Check for updates

OPEN

Integrated bioinformatics analysis to decipher molecular mechanism of compound Kushen injection for esophageal cancer by combining WGCNA with network pharmacology

Wei Zhou, Jiarui Wu✉, Jingyuan Zhang, Xinkui Liu, Siyu Guo, ShanShan Jia, Xiaomeng Zhang, Yingli Zhu & Miaomiao Wang

Compound Kushen injection (CKI), a medicine in widespread clinical use in China, has proven therapeutic effects on cancer. However, few molecular mechanism analyses have been carried out. To address this problem, bioinformatics approaches combining weighted gene co-expression network analysis with network pharmacology methods were undertaken to elucidate the underlying molecular mechanisms of CKI in the treatment of esophageal cancer (ESCA). First, the key gene modules related to the clinical traits of ESCA were analysed by WCGNA. Based on the results, the hub genes related to CKI treatment for ESCA were explored through network pharmacology. Molecular docking simulation was performed to recognize the binding activity of hub genes with CKI compounds. The results showed that the potential hub targets, including EGFR, ErbB2, CCND1 and IGF1R, are therapeutic targets of CKI for the treatment of ESCA. Moreover, these targets were significantly enriched in many pathways related to cancer and signalling pathways, such as the PI3K-Akt signalling pathway and ErbB signalling pathway. In conclusion, this research partially highlighted the molecular mechanism of CKI in the treatment of ESCA, offering great potential in the identification of the effective compounds in CKI and biomarkers for ESCA treatment.

Esophageal cancer (ESCA) is widespread worldwide. According to Global Cancer Statistics 2018, it ranks seventh in incidence and sixth in mortality[1]. The 5-year survival rate of ESCA is between 12 and 20% and differs substantially by sex[2]. China is a high-risk area for ESCA, especially in some rural areas, where the incidence rate far exceeds that of urban areas due to lifestyle and environmental reasons[3]. ESCA can be divided into esophageal adenocarcinoma (EAC) and esophageal squamous cell carcinoma (ESCC) according to histological classification. In recent decades, the incidence of EAC in Western countries has increased several times, and the proportion of ESCC has exceeded 90% throughout China[4,5]. The introduction of chemo(radio)therapy and surgical therapy led to increased survival rates and reduced the incidence of recurrence[6]. Because conventional methods do not adequately improve patient survival of ESCA, however, scientists are seeking more effective treatments. Recently, traditional Chinese medicine (TCM) has taken the world stage as complementary and alternative medicine[7]. Compound Kushen Injection (CKI) consists of two herbs, Kushen (Radix Sophorae Flavescentis) and Baituling (Rhizoma Smilacis Glabrae). CKI mainly contains various anticancer ingredients, such as matrine and oxymatrine, which can inhibit the growth of tumour cells, overcome resistance to metastasis and multidrug resistance, and protect human immunity[8]. CKI has been utilized in clinical practice for decades to treat various solid tumour types, including liver cancer, breast cancer, gastric cancer, and other cancer types[8,9]. The analysis of medical data on 2,550 ESCA patients from 22 large-scale hospitals in China confirmed that CKI has been relatively widely used in the clinical treatment of ESCA of different severities[10]. In addition, a previous study

Beijing University of Chinese Medicine, Beijing 100102, China. ✉email: exogamy@163.com

1

reported that CKI limited cancer pain both directly by blocking TRPV1 signalling and indirectly by reducing tumour growth[11]. For ESCA, it is worth noting that CKI used alone or combined with conventional radiotherapy can not only enhance antitumor efficacy but also reduce the toxicity induced by radiotherapy, thereby improving the quality of life[12]. Our early studies also found that combination with CKI can improve the clinical effectiveness rate and performance status of radiotherapy for ESCA. Furthermore, CKI can also provide treatment by reducing gastrointestinal reactions and radiation esophagitis[13].

Tumorigenesis is a complex process that is driven by a combination of networks of genes and environmental factors; there is a lack of effective methods to identify functional networks that Chinese medicine interferes with tumorigenesis[14–17]. To better analyse and predict the molecular mechanism of CKI in the treatment of ESCA, this study adopted weighted gene co-expression network analysis (WGCNA) integrated with the network pharmacology method. WGCNA can be used to find clusters (modules) of highly related genes, correlate modules and correlate with external sample traits and can be used to identify candidate biomarkers or therapeutic targets[18]. Network pharmacology not only caters to the "multi-component, multi-target" characteristics of TCM but also identifies drug-gene-disease links, explaining the therapeutic mechanism of drugs at the molecular level[19–21].

In the present work, we first used WGCNA to analyse ESCA mRNA datasets from The Cancer Genome Atlas (TCGA) to predict significant gene modules. Second, gene modules were combined with predicted targets of key CKI components to form a drug-gene-disease network and further analysed. In addition, molecular docking methods were adopted to confirm the degree of binding between the hub gene and the component. This study is intended to explain the mechanism of action of CKI in the treatment of ESCA at the molecular level and to provide a better basis for the diagnosis, treatment and prognosis of ESCA. Figure 1 depicts a flowchart of the technical strategy used in this study.

Results

WGCNA module construction.
A total of 161 samples and 5,000 genes were screened for the next WGCNA analysis. After normalization, no outlier samples were eliminated. In this study, the power of $\beta = 6$ (scale free $R^2 = 0.85$) was selected as the soft-thresholding parameter to ensure a scale-free network. (Fig. 2A) A total of 10 modules were identified via average linkage hierarchical clustering (Fig. 2B).

WGCNA hub module screening.
ME reflected the gene expression level of the entire module, and the relationship between ME and clinical traits was assessed by Pearson's test. The module and clinical traits were considered statistically significant when $p < 0.05$. The blue module and the turquoise module were considered to be hub modules through the association of modules with clinical traits (race, age, vital status, new tumour events, cancer status, histological type, pathologic T, pathologic N, pathologic M, stage, Barretts, smoking, alcohol) (Fig. 3). The topological overlap measure (TOM) was visualized with a heatmap that could depict adjacencies or topological overlaps (Fig. 4). Each module contained a set of RNAs that were co-expressed and had a high TOM. The same module genes could form networks and may participate in similar biological processes. The network building the key modules was filtered with a weight Cutoff = 0.1 between the genes. The blue module consists of 618 genes and 31,042 gene linkages. The turquoise module consists of 1,243 genes and 49,230 gene linkages. In addition, the top 100 genes in terms of degree were visualized using Cytoscape (Fig. 5).

Compound-predicted target network.
Following a literature search[8,22], the 23 active ingredients contained in CKI were selected for investigation, and the three-dimensional structure data of 16 active ingredients were obtained from the PubChem database[23].

The 16 active compounds of CKI are shown in Table 1. After Cytoscape visualization, 301 points (16 compound points and 285 gene points) and 636 edges were obtained (Fig. 6).

Potential target network for the treatment of ESCA with CKI.
The compound-predicted target network was combined with the blue and turquoise module genes. Thirty-two identical gene targets were considered potential targets for the CKI treatment of ESCA (Fig. 7A). To further unveil the therapeutic mechanism, STRING was used to construct a PPI network of 32 overlapping genes between the compound targets and key module targets. As shown in Fig. 7B, the potential therapeutic PPI network involved 41 nodes and 174 linkages between genes. Moreover, after network analysis, a target with a greater degree value than the mean based on the topological characteristics is a key gene for the CKI treatment of ESCA. The results of network analysis show 16 nodes with an average degree ≥ 8.49, including EGF, EGFR, ErbB2, HRAS, INS, STAT3, CCND1, IRS1, KRAS, IGF1R, IGF1, SHC1, GRB2, CBL, PTPN1 and CDKN1B.

