Based on network pharmacology, using methods such as molecular docking and gene difference analysis, small molecule compounds for the treatment of lung adenocarcinoma and lung squamous cell carcinoma were screened in saussurea involucrata, and new therapeutic targets were discovered

Dongdong Zhang
Shihezi University
Tieying Zhang
Shihezi University
Yao Zhang
Shihezi University
Zhongqing Li
Shihezi University
He Li
Shihezi University
Yueyang Zhang
Shihezi University
Chenggong Liu
Shihezi University
Zichao Han
Shihezi University
Jin Li (lijin@shzu.edu.cn)
Shihezi University
Jianbo Zhu
Shihezi University

Research Article

Keywords: Saussurea involucrata, LUAD (Lung adenocarcinoma), LUSC (Lung squamous cell carcinoma), Affinity, Non-small cell lung cancer, Survival analysis

Posted Date: September 21st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-917301/v1

Version of Record: A version of this preprint was published at BMC Complementary Medicine and Therapies on February 28th, 2022. See the published version at https://doi.org/10.1186/s12906-021-03501-0.

Abstract

Purpose: Based on systemic pharmacology and network pharmacology, to evaluate the therapeutic effect of saussurea involucrata (SAIN) on lung adenocarcinoma and lung squamous cell carcinoma (LUAD&LUSC) through clinical sample genetic difference analysis and compound-target molecular docking, and discover new The target of prevention or treatment of LUAD&LUSC. **Materials and methods:** Using the TCM System Pharmacology Database (TCMSP) as the starting point for preliminary selection of ingredients and targets (OB≥30%, DL≥0.18, n=9), with the GDC database as the end point, using Cytoscape 3.8, TBtools 1.082, AutoDock 4.2.6, R 4.0.4, PyMol and other tools have conducted a preliminary screening of the ingredients and targets of SAIN. In order to further screen the effective ingredients and targets, we used clinical samples from LUAD and LUSC from TCGA and GEPIA to perform genetic difference analysis (n=6), and perform biological process (BP) analysis (FDR) on these targets. ≤0.05, n=6), KEGG pathway analysis (FDR≤0.05, n=6), protein interaction network (PPI) analysis (n=6) and compounds-targets-pathways network analysis (n=6), obtain biological processes, disease pathways and various compounds regulated by targets-the relationship between targets and pathways. Through the precise molecular docking of ingredients and targets, we further screened the targets of the effective ingredients of SAIN (affinity≤-7.0 kcal/mol, H-Bond dist≤3.0, n=6) and visualized the data, Then these targets were verified by using PSORT, CELLO and BUSCA databases for subcellular localization prediction (n=6). Finally, use the large amount of TCGA clinical data provided by Cbiportal for the prognostic survival analysis of LUAD and LUSC for the genes obtained through the screening. , And consult a large number of documents to verify the results.

Results: After screening, comparing, analyzing and verifying a series of data, it is finally confirmed that there are three main active ingredients in SAIN. They are Quercetin, Luteolin and Kaempferol, which mainly act on 6 protein targets. The target mainly regulates 20 signal pathways including Pathways in cancer, Transcriptional misregulation in cancer, EGFR tyrosine kinase inhibitor resistance, Adherens junction, IL-17 signaling pathway, Melanoma, Non-small cell lung cancer and MicroRNAs in cancer. The preventive or therapeutic effects of LUSC. **Conclusion:** There are three active compounds of Q, L and K in SAIN, which play a role in the treatment and prevention of NSCLC by directly or indirectly regulating the expression of genes such as MMP1, MMP3 and EGFR.

1. Introduction

As of 2021, lung cancer is still among the top of many worrying cancers, not only because its mortality rate is as high as 22% in both men and women (1), but the more troublesome thing is that its treatment shows a lot of difficulty.

According to the morphological characteristics of lung cancer cells under the microscope, lung cancer can be divided into two major types, namely, small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). These two types of cancer cells have different growth and spread characteristics, so The treatment options are completely different, and they must be carried out separately when researching.

Studies have shown that the two-year survival rate of SCLC is very low, always hovering around 14%. Fortunately, the two-year survival rate of NSCLC can reach 42%. In addition, NSCLC accounts for 85% of all lung cancers. It has brought good news to cancer targeted therapy researchers and NSCLC patients (2, 3).

Currently, the targeted therapy drugs for NSCLC are mainly divided into four categories, namely, epidermal growth factor receptor (EGFR), c-ros oncogene 1, receptor tyrosine kinase (ROS1), Anaplastic Lymphoma Kinase (ALK) and Met proto. -oncogene (MET) gene mutation targeted drugs, such as Gefitinib (EGFR), Entrectinib (ROS1), Lorlatinib (ALK) and Capmatinib (MET). However, due to the high research and development costs of these drugs, the relatively expensive market price, and the long taking period, it is not possible for every NSCLC patient to use them normally. Lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) are the two most common types of NSCLC. Among them, LUAD accounts for much more than LUSC and has become the basic type of targeted therapy for NSCLC.

As we all know, Chinese medicine is a kind of effective medicine, which contains a variety of anti-cancer and anti-inflammatory active ingredients. It has a gratifying effect in the face of various intractable diseases that cannot be solved by Western medicine.

Saussurea involucrata (SAIN), also known as Snow lotus (SI), is mainly distributed in Xinjiang, China. In high-altitude areas such as Tibet, wild SAIN grows in cliffs and icy crevices above 4000 meters above sea level. This environment is characterized by extremely cold, minus 20℃ all year round. This unique growth environment makes SAIN natural compounds. And it is extremely rare, which has created its unique medicinal value. It has the titles of "the best in medicine" and "the king of herbs". At present, with the application of tissue culture technology, the whole plant of SAIN has been realized in the laboratory. Cultivation has brought great convenience to researchers.

People's research on the pharmacological effects of SAIN began in 1980 (4), and it has been more than 40 years until 2021. There are many known pharmacological effects of SAIN, but the most studied is its anti-inflammatory and antioxidant properties (5). Anti-cancer (prostate cancer, stomach cancer, breast cancer, etc.) (6), as for other diseases (lung cancer), no research has been done. Therefore, it is urgent for researchers to evaluate and supplement the efficacy of SAIN in the treatment of lung cancer.

Network pharmacology is an emerging research method in recent years. It was first proposed by Professor Hopkins, a pharmacologist at Dundee University in the United Kingdom. A large number of facts have proved that the anti-disease compounds and targets screened out by the network pharmacology method are effective for clinical trials. Therefore, we will use network pharmacology methods, combined with system pharmacology, gene difference analysis, molecular docking, subcellular localization prediction and patient prognostic survival analysis, etc., to treat and prevent the target mechanism of SAIN active ingredients in the treatment and prevention of LUAD and LUSC. For research, please refer to the flow chart of this experiment (Fig. 1).

2. Materials And Methods

2.1 Screening of active ingredients with drug-like properties in SAIN

TCMSP (https://tcmsp-e.com/) database is one of the most commonly used databases for screening of active ingredients in traditional Chinese medicine. Its advantage is that the database provides oral bioavailability (OB) and drug similarity (DL) Parameters. These two parameters play an important role in the evaluation of drug efficacy. Only when OB exceeds a certain value (OB≥30%) and DL is within a certain range (DL≥0.18), can it effectively reflect the effect of a certain compound drug-like properties.

Among them, the calculation of the DL value of the system follows the formula (1). In order to obtain the target drug, the DL value will be only when the lead compound is chemically easy to be synthesized and has the properties of ADME (absorption, distribution, metabolism, excretion) kick in.

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

The x in the formula represents the descriptive index of all ingredients in SAIN, and y represents the average drug similarity index of the ingredient from the DrugBank database (https://www.drugbank.ca/).

The fat-water partition coefficient logP_(O/W) refers to the partition coefficient of the drug in the n-octanol-water system. It is widely used as a measure of the hydrophobicity of chemical compounds. The main driving force of the biofilm composed of layers controls the compounds-targets binding effect, logP_(O/W) follows the following formula (2):

$$logP_{\frac{O}{W}} = \frac{\log C_O}{\log C_W} \qquad (2)$$

In the formula, CO represents the equilibrium concentration of the drug in the oil phase, and CW represents the equilibrium concentration of the drug in the water phase. The value of logP_(O/W) indicates the hydrophobicity of the solute. The larger the logP_(O/W), the stronger the hydrophobicity, and vice versa, the stronger the hydrophilicity. Therefore, we use logP_(O/W)≤5 as a screening criterion.

2.2 Target prediction and construction of active compounds-targets network

The structural formula of the SAIN active ingredient obtained above is drawn on the STP database (http://www.swisstargetprediction.ch) for target prediction, and the prediction result is combined with the NSCLC target retrieved in the TCMSP database to obtain the final required Disease target information. Use Cytoscape 3.8.0 software (7) to visually construct a compounds-targets network (C-T network) for the above active compounds and targets to obtain the C-T network relationship diagram we need.

So far, we have completed the preliminary screening of the active ingredients related to lung cancer in SAIN.

2.3 Gene difference analysis and LUAD&LUSC gene TPM data analysis

In order to further screen the active ingredients that we have already screened and their targets, we obtained a large number of clinical case samples of LUAD and LUSC from the GEO database (https://www.ncbi.nlm.nih.gov/geo/), through the analysis of gene differences of tens of thousands of samples, samples are randomly selected to draw a heat map of related gene differences, and a preliminary screening of targets is carried out. In order to screen the target for a second time to make the result more clear, we used the gene ID of the target and used the TCGA data provided by the GEPIA database (http://gepia.cancer-pku.cn/) as the basis to draw Quantitative scatter plots of transcripts per million (TPM) of the screened genes further screened out the effective compounds and targets of SAIN for the treatment of LUAD and LUSC.

2.4 Construction of protein interaction network, compounds-targets-pathways network and gene enrichment analysis

In order to evaluate the interaction between the targets screened above, the protein interaction network and the active compounds-targets-pathways network were constructed. In order to explore the biological processes and pathways that each target participates in the body, the STRING database (https://www.string-db.org) retrieved the target gene data, we performed biological process (BP) analysis in GO (gene ontology) analysis and KEGG (kyoto encyclopedia of genes and genomes) enrichment analysis. Among them, the FDR of GO analysis is less than or equal to 0.05, and the FDR of KEGG enrichment analysis is less than or equal to 0.05, both of which meet the requirements and statistical significance of significant gene enrichment in vivo.

The FDR value is to correct the value of P, and the results obtained by using it are more accurate. Therefore, this step aims to obtain the biological process and in vivo pathways of the target action, which provides a basis for subsequent research.

