

RESEARCH

Open Access



Screening the components of *Saussurea involucrata* for novel targets for the treatment of NSCLC using network pharmacology

Dongdong Zhang[†], Tieying Zhang[†], Yao Zhang, Zhongqing Li, He Li, Yueyang Zhang, Chenggong Liu, Zichao Han, Jin Li^{*} and Jianbo Zhu^{*}

Abstract

Background: *Saussurea involucrata* (SAIN), also known as Snow lotus (SI), is mainly distributed in high-altitude areas such as Tibet and Xinjiang in China. To identify novel targets for the prevention or treatment of lung adenocarcinoma and lung squamous cell carcinoma (LUAD&LUSC), and to facilitate better alternative new drug discovery as well as clinical application services, the therapeutic effects of SAIN on LUAD&LUSC were evaluated by gene differential analysis of clinical samples, compound target molecular docking, and GROMACS molecular dynamics simulation.

Results: Through data screening, alignment, analysis, and validation it was confirmed that three of the major active ingredients in SAIN, namely quercetin (Q), luteolin (L), and kaempferol (K), mainly act on six protein targets, which mainly regulate signaling pathways in cancer, transcriptional misregulation in cancer, EGFR tyrosine kinase inhibitor resistance, adherens junction, IL-17 signaling pathway, melanoma, and non-small cell lung cancer. In addition, micro-RNAs in cancer exert preventive or therapeutic effects on LUAD&LUSC. Molecular dynamics (MD) simulations of Q, L, or K in complex with EGFR, MET, MMP1, or MMP3 revealed the presence of Q in a very stable tertiary structure in the human body.

Conclusion: There are three active compounds of Q, L, and K in SAIN, which play a role in the treatment and prevention of non-small cell lung cancer (NSCLC) by directly or indirectly regulating the expression of genes such as MMP1, MMP3, and EGFR.

Keywords: *Saussurea involucrata*, LUAD (Lung adenocarcinoma), LUSC (Lung squamous cell carcinoma), Non-small cell lung cancer, Molecular dynamics

Introduction

In 2021, among many types of cancers, lung cancer is still among the tough, not only because its mortality is as high as 22% in both men and women [1], but also because its treatment is challenging.

According to the microscopic morphological characteristics of lung cancer cells, lung cancer can be divided into two types: small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC) [2]. In previous studies, it has been shown that the two-year survival rate of SCLC is very low, hovering around 14% [2]. Fortunately, the two-year survival

*Correspondence: lijin@shzu.edu.cn; zhujianboshz@163.com

[†]Dongdong Zhang and Tieying Zhang contributed equally to this work.
School of Life Sciences, Shihezi University, Xiangyang street,
Shihezi 832003, PR China



rate of NSCLC can reach 42% [2]. In addition, NSCLC accounts for 85% of all cases of lung cancer, which is good news for investigators in cancer-targeted therapy and NSCLC patients [3].

Currently, the world's targeted therapeutics for NSCLC have issues, including high R&D cost, a relatively expensive market price, and a long administration cycle. Therefore, these restrictions do not allow every NSCLC patient to use these drugs normally. Lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) are the two most common types of NSCLC, among which, LUAD accounts for far more cases than LUSC, and has become the major type of targeted therapy for NSCLC.

Saussurea involucrata (SAIN) is mainly distributed in high altitude areas such as Xinjiang and Tibet in China. Research on the pharmacological effects of SAIN started in 1980 [4] and has continued for more than 40 years now. There are many known pharmacological activities of SAIN, but the most studied effects are its anti-inflammatory, antioxidant [5], and anticancer (prostate, gastric, and breast cancer, etc.) effects [6]. The effects of SAIN on other diseases (e.g., lung cancer) have not yet been studied. Therefore, studies are urgently needed to evaluate and supplement the efficacy of SAIN for the treatment of lung cancer.

Network pharmacology is a new discipline that is based on the theory of systems biology to select specific signal nodes for multi-target drug molecule design [7]. Network pharmacology emphasizes the multi pathway regulation of signaling pathways to improve the therapeutic efficacy of drugs and to reduce toxic side effects, thereby improving the success rate of clinical trials of novel drugs and saving the cost of drug research and development [7]. Network pharmacology is a newly emerging research method in recent years, which was first proposed by Professor Hopkins, a pharmacologist at Dundee University, United Kingdom [7]. Therefore, network pharmacology methods were used, combining systems pharmacology, gene differential analysis, molecular docking, subcellular localization prediction, patient prognosis survival analysis, and molecular dynamics simulation. In this study, the mechanism of action of the targets for the treatment and prevention of LUAD and LUSC by the active ingredients in SAIN.

Materials and Methods

Screening of active ingredients with drug resistance in SAIN

Traditional Chinese Medicine Systems Pharmacology Database (TCMSP) (<https://tcmssp-e.com/>) [8] is one of the most commonly used databases for active ingredient screening of traditional Chinese medicine. The advantage of TCMSP lies in that the parameters of oral bioavailability (OB) and drug likeness (DL) are available from

this database. OB and DL are important for the evaluation of drug efficacy, only when the OB exceeds a certain value ($OB \geq 30\%$) and the DL is within a certain range ($DL \geq 0.18$) when only able to effectively reflect the class resistance of a certain ingredient.

Among them, the DL value for this system is calculated following Eq. (1), and to obtain the drug of interest, the DL value is functional only if the lead compound is chemically easily synthesized and has the properties of absorption, distribution, metabolism, excretion (ADME).

$$T(x, y) = \frac{x - y}{|x|^2 + |y|^2 - xy} \quad (1)$$

In the equation, x represents the descriptive index of all ingredients in SAIN, and y represents the descriptive index from the drug bank (<https://www.drugbank.ca>) database [9] for the average drug similarity index of an ingredient.

The lipid water partition coefficient $\log P_{O/W}$ refers to the partition coefficient of the drug in the *n*-octanol–water system and is widely used as a measure of the hydrophobicity of chemical constituents. This represents the main driving force for effective ingredient permeation through biological membranes composed of lipid bilayers, thereby controlling the ingredient target binding effect. The calculation of $\log P_{O/W}$ is presented in Eq. (2):

$$\log P_{\frac{O}{W}} = \frac{\log C_O}{\log C_W} \quad (2)$$

In the equation, C_O represents the equilibrium concentration of the drug in the oil phase and C_W represents the equilibrium concentration of the drug in the aqueous phase. The magnitude of the $\log P_{O/W}$ value represents the magnitude of solute hydrophobicity. The larger the $\log P_{O/W}$, the stronger the hydrophobicity and vice versa. Therefore, $\log P_{O/W} \leq 5$ was used as the screening criterion.

Target prediction and construction of an active ingredient target network.

The structural formula of the active ingredients of SAIN obtained above (Fig. 5c) was plotted on the STP (<http://www.swisstargetprediction.ch>) database [10] for target prediction. To obtain the ultimately required disease target information, the predicted results were merged with the NSCLC targets retrieved on the TCMSP database. Next, Cytoscape 3.8.0 software [11] was used for visual construction of the ingredient target network (C – T network) for the above active ingredients and targets to obtain the C – T network relationship diagram. Herein, a primary screen for lung cancer-related active ingredients of SAIN was completed.

Gene differential analysis and gene TPM data analysis for LUAD&LUSC

To further screen the active ingredients and their targets that were initially screened, a large number of clinical case samples of LUAD&LUSC were obtained from the GEO (<https://www.ncbi.nlm.nih.gov/geo/>) database [12]. Differential gene analyses of ten thousand samples, which were randomly selected samples used to draw the differential heat map of relevant genes, were performed for the primary screening of targets. To make the secondary screening of targets more explicit, gene IDs of targets were used to draw a quantitative scatter plot of transcripts per million (TPM) of the screened genes based on the TCGA data provided by the GEPIA (<http://gepia.cancer-pku.cn/>) database [13]. In addition, the effective ingredients and their action targets for the treatment of LUAD&LUSC in SAIN were further screened out.

Gene enrichment analysis, construction of protein–protein interaction network, and ingredients–targets pathway network

To evaluate the mutual influence among the targets screened above, a protein interaction network and an active ingredient target pathway network were constructed. The biological processes and pathways that each target participates in *in vivo* were explored, and data of target genes were retrieved through the David (<https://david.ncifcrf.gov/>) database [14]. Biological process (BP) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) [15] enrichment analysis in Gene Ontology (GO) analysis were performed. Of these, $FDR \leq 0.05$ for GO analysis and $FDR \leq 0.05$ for KEGG enrichment analysis, both meet the requirement of genes to be significantly enriched *in vivo* were statistically significant.

The FDR value is a corrected p-value, and the results screened out with FDR are also more precise. Therefore, this step aims to obtain the biological process and *in vivo* pathway of target action, which will provide the basis for subsequent studies.

Molecular docking and subcellular localization prediction

Molecular docking was performed by tools, including Autodock 4.2.6 software [16] and PyMol software [17]. In brief, the binding energy size of ingredients to targets was first used to verify whether the effect of the active ingredients from SAIN in treating LUAD&LUSC was reliable, and to exclude the ingredients from SAIN that were less effective on LUAD&LUSC.

Molecular docking is an approach for drug design that uses the characteristics of receptors and the mode of interaction between receptors and drug molecules [18].

One theoretical simulation approach is to primarily study the interactions between molecules, such as ligands and receptors, and to predict their binding modes and avidities. In recent years, the molecular docking method has become an important technique in the field of computer-assisted drug research [19].

In recent years, subcellular localization prediction has been a more popular method of subcellular localization. Existing data to create a database of the sequence relationships of the target sequences and subcellular structures of various genes and their regulation can accurately predict the location of the target protein on various organelles and cell membranes, and has multiple advantages. Currently used subcellular localization prediction tools include (1) the PSORT II (<https://psort.hgc.jp/form2.html>) database [20], which uses the k-nearest neighbor (k-NN) algorithm, a commonly used learning algorithm in data mining and machine learning. K-NN is widely used, and PSORT II can be used to identify a classical nuclear localization signal (CNLS) sequence. Accuracy of PSORT II is very high under conditions with a large sample size [21]; (2) The CELLO (<http://cello.life.nctu.edu.tw/>) database [22] uses the SVM-RFE algorithm (Support vector machine recursive feature elimination) to construct ranking coefficient based on weight vector w generated by SVM at training. IT removes one feature attribute with the smallest ranking coefficient for each iteration, thereby finally ranking all feature attributes in decreasing order [23]; (3) The BUSCA (<http://busca.biocomp.unibo.it/>) database [24] employs the betaware algorithm and aims to solve the detection of TMBB (transmembrane beta-barrels) in proteomes and the prediction of its topology [24, 25].

Combined with the analysis of biological processes, using PSORT II, CELLO, and BUSCA databases for subcellular location prediction, respectively, and to compare the prediction results of the three tools, the intersection part was selected and the main location of the target protein in the cell was obtained.

Prognostic survival analysis of patients

In clinical studies, to evaluate the efficacy of a drug and know the survival data, such as the survival time of patients after surgery, our analysis of these survival data is called survival analysis.

After a series of screening and analysis of the ingredients and action targets of SAIN, prognostic overall survival analysis was performed on the gene IDs of the targets of the final active ingredient actions in SAIN, so as to explore the length of time that patients can survive when drugs act on these targets over time.

Molecular dynamics simulation validation by GROMACS

Based on molecular docking studies, differential gene analysis, and the prediction of subcellular localization and survival analysis, the results were validated with molecular dynamics (MD) simulations, a means to computationally model the motion of a small molecule in an organismal environment [26].

