFD and the known therapeutic targets for BRCA (Table S4), which were the anti-BRCA targets in GFD with high confidence.

To further explore the biological mechanisms underlying these 38 anti-BRCA targets in GFD, the GO/KEGG pathway enrichment analysis was conducted. The top 10 most significant GO/KEGG pathways linked to these 38 targets were selected, as displayed in Fig. 2. Moreover, results of GO enrichment analysis



**Fig. 3.** Genetic alteration mining and survival analysis of 38 anti-BRCA targets in GFD based on BRCA studies in cBioPortal. (A) Overview of alterations in the 38 targets among genomic datasets available in 11 different BRCA studies. (B) Kaplan-Meier survival curve between groups with and without alterations. *P*-value < 0.05 indicated statistical significance.



**Fig. 4.** The heatmap and volcano plot of DEGs verified by HTS<sup>2</sup> assay. (A) The heatmap of DEGs. (B)The volcano plot of DEGs. The cut-off criteria were set at |Foldchange| > 2 and P-value < 0.05. The blue dots represented DEGs that reached the threshold, and the red dots stood for DEGs that did not reach the threshold. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

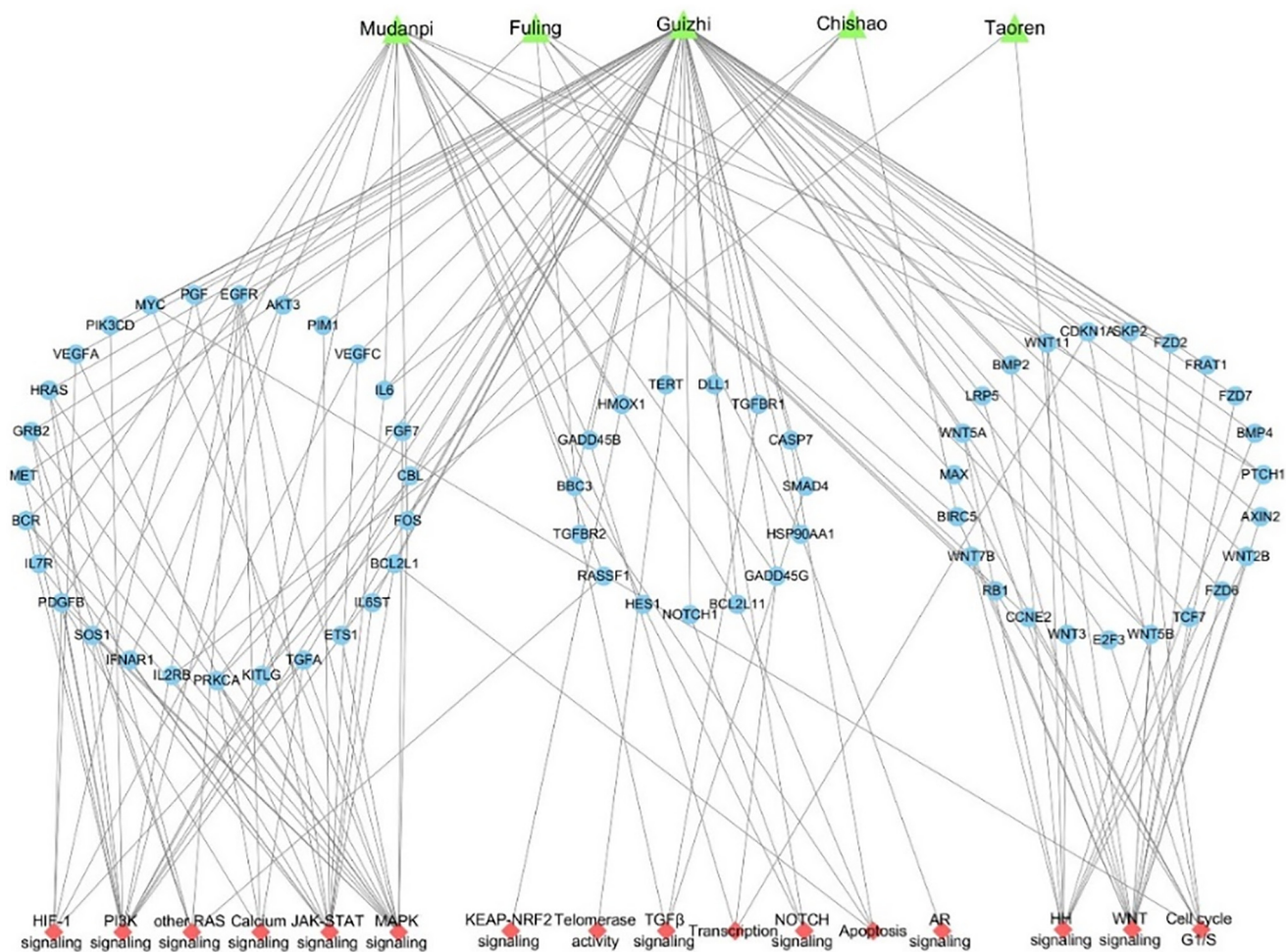


Fig. 5. The herb-target-signaling pathway interaction network.

indicated that, those targets were involved in multiple biological processes (BPs), including response to drug, negative regulation of apoptotic process, oxidation–reduction process, positive regulation of gene expression, xenobiotic metabolic process, response to estradiol, positive regulation of nitric oxide biosynthetic process, response to toxic substance, response to ethanol, and positive regulation of protein phosphorylation. Typically, response to drug, which is also referred to as drug susceptibility/resistance, is a key to the success or failure in cancer treatment [37]. Besides, the negative regulation of apoptotic process allows for the infinite cell growth and division, which is one of the fundamental factors in carcinogenesis [38]. Some evidence suggests that, the oxidation–reduction process has significant regulatory effects on tumor plasticity [39].

Further, results of KEGG enrichment analysis of the 38 targets revealed that, the anti-BRCA effect of GFD showed the highest correlation with “pathways in cancer”, followed by “microRNAs in cancer” and “proteoglycans in cancer”. Specifically, the “Pathways in cancer” involves multiple signaling pathways and it is tightly associated with cancer. Besides, cancer is also related to microRNAs (miRNAs), which have been reported to regulate the expression of various oncogenes or tumor suppressor genes and may serve as the diagnostic and prognostic biomarkers [40]. Proteoglycans exert multiple functions in cancer and angiogenesis, which is ascribed to their polyhedral nature and their capacity to interact with both ligands and receptors that regulate neoplastic growth and neovascularization [41,42].

### 3.5. Mining of genetic alterations and survival analysis

To further explore the molecular characteristics of 38 targets among different cohorts of BRCA patients, genetic alteration mining and survival analysis were performed by cBioPortal. According to the results, there were about 25%–80% genetic alterations among the 11 datasets of BRCA patients. To be specific, the genetic alterations of those 38 anti-BRCA targets in GFD included mutation, fusion, amplification, deep deletion, and multiple alterations, with mutation being the most frequently seen alteration (Fig. 3A). Notably, cases with genetic alterations were linked with poor survival compared with those without alterations (Fig. 3B). Such results suggested that, these 38 targets were closely correlated with the prognosis for BRCA, which supported the clinical application of GFD in treating BRCA.

### 3.6. Identification of DEGs by the HTS<sup>2</sup> assay

Due to the incompleteness of databases and the shortcomings of the systems pharmacology method, the predicted results might not entirely reflect the actual mechanism of GFD in treating BRCA. By means of HTS<sup>2</sup> assay, the above deficiencies might be partially corrected, and new research clues might be provided. As found in the studies mentioned above, the “pathways in cancer” was the most significant anti-BRCA signaling pathway in GFD. Thus, a total of 308 genes involved in this pathway were collected (Table S5),

and HTS<sup>2</sup> assay was carried out to detect the expression changes in these 308 genes within the BRCA MDA-MB-231 cells treated with the 5 herbs of GFD, respectively (Table S6). In the heatmap (Fig. 4A), almost all modules were blue, indicating that each of these five herbs suppressed the expression of most genes, and such finding was consistent with the anti-BRCA effect of GFD. In addition, the cut-off criteria were set as  $|FC| > 2$  and  $P$ -value  $< 0.05$  in this study to screen out DEGs. In the volcano plot (Fig. 4B), altogether 70 DEGs, including 32 up-regulated and 38 down-regulated ones, were identified in Guizhi; while 8 DEGs, including 5 up-regulated and 3 down-regulated ones, were found in Fuling. Besides, 33 DEGs were screened in Mudanpi, including 23 up-regulated and 10 down-regulated ones; whereas 8 DEGs were discovered in Chishao, including 5 up-regulated and 3 down-regulated ones. Not surprisingly, there were least DEGs in Taoren, which only included 2 down-regulated genes, with no up-regulated one. In conclusion, altogether 121 DEGs reached the threshold.

