

# The bioinformatics analysis of quercetin in octagonal lotus for the screening of breast cancer MYC, CXCL10, CXCL11, and E2F1

International Journal of  
Immunopathology and Pharmacology  
Volume 35: 1–17  
© The Author(s) 2021  
Article reuse guidelines:  
[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)  
DOI: 10.1177/20587384211040903  
[journals.sagepub.com/home/iji](https://journals.sagepub.com/home/iji)  
SAGE

Yuexing Ma<sup>1,2,3,4,\*</sup> , Zirong Peng<sup>2,3,4</sup>, Rongbin Pan<sup>2,5,\*</sup>, Zhixin Zhu<sup>2,3,4</sup>,  
Xiaoqi Meng<sup>2,3,4</sup>, Huiming Hu<sup>2,3,4</sup>, Xin Qiao<sup>2</sup>, Xuening Huang<sup>2</sup> and  
Mengyu Hou<sup>2</sup>

## Abstract

**Background:** Comprehensive bioinformatics analysis of the effective molecular screening of Podophyllum octagonal in breast cancer treatment by using network pharmacology.

**Methods:** We collected the active ingredients and target genes of Chinese medicine octagonal lotus through the Traditional Chinese Medicine System Pharmacology Analysis Platform (TCMSP); downloaded human protein annotation information on the protein database Uniport; and collected data from five databases: GeneCards, OMIM, PharmGkb, TDD, and DrugBank. Construct the practical ingredient-target gene data intersection to obtain the target gene-disease gene and draw the Venn diagram. We use Cytoscape 3.8.0 software to construct the effective component-target gene-disease gene network. The STRING database protein interaction (PPI) networks were erected, and we used Cytoscape 3.8.0 software to screen out its core sub-networks and hub gene networks. Through survival analysis, core genes and hub genes were screened to identify several key genes. We performed key target gene ontology (GO) analysis and gene interaction (KEGG) analysis, which were followed by molecular docking of the key active ingredients in the star anise corresponding to several key genes.

**Results:** 19 active ingredients, 444 drug targets, and 10,941 disease-related genes were obtained. The key active ingredient was quercetin. GO analysis revealed 2471 affected biological processes, and 167 pathways were obtained in KEGG enrichment analysis.

**Conclusion:** This study initially screened the key active ingredients of star anise lotus and analyzed key genes and several essential pathways. Traditional Chinese medicine is expected to provide new evidence and research ideas to prevent and treat breast cancer.

## Keywords

Octagonal lotus, breast cancer, network pharmacology, quercetin, MYC, CXCL10, CXCL11, E2F1

Date received: 2 August 2021; accepted: 3 August 2021

## Background

Breast cancer (BC) is a pathological process in which breast epithelial cells proliferate out of control under the action of a variety of carcinogens. The global incidence of breast cancer has been on the rise since the late 1970s, and

it is a significant cause of cancer-related deaths among women. The situation of breast cancer in China is not optimistic. More than 16,000 people are diagnosed with breast cancer every year, and 12,000 people die from breast cancer. At present, the cause of breast cancer is not clear. Studies have found that breast cancer incidence has a



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

particular pattern, among which risk factors are more important. Most breast cancer patients' risk factors are called high-risk factors of breast cancer, including aging, family history, reproductive factors (menarche, low age at first pregnancy, late menopause, etc.), estrogen (endogenous and exogenous estrogen), bad lifestyle habits, etc. Most of them involve changes in the expression of certain genes, such as microRNAs (miRNAs).<sup>1</sup> However, women with several of the abovementioned high-risk factors do not necessarily have breast cancer. It just means that their risk of breast cancer is higher than ordinary people. More and more evidence shows that lncRNA participates in the formation of breast cancer by influencing Notch, Wnt/ $\beta$ -catenin, transforming growth factor- $\beta$ , and mitogen-activated protein kinase pathways and this kind of participation has important clinical diagnostic and prognostic value.<sup>2</sup> So far, significant progress has been made in the clinical and theoretical research of breast cancer. Current prevention methods include screening, chemoprevention, and biological prevention, which are more direct and effective than ever.

The mortality rate of breast cancer has decreased. However, breast cancer remains the leading cause of cancer death among women. Therefore, more input is needed to study breast cancer's pathogenesis and prevention.<sup>1-6</sup> Podophyllum was first published in "Shen Nong's Materia Medica." The octagonal genus lotus is a genus endemic to our country. It is the rhizome of hexagonal lotus and octagonal lotus in the Berberis family.<sup>5</sup> Its effects involve clearing heat and detoxification, removing phlegm, removing blood stasis, and reducing swelling.<sup>6</sup> Modern pharmacological research shows that star aniseed lotus is widely used in cancer treatment (such as breast cancer, testicular cancer, etc.).<sup>7,8</sup>

In this article, network pharmacology is used to obtain the targets and active ingredients of octagonal lotus from the TCMSP database and obtain breast cancer disease-related genes from GeneCards, OMIM, PharmGkb, TDD, and DrugBank to establish a dynamic ingredient-

target gene network. We establish the PPI network of the octagonal lotus' key targets through the STRING database, and the core genes and pivot genes were screened for survival analysis and verification. GO enrichment analysis and KEGG pathway enrichment analysis were performed on key targets. Finally, the selected essential genes are molecularly docked with effective ingredients. In this study, four key genes (MYC, CXCL10, CXCL11, and E2F1) affected by the octagonal lotus were initially screened. Quercetin is speculated to be the main effective component of octagonal lotus to treat and prevent breast cancer. Through GO enrichment analysis and KEGG pathway enrichment analysis, we have obtained important pathways of star aniseed lotus and discussed the relationship among key genes, active ingredients, pathways, and breast cancer, which, to a certain extent, provide a relevant reference to the prevention and treatment of breast cancer with traditional Chinese medicine.

## Material and methods

### Acquisition of active ingredients and targets

We use the Traditional Chinese Medicine System Pharmacology Database (TCMSP) (<https://tcmospw.com/tcmosp.php>) to search for the dataset of the chemical constituents and targets of octagonal lotus. We use drug-like properties (DL)  $\geq 0.18$  and bioavailability (OB)  $\geq 30\%$  as the evaluation criteria to screen the effective ingredients in the octagonal lotus, and obtain the target of the octagonal lotus, and then use Strawberry Perl-5.30.0.1 to act on it. The target is screened, and the target of the active ingredient is obtained. The human protein annotation information is obtained from the Uniprot website (<https://www.uniprot.org/>) to convert the star aniseed lotus target (Annex.1).

### The collection of BC genes

We use "Breast Cancer" as a keyword in GeneCards (<https://www.genecards.org/>), OMIM (<https://omim.org/>), PharmGkb (<https://www.pharmgkb.org/>), TDD (<http://db.idrblab.net/ttd/>), and DrugBank (<https://go.drugbank.com/>); these five disease gene databases collect breast cancer-related genes. All genes were integrated to remove duplicate data to obtain the set of disease genes for breast cancer, and we use the software R 4.0.2 and its package ("Venn") to draw the union Venn diagram (Venn) (Figure 1(a)).

### Dysosma Versipellis targets and construction of BC-related gene network

We use the software R 4.0.2 to pair the targets corresponding to the active ingredients of the octagonal lotus with the collection of disease genes for breast cancer to obtain the

<sup>1</sup>China Pharmaceutical University, Nanjing, China

<sup>2</sup>Jiangxi University of Chinese Medicine, Nanchang, China

<sup>3</sup>Science and Technology College, Jiangxi University of Chinese Medicine, Nanchang, China

<sup>4</sup>Nanchang Medical College, Nanchang, China

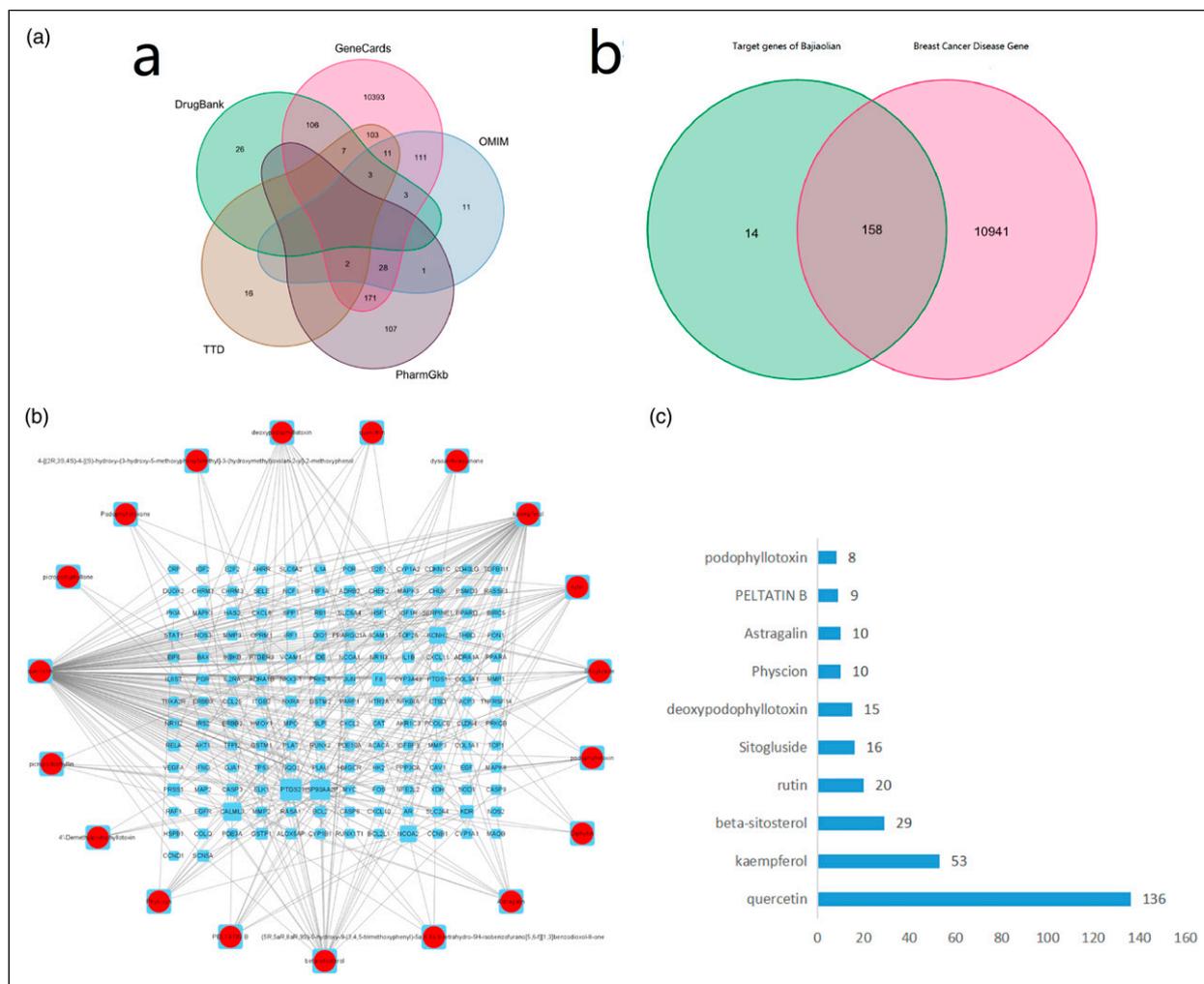
<sup>5</sup>Jiangzhong Cancer Research Center, Jiangxi University of Chinese Medicine, Nanchang, China