MD simulations were performed with GROMACS software [27], thereby setting the physical conditions to a constant temperature (305 K), constant pressure (101 kPa), and periodic boundary conditions to simulate the human body using the TIP3P water model in a neutral sodium chloride solution of 0.145 M.

After state equilibration of all environments, we performed 50 ns MD simulations of the ingredient target complex system using epidermal growth factor receptor (EGFR) with L, Q, matrix metalloproteinase 3 (MMP3) with Q, hepatocyte growth factor receptor (MET) with L, and matrix metalloproteinase 1 (MMP1) with L, Q, and K, in which stored conformations were calculated every 10 ps (1 s = 10¹² ps). In addition, the root mean square deviation (RMSD) of the MD simulation results analysis and visualization were performed using a GROMACS inbuilt program with the RMSD formula as follows (3).

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^{i=N} (R_t - R_{ref})^2} \quad (3)$$

In this equation, $R_t - R_{ref}$ represents the position of the t -th atom at a certain frame minus its position in the reference conformation ref (positional offset), and N refers to the number of atoms.

Results

Screening results of active ingredients in SAIN

In this study, a total of 55 known active ingredients in SAIN were collected from the TCMSP database, including alkaloids, lipids, flavonoids, and flavonoids. Subsequently, six potent drug-resistant ingredients from 55 compounds were screened under the following screening conditions of $OB \geq 30\%$ & $DL \geq 0.18$ & $\log Po/w \leq 5$. Details of the screening process of these six components, Flazin

(F), Q, K, L, Alloisioimperatorin (A), and Hispidulin (H), are shown in Table 1.

Results of targets screening and ingredients-targets network analysis

By merging 1023 targets screened from the STP database and 764 LUAD and LUSC targets screened from the TCMSP database, a total of 9 targets with six ingredient roles was obtained. The nine protein targets interacted with the ingredients of six SAIN, and an ingredient target network was constructed using Cytoscape, resulting in an ingredient target network characterized by 15 nodes and 24 edges (Fig. 1).

As determined by the C-T network, the ranking of six components with strong and weak target effects of SAIN was as follows: Q, L, K, and H, F, and A. Moreover, the rank order of the nine target and component effects was as follows: prostaglandin endoperoxide synthase 2 (PTGS2), 90 kDa heat shock protein AA1 (Hsp90AA1), MMP1, and EGFR.

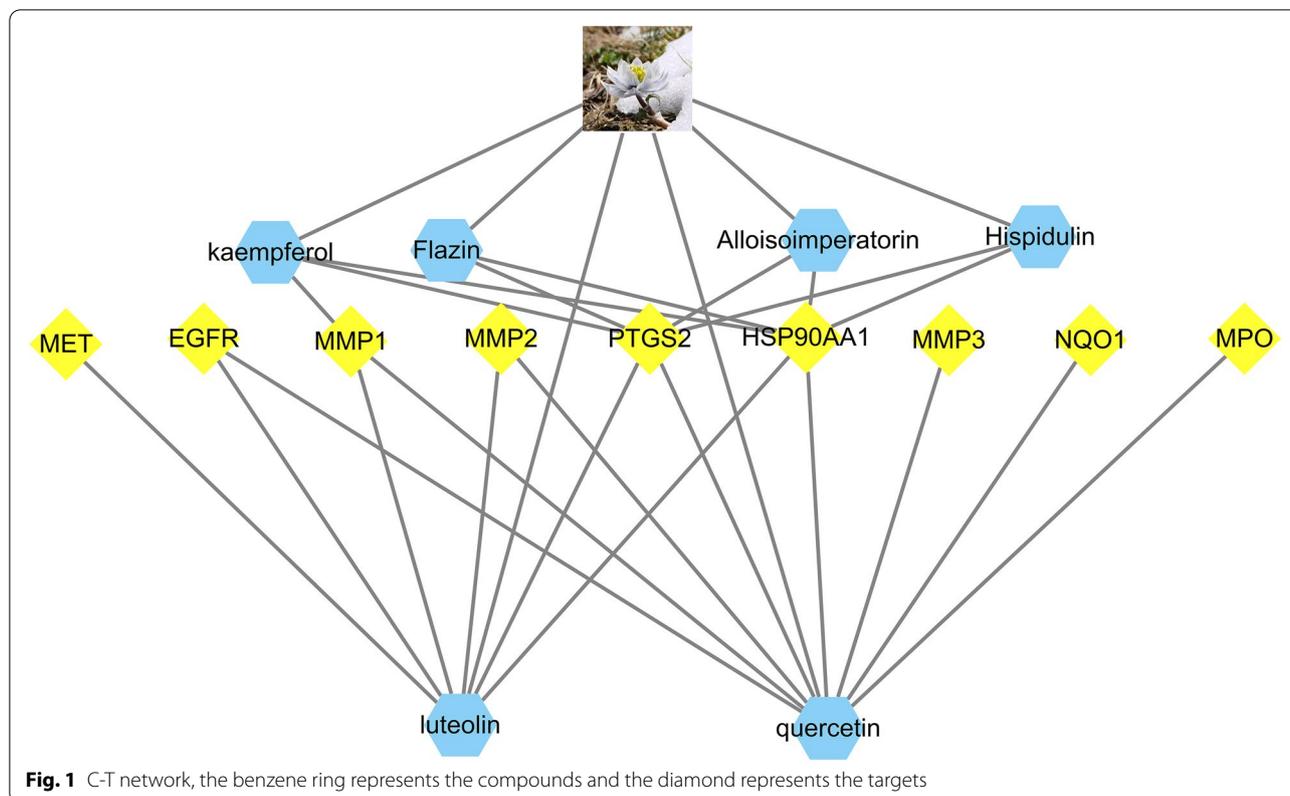
Differential analysis of genes and TPM data analysis results

To ensure the accuracy of the experimental data, further screening for ingredients and targets was performed. In this study, data from a total of 22,589 LUAD and 22,697 LUSC related disease genes was obtained through the GEO database and R 4.0.4, TCGA data through the GDC (<https://portal.gdc.cancer.gov/>) database [28]. A total of 535 LUAD samples and 59 LUAD control group samples was obtained, respectively, and 502 LUSC samples and 49 LUSC control samples were obtained. Undeniably, these data are all highly informative [28].

To make the experiment go smoothly, data were selected by combining the targets and clinical samples using R language with the nine targets under study as the targets of interest by using the p-value range ($P \leq 0.05$) of these data and the patient samples to obtain the gene expression matrix of two groups, LUAD & Normal and LUSC & Normal, corresponding to the nine targets. In this study, 15 samples from each of the two groups within the two cohorts were selected, including 30 clinical samples from each of the LUSC & control group in

Table 1 Screening information of six compounds in SAIN

Molecule Name	MW(KDa)	logP(o/w)	OB(%)	DL	PubChem CID
Flazin	308.31	3.235	94.27575085	0.38559	5,377,686
Quercetin	302.25	1.504	46.43334812	0.27525	5,280,343
Kaempferol	286.25	1.771	41.88224954	0.24066	5,280,863
Luteolin	286.25	2.067	36.16262934	0.24552	5,280,445
Alloisioimperatorin	270.3	3.793	34.80406732	0.21854	5,317,436
Hispidulin	300.28	2.318	30.97205344	0.27025	5,281,628



LUAD and LUSC, and the gene differential heat maps of LUAD (Fig. 2B) and LUSC (Fig. 2a The direction of the ordinate is “←”) were plotted using TBtools v1.082 software [29]. Since the nine targets can act on both types of lung cancer, a gene cluster was created, since our samples were divided into two groups (experimental and normal), to make the sample dispersion uniform, and without affecting the experimental results, we do not cluster the samples.

The results of gene difference analysis showed that the difference of Hsp90AA1 between LUAD and LUSC was least obvious, therefore, this was discarded. The gene IDs of the remaining eight targets were TPM analyzed by GEPIA database, input gene ID, one-way analysis of variance (ANOVA) with Log_2FC value, and Q value of 1 and 0.01 were selected by the difference method [30], and the total number of involved samples were as follows: LUAD (T=483, N=347), LUSC (T=486, N=338), in which the ‘T’ represented tumor patients and the ‘N’ represented normal controls (Fig. 3).

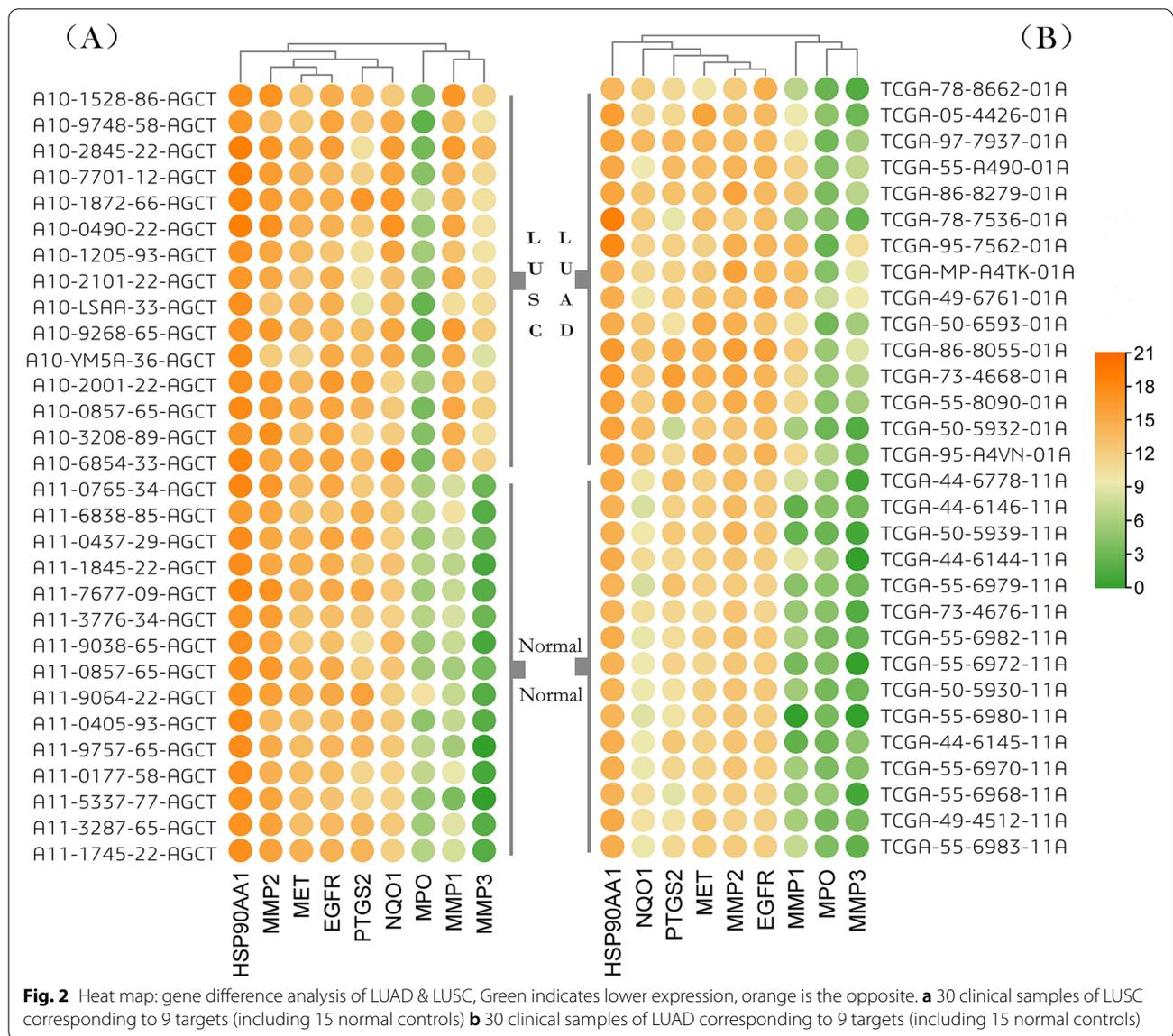
When combining the correspondence between components and targets, and the results of gene difference analysis and TPM analysis, the data showed that the expression and changes of MMP2 and PTGS2 in group T were consistent with that in group N, therefore, they were discarded.

