

STAT1: A novel candidate biomarker and potential therapeutic target of the recurrent aphthous stomatitis

Ming-Chen CAO

Affiliated Hospital of Qingdao University

Meng-Xiang Fang

Huangdao District Second Hospital of Traditional Chinese Medicine

Lei Li

Affiliated Hospital of Qingdao University

XING Xiao-Min

Affiliated Hospital of Qingdao University

Chang-Kai ZHOU

Affiliated Hospital of Qingdao University

Long-Yuan WANG

Affiliated Hospital of Qingdao University

Long Xu

Affiliated Hospital of Qingdao University

Wei Ren

Affiliated Hospital of Qingdao University

Fan-Bo JING (✉ 383642665@qq.com)

Affiliated Hospital of Qingdao University

Research Article

Keywords: recurrent aphthous stomatitis (RAS), diagnostic biomarkers, treatment, STAT1 protein, therapeutic targets

Posted Date: April 2nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-355780/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

STAT1: A novel candidate biomarker and potential therapeutic target of the recurrent aphthous stomatitis

Ming-Chen CAO 1, Meng-Xiang FANG 2, Lei LI 1, Xiao-Min XING 1, Chang-Kai ZHOU 1, Long-Yuan WANG, Long XU 1, Wei REN 1, Fan-Bo JING 1 (Corresponding Author) *

(1, The Affiliated Hospital of Qingdao University;
2, West Coast New District Second Hospital of Traditional Chinese Medicine)

Abstract: At present, the recurrent aphthous stomatitis (RAS) are not completely clear. Therefore, identifying the underlying diagnostic biomarkers of RAS can provide new ideas for the diagnosis and treatment of RAS. The gene chip data of RAS (GSE37265) were downloaded from the NCBI Gene Expression Omnibus (GEO) database. Weighted Gene Co-Expression Network Analysis (WGCNA) was used to construct a co-expression module. A total of 16 co-expression modules were identified, and 30 hub genes in the turquoise module were identified. In addition, functional analysis of hub genes in modules of interest was performed, which indicated that such hub genes were mainly involved in pathways related to immune response, virus infection, epithelial cell, signal transduction. module two clusters (highly interconnected regions) were determined in the network, with score=17.647 and 10, respectively, cluster 1 and cluster 2 are linked by STAT1 and ICAM1, it is speculated that STAT1 may be a primary gene of RAS. Finally, genistein, daidzein, kaempferol, resveratrol, rosmarinic acid, triptolide, quercetin and (-)-epigallocatechin-3-gallate were selected from the TCMSP database, and both of them is the STAT-1 inhibitor. The results of reverse molecular docking suggest that in addition to triptolide, (-)-Epigallocatechin-3-gallate and resveratrol, the other 5 compounds (flavonoids) with similar structures may bind to the same position of STAT1 protein with different docking score. In conclusion, our results screened potential biomarkers that might contribute to the diagnosis and treatment of RAS, STAT1 protein is one of the potential therapeutic targets of RAS, and this target can be used to screen potential compounds.

Introduction

Recurrent aphthous stomatitis (RAS) is recognized as the most common oral mucosal disease [1]. RAS is a painful (include prodromal burning sensation), well-circumscribed, and round-shaped ulceration that is covered by a yellow-grayish pseudomembrane and surrounded by an erythematous halo, or become confluent to produce larger, plaqueform, irregular lesions throughout the oral cavity [2-3]. The classic presentation of RAS is recurrent, self-limiting ulcers that mainly affect nonkeratinized oral mucosa (typically located on the buccal, labial mucosa, tongue and floor of the mouth). Involvement of the heavily keratinized mucosa of the palate and gingiva is less common [4]. Since oral disorders frequently have detrimental effects on speech, nutrition, physical appearance, self-esteem and social interaction, especially RAS frequently affects patient quality of life as a result of long lasting and recurrent episodes of burning pain [5].

Although the molecular mechanism of RAS pathogenesis is not yet clear, it may involve biological processes such as immune response, chronic inflammation, oxidative stress, extracellular matrix, et al [5-7]. Therefore, elucidating the molecular mechanism of RAS pathogenesis and exploring the potential molecular targets of medical therapies for RAS is of vital importance. With the widespread application of gene chips and high-throughput sequencing technologies, databases

related to genomes have accumulated a large amount of data [8]. Computerization methodologies have been applied into the discovery of signature genes as potential biomarkers of diseases [9]. How to use bioinformatics technology to deeply explore the potential value of these data has become one of the important directions for studying the molecular mechanisms of diseases. The bioinformatics analysis methods can help us study the molecular mechanism of diseases and discover potential therapeutic targets from a systematic perspective [8-9]. Among all the bioinformatics analysis methods, the weighted gene co-expression network analysis (WGCNA) is a useful advanced and comprehensive algorithm approach for the analysis of the gene expression patterns of multiple samples [10]. The unique advantage of WGCNA is the ability to analyze gene expression profiling to cluster genes and form co-expression modules by similarly behaving genes with a common biological relationship or function, that reveal the gene networks and signaling pathways and identify intramodular hub genes [11]. It has been successfully used to study various biological processes, proving to be quite helpful for the identification of candidate biomarkers and potential therapeutic targets [11-12].

Management of RAS depends upon the frequency and severity of the lesions [5]. Most RAS cases can be adequately managed with topical therapy, the current treatment methods include pain relief, anti-inflammatory, and promotion of ulcer healing, while mainly include antibiotic therapy, hormonal therapy, medicine mouthwash, and laser therapy [5-7]. However, there are temporarily few available effective medical therapies to treat RAS currently. Traditional Chinese medicine has accumulated many natural medicines for the treatment of diseases, molecular biology and drug molecular target identification techniques have been more and more widely used in current Chinese herbal medicine research [13]. Drug target identification, which includes many distinct algorithms for finding genes and proteins, is the first step in drug discovery, the problem of target identification is usually finding the best interaction mode between the potential target candidates and probe small molecules [14]. Many computer simulation analysis technologies have been developed for the confirmation of lead compounds, such as structure-based target discovery methods (such as pharmacophores, similar binding sites, fingerprint-based interaction methods, and molecular docking), representative databases such as TCMSP, Pharmmapper and others, calculate and save a large number of target data of natural active chemical components [13-15].

In this study, we used a variety of bioinformatics analysis tools to conduct in-depth data mining on the gene chip data of 28 RAS patient samples, and finally determined that STAT1 may be a key target affecting the RAS process and a potential therapeutic target at the same time, based on this target, the natural chemical components of 8 herbs were screened, which may become potential drugs for local treatment of RAS, providing new directions for follow-up research.

MATERIALS & METHODS

Data Collection and Validation of the datasets

The gene expression dataset used for our analysis was screened from the Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo/>), "recurrent aphthous stomatitis" was used as the search keyword. A dataset, with a GEO tracking number GSE37265 and a platform entry number GPL570 provided by Baccaglini L et al., was screened out and download. Sample collection and microarray dataset were performed by the Microarray lab 103, Molecular

Genetics and Microbiology, University of Florida. In this dataset, transcription profiles were established from normal tissue from control individuals and ulcer and non-ulcer tissue from afflicted individuals. The transcriptional profiles were measured by Affymetrix Human Genome U133 Plus 2.0 Array.

Differential expression genes (DEGs) analysis

The matrix file was annotated with an official gene symbol using the data table of the microarray platform, the “sva” R package was used to conduct batch normalization of the original expression data, and a normalized gene expression matrix file containing data was obtained for DEGs analysis. The “limma” R package was used to conduct DEGs analysis. The P-value of genes was calculated using t test method, and Benjamini and Hochberg's method was used to calculate the adjusted P-value.

Construct the miRNA-gene interaction network

MicroRNA (miRNA) are identified to play a key role in regulating development in mammalian organisms. The analysis of miRNA and protein coding genes were studied based on the TarBase (<http://www.microrna.gr/tarbase>), the largest available manually curated target database, indexed targets derived from high throughput experiments, provides millions of high quality manually curated experimentally validated miRNA-gene interactions[16].

Construct the signaling information network

The regulation of gene and protein expression in organisms is inseparable from the extensive participation of chemicals such as signaling molecule, the signaling information network was constructed based on the SIGNOR 2.0 (<https://signor.uniroma2.it/>), a public repository that stores manually-annotated causal relationships between proteins and other biologically relevant entities (chemicals, phenotypes, complexes, etc.) that participate in signal transduction relationship, represented graphically as a signed directed graph[17].

Weighted gene Co-Expression network analysis and co-expression network construction

The weighted gene correlation network analysis was performed to construct a co-expression network via R (3.6.2) WGCNA package, a typical system biology algorithm. First, we performed cluster analysis of the samples to detect the outliers by the hclust function [10-12].

Gene Set Enrichment Analysis (GSEA) of gene modules

The constructed modules were consisted of a number of genes and functional enrichment analysis was then performed on the DEGs in those modules. To obtain the biological functions and signaling pathways involved in those modules, DEGs in modules were subjected to gene ontology (GO) analysis and Kyoto encyclopedia of genes genomes (KEGG) pathway analysis using the GSEA software (GSEA version 4.0.3) [18]. After multiple test calibration, we used “adjusted P < 0.002” and “FDR <0.05” as the threshold value to identify the enriched terms, and the top 10 most important terms were screened.

PPI network

A PPI network was constructed to evaluate the interactions between genes, which helps us to explore novel molecular mechanism. Modules of interest were visualized using STRING 11.0

(<https://string-db.org/>), an online database to search the interaction among different proteins [19]. The common genes in the preserved modules that were obtained from WGCNA, and the DEGs with significant consistency, were selected to construct a PPI network, visualized using Cytoscape 3.7.2 software [20]. In the PPI network, a node represents a gene; the undirected link between two nodes is an edge, denoting the interaction between two genes; and the degree of a node corresponds to the number of interactions of a gene with other genes in the network, and only experimentally validated interactions with a combined score of more than 0.9 were selected as significant. Using node degree and interaction score as the key topological parameters, the maximally connected genes were informally referred to as hub genes.