2.5 Molecular docking and subcellular localization prediction

Through AutoDock 4.2.6 and PyMol and other tools for molecular docking, the binding energy of the compound and the target is first used to verify whether the effective compounds of SAIN are reliable in the treatment of LUAD&LUSC, and further exclude the compounds of SAIN that have poor effects on LUAD&LUSC.

Molecular docking (8) is a method of drug design based on the characteristics of the receptor and the interaction between the receptor and the drug molecule. A theoretical simulation method that mainly studies the interaction between molecules (such as ligands and receptors), and predicts its binding mode and affinity. In recent years, molecular docking methods have become an important technology in the field of computer-aided drug research (9).

Subcellular localization prediction is a popular subcellular localization method in recent years. It uses existing data to create a database of the sequence relationship between various genes and their regulatory target sequences and subcellular structures, which can accurately predict the target protein The location of various organelles and cell membranes has brought great convenience to scientific researchers. Currently commonly used subcellular location prediction tools are (1) PSORT Ⅱ (https://psort.hgc.jp/form2.html), the database uses k-Nearest Neighbor (K-NN) algorithm, K-NN is data mining Compared with the commonly used learning algorithms in machine learning, K-NN has a wide range of applications. PSORT II can identify a classic nuclear localization signal (cNLS) sequence, and its accuracy is very high when the sample size is large enough (10,11 ,12); (2) CELLO (http://cello.life.nctu.edu.tw/), the database uses the support vector machine recursive feature elimination algorithm (SVM-RFE), and the SVM-RFE algorithm is trained on the basis of SVM The weight vector w generated at time is used to construct the sorting coefficient, and each iteration removes a characteristic attribute with the smallest sorting coefficient, and finally obtains the descending order of all the characteristic attributes (13,14); (3) BUSCA (http://busca. biocomp.unibo.it/), the database uses betaware algorithms to solve the detection of transmembrane beta-barrels (TMBB) in the proteome and the prediction of its topology (15,16).

Combined with the analysis of biological processes, PSORT Ⅱ, CELLO and BUSCA databases were used to predict subcellular localization, and the prediction results of the three tools were compared. We selected the overlapping parts to obtain the main location of the target protein in the cell.

2.6 Survival analysis of patient prognosis

In medical research, in order to evaluate the efficacy of a certain drug and understand survival data such as survival time of patients after surgery, the analysis of these survival data is called survival analysis.

After a series of screening and analysis of the compounds and targets of SAIN, we performed prognostic overall survival analysis (17) on the gene IDs of the targets of the effective compounds of SAIN (17) to explore when the drugs act on these targets, the length of time the patient can survive over time.

3. Results

3.1 Screening results of SAIN active ingredients

We collected a total of 55 known active ingredients of SAIN from the TCMSP database, including alkaloids, lipids, flavonoids and flavonoids, etc. We selected 55 from 55 under the screening conditions of OB ≥ 30%&DL ≥ 0.18&log$P_{o/w}$≤5, six effective drug-like ingredients were screened out of the SAIN, namely flazin, quercetin, kaempferol, luteolin, alloisoimperatorin and hispidulin. See (Table 1) for the screening details of these six ingredients.

Table 1
Screening information of six compounds in SAIN

Molecule Name	MW(KDa)	logP(o/w)	OB(%)	DL
Flazin	308.31	3.235	94.27575085	0.38559
Quercetin	302.25	1.504	46.43334812	0.27525
Kaempferol	286.25	1.771	41.88224954	0.24066
Luteolin	286.25	2.067	36.16262934	0.24552
Alloisoimperatorin	270.3	3.793	34.80406732	0.21854
Hispidulin	300.28	2.318	30.97205344	0.27025

3.2 Target screening and compounds-targets network analysis results

By merging the 1023 targets screened from the STP database and the 764 LUAD and LUSC targets screened from the TCMSP database, we obtained a total of 9 targets with six compounds effects. These 9 protein targets were combined with The six SAIN compounds interact, and Cytoscape is used to construct a compounds-targets network, and a C-T network with 15 nodes and 24 edges as the characteristics is obtained (Fig. 2).

Through the C-T network, it can be found that the six compounds of SAIN that have strong and weak effects on the target are ranked quercetin (Q), luteolin (L), kaempferol (K) and hispidulin (H), flazin (F), alloisoimperatorin (A), The size order of the 9 targets and compounds is prostaglandin-endoperoxide synthase 2 (PTGS2), 90kDa heat shock protein AA1 (HSP90AA1), Matrix metalloproteinase 1 (MMP1), Epidermal growth factor receptor (EGFR), etc.

3.3 Gene difference analysis and TPM data analysis results

In order to ensure the accuracy of the experimental data, we further screened the compounds and targets. Through the GEO database and R 4.0.4, we obtained a total of 22,589 LUAD and 22697 LUSC-related disease genes through the GDC database. (Https://portal.gdc.cancer.gov/) TCGA data, we obtained 535 LUAD samples and 59 LUAD control samples, 502 LUSC samples and 49 LUSC control samples. It is undeniable that these data all have high reference value. However, in order to make the experiment proceed smoothly, we aim at the 9 target points studied. Through the P value range ($P \leq 0.05$) of these data and patient samples, the target points are determined by R. Combined with clinical samples, the data was screened, and the gene expression matrices of LUAD&Control and LUSC&Control corresponding to 9 targets were obtained. We randomly selected 15 samples from each of the two groups in the two groups, namely the LUAD&Control group, the LUSC&Control group Thirty clinical samples were randomly selected, and TBtools v1.082 (18) software was used to draw genetic difference heat maps of LUAD (Fig. 3b) and LUSC (Fig. 3a The direction of the ordinate is "←"). In the figure, since 9 targets can act on both types of lung cancer, we cluster genes. Since our samples are divided into two groups (experimental

group and control group), in order to make the samples evenly distributed, and on the premise of not affecting the experimental results, we do not allow the samples to cluster.

According to the results of genetic difference analysis, we can find that HSP90AA1 has the least obvious effect on LUAD and LUSC, so we discard it. Through the GEPIA database, input the gene ID, and select the analysis of variance (ANOVA) for the difference method. The Log_2FC value and Q value are 1 and 0.01 respectively. The gene IDs of the remaining 8 targets are analyzed by TPM. The total number of samples involved is LUAD (T = 483, N = 347), LUSC (T = 486, N = 338), where T stands for tumor patients and N stands for normal control group (Fig. 4).

Combining the corresponding relationship between the binding compounds and the target, the results of gene difference analysis and the results of TPM analysis, we found that the expression and changes of MMP2 and PTGS2 in the T group were consistent with those in the N group, so they were discarded.

Therefore, the SAIN compounds that play a major role in LUAD and LUSC are Q, L, and K, and the genes that play a major role in LUAD and LUSC are MMP1, MMP3, EGFR, MET, NQO1 and MPO.

For the analysis of six genes, we came to the conclusion that the expression of EGFR and MMP3 genes in LUSC far exceeds that in LUAD, while MET is the opposite. Although the other three genes are consistent in the occurrence and development of the two diseases, MPO is worthy of attention. This gene showed high expression in all stages in the N group of cases, and low expression in the T group. Expression status, therefore, we speculate that excessive activation of MPO expression in tissues may be effective in inhibiting the development of LUAD and LUSC.

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

According to the above analysis, we have excluded the three targets of HSP90AA1, PTGS2 and MMP2. Studies have shown that HSP90AA1 and the same type of HSP90AB1 may be required for the treatment of the two subtypes of breast cancer (19). PTGS2 is used in colorectal cancer. It is needed in other diseases (20), and MMP2 mostly plays a role in inflammation and immune response (21, 22).

The remaining 6 targets were subjected to GO biological process (BP) analysis (Fig. 5a), KEGG enrichment analysis (Fig. 5b) and PPI analysis (Fig. 6).

In order to achieve the credibility and statistical significance of the data, we performed GO (BP) analysis and KEGG enrichment analysis with FDR $\leq$ 0.05. The results are more rigorous and reliable than the results obtained by using P values. Therefore, the results are very great reference value, see the attachment (Table 2 and Table 3) for detailed information.

With reference to the results of KEGG enrichment analysis, in order to make it easy to understand, we use Cytoscape to link the 3 compounds, 6 targets and 20 pathways, and draw a compounds-targets-pathways diagram (Fig. 7) to make the compounds The relationship with the target and the relationship

between the target and the pathway are visualized. A total of 20 metabolic regulation pathways including 29 nodes and 53 edges are obtained.

First, from the C-T-P network diagram, we sort the targets into EGFR, MET, MMP1, MMP3, NQO1, and MPO according to the number of target regulatory pathways and the number of corresponding compounds, and sort the pathways into Pathways in cancer, Transcriptional misregulation in cancer, Hepatocellular carcinoma, Bladder cancer, EGFR tyrosine kinase inhibitor resistance, Adherens junction, IL-17 signaling pathway, Melanoma, Non-small cell lung cancer, Relaxin signaling pathway and MicroRNAs in cancer, etc. However, despite these pathways both LUAD and LUSC have contributed, but the strength of the contribution is still not easy to distinguish. Therefore, further analysis and verification are needed.

3.5 Analysis of molecular docking results

After the above series of analysis, screening and verification work, we have determined the target sites of the three compounds respectively. These three compounds can act on multiple targets. Among them, Q can act on MMP1, MMP3, EGFR, MPO and NQO1, L can act on MMP1, EGFR and MET, K only acts on MMP1.

After obtaining this information, we used Chemdraw 19.0 software to draw the structure of the 6 compounds, saved as a mol2 file, and used the PDB database (https://www.rcsb.org/) to select protein tertiary structure of ligand-containing and ligand according to the structure of the compounds and download the pdb file, process the pdb protein file through Discovery Studio 4.5 Client software (remove water molecules and redundant structures, etc.), use AutoDock 1.5.6 software to save the two file formats as pdbqt files, select the configuration Body and coordinate position (x, y, z), the compounds are screened and sorted by the distance of the hydrogen bond. These tertiary structures are all predicted using the X-RAY DIFFRACTION method, and the resolution of the six proteins is 1.90 Å, and the highest is 2.40 Å. Then, we use AutoDock 1.5.6 and AutoDock Vina to do the molecular docking of the compound and the target, and finally use PyMol to process the docking result (Fig. 8).

In the figure, the coordinates of each compound 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), NQO1(22, 22, 22), where the greater the value in this coordinate, the greater the affinity for docking, and the greater the affinity, which proves that the binding of the compound to the target is tighter, and the compound is also The more effective. However, in order to ensure the availability and reference value of the data, we set the coordinates (x, y, z) according to the structural size of the compound itself. Therefore, the coordinate values we use are just suitable for the compound itself. The results show that the binding energy (kcal/mol) of the compound and the target is the best is -10.1 kcal/mol, and the worst is -7.9 kcal/mol.