### 3.7. Interaction network construction and analysis

The “pathways in cancer” is composed by diverse signaling pathways, such as the MAPK, p53, and PI3K-Akt signaling pathways; therefore, it is needed to understand the possible underlying therapeutic mechanism further. The Cytoscape software was utilized to visualize the interaction network among the five herbs of GFD, the anti-BRCA targets and the signaling pathways from “pathways in cancer” (Fig. 5). As shown in Fig. 5, these DEGs were significantly enriched in the PI3K, WNT, MAPK, and JAK-STAT signaling pathways. These pathways have been well known to be related to cancer initiation and progression. Interestingly, these results showed that Guizhi exerted a greater role in affecting cancer targets than the other 4 herbs in GFD (Fig. 5), and such observation was consistent with the TCM theory suggesting that the monarch

drug played the most important role in treating diseases. Notably, the theory of monarch, minister, assistant, and guide embodies the basic principle regarding the composition of TCM prescriptions, and Guizhi is the monarch drug in the whole prescription of GFD.

To further explore the biological phenotypes related to the above-mentioned signaling pathways, another network was constructed to indicate the interactions among the five herbs in GFD, signaling pathways and the related biological phenotypes (Fig. 6). Unsurprisingly, as the monarch drug, Guizhi was involved in most signaling pathways. Besides, Mudanpi, which played the role of minister drug in treating BRCA, exerted a synergistic effect with Guizhi in 10 signaling pathways. In addition, Fuling, Chishao, and Taoren, which might be used as the assistant or guide drugs, mainly functioned to assist the monarch drug and the minister drug in strengthening their therapeutic effects or the guide drugs in reaching the lesions. In conclusion, all the five herbs contained in GFD possessed their specific functions and synergistically exerted their anti-BRCA effects at the molecular level.

Noteworthy, the biological phenotypes mainly focused on proliferation, evading apoptosis, and sustained angiogenesis. The above results suggested that GFD might execute the anti-BRCA therapeutic effect mainly through its anti-proliferation, pro-apoptosis, and anti-angiogenesis activities.

### 3.8. Construction of the PPI network and topological analysis

To seek the most critical anti-BRCA target of GFD, the PPI network should be analyzed based on the DEGs verified by HTS<sup>2</sup> assay. To this end, a PPI network containing 90 interactive targets was constructed (Fig. 7). Afterwards, topological analysis of this network was conducted using three centrality algorithms, including degree centrality, closeness centrality, and betweenness centrality. Moreover, these 3 algorithms were adopted to calculate the whole

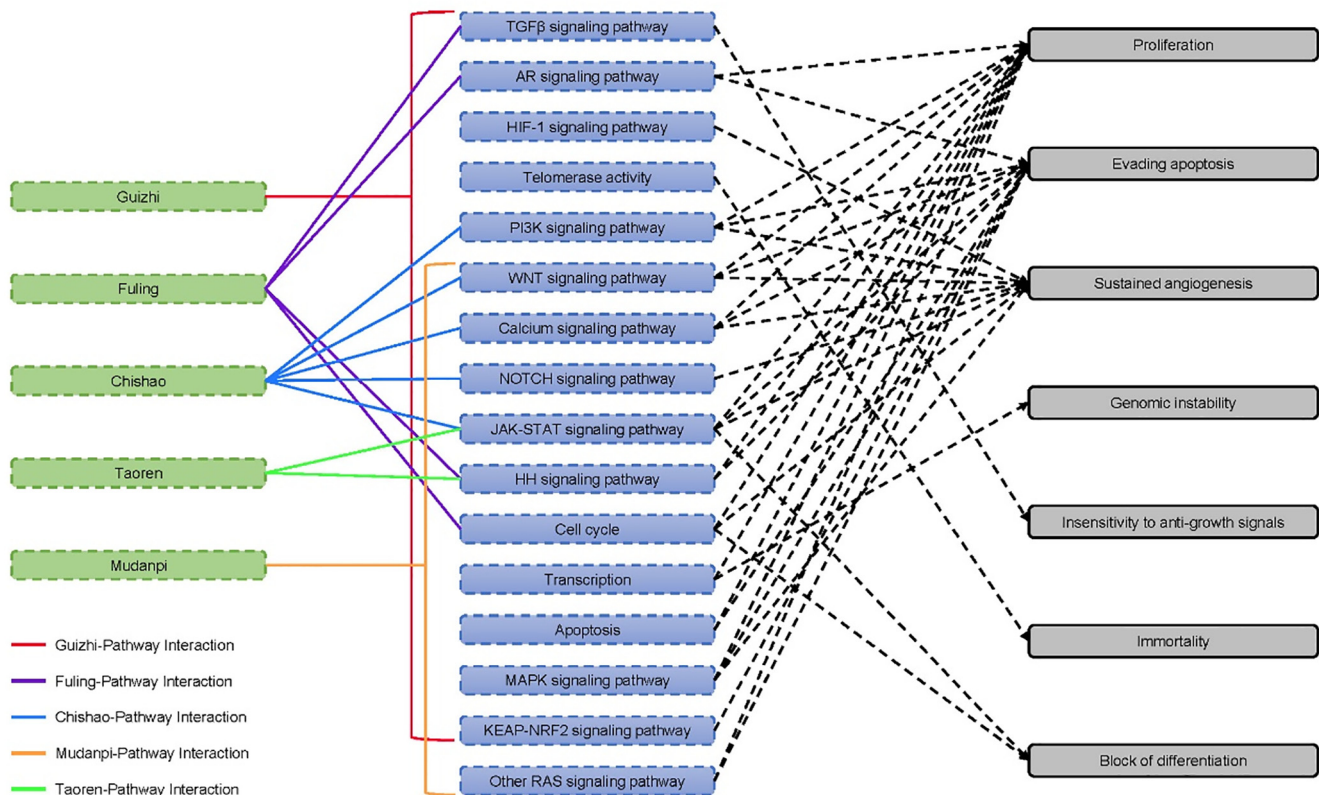
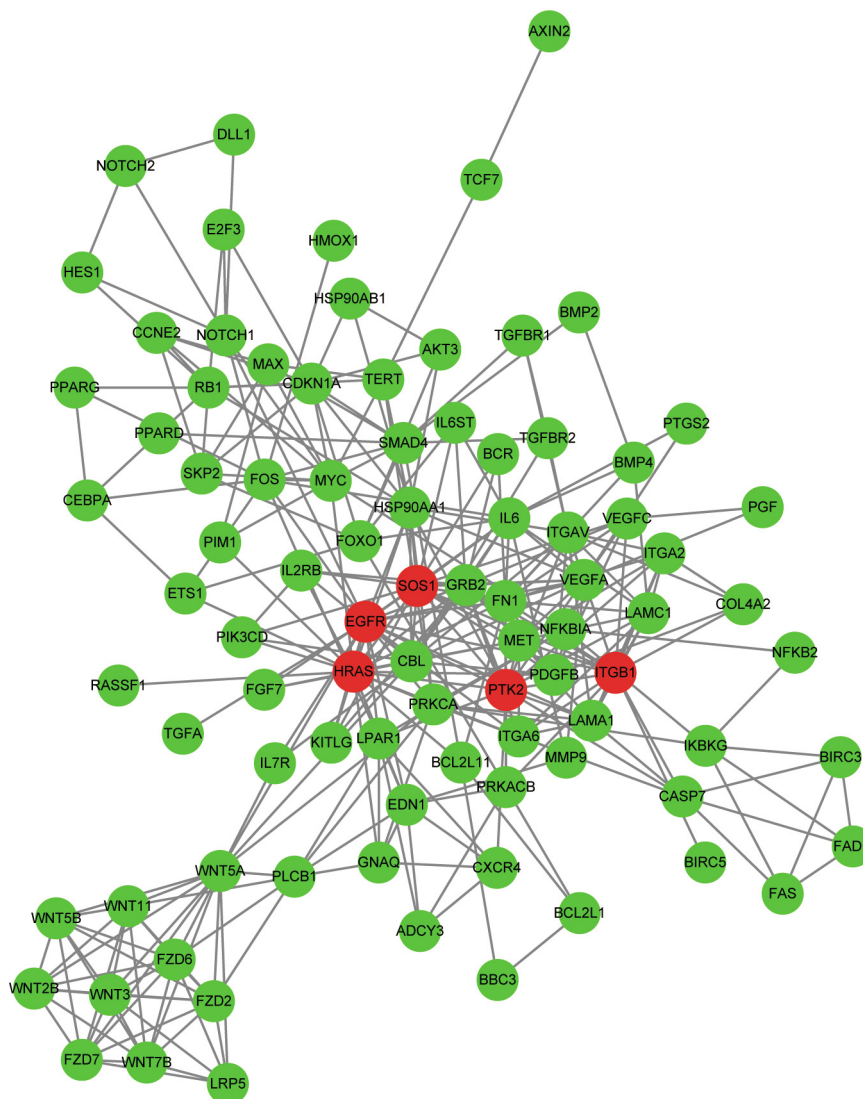


Fig. 6. The herb-signaling pathway-phenotype interaction network.