\*Co-corresponding author

#### Corresponding authors:

Yuexing Ma, China Pharmaceutical University, 639 Longmian Avenue, Jiangning District, Nanjing 330000, China.  
Email: [ma-yuexing@qq.com](mailto:ma-yuexing@qq.com)

Rongbin Pan, Jiangzhong Cancer Research Center, Jiangxi University of Traditional Chinese Medicine, 1688 Meiling Road, Nanchang 330004, Jiangxi, China.  
Email: [PRB2019@jxutcm.edu.cn](mailto:PRB2019@jxutcm.edu.cn)



**Figure 1.** A: Dysosma Versipellis-BC Venn diagram. (a) is the Venn diagram of the breast cancer data set in the five gene databases of GeneCards, OMIM, PharmGkb, TDD, and DrugBank, and (b) is the Venn diagram of the intersection of the target of the Dysosma Versipellis active ingredient and the breast cancer data set. B: Dysosma Versipellis’s active ingredients in BC target gene network diagram. The red node represents the affective component, and the blue node represents the target. The larger the node, the more complex the connection relationship. C: Active ingredients-the top 10 active ingredients with the number of connections in the target gene network diagram.

intersection genes of the two and use the R package (“Venn”) to draw the Venn diagram (Venn) for the intersection (Figure 1(b)). We then use Cytoscape 3.8.0 software to construct an effective gene network of ingredient-target gene-disease (Figure 1(b)).

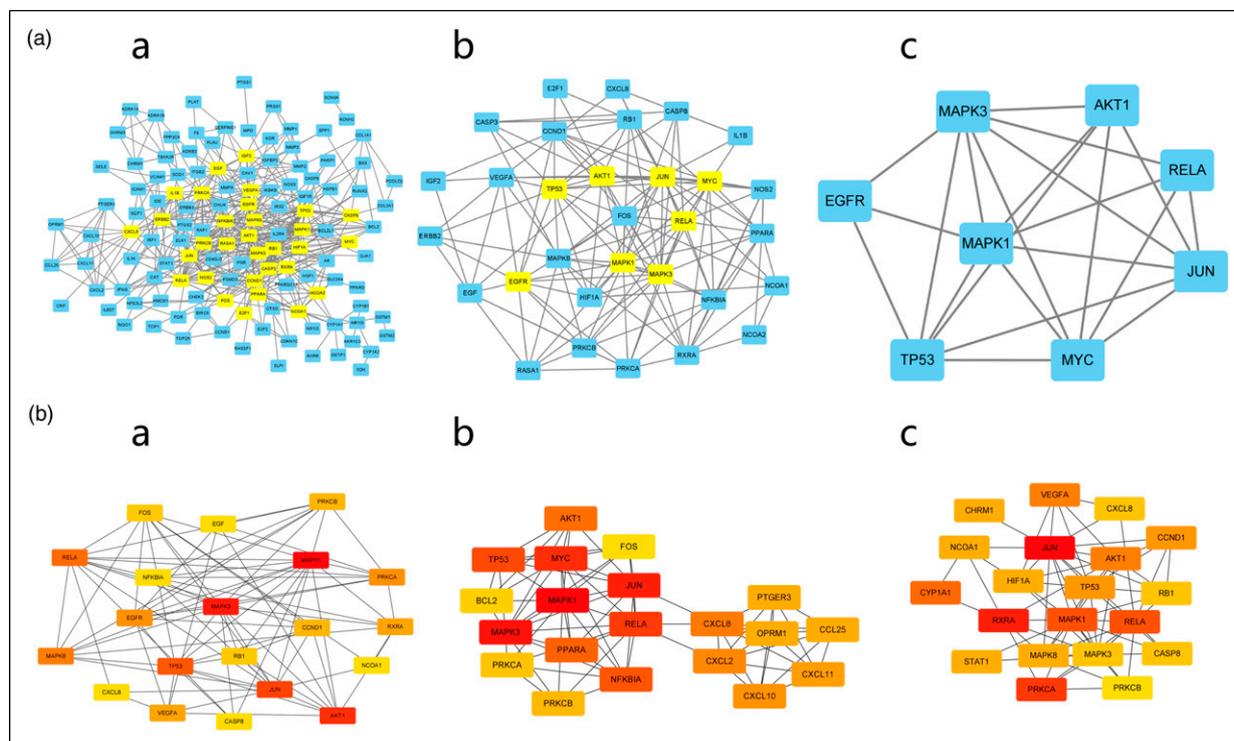
### Construction of PPI network

We import the key target genes into the STRING database (<https://string-db.org/>) to construct a protein interaction network (PPI), select *Homo sapiens* as the research species, set the required minimum interaction score as the highest confidence degree [highest confidence (0.900)], and hide disconnected nodes in the network. Other parameters are the default values to obtain the protein interaction network of octagonal lotus for breast cancer (Figure 2).

### Screening of core genes and pivot genes

**Screening of core genes:** Import the data of the constructed protein interaction network (PPI) into Cytoscape 3.8.0, use the plug-in CytoNCA to filter the network to obtain the first-level sub-network, then filter the first-level sub-network to get the second-level sub-network (Figure 3(a)), and select genes in the first-level sub-network as core genes. **Screening of hub genes:** Import the protein interaction network (PPI) data into Cytoscape 3.8.0 and use the plug-in CytoHubba to screen the hubs. Here, three scores of Degree, MCC, and Bottle Neck are used as the screening basis; the top 20 hubs are selected; genes were screened for three scores; and three pivot gene groups were obtained (Figure 3(b)).





**Figure 3.** A: *Dysosma Versipellis*-breast cancer protein network diagram (a) is the protein network diagram, b is the first-level sub-network diagram of the core network, and c is the second-level sub-network diagram of the core network. B: *Dysosma Versipellis*-BC hub gene network diagram (a) is the Degree scoring hub gene network diagram, (b) is the MCC scoring hub gene network diagram, c is the Bottle Neck scoring hub gene network diagram.

right side are the key genes. The cancer genomics visualization tool is used to visually summarize several genes (cBioPortal) (<https://www.cbioportal.org/>) whose logarithmic rank  $p$ -value is less than 0.05. We use the GEPIA2 online database (<http://gepia.cancer-pku.cn/index.html>) to verify genes with  $p$ -value <0.05, select boxplot to add dataset BRCA, set  $|\text{Log}_2\text{FC}|$  cutoff values and  $p$ -value cutoff value to the default values of one, and set 0.01 for gene analysis. We use GEPIA2 online database (<http://gepia.cancer-pku.cn/index.html>) to verify genes with  $p$ -value <0.05, select boxplot to add dataset BRCA, set  $|\text{Log}_2\text{FC}|$  cutoff values and  $p$ -value cutoff value as default values of one, set 0.01 to analyze the genes, select stage plot to add the dataset BRCA, and then analyze to get the analysis of the genes. The survival analysis and verification of the three pivot gene groups were carried out in the same way.

### Gene copy number variation rate

The METABRIC data in cBioportal ([https://www.cbioportal.org/study/summary?id=brca\\_metabric](https://www.cbioportal.org/study/summary?id=brca_metabric)) were used to perform gene copy number variation rate analysis (CNA) on the screened core genes and pivot genes. Several evaluations of genes were carried out through the CNA genome group: ① Survival analysis of multiple genes in the total breast patient population, ② survival analysis of multiple genes in breast

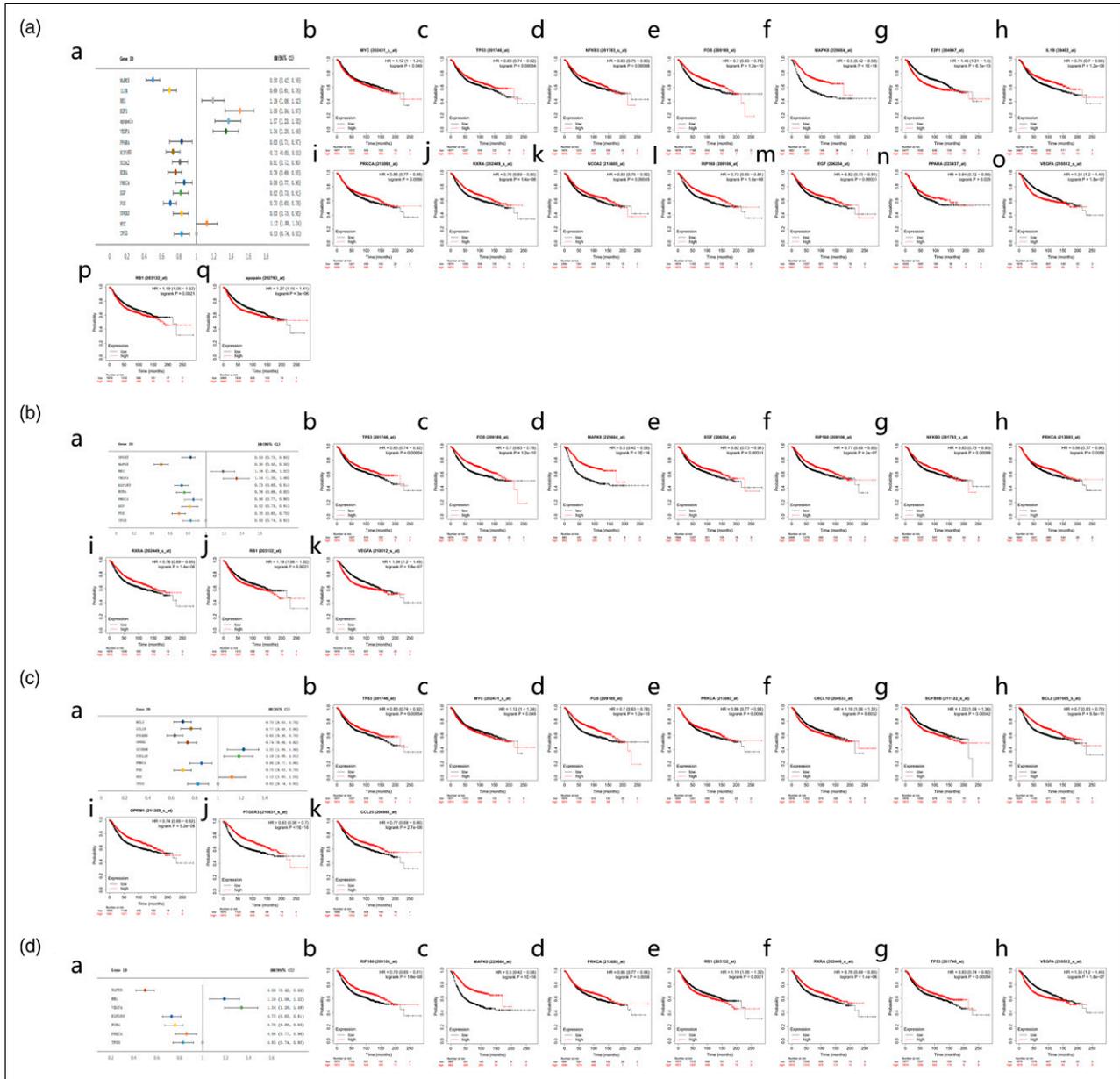
cancer patients without recurrence, ③ age at diagnosis, ④ Nottingham prognostic index, ⑤ integrative cluster, ⑥ patient's vital status, and ⑦ tumor and other histologic subtype.