According to the results of molecular docking, we selected items with affinity within a certain range (affinity≤-7.0 kcal/mol). The specific molecular docking results of 3 SAIN compounds and 6 targets can be found in the attachment (Table 4).

In order to visualize the results, we use heat maps to show the results in detail below (Fig. 9).

3.6 Analysis of subcellular location prediction results

In order to explore the specific positions of the 6 proteins we screened out in the cell, we used the PSORT Ⅱ, CELLO and BUSCA databases to predict the subcellular location, and obtained the prediction results of the three databases PSORT Ⅱ (Table 5), CELLO (see attachment for Table 5) and BUSCA (see attachment for Table 5).

Comparing the results of the three subcellular location prediction databases, we found that PSORT Ⅱ has higher compatibility and reliability than the other two databases, so 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 subcellular location prediction.

Table 5
Prediction results of subcellular localization 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%		

The prediction results of subcellular localization show that MMP1, MMP3, NQO1, EGFR, MET and MPO can all be found in the cytoplasm. Among them, MMP1, MMP3, NQO1 and MET are also present in mitochondria. In addition to MMP3, the other 5 proteins can be found. Found in the nucleus.

3.7 Analysis of prognosis and survival of patients with 6 target genes

Survival analysis in the traditional sense, its purpose is to analyze the genetic data related to the disease before the patient undergoes surgery or other treatment. In this sense, it provides patients with a choice of whether to continue treatment (23).

After analyzing the survival of these 6 targets, we found that the 6 genes all showed statistical significance in the regulation of LUAD ($P < 0.05$), while in the regulation of LUSC, only EGFR and MMP3 had statistical significance ($P < 0.05$), which is consistent with our previous TPM analysis results (Fig. 11).

4. Discussion

4.1 Current status of LUSC treatment targets and MMP3

Compared with LUAD, LUSC has a unique pathological morphology (the growth position, direction and variation of cancer cells) in NSCLC. Therefore, because there are few carcinogenic aberrations that can target LUSC, no targeted therapy method for LUSC has been developed.

Fibroblast growth factor (FGFR) and EGFR belong to the receptor tyrosine kinase (RTK) family. The FGFR family contains four members FGFR1-4. We have observed FGFR1-4 amplification or mutation in NSCLC. It plays a vital role in tumor development and maintenance. In 2016, a gene sequencing study involving 4853 cancer patients showed that 7.1% of cancer patients had abnormalities in FGFR1-4 genes, and the most common was the amplification of FGFR-1 gene, accounting for 50% (24). A large amount of clinical evidence shows that (25), FGFR-1 inhibitors have a significant effect on the symptoms caused by LUSC, especially Infigratinib (BGJ398), BGJ398 treatment of LUSC, the disease control rate reached 47.6% (24). Therefore, FGFR-1 may be the next effective target for selective treatment of LUSC.

Matrix metalloproteinase 3 (MMP3) is the third member of the matrix metalloproteinase (MMP) family. Studies have shown that it plays many roles in normal humans and patients, such as promoting epithelial-mesenchymal transition and increasing pulmonary fibrotic mediators. Level or activity or reduce anti-fibrotic mediators and promote abnormal epithelial cell migration and other abnormal repair processes (26). In addition, adipocytes can increase the ability of tumor cells to invade the body by producing exosomes with high levels of MMP3, and increase lung cancer. The risk of metastasis of local tumors (27, 28).

We already know that insulin enhances the proliferation, migration and drug resistance of non-small cell lung cancer cells by activating the PI3K/Akt pathway. Studies have shown that insulin can up-regulate the expression of MMP3 genes in NSCLC cells, but its expression will be affected by PI3K/Akt. Inhibited by pathway inhibitors (29).

Studies have shown that FGFR-1 and MMP3 genes and proteins are highly expressed in esophageal squamous cell carcinoma, and both may be related to the invasion and metastasis of esophageal cancer (30). Our analysis results show that MMP3 can positively regulate NSCLC, especially LUSC, by participating in the transcriptional error regulation in cancer, IL-17 signaling pathway and rheumatoid arthritis pathway. The molecular docking results show that Q can interact with MMP3 protein. For binding, the binding affinity is as high as -8.9 kcal/mol. Therefore, we speculate that MMP3 and FGFR-1 may have the same or homology in the treatment of LUSC, that is, MMP3 can be used as a biomarker for the diagnosis and prognosis of LUSC (31) However, more detailed results still need a large number of clinical cases to verify.

4.2 Application of EGFR and MET in LUAD targeted therapy

The latest NSCLC guidelines 2021 V2 released by the National Comprehensive Cancer Network (NCCN) show that the detected genes related to NSCLC targeted therapy include EGFR, KRAS, ALK, ROS1, MET, BRAF, RET, and NTRK (32). However, these genes are mainly for LUAD.

Receptor tyrosine kinase (RTK) is a high-affinity cell surface receptor for many hormones, cytokines and polypeptide growth factors. Among the 90 unique tyrosine kinase genes identified in the human genome, 58 encode receptor tyrosine kinase proteins. Receptor tyrosine kinases have not only been shown to be key regulators of normal cellular processes, but also play a key role in the occurrence and development of many types of cancer (33).

Epidermal growth factor receptor (EGFR) and hepatocyte growth factor (HGF) are members of the RTK family, and MET is the tyrosine kinase transmembrane receptor of HGF, a protein encoded by the c-MET proto-oncogene The product, as a potential therapeutic target for many types of tumors, has attracted attention in the medical field.

At present, the world has developed a variety of targeted drugs for NSCLC against EGFR mutations and MET exon14 changes. Among them, drugs for EGFR target mutations are mainly divided into small molecule tyrosine kinase inhibitors (TKI) and monoclonal antibodies. Major categories, the TKI category mainly includes Gefitinib, Erlotinib, Afatinib, Osimertinib, Dacomitinib, Lcotinib and Brigatinib, etc. Monoclonal antibodies are relatively few, mainly including Necitumumab, Amivantamab and Ramucirumab, of which Amivantamab targets EGFR exon 20 insertion.

There are also two types of drugs that are modified for MET exon14: TKI and monoclonal antibodies, including multiple kinase inhibitors including TKI (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 directly shows that our analysis results are correct, that is, the expression of the MET gene in LUAD patients is far more than that of LUSC patients, and the expression of EGFR gene in the two types of patients is roughly the same. Occasionally, the expression of LUSC

patients may be even higher. Much phenomenon. The results of molecular docking show that the compound L and Q of SAIN can bind to the EGFR protein stably with an affinity of -8.8 kcal/mol, while the MET protein can only bind to the compound L with an affinity of -7.9 kcal/mol. Through survival analysis, we know that among LUAD patients, EGFR mutation patients ($P < 0.001$) have a significantly shorter prognostic survival period than MET patients ($P < 0.02$).

4.3 The role of MPO and NQO1 in NSCLC patients

We all know that smoking increases the risk of lung cancer. In the early years, researchers found that in a large number of patients with lung adenocarcinoma combined with smoking and non-smoking, Nad(p)h: quinone oxidoreductase 1 (NQO1) there was no significant difference in genotypes, and smokers with myeloperoxidase (MPO) genotype had a lower prevalence of lung cancer than non-smokers (34).

MPO is a member of the heme peroxidase superfamily and is mainly expressed in neutrophils and monocytes. Studies have shown that MPO targeted therapy can show a good prognostic survival status in LUAD patients (35), which is consistent with the results of our survival analysis ($P < 0.001$). The results of molecular docking show that the compound Q of SAIN can specifically bind to MPO protein (affinity = -8.8 kcal/mol). Although there is evidence that MPO has an effect on NSCLC patients, the results of genetic difference analysis and TPM analysis show that it is different from other targets. The expression of MPO in tumor tissues of LUAD and LUSC is much lower than its expression in normal tissues. The amount ($TPM < 2$), that is, MPO is basically not expressed in NSCLC patients. According to the results of KEGG enrichment analysis, we speculate that MPO may play a certain role in LUAD and LUSC patients through the transcriptional misregulation in cancer pathway.

The results of PPI analysis showed that MPO and NQO1 may have similarities in function, and NQO1 has been confirmed to be overexpressed in a variety of solid tumors, including lung (36, 37), which is consistent with our analysis results. Currently, β-lap-dC3 prodrug micelles have been developed to target NSCLC overexpressing NQO1. Studies have shown that the drug can significantly prolong the prognostic survival time of NSCLC patients (38), which is consistent with the results of our NQO1 survival analysis. It is consistent, that is, when NQO1 is overexpressed, the patient's survival time is greatly reduced. The molecular docking results showed that the affinity of compound Q when binding to NQO1 protein was −8.5 kcal/mol, which proved the effectiveness of this compound as a drug.

4.4 The potential of high expression of MMP1 in the treatment of NSCLC

The results of the compounds-targets molecular docking show that compound Q, L, and K can bind to the MMP1 protein, and the binding affinity is better than any other compounds, respectively Q (affinity = -10.0 kcal/mol), L (affinity = -10.1 kcal/mol) and K (affinity = -9.6 kcal/mol). We analyzed the survival of LUAD and LUSC patients after MMP1 overexpression and found that compared with the survival time of patients after EGFR mutation, MMP1 overexpression showed a worse prognostic survival. The PPI network analysis results show that MMP1 forms three triangular structures with EGFR, MMP3, NQO1 and

MPO **(Fig. 12)**, which suggests that when MMP1 functions, it may form connections with four other targets.

MMP1, like MMP3, belongs to the MMP family. However, because the expression of MMP1 in rodent orthologs is not conservative (39), the research on MMP1 is still limited. However, people still use mice, cell models and other methods to study the function of MMP1. Related results show that (1) MMP1 can regulate the mechanism of LUAD cell proliferation, migration and invasion under the regulation of miR-202-3p (40), (2) lung cancer caused by XPC gene (xeroderma pigmentosum) defect P53 dysfunction may enhance tumor metastasis by increasing the expression of MMP1 (41), and (3) the specific inhibition of MMP1 secretion by macrophages may be a potential therapy for reducing lung metastasis in smoking cancer patients (42).

Therefore, combined with the above results, we speculate that MMP1 may be a new target for NSCLC treatment, and L, Q, and K may be related lead compound.

4.5 Verification of the effectiveness of the three compounds Q, L and K

Finally, we screened out three active compounds Q, L and K from SAIN. However, whether these three ingredients are effective for NSCLC requires further verification by cell experiments or animal experiments. Before doing in vitro and in vivo verification, we found a large number of experimental verification documents about these ingredients in the treatment of NSCLC, so we did not conduct in vitro or in vivo verification.