**Fig. 7.** Protein-protein interaction network. Red nodes represent the top 5 targets which were calculated by three centrality algorithms. Green nodes represent other 85 targets that make up the network. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 1**  
The topological analysis of the PPI network.

Name	Degree Centrality	Name	Betweenness Centrality	Name	Closeness Centrality
<b>EGFR</b>	21	<b>EGFR</b>	0.205218	<b>EGFR</b>	0.517442
<b>HRAS</b>	21	WNT5A	0.160696	<b>HRAS</b>	0.478495
<b>SOS1</b>	19	<b>HRAS</b>	0.139867	HSP90AA1	0.470899
<b>PTK2</b>	17	PRKCA	0.103429	<b>PTK2</b>	0.468421
<b>ITGB1</b>	17	HSP90AA1	0.100175	<b>SOS1</b>	0.451777
VEGFA	16	FOS	0.079515	PRKCA	0.449495
FN1	16	CBL	0.070176	FN1	0.445
GRB2	16	NOTCH1	0.068606	CBL	0.438424
CBL	15	MYC	0.068045	VEGFA	0.436275
WNT5A	14	SMAD4	0.063676	FOS	0.429952
HSP90AA1	13	<b>PTK2</b>	0.059873	<b>ITGB1</b>	0.429952
IL6	13	<b>ITGB1</b>	0.057459	GRB2	0.429952
ITGAV	13	TERT	0.055979	ITGA6	0.42381
PRKCA	11	IL6	0.048006	IL6	0.421801
MYC	11	VEGFA	0.047665	ITGAV	0.413953
MET	11	FN1	0.047118	MYC	0.410138
FOS	10	<b>SOS1</b>	0.040494	WNT5A	0.408257
SMAD4	10	CASP7	0.039018	LPAR1	0.408257
CDKN1A	10	BCL2L11	0.035486	LAMA1	0.406393
LAMA1	10	LPAR1	0.032512	MET	0.400901

network, and the top 20 targets were summarized according to the algorithm results (Table 1). In the degree centrality algorithm, a higher degree of a node indicated a more significant impact on the whole network. Noteworthily, the top five targets in the degree centrality algorithm were among the top 20 in both the closeness centrality and the betweenness centrality algorithms. Thus, these top 5 targets in the degree centrality algorithm, including EGFR, HRAS, SOS1, PTK2, and ITGB1, were selected as the key targets.

### 3.9. Survival analysis on the key anti-BRCA targets in GFD

To assess the potential clinical significance of the critical targets, it is necessary to analyze their possible impacts on prognosis. Therefore, survival analysis was performed on the key anti-BRCA targets in GFD using Kaplan-Meier Plotter. As shown in Fig. 8, HRAS, SOS1, PTK2, and ITGB1 were significantly related to survival, while EGFR showed no obvious correlation with survival. Meanwhile, patients with high expression of HRAS, PTK2, and ITGB1 were associated with a lower survival rate than those with low expression of these genes. Besides, BRCA with low SOS1 expression was linked with a lower risk of death, which was in line with the report that SOS1 deletion prolonged the survival of *Kras*<sup>G12D/+</sup> mice [43].

### 3.10. Molecular docking

Virtual screening is one of the rapidly developing drug screening technologies, while molecular docking is an indispensable means for active compound screening. It is helpful to calculate

the free binding energy of a compound to a target for inferring the binding stability. Generally, the lower free energy indicates the more stable binding. Table 2 shows the five compounds that best interface with the key targets. Besides, quercetin, (–)-taxifolin, kaempferol, (+)-catechin, and ellagic acid possessed the lowest binding energy towards EGFR, HRAS, SOS1, PTK2, and ITGB1, respectively.

To visualize the docking results, the 3D interaction diagrams of key targets and their corresponding best-matched compounds were drawn, as shown in Fig. 9. Clearly, compared with other key targets, quercetin displayed a high affinity towards EGFR, and the interaction diagram at the active site of target protein (Fig. 9A) revealed the formation of two hydrogen bonds with MET793. Additionally, another two hydrogen bonds formed with residues ASP855 and GLU762 also contributed to stabilizing the interaction. Further, the 3D interaction diagram of (–)-taxifolin at the active site of HRAS (Fig. 9B) was investigated, which revealed that its interaction was stable through forming hydrogen bonds with the key residues ASP30, LYS147, and LEU120. In the meantime, a pi-pi stacking interaction of PHE28 with the aromatic ring of (–)-taxifolin was also observed. Besides, kaempferol showed a high affinity towards SOS1 compared with other investigated targets, and its interaction diagram (Fig. 9C) revealed the formation of two hydrogen bonds with the residues LEU901 and ASP887, along with a pi-pi stacking interaction with TYR884. The interaction diagram of (+)-catechin at the active site of PTK2 (Fig. 9D) showed that four hydrogen bonds were formed with the residues GLU506, GLU430, GLU500, and CYS502, which contributed to stabilizing the ligand at the active site of the target protein. Similarly, ellagic

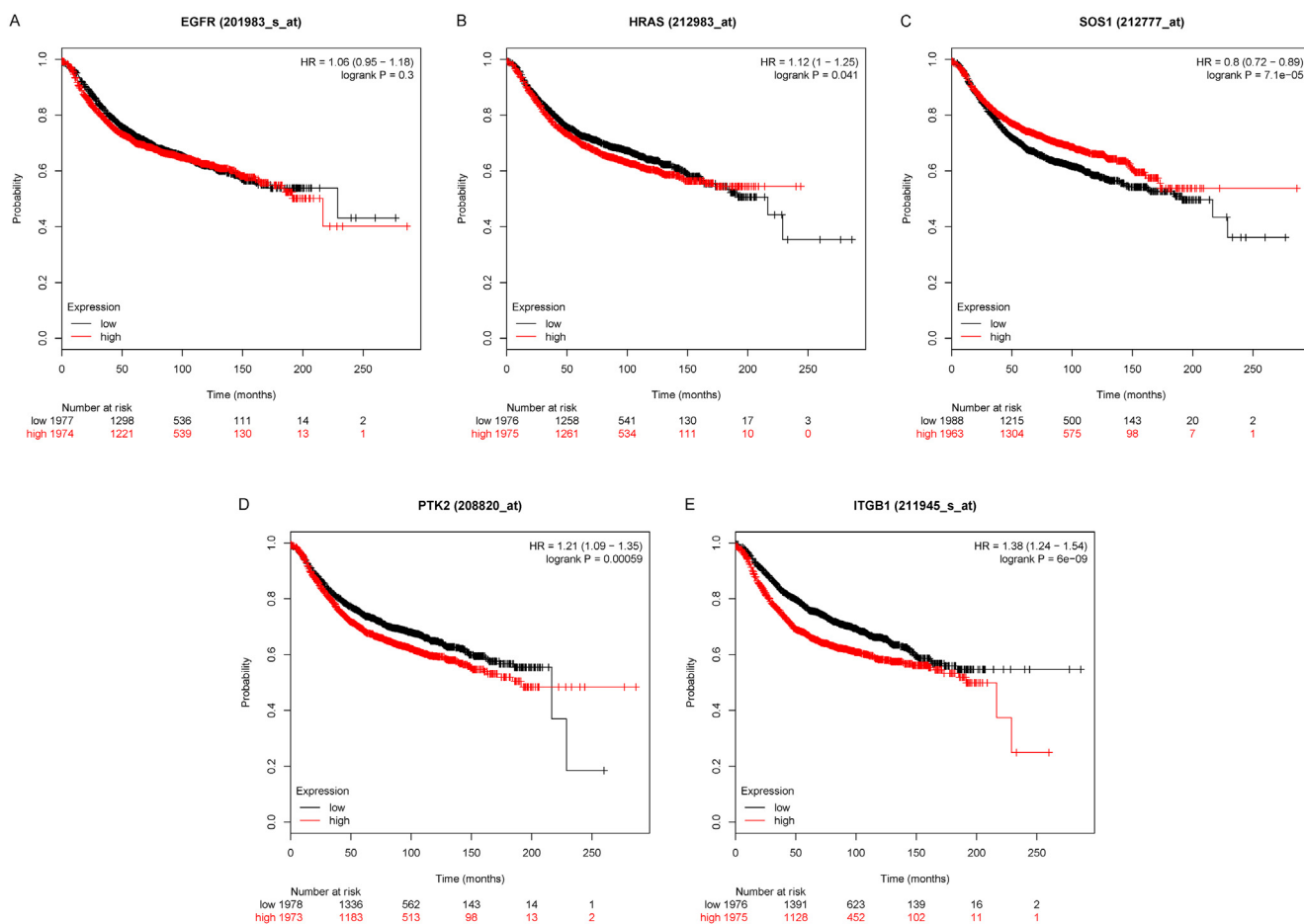


Fig. 8. Overall survival analyses on EGFR, HRAS, ITGB1, PTK2, and SOS1.