### Key gene ontology (GO) analysis and gene interaction (KEGG) pathway analysis

We first prepare packages such as “colorspace,” “stringi,” “ggplot2,” “BiocManager,” “DOSE,” “clusterProfiler,” “enrichplot,” “org.Hs.eg.db,” and “pathview”; then we use R 4.0.2 and reference packages such as “clusterProfiler,” “org.Hs.eg.db,” “enrichplot,” and “ggplot2” to perform GO enrichment analysis on key genes; and use packages such as “clusterProfiler,” “org.Hs.eg.db,” “enrichplot,” “ggplot2,” and “pathview” for KEGG pathway enrichment analysis; the results are shown in Figure 7(b) (Annex.2).

### Molecular docking with key active ingredients

The key genes are converted into key active ingredients in the active ingredient-target gene-disease gene network. Preparation of small molecule ligand: Collect the small molecule ligand 2D structure of the key active ingredient in PubChem database (<https://pubchem.ncbi.nlm.nih.gov/>) and import the small molecule ligand 2D



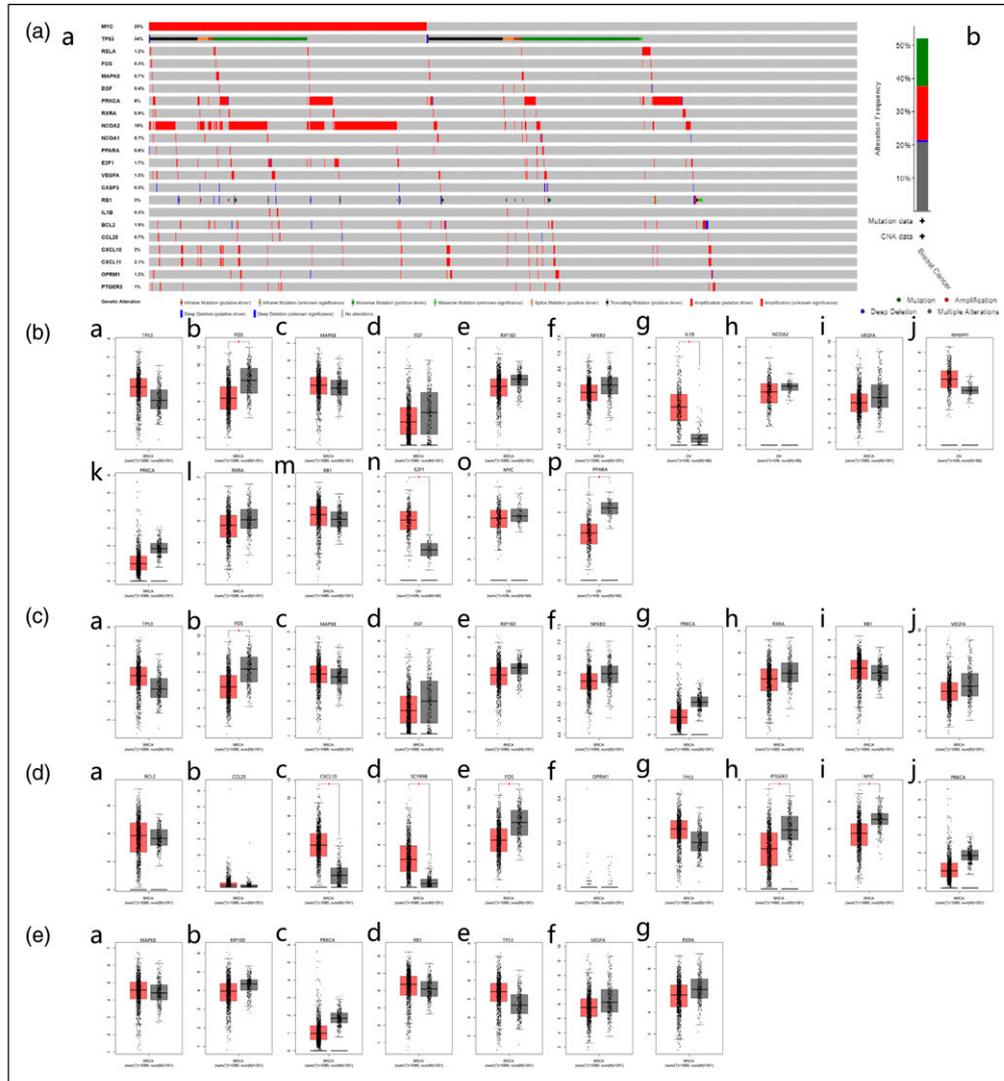
**Figure 4.** A: (a) is a forest map of 16 core genes, (b)-(q) is a survival analysis map of 16 core genes. B: (a) is the forest plot of the pivot gene's degree score, and (b)-(k) is the survival analysis plot of the degree score of the pivot gene. C: (a) is the forest plot of the pivot gene MCC score, (b)-(k) is the survival analysis plot of the pivot gene MCC score of the pivot genes. D: (a) is the forest plot of the Hub gene Bottle Neck score, (b)-(h) is the survival analysis plot of the Bottle Neck score of the pivot genes.

structure of Chem3D to calculate the 3D structure from which its minimum free energy is derived. Preparation of protein receptor: Find the human protein ID corresponding to the key gene in Uniprot, enter the protein ID in the PDB database (<http://www.rcsb.org/>) to search for the corresponding protein receptor, and use PyMOL to remove the protein receptor water molecules and small molecule ligands. Determine the activity pocket: Use the software AutoDockTools-1.5.6 to determine the activity range of the protein receptor. Molecular docking: Use vina for molecular docking; then, the docking

results are imported into PyMOL to get the effect picture of molecular docking (Figure 7(c)).

#### Verification of the differential gene in the TIMER database

Enter the TIMER database (Tumor Immune Estimation Resource) (<https://cistrome.shinyapps.io/timer/>), search for the differential expression of key genes (MYC, CXCL10, CXCL11, and E2F1) in different tumor tissues, and verify