Identification and validation of hub genes

The intra-module connectivity of a gene is equal to the sum of the degree of correlation between this gene and other genes in that module. The top 30 genes with the highest intra-module connectivity were selected as hub genes. After screening out the interested modules, the weighted gene co-expression network was constructed using Cytoscape, and the hub genes were identified by the Molecular Complex Detection (MCODE) plugin. Gene regulatory network could help us accurately screen candidate genes that were potentially involved in the regulation of target genes, and could use the function of known genes to predict unknown gene function.

Screening of active ingredients of natural medicines acting on hub genes

TCMSP is a pharmacology platform of Chinese herbal medicines that focus on the exploration of the active ingredients and targets, which had collected 499 herbs, with a total of 12144 chemicals, as well as pharmacokinetic properties for natural compounds [21]. The drug-target were obtained from two sources: (1) experimental validated drug-target pairs were retrieved from HIT database (2) the SysDT model constructed was used to predict the potential targets of a compound [21]. In order to obtain the related ingredients based on the TCMSP database, we selected the search category as "targets name" and the keyword setting as "signal transducer and activator of transcription 1-alpha/beta " to search, $DL \geq 0.1$ as the filter condition.

Predict potential targets

PharmMapper[22] is designed to identify potential target candidates for the given probe small molecules using pharmacophore mapping approach. Upload Query File: .Mol2, parameter set: Generate Conformers: Yes; Maximum Generated Conformations: 300; Select Targets Set: Druggable Pharmacophore Models (v2017, 16159); Number of Reserved Matched Targets (Max 1,000): 500. After submitting and waiting for the calculation to be completed, the results are saved in csv file format.

Reverse molecular docking verification

Molecular docking was performed by AutoDock Vina [23]. All visualizations of biomolecules were conducted by PyMol Software [24].

Results

Validation of the datasets

We normalized the raw data of GSE37265 before analysis, the box plot showing distribution of raw read counts in the dataset (Figure 1A). To further validate the intra-group data repeatability, we employed the Pearson's correlation test and principal component analysis (PCA). The color reflects the intensity of the correlation, when $0 < \text{correlation} < 1$, there exists a positive correlation. When $-1 < \text{correlation} < 0$, there exists a negative correlation, the larger the absolute value of a number the stronger the correlation, which showed that there were strong correlations among the samples in the health group and RAS group in the GSE37265 dataset (Figure 1A). Based on the PCA, the intra-group data repeatability for GSE37265 dataset was acceptable. In the PCA diagram, principal component 1 (PC1) and principal component 2 (PC2) are used as the X-axis and Y-axis, respectively, to draw the scatter diagram, where each point represents a sample, the farther the two samples are from each other, the greater the difference is between the two samples in gene expression patterns. The distances between per samples in the control group and the recurrent aphthous stomatitis group were acceptable in the dimension of principal component-1 (PC1) (Figure 1B). The diagnostic plot summarizing the standard deviation versus mean measures of reads in the samples for each gene, which showed the dependence between counts and variance was acceptable. The plot of density against log2 of read counts displays the relative distribution of different counts in the health group and RAS group.

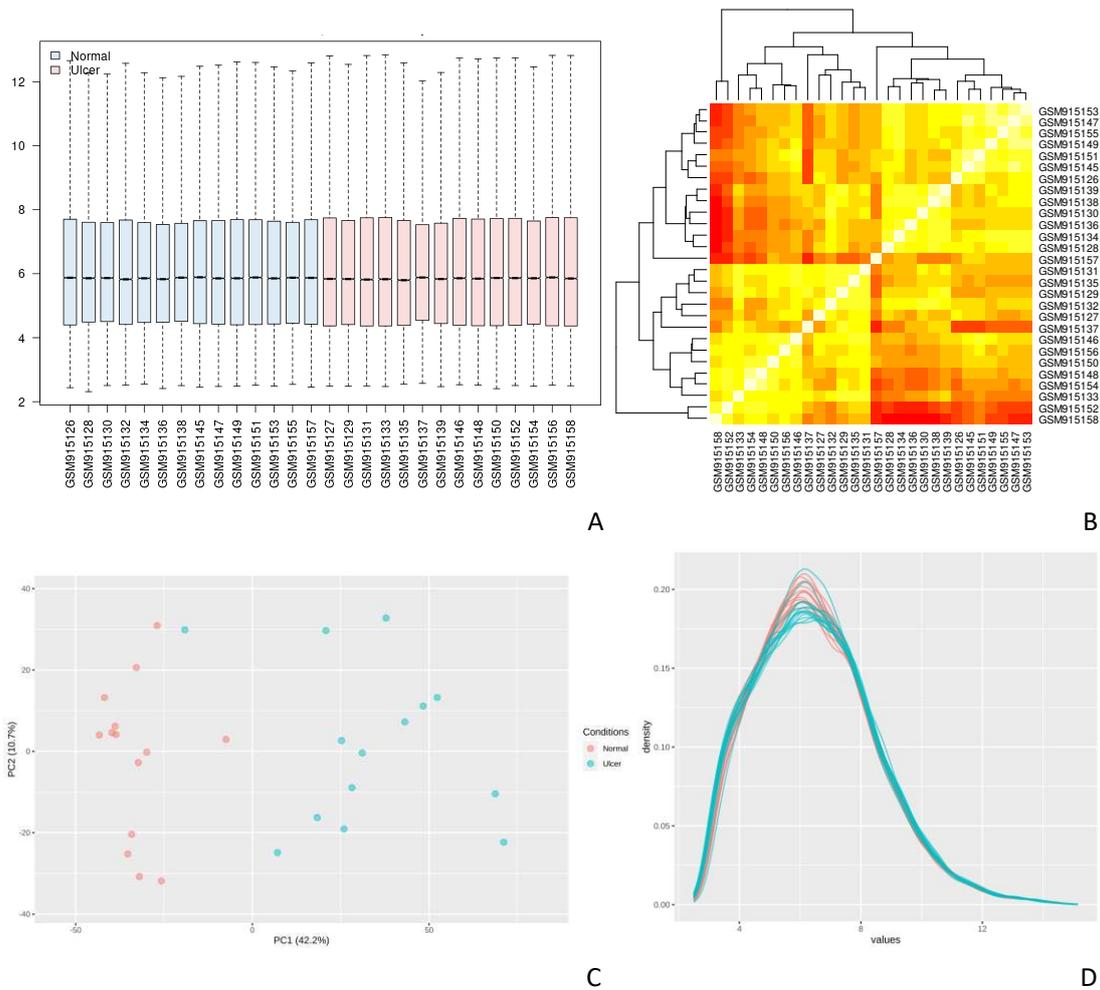


Figure 1. (A) box plot showing distribution of raw read counts in the GSE37265 dataset (B) Pearson's correlation analysis of samples from the GSE37265 dataset. (C) PCA of samples from the GSE37265 dataset. (D) The diagnostic plot

Differentially expressed genes (DEGS) between RAS and healthy control

The recurrent aphthous stomatitis samples from cohort GSE37265 were analyzed using R software and its extension packages. The gene expression matrix was obtained after data preprocessing (included 12,548 genes). A total of 187 DEGs were identified with the threshold at $|\log_2(\text{fold-change})| > 2$ and $p < 0.05$, which consisted of 125 down-regulated genes and 18 up-regulated genes, and the volcano plot of all probesets is shown in Figure 2. The 50 most significant down-regulated genes and up-regulated genes were visualized using a heatmap (Figure 3). Red represents increased expression, whereas blue represents decreased expression. The most up-regulated genes included DAPL1, TSPAN8, ELOVL4, KRT31, WNK4, CTTNBP2, CALB2, GYS2, ETNK2, KRTAP3-2, whereas MMP1, CXCL11, MMP3, DEFB4A, CXCL10, CXCL9, CXCL1, KRT24, CXCL6, S100A7, MMP10, CXCL8, SLC6A14, CCL8, MMP12 were the most down-regulated genes in the RAS samples.

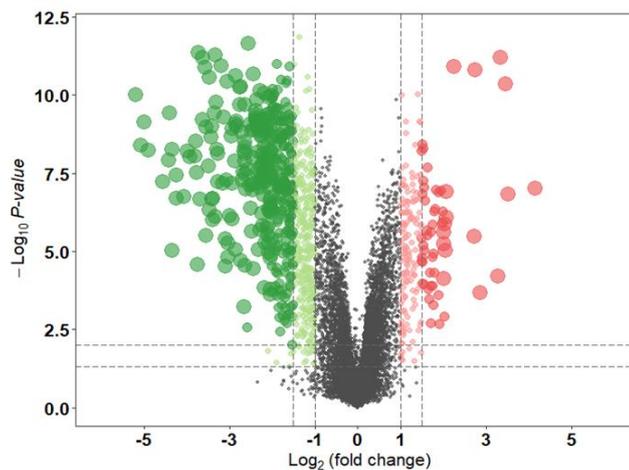


FIGURE 2 The volcano plot shows the up-regulated and down-regulated genes in RAS. The horizontal axis represents the fold change between health and RAS. The vertical axis represents the P value of t test for the differences between health and RAS.

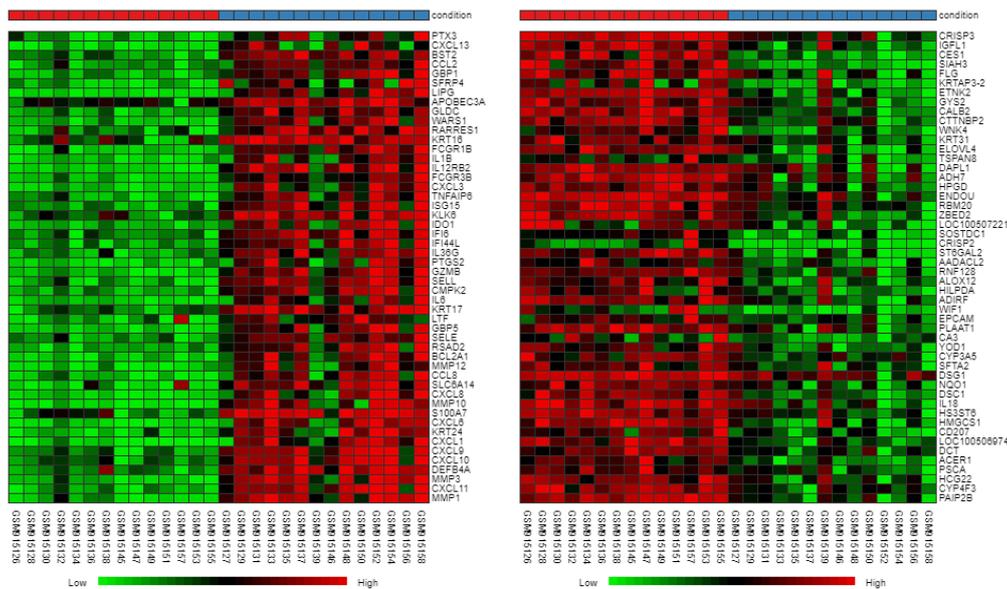


FIGURE 3 The heatmap shows the top 50 differential gene expression ($P < 0.05$) between health and RAS.