**Table 2**  
The results of molecular docking.

Mol ID	Chemical name	Glide gscore
<i>EGFR (PDB ID:6JZ0)</i>		
MOL000098	quercetin	-7.869
MOL000422	kaempferol	-7.816
MOL007022	evofolinB	-7.744
MOL007005	Albiflorin Qt	-7.647
MOL004576	taxifolin	-7.578
<i>HRAS (PDB ID:6E6C)</i>		
MOL001736	(-)-taxifolin	-7.121
MOL000073	ent-Epicatechin	-7.081
MOL007374	5-[[5-(4-methoxyphenyl)-2-furyl]methylene]barbituric acid	-6.528
MOL002776	Baicalin	-6.403
MOL004576	taxifolin	-5.771
<i>SOS1 (PDB ID:5OVE)</i>		
MOL000422	kaempferol	-6.736
MOL000098	quercetin	-6.599
MOL000073	ent-Epicatechin	-6.465
MOL004576	taxifolin	-6.374
MOL000492	(+)-catechin	-6.222
<i>PTK2 (PDB ID:618Z)</i>		
MOL000492	(+)-catechin	-8.826
MOL000098	quercetin	-8.77
MOL001736	(-)-taxifolin	-8.683
MOL004576	taxifolin	-8.507
MOL000422	kaempferol	-8.499
<i>ITGB1 (PDB ID:5XQ0)</i>		
MOL001002	ellagic acid	-7.236
MOL004576	taxifolin	-7.013
MOL007374	5-[[5-(4-methoxyphenyl)-2-furyl]methylene]barbituric acid	-6.881
MOL000422	kaempferol	-6.583
MOL000098	quercetin	-6.549

acid also relied on multiple hydrogen bonds to maintain its favorable binding with ITGB1 (Fig. 9E).

### 3.11. Molecular dynamics (MD) simulation

MD simulation has been proven to be a good strategy for improving virtual screening [44]. In order to optimize the binding pattern of compounds and key anti-BRCA targets in GFD, 100 ns MD simulations were carried out and the lowest energy structures were extracted. As shown in Fig. 10A, hydrogen bonds were formed between quercetin and four amino acid residues at the EGFR active site, including ASP 855, GLN 791, MET 793 and LEU 718. Similarly, there were also four amino acid residues (SER 17 · ASN 26 · ALA 146 and ASP 119) at the HRAS active site interacted with (-)-taxifolin to form hydrogen bonds (Fig. 10B). Further, the orthogonal view of Fig. 10C revealed the formation of hydrogen bonds between kaempferol and SOS1 with the residues ASN 879 and LYS 898, along with a pi-pi stacking interaction with PHE 890. Meanwhile, (+)-catechin maintained its favorable binding with PTK2 through forming hydrogen bonds with the residues GLU 500, SER 580, and LYS 581 (Fig. 10D). Ellagic acid interacted with ITGB1 through hydrogen bonds formed with the residues ASN 249 and GLY 569 and pi-pi stacking interaction formed with the residue TYR 277, which all contributed to the binding stability. Besides, the root-mean-square deviation (RMSD) values were shown in Fig. S1 to evaluate the overall structural stability. The results of the binding free energies of studied systems were summarized in Table S7.

### 3.12. Mechanism diagram of GFD

Based on all the above results, an integrated mechanism diagram was plotted to reflect the anti-BRCA mechanism of GFD more

clearly. As shown in Fig. 11, GFD might exert its anti-BRCA effect mainly through the PI3K and the MAPK signaling pathways. As shown in Fig. S2, the HTS<sup>2</sup> results showed that the expression of a number of genes in PI3K pathway and MAPK pathway is down-regulated in cells treated by each of these five herbs in GFD. Particularly, the expression of *PI3K*, which is a well-known key gene in PI3K pathway [45,46], was down-regulated by Guizhi, Mudanpi and Taoren, respectively. The expression of *AKT* gene, another key gene in PI3K pathway, was also down-regulated by Guizhi, Fuling, Mudanpi and Chishao, respectively. Meanwhile, among the key genes of MAPK pathway [47,48], the expression of *MEK1* was down-regulated by Guizhi and Fuling; *MEK2* was down-regulated by Chishao, Mudanpi and Taoren; and *ERK* was down-regulated by each of these five herbs.

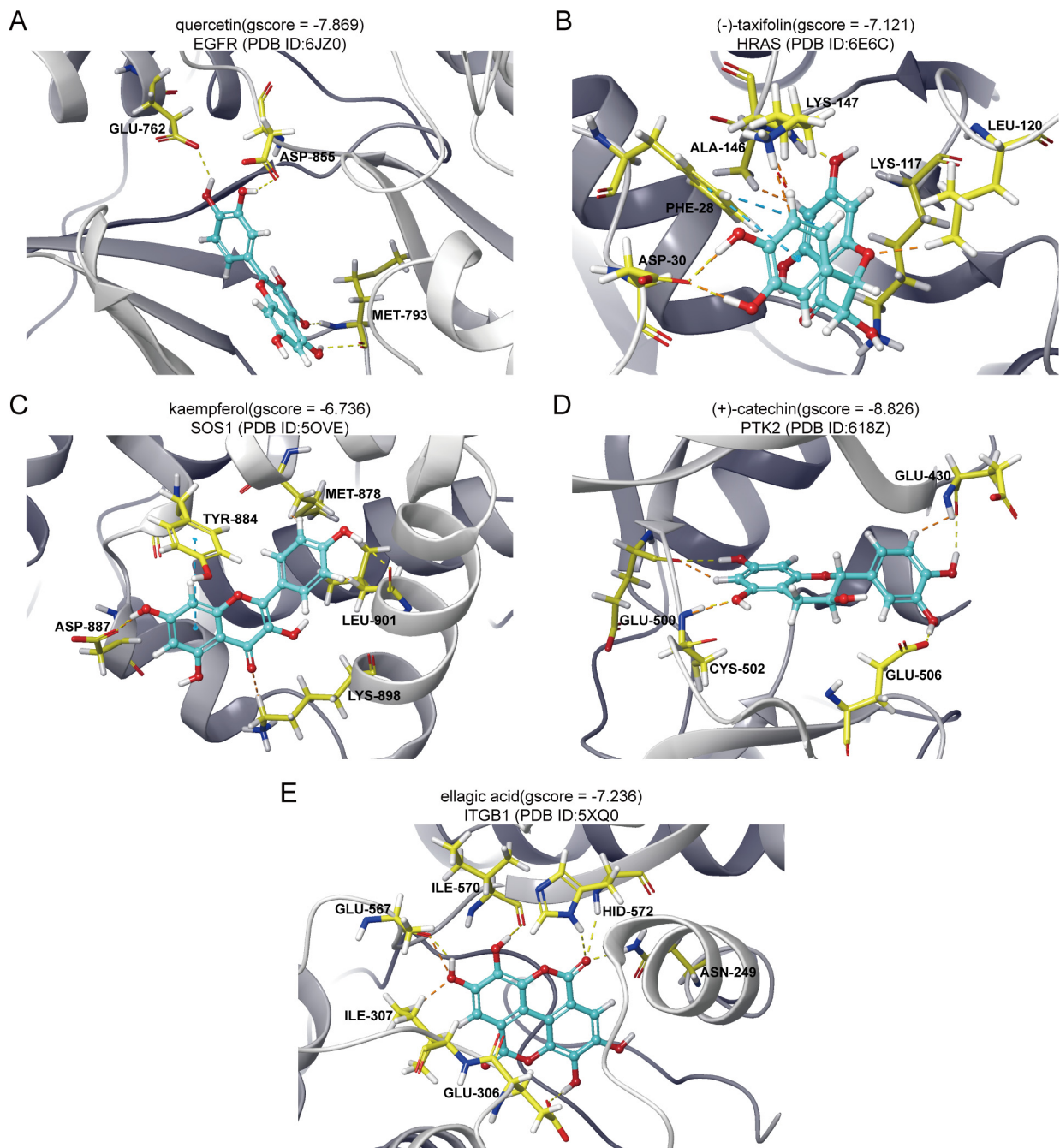
In addition, our finding is also consistent with published reports. It is reported that the crude terpene glycoside component from *Radix Paeoniae Rubra* (Chishao) could protect against isoproterenol-induced myocardial ischemic injury via activation of the PI3K/AKT/mTOR signaling pathway [49]. Moreover, GFD could sensitize the cisplatin-resistant ovarian carcinoma cell line (SKOV3/DDP) through inactivation of the PI3K/AKT/mTOR pathway [50].