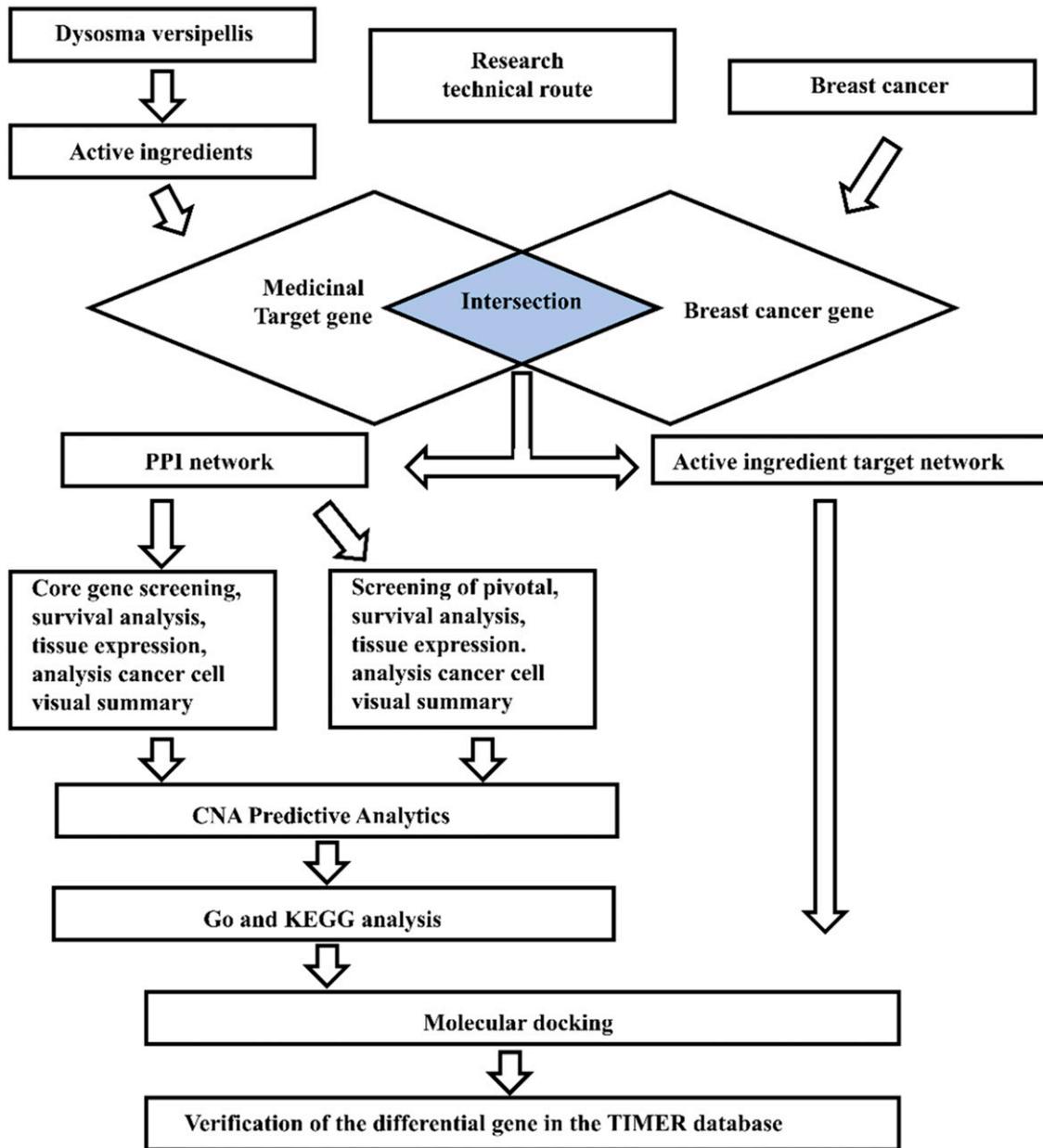


**Figure 5.** A. Visual summary and overview of 22 significant genes. (a) is a visual summary diagram of 22 significant genes in the survival analysis, (b) is a summary diagram of changes in 22 significant genes. B. The expression level of hub gene in BC tissues and normal tissues of patients (a)-(p) is the expression level of core genes ( $p < 0.01$  is considered statistically significant). C. The expression level of hub gene in BC tissues and normal tissues of patients (a)-(j) are the expression levels of the degree score of the pivot gene ( $p < 0.01$  is considered statistically significant). D. The expression level of hub gene in BC tissues and normal tissues of patients. (a)-(j) are the expression levels of the pivot gene MCC score ( $p < 0.01$  is considered statistically significant). E. The expression level of hub gene in BC tissues and normal tissues of patients. (a)-(g) are the expression levels of the hub gene Bottleneck score ( $p < 0.01$  is considered statistically significant).

the copy number of key genes. The correlation between changes and the abundance of immune cells, to prove the relationship between survival analysis and immune infiltration, and the correlation between TP53 gene mutation and immune infiltration are as follows:

- (1) Differential gene expression in different tumor tissues (Figure 8(a));

- (2) Correlation between differential gene copies number changes and immune cell abundance (Figures 8B and 8C);
- (3) The relationship between survival analysis and immunity infiltration (Figure 9); and
- (4) The relationship between TP53 gene mutation and immune infiltration which is shown in Figure 10.



## Results

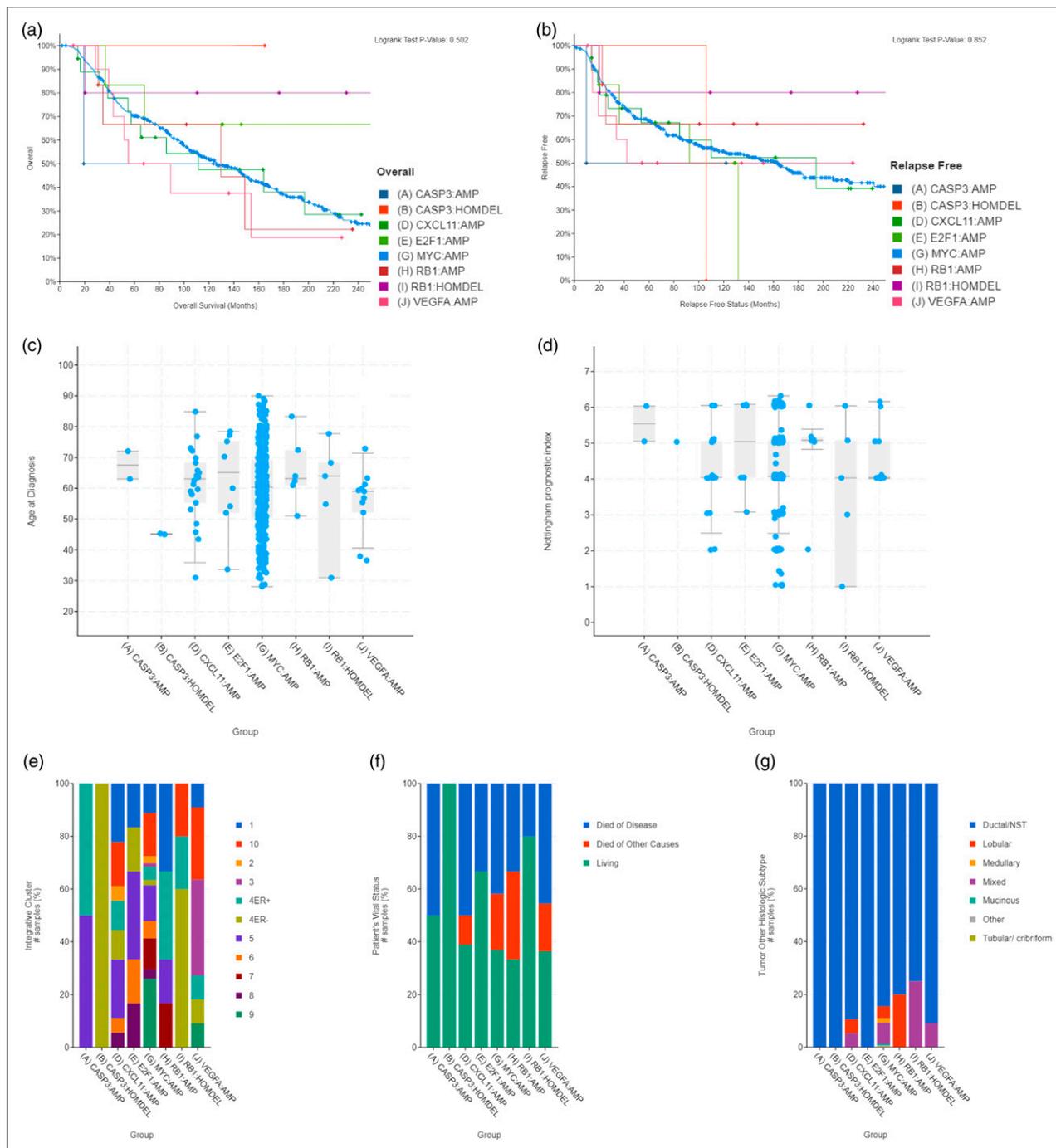
### Screening of *Dysosma Versipellis* active ingredients and targets

Search all the chemical components in *Dysosma Versipellis* through the TCMSP database. And search for 22 effective compounds with drug-like properties (DL)  $\geq 0.18$  and bioavailability (OB)  $\geq 30\%$  as the screening thresholds, then retrieve all the components corresponding to *Dysosma Versipellis* through the TCMSP database, and then screen the target points through Strawberry Perl-5.30.0.1; finally, 443 target points corresponding to 22 active ingredients

were obtained. Then, obtain 20,386 human protein annotation pieces of information from the Uniprot database, use Strawberry Perl-5.30.0.1 to convert 443 *Dysosma Versipellis* targets with the target names corresponding to the Uniprot database, and merge and delete duplicates to obtain 348 target genes.

### *Dysosma Versipellis* ingredients improve the prediction of potential targets for breast cancer

We use "Breast Cancer" as a keyword to search in five databases: GeneCards, OMIM, PharmGkb, TDD, and



**Figure 6.** (a) is Survival analysis in the overall breast patient population; (b) is Survival analysis in recurrence-free breast cancer patients; (c) is Age distribution is diagnosed in each gene; (d) is Nottingham prognostic index; (e) is Heat map of gene copy number variation rate; (f) is Patient’s Vital Status; (g) is Tumor Other Histologic Subtype.

DrugBank. The disease targets obtained from the five databases are integrated, and duplicate data are removed to obtain 10,941 targets compared with those obtained in 2.1. The intersection of 348 targets accepts 158 intersection targets (Figures 1(a) and (b)).

### Active ingredient-target gene network

The active ingredient-target gene network was constructed by using Cytoscape 3.8.0 software. A network containing 177 nodes (19 effective component nodes and 158 target

**Table 1.** Top six biological functions according to the ranking (Count) in GO enrichment analysis.

Description	geneID	Count
Response to lipopolysaccharide	PTGS2/RELA/MAPK3/MAPK1/MAPK8/IL1B/OPRM1/CASP9/CASP3/CASP8/PRKCA/CXCL8/GSTP1/TBXA2R/NOS2/AKT1/PPARGC1A/CYP1A2/CYP1A1/ICAM1/SELE/SLPI/MAOB/FOS/NFKBIA/GJA1/NOS3/THBD/SERPINE1/IL1A/MpO/CXCL11/CXCL2/PPARD/HSF1/CXCL10/CHUK	37
Response to molecule of bacterial origin	PTGS2/RELA/MAPK3/MAPK1/MAPK8/IL1B/OPRM1/CASP9/CASP3/CASP8/PRKCA/CXCL8/GSTP1/TBXA2R/NOS2/AKT1/PPARGC1A/CYP1A2/CYP1A1/ICAM1/SELE/SLPI/MAOB/FOS/NFKBIA/GJA1/NOS3/THBD/SERPINE1/IL1A/MpO/CXCL11/CXCL2/PPARD/HSF1/CXCL10/CHUK	37
Response to drug	KCNH2/PTGS2/NCOA1/RELA/IL1B/PDE3A/ADRA1A/SLC6A4/BCL2/CASP3/POR/SOD1/CAT/TBXA2R/NOS2/SLC6A2/STAT1/PPARGC1A/CYP1A2/CYP1A1/ICAM1/NR1I2/PPP3CA/MAOB/EGFR/CCND1/FOS/TP53/TOPI/MYC/CCNB1/COL1A1/NFE2L2/CHEK2/HSF1/CHUK/NKX3-1	37
Response to nutrient levels	PTGS2/NCOA1/RELA/MAPK3/MAPK1/MAPK8/ADRB2/SLC6A4/OPRM1/BCL2/JUN/PONI/POR/SOD1/CAT/HMGCR/GSTP1/TBXA2R/AKT1/STAT1/PPARGC1A/HMOX1/CYP1A1/ICAM1/AKR1C3/EGFR/CCND1/TP53/COL1A1/MPO/NFE2L2/NQO1/PPARA/PPARD/HSF1/CXCL10/SPP1	37
Response to metal ion	PTGS2/MMP9/MAPK3/MAPK1/MAPK8/SCN5A/BCL2/CASP9/JUN/CASP3/CASP8/SOD1/CAT/ALOX5AP/AKT1/PPARGC1A/HMOX1/CYP1A2/CYP1A1/ICAM1/PPP3CA/AKR1C3/MAOB/EGFR/CCND1/FOS/HIF1A/CAV1/CCNB1/IL1A/NCF1/NFE2L2/NQO1/PARP1/HSF1/CHUK	36
Response to oxidative stress	PTGS1/PTGS2/RELA/MMP2/MMP9/MAPK3/MAPK1/MAPK8/BCL2/JUN/CASP3/SOD1/CAT/GSTP1/AKT1/STAT1/PPARGC1A/HMOX1/CYP1B1/AKR1C3/MMP3/EGFR/FOS/TP53/HIF1A/DUOX2/NOS3/HSPB1/COL1A1/MPO/NCF1/NFE2L2/NQO1/PARP1/HSF1/CHUK	36

nodes) and 328 connections (i.e., the connection between the effective component and the target) is obtained (Figure 1B). The Degree value is used as main reference basis for the active ingredient-target gene network diagram in the topological analysis; the top 10 active ingredients in the order of the number of connections, namely, quercetin, kaempferol, beta-sitosterol, Rutin, Sitogluside, etc., are docked with 136,53,29,20, and 16 targets, respectively, (for details see Figure 1C). This analysis shows the mechanism of *Dysosma Versipellis*'s multi-component multi-target effect on breast cancer.