GO functional and KEGG pathway enrichment analysis.
To clarify the multiple mechanisms of CKI on ESCA on a systematic level, we performed GO enrichment analysis including the biological process (BP), molecular function (MF), and cellular component (CC) and KEGG functional enrichment analysis of the selected PPI targets. Eventually, 559 enriched GO terms were identified, of which 477 were BPs, 54 were MFs, and 27 were CCs (FDR < 0.01 and $P < 0.01$). Figure 8A shows the top ten entries for BP, MF and CC, most of which were related to the cell cycle. To gain insights into the pharmacological mechanisms of CKI on ESCA, we performed KEGG analysis. The results demonstrated that 92 entries satisfy FDR < 0.01 and $P < 0.01$. Moreover, these targets were significantly enriched in many pathways related to cancer and signalling pathways, such as the PI3K-Akt signalling pathway, ErbB signalling pathway and FoxO signalling pathway (Figs. 8B, 9).
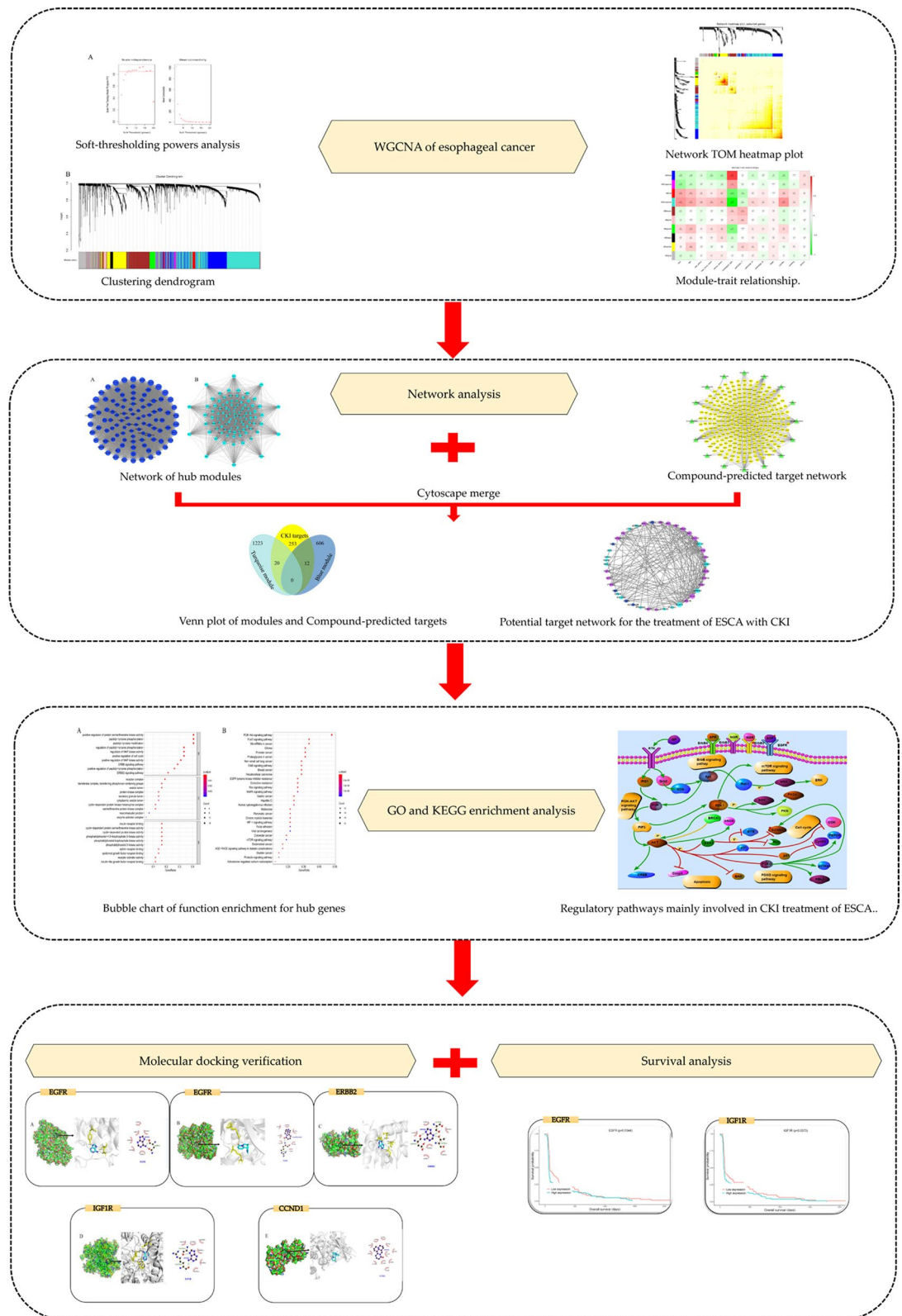
Figure 1. Workflow of this integrated bioinformatics analysis.

Molecular docking verification. Four of the hub genes (EGFR, ErbB2, CCND1 and IGF1R) were directly related to the CKI active ingredients (adenine, N-methylcytisine and matrine). These 4 potential target proteins and their corresponding small-molecule ligand components were docked by AutoDock Vina. The docking
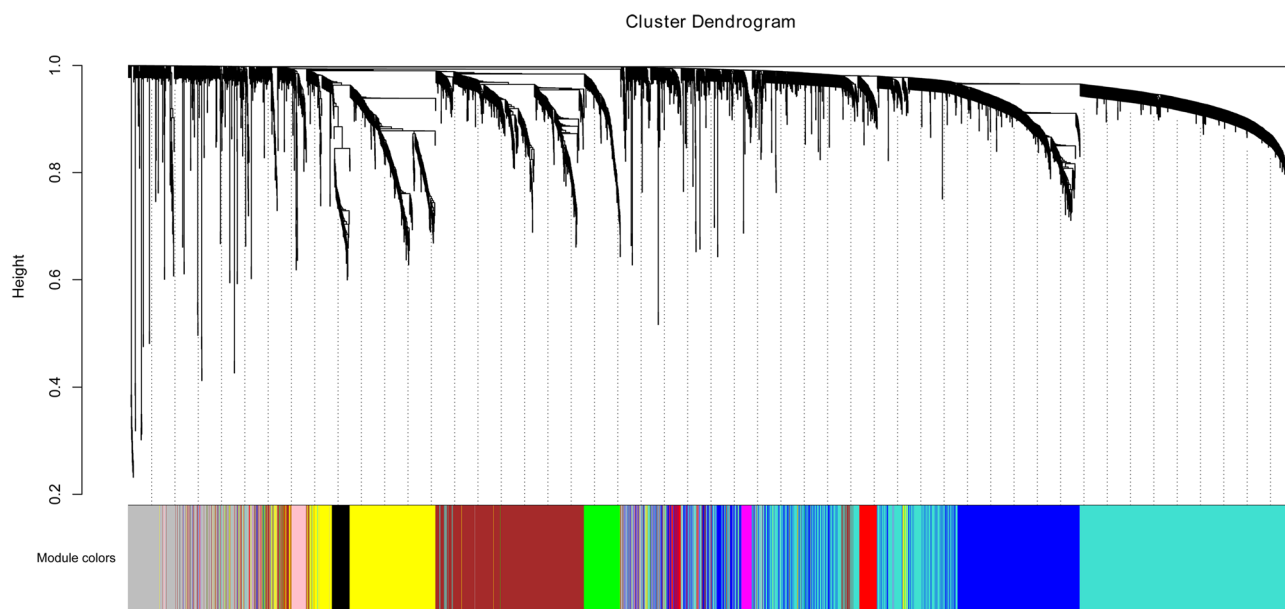
Figure 2. Clustering dendrogram. corFnc = "pearson"; power = 6; min. module size = 30; mergeCutHeight of 0.2.
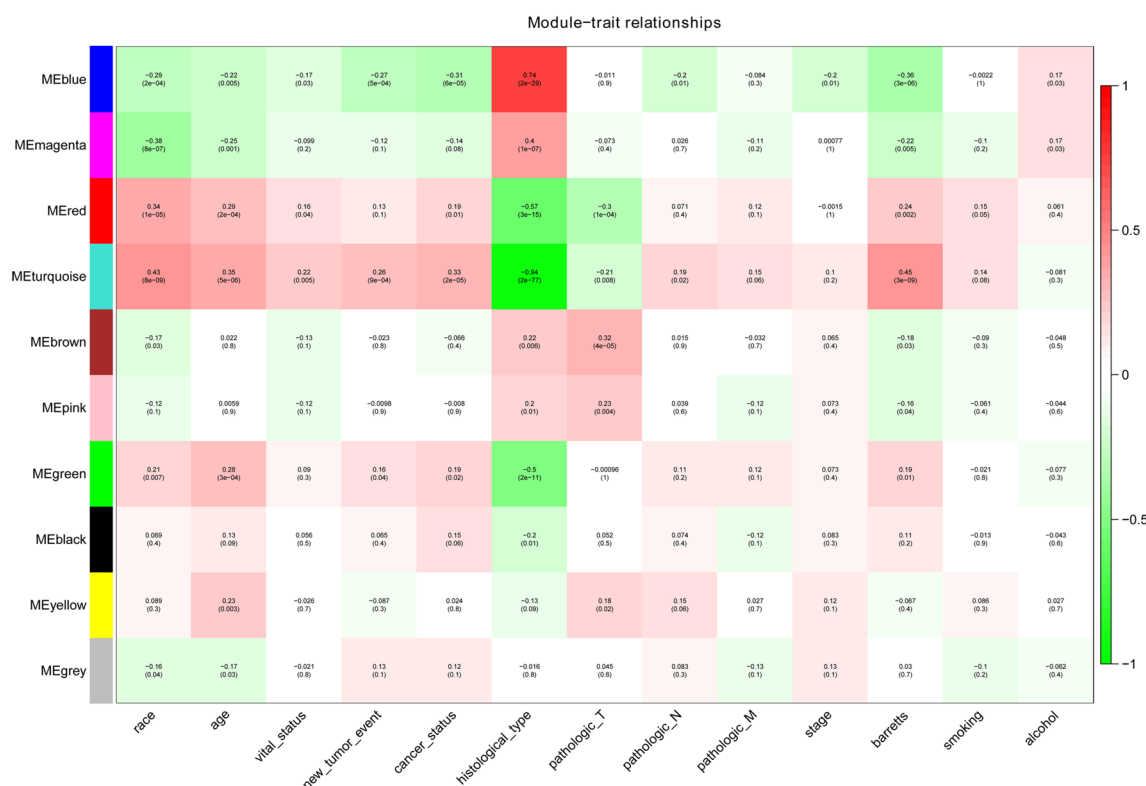
Figure 3. Module-trait relationship. Each row corresponds to an ME, and each column corresponds to a clinical trait. Each cell contains a corresponding correlation.