Some studies have shown that the PI3K signaling pathway plays a vital role in tumor invasion and metastasis, as well as cell proliferation and apoptosis [51,52]. In this study, bioinformatic analysis was carried out, which discovered that the PI3K signaling pathway might be closely related to apoptosis, proliferation, and angiogenesis of BRCA, as verified by other literature reports [53–56]. More importantly, it has been reported that over 30% of BRCA patients harbor the *PI3KCA* gene mutations [57]. Moreover, the process responsible for BRCA is also regulated by the MAPK signaling pathway, which is involved in a series of cell physiological processes, such as cell growth, development, differentiation, and apoptosis [58–60]. In conclusion, the anti-BRCA effect of GFD might be performed via multiple key genes in the PI3K and the MAPK signaling pathways, which related to evading apoptosis, proliferation, and sustained angiogenesis.

### 3.13. Difference in the anti-BRCA targets of GFD between systems pharmacology prediction and HTS<sup>2</sup> assay verification

We predicted 38 GFD's anti-BRCA targets by systems pharmacology and discovered 5 GFD's anti-BRCA targets by the experimental HTS<sup>2</sup> assay. As shown in Fig. 12, EGFR was the only common target in both analyses. It is well known that EGFR exerts a vital role in tumor cell growth and drug resistance [61,62].

Meaningfully, the results indicate that the combined application of systems pharmacology prediction and HTS<sup>2</sup> experimental verification can help us to examine the underlying mechanisms of TCM effectively and accurately. Typically, HTS<sup>2</sup> assay can avoid the shortcomings of systems pharmacology to a certain extent. For example, we identified five key targets by HTS<sup>2</sup>, including EGFR, HRAS, ITGB1, PTK2, and SOS1, but only EGFR was predicted by systems pharmacology. Recent studies have demonstrated that somatic mutations of HRAS are closely associated with the initiation and development of several cancer types [63,64]. Furthermore, SOS1 participates in the migration and invasion of breast cancer stem cells (CSCs), as well as the occurrence and development of some cancers [65–67]. As one of the integrin subunits that regulate cell survival and proliferation, ITGB1 exerts an essential part in tumor development. Nonetheless, little research shows that linc-ITGB1 can be used as a potential prognostic biomarker for BRCA [68]. Besides, PTK2 is over-expressed in diverse primary and metastatic tumor tissues, which regulates cytoskeletal organization, as well as cell proliferation, migration, and invasion [69]. As sug-



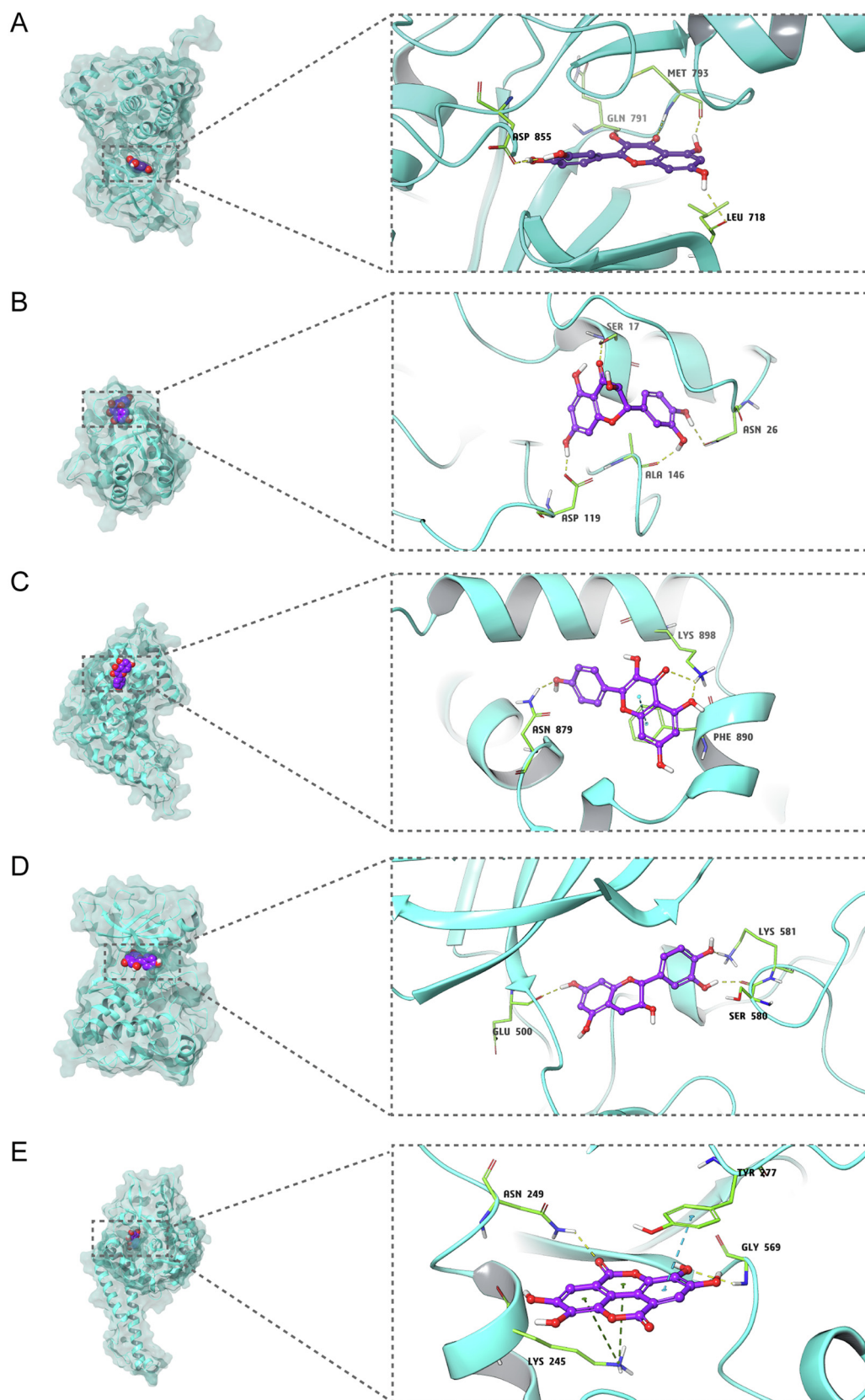
**Fig. 9.** 3D interaction diagrams of the lowest gscore chemicals in the active sites of key anti-BRCA targets in GFD. (A) 3D interaction diagram of quercetin in the active site of EGFR (PDB ID: 6JZO). (B) 3D interaction diagram of (-)-taxifolin in the active site of HRAS (PDB ID: 6E6C). (C) 3D interaction diagram of kaempferol in the active site of SOS1 (PDB ID: 5OVE). (D) 3D interaction diagram of (+)-catechin in the active site of PTK2 (PDB ID: 618Z). (E) 3D interaction diagram of ellagic acid in the active site of ITGB1 (PDB ID: 5XQ0). The lines connecting the compound to the protein are yellow for hydrogen bonds and blue for pi-pi stacking. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

gested in one study, blocking the PTK2 activity suppresses BRCA metastasis [70].