### Construction and analysis of PPI network

To further study how the mechanism of *Dysosma Versipellis* prevents and treats breast cancer, this study introduced 158 key targets of *Dysosma Versipellis* to the STRING database platform (<https://string-db.org/>) to predict protein interaction relationships and draw protein relationship network diagrams and set the minimum interaction score required to be the highest confidence level [highest confidence (0.900)]. The research species is limited to "humans" ("*Homo sapiens*"). Hide the unconnected nodes in the network and keep the remaining parameters at the default settings. Get *Dysosma Versipellis* and the PPI networks acting on breast cancer (Figure 2). As shown in Figure 2, the *Dysosma Versipellis* targets protein PPI network which has 470 interaction lines,

of which there are 157 nodes (1 target protein is not involved in the interaction between proteins).

The light blue line represents the protein-protein interactions obtained in the database, and the purple line represents the protein-protein interactions that have been experimentally verified.

### Survival analysis and verification of core genes and pivot genes

Survival analysis of core genes and hub genes obtained in 1.5 showed that 16 of the 32 core genes were significantly related to breast cancer (log-rank  $p < 0.05$ ) (Figure 4A). In the three hub gene groups (Degree, MCC, and Bottle Neck), there are 10, 10, and seven genes significantly related in each of 20 genes in Bottle Neck and Bottle Neck (Figures 4B-4D), and there are 22 genes to remove duplicate values: MYC, TP53, RELA (NFKB3), FOS, MAPK8, EGF, PRKCA, RXRA, NCOA2, NCOA1 (RIP160), PPARA, E2F1, VEGFA, CASP3 (apopain), RB1, IL1B, BCL2, CCL25, CXCL10, CXCL11 (SCYB9B), OPRM1, and PTGER3, of which there are seven highly expressed genes: VEGFA, E2F1, RB1, CXCL10, CXCL11 (SCYB9B), MYC, and CASP3 (apopain). The mutation information about 22 genes is shown in Figure 5A. In 2173 cases (50%) of patients or samples, mutations in central genes occurred. MYC, TP53, PRKCA, and NCOA2 were most frequently mutated (25%,

34%, 8%, 16%, respectively). These mutations include inframe mutation, missense mutation, splice mutation, truncating mutation, amplification, and deep deletion. Among different types of mutations, multiple mutations account for the highest percentage. The GEPIA2 database was used to verify the expression levels of 22 genes in tumors and normal tissues. According to the Tumor Genome Atlas (TCGA) gene expression profile and Genotype-Tissue Expression (GTEx) project, there are 22 genes in breast cancer tissues, and seven genes in normal tissues (MYC, FOS, E2F1, IL1B, CXCL10, SCYB9B, and the expression level of PTGER3) were statistically significant ( $p < 0.01$ ) (Figures 5B–5E).

### Gene copy number variation rate

The analysis of six genes VEGFA, E2F1, RB1, CXCL11, CASP3, and MYC (CXCL10 is not in groups from CNA genes) for gene copy number variation rate analysis (CNA): ① The survival analysis of multiple genes in the total breast patient population, with a  $p$ -value of 0.502 (Figure 6(a)). ② The Survival analysis of multiple genes in breast cancer patients without recurrence, and the  $p$ -value is 0.852 (Figure 6(b)). ③ Age at diagnosis, the diagnosis of patients is mainly concentrated on those between 50 and 70 years old (Figure 6(c)). ④ Nottingham prognostic index, the overall index is relatively high (Figure 6(d)). ⑤ The resulting integrated cluster  $p$ -value is less than one billionth. The heat map of gene copy number variation rate is shown in Figure 6(e). ⑥ Patient's vital status, the percentage of patients who died of breast cancer is relatively large (Figure 6(f)). ⑦ Tumor and other histologic subtypes, there

are more subtypes regulated by MYC (Figure 6(g)). This analysis shows that VEGFA, E2F1, RB1, CXCL11, CASP3, and MYC play a certain role in regulating breast cancer.

### GO enrichment analysis

GO enrichment analysis of 158 key target genes of *Dysosma Versipellis* in breast cancer treatment showed that 2471 biological processes were affected. According to the ranking (Count), the top six biological functions (Table 1) are for lipopolysaccharide reaction to the bacteria originating molecule, the reaction to the drug, the reaction to the nutrient level, the response to the metal ion, the reaction to the oxidative stress, and the analysis results (Figure 7A(a)).

### KEGG enrichment pathway analysis

KEGG pathway enrichment analysis was performed on the 158 key target genes of *Dysosma Versipellis* for breast cancer treatment. A total of 167 pathways were obtained. The top six pathways were ranked according to the Gene Count (Table 2) and the analysis results (Figure 7A(b)). The results show that key targets are mainly enriched in fluid shear stress and atherosclerosis, Kaposi sarcoma-associated herpesvirus infection, human cytomegalovirus infection, MAPK signaling pathway, hepatitis B, and PI3K-Akt signaling pathway wait. Figure 7B(a) shows the fluid shear stress and atherosclerosis pathway and Figure 7B(b) shows the breast cancer pathway. KEGG pathway enrichment analysis shows that *Dysosma Versipellis* exerts an effect on BC through multiple signaling pathways.

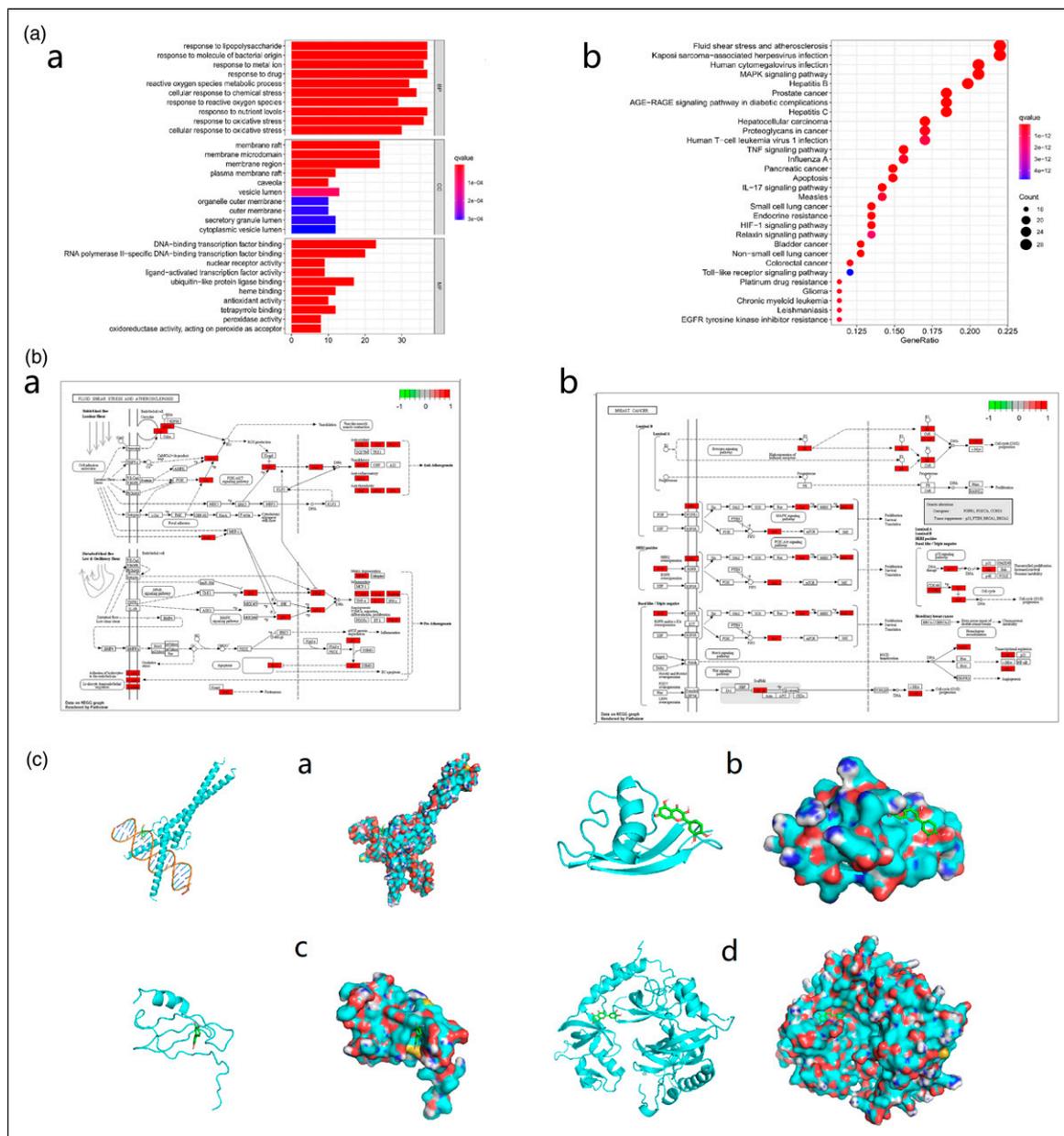
**Table.2.** The top six pathways were ranked according to the ranking (Count) in the enrichment analysis of the KEGG pathway.