L (2-(3,4-dihydroxyphenyl)-5,7-dihydroxy-4-chromenone) is a natural flavonoid compound. According to research, the effect of L on NSCLC is mainly reflected in (1) In lung cancer A549 cell nuclear H460 xenotransplanted mice, the mRNA and protein expression levels of L significantly reduced the expression of melanoma 2 (AIM2) deletion, thereby inhibiting the activation of AIM2 inflammasomes, thereby inducing cell cycle arrest in G2/M phase. Inhibit the epithelial-mesenchymal transition (EMT) process in NSCLC (43). (2) At the cellular and animal level, L has a significant anti-tumor effect on the L858R/T790M mutation of EGFR and erlotinib-resistant NSCLC. The mechanism is that L inhibits the binding of Hsp90 to the mutant EGFR protein. Induces the degradation of EGFR and further prevents PI3K/Akt/mTOR signal transduction, leading to NSCLC cell apoptosis (44, 45). (3) At physiological concentrations, L significantly makes A549 cells more sensitive to the anticancer drugs oxaliplatin, bleomycin and doxorubicin, and can be used as a natural sensitizer in chemotherapy to achieve better Drug availability (46, 47). These experimental verification results are basically consistent with our analysis, prediction, and verification results, which shows that L has great potential in the treatment of NSCLC.

Q (2-(3,4-dihydroxyphenyl)-3,5,7-trihydroxy-4-chromenone) and L are very similar in structure, the only difference is that Q has one more at position 3 of the chromenone core Hydroxy. Epidemiological studies have shown that Q has the effect of preventing lung cancer. This is mainly reflected in (1) Q can

significantly enhance the cytotoxicity induced by tumor necrosis factor-related apoptosis-inducing ligand (TRAIL) in NSCLC cells, thereby accelerating Death of NSCLC cells (48). (2) Starting H520 cells (LUSC cells) with Q can increase the apoptosis induced by cisplatin by 30.2%. This process is accompanied by the expression of various apoptosis-related genes induced by Q, which can be used as an effective chemical increase Sensitizer (49). (3) Q can increase or decrease the expression of many miRNAs or TKIs, including increasing the expression of miR-16 family members miR-16-5p, decreasing the expression of WEE1, and inhibiting the expression of the Src family (50, 51). Similarly, our prediction result is that Q may treat or alleviate disease symptoms by reducing the expression of MMP3, EGFR, NQO1 and MMP1 in NSCLC tissue.

Studies have shown that dietary flavonoid K (3,5,7-trihydroxy-2-(4-hydroxyphenyl)-4-chromenone) can effectively prevent and treat lung cancer, mainly through the following mechanisms: (1) K can induce tumor suppressor genes -Overexpression of antioxidant enzyme (SOD-2) to inhibit apoptosis of H460 cells (NSCLC cells) (52). (2) K inhibits Akt1-mediated phosphorylation of Smad3 at threonine 179 (Thr179), thereby inhibiting transforming growth factor-β1 (TGF-β1)-induced epithelial to mesenchymal transformation of A549 lung cancer cells and Migration (53). (3) K makes chemically resistant cancer cells sensitive to multi-target antifolate (MTA) by inhibiting the epithelial-mesenchymal transition (EMT) signal pathway. This makes K reverse the NSCLC's ability to resist MTA and achieve the ability to pass EMT Pathway inhibits the effect of lung cancer chemotherapy response (54). Our results are that K plays an anti-NSCLC effect by inhibiting the MMP1-mediated IL-17 signaling pathway, rheumatoid arthritis pathway, and relaxin signaling pathway. These conclusions are consistent with our results.

5. Conclusion

We know that the etiology of NSCLC is very complicated, and we cannot explain it with a single gene change. Therefore, the treatment of lung cancer cannot be achieved by one method. It also requires a combination of multiple methods and multiple targets, although SAIN The three compounds of Q, L and K can prevent and treat NSCLC through the targets of MMP1, MMP3, EGFR, MPO, MET and NQO1, but none of us can guarantee that the cause of NSCLC is the result of these genes. Therefore, the complete cure of NSCLC still requires the joint efforts of researchers in various fields to achieve.

Abbreviations

SAIN, Saussurea involucrate
NSCLC, Non-small cell lung cancer
LUAD, Lung adenocarcinoma
LUSC, Lung squamous cell carcinoma
SI, Snow lotus
Q, Quercetin
L, Luteolin
K, Kaempferol

H, Hispidulin
A, Alloisoimperatorin
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.

References

1. Siegel RL, Miller KD, Fuchs H, Jemal A. Cancer Statistics. CA Cancer J Clin. 2021;71:7–33.
2. Howlader N, Forjaz G, Mooradian MJ, Meza R, Feuer EJ. The effect of advances in lung-cancer treatment on population mortality. New England Journal of Medicine. 2020;383:640–649.
3. 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.
4. Li G-h, Liu F, Zhao R-c. Studies on pharmacological actions of Saussurea involucrata Kar et Kir ex Maxim (author's transl). Acta pharmaceutica Sinica. 1980;15(6):368.
5. 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 Involucrata Injection against Severe Acute Pancreatitis- Induced Brain Injury in Rats. Biomedicine & Pharmacotherapy. 2018;100:564–574
6. 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 involucrata in the lipopolysaccharide-stimulated inflammation in cultured RAW 264.7 cells. Journal of Ethnopharmacology. 2020; 251:112532–112532.
7. 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.
8. Pinzi L, Rastelli G. Molecular Docking: Shifting Paradigms in Drug Discovery. Int J Mol Sci. 2019;20(18):4331.
9. Kaur T, Madgulkar A, Bhalekar M, Asgaonkar K. Molecular Docking in Formulation and Development. Curr Drug Discov Technol. 2019;16(1):30–39.
10. Ying H, Li Y. Prediction of protein subcellular locations using fuzzy k-NN method. Bioinformatics, 2004, 20(1):21–28.
11. 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.
12. Azar WJ, Zivkovic S, Werther GA, Russo VC. IGFBP-2 nuclear translocation is mediated by a functional NLS sequence and is essential for its pro-tumorigenic actions in cancer cells. Oncogene. 2014;33(5):578–588.

13. 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 Science. 2004, 13:1402–1406.
14. Yu C-s, Chen Y-c, Lu C-h, Hwang J-k. Prediction of protein subcellular localization. Proteins: Structure, Function and Bioinformatics. 2006, 64:643–651.
15. Castrense Savojardo, Pier Luigi Martelli, Piero Fariselli, Giuseppe Profiti, Rita Casadio, BUSCA: an integrative web server to predict subcellular localization of proteins. Nucleic Acids Research. 2018;46: W459–W466
16. Castrense S, Piero F, Rita C. BETAWARE: a machine-learning tool to detect and predict transmembrane beta-barrel proteins in prokaryotes. Bioinformatics. 2013(4):504–505.
17. 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 Research, 2017(W1):W1.
18. 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.
19. 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.
20. 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):1490–1497.
21. 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.
22. Muniz-Bongers LR, McClain CB, Saxena M, Bongers G, Merad M, Bhardwaj N. MMP2 and TLRs modulate immune responses in the tumor microenvironment. JCI Insight. 2021;6(12):e144913.
23. Jung SH, Lee HY, Chow SC. Statistical Methods for Conditional Survival Analysis. Journal of Biopharmaceutical Statistics. 2018;28(5):927–938.
24. Desai A, Adjei AA. FGFR Signaling as a Target for Lung Cancer Therapy. J Thorac Oncol. 2016;11(1):9–20.
25. 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):116–120.
26. 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.
27. 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):81880–81891.
28. 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):15164–15179.

29. 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):447–452.
30. Tan Yong. Expression and clinical significance of FGFR1 and MMP3 in esophageal squamous cell carcinoma. J Mod Oncol. 2017(05):729–733.
31. 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):480–489.
32. 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, Lanuti M, Leal TA, Lin J, Jr BWL, 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):254–266.
33. Robinson DR, Wu YM, Lin SF. The protein tyrosine kinase family of the human genome. Oncogene. 2000;19(49):5548–5557.
34. Kiyohara C, Yoshimasu K, Takayama K, Nakanishi Y. NQO1, MPO, and the risk of lung cancer: a HuGE review. Genet Med. 2005;7(7):463–478.
35. Hao Y, Hu P, Zhang J. Genomic analysis of the prognostic effect of tumor-associated neutrophil-related genes across 15 solid cancer types: an immune perspective. Ann Transl Med. 2020;8(22):1507.
36. Oh ET, Park HJ. Implications of NQO1 in cancer therapy. BMB Rep. 2015;48(11):609–617.
37. Ma Y, Kong J-n, Yan G-h, Ren X-s, Jin D, Jin T-f, Lin L-j, Lin Z-h. NQO1 overexpression is associated with poor prognosis in squamous cell carcinoma of the uterine cervix. BMC Cancer. 2014;14:414.
38. 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;200:201–211.
39. 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):153–159.
40. 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):406–416.
41. 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):10422–10432
42. 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.

43. 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.
44. Hong Z, Cao X, Li N, Zhang Y-z, Lan L, Zhou Y, Pan X-l, 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):2842–2853.
45. 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):1385–1391.
46. 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):1599–1609.
47. 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. International journal of oncology. 2015;46(3):1149–1158.
48. 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):2114–2121.
49. 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):1297–1303.
50. 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. IUBMB Life. 2020;72(5):1012–1022.
51. 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-κB) pathway. Med Sci Monit. 2020;26:e920537.
52. 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):2005–2013.
53. Jo E, Park SJ, Choi YS, Jeon WK, Kim BC. Kaempferol Suppresses Transforming Growth Factor-β1-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):525–537.
54. 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.

Tables

Due to technical limitations, tables 2, 3, and 4 are only available as downloads in the supplementary files section.