The SAIN components with major effects on LUAD and LUSC were Q, I, and K, and the genes with major effects on LUAD and LUSC were MMP1, MMP3, EGFR, MET, NAD (P) H: quinone oxidoreductase 1 (NQO1), and myeloperoxidase (MPO).

For the analysis of six genes, we concluded that the expression of two genes, EGFR and MMP3, were much more effective in LUSC than in LUAD, whereas MET had the opposite effect. Although the other three genes (MMP1, NQO1 and MPO) showed consistent expression in the occurrence and development of LUAD and LUSC, MPO was of interest as this gene showed full-stage high expression in cases in the N group and low expression in cases in the T group. Therefore, we speculated that overexpression of MPO in tissues might be effective in inhibiting the development of LUAD and LUSC.

GO (BP), KEGG enrichment analysis, PPI analysis, and C-T pathway analysis results

Based on the above-mentioned analysis, three targets, included Hsp90AA1, PTGS2, and MMP2 were excluded, and previous studies have shown that Hsp90AA1 and the isotype Asp90AB1 may be required for the treatment of two subtypes of breast cancer (MCF-7 and BT-474) [31]. PTGS2 is required in diseases, such as colorectal cancer



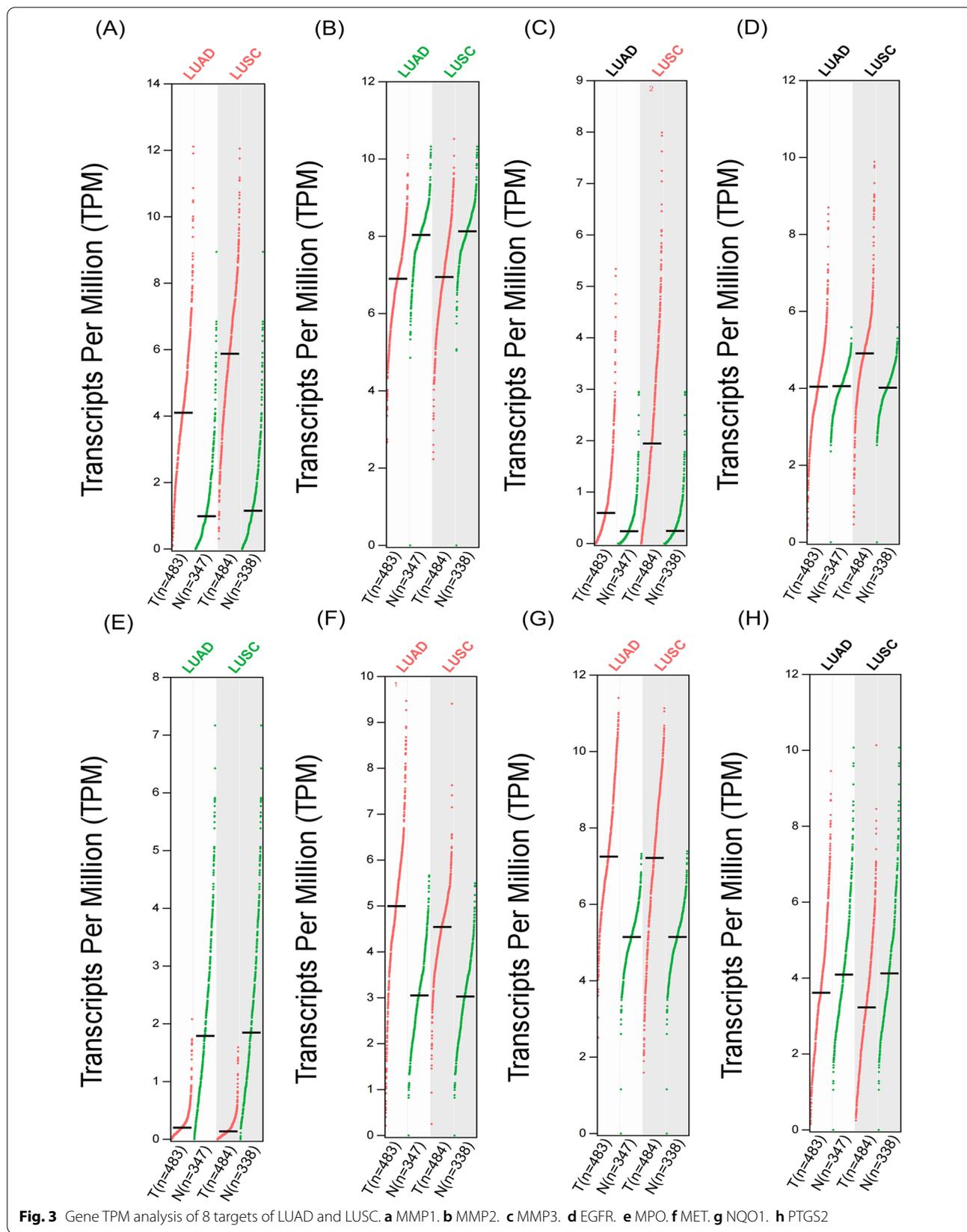
[32], while MMP2 mostly plays a role in inflammatory responses and immune responses [33].

The remaining six targets (EGFR, MET, MMP1, MMP3, NQO1 and MPO) were subjected to undergo BP analysis (Fig. 4a), KEGG enrichment analysis (Fig. 4b) [15], and PPI (protein protein interaction) analysis (Fig. 5a).

To achieve credibility and statistical significance of the data, we performed GO (BP) analysis and KEGG enrichment analysis at $FDR \leq 0.05$. Referring to the results of KEGG enrichment analysis, and to make it easy to understand, we used Cytoscape to associate the three components, six targets, and 20 pathways to draw a component target pathway diagram (Fig. 5b). This diagram shows the relationship between components and targets. The

relationship between targets and pathways is visualized, and a total of 20 metabolic regulatory pathways, including 29 nodes and 53 edges were obtained.

First, from the C-T-P (compounds-target-pathway) network diagram, targets were sorted into EGFR, MET, MMP1, MMP3, NQO1, and MPO based on the number of regulatory pathways and corresponding components, and pathways were ranked into pathways in cancer, transcriptional misregulation in cancer, hepatocellular carcinoma, bladder cancer, and EGFR tyrosine kinase inhibitor resistance. According to the degree of correlation adherens junction, IL-17 signaling pathway, melanoma, non-small cell lung cancer, relaxin signaling pathway and microRNAs in cancer, etc. However,



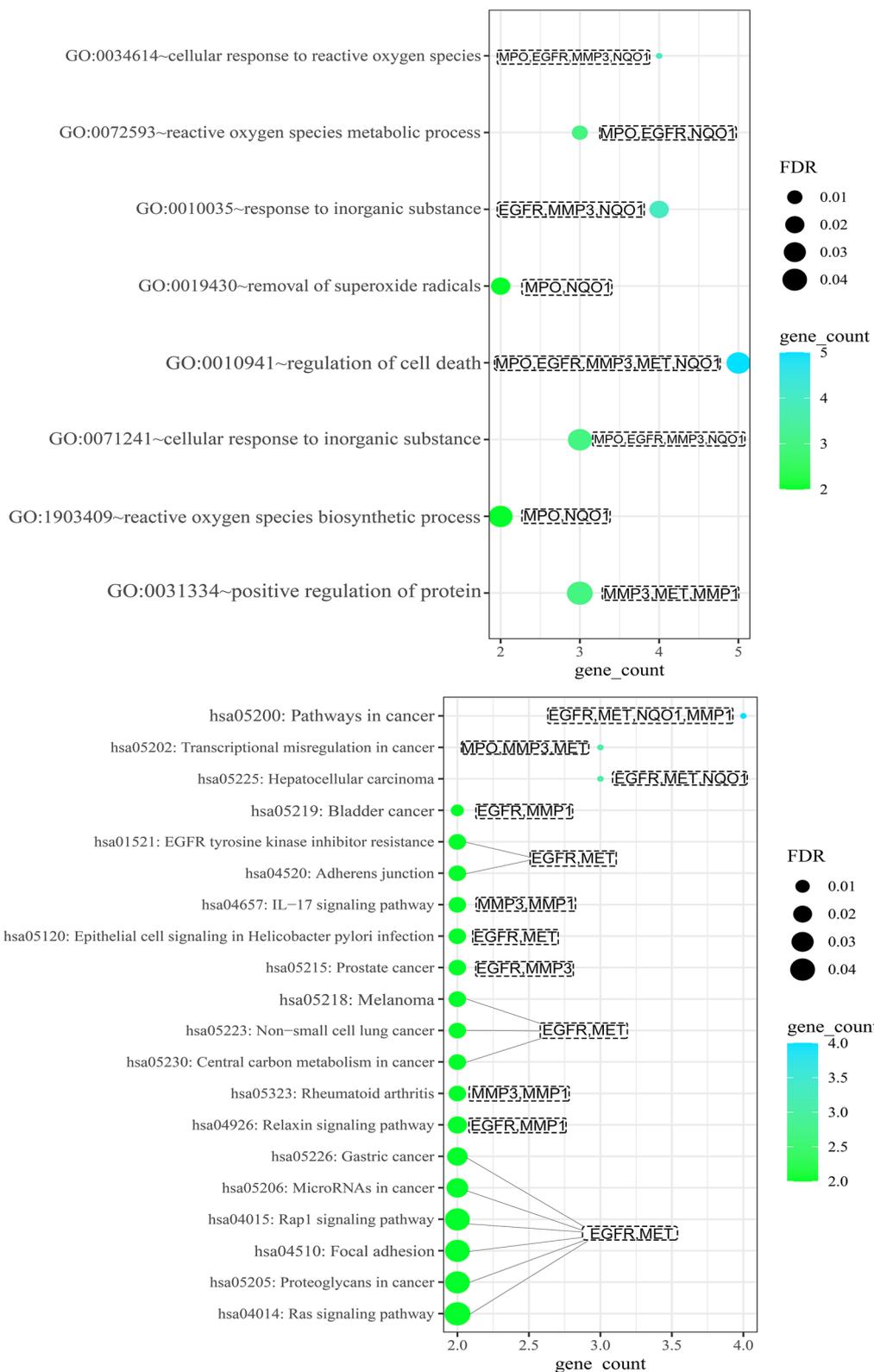
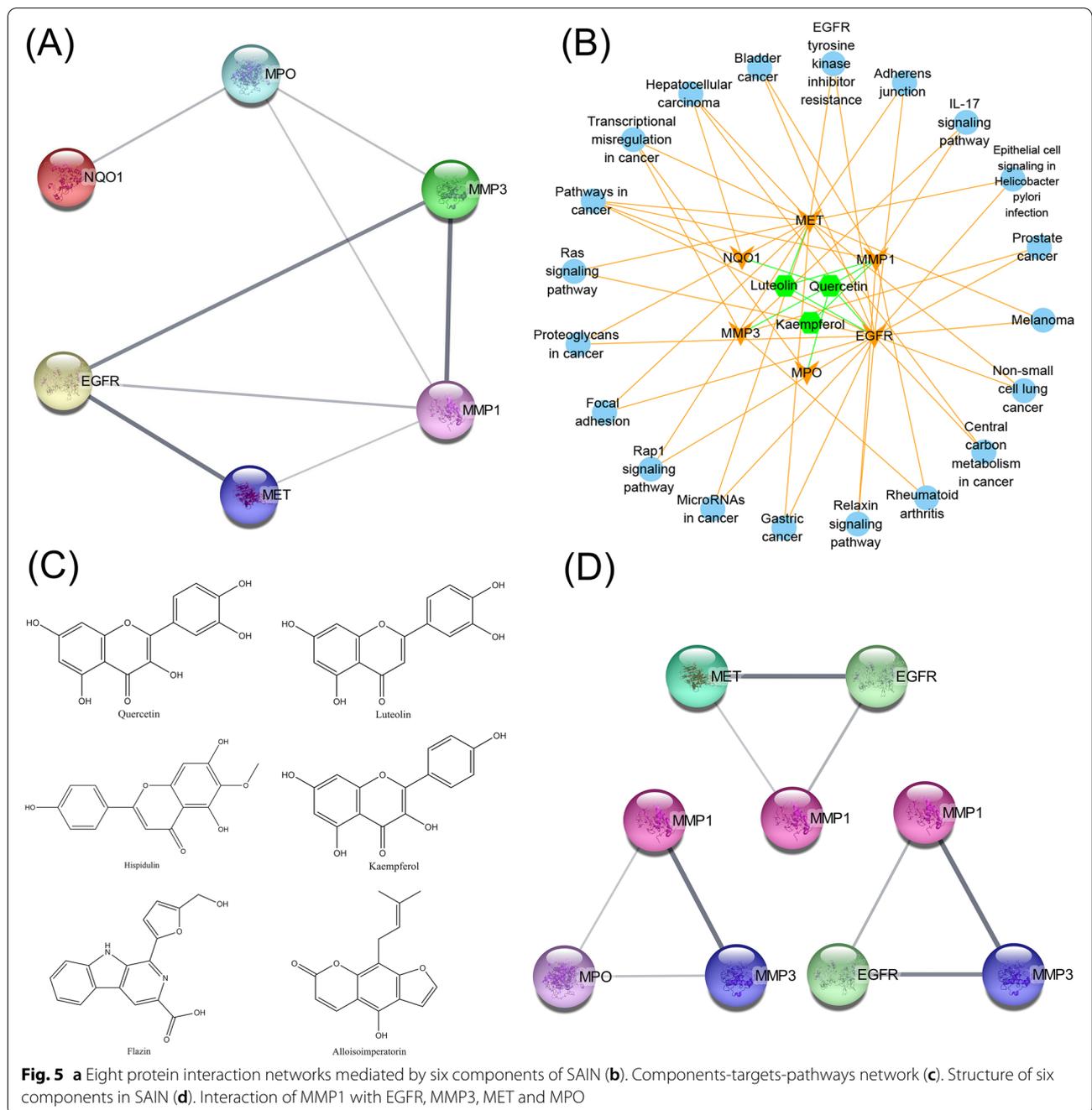