Description	genelD	Count
Fluid shear stress and atherosclerosis	KDR/CALML3/RELA/MMP2/MMP9/VEGFA/MAPK8/IL1B/BCL2/JUN/GSTP1/IKBKB/AKT1/HMOX1/ICAM1/SELE/VCAM1/GSTM1/GSTM2/FOS/TP53/CAVI/NOS3/PLAT/THBD/IFNG/IL1A/NCF1/NFE2L2/NQO1/CHUK	31
Kaposi sarcoma-associated herpesvirus infection	PTGS2/CALML3/RELA/MAPK3/MAPK1/VEGFA/MAPK8/BAX/CASP9/JUN/CASP3/CASP8/IL6ST/CXCL8/IKBKB/AKT1/STAT1/ICAM1/PPP3CA/CCND1/FOS/RB1/TP53/NFKBIA/RAFI/HIF1A/MYC/CXCL2/CHUK/E2F1/E2F2	31
Human cytomegalovirus infection	PTGS2/CALML3/RELA/MAPK3/MAPK1/VEGFA/IL1B/BAX/CASP9/CASP3/CASP8/PRKCA/CXCL8/PRKCB/IKBKB/AKT1/PPP3CA/EGFR/CCND1/RB1/TP53/ELK1/NFKBIA/RAFI/MYC/PTGER3/CHUK/E2F1/E2F2	29
MAPK signaling pathway	KDR/RELA/MAPK3/MAPK1/VEGFA/IGF1R/MAPK8/IL1B/JUN/CASP3/PRKCA/PRKCB/IKBKB/AKT1/PPP3CA/EGFR/FOS/EGF/TP53/ELK1/RAFI/ERBB2/MYC/HSPB1/IL1A/CHUK/IGF2/ERBB3/RASA1	29
Hepatitis B	RELA/MMP9/MAPK3/MAPK1/MAPK8/BCL2/BAX/CASP9/JUN/CASP3/CASP8/PRKCA/CXCL8/PRKCB/IKBKB/AKT1/STAT1/FOS/RB1/TP53/ELK1/NFKBIA/RAFI/MYC/BIRC5/CHUK/E2F1/E2F2	28
PI3K-Akt signaling pathway	KDR/RELA/MAPK3/MAPK1/VEGFA/IGF1R/CHRM1/RXRABCL2/CASP9/PRKCA/IKBKB/AKT1/EGFR/CCND1/BCL2L1/EGF/TP53/RAFI/ERBB2/MYC/NOS3/IL2RA/COL1A1/CHUK/SpP1/IGF2/ERBB3	28

**Docking analysis of key targets and key active molecules**

The quercetin was docked with MYC, CXCL10, CXCL11, and E2F1, respectively, and the results are as follows (Figure 7C): ① MYC score was  $-7.7\text{kcal/mol}$ , and the binding force was relatively poor ( $-7.7\text{kcal/mol} > -40\text{kcal/mol}$ ). ② The score of CXCL10 is  $-6.1\text{kcal/mol}$ , and the

binding force is rather poor ( $-6.1\text{kcal/mol} > -40\text{kcal/mol}$ ). ③ The score of CXCL11 is  $-6.4\text{kcal/mol}$ , and the pressing force is relatively poor ( $-6.4\text{kcal/mol} > -40\text{kcal/mol}$ ). ④ The score of E2F1 is  $-8.1\text{kcal/mol}$  and the binding energy is rather poor ( $-8.1\text{kcal/mol} > -40\text{kcal/mol}$ ). The molecules of the four receptors are small. The hydrophobic interaction area is small, the Van der Waals force is small, the electrostatic force is weak, and the binding energy lacks; this



**Figure 7.** A. *Dysosma Versipellis*-pathway analysis of breast cancer. (a) is a histogram of key gene ontology (GO) analysis of *Dysosma Versipellis* action, (b) is a bubble chart of gene interaction (KEGG) pathway analysis. B. KEGG pathway diagram (a) is a diagram of fluid shear stress and atherosclerosis, (b) is a diagram of the breast cancer pathway. C. The docking of quercetin and gene molecules (a) is the molecular docking diagram of MYC and Quercetin, (b) is the molecular docking diagram of CXCL10 and Quercetin, (c) is the molecular docking diagram of CXCL11 and Quercetin, (d) is the molecular docking diagram of E2F1 and Quercetin.

requires preparations to strengthen the binding capacity of ligands and receptors to prevent and treat BC.

**Verification of the differential genes in the TIMER database**

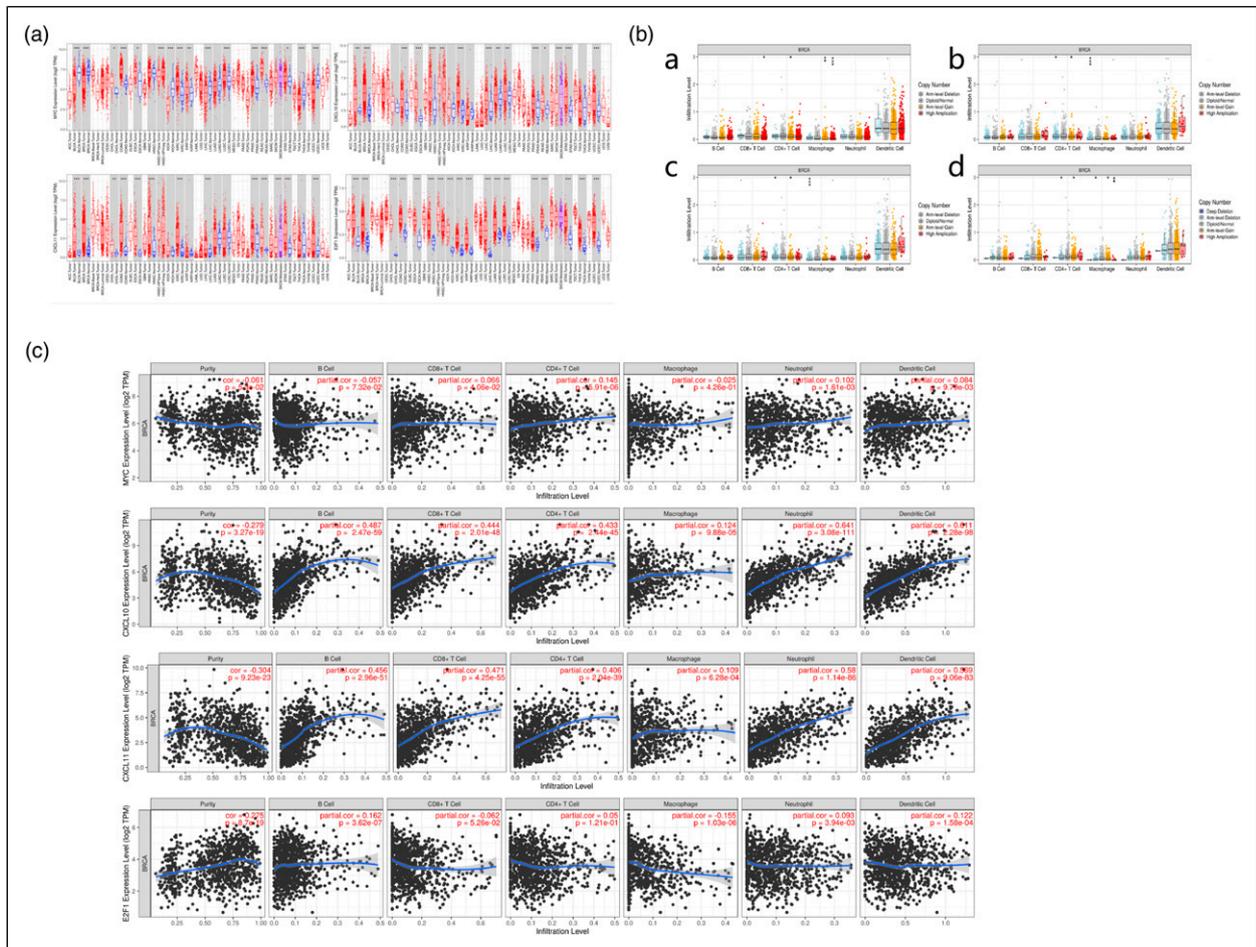
MYC, CXCL10, CXCL11, and E2F1 have different expressions in different cancer types, and \* indicates a very significant difference in expression (Figure 8A). MYC, CXCL10, CXCL11, and E2F1 gene copy number changes are correlated with immune abundance (Figure 8B), and \* indicates that the gene copy number has a significant correlation with the immune microenvironment. For the correlation between MYC, CXCL10, CXCL11, and E2F1 gene expression and the abundance of immune cells (Figure 8C), \* indicates that gene expression has a significant correlation with the immune microenvironment. For the correlation between MYC, CXCL10,

CXCL11, and E2F1 gene expression and immune cell survival analysis (Figure 9), the correlation between TP53 gene mutation, and immune infiltration (Figure 10), \* indicates that TP53 gene mutation is highly correlated with immune cells.

**Discussion**

**Active ingredient-target gene network**

In this study, we constructed a network of key active components of *Dysosma Versipellis* and breast cancer through the TCMSP database, quercetin was selected through five disease gene databases (GeneCards, OMIM, PharmGkb, TDD, and DrugBank), and bioinformatics methods and quercetin were screened out: quercetin, kaempferol, beta-sitosterol, rutin, and other active ingredients (degree scores decrease in order). ① Quercetin is



**Figure 8.** A. Differences in gene expression in different tumor tissues. B. Correlation between gene copy number changes and immune abundance (a) is the correlation between the copy number change of gene MYC and the abundance of immune cells, (b) is the correlation between the copy number change of gene CXCL10 and the abundance of immune cells, (c) is the correlation between the copy number change of gene CXCL11 and the abundance of immune cells, (d) is The correlation between the copy number of gene E2F1 and the abundance of immune cells. C. Correlation between screening genes and abundance of immune cells.