results showed that the binding affinity was not greater than −6.0 kcal/mol, which proved that CKI components had good binding ability to the targets (Table 2). Figure 10 illustrates the interaction of the target compounds of the docking simulation. Adenine mainly forms two hydrogen bonds with the TYR-801 and ASN-808 residues on the EGFR protein, and a total of 6 residues are bound to the protein by hydrophobic interaction. In addition, the same EGFR protein formed two hydrogen bonds with residues LYS-852 and ARG-776 in addition to 7 hydrophobic bonds. Adenine formed 4 hydrogen bonds and 6 hydrophobic bonds with the ERBB2 protein and IGF1R protein, respectively, which proved that they are relatively tightly bound. Finally, although matrine does not form a hydrogen bond with CCND1, it binds to 10 residues of the protein by hydrophobic interaction.
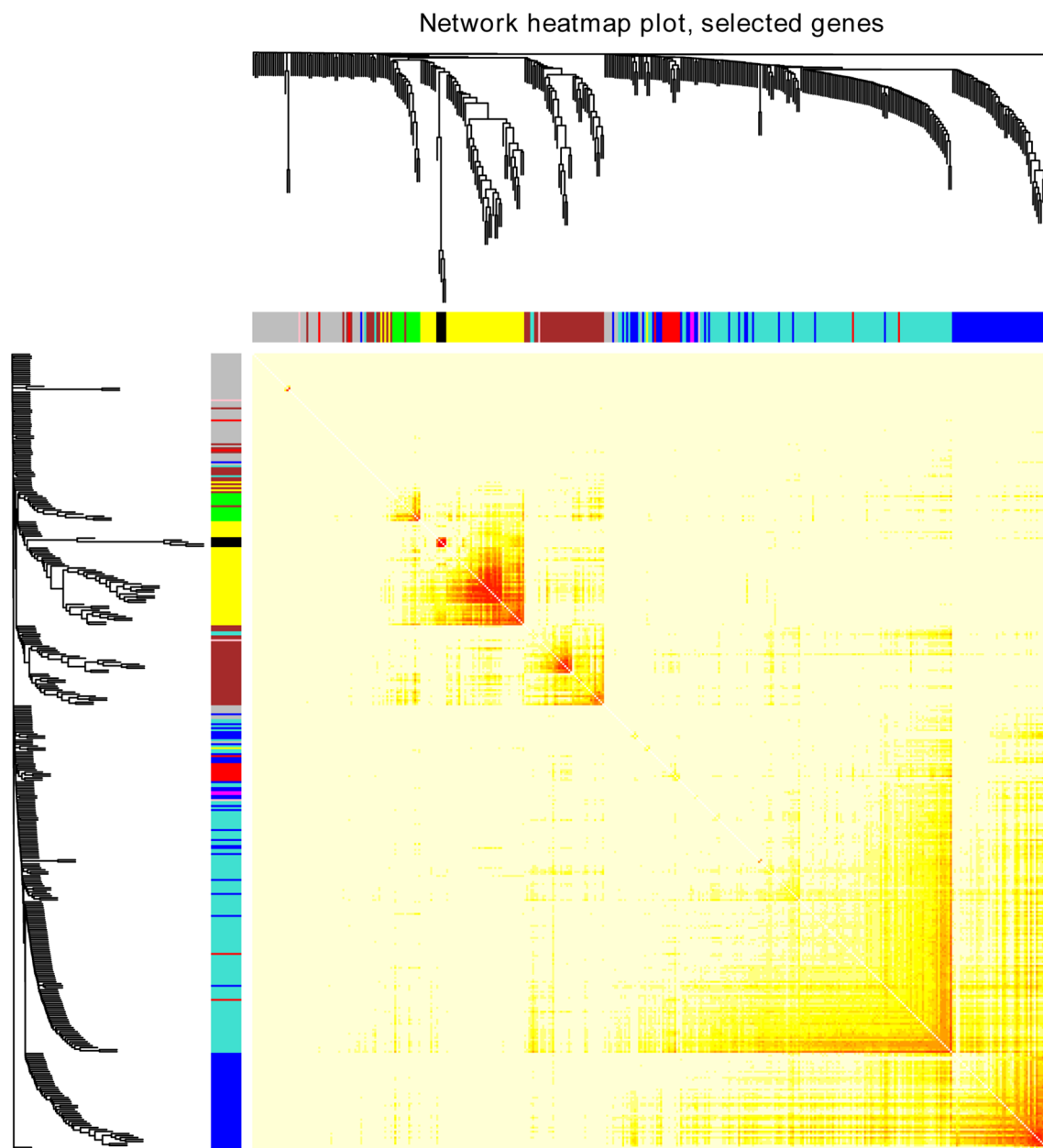
Network heatmap plot, selected genes

Figure 4. Network TOM heatmap plot. The TOM plot is made up of 400 randomly selected genes. Each row and column represents a module and the genes of the module. This diagram shows the degree of correlation within the module.

Survival analysis. Kaplan–Meier survival analysis was performed to investigate overall survival. The results demonstrated that high expression of EGFR and IGF1R may be considered an effective prognostic indicator for ESCA patients (Fig. 11).

Discussion

ESCA is one of the most common malignant tumours worldwide. With the development of medical technology, the treatment and prognosis of ESCA have improved, but it also poses a great threat to human life[24]. Although the pathogenesis of ESCA is still unclear, there are many causes, such as eating excessively hot food, smoking and drinking, obesity and gastroesophageal reflux disease[25,26]. ESCA also differs by region, sex, and race[27]. CKI has been proven to be immensely useful in the treatment of various cancers and also has relief and treatment effects for cancer pain[28]. According to real-world research in hospital systems, the application of CKI in the clinical treatment of ESCA is also famously used[10].

In this study, we used integrated bioinformatics methods to explore the molecular mechanisms of CKI in the treatment of ESCA. WGCNA was employed to analyse key genes in ESCA and combined with network pharmacology to predict the therapeutic mechanism of CKI. Moreover, molecular docking methods were performed to verify the binding affinity of CKI with hub targets to validate the medicinal effects of CKI.
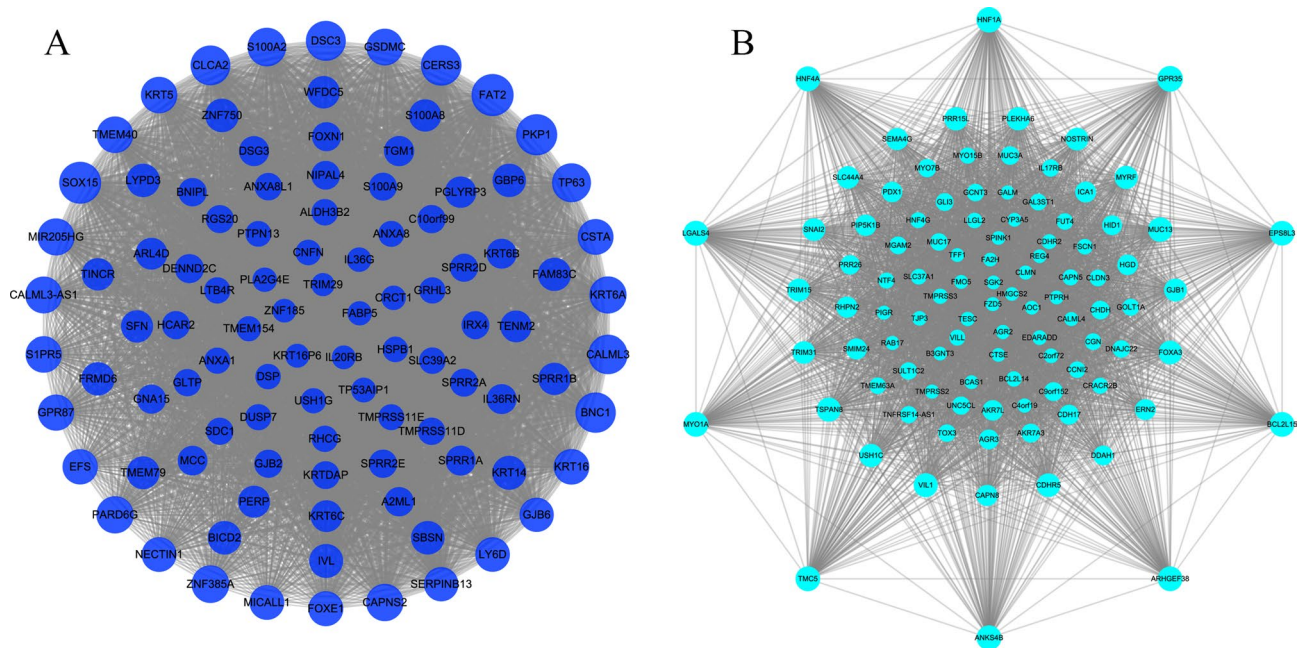
Figure 5. The top 100 genes in the degree of hub modules. (**A**) Blue module. (**B**) Turquoise module.