Fig. 4 BP analysis and KEGG analysis of 8 targets affected by SAIN, the size of the target is represented by the counts of participating targets (a). BP analysis of secondary screened targets (b). KEGG analysis of secondary screened targets



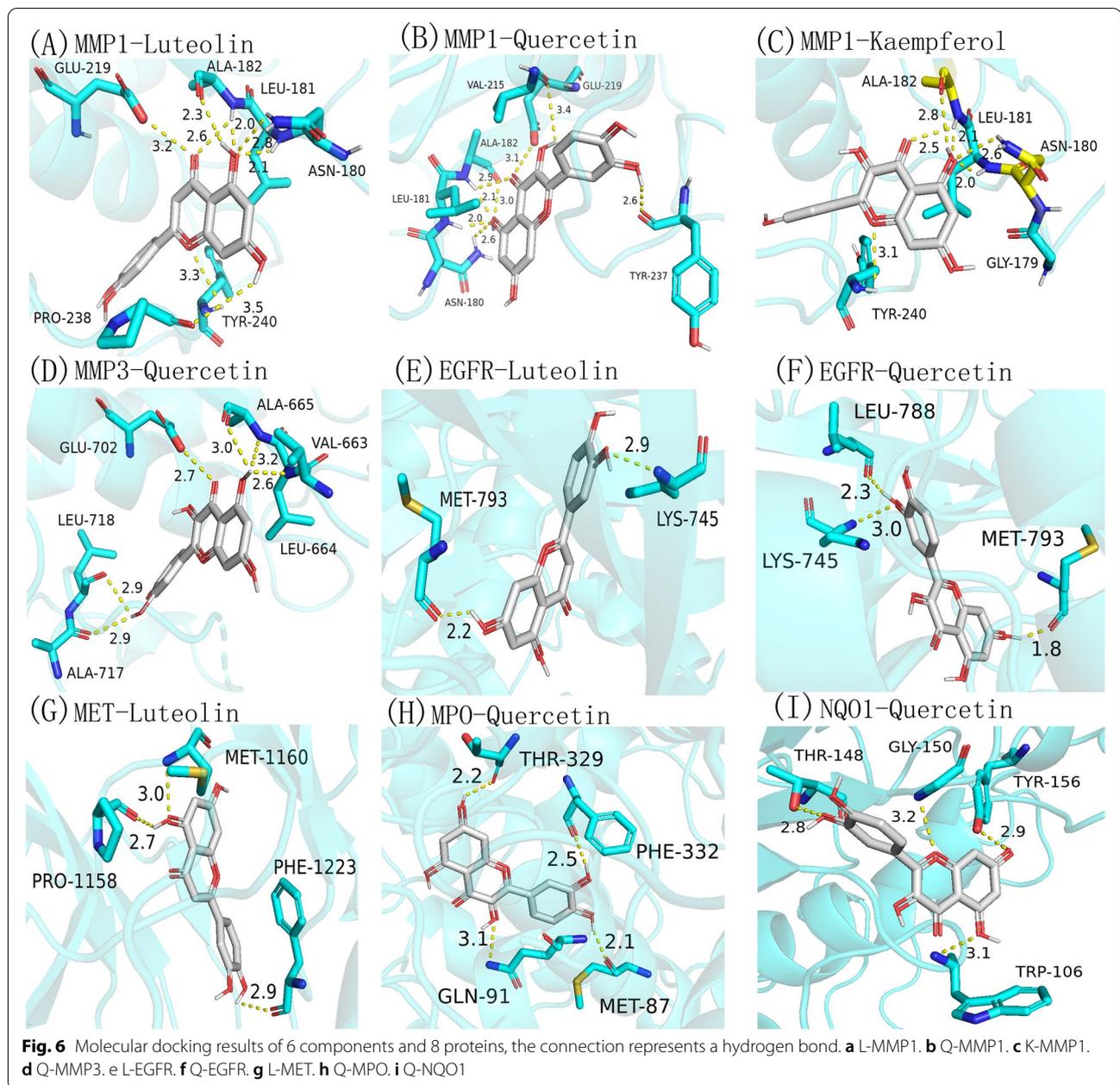
although these pathways contribute to both LUAD and LUSC, the strength of the contribution is still indistinguishable, and therefore, further analysis and validation (immunofluorescence, etc.) are needed.

Analysis of molecular docking results

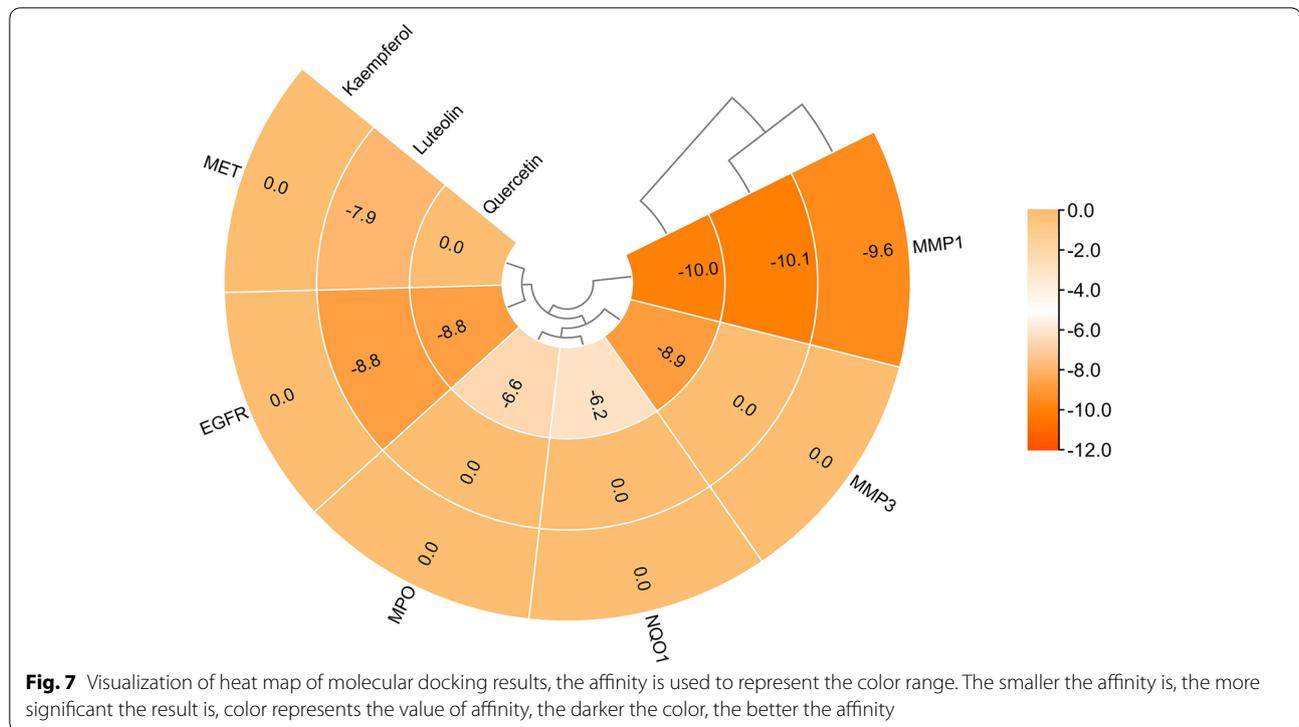
After a series of analysis, screening, and validation efforts as mentioned above, we identified the target sites where the three components (L, Q and K) act separately, where

the three components can act on multiple targets, where, Q can act on MMP1, MMP3, EGFR, MPO, and NQO1; L can act on MMP1, EGFR, and MET; and K only acts on MMP1.