Construct miRNA-gene interaction network

As shown in the Figure 4, we constructed a miRNA-gene interaction network of DEGs based on TarBase, Table 1 lists the top 20 high-level genes according to their interaction degrees, which reveals that SOD2, SLC2A3, PXDN, PTGS2, IRF1, COL4A1, MICB, CXCL8, etc. may play an important role in the miRNA regulatory network.

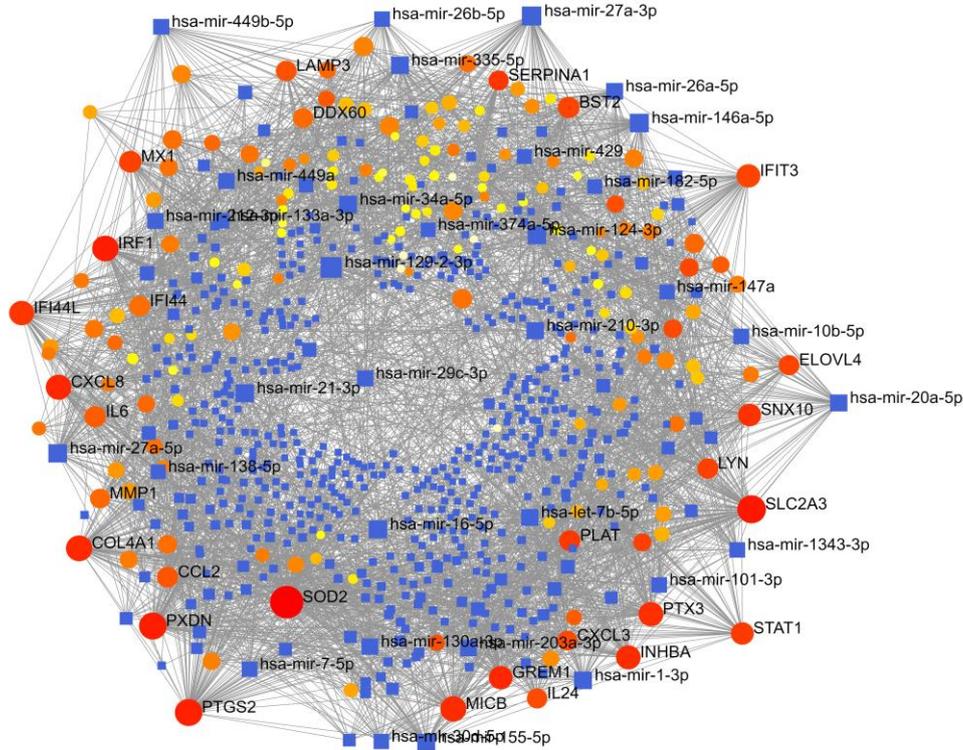


Figure 4 The miRNA-gene interaction network of RAS

Table 1 The top 20 genes of miRNA-gene interaction network of RAS

Id	Label	Degree	Betweenness	Expression
6648	SOD2	156	53322.71	-2.45583
6515	SLC2A3	95	23070.28	-2.24469
7837	PXDN	85	19402.04	-2.22854
5743	PTGS2	85	17813.26	-3.55500
3659	IRF1	73	18693.35	-2.20671
1282	COL4A1	73	13650.85	-2.32325
4277	MICB	70	12782.87	-2.23157
3576	CXCL8	67	13077.81	-4.33586
10964	IFI44L	62	8806.11	-3.49043
5806	PTX3	61	11559.62	-3.03014
3624	INHBA	59	12948.04	-2.57143
3437	IFIT3	54	5974.13	-2.93725
26585	GREM1	53	13463.77	-2.47693
29887	SNX10	49	10198.16	-2.62950
6772	STAT1	48	7937.42	-2.06146
4599	MX1	44	6582.67	-2.03886

684	BST2	41	5885.18	-3.04879
3569	IL6	40	2793.91	-3.67836
5327	PLAT	38	8060.72	-2.85614
6347	CCL2	38	3233.80	-3.07236

Construct the signaling information network based on SIGNOR

As shown in the Figure 5, we constructed the signaling information network of DEGs based on SIGNOR2.0, Table 2 lists the top 20 high-level genes according to their interaction degrees, which reveals that STAT1, IL6, LYN, PTGS2, IL1B, IFNG, HCK, CXCL8, CCL2, CXCR4, etc. participated extensively in the regulation of chemical signaling substances in this network.

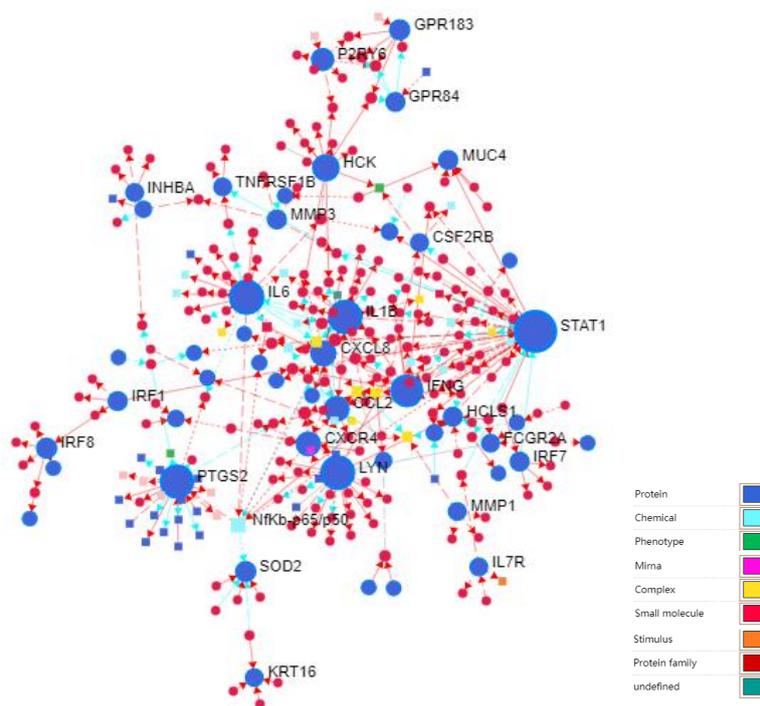


Figure 5 The signaling information network based on SIGNOR

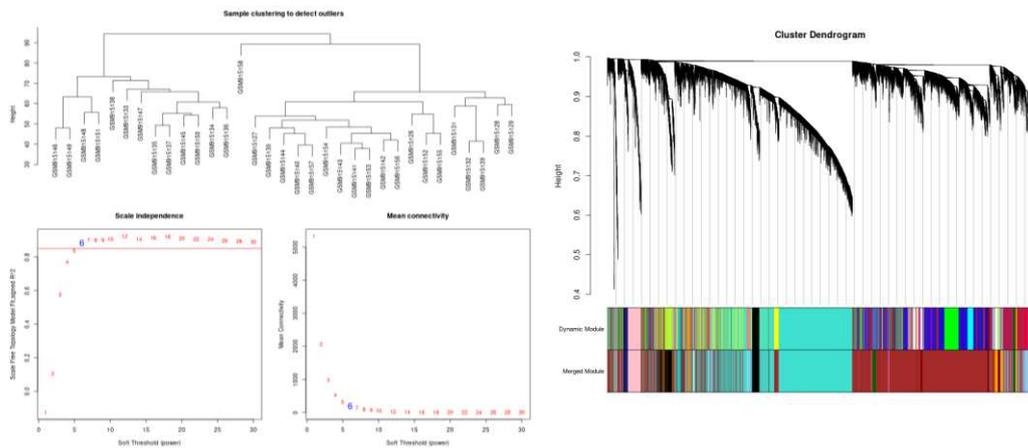
Table 2 The top 20 genes of signaling information network of RAS

Id	Label	Degree	Betweenness	Expression
6772	STAT1	48	22698.35	-2.06146
3569	IL6	27	8361.34	-3.67836
4067	LYN	26	8816.84	-2.37319
5743	PTGS2	25	7101.68	-3.55500
3553	IL1B	25	5464.20	-3.32514
3458	IFNG	22	5069.17	-2.72686
3055	HCK	13	6563.89	-2.36321
3576	CXCL8	11	3444.47	-4.33586
6347	CCL2	11	1569.33	-3.07236
7852	CXCR4	10	3602.66	-2.03314
5031	P2RY6	8	1444.38	-2.05379
6648	SOD2	6	2492.00	-2.45583
3394	IRF8	6	1676.00	-2.35421

1880	GPR183	6	743.57	-2.23414
3659	IRF1	5	2754.00	-2.20671
4314	MMP3	5	1947.00	-6.14321
3059	HCLS1	5	1258.00	-2.28786
4585	MUC4	5	1204.32	-2.07303
3665	IRF7	5	843.00	-2.11757
53831	GPR84	5	602.76	-2.15514

WGCNA Co-Expression Network and Construction of coexpression modules

We performed network topology analysis to determine candidate power values for relative, balanced scale independence, and mean connectivity in the WGCNA. As a result, the 6208 DEGs (adjust P values <0.05) of the RAS samples were used to construct co-expression modules using the WGCNA algorithms. Subsequently, hierarchical clustering analysis was performed with the flashClust function and the results are presented in Figure 6A. The soft-power threshold β was determined by the function “sft\$powerEstimate”, as shown in Figure 6B, a power value of 6 was the lowest power for which scale independence was below 0.8, and this was selected to produce a hierarchical clustering tree of the 6208 genes. Finally, 16 modules were identified based on average hierarchical clustering and dynamic tree clipping, each module had different color and genes. All the modules were significantly independent of each other, eigengene module values were calculated in each module and a clustering tree is presented in Figure 6C. Among all the modules, the turquoise module had the highest number of hub genes. Then, gene modules were detected based on the TOM matrix, Interactions between the 16 modules were then analyzed (Figure 6D). In addition, the eigengene dendrogram and heatmap were used to quantify module similarity by eigengene correlation (Figure 6E).