First, this study used the WGCNA method to explore the pathogenesis of ESCA and to study the link between gene expression data and clinical traits to screen for important gene modules. According to the WGCNA results, a total of 10 related modules were obtained, among which the blue module and the turquoise module were strongly related to multiple clinical traits, such as blue-new tumour event ($p = 5e−04$), blue-histological type ($p = 2e−29$), turquoise-cancer status ($p = 2e−05$) and turquoise-pathologic T ($p = 0.008$). Blue modules and turquoise modules were selected for network pharmacology analysis with the CKI predicted targets. Through network merging and PPI, a total of 16 gene targets were obtained with topological characteristics degree greater than the average. Among them, EGFR (degree = 21), ERBB2 (degree = 18), CCND1 (degree = 16) and IGF1R (degree = 15) were the predicted targets directly corresponding to the CKI components that were considered to be the hub genes related to CKI treatment of ESCA, and detailed discussions were conducted.

Epidermal growth receptor factor (EGFR) is an expression product of the proto-oncogene c-erb-B1 and is a member of subtype 1 of the receptor tyrosine kinase (RTK) family. The family also includes ErbB2, ErbB3 and ErbB4[29]. The binding of EGFR to a ligand activates intracellular tyrosine protein kinase activity, which phosphorylates the terminal tyrosine to activate downstream enzymes *and* initiate downstream signalling[30]. EGFR is overexpressed or mutated in most tumours, resulting in dysregulation of the signal transduction pathway, uncontrolled cell growth, and inhibition of cancer cell apoptosis[31]. Therefore, EGFR-targeted drugs are clinically used in a variety of cancers, and EGFR is also a hot target for tumour diagnosis and treatment. EGFR is highly expressed in both EA and ESCC. In addition, high expression of EGFR is closely related to the proliferation, infiltration and poor prognosis of ESCA cells[32–34]. Thus, abnormal EGFR expression is one of the serious pathogenic factors of ESCA. CKI can treat ESCA by regulating EGFR. ErbB2 is a transmembrane glycoprotein with protein tyrosine kinase (PTK) activity and a proto-oncogene of the human epidermal growth factor receptor family[35]. Overexpression of ErbB2 accelerates tumour growth, metastasis, and tumour blood vessel formation, increasing its invasion in vitro[36]. Equally important, ErbB2 can further improve the ability of tumour cells to migrate and adhere, promote tumour invasion and encourage local and/or distant metastasis[37]. Hoffmann[38] proposed a diagnostic method to detect ErbB2 amplification in single disseminated cancer cells, demonstrating that ErbB2 amplification in esophageal cancer patients is significantly correlated with short-term survival. Previous studies have shown that ESCA can be treated by taking ErbB2 inhibitors such as trastuzumab and ramiximab[39]. The molecular docking results obtained in this study are similar to the docking results of the ErbB2 small-molecule antibody drugs erlotinib and lapatinib in the Rambabu Gundla study[40]. SER-783, THR-862, THR-798 and ASP-863 residues can all generate hydrogen bonds to connect with the compound. Therefore, we believe that CKI can treat ESCA by inhibiting ErbB2. CCND1 is a cell cycle regulating protein that can control the transition of the cell cycle from G1 to S phase, which is closely related to the occurrence and development of many tumours[41]. Matrine is one of the active components of CKI and has been shown to have anti-inflammatory, immunosuppressive, antitumour and antifibrotic effects[42]. Studies have shown that matrine can mediate the expression of CCND1 in breast cancer cells and thus inhibit cancer cells[43]. Guo[44] detected rhabdomyosarcoma cells treated with matrine at different concentrations by MTT, flow cytometry, and RT-PCR and found that matrine significantly inhibits the proliferation of rhabdomyosarcoma cells by reducing the expression of CCND1 mRNA and blocking the cell cycle of the G0/G1 phase. Genome-wide screening revealed that the amplification of cyclin D1 is one of many genetic changes in ESCC. Accordingly, CKI may be used as an inhibitor of CCND1 to treat ESCA[45]. Insulin-like growth factor-1 receptor (IGF1R) is a tyrosine kinase that is involved in the pathogenesis of many cancers. After binding to the ligand, IGF1R can activate PI3K/AKT/mTOR and Ras/Raf/MEK/to activate the MAPK pathway,
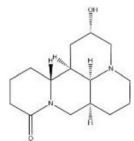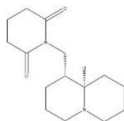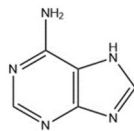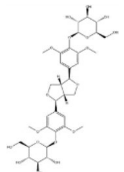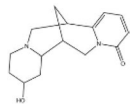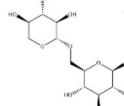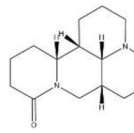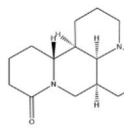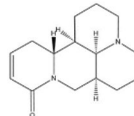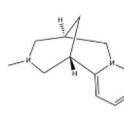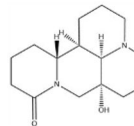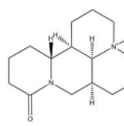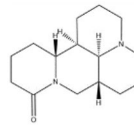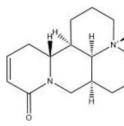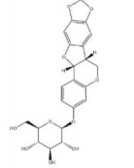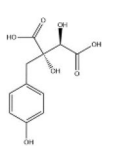
PubChem CID	Compound	Structure	PubChem CID	Compound	Structure
15,385,684	9α-Hydroxymatrine		87,752	Lamprolobine	
190	Adenine		226,371	Liriodendrin	
621,307	Baptifoline		9,576,780	Macrozamin	
5,271,984	Isomatrine		91,466	Matrine	
115,269	Sophocarpine		670,971	N-methylcytisine	
12,442,899	Sophoranol		24,864,132	Oxymatrine	
165,549	Sophoridine		24,721,085	Oxysophocarpine	
442,827	Trifolirhizin		6,710,641	Piscidic acid	

Table 1. Information on the active ingredients of CKI.

which can regulate cell proliferation, survival, differentiation, movement, invasion and angiogenesis[46,47]. Studies have shown that IGF1R is overexpressed in cancer tissues compared to normal tissues adjacent to the cancer. In addition, a mouse xenograft model was used to test the function of IGF-1R in vitro and in vivo. IGF1R was found to have carcinogenic effects in regulating cell proliferation, colony formation, the cell cycle and apoptosis[48,49].

The enriched GO analysis indicated that the hub genes localized mainly to the cyclin-dependent protein kinase holoenzyme complex, protein kinase complex and serine/threonine protein kinase complex, while their molecular functions were associated with the cell cycle. Similarly, KEGG pathway analysis showed enrichment in some cancer pathways and signalling pathways, such as the PI3K-Akt signalling pathway and ErbB signalling pathway. The ErbB family, after binding to its corresponding ligands (EGF, TGF, AR, etc.), downstream related genes such as PI3K/AKT and MAPK can be activated, thereby regulating cell proliferation, differentiation, migration, and apoptosis activities[50]. Studies have shown that the PI3K/AKT pathway is abnormally activated in a variety of cancers, such as esophageal, gastric, and breast cancer[51–53].

Upstream genes such as ErbB2 and various growth factors such as EGF and IGF1 can activate PI3K, resulting in aberrant activation of the PI3K/AKT pathway[54]. Additionally, the deviant activation of the PI3K/AKT pathway inhibits the degradation of CCND1, increases its expression, promotes its shift to the nucleus, and interferes with the transition from G1 to S phase of the cell cycle[55]. The ligands IGF-1 and IGF-2 bind to IGF1R, leading

7

Figure 6. Compound-predicted target network. Sixteen compound nodes are green, and 285 target nodes are yellow.