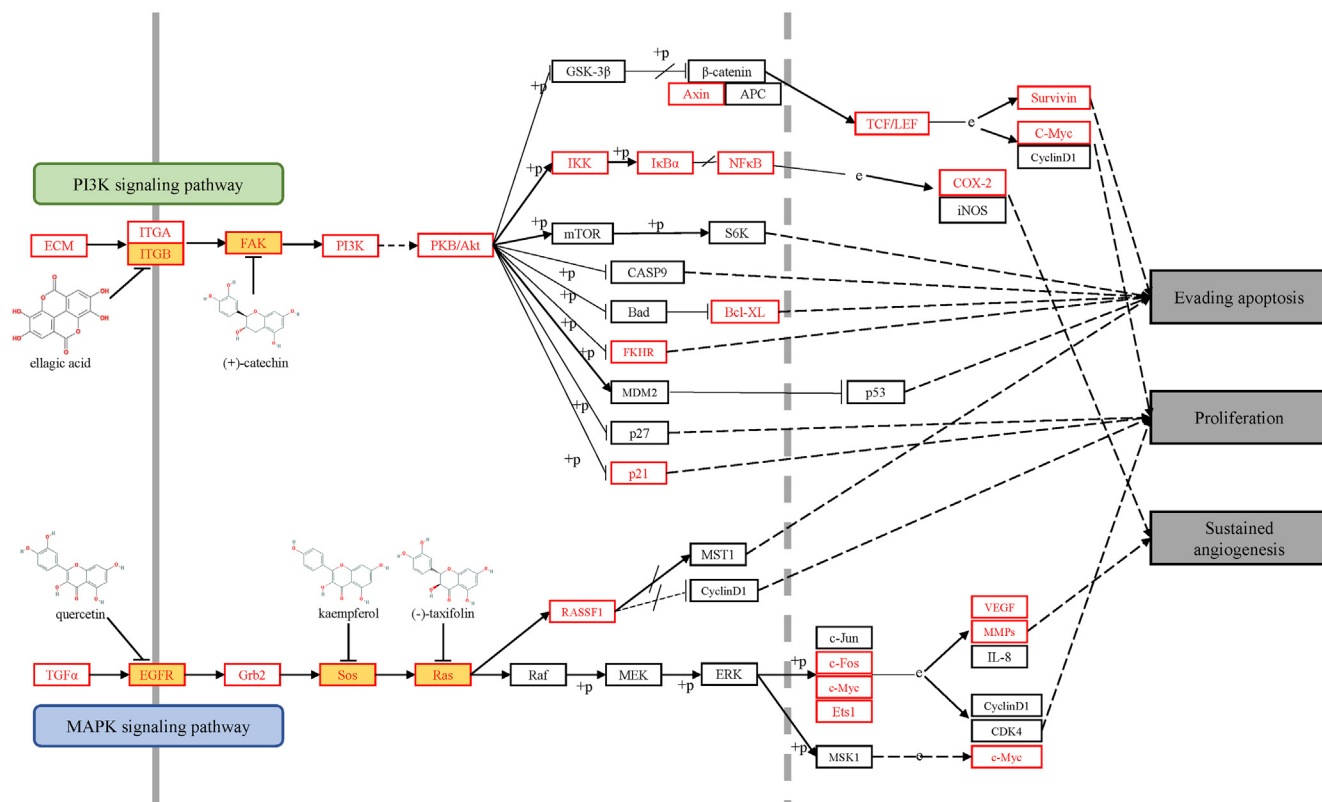
#### 4. Conclusions

In this work, systems pharmacology technologies are used in combination with the HTS<sup>2</sup> assay and bioinformatic analysis to elucidate the molecular mechanisms of GFD in treating BRCA (Fig. 13). The application of the high-throughput drug screening technology contributes to partially overcoming the deficiencies of systems

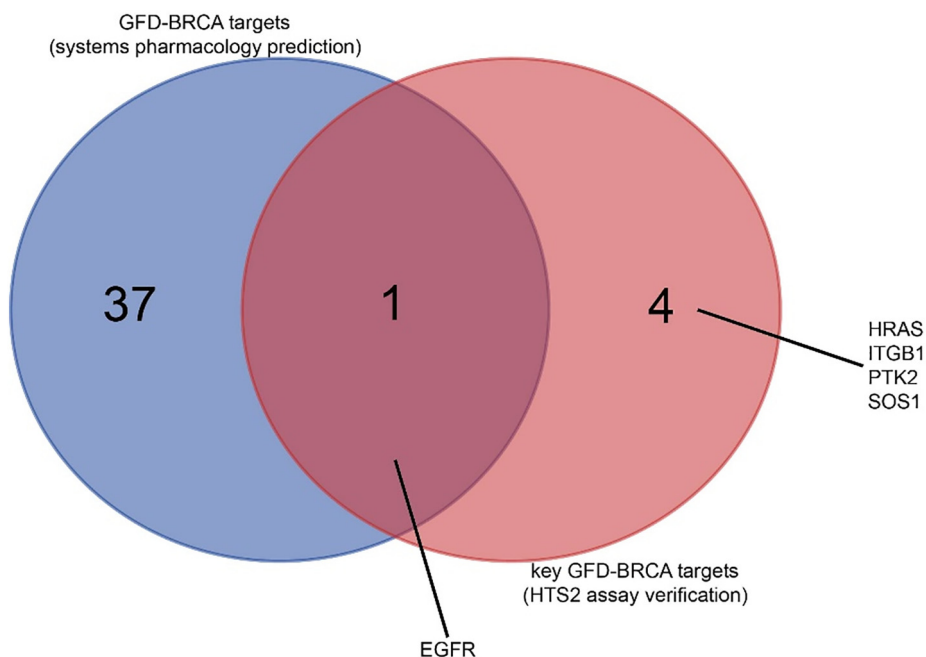
pharmacology and providing new research clues. Firstly, 76 active ingredients of GFD are screened out, followed by the identification of 38 anti-BRCA targets, with the latter are mostly enriched in the “pathways in cancer”. Secondly, the HTS<sup>2</sup> assay is conducted, and the results suggest that Guizhi plays the most important role in treating BRCA, and that Guizhi exerts a synergistic effect with Mudanpi in 10 signaling pathways. Besides, Fuling, Chishao, and Taoren play a supporting role in treating BRCA or guiding other herbs to reach the lesions. Thirdly, five key targets, including HRAS, EGFR, PTK2, SOS1, and ITGB1, are identified by network construction and topological analysis, among which, HRAS and SOS1 may



**Fig. 10.** Structures and orthogonal views of the pocket of binding between the lowest gscore chemicals and key anti-BRCA targets in the last step of molecular dynamics simulation. (A) Structures and orthogonal views of the pocket of binding between quercetin and EGFR (PDB ID: 6JZ0). (B) Structures and orthogonal views of the pocket of binding between (–)-taxifolin and HRAS (PDB ID: 6E6C). (C) Structures and orthogonal views of the pocket of binding between kaempferol and SOS1 (PDB ID: 5OVE). (D) Structures and orthogonal views of the pocket of binding between (+)-catechin and PTK2 (PDB ID: 618Z). (E) Structures and orthogonal views of the pocket of binding between ellagic acid and ITGB1 (PDB ID: 5XQ0). The lines connecting the compound to the protein are yellow for hydrogen bonds and blue for pi-pi stacking. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



**Fig. 11.** The specific BRCA-related pathways affected by the major chemicals in GFD. Arrows represent the activation effect; T-arrows stand for inhibition effect, and bias indicates interruption. In addition, the targets correlated with GFD in treating BRCA are marked in red, while the most critical targets are labeled in yellow. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



**Fig. 12.** The Venn diagram of 38 targets predicted by systems pharmacology and 5 key targets verified by the HTS<sup>2</sup> assay.

be the new clues for treating BRCA in the future. Finally, GFD may exert its anti-BRCA effect via the PI3K and the MAPK signaling pathways.

Based on the above findings, a new research paradigm combining information mining and HTS<sup>2</sup> assay is proposed, which helps to

elucidate the complex drug mechanisms better. Typically, TCM is characterized by the synergistic effects of the multi-component and multi-target drugs. Taken together, this study contributes to deepening our understanding towards TCM theory and applying this research model in interpreting other complex prescriptions.