A

B

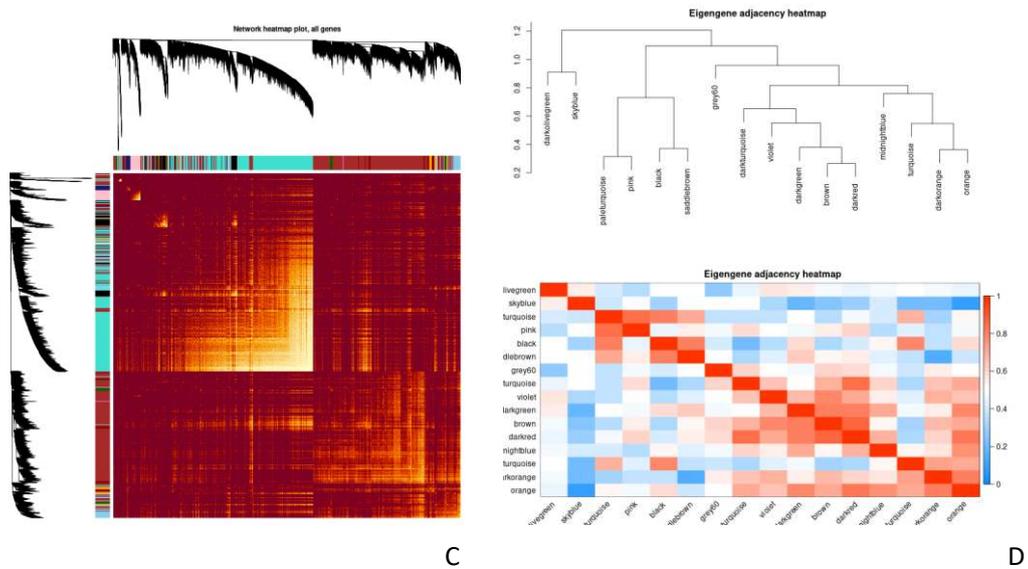


Figure 6 The result of WGCNA analysis

A: Sample clustering to detect outliers; B: Analysis of network topology for a set of soft-thresholding powers. Scale independence and mean connectivity of various soft-thresholding values (β). The left picture displays the scale free fit index (y-axis) as a function of the soft-thresholding power (x-axis). The right picture shows the mean connectivity (degree, y-axis) as a function of the soft- thresholding power (x-axis); C: Clustering dendrograms of genes, with dissimilarity based on topological overlap, together with assigned module colors. Cluster dendrogram of all filtered genes enriched based on the dissimilarity measure and the cluster module colors; D: The heatmap plot describes the Topological Overlap Matrix (TOM) among DEGs in the analysis; E: The eigengene dendrogram and heatmap identify groups of correlated eigengenes termed meta modules.

Define the DEGs in co-expression modules

Interestingly, we found that almost all of the differentially expressed genes, especially the key nodes in the above two networks analysis are involved in the Turquoise module, including SOD2, STAT1, PTGS2, IL6, etc. Therefore, we use the R software to obtain the DEGs ($\log_2(\text{fold-change}) > 1.5$ and $p < 0.05$) of the Turquoise module, there are 254 DEGs among the 677 co-expressed genes in the Turquoise module.

GSEA enrich analysis

The Figure 7 showed the result of GSEA enrichment analysis based on Go (biological process). As shown in Figure 8, The pathways for the DEGs in the Turquoise modules mainly focus on immune response, virus infection, epithelial cell, signal transduction, which the pathways that are highly related to RAS mainly include positive regulation of GTPase activity, T cell activation involved in immune response, epithelial cell differentiation, positive regulation of organelle organization, cell substrate adhesion, regulation of defense response to virus by host, regulation of calcium mediated signaling, interleukin 1 production ,etc. A total of 56 core targets were enriched, including ICAM1, CCR7, IL1B, PLEK, CCL4, NCKAP1L, GPR65, ZC3H12A, P2RY6, CCL8, RGS1, CCL2, ARHGAP9, ADAP2, RGS18, ITGAL, LCP1, LILRB1, STAT1, MSN, CORO1A participates in more than 2 GO pathways, and the ICAM1 with the highest frequency which participates in 5 pathways.

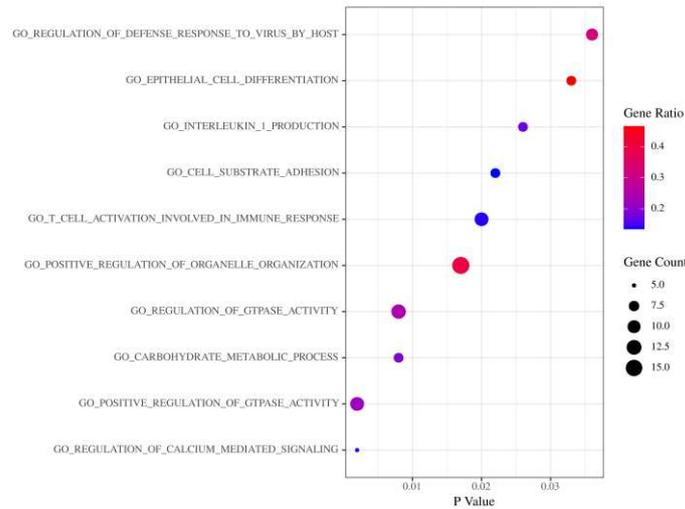


Figure 7 The result of GSEA enrichment analysis based on Go (biological process)

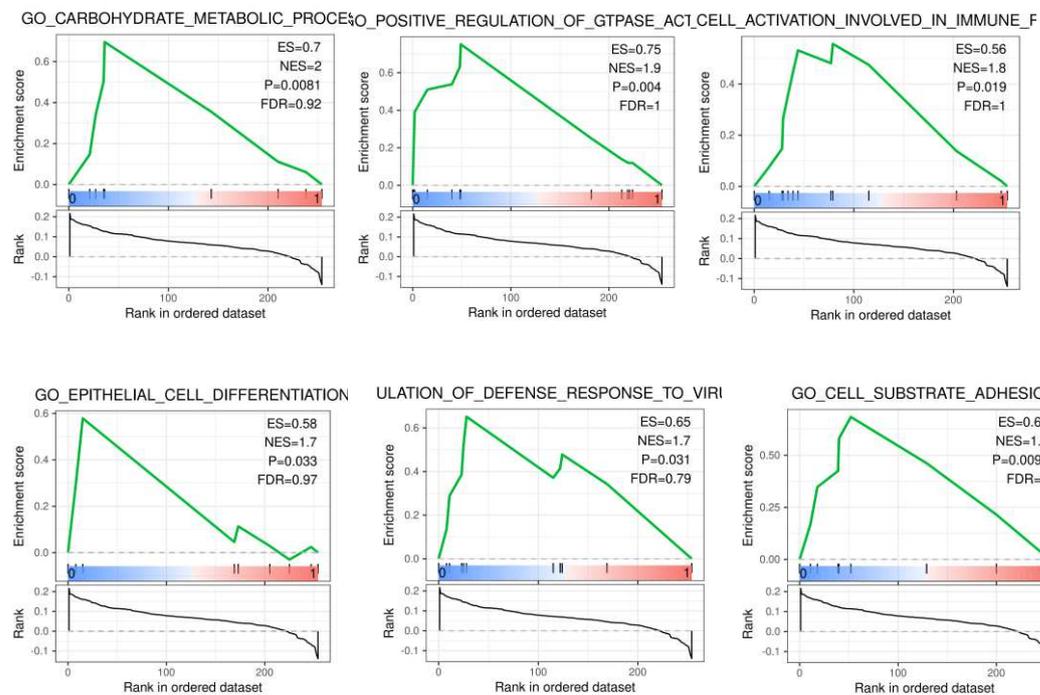


Figure 8 The Go (biological process) pathways highly related to RAS

The Figure9 showed the result of GSEA enrichment analysis based on based on KEGG. The pathways for the DEGs in the Turquoise modules that are highly related to recurrent RAS mainly include cell adhesion molecules cams, cytosolic DNA sensing pathway, natural killer cell mediated cytotoxicity, type I diabetes mellitus, nod like receptor signaling pathway, hematopoietic cell lineage, graft versus host disease, leukocyte transendothelial migration, JAK/STAT signaling pathway, TOLL like receptor signaling pathway, etc. A total of 55 key genes were enriched, including IL1B, IL6, CD86, HLA-B, ICAM1, ITGAL, HLA-DMA, HLA-F, HLA-DMB, GZMB, CD2, CCL4, CXCL10, IRF7, IL18, CXCL8, IL7R, CD14, CSF3R, STAT1 participates in more than 2 pathways, and the IL1B and IL6 target with the highest frequency participates in 6 pathways.