Figure 7. (**A**) Venn plot of modules and compound-predicted targets. (**B**) Potential target network for the treatment of ESCA with CKI. Blue and turquoise nodes indicate their corresponding colour module genes, and purple nodes represent PPI-related human targets.

to receptor autophosphorylation and the activation of various signalling pathways, including the PI3K/AKT pathway, which leads to cell proliferation and prevents apoptosis[56]. Abnormal activation of pathways caused by abnormal expression of these proteins is a critical factor affecting the progression of esophageal cancer. Zhang[57] found that CKI can increase the ability to inhibit lung cancer cell proliferation and increase sensitivity to gefitinib by downregulating the PI3K/AKT pathway. A study has also shown that matrine derivatives, which are one of the main components of CKI, can be downregulated, CCND1, and attenuated the PI3K/Akt pathway to induce
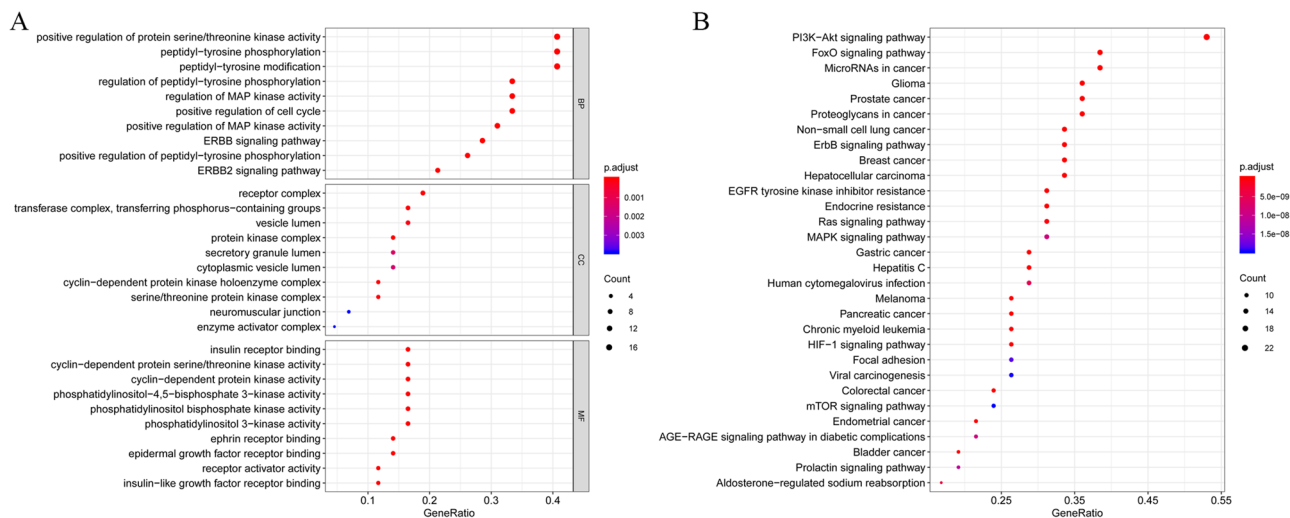
Figure 8. Bubble chart of functional enrichment for hub genes. (**A**) GO function enrichment (**B**) KEGG function enrichment.
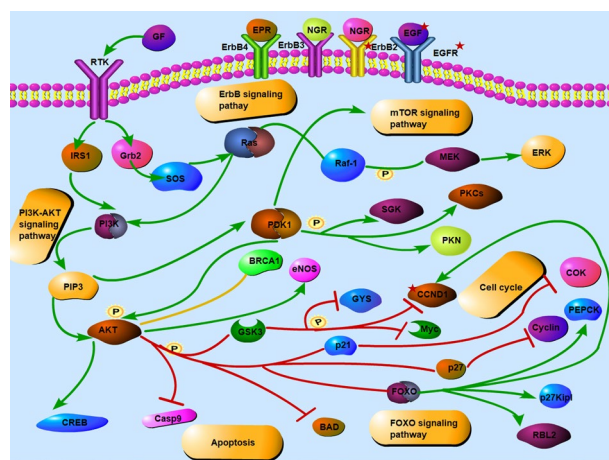

Figure 9. Regulatory pathways are mainly involved in the CKI treatment of ESCA. Green arrows indicate activation, red indicates inhibition, and yellow lines denote binding. P represents phosphorylation.

G1 cell cycle arrest and autophagy in cancer cells through immunofluorescence analysis, western blotting and murine models[58]. Therefore, these potential targets and pathways may be the key to the CKI treatment of ESCA.

The central idea of TCM has a lot in common with network pharmacology, which can explain the treatment process of many complex diseases in a system manner[59]. In previous studies, Li[60] provided a powerful means for identifying mechanisms of Ge–Gen–Qin–Lian decoction in the treatment of type 2 diabetes through the network pharmacology strategy. Liang[61] exploited drugCIPHER to incorporate the traditional network pharmacology concept to analyse the target network of the TCM traditional prescription Liu-Wei-Di-Huang pill. This study was based on the network pharmacology method combined with WGCNA analysis, aiming to accurately detect the genes related to ESCA from the aspect of close to clinical traits to analyse the mechanism of CKI treatment of ESCA. However, there are some limitations in this method. First, our data collection is based on existing database information, so it may produce deviations and incomplete results. Second, biological experiments are urgently needed to validate our results because our study was performed based on data analysis.

Conclusion

In summary, by combining WGCNA and a network pharmacology method, we revealed that CKI controlled the growth of ESCA by regulating potential hub genes, such as EGFR, ErbB2, CCND1 and IGF1R, as well as important related pathways. The study preliminarily verified and predicted the molecular mechanism of CKI against ESCA but still needs further experimental verification. These findings provide insights into the underlying mechanism of CKI for the treatment of ESCA and provide a reference for the study of the more complex mechanism of action of this Chinese herbal compound.
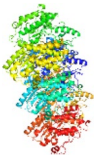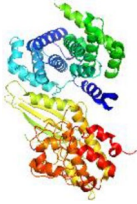
No	Protein name	PDB ID	Protein structure	Test compounds	Affinity (kcal/mol)
1	EGFR	6DUK		Adenine	− 6.3
2	EGFR	6DUK		N-methylcytisine	− 6.7
3	ErbB2	3PP0		Adenine	− 6.2
4	IGF1R	5HZN		Adenine	− 6.0
5	CCND1	6P8G		Matrine	− 6.1

Table 2. Molecular docking information.

Methods

Data collection and preprocessing.   RNA sequencing data in fragments per kilobase million (FPKM) of ESCA were obtained from the TCGA data portal (https://portal.gdc.cancer.gov) in September 2019, with a total of 164 samples. The clinical metadata of the 164 samples were also downloaded and filtered for useful information. After the removal of samples containing incomplete analytical data and/or other malignancies, 161 samples were retained. Since some genes lacked significant changes in expression between samples, we chose the top 5,000 genes that were most important in terms of differential expression for the next WGCNA analysis.

Weighted gene co-expression network analysis and module preservation.   The gene co-expression networks were constructed by the WGCNA package. We used the similarity between gene expression profiles to construct a similarity matrix based on pairwise Pearson correlation coefficient matrices. The similarity matrix was transformed into an adjacency matrix using a power adjacency function[18,62]. The appropriate soft threshold power β was selected by using the integration function (pickSoftThresshold) in the WGCNA software package. With this soft threshold function, the co-expression similarity was improved to achieve a scale-free topology[63,64]. Then, we reconstructed the topological overlap matrix by calculating the topological overlap measure (TOM), which is a robust measure of network interconnectedness[65,66]. The dynamic tree-cut algorithm method was adopted to identify the module of gene co-expression with the maxBlockSize of 6,000, minModuleSize of 30 and mergeCutHeight of 0.2.

Identification of clinically significant modules.   Module eigengene (ME) is the first principal component of each gene module, and the expression of ME is considered representative of all genes in one module. The module membership (MM) is the correlation between the ME and the gene expression profile. Gene Significance (GS) is the absolute value of the correlation between a specific gene and a clinical trait. According to ME, GS, and MM, we can associate modules with clinical traits not only to calculate the correlation between ME and clinical traits but also to analyse clinically vital modules[18].