Figures

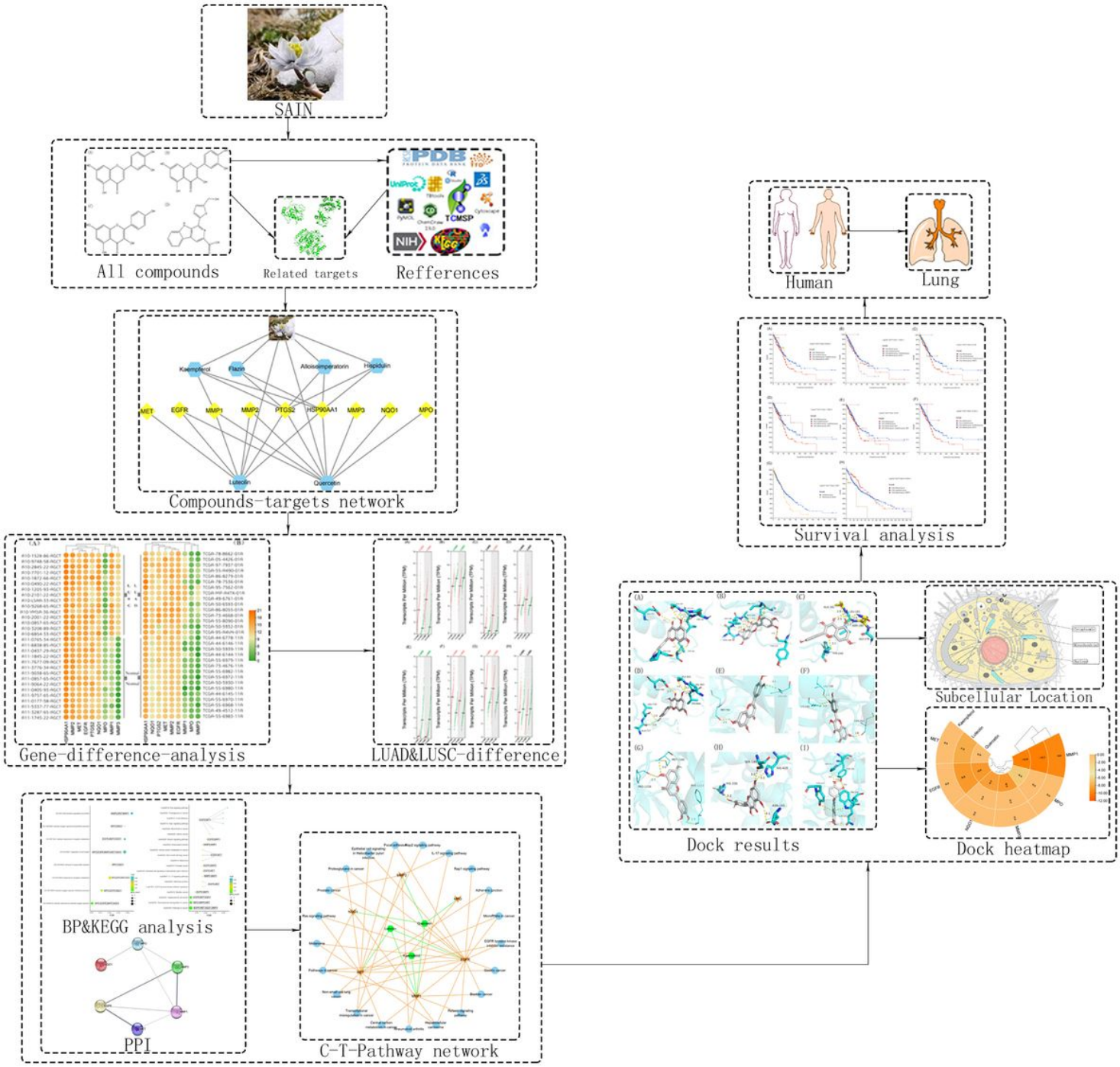

Figure 1

Workflow of experiment to predict the mechanism of SAIN in the treatment of LUAD&LUSC.

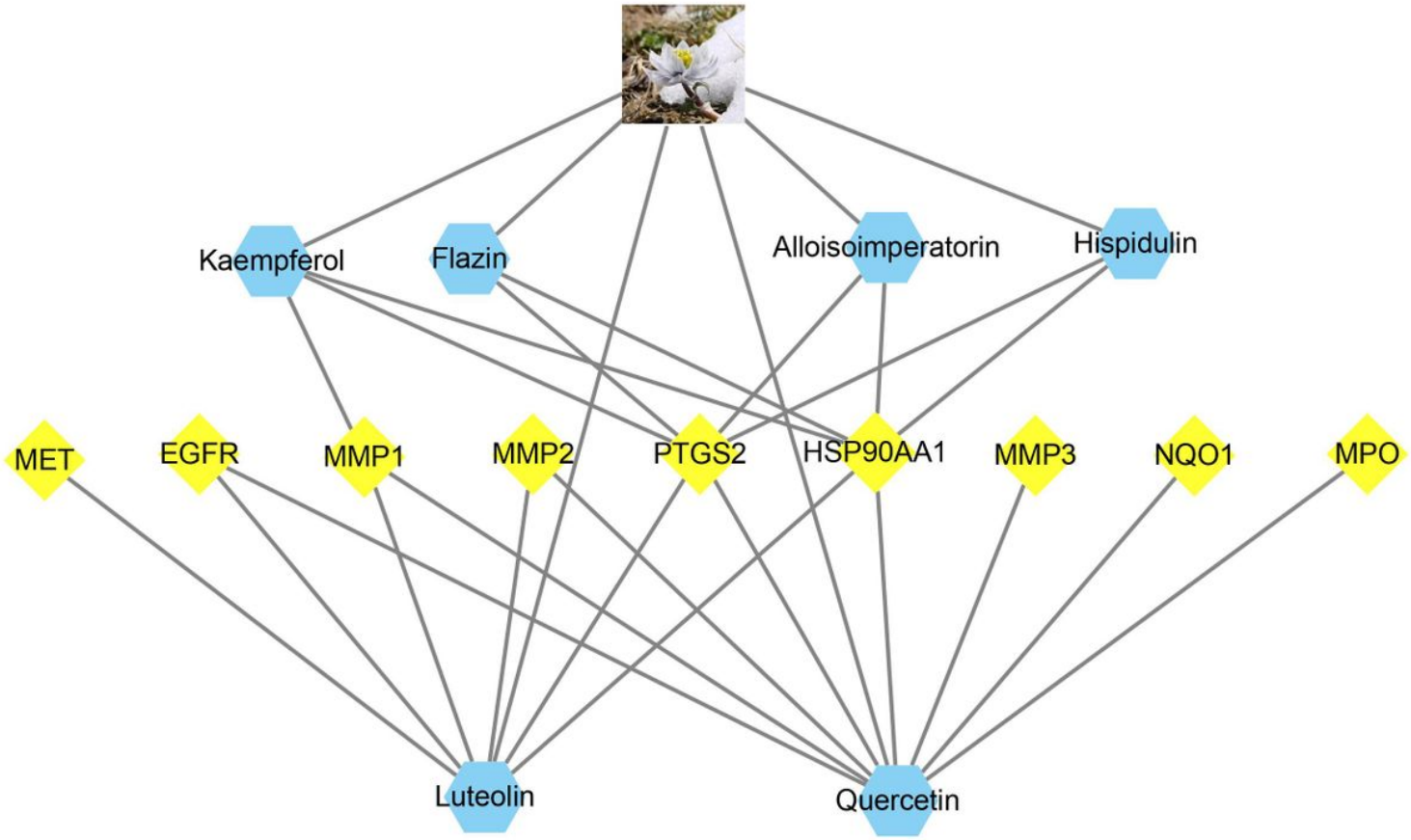

Figure 2

C-T network, the benzene ring represents the compounds and the diamond represents the targets

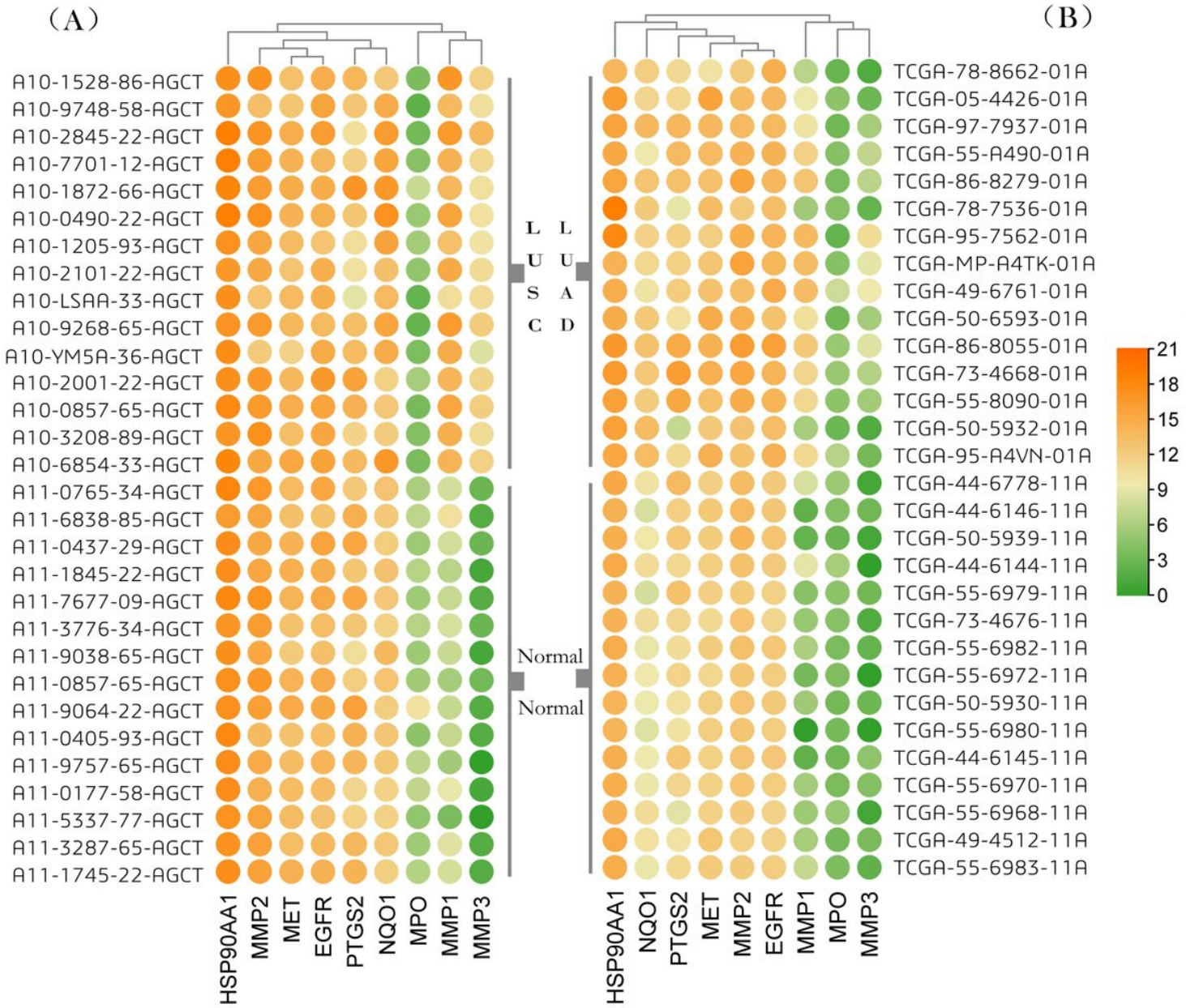

Figure 3

Heat map: gene difference analysis of LUAD & LUSC, number represents the log value of the difference of each gene and green indicates lower expression, orange is the opposite. (a). 30 clinical samples of LUSC corresponding to 9 targets (including 15 normal controls) (b). 30 clinical samples of LUAD corresponding to 9 targets (including 15 normal controls)