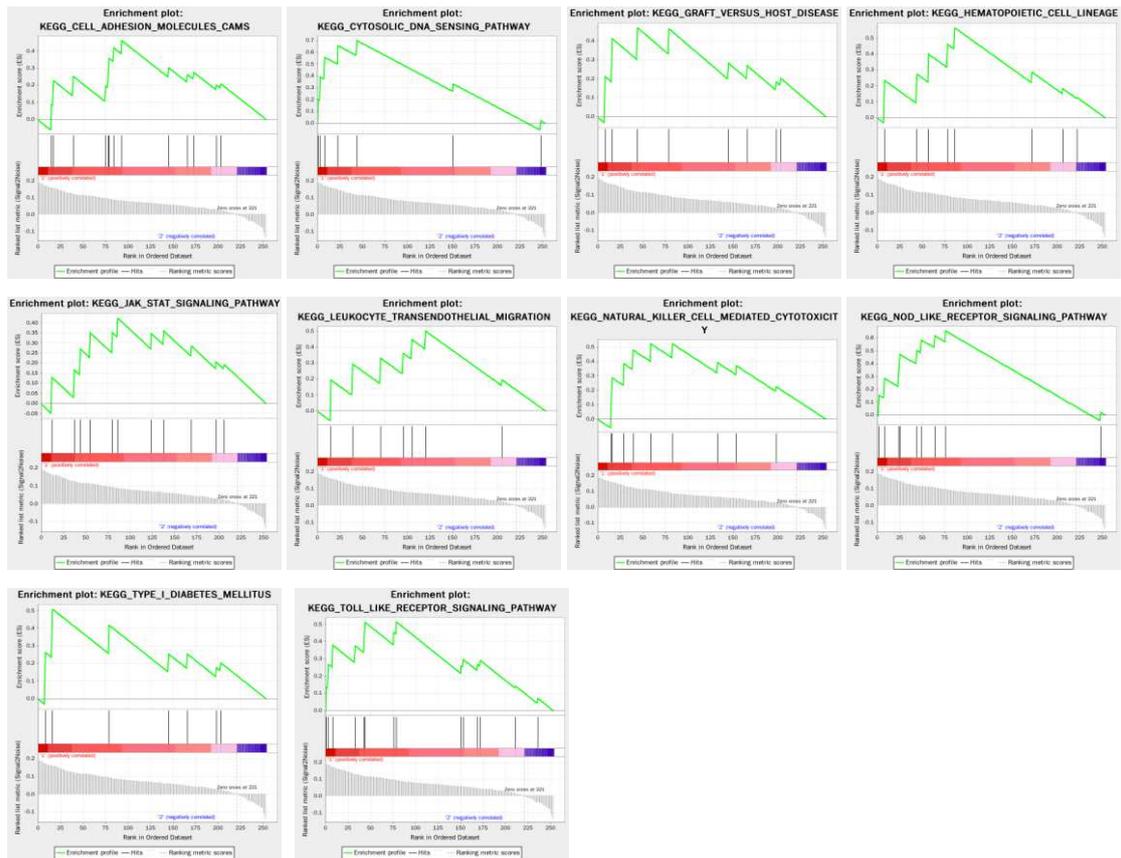


Figure 9 The result of GSEA enrichment analysis based on KEGG

Identification and Validation of Hub Genes

After merging the DEGs involved in the relevant pathways of the Turquoise module through the GSEA enrichment analysis, we constructed the protein-protein interaction (PPI) network of enriched DEGs based on the String database (Figure 10). Table 3 lists the network parameters of top 20 DEGs, such as HLA-B, IRF7, HLA-F, IFIT3, OASL, CXCL8, OAS2, MX1, ISG15, RSAD2, IRF1, etc.

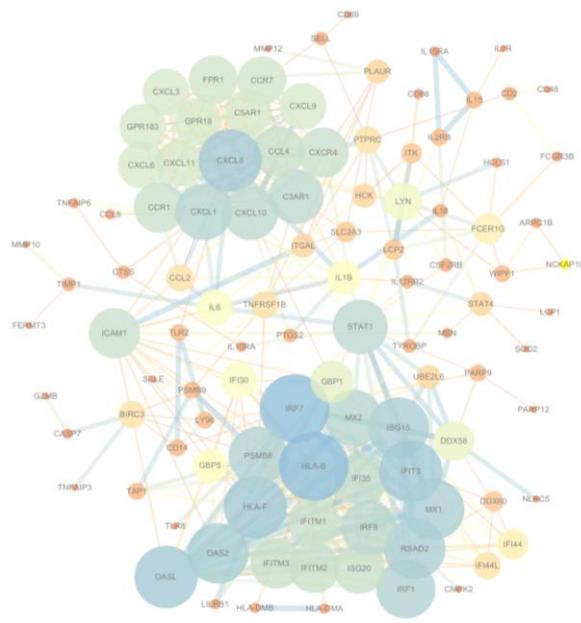


Figure 10 The protein-protein interaction (PPI) network of enriched DEGs

Table 3 The parameters of node in the PPI network

node	Degree	DMNC	EPC	Closeness	Betweenness	Stress	Clustering Coefficient
HLA-B	26	0.80384	51.600	54.68333	960.1928	4462	0.51385
IRF7	26	0.77479	51.902	56.48333	909.1212	5070	0.49538
HLA-F	23	0.80384	51.638	52.41667	222.4021	2268	0.65613
IFIT3	23	0.81353	51.946	53.65000	316.7897	2992	0.66403
OASL	23	0.76995	51.767	52.58333	155.9646	1834	0.62846
CXCL8	23	0.59562	49.514	54.95000	1155.311	4654	0.48617
OAS2	22	0.79905	51.740	51.91667	124.6154	1626	0.66234
MX1	22	0.83561	51.744	53.15000	229.5482	2374	0.69264
ISG15	22	0.83561	51.509	53.15000	229.5482	2374	0.69264
RSAD2	22	0.86481	51.910	48.93333	280.9047	2708	0.66234
IRF1	21	0.85350	51.580	53.48333	231.3252	2974	0.71905
PSMB8	21	0.78567	51.550	50.73333	725.2455	3582	0.66190
MX2	20	0.92117	51.595	52.15000	166.9635	1952	0.78947
CXCL1	20	0.83742	49.030	48.90000	448.4261	2204	0.60526
IRF8	19	0.84429	51.163	49.66667	71.51968	1218	0.73684
STAT1	19	0.45368	49.415	56.78333	3186.268	16486	0.19883
CXCL10	19	0.74378	48.905	52.70000	828.0069	5742	0.64912
C3AR1	19	0.75048	48.447	49.28333	510.6841	1548	0.65497
IFI35	18	0.99902	51.134	46.76667	16.85864	158	0.88889
IFITM2	17	1.07668	51.137	46.26667	0.740030	16	0.97794
IFITM1	17	1.07668	50.947	46.26667	0.740030	16	0.97794
IFITM3	17	1.07668	51.047	46.26667	0.740030	16	0.97794
ISG20	17	1.07668	50.755	46.26667	0.740030	16	0.97794
ICAM1	17	0.81806	49.857	53.95000	1480.959	9140	0.34559
CCL4	17	0.87430	48.431	47.06667	79.97945	364	0.79412
CXCR4	17	0.95127	48.669	50.70000	588.8160	3532	0.77941
CCR1	17	0.88240	48.938	46.43333	61.21627	226	0.80147
FPR1	16	0.95127	48.054	46.10000	82.02197	356	0.88333
CCR7	16	1.05156	48.012	46.10000	237.2305	1468	0.87500
CXCL9	15	1.05156	47.915	45.43333	0.000000	0	1.00000

EPC: Edge Percolated Component; MNC: Maximum Neighborhood Component; DMNC: Density of Maximum Neighborhood Component; EcCentricity, Closeness, Betweenness, and Stress

As show in Figure 11, the MCODE module determines two clusters (highly interconnected regions) in the network, with score=17.647 and 10, respectively. Clusters in a protein-protein interaction network are often protein complexes and parts of pathways, which mean different things in different types of networks. Interestingly, cluster 1 and cluster 2 are linked by STAT1 and ICAM1. Current studies have shown that ICAM1 is a downstream gene of STAT1, and activation of STAT1 induces the expression of ICAM1[40]. Based on the scoring parameters in Table 4,

simultaneously combine the comprehensive analysis of signaling information network and miRNA-gene interaction network we constructed before, it is speculated that STAT1 may be a key gene affecting the process of RAS.

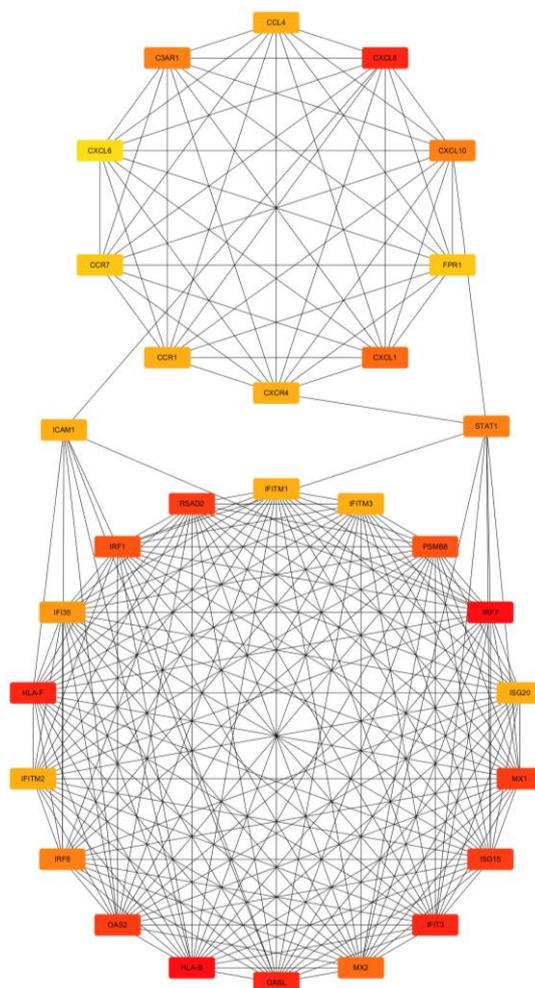


Figure 11 The Clusters of PPI network determined by MCODE module

Screening of potential natural compounds

Finally, eight natural components were finally screened to have high affinity with STAT1 based on the TCMSP database, including genistein, daidzein, kaempferol, resveratrol, rosmarinic acid, triptolide, quercetin and (-)-epigallocatechin-3-gallate. Table 4 shows the pharmacokinetic characteristics of above ingredients.