Construction of predictive target network for CKI components.   The 3D chemical structure data of 16 active ingredients were imported into the Search Tool for Interactions of Chemicals (STITCH)[67], SuperPred[68], SwissTargetPrediction[69] and Traditional Chinese Medicine Systems Pharmacology Database and Analysis Plat-
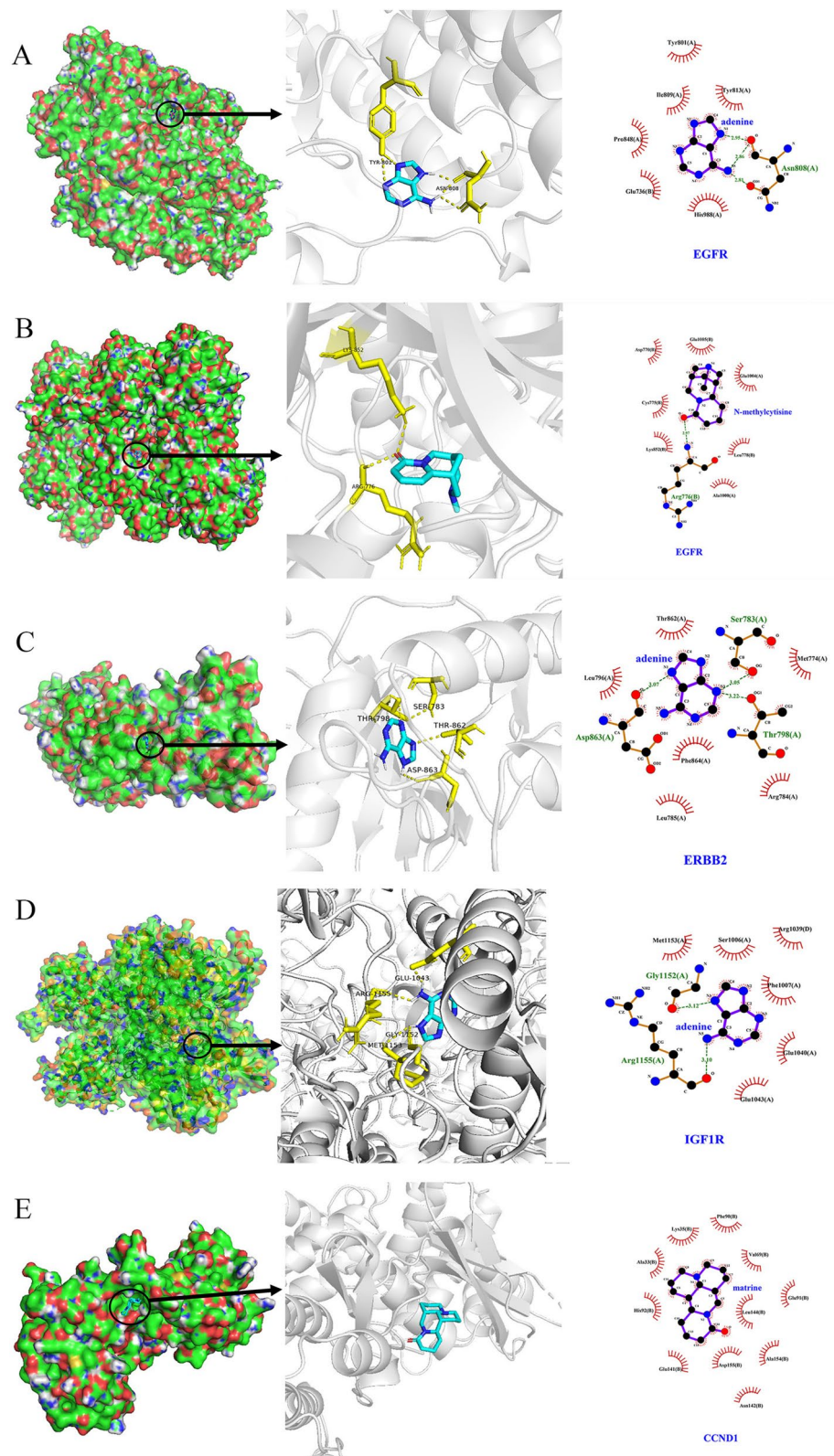
Figure 10. Molecular docking of the hub gene with its corresponding component. (**A**) EGFR with adenine; (**B**) EGFR with N-methylcytisine; (**C**) ErbB2 with adenine; (**D**) IGF1R with adenine; (**E**) CCND1 with matrine.
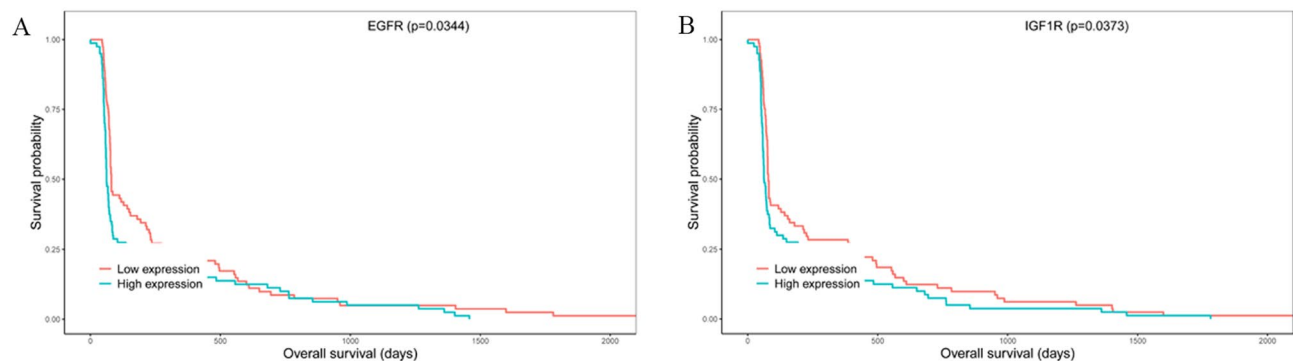
Figure 11. Survival analysis of hub genes. (**A**) EGFR; (**B**) IGF1R.

form (TCMSP) [70] databases for retrieval. The predicted multiple target information of the compounds and the obtained information were introduced into Cytoscape 3.6.1 (https://www.cytoscape.org/) to obtain a compound-predicted target network map. Cytoscape is bioinformatics analysis software that visualizes biological pathways and intermolecular interaction networks and provides a basic set of data integration, analysis and visualization capabilities for complex network analyses[71].

Network construction and analysis of CKI in the treatment of ESCA. The compound-predicted target network and clinically important module network obtained from WGCNA analysis were merged in Cytoscape. The overlapping proteins in the two networks are likely to be potential targets for the treatment of ESCA by the active ingredients of CKI. The Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) 10.5 (https://string-db.org/) is a database of known and predicted protein interactions that contains direct and indirect protein associations[72]. The overlapping genes were input into the STRING 10.5 database, and the species selection "Homo sapiens" was selected as the confidence data with a scoring value greater than 0.7 and 1_{st} shell no more than 20. Then, the data were introduced into Cytoscape to construct a protein–protein interaction (PPI) network.

Gene ontology (GO) functional and kyoto encyclopedia of genes and genomes (KEGG) pathway enrichment analysis. The GO database (https://geneontology.org/) was used to identify the possible biological mechanisms using high-throughput genome or transcriptome data[73]. KEGG is a reference knowledge base for biological interpretation of genome sequences and other high-throughput data, which is a reference for biological interpretation of genome sequences and other high-throughput data[74]. In addition, the R package clusterProfiler was used to perform GO and KEGG functional enrichment analysis[75].

Molecular docking simulation. Molecular docking can reflect the binding energetics of drug molecules to protein receptors by calculating the binding affinity between ligands and receptors and the corresponding intermolecular interactions[76,77]. The potential targets that were directly related to the CKI active components were imported into the Protein Data Bank (PDB) (https://www.rcsb.org/) database to find their 3D structure[78]. Proteins that met the following conditions were considered appropriate protein conformations: (1) The 3D protein structures were determined by X-ray crystallography. (2) Crystal resolution Protein was less than 3 Å; (3) Genotype protein analysis was reliable. Molecular docking simulation of potential targets and their corresponding components was performed using AutoDock 4.2 and AutoDock Vina software according to published methods[79,80]. A suitable grid box size with a spacing of 1.0 Å between grid points was generated to cover almost the entire favourable protein binding site. The X, Y and Z centres are adjusted according to different macromolecular forms. The results of docking are displayed by Pymol and Ligplot[81].

Survival analysis of hub genes. Survival analysis was conducted using the Kaplan–Meier method and log-rank tests. Hazard ratios (HRs) were calculated using a Cox proportional hazards model with R software and the "survival" R package. In the survival analysis, death from any cause was considered an event[82].

Ethics approval and consent to participate. Ethical approval was not necessary in the current study because our study gathered data from TCGA, and this procedure did not address any patients' personal data or harm any patient.

Data availability
CKI compounds, cut off 0.1 blue network nodes and cut off 0.1 turquoise network nodes are available in the Supplementary Source files.