After obtaining this information, the structures of the 6 components were plotted using ChemDraw ultra 12.0 software [34], saved as mol2 files, and the corresponding tertiary structures of proteins with ligands were selected according to the structures of the components using the

**Table 2** The specific molecular docking results of 3 compounds and 6 targets

Compounds	Targets	PDB ID	Resolution	Affinity (kcal/mol)	H-Bond (Dist ≤ 3.0)	UniProt ID	Method	dist from rmsd l.b	best mode rmsd u.b
Quercetin	MMP1	966C	1.90 Å	-10.0	6	P03956	X-RAY DIFFRACTION	0	0
	MMP3	1HY7	1.50 Å	-8.9	5	P08254	X-RAY DIFFRACTION	0	0
	EGFR	3W2S	1.90 Å	-8.8	3	P00533	X-RAY DIFFRACTION	0	0
	NQO1	5FUQ	2.04 Å	-6.2	2	P15559	X-RAY DIFFRACTION	0	0
	MPO	5WDJ	2.40 Å	-8.9	3	P05164	X-RAY DIFFRACTION	0	0
Luteolin	MMP1	966C	1.90 Å	-10.1	5	P03956	X-RAY DIFFRACTION	0	0
	EGFR	3W2S	1.90 Å	-8.8	2	P00533	X-RAY DIFFRACTION	0	0
	MET	4EEV	1.80 Å	-7.9	3	P08581	X-RAY DIFFRACTION	0	0
Kaempferol	MMP1	966C	1.90 Å	-9.6	5	P03956	X-RAY DIFFRACTION	0	0

**Table 3** The results of PSORT II

Compounds	Protein	PDB ID	Location (k=23)	Certainty (%)	Approach	UniProt ID
Luteolin	MMP1	966C	cytoplasmic	39.1%	k-NN	P03956
Quercetin			mitochondrial	26.1%		
Kaempferol			nuclear	21.7%		
Quercetin	MMP3	1HY7	mitochondrial	56.5%	k-NN	P08254
			cytoplasmic	30.4%		
Quercetin	NQO1	5FUQ	mitochondrial	47.8%	k-NN	P15559
			cytoplasmic	21.7%		
			nuclear	17.4%		
Quercetin	EGFR	3W2S	cytoplasmic	69.6%	k-NN	P00533
Luteolin			nuclear	21.7%		
Luteolin	MET	4EEV	cytoplasmic	43.5%	k-NN	P08581
			nuclear	30.4%		
			mitochondrial	26.1%		
Quercetin	MPO	5WDJ	nuclear	78.3%	k-NN	P05164
			cytoplasmic	13.0%		

PDB (<https://www.rcsb.org/>) database [35]. PDB files were downloaded, and the PDB protein files were processed by discovery studio 4.5 client software [36] (to remove water molecules and excess structures, etc.). Both file formats were saved as pdbqt files using autodock 1.5.6 software, ligands and coordinate positions (x, y, z) were selected for component screening and ranked by the distance of hydrogen bonds. Tertiary structures were predicted using an X-ray diffraction method, and the resolution of 6 proteins was the lowest at 1.90 Å and the

highest at 2.40 Å. Next, Autodock 1.5.6 and Autodock Vina were used for molecular docking of ingredients with targets, and finally, the docking results were processed using PyMol (Fig. 6).

The coordinates of each component in the tertiary structure are EGFR (12,12,14), MET (16,14,12), MMP1 (12,12,12), MMP3 (16,16,16), MPO (16,14,10), and NQO1 (22,22,22). The larger the value in this coordinate, the greater the affinity for docking, and the greater the affinity (interactions existing between ligand

Table 4 The results of CELLO

Compounds	Protein	PDB ID	Location	RELIABILITY	Approach	UniProt ID
Luteolin	MMP1	966C	Extracellular	2.122*	SVM-RFE	P03956
Quercetin			Cytoplasmic	2.074*		
Kaempferol						
Quercetin	MMP3	1HY7	Cytoplasmic	3.373*	SVM-RFE	P08254
Quercetin	NQO1	5FUQ	InnerMembrane	1.851*	SVM-RFE	P15559
			Cytoplasmic	1.311*		
			Periplasmic	1.209*		
Quercetin	EGFR	3W2S	Cytoplasmic	2.510*	SVM-RFE	P00533
Luteolin			Nuclear	1.183		
Luteolin	MET	4EEV	Cytoplasmic	3.156*	SVM-RFE	P08581
			InnerMembrane	1.166		
Quercetin	MPO	5WDJ	Nuclear	1.215*	SVM-RFE	P05164

Table 5 The results of BUSCA

Compounds	Protein	PDB ID	Location	Score	Approach	UniProt ID
Luteolin	MMP1	966C	cytoplasm	0.7	BetAware etc.	P03956
Quercetin						
Kaempferol						
Quercetin	MMP3	1HY7	mitochondrion	1	BetAware etc.	P08254
Quercetin	NQO1	5FUQ	mitochondrion	1	BetAware etc.	P15559
Quercetin	EGFR	3W2S	cytoplasm	1	BetAware etc.	P00533
Luteolin						
Luteolin	MET	4EEV	cytoplasm	1	BetAware etc.	P08581
Quercetin	MPO	5WDJ	cytoplasm	0.7	BetAware etc.	P05164

and receptor), thereby demonstrating that the tighter the binding of the component to the target, the more effective the component is. However, to guarantee the availability and reference value of the data, we set the coordinates (x , y , z) according to the structure size of the ingredients. Therefore, the coordinate values that were used in this study were appropriate for the ingredients. The best value was -10.1 kcal/mol and the worst was -7.9 kcal/mol.

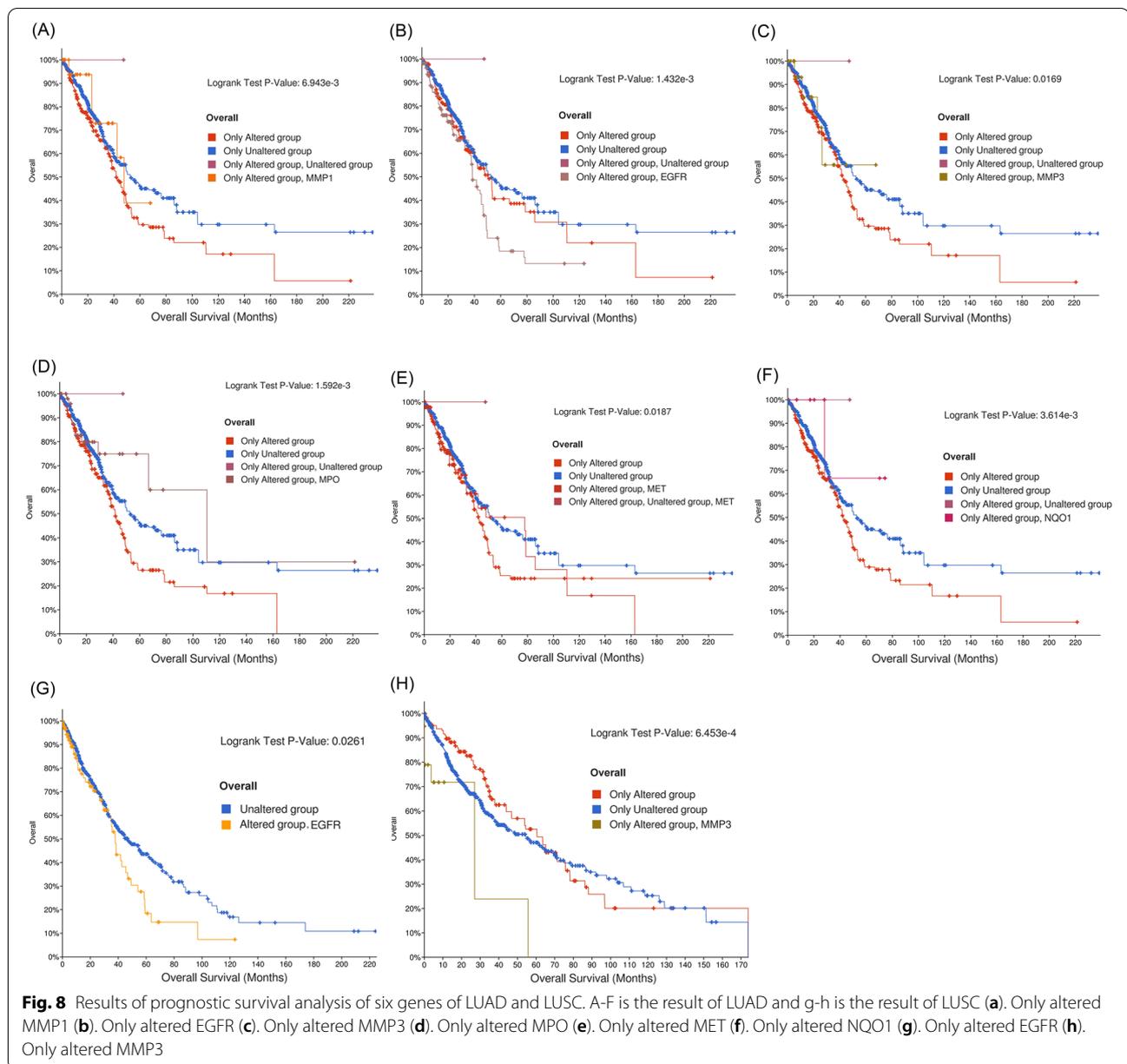
Based on the molecular docking results, items with affinities ranging from values (affinity ≤ -7.0 kcal/mol) were used, and the 3 SAIN components and 6 targets for which detailed results of molecular docking are given in Table 2 were selected. To visualize the results, the results are presented in detail in a Heatmap (Fig. 7).

Analysis of subcellular localization prediction results

To probe the specific location in the cell of the six proteins that were screened, the PSORT II, CELLO, and

BUSCA databases were used for subcellular localization predictions and predictions from three databases PSORT II (Tables 3, 4 and 5), CELLO (Tables 3, 4 and 5), and BUSCA (Tables 3, 44 and 5) were obtained. By comparing the results of the three subcellular localization prediction databases, we found that the PSORT II had a higher compatibility and reliability than the other two databases. Therefore, we chose to use the prediction results of PSORT II. The database uses the k-NN algorithm ($k = 23$), which is currently a relatively accurate algorithm for the prediction of subcellular localization.

Subcellular localization prediction showed that MMP1, MMP3, NQO1, EGFR, MET, and MPO were found in the cytoplasm, and among them, MMP1, MMP3, NQO1, and MET were also found in the mitochondria, in addition to MMP3. The other five proteins were all found in the nucleus.



Results of patient prognostic survival analysis of genes corresponding to the 6 targets

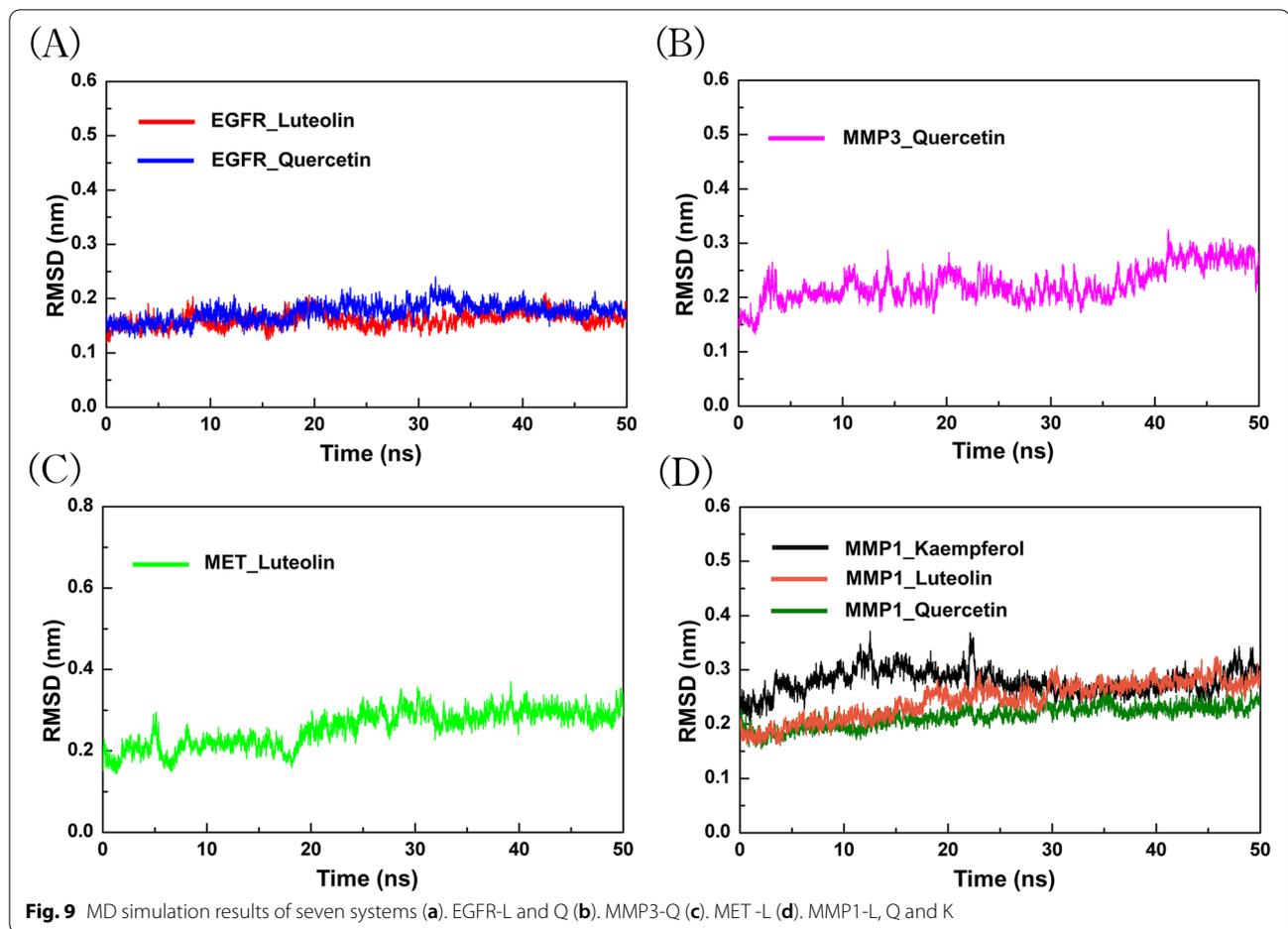
The purpose of survival analysis is to provide a patient with the option to continue treatment before undergoing surgery or other treatments, and is derived from the analysis of genetic data associated with the disease when the patient can survive after treatment with the corresponding target [37].