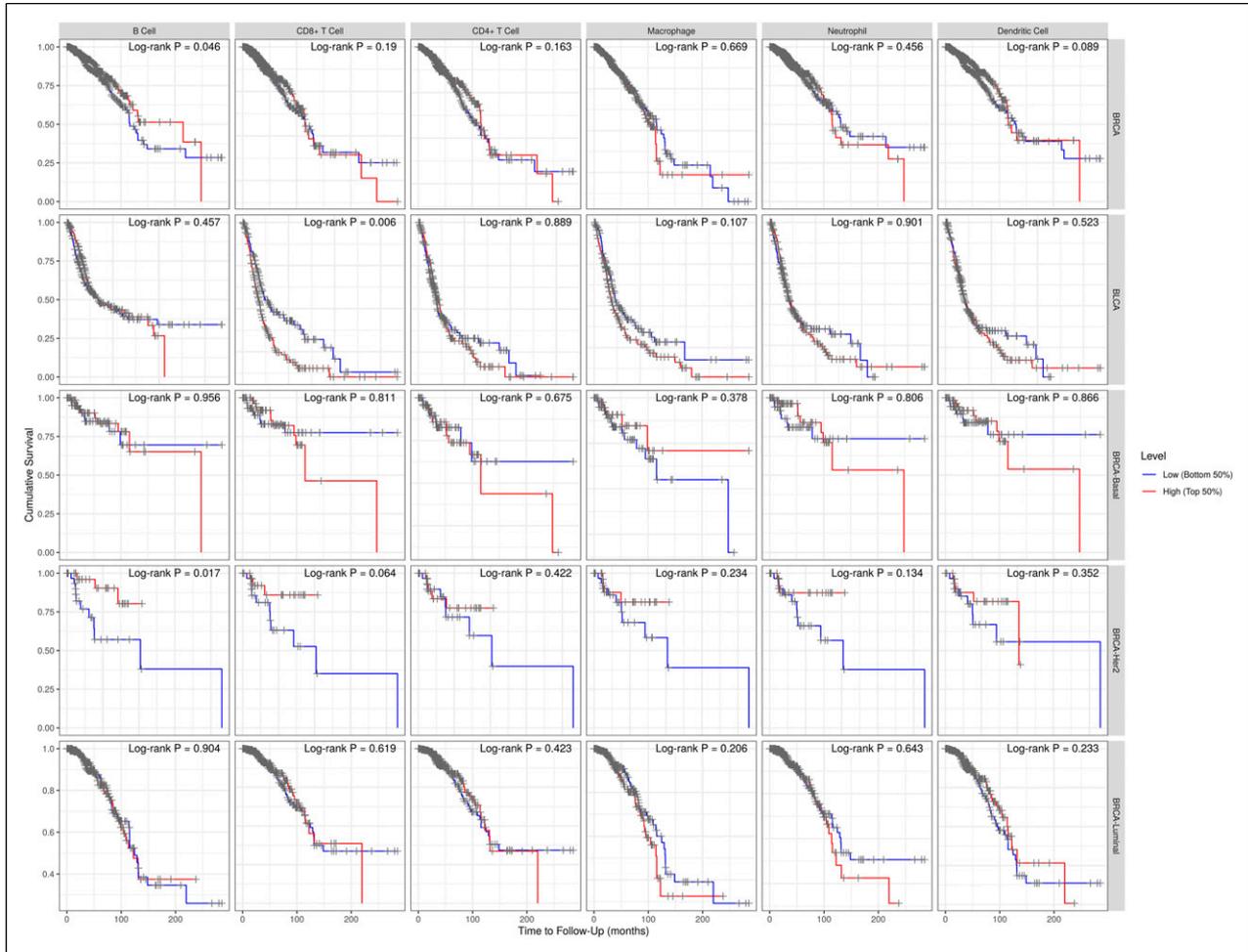


Figure 9. The relationship between survival analysis and immune infiltration.

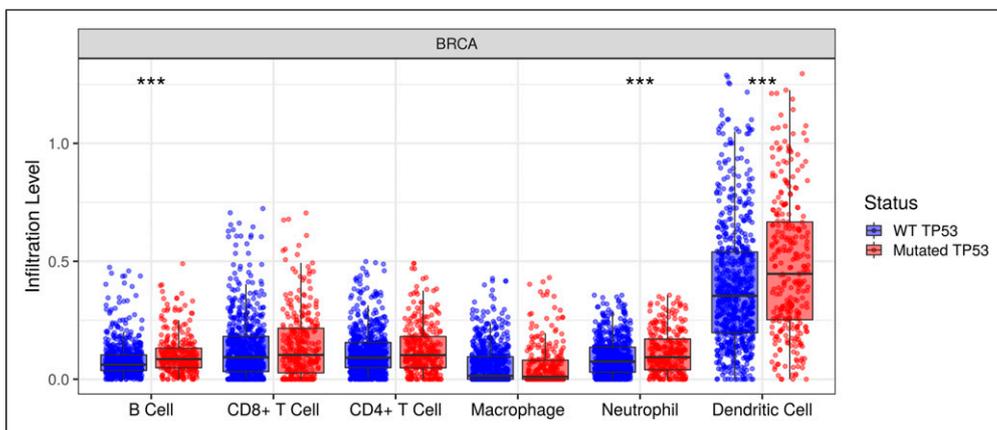


Figure 10. The correlation between TP53 gene mutation and immune infiltration.

widely distributed in the plant kingdom and is a flavonol compound with multiple biological activities. Studies have shown that quercetin inhibits tumor cell proliferation, can induce apoptosis and cell death, and has anti-inflammatory and anti-oxidant effects. Quercetin can also regulate a variety of pathways to slow down the progression of cancer. Eating flavonoids and canceler risk is positively correlated, but its intrinsic activity is low, water solubility is poor, metabolic rate is high, and oral bioavailability and absorption are poor. To improve its intrinsic activity, oral bioavailability, and absorption rate, people have begun scientific research using biodegradable and biocompatible carriers as delivery systems, including liposomes, PLGA, PLA, chitosan, and silica. Quercetin's delivery system will effectively improve quercetin's effect in breast cancer treatment.<sup>9,10</sup> ②Kaempferol is a class of flavonoids, which can inhibit proliferation and induce cell cycle and arrest apoptosis and DNA damage of breast cancer cells. Kaempferol can effectively inhibit triple-negative breast cancer (TNBC) MDA-MB-231 cells. Proliferation and the inhibitory effect of kaempferol on the proliferation of triple-negative breast cancer cells are stronger than that of the estrogen receptor-positive BT474 cell line.<sup>11</sup> ③β-sitosterol is a plant sterol that can induce apoptosis of breast cancer cells, and it has a certain effect on the two pathways leading to apoptosis; according to the location of the signal triggered by the pathway, it is divided into the external pathway and the internal pathway. Caspases eight and nine catalyze the external and internal pathways, respectively, which can lead to caspase 3.<sup>12</sup> Studies have also shown that β-sitosterol can enhance tamoxifen's effectiveness on breast cancer cells by affecting ceramide metabolism. SIT effectively activates MCF-7 and MDA-MB-231 De Novo CER synthesis in cells by stimulating serine palmitoyltransferase activity in cells; TAM promotes CER accumulation in both cells by inhibiting CER glycosylation.<sup>13</sup> ④Rutin is a common flavonoid compound with anti-oxidation, anti-inflammatory, anti-diabetic, and anti-adipogenic effects.<sup>14</sup> Studies have found that rutin is an inhibitor of breast cancer resistance protein (BCRP) transporter and can be used as an oral bioavailability enhancer for drugs such as diclofenac,<sup>15</sup> which can prevent and control c-Met-dependent breast malignancies.<sup>16</sup>

### GO enrichment analysis

From the GO enrichment analysis results, *Dysosma Versipellis* uses multiple biological processes to prevent and treat breast cancer. The more robust biological processes are the response to lipopolysaccharides, the response to bacteria originating in molecules, the response to drugs and these nutritional level responses, the answer to metal ions, the response to oxidative stress, etc.; this one corresponds to the risk factors of breast cancer.

### KEGG pathway enrichment analysis

KEGG pathway enrichment analysis found that 158 targets were involved pathways which include fluid shear stress and atherosclerosis (fluid shear stress and atherosclerosis), Kaposi sarcoma-associated herpesvirus infection (Kaposi sarcoma-associated herpesvirus infection), human cytomegalovirus infection, MAPK signaling pathway, hepatitis B, PI3K-Akt signaling pathway, etc. ①Novak CM et al.<sup>17</sup> designed and revealed a bioreactor that can apply shear stress to cells in the 3D extracellular matrix, and studied the effect of shear stress on breast and lung pleural space by applying continuous shear stress to breast cancer. Research results show that pulsating shear stress promotes breast cancer cell proliferation, invasion potential, chemoresistance, and PLAU signaling. Atherosclerosis is a chronic inflammatory disease characterized by lipid accumulation, smooth muscle cell proliferation, apoptosis, necrosis, fibrosis, and local inflammation.<sup>18</sup> Studies have shown that breast cancer 1/2 (BRCA1/2) gene-invalid cells are more sensitive to oxidative stress. Overexpression or silencing of BRCA1 protects or amplifies endothelial cell apoptosis induced by inflammation and DOX, respectively,<sup>19</sup> which may be related to atherogenesis. ②Molecular epidemiological evidence shows an association between human cytomegalovirus (HCMV) and breast cancer. HCMV is a β-herpes virus that can affect 70–90% of the global population and cause general, acute, persistent, or lifelong latent infections.<sup>20,21</sup> Gong Yaya et al.<sup>22</sup> detected HCMV protein and DNA in ductal carcinoma in situ and invasive ductal carcinoma tissues of the breast. They found that HCMV infection can induce the secretion of inflammatory cells and growth factors, thereby accelerating carcinogenesis. The transcription factors produce tumor suppressor protein which can control the cell cycle in normal physiological processes and has the potential of tumor control. It promotes the occurrence and development of tumors. ③Breast cancer patients will activate the hepatitis B virus (HBV) during or after chemotherapy.<sup>23</sup> The infection rate of HBV in Chinese women with breast cancer is very high.<sup>24</sup> Hu N, Zhang J, and others<sup>25</sup> found that miR-520b directly targets the 3' untranslated region (3'UTR) of hepatitis B X interacting protein (HBXIP) or interleukin-8 (IL-8), which can help cell migration, and HBXIP is considered as a potential therapeutic target for breast cancer. ④Abnormal activation of PI3K pathway activity is often observed in breast cancer, leading to uncontrolled tumor cell growth and drug resistance.<sup>26</sup> Rahmani F, Ferns GA, etc.<sup>27</sup> believed that the uncontrolled activity of oncogenic PI3K/AKT signals is related to the poor prognosis and tumor metastasis of breast cancer patients.