TABLE 4 The pharmacokinetic characteristics of potential natural compounds for RAS treatment

Molecule name	MW	AlogP	Hdon	Hacc	OB (%)	Caco-2	BBB	DL	FASA-	TPSA	RBN
Genistein	270.25	2.07	3	5	17.93	0.43	-0.40	0.21	0.00	90.90	1
Daidzein	254.25	2.33	2	4	19.44	0.59	-0.22	0.19	0.00	70.67	1
Kaempferol	286.25	1.77	4	6	41.88	0.26	-0.55	0.24	0.00	111.1	1
Resveratrol	228.26	3.01	3	3	19.07	0.80	-0.01	0.11	0.49	60.69	2
Rosmarinic acid	360.34	2.69	5	8	1.38	-0.54	-1.24	0.35	0.47	144.5	7
Triptolide	360.44	0.87	1	6	51.29	0.25	-0.19	0.68	0.28	84.12	1
Quercetin	302.25	1.50	5	7	46.43	0.05	-0.77	0.28	0.38	131.4	1
(-)-Epigallocatechin											
-3-gallate	458.40	2.89	8	11	55.09	-0.57	-1.70	0.77	0.37	197.4	4

MW : molecular weight; AlogP: the critical for measuring hydrophobicity of molecule; Hdon and Hacc: the measures of the hydrogen-bonding ability of a molecule expressed in terms of number of possible hydrogen-bond donors and acceptors, respectively; OB : Oral bioavailability; Caco-2 : the ingredients' transport rates (nm/s) in Caco-2 monolayers to represent the intestinal epithelial permeability; BBB : blood-brain barrier; DL :Drug-likeness, a qualitative concept used in drug design for an estimate on how "drug-like" a prospective compound is; FASA- : Fractional water accessible surface area of all atoms with negative partial charge, can be used as drug-likeness evaluation for drug-like molecules; TPSA : a physico chemical property describing the polarity of molecules; RBN :description for molecular flexibility, the number of bonds which allow free rotation around themselves, and roughly proportional to molecular size for many "drug-like" compounds.

Predict potential targets based on Pharmmapper and enrichment analysis

We predicted the potential targets of the above 8 compounds based on Pharmmapper. The Table 5 lists the intersection of predicted targets and DEGs of RAS, and Figure 12 indicates that the potential targets of these compounds affecting RAS are very similar, and it also suggests that these compounds may affect the process of RAS through multiple targets.

TABLE 5 The predicted targets of RAS treatment for each compound

Molecule	Gene
Genistein	STAT1/ICAM1/VCAM1/SOD2/CRYAB/C3/SELE
Daidzein	STAT1/ICAM1/VCAM1/SOD2/IL6/C3/CAMK4/CSF2RB/HCK/HLA-B/HLA-E/IFNGR1/INHBA/LTF/MSN/NT5C3A/PANK1/THBS1/TPK1
Quercetin	STAT1/ICAM1/VCAM1/SELE/CYP1B1/NQO1/IRF1/CAMK4/CSF2RB/HCK/HLA-B/HLA-E/HLA-G/HMGCS1/IFNGR1/LTF/LYZ/MMP1/MMP9/NPR3/NT5C3A/PANK1/THBS1/TLR1/MMP3/IL6
Kaempferol	STAT1/ICAM1/VCAM1/SELE/CYP1B1/C3/CD74/GZMB/HLA-E/IFNGR1/IL10RA/LTF/MS4A1/MSN/THBS1/THBS2/TPK1/WARS1
Resveratrol	IL6/STAT1/ICAM1/SELE/VCAM1/SOD2/CYP1B1/NQO1/BIRC3/C1R/C3/CSF2RB/HCK/HLA-E/ME1/MSN/NPR3/SAMHD1/TLR1
(-)-Epigallocatechin-3-gallate	IL6/MMP3/STAT1/TLR4/BTK/C3/CAMK4/FAP/GCH1/GLUL/GZMB/GZMK/HCK/HLA-B/HLA-E/HMGCS1/IFNGR1/KLF10/LTF/MMP1/MSN/NT5C3A/PANK1/PLA2G2A/THBS1/TRIM21/WARS1
Rosmarinic Acid	STAT1/IDO1/ALOX12/ANXA1/BTK/CD38/HCK/HLA-E/LAP3/LYZ/MS4A1/NT5C3A/RGS18/SOD2/TPK1/TRIM21
Triptolide	STAT1/CXCR4/BIRC3/CCR7/ALOX12/ANXA1/BTK/CAMK4/GBP1/HLA-E/HMGCS1/LAP3/LTF/MS4A1/MSN/PLA2G2A/RGS18/SOD2/TPK1/TRIM21/WARS1

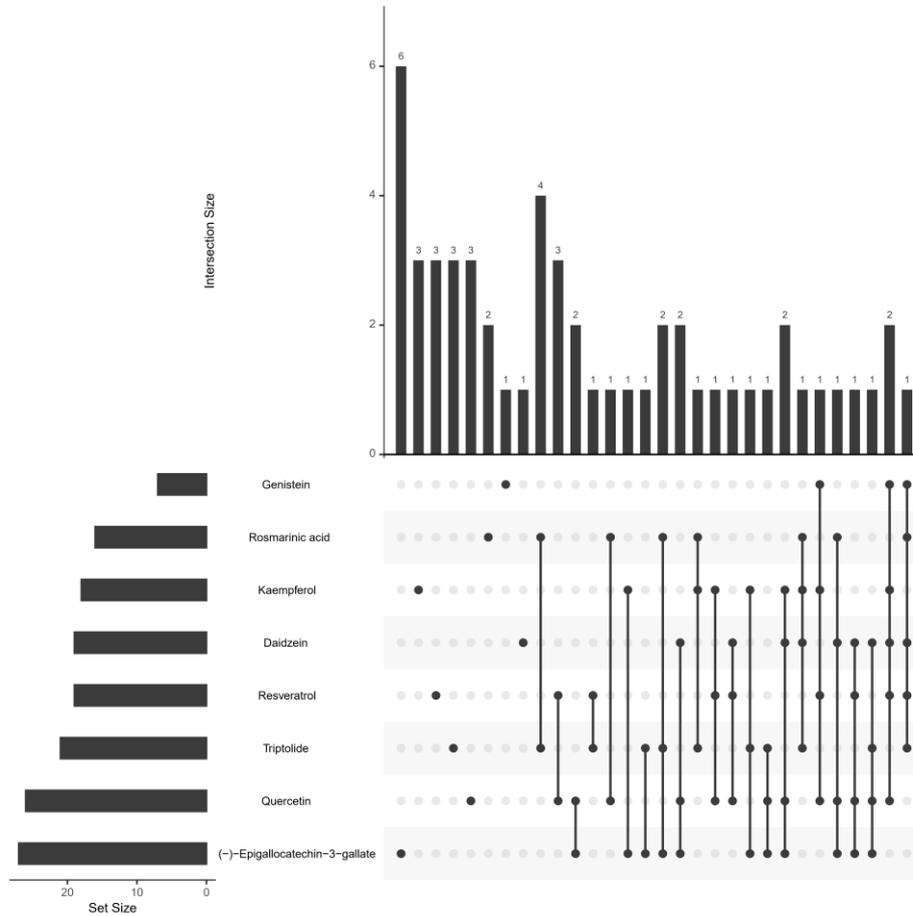
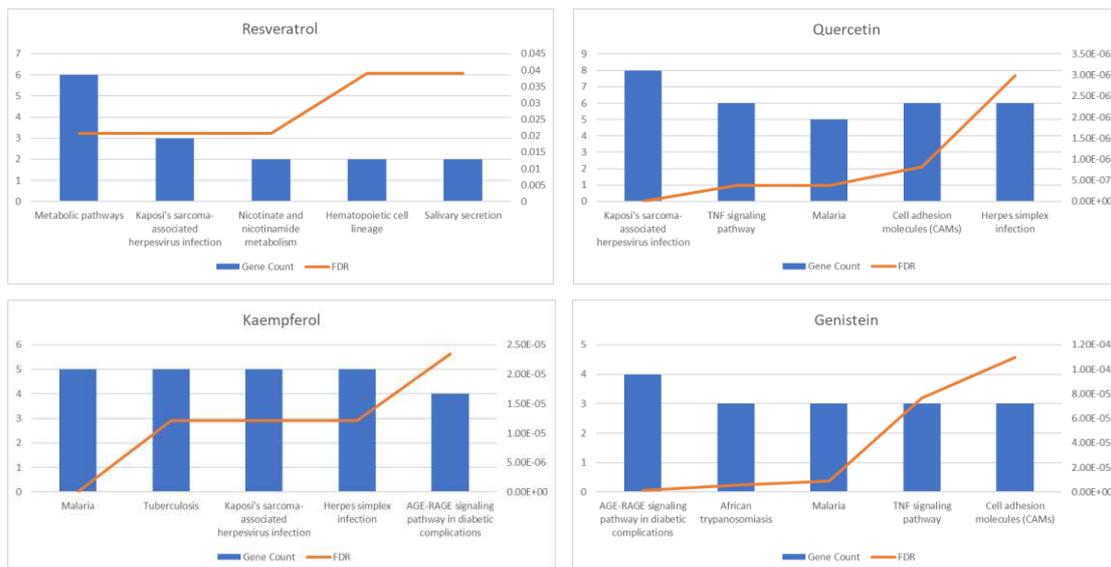


Figure 12 The intersection of the potential targets of 8 compounds

Figure 13 shows the results of the KEGG enrichment analysis on the predicted targets of these compounds, suggesting that the potential targets of these compounds to affect the RAS process are mainly concentrated in a variety of viral infection-related immune pathways, TNF pathways, cell adhesion and other biological pathways.