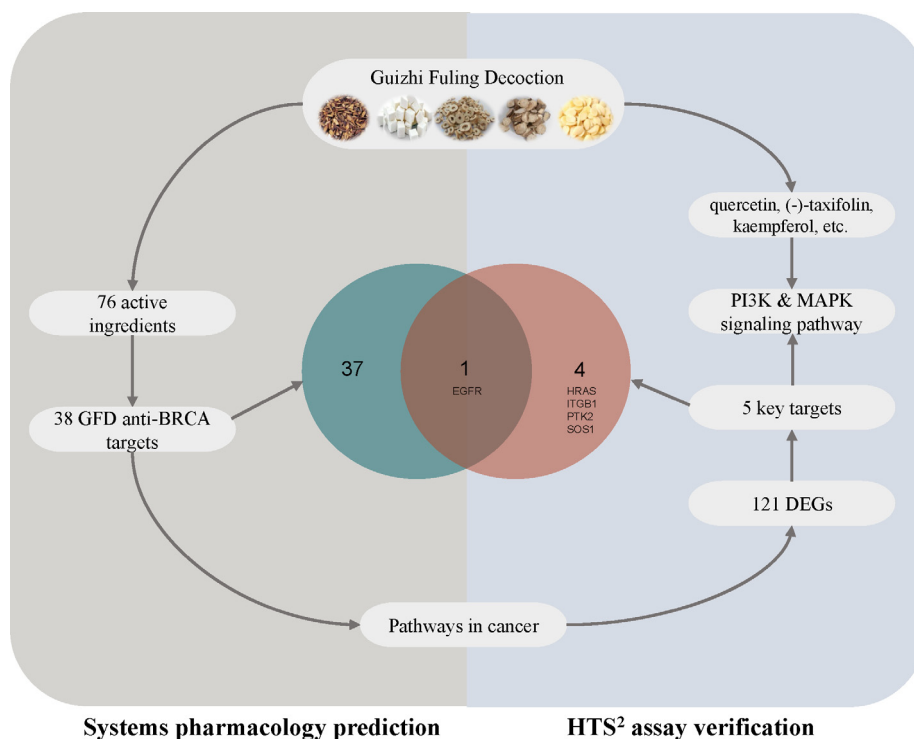


Fig. 13. Summary diagram regarding the molecular mechanisms of GFD on BRCA.

### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgments

We thank Huili Wang, Lu Li and Qianyu Wang (School of Life Sciences, Tsinghua University) for discussions. We also thank Wenting Huang (School of Medicine, Tsinghua University) for technical assistance. This work was supported by grants from the National Natural Science Foundation of China (Grant No. 81673460) and Science & Technology Department of Sichuan Province (Grant No. 2020JDTD0022).

### Appendix A. Supplementary data

Supplementary data (Table S1 Chemical information of 76 active ingredients in GFD and their parameters. Table S2 Potential targets of the active ingredients in GFD. Table S3 Detailed information of the 182 known therapeutic targets for BRCA. Table S4 Interactions between GFD and the therapeutic targets for BRCA. Table S5 308 genes involved in the “pathways in cancer”. Table S6 Raw experimental data of HTS<sup>2</sup> assay after treatments with DMSO and herb extract. Fig. S1 Root-mean-square deviation (RMSD) plots of receptors in free and bound states with ligands. Table S7 Binding affinities of compounds to proteins computed with MM-GBSA method ( $\Delta G$ : kcal/mol). Fig. S2 The volcano plot of differentially expressed genes in (DEGs) PI3K and MAPK pathway) to this article can be found online at <https://doi.org/10.1016/j.csbj.2020.05.004>.

### References

- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;68:394–424.
- Adams M, Jewell AP. The use of complementary and alternative medicine by cancer patients. *Int Semin Surg Oncol* 2007;4–10.
- Wang W, Xu L, Shen C. Effects of traditional Chinese medicine in treatment of breast cancer patients after mastectomy: A meta-analysis. *Cell Biochem Biophys* 2015;71:1299–306.
- Su Z, Li N, Cao L, Wang T, Zhang C, Ding G, et al. Main progress on studies of pharmacological activities and clinical applications of Guizhi Fuling capsule. *China J Chin Mater Med* 2015;40:989–92.
- Ziru Z, Li L, Jinhua W, Zhenzhong W, Wei X, Guanhua D. Effect of Guizhi Fuling capsule and its ingredients on human breast cancer cells proliferation. *Chin J Integr Trad West Med* 2018;38:202–7.
- Songlin M, Yan L, Tiezhu Y, Liangye L. The effect of Guizhifuling Decoction on related protein of the mitochondrial pathway of apoptosis of breast cancer cell lines and cell cycle. *Guid J Trad Chin Med Pharm* 2015;21:23–6.
- Huang C, Zheng C, Li Y, Wang Y, Lu A, Yang L. Systems pharmacology in drug discovery and therapeutic insight for herbal medicines. *Brief Bioinf* 2014;15:710–33.
- Dai Y, Sun L, Qiang W. A new strategy to uncover the anticancer mechanism of chinese compound formula by integrating systems pharmacology and bioinformatics. *Evid-Based Complementary Altern Med* 2018;2018:6707850.
- Shao W, Li S, Li L, Lin K, Liu X, Wang H, et al. Chemical genomics reveals inhibition of breast cancer lung metastasis by Ponatinib via c-Jun. *Protein Cell* 2019;10:161–77.
- Li H, Zhou H, Wang D, Qiu J, Zhou Y, Li X, et al. Versatile pathway-centric approach based on high-throughput sequencing to anticancer drug discovery. *PNAS* 2012;109:4609–11614.
- Ru J, Li P, Wang J, Zhou W, Li B, Huang C, et al. TCMSPP: a database of systems pharmacology for drug discovery from herbal medicines. *J Cheminform* 2014;6:13.
- Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, et al. PubChem substance and compound databases. *Nucleic Acids Res* 2016;44:D1202–13.
- Xu X, Zhang W, Huang C, Li Y, Yu H, Wang Y, et al. A novel chemometric method for the prediction of human oral bioavailability. *Int J Mol Sci* 2012;13:6964–82.
- Tao W, Xu X, Wang X, Li B, Wang Y, Li Y, et al. Network pharmacology-based prediction of the active ingredients and potential targets of Chinese herbal *Radix Curcumae* formula for application to cardiovascular disease. *J Ethnopharmacol* 2013;145:1–10.
- Yu H, Chen J, Xu X, Li Y, Zhao H, Fang Y, et al. A systematic prediction of multiple drug-target interactions from chemical, genomic, and pharmacological data. *PLoS ONE* 2012;7:e37608.
- Szklarczyk D, Santos A, Von Mering C, Jensen LJ, Bork P, Kuhn M. STITCH 5: Augmenting protein-chemical interaction networks with tissue and affinity data. *Nucleic Acids Res* 2016;44:D380–4.
- Yang H, Qin C, Li YH, Tao L, Zhou J, Yu CY, et al. Therapeutic target database update 2016: Enriched resource for bench to clinical drug target and targeted pathway information. *Nucleic Acids Res* 2016;44:D1069–74.