Survival analysis of these six targets showed that all six genes showed statistical significance in the regulation of LUAD ($P < 0.05$), while only EGFR and MMP3 were statistically significant in the regulation of LUSC ($P < 0.05$), which was consistent with our previous TPM analysis (Fig. 8).

Analysis of MD simulation results

RMSD data are an important basis for measuring the stability of the system and can be used to measure the stability of the tertiary structure of a protein after incorporating small molecules [38]. To make the RMSD data behave more intuitively, data were visualized (Fig. 9).

EGFR, NSCLC targets have been used in clinic, and the RMSD change trend and starting value in the time period of 0–50 ns are similar to those of the MMP3 protein. In addition, EGFR and MMP3 exhibit a very stable conformation (average RMSD < 0.15) after forming a complex under the action of L or Q. Similarly, for clinical use with the target MET, its complex formed under the action



of L was relatively stable, but the MMP1 protein had a stronger structural stability than MET when in complex with L or Q. When K is in a complex with MMP1, the stability of the complex in the human setting is general, but it does not show major unusual fluctuations (Fig. 9).

Therefore, L, Q, and K, the active ingredients of SAIN, may play an important role in improving or treating NSCLC (LUAD and LUSC) after acting on the targets EGFR, MET, MMP1, and MMP3.

Discussion

Current status of therapeutic targets and MMP3 for LUSC

Compared with LUAD, LUSC has a unique pathological morphology (cancer cell growth location, direction, and variation) in NSCLC, therefore, no targeted therapy has been developed for LUSC due to few oncogenic aberrations that can target LUSC.

Fibroblast growth factors (FGFRs) belong to the family of receptor tyrosine kinases (RTKs) along with EGFR, which comprises the four members FGFR 1–4. In our study, we observed amplification or mutations of

FGFR1–4 in NSCLC, which play a crucial role in tumor development and maintenance. In 2016, a gene sequencing study that included 4853 cancer patients revealed that 7.1% of cancer patients had abnormalities in both FGFR 1–4 genes. The most commonly found amplification of the FGFR-1 gene accounted for 50% [39]. A large body of clinical evidence has shown that FGFR-1 inhibitors play a key role in targeting the symptoms caused by LUSC, especially Infigratinib (BGJ398), BGJ398 for the treatment of LUSC, with a disease control rate of 47.6% [40]. Therefore, FGFR-1 may be the next effective target for LUSC selective therapy.

Matrix metalloproteinase 3 (MMP3) is the third member of the matrix metalloproteinase (MMP) family, and studies have shown that it plays many roles in healthy humans and patients, such as promoting epithelial mesenchymal transition, increasing the levels or activity of pulmonary profibrotic mediators, or decreasing antifibrotic mediators, and promoting abnormal epithelial cell migration and other abnormal repair processes [41]. In addition, adipocytes can increase the invasive ability of tumor cells to the body and increase the metastatic risk

of lung tumors by producing exosomes with high levels of MMP3 [42, 43].

It is well known that insulin enhances the proliferation, migration, and drug resistance of NSCLC cells by activating the PI3K / Akt pathway. In previous studies, it has been shown that insulin can upregulate gene expression of MMP3 in NSCLC cells, and that its expression can be inhibited by PI3K / Akt pathway inhibitors [44].

In previous studies, it has been shown that the genes and proteins of FGFR-1 and MMP3 are highly expressed in esophageal squamous cell carcinoma, and may be associated with esophageal cancer invasion and metastasis [45].

Our findings showed that MMP3 positively regulated NSCLC, especially LUSC by participating in transcriptional misregulation in cancer, the IL-17 signaling pathway, and the rheumatoid arthritis pathway. Furthermore, molecular docking results showed that Q could bind to MMP3 protein with high binding affinity of -8.9 kcal/mol. Therefore, we speculated that there may be consistency or homology between MMP3 and FGFR-1 in treating LUSC. MMP3 may serve as a biomarker for the diagnosis and prognosis of LUSC [46]. However, more data and a large number of clinical cases are needed to verify this conclusion.

EGFR and MET in LUAD targeted therapy

The latest NSCLC guideline 2021 V2 edition issued by the national comprehensive cancer network (NCCN) indicated that genes associated with targeted therapy of NSCLC include EGFR, KRAS, ALK, ROS1, MET, BRAF, RET, and NTRK [47]. However, these genes were mainly targeted for LUAD.

Receptor tyrosine kinases (RTKs) are high affinity cell surface receptors for many hormones, cytokines, and polypeptide growth factors. Of the 90 unique tyrosine kinase genes identified in the human genome, 58 encode RTK proteins. RTKs have not only been shown to be key regulators of normal cellular processes but also play a key role in the initiation and progression of many types of cancer [47].

EGFR and MET belong to members of the RTK family, and MET, the tyrosine kinase transmembrane receptor for MET, is a protein product encoded by the c-MET proto oncogene that has been of interest in the clinic as a potential therapeutic target in multiple tumor types [48].

Currently, a variety of targeted drugs have been developed for NSCLC targeting EGFR mutations and MET exon14 alterations, among which, drugs targeting EGFR mutations are mainly classified into two major classes: small molecule tyrosine kinase inhibitors (TKI) and monoclonal antibodies, and the TKI class mainly includes *Gefitinib*, *Erlotinib*, *Afatinib*, *Osimertinib*, *Dacomitinib*,

Lcotinib, and *Brigatinib*, among others, monoclonal antibodies are comparatively rare and mainly include *Necitumumab*, *Amivantamab*, and *Ramucirumab*, of which *Amivantamab* targets the EGFR exon20 insertion [49].

Drugs targeting alterations in MET exon14 are two major classes of TKIs and monoclonal antibodies, multi kinase inhibitors, including TKIs (*Crizotinib*, *Cabozantinib*, *MGCD265*, *AMG208*, *Altiratinib*, *Golvatinib*) and selective MET inhibitors (*Capmatinib*, *Tepotinib*, *Tivantinib*). Monoclonal antibodies can be further divided into Anti-MET antibodies (*Onartuzumab*, *Emibetuzumab*) and anti-HGF antibodies (*Ficlatuzumab*, *Rilotumumab*).

The development of these drugs supported the results of our analysis, that is, MET gene expression in LUAD patients was much higher than that in LUSC patients, EGFR gene expression was roughly the same in the two groups of patients, and occasionally LUSC patients showed a higher expression. The results of molecular docking studies showed that SAIN's component I to Q could stably bind to the EGFR protein with an affinity of both -8.8 kcal/mol, while the MET protein could only bind to component I with an affinity of -7.9 kcal/mol. Survival analysis showed that among LUAD patients, EGFR mutated patients ($P < 0.001$) had significantly shorter prognostic survival relative to patients with MET mutations ($P < 0.02$).

Role of MPO in NSCLC patients

Smoking confers an increased risk of lung cancer, and in an earlier study, no significant difference by NQO1 genotype was found in a large cohort of patients with combined smoking and nonsmoking lung adenocarcinoma. The data showed that smokers with combined MPO genotype had a lower lung cancer prevalence than nonsmokers [50].

MPO is a member of the heme peroxidase superfamily and is mainly expressed in neutrophils and monocytes [51]. It has been shown that MPO targeted therapy can exhibit a favorable prognostic survival status in LUAD patients [52], which is consistent with the results of our survival analysis ($P < 0.001$). Molecular docking studies showed that component Q of SAIN could specifically bind the MPO protein (affinity = -8.8 kcal/mol). Despite the evidence that MPO is functional in NSCLC patients, both gene differential analysis and TPM analysis results were different from that of other targets. MPO expression in tumor tissues of LUAD and LUSC was much lower compared to its expression in healthy tissues ($TPM < 2$), and MPO was not expressed in NSCLC patients. Based on the results from KEGG enrichment analysis, we speculated that MPO may play a role in LUAD and LUSC patients through transcriptional misregulation in cancer pathways.

High expression of MMP1 has potential in NSCLC treatment.

Molecular docking of C-T revealed that components Q, L, and K could bind to MMP1 protein with better affinity than any other component, Q (affinity = -10.0 kcal/mol), L (affinity = -10.1 kcal/mol), and K (affinity = -9.6 kcal/mol), respectively. Our survival analysis of LUAD and LUSC patients after comparing MMP1 overexpression revealed that the overexpression of MMP1 exhibited a worse prognostic survival compared with the survival time of patients after an EGFR mutation. Moreover, PPI network analysis showed that MMP1 formed three triangular structures with EGFR, MMP3, MET, and MPO (Fig. 5d), suggesting a possible link with four additional targets when MMP1 functions.

MMP1, like MMP3, belongs to the MMP family. Only few studies have been performed on MMP1 because the expression of MMP1 in rodent orthologs is not conserved [53]. However, there are several ways to study the function of MMP1 in mice and cell systems. Results from previous studies have suggested that (1) MMP1 could regulate the mechanisms of LUAD cell proliferation, migration, and invasion under the regulation of mir-202-3p [54], (2) p53 dysfunction caused by XPC gene (xeroderma pigmentosum) deficiency in lung cancer may enhance tumor metastasis by increasing MMP1 expression [55], (3) macrophage-specific inhibition of MMP1 secretion may be a potential therapy to reduce lung metastasis in smoking cancer patients [56].

Therefore, combined with the above-mentioned results, we speculate that MMP1 may be a novel target for NSCLC treatment, and I, Q, and K may be the relevant lead compounds.

Validation of the validity of Q, L, and K

In this study, we screened the three active ingredients Q, L, and K from SAIN. However, as to whether these three ingredients have curative effects on NSCLC, experiments in cell systems or animal experiments are needed for further verification. Before performing in vitro and in vivo validation, extensive experimental validation for the treatment of NSCLC on several of these components was found in the literature, therefore, we did not perform in vitro or in vivo validation.