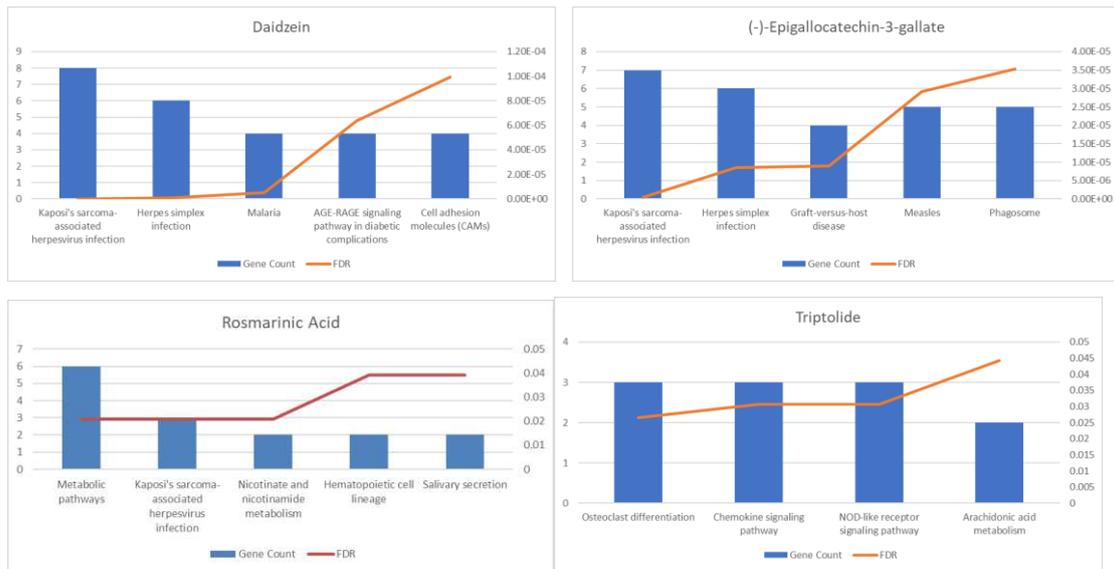
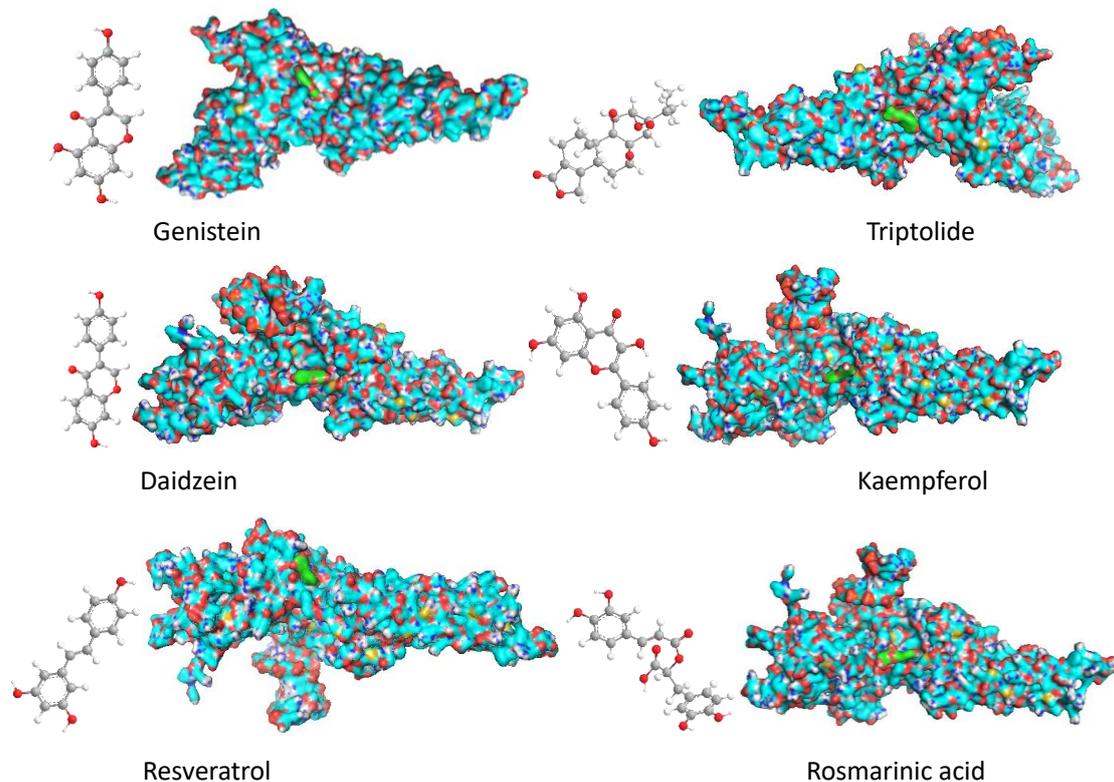


Figure 13 The result of KEGG enrichment

Confirm the target of RAS based on reverse docking technology

As shown in Figure 14, the results of reverse molecular docking suggest that different compounds may bind to different parts of the STAT1 protein, while compounds with similar molecular structures bind to similar positions of the STAT1 protein. The higher the docking score, the stronger the binding force to the protein. According to the docking score, it is sorted from high to low: triptolide (-9.1 kcal/mol), (-)-epigallocatechin-3-gallate (-8.1 kcal/mol), rosmarinic acid (-7.3 kcal/mol), quercetin (-7.3 kcal/mol), genistein (-7.1 kcal/mol), daidzein (-6.9 kcal/mol), kaempferol (-6.9 kcal/mol), resveratrol (-5.6 kcal/mol).



Cluster 1 may be the initiation factors of RAS, while Cluster 2 mainly contains chemokines and their receptors. The crosstalk between the matrix metalloprotease system and the chemokine network had been proved, and chemokines and their receptors may regulate the activity of matrix metalloproteinases[43-44], which may affect the synthesis and degradation of oral epithelial collagen, and may also be one of the mechanisms of RAS. According to previous studies and the DEGs in this chip, many types of MMP or TIMP in RAS have been confirmed to be differentially expressed compared with normal tissues[45-47]. Therefore, inhibition of STAT1 may cut off some abnormalities in the TNF pathway and inhibit chemokines activity, which in turn affects the related activities of matrix metalloproteinases and affects the decomposition of collagen in the oral cavity.

Finally, we have also confirmed that STAT1 protein is one of the potential therapeutic targets of RAS, and this target can be used to screen potential therapeutic compounds. Finally, genistein, daidzein, kaempferol, resveratrol, rosmarinic acid, triptolide, quercetin and (-)-epigallocatechin-3-gallate were selected from the TCMSP database, and both of them is the STAT-1 inhibitor[48-52]. Interestingly, some of those ingredients, such as rosmarinic acid, quercetin, (-)-Epigallocatechin-3-gallate, resveratrol, etc., have already been made into topical formulations for the treatment of oral ulcers, such as quercetin, (-)-Epigallocatechin-3-gallate, resveratrol [53-55]. The results of reverse molecular docking suggest that in addition to triptolide, (-)-Epigallocatechin-3-gallate and resveratrol, the other 5 compounds (flavonoids) with similar structures bind to STAT1 at almost the same position, that is, this position may be It is the key position for flavonoids to inhibit stat1 protein.

Declarations

Ethics approval and consent to participate

The experimental protocol was established, according to the ethical guidelines of the Helsinki Declaration and was approved by the Human Ethics Committee of the affiliated hospital of Qingdao University. Written informed consent was obtained from individual or guardian participants.

Consent for publication

Not applicable.

Availability of data

All data generated or analysed during this study are included in this published article and material

Competing interests

The authors declare no competing non-financial/financial interests.

Funding

This study was supported in part by grants from 2021 Qingdao Traditional Chinese Medicine Development Project (2021-zyyq22)

Authors' contributions

Ming-Chen CAO and Fan-Bo JING contributed to the conception of the study; Ming-Chen CAO and Meng-Xiang FANG retrieving and launching gene chip data mining work; Lei LI and Xiao-Min XING contributed significantly to manuscript preparation; Chang-Kai ZHOU and Long-Yuan WANG helped perform the analysis with constructive discussions; Long XU and Wei REN organized the tables and pictures of the full text.

Acknowledgements

First and foremost, I would like to show my deepest gratitude to my supervisor, Mr. JING, a respectable, responsible and resourceful scholar, who has provided me with valuable guidance in every stage of the writing of this thesis. Without his enlightening instruction, impressive kindness and patience, I could not have completed my thesis. His keen and vigorous academic observation enlightens me not only in this thesis but also in my future study. I would also like to thank all my teachers who have helped me to develop the fundamental and essential academic competence. My sincere appreciation also goes to the teachers and students from the affiliated hospital of Qingdao University, who participated this study with great cooperation. Last but not least, I'd like to thank all my friends, for their encouragement and support.

References

- [1] Saikaly, S. K., Saikaly, T. S., & Saikaly, L. E. (2018). Recurrent aphthous ulceration: a review of potential causes and novel treatments. *Journal of Dermatological Treatment* , 29 (6), 542-552.
- [2] Edgar, N. R., Saleh, D., & Miller, R. A. (2017). Recurrent aphthous stomatitis: a review. *The Journal of clinical and aesthetic dermatology*, 10 (3), 26.
- [3] Ślebioda, Z., Szponar, E., & Kowalska, A. (2014). Etiopathogenesis of recurrent aphthous stomatitis and the role of immunologic aspects: literature review. *Archivum immunologiae et therapiae experimentalis* , 62 (3), 205-215.
- [4] Ślebioda, Z., Szponar, E., & Kowalska, A. (2013). Recurrent aphthous stomatitis: genetic aspects of etiology. *Advances in Dermatology and Allergology/Postępy Dermatologii I Alergologii* , 30 (2), 96.
- [5] Tarakji, B., Gazal, G., Al-Maweri, S. A., Azzeghaiby, S. N., & Alaizari, N. (2015). Guideline for the diagnosis and treatment of recurrent aphthous stomatitis for dental practitioners. *Journal of international oral health: JIOH*, 7 (5), 74.
- [6] Albrektson, M., Hedström, L., & Bergh, H. (2014). Recurrent aphthous stomatitis and pain management with low-level laser therapy: a randomized controlled trial. *Oral surgery, oral medicine, oral pathology and oral radiology*, 117 (5), 590-594.
- [7] Belenguer-Guallar, I., Jiménez-Soriano, Y., & Claramunt-Lozano, A. (2014). Treatment of recurrent aphthous stomatitis. A literature review. *Journal of clinical and experimental dentistry*, 6 (2), e168.
- [8] Mangul, S., Martin, L. S., Langmead, B., Sanchez-Galan, J. E., Toma, I., Hormozdiari, F., ... & Eskin,

E. (2019). How bioinformatics and open data can boost basic science in countries and universities with limited resources. *Nature biotechnology*, 37 (3), 324-326.

[9] Attwood, T. K., Blackford, S., Brazas, M. D., Davies, A., & Schneider, M. V. (2019). A global perspective on evolving bioinformatics and data science training needs. *Briefings in Bioinformatics*, 20 (2), 398-404.

[10] Pei, G., Chen, L., & Zhang, W. (2017). WGCNA application to proteomic and metabolomic data analysis. In *Methods in enzymology* (Vol. 585, pp. 135-158). Academic Press.

[11] Tian, Z., He, W., Tang, J., Liao, X., Yang, Q., Wu, Y., & Wu, G. (2020). Identification of Important Modules and Biomarkers in Breast Cancer Based on WGCNA. *OncoTargets and therapy*, 13, 6805.

[12] Pei, G., Chen, L., & Zhang, W. (2017). WGCNA application to proteomic and metabolomic data analysis. In *Methods in enzymology* (Vol. 585, pp. 135-158). Academic Press.