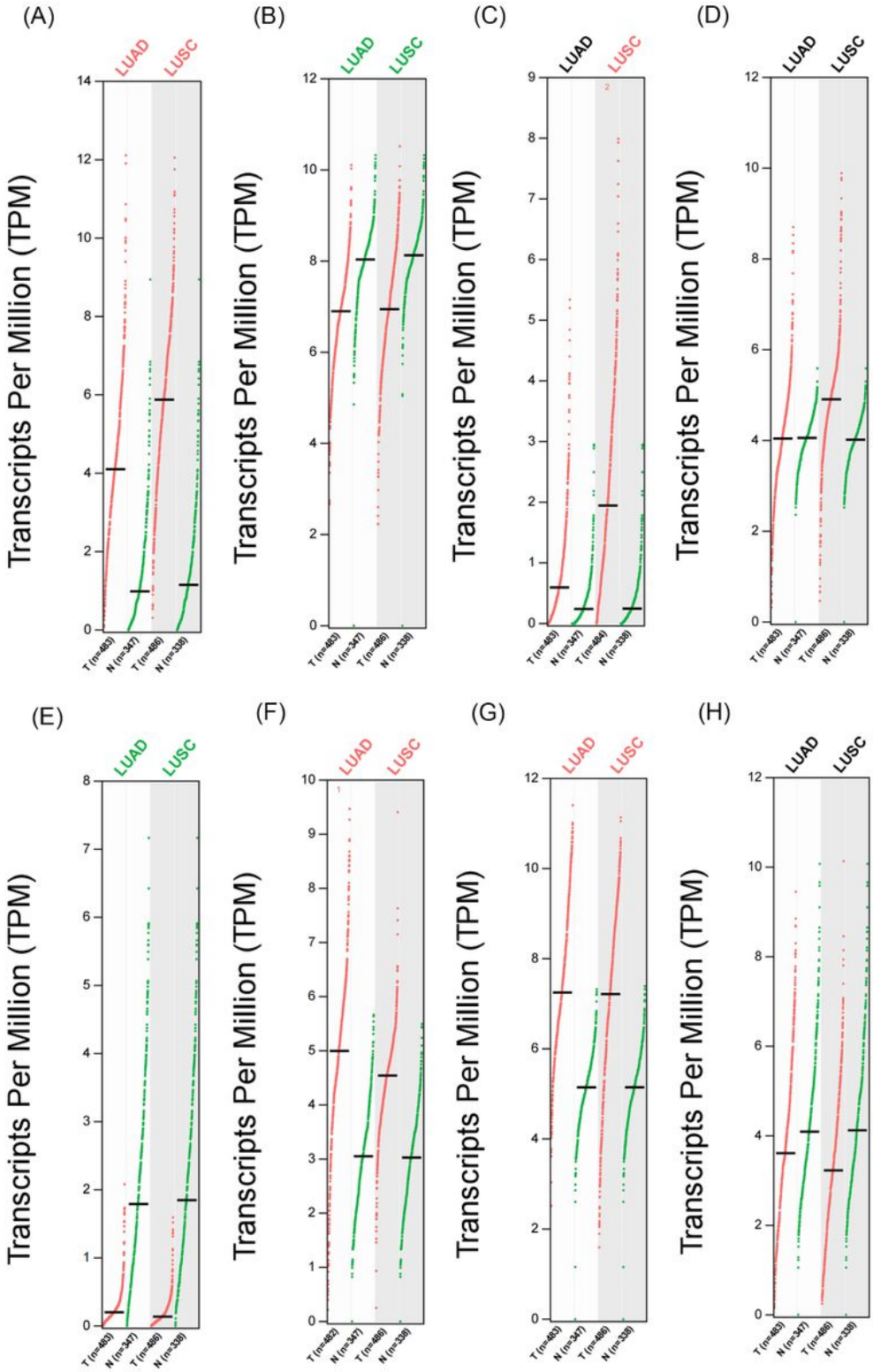

Figure 4

Gene TPM analysis of 8 targets of LUAD and LUSC.(a).MMP1 (b).MMP2 (c).MMP3 (d).EGFR (e).MPO (f).MET (g).NQO1 (h).PTGS2