### Survival analysis and verification of key genes

Through the establishment of a PPI network, screening of core genes, pivot genes, and survival analysis and verification, the results show that *Dysosma Versipellis*' key target genes of breast cancer involve MYC, CXCL10, CXCL11, E2F1, etc. ① Lourenco C and Kalkat M et al.<sup>28</sup> simulated the in vivo model of human diseases to verify that MYC is out of control; ectopic expression of phosphoinositide 3-kinase pathway (PIK3CA H1047R) and ectopic expression of common breast cancer mutations in MCF10A cells lead to mouse tissues development of sex acinar structure. Simultaneous expression (PIK3CA H1047R) and dysregulation of MYC lead to the development of invasive ductal carcinoma. Therefore, the uncontrolled expression of MYC will produce an MYC-dependent normal-to-tumor transition, which can be measured in vivo. These MYC-driven tumors show the classic hallmarks of human breast cancer at both the pathological and molecular levels. Liang ZR et al.<sup>29</sup> found that the key oncogene protein c-MYC was significantly inhibited in estrogen receptor  $\alpha$  (ER- $\alpha$ )-positive breast cancer cells, but not in ER- $\alpha$ -negative cells. ② CXCL10 and CXCL11 (chemokines) are proteins that induce chemotaxis, promote the differentiation of immune cells, and cause tissue extravasation. They can participate in the migration, differentiation, and activation of white blood cells, leading to tumor suppression.<sup>30,31</sup> ③ E2F1 is a transcription factor involved in cell cycle regulation and apoptosis.<sup>32</sup> Studies have found that E2F1 and EIF4A3 may promote the expression of circSEPT9 and significantly inhibit the proliferation, migration, and invasion of TNBC cells, induce TNBC cell apoptosis and autophagy, and inhibit the growth and metastasis of tumors in vivo.<sup>33</sup>

### Limitations

The article uses TCMSP (DL>0.18 and OB>30%) to screen the chemical components of traditional Chinese medicine, which can reflect the anti-cancer effect of the drug to a certain extent. However, it is also possible to delete individual chemical components with anti-tumor activity. The article fully analyzes the anti-cancer mechanism of quercetin, but it cannot explain that the traditional Chinese medicine octagonal lotus also has the same pharmacological effects. The results of the article are obtained based on analyzing a large amount of data, and necessary data verification has been carried out. If we want to make the data more reliable, I think we need to conduct experimental verification.

### Conclusion

*Dysosma Versipellis* is used in breast cancer through complex mechanisms such as multiple targets, multiple

biological processes, and multiple pathways. Through analysis, we have obtained quercetin, a valuable component in the prevention and treatment of breast cancer, and several key genes: MYC, CXCL10, CXCL11, and E2F1. Many studies have shown that these genes play a role in regulating breast cancer to a particular effect. This study provides a reference for *Dysosma Versipellis* to prevent and treat breast cancer to a certain extent. The common genes are compared through the chemical composition database and the breast cancer database, and the gene sets with significant effects are screened out. And perform survival verification on the gene set pathway enrichment analysis to screen out the key role pathways. The critical gene is molecularly docked with the compound quercetin to obtain a more stable binding mode. Key genes have been fully verified in a variety of immune microenvironments.

Due to incomplete information, incomplete retrieval, unclear function mechanism, and incomplete analysis, the research is limited to the research on the condition of network database, hoping to combine network pharmacology with biological experiments and clinical data in future research. We continue to explore the complex mechanism of traditional Chinese medicine prevention and treatment of diseases.

### Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Science and Technology Project of Education Department of Jiangxi Province, GJJ191636. Science and Technology Plan of Health and Family Planning Commission of Jiangxi Province, 20175542,202131039. College Student Innovation and Entrepreneurship Training Program, S202113437009. JXUTCM Chinese Medicine Discipline Construction Project, 5251800240

### ORCID iD

Yuxing Ma  <https://orcid.org/0000-0001-8234-762X>

### References

1. Abolghasemi M, Tehrani SS, Yousefi T, et al. (2020) MicroRNAs in breast cancer: Roles, functions, and mechanism of actions. *Journal of Cellular Physiology* 235: 5008–5029.
2. Abolghasemi M, Tehrani SS, Yousefi T, et al. (2020) Critical roles of long noncoding RNAs in breast cancer. *Journal of Cellular Physiology* 235: 5059–5071.

3. Fan L, Strasser-Weippl K, Li J-J, et al. (2014) Breast cancer in China. *The Lancet Oncology* 15(7): e279–e289.
4. Sun Y-S, Zhao Z, Yang Z-N, et al. (2017) Risk factors and preventions of breast cancer. *International Journal of Biological Sciences* 13(11): 1387–1397.
5. Yuan Y (2011) *Study on the Extraction and Analysis of Podophyllotoxin and the New Method of Quality Control [D]*. Changsha, China: Hunan University.
6. Lu J (2009) The pharmacology and clinical application of Dysosma Versipellis[J]. *Modern Medicine and Health* (23): 3608–3609.
7. Mukherjee A, Basu S, Sarkar N, et al. (2001) Advances in cancer therapy with plant based natural products. *Current Medicinal Chemistry* 8: 1467–1486.
8. Mantle D, Lennard TW, and Pickering AT (2000) Therapeutic applications of medicinal plants in the treatment of breast cancer: a review of their pharmacology, efficacy and tolerability. *Adverse drug reactions and toxicological reviews* 19: 223–240.
9. Ezzati M, Yousefi B, Velaei K, et al. (2020) A review on anti-cancer properties of Quercetin in breast cancer. *Life Sciences* 248: 117463.
10. Kasiri N, Rahmati M, Ahmadi L, et al. (2020) Therapeutic potential of Quercetin on human breast cancer in different dimensions. *Inflammopharmacology* 28(1): 39–62.
11. Zhu L and Xue L (2019) Kaempferol suppresses proliferation and induces cell cycle arrest, apoptosis, and DNA damage in breast cancer cells. *Oncology Research Featuring Preclinical and Clinical Cancer Therapeutics* 27(6): 629–634.
12. Awad AB, Roy R and Fink CS (2003) Beta-sitosterol, a plant sterol, induces apoptosis and activates key caspases in MDA-MB-231 human breast cancer cells. *Oncology reports* 10(2): 497–500.
13. Awad AB, Barta SL, Fink CS, et al. (2008)  $\beta$ -Sitosterol enhances tamoxifen effectiveness on breast cancer cells by affecting ceramide metabolism. *Molecular Nutrition & Food Research* 52(4): 419–426.
14. Chua LS A review on plant-based rutin extraction methods and its pharmacological activities. *Journal of Ethnopharmacology*. 2013;150(3):805-817.
15. Dogra A, Gour A, Bhatt S, et al. (2013) A review on plant-based rutin extraction methods and its pharmacological activities. *Journal of Ethnopharmacology* 150(3): 805–817.
16. Elsayed HE, Ebrahim HY, Mohyeldin MM, et al. (2017) Rutin as a novel c-met inhibitory lead for the control of triple negative breast malignancies. *Nutrition and Cancer* 69(8): 1256–1271.
17. Novak CM, Horst EN, Taylor CC, et al. (2019) Fluid shear stress stimulates breast cancer cells to display invasive and chemoresistant phenotypes while upregulating PLA2 in a 3D bioreactor. *Biotechnology and Bioengineering* 116(11): 3084–3097.
18. Wu M-Y, Li C-J, Hou M-F, et al. (2017) New insights into the role of inflammation in the pathogenesis of atherosclerosis. *International Journal of Molecular Sciences* 18(10): 2034.
19. Zhou S, Jin J, Wang J, et al. (2021) Effects of breast cancer genes 1 and 2 on cardiovascular diseases. *Current Problems in Cardiology* 46(3): 100421.
20. Sepahvand P, Makvandi M, Samarbafzadeh A, et al. (2019) Human Cytomegalovirus DNA among Women with Breast Cancer. *Asian Pacific Journal of Cancer Prevention* 20(8): 2275–2279.
21. Goodrum F, Caviness K and Zagallo P (2012) Human cytomegalovirus persistence. *Cellular Microbiology* 14(5): 644–655.
22. Gong Y, Zhao J and Wang M (2016) Research progress on the relationship between human cytomegalovirus infection and breast cancer[J]. *Microbes and Infection* 11(01): 55–58.
23. Liu Z, Jiang L, Liang G, et al. (2017) Hepatitis B virus reactivation in breast cancer patients undergoing chemotherapy: a review and meta-analysis of prophylaxis management. *Journal of Viral Hepatitis* 24(7): 561–572.
24. Wu H, Zhao C, Adhikari VP, et al. (2017) The prevalence and clinicopathological features of breast cancer patients with hepatitis B virus infection in China. *Oncotarget* 8(11): 18185–18190.
25. Hu N, Zhang J, Cui W, et al. (2011) miR-520b regulates migration of breast cancer cells by targeting hepatitis B X-interacting protein and interleukin-8. *Journal of Biological Chemistry* 286(15): 13714–13722.
26. Ellis H and Ma CX (2019) PI3K inhibitors in breast cancer therapy. *Current Oncology Reports* 21(12): 110.
27. Rahmani F, Ferns GA, Talebian S, et al. (2020) Role of regulatory miRNAs of the PI3K/AKT signaling pathway in the pathogenesis of breast cancer. *Gene* 737: 144459.
28. Lourenco C, Kalkat M, Houlihan KE, et al. (2019) Modelling the MYC-driven normal-to-tumour switch in breast cancer. *Disease models & mechanisms* 12(7): dmm038083.
29. Liang Z-R, Qu L-H and Ma L-M (2020) Differential impacts of charcoal-stripped fetal bovine serum on c-Myc among distinct subtypes of breast cancer cell lines. *Biochemical and Biophysical Research Communications* 526(1): 267–272.
30. Tokunaga R, Zhang W, Naseem M, et al. (2018) CXCL9, CXCL10, CXCL11/CXCR3 axis for immune activation - a target for novel cancer therapy. *Cancer Treatment Reviews* 63: 40–47.
31. Liu L, Callahan MK, Huang D, et al. (2005) Chemokine receptor CXCR3: an unexpected enigma. *Current Topics in Developmental Biology* 68: 149–181.
32. Ertosun MG, Hapil FZ and Osman Nidai O (2016) E2F1 transcription factor and its impact on growth factor and cytokine signaling. *Cytokine & Growth Factor Reviews* 31: 17–25.
33. Zheng X, Huang M, Xing L, et al. (2020) The circRNA circSEPT9 mediated by E2F1 and EIF4A3 facilitates the carcinogenesis and development of triple-negative breast cancer. *Molecular Cancer* 19(1): 73.