12

References

1. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: Cancer J. Clin.* **68**, 394–424. https://doi.org/10.3322/caac.21492 (2018).
2. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2016. *CA: Cancer J. Clin.* **66**, 7–30. https://doi.org/10.3322/caac.21332 (2016).
3. Chen, W. *et al.* Cancer incidence and mortality in China, 2014. *Chin. J. Cancer Res. = Chung-kuo yen cheng yen chiu* **30**, 1–12. https://doi.org/10.21147/j.issn.1000-9604.2018.01.01 (2018).
4. Coleman, H. G., Xie, S. H. & Lagergren, J. The epidemiology of esophageal adenocarcinoma. *Gastroenterology* **154**, 390–405. https://doi.org/10.1053/j.gastro.2017.07.046 (2018).
5. Esophageal cancer: epidemiology, pathogenesis and prevention. *Nat. Clin. Pract. Gastroenterol. Hepatol.* **5**, 517–526. doi:10.1038/ncpgasthep1223 (2008).
6. Lagergren, J. & Lagergren, P. Oesophageal cancer. *BMJ* **341**, c6280. https://doi.org/10.1136/bmj.c6280 (2010).
7. Parekh, H. S., Liu, G. & Wei, M. Q. A new dawn for the use of traditional Chinese medicine in cancer therapy. *Mol. Cancer* **8**, 21. https://doi.org/10.1186/1476-4598-8-21 (2009).
8. Wang, W. *et al.* Anti-tumor activities of active ingredients in compound Kushen Injection. *Acta Pharmacol. Sin.* **36**, 676–679. https://doi.org/10.1038/aps.2015.24 (2015).
9. Yang, Y. *et al.* Compound kushen injection relieves tumor-associated macrophage-mediated immunosuppression through TNFR1 and sensitizes hepatocellular carcinoma to sorafenib. *J. Immunother. Cancer* https://doi.org/10.1136/jitc-2019-000317 (2020).
10. Liu, Y. X. *et al.* Clinical application features of Fufang Kushen injection in treating malignant esophageal tumor: real world study based on hospital information system. *Zhongguo Zhong Yao Za Zhi* **42**, 2877–2882. https://doi.org/10.19540/j.cnki.cjcmm.20170705.005 (2017).
11. Zhao, Z. *et al.* Fufang Kushen injection inhibits sarcoma growth and tumor-induced hyperalgesia via TRPV1 signaling pathways. *Cancer Lett.* **355**, 232–241. https://doi.org/10.1016/j.canlet.2014.08.037 (2014).
12. Shao, Q. 987 Poster the recent effect of radiotherapy combined with compound Kushen injection for elderly patients with esophageal cancer. *Radiother. Oncol.* **99**, S372–S373 (2011).
13. Zhang, D. *et al.* The optimal chinese herbal injections for use with radiotherapy to treat esophageal cancer: a systematic review and Bayesian network meta-analysis. *Front. Pharmacol.* **9**, 1470. https://doi.org/10.3389/fphar.2018.01470 (2018).
14. Guo, Y. C. *et al.* Network-based combinatorial CRISPR-Cas9 screens identify synergistic modules in human cells. *ACS Synth. Biol.* **8**, 482–490. https://doi.org/10.1021/acssynbio.8b00237 (2019).
15. Zhang, P. *et al.* Dissecting the single-cell transcriptome network underlying gastric premalignant lesions and early gastric cancer. *Cell. Rep.* **27**, 1934–1947 (2019).
16. Guo, Y. C. *et al.* Multiscale modeling of inflammation-induced tumorigenesis reveals competing oncogenic and onco-protective roles for inflammation. *Cancer Res.* **77**, 6429–6441 (2017).
17. Jia, X. *et al.* Tongue coating microbiome as a potential biomarker for gastritis including precancerous cascade. *Protein Cell.* **10**, 496–509 (2018).
18. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform.* **9**, 559.
19. Li, S. & Zhang, B. Traditional Chinese medicine network pharmacology: theory, methodology and application. *Chin. J. Nat. Med.* **11**, 110–120. https://doi.org/10.1016/s1875-5364(13)60037-0 (2013).
20. Hopkins, A. L. Network pharmacology: the next paradigm in drug discovery. *Nat. Chem. Biol.* **4**, 682–690. https://doi.org/10.1038/nchembio.118 (2008).
21. Zheng, J. H. *et al.* Network pharmacology to unveil the biological basis of health-strengthening herbal medicine in cancer treatment. *Cancers* **10**, 461 (2018).
22. Ma, Y. *et al.* Identifcation and determination of the chemical constituents in a herbal preparation, Compound Kushen Injection, by Hplc and Lc-Dad-Ms/Ms. *J. LIQ Chromatogr. Relat. Technol.* **37**, 207–220 (2014).
23. Kim, S. *et al.* PubChem substance and compound databases. *Nucleic Acids Res.* **44**, D1202-1213. https://doi.org/10.1093/nar/gkv951 (2016).
24. Pennathur, A., Gibson, M. K., Jobe, B. A. & Luketich, J. D. Oesophageal carcinoma. *Lancet (London, England)* **381**, 400–412. https://doi.org/10.1016/s0140-6736(12)60643-6 (2013).
25. Toh, Y. *et al.* Alcohol drinking, cigarette smoking, and the development of squamous cell carcinoma of the esophagus: molecular mechanisms of carcinogenesis. *Int. J. Clin. Oncol.* **15**, 135–144 (2010).
26. Arnal, M. J. D., Arenas, ÁF. & Gastroenterology, ÁL. Esophageal cancer: risk factors, screening and endoscopic treatment in Western and Eastern countries. *World J. Gastroenterol.* **21**, 15–25 (2015).
27. Torre, L. A. *et al.* Global cancer statistics, 2012. *CA Cancer J Clin* **65** (2015).
28. Guo, Y. M. *et al.* Efficacy of compound Kushen injection in relieving cancer-related pain: a systematic review and meta-analysis. *Evid.-Based Complement. Altern. Med.: eCAM* **2015**, 840742. https://doi.org/10.1155/2015/840742 (2015).
29. Yarden, Y. & Sliwkowski, M. X. Untangling the ErbB signaling network. *Nat. Rev. Mol. Cell. Biol.* **2**, 127–137 (2001).
30. Jorissen, R. N. *et al.* Epidermal growth factor receptor: mechanisms of activation and signalling. *Exp. Cell. Res.* **284**, 0–53 (2003).
31. Liu, D., Ghiso, J. A. A., Estrada, Y. & Ossowski, L. EGFR is a transducer of the urokinase receptor initiated signal that is required for in vivo growth of a human carcinoma. *J. Cancer Cell.* **1** (2002).
32. Wang, K. L. *et al.* Expression of epidermal growth factor receptor in esophageal and esophagogastric junction adenocarcinomas: association with poor outcome. *Cancer* **109**, 658–667. https://doi.org/10.1002/cncr.22445 (2007).
33. Aichler, M. *et al.* Epidermal growth factor receptor (EGFR) is an independent adverse prognostic factor in esophageal adenocarcinoma patients treated with cisplatin-based neoadjuvant chemotherapy. *Oncotarget* **5**, 6620–6632. https://doi.org/10.18632/oncotarget.2268 (2014).
34. Jiang, D. *et al.* The prognostic value of EGFR overexpression and amplification in Esophageal squamous cell Carcinoma. *BMC Cancer* **15**, 377. https://doi.org/10.1186/s12885-015-1393-8 (2015).
35. Marotta, M. *et al.* Palindromic amplification of the ERBB2 oncogene in primary HER2-positive breast tumors. *Sci. Rep.* **7**, 41921. https://doi.org/10.1038/srep41921 (2017).
36. T, A. *et al.* Significance of serum c-erbB-2 oncoprotein, insulin-like growth factor-1 and vascular endothelial growth factor levels in ovarian cancer. *Bratislavske lekarske listy* **117** (2016).
37. Breyer, J. *et al.* ESR1, ERBB2, and Ki67 mRNA expression predicts stage and grade of non-muscle-invasive bladder carcinoma (NMIBC). *Virchows Arch.* **469**, 547–552 (2016).
38. Hoffmann, M. *et al.* Diagnostic pathology of early systemic cancer: ERBB2 gene amplification in single disseminated cancer cells determines patient survival in operable esophageal cancer. *Int. J. Cancer* **142**, 833–843. https://doi.org/10.1002/ijc.31108 (2018).
39. Ilson, D. H. & van Hillegersberg, R. Management of patients with adenocarcinoma or squamous cancer of the esophagus. *Gastroenterology* **154**, 437–451 (2018).
40. Gundla, R. *et al.* Discovery of novel small-molecule inhibitors of human epidermal growth factor receptor-2: combined ligand and target-based approach. *J. Med. Chem.* **51**, 3367–3377. https://doi.org/10.1021/jm7013875 (2008).
41. Hall, M. & Peters, G. Genetic alterations of cyclins, cyclin-dependent Kinases, and Cdk Inhibitors in Human Cancer. *Adv. Cancer Res.* **68** (1996).

42. Yi, F. *et al.* Novel matrine derivative MD-1 attenuates hepatic fibrosis by inhibiting EGFR activation of hepatic stellate cells. *Protein & Cell* **7** (2016).

43. Li, X. *et al.* Matrine suppression of self-renewal was dependent on regulation of LIN28A/Let-7 pathway in breast cancer stem cells. *J. Cell. Biochem.* https://doi.org/10.1002/jcb.29396 (2019).

44. Guo, L., Xue, T. Y., Xu, W. & Gao, J. Z. Matrine promotes G0/G1 arrest and down-regulates cyclin D1 expression in human rhabdomyosarcoma cells. *Panminerva Med.* **55**, 291–296 (2013).