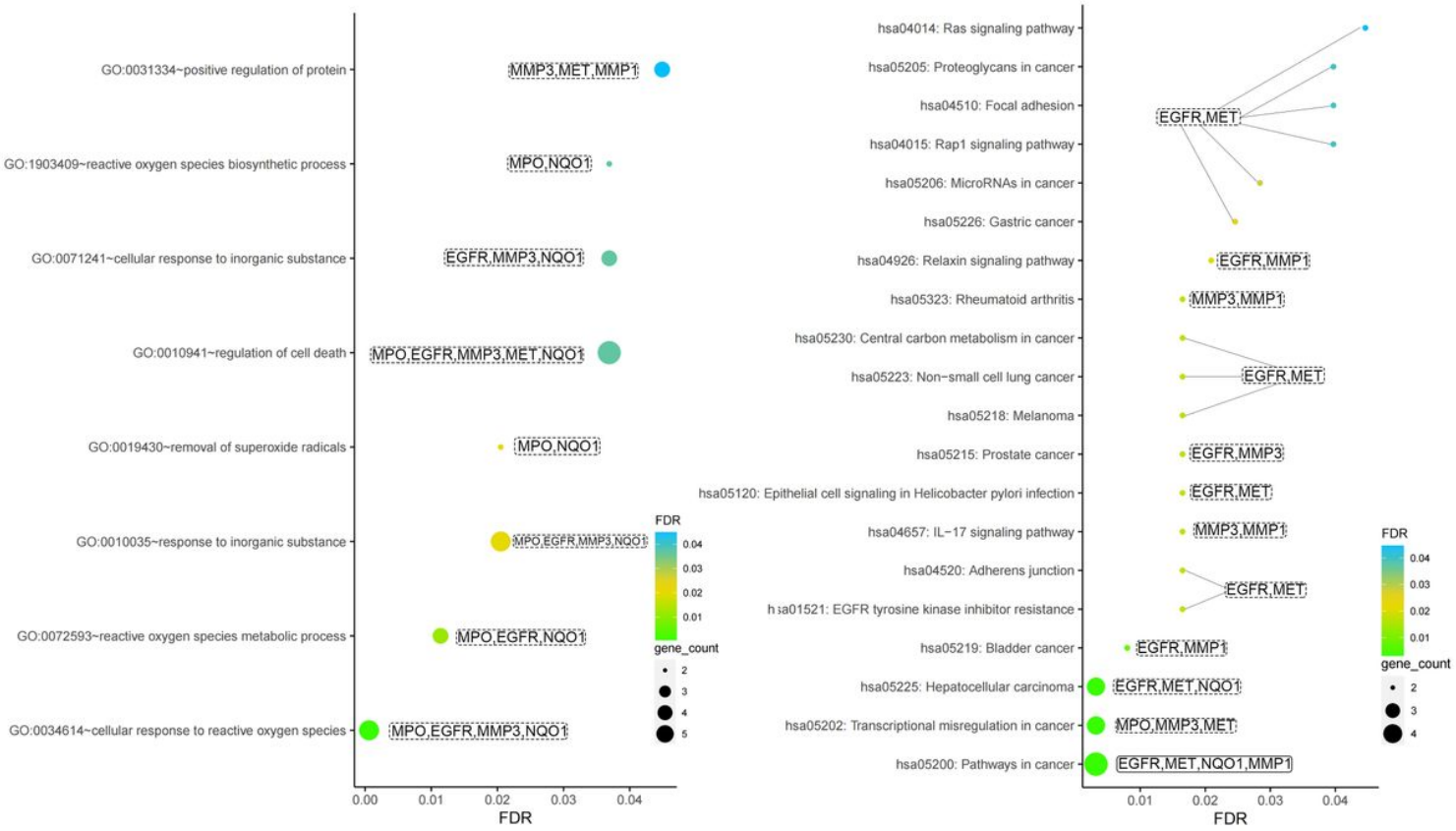

Figure 5

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

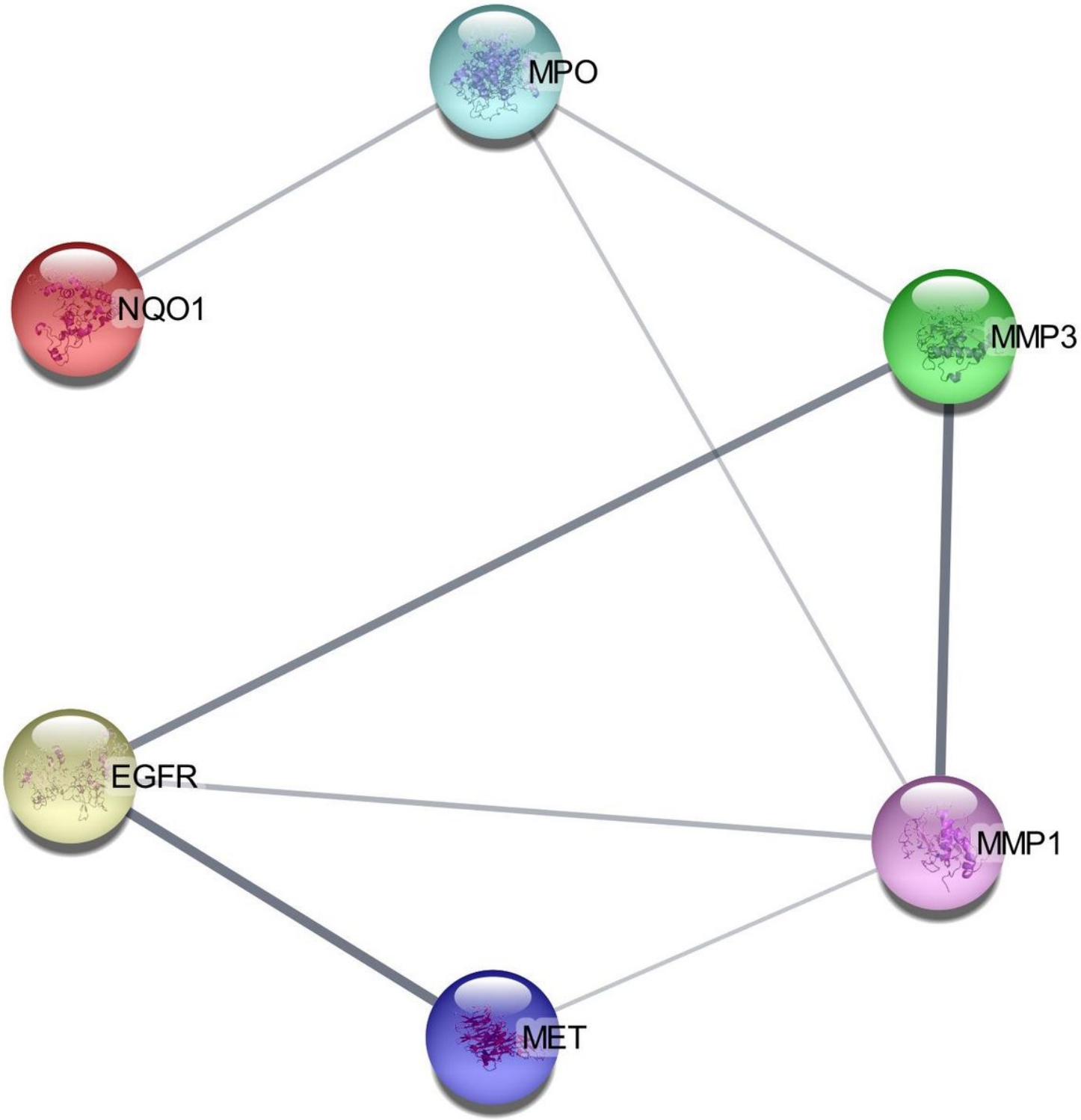

Figure 6

Eight protein interaction networks mediated by six compounds of SAIN, the thickness of the line represents the strength and weakness relationship between the proteins.

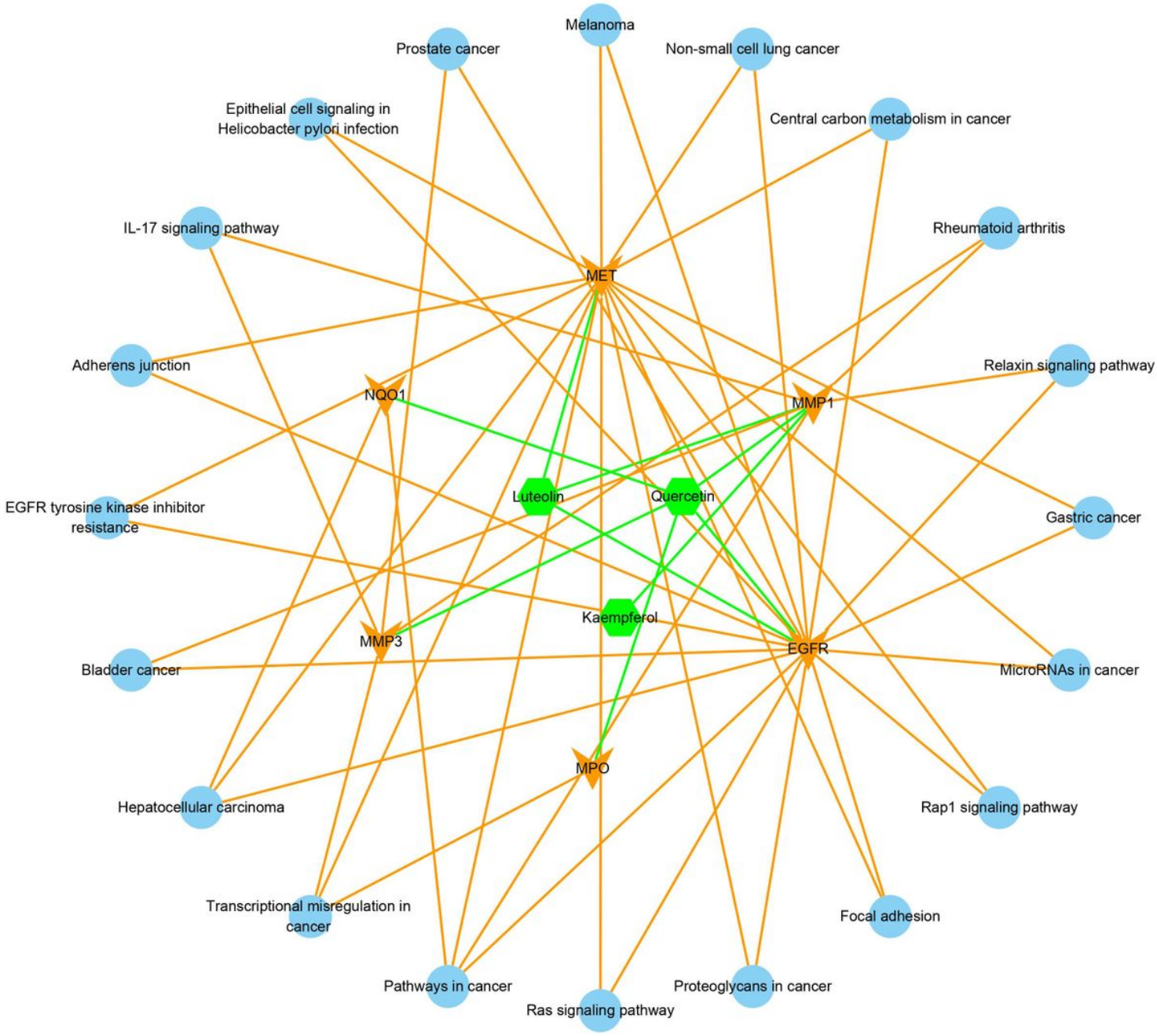

Figure 7

Compounds-targets-pathways network, and the green line represents the C-T relationship, and the orange line represents the T-P relationship

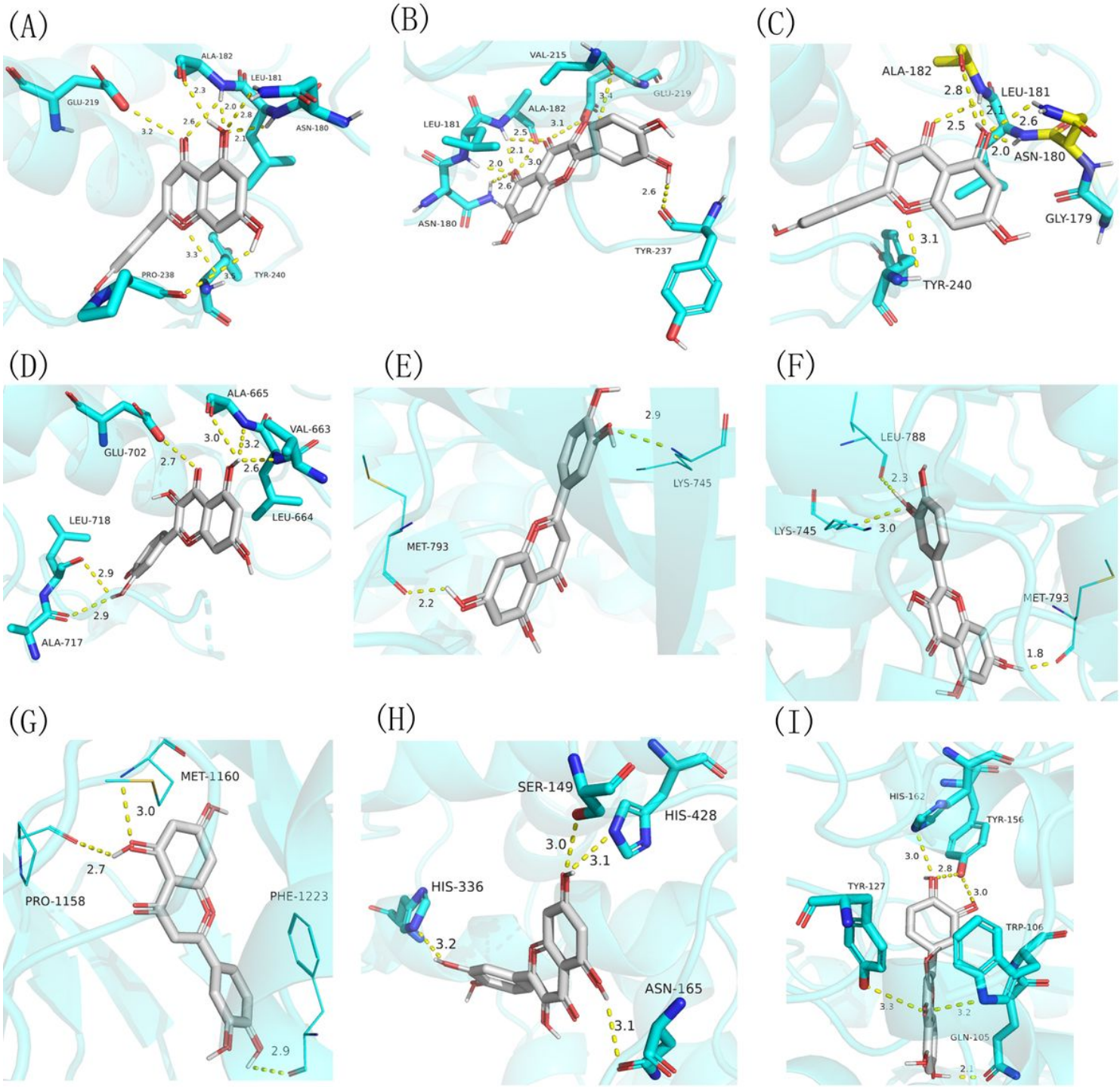

Figure 8

Molecular docking results of 6 compounds and 8 proteins, the connection represents a hydrogen bond. (a).L-MMP1 (b).Q-MMP1 (c).K-MMP1 (d).Q-MMP3 (e).L-EGFR (f).Q-EGFR (g).L-MET (h).Q-MPO (i).Q-NQO1