L (2-(3,4-dihydroxyphenyl)-5,7-dihydroxy-4-chromenone) is a natural flavonoid compound, and it has been reported that the effects of L on NSCLC are mainly reflected in (1) in lung cancer A549 cells and nuclear H460 xenograft mice. The expression of L in melanoma 2 (AIM2) was significantly reduced at the mRNA and protein expression levels, which inhibited AIM2 inflammasome activation, and in turn induced G2/M phase cell cycle arrest and inhibited the epithelial

mesenchymal transition (EMT) process in NSCLC [57]. (2) At both the cellular and animal levels, L had a significant antitumor effect against NSCLC with L858R/T790M mutations in EGFR and erlotinib resistance. In addition, the mechanism is that L induces the degradation of EGFR by inhibiting the binding of Hsp90 to the mutated EGFR protein, thereby further preventing PI3K/Akt/mTOR signaling, which leads to NSCLC cell apoptosis [58, 59]. (3) At physiological concentrations, L significantly sensitizes A549 cells to the anticancer drugs oxaliplatin, bleomycin, and doxorubicin and can be used in chemotherapy as a natural sensitizer to achieve better drug availability [60, 61]. These experimental validation results are largely consistent with our findings, prediction, and MD validation results, and indicate that L has great potential for being used in the treatment of NSCLC.

Q (2-(3,4-dihydroxyphenyl)-3,5,7-trihydroxy-4-chromenone) and L are structurally similar, the only difference being that Q has more than one hydroxyl group in position 3 of the chromenone parent nucleus. Epidemiologically, studies have shown that Q has the effect of preventing lung cancer, which is mainly reflected by (1) Q can significantly enhance tumor necrosis factor-related apoptosis inducing ligand (TRAIL)-induced cytotoxicity in NSCLC cells, thereby accelerating the death of NSCLC cells [62, 63]. (2) Priming H520 cells (LUSC cells) with Q increased cisplatin-induced apoptosis by 30.2%, and this process was accompanied by Q-induced expression of multiple apoptosis-related genes, which could serve as a potent chemosensitizer [64]. (3) Q can increase or decrease the expression of many miRNAs or TKIs, including increasing the expression of mir-16-5p, a member of the mir-16 family, decreasing the expression of WEE1, and inhibiting the expression of the SRC family [65]. Similarly, our prediction was that Q might treat or alleviate disease symptoms by decreasing the expression of MMP3, EGFR, and MMP1 in NSCLC tissues.

In previous studies, it was shown that dietary flavone K (3,5,7-trihydroxy-2-(4-hydroxyphenyl)-4-chromenone) can effectively prevent and treat lung cancer, mainly through the following mechanisms: (1) K inhibits the apoptosis of H460 cells (NSCLC cells) by inducing the overexpression of tumor suppressor gene antioxidant enzyme (Sod-2) [66]. (2) K mediates Smad3 phosphorylation at threonine 179 (THR-179) by inhibiting AKT1, which in turn inhibits transforming growth factor- β 1 (TG- β 1)-induced epithelial to mesenchymal transition and migration of A549 lung cancer cells [67]. (3) K sensitizes otherwise chemoresistant cancer cells to multitarget antifolate (MTA) by inhibiting the EMT signaling pathway, which enables K to reverse the ability of NSCLC to resist MTA, thereby realizing the role of inhibiting lung cancer chemotherapy responses through the EMT

pathway [68]. Our results that show that K exerts its anti NSCLC effects by inhibiting the MMP1 mediated IL-17 signaling pathway, rheumatoid arthritis pathway, and relaxin signaling pathway are consistent with our docking results and MD results.

Conclusion

It is well known that the complexity of NSCLC etiology does not allow us to explain it by the alteration of a single gene, therefore. The treatment of lung cancer is not achieved by one means, but requires combination therapy of multiple means and multiple targets. Although the three components Q, L, and K in SAIN can do so by binding to MMP1, MMP3, EGFR, and MET, and are capable of acting as inhibitors by participating in EGFR tyrosine kinase inhibitor resistance, adherens junction, IL-17 signaling pathway, NSCLC, and the relaxin signaling pathway, exerting their functions together to prevent and treat NSCLC (LUAD and LUSC). Nobody can predict which genetic mutations are responsible for NSCLC, and therefore, a concerted effort from researchers in all fields is required for the complete cure of NSCLC.

Abbreviations

SAIN: Saussurea involucrate; NSCLC: Non-small cell lung cancer; LUAD: Lung adenocarcinoma; LUSC: Lung squamous cell carcinoma; Sl: Snow lotus; Q: Quercetin; L: Luteolin; K: Kaempferol; H: Hispidulin; A: Alloisoperatorin; F: Flazin; TCMSP: Traditional Chinese Medicine Systems Pharmacology Database; OB: Oral bioavailability; DL: Drug-like; GDC: Genomic Data Commons; GEO: Gene Expression Omnibus; GEPIA: Gene Expression Profiling Interactive Analysis; TCGA: The Cancer Genome Atlas; FDR: False discovery rate; GO: Gene Ontology; BP: Biology Progress; KEGG: Kyoto Encyclopedia of Genes and Genomes; PPI: Protein-Protein Interaction; Dist: Distance; IL-17: Interleukin-17; MMP1: Matrix metalloproteinase 1; MMP3: Matrix metalloproteinase 3; MET: Mesenchymal to epithelial transition factor; EGFR: Epidermal growth factor receptor; MPO: Myeloperoxidase; NQO1: Nad(p)h: quinone oxidoreductase 1; PTGS2: Prostaglandin G/H synthase 2; HSP90AA1: Heat Shock Protein 90 Alpha Family Class A Member 1; TPM: Transcripts Per Million; ALK: Anaplastic Lymphoma Kinase; ROS1: C-ros oncogene 1, receptor tyrosine kinase; C-T-P: Compounds-Targets-Pathways; cNLS: Classic nuclear localization signal; k-NN: K-Nearest Neighbor; SVM-RFE: Support vector machine recursive feature elimination algorithm; STP: Swisstargetprediction; ANOVA: Analysis of Variance; PDB: Palm DataBase; FGFR: Fibroblast growth factor; RTK: Receptor tyrosine kinase; KRAS: V-Ki-ras2 Kirsten ratsarcoma viral oncogene homolog; BRAF: V-raf murine sarcoma viral oncogene homolog B1; NTRK: Neurotrophic factor receptor tyrosine kinase; HGF: Hepatocyte growth factor; TKI: Tyrosine kinase inhibitors; AIM2: Melanoma 2; EMT: Epithelial-mesenchymal transition; XPC: Xeroderma pigmentosum; MTA: Multi-target antifolate; TMBB: Transmembrane beta-barrels.

Acknowledgements

Not applicable

Authors' contributions

Dongdong Zhang and Tieying Zhang contributed equally to this research and Dongdong Zhang and Tieying Zhang wrote the main manuscript text. Yao Zhang, Zhongqing Li, He Li, Yueyang Zhang, Chenggong Liu and Zichao Han prepared all figures. Jin Li and Jianbo Zhu are corresponding authors. All authors reviewed the manuscript. The author(s) read and approved the final manuscript.

Funding

Not applicable.

Availability of data and materials

The datasets generated and/or analysed during the current study are not publicly available due the limited scope of data availability, these data are used under the license of this study and are not disclosed, but are available from the author[sedate@stu.shzu.edu.cn (Dongdong Zhang)] on reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

Not applicable.

Received: 18 September 2021 Accepted: 30 December 2021

Published online: 28 February 2022

References

- Siegel RL, Miller KD, Fuchs H, Jemal A. Cancer Statistics. *CA Cancer J Clin*. 2021;71:7–33.
- Howlader N, Forjaz G, Mooradian MJ, Meza R, Feuer EJ. The effect of advances in lung-cancer treatment on population mortality. *N Engl J Med*. 2020;383:640–9.
- Li Y, Appius A, Pattipaka T, Feyereislova A, Cassidy A, Ganti AK. Real-world management of patients with epidermal growth factor receptor (EGFR) mutation-positive non-small-cell lung cancer in the USA. *PLoS One*. 2019;14:e0209709.
- Li G-h, Liu F, Zhao R-c. Studies on pharmacological actions of Saussurea involucreta Kar et Kir ex Maxim (author's transl). *Acta pharmaceutica Sinica*. 1980;15(6):368.
- Wang X-h, Chu L, Liu C, Wei R-l, Xue X-l, Xu Y-f, Wu M-j, Miao Q. Therapeutic Effects of Saussurea Involucreta Injection against Severe Acute Pancreatitis- Induced Brain Injury in Rats. *Biomed Pharmacother*. 2018;100:564–74.
- Gong G-w, Xie F, Zheng Y-z, Hu W-h, Qi B-h, He H, Dong TT, Tsim KW. The effect of methanol extract from Saussurea involucreta in the lipopolysaccharide-stimulated inflammation in cultured RAW 264.7 cells. *J Ethnopharmacol*. 2020;251:112532–112532.
- Hopkins AL. Network pharmacology. *Nat Biotechnol*. 2007;25(10):1110–1.
- Ru J, Li P, Wang J, Zhou W, Li B, Huang C, Li P, Guo Z, Tao W, Yang Y, Xu X, Li Y, Wang Y, Yang L. TCMSP: a database of systems pharmacology for drug discovery from herbal medicines. *J Cheminform*. 2014;16(6):13.
- Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Sajed T, Johnson D, Li C, Sayeeda Z, Assempour N, Iynkkaran I, Liu Y, Maciejewski A, Gale N, Wilson A, Chin L, Cummings R, Le D, Pon A, Knox C, Wilson M. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res*. 2018;46(D1):D1074–82.
- Daina A, Michielin O, Zoete V. SwissTargetPrediction: updated data and new features for efficient prediction of protein targets of small molecules. *Nucleic Acids Res*. 2019;47(W1):W357–64.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498–504.
- Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, Soboleva A. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res*. 2013;41(Database issue):D991–5.
- Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res*. 2017;45(W1):W98–102.

14. Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* 2003;4(5):P3 Epub 2003 Apr 3.
15. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000;28(1):27–30.
16. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multi-threading. *J Comput Chem.* 2010;31(2):455–61.
17. Seeliger D, de Groot BL. Ligand docking and binding site analysis with PyMOL and Autodock/Vina. *J Comput Aided Mol Des.* 2010;24(5):417–22.
18. Pinzi L, Rastelli G. Molecular Docking: Shifting Paradigms in Drug Discovery. *Int J Mol Sci.* 2019;20(18):4331.
19. Kaur T, Madgulkar A, Bhalekar M, Asgaonkar K. Molecular Docking in Formulation and Development. *Curr Drug Discov Technol.* 2019;16(1):30–9.
20. Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K. WoLF PSORT: protein localization predictor. *Nucleic Acids Res.* 2007;35(Web Server issue):W585–7.
21. Ying H, Li Y. Prediction of protein subcellular locations using fuzzy k-NN method. *Bioinformatics.* 2004;20(1):21–8.
22. Bernstien MN, Ma Z, Gleicher M, Dewey CN. Cello: comprehensive and hierarchical cell type classification of human cells with the Cell Ontology. *iScience.* 2020;24(1):101913.
23. Yu C-s, Lin C-j, Hwang J-k: Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. *Protein Sci.* 2004;13:1402–6.
24. Savojardo C, Martelli PL, Fariselli P, Profti G, Casadio R. BUSCA: an integrative web server to predict subcellular localization of proteins. *Nucleic Acids Res.* 2018;46(W1):W459–66.
25. Castrense S, Piero F, Rita C. BETAWARE: a machine-learning tool to detect and predict transmembrane beta-barrel proteins in prokaryotes. *Bioinformatics.* 2013;4:504–5.
26. Collier TA, Piggot TJ, Allison JR. Molecular Dynamics Simulation of Proteins. *Methods Mol Biol.* 2020;2073:311–27.
27. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS: fast, flexible, and free. *J Comput Chem.* 2005;26(16):1701–18.
28. Wang Z, Jensen MA, Zenklusen JC. A Practical Guide to The Cancer Genome Atlas (TCGA). *Methods Mol Biol.* 2016;1418:111–41.
29. Chen C, Rui X, Hao C, He Y. TBtools, a Toolkit for Biologists integrating various HTS-data handling tools with a user-friendly interface. 2018.
30. Armstrong RA, Eperjesi F, Gilmartin B. The application of analysis of variance (ANOVA) to different experimental designs in optometry. *Ophthalmic Physiol Opt.* 2002;22(3):248–56.
31. Wang Y, Shen SY, Liu L, Zhang XD, Liu DY, Liu N, Liu BH, Shen L. Jolkinolide B inhibits proliferation or migration and promotes apoptosis of MCF-7 or BT-474 breast cancer cells by downregulating the PI3K-Akt pathway. *Journal of Ethnopharmacology.* 2021;282:114581.
32. Kunzmann AT, Murray LJ, Cardwell CR, McShane CM, McMenamin UC, Cantwell MM. PTGS2 (Cyclooxygenase-2) expression and survival among colorectal cancer patients: a systematic review. *Cancer Epidemiol Biomarkers Prev.* 2013;22(9):1490–7.
33. Andries L, Masin L, Navarro MS, Zaunz S. MMP2 Modulates Inflammatory Response during Axonal Regeneration in the Murine Visual System. *Cells.* 2021;10(7):1672.
34. Cousins KR. Computer review of ChemDraw Ultra 12.0. *J Am Chem Soc.* 2011;133(21):8388.
35. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. *Nucleic Acids Res.* 2000;28(1):235–42.
36. Wang S, Jiang JH, Li RY, Deng P. Docking-based virtual screening of TBR1 inhibitors: evaluation of pose prediction and scoring functions. *BMC Chem.* 2020;14(1):52.
37. Jung SH, Lee HY, Chow SC. Statistical Methods for Conditional Survival Analysis. *J Biopharm Stat.* 2018;28(5):927–38.
38. Coutsiadis EA, Wester MJ. RMSD and Symmetry. *J Comput Chem.* 2019;40(15):1496–508. <https://doi.org/10.1002/jcc.25802> Epub 2019 Mar 3 PMID: 30828834.
39. Desai A, Adjei AA. FGFR Signaling as a Target for Lung Cancer Therapy. *J Thorac Oncol.* 2016;11(1):9–20.
40. Dong M, Li T, Chen J. Progress on the Study of Targeting FGFR in Squamous Non-small Cell Lung Cancer. *Zhongguo Fei Ai Za Zhi.* 2018;21(2):116–20.
41. Craig VJ, Zhang L, Hagood JS, Owen CA. Matrix metalloproteinases as therapeutic targets for idiopathic pulmonary fibrosis. *Am J Respir Cell Mol Biol.* 2015;53(5):585–600.
42. Wang J, Wu Y, Guo J, Fei X, Yu L, Ma S. Adipocyte-derived exosomes promote lung cancer metastasis by increasing MMP9 activity via transferring MMP3 to lung cancer cells. *Oncotarget.* 2017;8(47):81880–91.
43. Banik D, Netherby CS, Bogner PN, Abrams SI. MMP3-mediated tumor progression is controlled transcriptionally by a novel IRF8-MMP3 interaction. *Oncotarget.* 2015;6(17):15164–79.
44. Jiang J, Ren H-y, Geng G-j, Mi Y-j, Liu Y, Li N, Yang S-y, Shen D-y. Oncogenic activity of insulin in the development of non-small cell lung carcinoma. *Oncol Lett.* 2018;15(1):447–52.
45. Yong T. Expression and clinical significance of FGFR1 and MMP3 in esophageal squamous cell carcinoma. *J Mod Oncol.* 2017;05:729–33.
46. Mehner C, Miller E, Nassar A, Bamlet WR, Radisky ES, Radisky DC. Tumor cell expression of MMP3 as a prognostic factor for poor survival in pancreatic, pulmonary, and mammary carcinoma. *Genes Cancer.* 2015;6(11–12):480–9.
47. Ettinger DS, Wood DE, Aggarwal C, Aisner DL, Akerley W, Bauman JR, Bharat A, Bruno DS, Chang JY, Chirieac LR, D'Amico TA, Dilling TJ, Dobelbower M, Gettinger S, Govindan R, Gubens MA, Hennon M, Horn L, Lackner RP, anuti LM, Leal TA, Lin J, Loo BW Jr, Martins RG, Otterson GA, Patel SP, Reckamp KL, Riely GJ, Schild SE, Shapiro TA, Stevenson J, Swanson SJ, Tauer KW, Yang SC, OCN KG, Hughes M. NCCN Guidelines Insights: Non-Small Cell Lung Cancer, Version 2.2021. *J Natl Compr Canc Netw.* 2021;19(3):254–66.
48. Mo HN, Liu P. Targeting MET in cancer therapy. *Chronic Dis Transl Med.* 2017;3(3):148–53.
49. Cardona AF, Rojas L, Zatarain-Barrón ZL, Freitas HC, Granados ST, Castillo O, Oblitas G, Corrales L, Castro CD, Ruiz-Patiño A, Martín C, Pérez MA, González L, Chirinos L, Vargas C, Carranza H, Otero J, Rodríguez J, Rodríguez J, Archila P, Lema M, Acosta Madiedo J, Karachaliu N, Wills B, Pino LE, de Lima V, Rosell R, Arrieta O, CLICaP. EGFR exon 20 insertion in lung adenocarcinomas among Hispanics (geno1.2-CLICaP). *Lung Cancer.* 2018;125:265–72.
50. Kiyohara C, Yoshimasu K, Takayama K, Nakanishi Y. NQO1, MPO, and the risk of lung cancer: a HuGE review. *Genet Med.* 2005;7(7):463–78.
51. Jennette JC, Nachman PH. ANCA Glomerulonephritis and Vasculitis. *Clin J Am Soc Nephrol.* 2017;12(10):1680–91.
52. Ma X-p, Huang X-m, Moore Z, Huang G, Kilgore JA, Wang Y-g, Hammer S, Williams NS, Boothman DA, Gao J-m. Esterase-activatable β -lapachone prodrug micelles for NQO1-targeted lung cancer therapy. *J Control Release.* 2015;200:201–11.
53. Carver PI, Anguiano V, D'Armiento JM, Shiomi T. Mmp1a and Mmp1b are not functional orthologs to human MMP1 in cigarette smoke induced lung disease. *Exp Toxicol Pathol.* 2015;67(2):153–9.
54. Li Y, Huang H-q, Ye X-l, Huang Z-h, Chen X-q, Wu F, Lin T-y. miR-202–3p negatively regulates MMP-1 to inhibit the proliferation, migration and invasion of lung adenocarcinoma cells. *Cell Cycle.* 2021;20(4):406–16.
55. Wu Y-h, Wu T-c, Liao J-w, Yeh K-t, Chen C-y, Lee H. p53 dysfunction by xeroderma pigmentosum group C defects enhance lung adenocarcinoma metastasis via increased MMP1 expression. *Cancer Res.* 2010;70(24):10422–32.
56. Morishita A, Gerber A, Gow CH, Zelonina T, Chada K, D'Armiento J. Cell Specific Matrix Metalloproteinase-1 Regulates Lung Metastasis Synergistically with Smoke Exposure. *J Cancer Res Forecast.* 2018;1(2):1014.
57. Yu Q, Zhang M-d, Ying Q-d, Xie X, Yue S-w, Tong B-d, Wei Q, Bai Z-s, Ma L-m. Decrease of AIM2 mediated by luteolin contributes to non-small cell lung cancer treatment. *Cell Death Dis.* 2019;10(3):218.
58. Hong Z, Cao X, Li N, Zhang Y-z, Lan L, Zhou Y, Pan X-i, Shen L, Yin Z-m, Luo L. Luteolin is effective in the non-small cell lung cancer model with L858R/T790M EGF receptor mutation and erlotinib resistance. *Br J Pharmacol.* 2014;171(11):2842–53.
59. Cai X-t, Ye T-m, Liu C, Lu W-g, Lu M, Zhang J, Wang M, Cao P. Luteolin induced G2 phase cell cycle arrest and apoptosis on non-small cell lung cancer cells. *Toxicol In Vitro.* 2011;25(7):1385–91.
60. Tang X-w, Wang H-y, Fan L-f, Wu X-y, Xin A, Ren H-y, Wang X-j. Luteolin inhibits Nrf2 leading to negative regulation of the Nrf2/ARE pathway and sensitization of human lung carcinoma A549 cells to therapeutic drugs. *Free Radic Biol Med.* 2011;50(11):1599–609.

61. Cho HJ, Ahn KC, Choi JY, Hwang SG, Kim WJ, Um HD, Park JK. Luteolin acts as a radiosensitizer in non-small cell lung cancer cells by enhancing apoptotic cell death through activation of a p38/ROS/caspase cascade. *Int J Oncol.* 2015;46(3):1149–58.
62. Che W-s, Wang X, Zhuang J-g, Zhang L, Lin Y. Induction of death receptor 5 and suppression of survivin contribute to sensitization of TRAIL-induced cytotoxicity by quercetin in non-small cell lung cancer cells. *Carcinogenesis.* 2007;28(10):2114–21.
63. Kuhar M, Sen S, Singh N. Role of mitochondria in quercetin-enhanced chemotherapeutic response in human non-small cell lung carcinoma H-520 cells. *Anticancer Res.* 2006;26(2A):1297–303.
64. Wang Q, Chen Y-kun, Lu H-j, Wang H-j, Feng H, Xu J-p, Zhang B-y. Quercetin radiosensitizes non-small cell lung cancer cells through the regulation of miR-16–5p/WEE1 axis. *IJBMB Life.* 2020;72(5):1012–22.
65. Dong Y, Yang J, Yang L-y, Li P. Quercetin Inhibits the Proliferation and Metastasis of Human Non-Small Cell Lung Cancer Cell Line: The Key Role of Src-Mediated Fibroblast Growth Factor-Inducible 14 (Fn14)/ Nuclear Factor kappa B (NF- κ B) pathway. *Med Sci Monit.* 2020;26:e920537.
66. Leung H-w, Lin C-j, Hour M-j, Yang W-h, Wang M-y, Lee H-z. Kaempferol induces apoptosis in human lung non-small carcinoma cells accompanied by an induction of antioxidant enzymes. *Food Chem Toxicol.* 2007;45(10):2005–13.
67. Jo E, Park SJ, Choi YS, Jeon WK, Kim BC. Kaempferol Suppresses Transforming Growth Factor- β 1-Induced Epithelial-to-Mesenchymal Transition and Migration of A549 Lung Cancer Cells by Inhibiting Akt1-Mediated Phosphorylation of Smad3 at Threonine-179. *Neoplasia.* 2015;17(7):525–37.
68. Liang S-q, Marti TM, Dorn P, Froment L, Hall SRR, Berezowska S, Kocher G, Schmid RA, Peng R-w. Blocking the epithelial-to-mesenchymal transition pathway abrogates resistance to anti-folate chemotherapy in lung cancer. *Cell Death Dis.* 2015;6(7):e1824.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