45. Chen, X. X. *et al.* Genomic comparison of esophageal squamous cell carcinoma and its precursor lesions by multi-region whole-exome sequencing. *Nat. Commun.* **8**, 524. https://doi.org/10.1038/s41467-017-00650-0 (2017).

46. Sharmila, G. *et al.* Chemopreventive effect of quercetin, a natural dietary flavonoid on prostate cancer in in vivo model. *Clin. Nutr.* **33**, 718–726. https://doi.org/10.1016/j.clnu.2013.08.011 (2014).

47. Wilson, S. & Chia, S. K. IGF-1R inhibition: right direction, wrong pathway?. *Lancet Oncol.* **14**, 182–183. https://doi.org/10.1016/S1470-2045(13)70019-6 (2013).

48. Ye, P., Qu, C.-F. & Hu, X.-L. Impact of IGF-1, IGF-1R, and IGFBP-3 promoter methylation on the risk and prognosis of esophageal carcinoma. *Tumour Biol.* **37**, 6893–6904. https://doi.org/10.1007/s13277-015-4489-5 (2016).

49. Ma, W. *et al.* Assessment of insulin-like growth factor 1 receptor as an oncogene in esophageal squamous cell carcinoma and its potential implication in chemotherapy. *Oncol. Rep.* **32**, 1601–1609. https://doi.org/10.3892/or.2014.3348 (2014).

50. Bublil, E. M. & Yarden, Y. The EGF receptor family: spearheading a merger of signaling and therapeutics. *Curr. Opin. Cell Biol.* **19**, 124–134 (2007).

51. Tokunaga, E. *et al.* Activation of PI3K/Akt signaling and hormone resistance in breast cancer. *Breast Cancer* **13**, 137–144 (2006).

52. Beales, I. L. P. *et al.* Activation of Akt is increased in the dysplasia-carcinoma sequence in Barrett's oesophagus and contributes to increased proliferation and inhibition of apoptosis: a histopathological and functional study. *BMC Cancer* **7**, 97 (2007).

53. Ang, K. L., Shi, D. L., Keong, W. W. & Epstein, R. J. Upregulated Akt signaling adjacent to gastric cancers: implications for screening and chemoprevention. *Cancer Lett.* **225**, 53–59 (2005).

54. Pene, F. *et al.* Role of the phosphatidylinositol 3-kinase/Akt and mTOR/P70S6-kinase pathways in the proliferation and apoptosis in multiple myeloma. *Oncogene* **21**, 6587–6597 (2002).

55. Gul, A., Leyland-Jones, B., Dey, N. & De, P. A combination of the PI3K pathway inhibitor plus cell cycle pathway inhibitor to combat endocrine resistance in hormone receptor-positive breast cancer: a genomic algorithm-based treatment approach. *Am J. Cancer Res.* **8**, 2359–2376 (2018).

56. Doyle, S. L. *et al.* IGF-1 and its receptor in esophageal cancer: association with adenocarcinoma and visceral obesity. *Am. J. Gastroenterol.* **107**, 196–204. https://doi.org/10.1038/ajg.2011.417 (2012).

57. Zhang, J. *et al.* An effective drug sensitizing agent increases gefitinib treatment by down regulating PI3K/Akt/mTOR pathway and up regulating autophagy in non-small cell lung cancer. *Biomed. Pharmacother* **118**, 109169. https://doi.org/10.1016/j.biopha.2019.109169 (2019).

58. Wu, L. *et al.* Synthesis and biological evaluation of matrine derivatives containing benzo-α-pyrone structure as potent anti-lung cancer agents. *Sci. Rep.* **6**, 35918. https://doi.org/10.1038/srep35918 (2016).

59. Zheng, J. *et al.* Network pharmacology to unveil the biological basis of health-strengthening herbal medicine in cancer treatment. *Cancers (Basel)* https://doi.org/10.3390/cancers10110461 (2018).

60. Li, H. *et al.* A network pharmacology approach to determine active compounds and action mechanisms of ge-gen-qin-lian decoction for treatment of type 2 diabetes. *Evid. Based Complement. Alternat. Med.* **2014**, 495840. https://doi.org/10.1155/2014/495840 (2014).

61. Liang, X., Li, H. & Li, S. A novel network pharmacology approach to analyse traditional herbal formulae: the Liu–Wei–Di–Huang pill as a case study. *Mol. Biosyst.* **10**, 1014–1022. https://doi.org/10.1039/c3mb70507b (2014).

62. Yepes, S. *et al.* Co-expressed miRNAs in gastric adenocarcinoma. *Genomics* **108**, 93–101. https://doi.org/10.1016/j.ygeno.2016.07.002 (2016).

63. Jeong, H., Mason, S. P., Barabási, A. L. & Oltvai, Z. N. Lethality and centrality in protein networks. *Nature* **411**, 41–42. https://doi.org/10.1038/35075138 (2001).

64. Tang, J. *et al.* Prognostic genes of breast cancer identified by gene co-expression network analysis. *Front. Oncol.* **8**, 374. https://doi.org/10.3389/fonc.2018.00374 (2018).

65. Yip, A. M. & Horvath, S. Gene network interconnectedness and the generalized topological overlap measure. *BMC Bioinform.* **8**, 22. https://doi.org/10.1186/1471-2105-8-22 (2007).

66. Li, A. & Horvath, S. Network neighborhood analysis with the multi-node topological overlap measure. *Bioinformatics (Oxford, England)* **23**, 222–231. https://doi.org/10.1093/bioinformatics/btl581 (2007).

67. Szklarczyk, D. *et al.* STITCH 5: augmenting protein-chemical interaction networks with tissue and affinity data. *Nucleic Acids Res.* **44**, D380-384. https://doi.org/10.1093/nar/gkv1277 (2016).

68. Nickel, J. *et al.* SuperPred: update on drug classification and target prediction. *Nucleic Acids Res.* **42**, W26-31. https://doi.org/10.1093/nar/gku477 (2014).

69. Gfeller, D. *et al.* SwissTargetPrediction: a web server for target prediction of bioactive small molecules. *Nucleic Acids Res.* **42**, W32-38. https://doi.org/10.1093/nar/gku293 (2014).

70. Ru, J. *et al.* TCMSP: a database of systems pharmacology for drug discovery from herbal medicines. *J. Cheminform.* **6**, 13. https://doi.org/10.1186/1758-2946-6-13 (2014).

71. Franz, M. *et al.* Cytoscape.js: a graph theory library for visualisation and analysis. *Bioinformatics (Oxford, England)* **32**, 309–311. https://doi.org/10.1093/bioinformatics/btv557 (2016).

72. Szklarczyk, D. *et al.* The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* **45**, D362-d368. https://doi.org/10.1093/nar/gkw937 (2017).

73. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29. https://doi.org/10.1038/75556 (2000).

74. Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K. & Tanabe, M. New approach for understanding genome variations in KEGG. *Nucleic Acids Res.* **47**, D590-d595. https://doi.org/10.1093/nar/gky962 (2019).

75. Sun, M. *et al.* Antitumor activities of kushen: literature review. *Evid. Based Complement. Alternat. Med.* **2012**, 373219. https://doi.org/10.1155/2012/373219 (2012).

76. Ferreira, L. G., Dos Santos, R. N., Oliva, G. & Andricopulo, A. D. Molecular docking and structure-based drug design strategies. *Molecules (Basel, Switzerland)* **20**, 13384–13421. https://doi.org/10.3390/molecules200713384 (2015).

77. Huang, S. Y. & Zou, X. Advances and challenges in protein-ligand docking. *Int. J. Mol. Sci.* **11**, 3016–3034. https://doi.org/10.3390/ijms11083016 (2010).

78. Rose, P. W. *et al.* The RCSB protein data bank: integrative view of protein, gene and 3D structural information. *Nucleic Acids Res.* **45**, D271-d281. https://doi.org/10.1093/nar/gkw1000 (2017).

79. Forli, S. *et al.* Computational protein-ligand docking and virtual drug screening with the AutoDock suite. *Nat. Protoc.* **11**, 905–919. https://doi.org/10.1038/nprot.2016.051 (2016).

80. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461. https://doi.org/10.1002/jcc.21334 (2010).

81. Laskowski, R. A. *et al.* LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J. Chem. Inf. Model.* **51**, 2778–2786 (2011).
82. Zhang, Z. *et al.* Integrating clinical and genetic analysis of perineural invasion in head and neck squamous cell carcinoma. *Front. Oncol.* **9**, 434. https://doi.org/10.3389/fonc.2019.00434 (2019).

Acknowledgements

Author contributions

Z.W. and W.J.R. conceived and designed the study. J.S.S., Z.J.Y., G.S.Y. and L.X.K. collected the data. Z.X.M., Z.Y.L. and W.M.M. performed the data analysis, Z.W. and W.J.R. wrote and revised the manuscript. All authors were responsible for reviewing the data. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-69708-2.

Correspondence and requests for materials should be addressed to J.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.