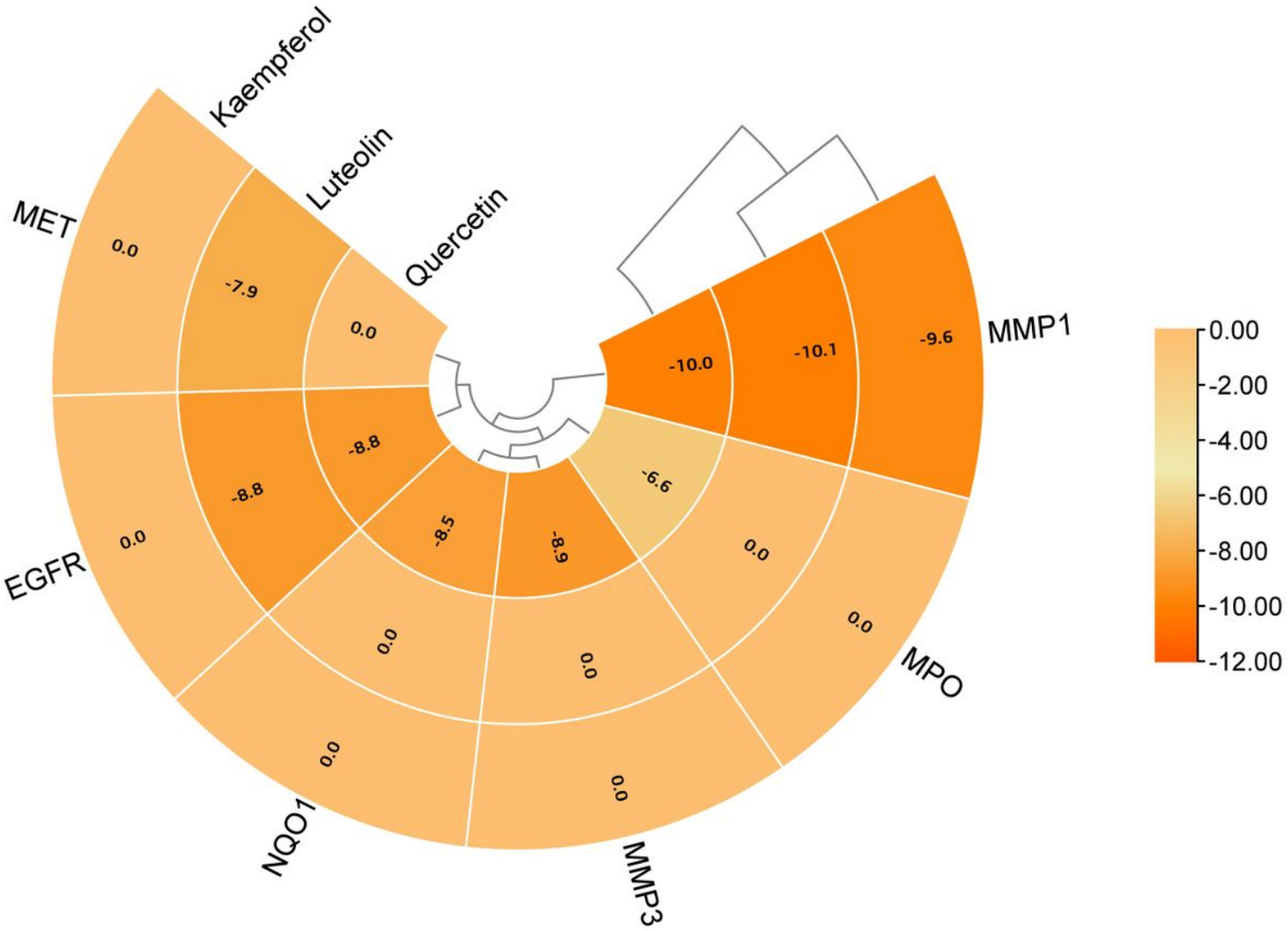

Figure 9

Visualization of heat map of molecular docking results, the affinity is used to represent the color range. The smaller the affinity is, the more significant the result is, color represents the value of affinity, the darker the color, the better the affinity

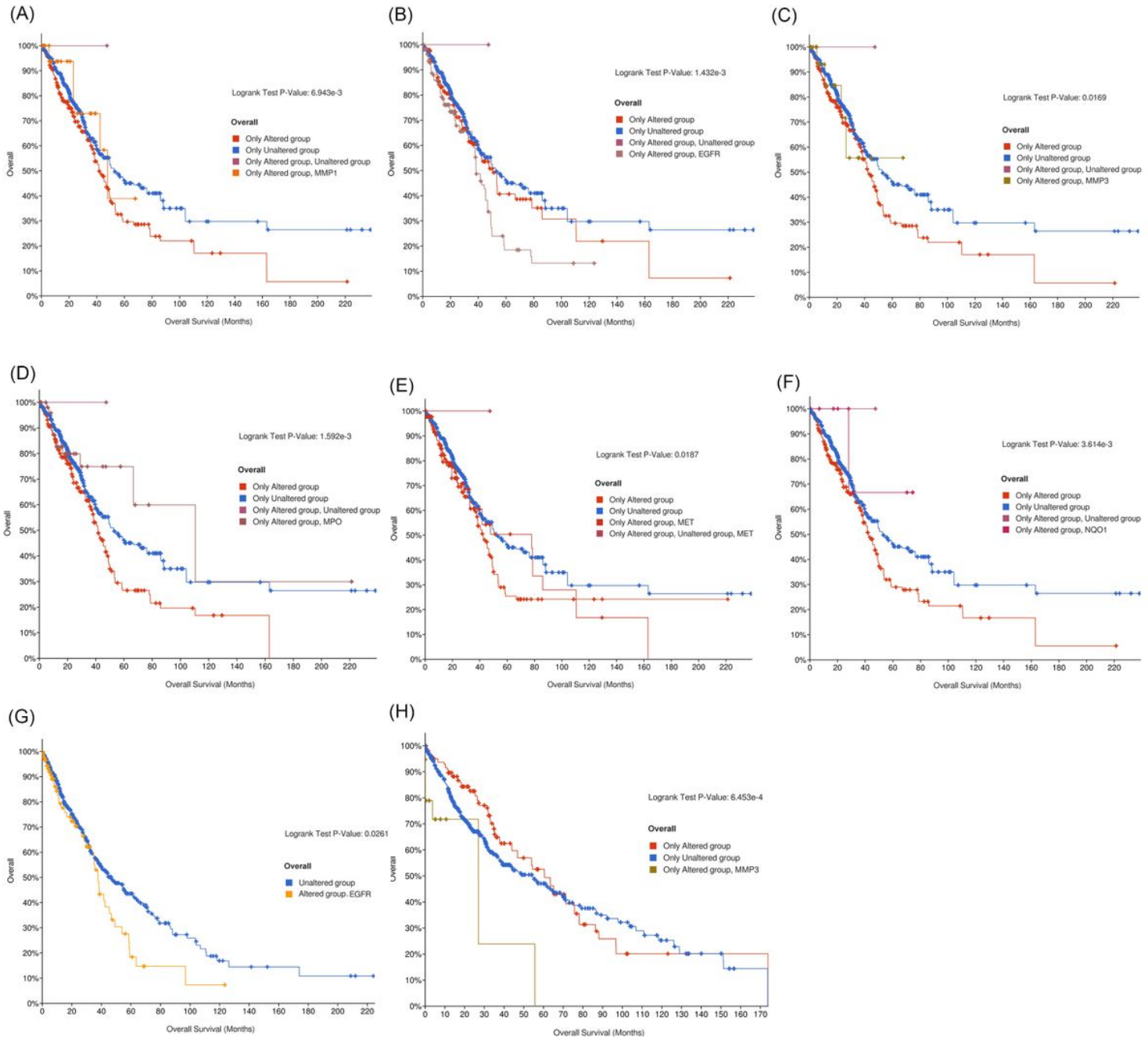

Figure 10

Results of prognostic survival analysis of six genes of LUAD and LUSC. A-F is the result of LUAD and g-h is the result of LUSC (a). Only altered MMP1 (b). Only altered EGFR (c). Only altered MMP3 (d). Only altered MPO (e). Only altered MET (f). Only altered NQO1 (g). Only altered EGFR (h). Only altered MMP3

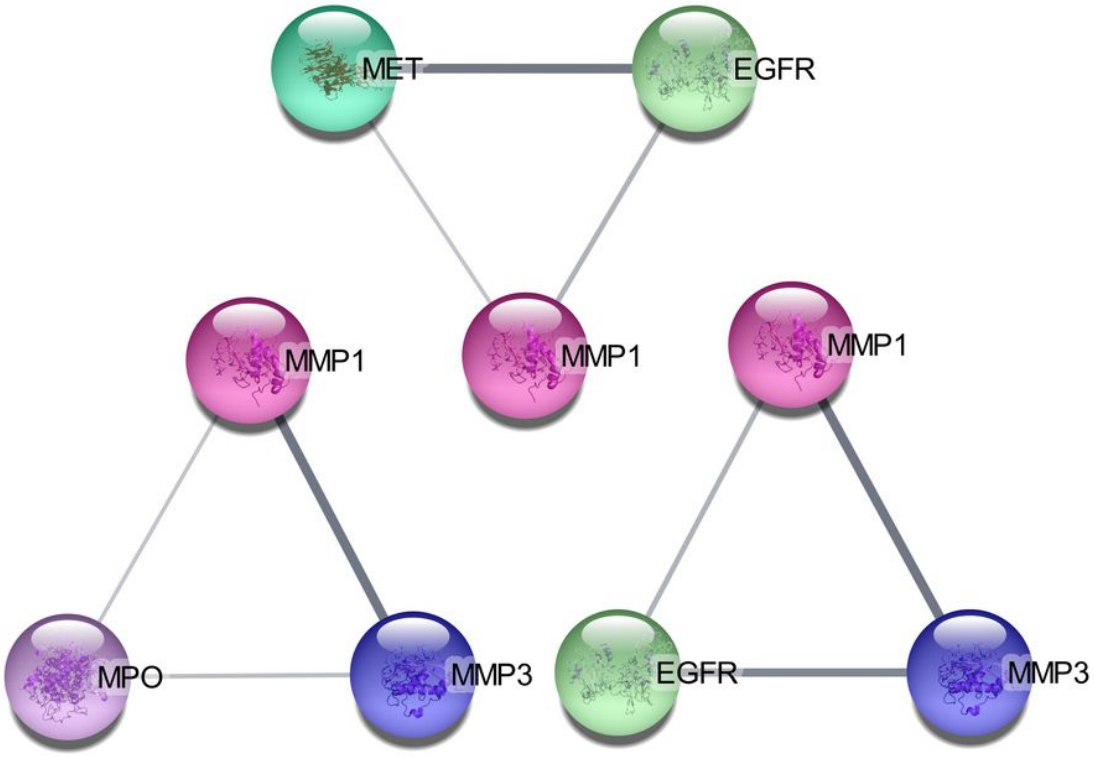

Figure 11

Interaction of MMP1 with EGFR, MMP3, NQO1 and MPO, the thickness of the line represents the strength and weakness relationship between the proteins.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Table1.xlsx
- Table2.xlsx
- Table3.xlsx
- Table4.xlsx
- Table5.xlsx