- [18] Bateman A, Martin MJ, O'Donovan C, Magrane M, Alpi E, Antunes R, et al. UniProt: The universal protein knowledgebase. *Nucleic Acids Res* 2017;46:2699.
- [19] Hamosh A, Scott AF, Amberger J, Valle D, McKusick VA. Online Mendelian Inheritance in Man (OMIM). *Hum Mutat* 2000;15:57–61.
- [20] Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Res* 2018;46:D1074–82.
- [21] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;13:2498–504.
- [22] Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009;4:44–57.
- [23] R Development Core Team R. R: A Language and Environment for Statistical Computing. 2011.
- [24] Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio Cancer Genomics Portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Disc* 2012;2:401–4.
- [25] Li H, Qiu J, Fu XD. RASL-seq for massively parallel and quantitative analysis of gene expression. *Curr Prot Mol Biol* 2012;98.4.13.1–4.13.9.
- [26] Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000;28:27–30.
- [27] Love MI, Anders S, Huber W. Differential analysis of count data – the DESeq2 package. *Genome Biol* 2014;15:10–1186.
- [28] Kolde R. pheatmap: Pretty Heatmaps. R Package Version 1.08.2015.
- [29] Ginestet C. ggplot2: elegant graphics for data analysis. *J R Stat Soc: Ser A (Stat Soc)* 2011;174:245–6.
- [30] Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* 2019;47:D607–13.
- [31] Györfy B, Lanczky A, Eklund AC, Denkert C, Budczies J, Li Q, et al. An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Res Treat* 2010;123:725–31.
- [32] Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, et al. The protein data bank. *Acta Crystallogr D Biol Crystallogr* 2002;58:899–907.
- [33] Schrödinger. Maestro | Schrödinger. Schrödinger Release 2018-1 2018.
- [34] Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA, et al. Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J Med Chem* 2006;49:6177–96.
- [35] Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. GROMACS: Fast, flexible, and free. *J Comput Chem* 2005;26:1701–18.
- [36] DeLano WL. Pymol: An open-source molecular graphics tool. *CCP4 Newsletter On Protein Crystallography* 2002;40:82–92.
- [37] Chisholm RH, Lorenzi T, Clairambault J. Cell population heterogeneity and evolution towards drug resistance in cancer: biological and mathematical assessment, theoretical treatment optimisation. *Biochim Biophys Acta – Gen Subj* 2016;1860:2627–45.
- [38] Kohoutova D, Pejchal J, Bures J. Mitotic and apoptotic activity in colorectal neoplasia. *BMC Gastroenterol* 2018;18:65.
- [39] Chio IIC, Tuveson DA. ROS in cancer: the burning question. *Trends Mol Med* 2017;23:411–29.
- [40] Reddy KB. MicroRNA (miRNA) in cancer. *Cancer Cell Int* 2015;15:38.
- [41] Iozzo RV, Sanderson RD. Proteoglycans in cancer biology, tumour microenvironment and angiogenesis. *J Cell Mol Med* 2011;15:1013–31.
- [42] Garusi E, Rossi S, Perris R. Antithetic roles of proteoglycans in cancer. *Cell Mol Life Sci* 2012;69:553–79.
- [43] You X, Kong G, Ranheim EA, Yang D, Zhou Y, Zhang J. Unique dependence on Sos1 in KrasG12D-induced leukemogenesis. *Blood* 2018;132:2575–9.
- [44] Durrant JD, McCammon JA. Molecular dynamics simulations and drug discovery. *BMC Biol* 2011;9:71.
- [45] Wu Q, Chen JX, Chen Y, Cai LL, Wang XZ, Guo WH, et al. The chemokine receptor CCR10 promotes inflammation-driven hepatocarcinogenesis via PI3K/Akt pathway activation article. *Cell Death Dis* 2018;9:232.
- [46] Yan W, Ma X, Zhao X, Zhang S. Baicalein induces apoptosis and autophagy of breast cancer cells via inhibiting PI3K/AKT pathway in vivo and vitro. *Drug Des Dev Ther* 2018;12:3961–72.
- [47] Ding K, Liao Y, Gong D, Zhao X, Ji W. Effect of long non-coding RNA H19 on oxidative stress and chemotherapy resistance of CD133+ cancer stem cells via the MAPK/ERK signaling pathway in hepatocellular carcinoma. *Biochem Biophys Res Commun* 2018;502:194–201.
- [48] Zheng G, Shen Z, Chen H, Liu J, Jiang K, Fan L, et al. Metapristone suppresses non-small cell lung cancer proliferation and metastasis via modulating RAS/RAF/MEK/MAPK signaling pathway. *Biomed Pharmacother* 2017;90:437–45.
- [49] Ke Z, Wang G, Yang L, Qiu H, Wu H, Du M, et al. Crude terpene glycoside component from *Radix paeoniae rubra* protects against isoproterenol-induced myocardial ischemic injury via activation of the PI3K/AKT/mTOR signaling pathway. *J Ethnopharmacol* 2017;12:160–9.
- [50] Han L, Guo X, Bian H, Yang L, Chen Z, Zang W, et al. Guizhi Fuling Wan, a traditional Chinese herbal formula, sensitizes cisplatin-resistant human ovarian cancer cells through inactivation of the PI3K/AKT/mTOR Pathway. *Evid-Based Complementary Altern Med* 2016;2016:4651949.
- [51] Wu X, Sun L, Wang X, Su P, Li Z, Zhang C, et al. Breast cancer invasion and metastasis by mPR $\alpha$  through the PI3K/Akt signaling pathway. *Pathol Oncol Res* 2016;22:471–6.
- [52] Fan L, Zhang Y, Zhou Q, Liu Y, Gong B, Lü J, et al. Casticin inhibits breast cancer cell migration and invasion by downregulation of PI3K/Akt signaling pathway. *Biosci Rep* 2018;38. BSR20180738.
- [53] Zhong JT, Yu J, Wang HJ, Shi Y, Zhao TS, He BX, et al. Effects of endoplasmic reticulum stress on the autophagy, apoptosis, and chemotherapy resistance of human breast cancer cells by regulating the PI3K/AKT/mTOR signaling pathway. *Tumor Biol* 2017;1–9.
- [54] Moradzadeh M, Hosseini A, Erfanian S, Rezaei H. Epigallocatechin-3-gallate promotes apoptosis in human breast cancer T47D cells through downregulation of PI3K/AKT and Telomerase. *Pharmacol Rep* 2017;69:924–8.
- [55] Miao Y, Zheng W, Li N, Su Z, Zhao L, Zhou H, et al. MicroRNA-130b targets PTEN to mediate drug resistance and proliferation of breast cancer cells via the PI3K/Akt signaling pathway. *Sci Rep* 2017;7:41942.
- [56] Yang Y, Gao M, Lin Z, Chen L, Jin Y, Zhu G, et al. DEK promoted EMT and angiogenesis through regulating PI3K/AKT/mTOR pathway in triple-negative breast cancer. *Oncotarget* 2017;8:98708–22.
- [57] Bachman KE, Argani P, Samuels Y, Silliman N, Ptak J, Szabo S, et al. The PIK3CA gene is mutated with high frequency in human breast cancers. *Cancer Biol Ther* 2004;3:772–5.
- [58] Yang XL, Liu KY, Lin FJ, Shi HM, Ou ZL. CCL28 promotes breast cancer growth and metastasis through MAPK-mediated cellular anti-apoptosis and prometastasis. *Oncol Rep* 2017;38:1393–401.
- [59] Chen X, Chen J. miR-3188 regulates cell proliferation, apoptosis, and migration in breast cancer by targeting TUSC5 and regulating the p38 MAPK signaling pathway. *Oncol Res* 2018;26:363–72.
- [60] Fan ZM, Wang DY, Yang JM, Lin ZX, Lin YX, Yang AL, et al. *Dalbergia odorifera* extract promotes angiogenesis through upregulation of VEGFRs and PI3K/MAPK signaling pathways. *J Ethnopharmacol* 2017;204:132–41.
- [61] Kjaer IM, Bechmann T, Brandslund I, Madsen JS. Prognostic and predictive value of EGFR and EGFR-ligands in blood of breast cancer patients: A systematic review. *Clin Chem Lab Med* 2018;56:688–701.
- [62] Alanazi IO, Khan Z. Understanding EGFR signaling in breast cancer and breast cancer stem cells: Overexpression and therapeutic implications. *Asian Pac J Cancer Prev* 2016;17:445–53.
- [63] Chiosea SI, Williams L, Griffith CC, Thompson LDR, Weinreb I, Bauman JE, et al. Molecular characterization of apocrine salivary duct carcinoma. *Am J Surg Pathol* 2015;39:744–52.
- [64] Chiosea SI, Miller M, Seethala RR. HRAS mutations in epithelial-myoepithelial carcinoma. *Head Neck Pathol* 2014;8:146–50.
- [65] Xiao ZH, Linghu H, Liu QF. Expressions of Ras and Sos1 in epithelial ovarian cancer tissues and their clinical significance. *J South Med Univ* 2016;36:1502–7.
- [66] Cai D, Choi PS, Gelbard M, Meyerson M. Identification and characterization of oncogenic SOS1 mutations in lung adenocarcinoma. *Mol Cancer Res* 2019;17:1002–12.
- [67] Lin C, Gao B, Yan X, Lei Z, Chen K, Li Y, et al. MicroRNA 628 suppresses migration and invasion of breast cancer stem cells through targeting SOS1. *Oncotargets Ther* 2018;11:5419–28.
- [68] Li WX, Sha RL, Bao JQ, Luan W, Su RL, Sun SR. Expression of long non-coding RNA linc-ITGB1 in breast cancer and its influence on prognosis and survival. *Eur Rev Med Pharmacol Sci* 2017;21:3397–401.
- [69] Tancioni I, Miller NLG, Uryu S, Lawson C, Jean C, Chen XL, et al. FAK activity protects nucleostemin in facilitating breast cancer spheroid and tumor growth. *Breast Cancer Res* 2015;17:47.
- [70] Woo JK, Jung HJ, Park JY, Kang JH, Il Lee B, Shin DY, et al. Daurinol blocks breast and lung cancer metastasis and development by inhibition of focal adhesion kinase. (FAK). *Oncotarget* 2017;8:57058–71.