[13] Yadav, B. S., & Tripathi, V. (2018). Recent advances in the system biology-based target identification and drug discovery. *Current Topics in Medicinal Chemistry*, 18 (20), 1737-1744.

[14] Katsila, T., Spyroulias, G. A., Patrinos, G. P., & Matsoukas, M. T. (2016). Computational approaches in target identification and drug discovery. *Computational and structural biotechnology journal*, 14, 177-184.

[15] Katara, P. (2020). Computational Approaches for Drug Target Identification. In *Computer-Aided Drug Design*, Springer, Singapore. pp. 163-185.

[16] Karagkouni, Dimitra, et al. (2018) DIANA-TarBase v8: a decade-long collection of experimentally supported miRNA–gene interactions. *Nucleic acids research*, 46.D1: D239-D245.

[17] Licata, L. , Surdo, P. L. , Iannuccelli, M. , Palma, A. , & Cesareni, G. . (2019). Signor 2.0, the signaling network open resource 2.0: 2019 update. *Nucleic Acids Research*, 48(database).

[18] Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., ... & Mesirov, J. P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, 102 (43), 15545-15550.

[19] Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., ... & Jensen, L. J. (2019). STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic acids research*, 47 (D1), D607-D613.

[20] Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., ... & Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, 13 (11), 2498-2504.

[21] Ru, J., Li, P., Wang, J., Zhou, W., Li, B., Huang, C., ... & Xu, X. (2014). TCMSP: a database of systems pharmacology for drug discovery from herbal medicines. *Journal of cheminformatics*, 6 (1), 13.

[22] Wang, X., Shen, Y., Wang, S., Li, S., Zhang, W., Liu, X., ... & Li, H. (2017). PharmMapper 2017 update: a web server for potential drug target identification with a comprehensive target pharmacophore database. *Nucleic acids research*, 45 (W1), W356-W360.

[23] Trott, O., & Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, 31 (2), 455-461.

[24] DeLano, W. L. The PyMOL molecular graphics system. Schrödinger LLC. 2002; Version 2.0.

[25] Mogensen, T. H. . (2019). Irf and stat transcription factors - from basic biology to roles in

- infection, protective immunity, and primary immunodeficiencies. *Frontiers in Immunology*, 9.
- [26] Krzysztof Sikorski, Stefan Chmielewski, Adam Olejnik, Joanna Z. Wesoly, Uwe Heemann, Marcus Baumann & Hans Bluysen (2012) STAT1 as a central mediator of IFN γ and TLR4 signal integration in vascular dysfunction, *JAK-STAT*, 1:4, 241-249, DOI: 10.4161/jkst.22469
- [27] Yarilina, A. , Park-Min, K. H. , Antoniv, T. , Hu, X. , & Ivashkiv, L. B. . (2008). Tnf activates an irf1-dependent autocrine loop leading to sustained expression of chemokines and stat1-dependent type i interferon-response genes. *Nature Immunology*, 9(4), 378.
- [28] Ohmori, Y. , & Hamilton, T. A. . (2001). Requirement for stat1 in lps-induced gene expression in macrophages. *J Leukoc Biol*, 69(4), 598-604.
- [29] Lehtonen, A. , Matikainen, S. , & Julkunen, I. . (1997). Interferons up-regulate stat1, stat2, and irf family transcription factor gene expression in human peripheral blood mononuclear cells and macrophages. *Journal of Immunology*, 159(2), 794.
- [30] Li, X. , Leung, S. , Qureshi, S. , Darnell, J. E. , & Stark, G. R. . (1996). Formation of stat1-stat2 heterodimers and their role in the activation of irf-1 gene transcription by interferon-. *Journal of Biological Chemistry*, 271(10), 5790-5794.
- [31] Thibault, D. L. , Chu, A. D. , Graham, K. L. , Balboni, I. , Lee, L. Y. , & Kohlmoos, C. , et al. (2008). IRF9 and STAT1 are required for IgG autoantibody production and B cell expression of TLR7 in mice. *Journal of Clinical Investigation*, 118(4), 1417.
- [32] Huang, Y. , Walstrom, A. , Zhang, L. , Zhao, Y. , & Zheng, J. C. . (2009). Type i interferons and interferon regulatory factors regulate tnf-related apoptosis-inducing ligand (trail) in hiv-1-infected macrophages. *PLoS ONE*, 4(4), e5397-.
- [33] David, Langlais, Luis, B., Barreiro, & Philippe, et al. (2016). The macrophage irf8/irf1 regulome is required for protection against infections and is associated with chronic inflammation. *The Journal of Experimental Medicine*, 213(4).
- [34] Piaszyk-Borychowska, A. , Joanna Wesoly, & Bluysen, H. . (2015). Id: 243 : pro-atherogenic roles of stat1 and irf8 in cardiovascular disease. *Cytokine*, 76(1), 110-110.
- [35] Lui, A. , Choi, H. J. , & Lewis-Wambi, J. . (2015). Abstract 717: overexpression of interferon-stimulated genes is critical for the survival of aromatase inhibitor-resistant breast cancer cells. *Cancer Research*, 75(15 Supplement), 717-717.
- [36] Tymoszek, P. , Charoentong, P. , Hackl, H. , Spilka, R. , Müller-Holzner, Elisabeth, & Trajanoski, Z. , et al. (2014). High stat1 mrna levels but not its tyrosine phosphorylation are associated with macrophage infiltration and bad prognosis in breast cancer. *Bmc Cancer*, 14(1), 257.
- [37] Herranz, Maria Gómez, Nekulova, M. , Faktor, J. , Hernychova, L. , Kote, S. , & Sinclair, E. H. , et al. (2019). The effects of ifitm1 and ifitm3 gene deletion on ifn γ stimulated protein synthesis. *Cellular Signalling*.
- [38] Hu, X. , Park-Min, K. H. , Ho, H. H. , & Ivashkiv, L. B. . (2005). Ifn- γ -primed macrophages exhibit increased ccr2-dependent migration and altered ifn- γ responses mediated by stat1. *Journal of Immunology*, 175(6), 3637-3647.
- [39] Travis, Laver, Susan, E. , Nozell, Ety, N. , & Benveniste. (2008). Ifn-beta-mediated inhibition of il-8 expression requires the isgf3 components stat1, stat2, and irf-9. *Journal of Interferon & Cytokine Research : the official journal of the International Society for Interferon and Cytokine Research*.
- [40] Ryan, N. , Anderson, K. , Volpedo, G. , Hamza, O. , & Oghumu, S. . (2020). Abstract A11: STAT1 mediates resistance to experimental oral cancer that is associated with enhanced antitumor T-cell

responses. Abstracts: AACR-AHNS Head and Neck Cancer Conference: Optimizing Survival and Quality of Life through Basic, Clinical, and Translational Research; April 29-30, 2019; Austin, TX.

[41] Yockell-Lelievre, J., Spriet, C., Cantin, P., Malenfant, P., Heliot, L., De Launoit, Y., & Audette, M. (2009). Functional cooperation between Stat-1 and ets-1 to optimize icam-1 gene transcription. *Biochemistry and Cell Biology* , 87 (6), 905-918.

[42] Janowska-Wieczorek, A., Marquez, L. A., Dobrowsky, A., Ratajczak, M. Z., & Cabuhat, M. L. (2000). Differential MMP and TIMP production by human marrow and peripheral blood CD34+ cells in response to chemokines. *Experimental hematology* , 28 (11), 1274-1285.

[43] Robinson, S. C., Scott, K. A., & Balkwill, F. R. (2002). Chemokine stimulation of monocyte matrix metalloproteinase - 9 requires endogenous TNF - α . *European journal of immunology* , 32 (2), 404-412.

[44] J Hatfield, K., Reikvam, H., & Bruserud, O. (2010). The crosstalk between the matrix metalloprotease system and the chemokine network in acute myeloid leukemia. *Current medicinal chemistry* , 17 (36), 4448-4461.

[45] Karasneh, J. A., Bani - Hani, M. E., Alkhateeb, A. M., Hassan, A. F., & Thornhill, M. H. (2014). Association of MMP but not TIMP - 1 gene polymorphisms with recurrent aphthous stomatitis. *Oral diseases* , 20 (7), 693-699.

[46] Kang, Y., & Zhany, Y. (2000). Expression and significance of MMP-9 in recurrent aphthous ulcer. *Journal of Practical Stomatology* , (05).

[47] KANG, Y. Y., ZHANG, Y., & SUN, Y. (2011). Study of expression and significance of MMP-28 in RAU, OLP, OLK. *Stomatology* , (9), 5.

[48] Hämäläinen, M., Nieminen, R., Vuorela, P., Heinonen, M., & Moilanen, E. (2007). Anti-inflammatory effects of flavonoids: genistein, kaempferol, quercetin, and daidzein inhibit STAT-1 and NF- κ B activations, whereas flavone, isorhamnetin, naringenin, and pelargonidin inhibit only NF- κ B activation along with their inhibitory effect on iNOS expression and NO production in activated macrophages. *Mediators of inflammation* , 2007 .

[49] Hamed, F., Mcdonagh, A., Almaghribi, S., Bakri, Y., & Tazi-Ahnini, R. (2018). Epigallocatechin-3 gallate inhibits stat-1/jak2/irf-1/hla-dr/hla-b and reduces cd8 mkg2d lymphocytes of alopecia areata patients. *International Journal of Environmental Research and Public Health* , 15(12), 2882.

[50] Hongqin, T., Xinyu, L., Heng, G., Lanfang, X., Yongfang, W., & Shasha, S. (2011). Triptolide inhibits IFN - γ signaling via the Jak/STAT pathway in HaCaT keratinocytes. *Phytotherapy Research* , 25 (11), 1678-1685.

[51] Ma, C., Wang, Y., Dong, L., Li, M., & Cai, W. (2015). Anti-inflammatory effect of resveratrol through the suppression of NF- κ B and JAK/STAT signaling pathways. *Acta biochimica et biophysica Sinica* , 47 (3), 207-213.

[52] Yang, J. H., Yoo, J. M., Lee, E., Lee, B., Cho, W. K., Park, K. I., & Ma, J. Y. (2018). Anti-inflammatory effects of Perillae Herba ethanolic extract against TNF- α /IFN- γ -stimulated human keratinocyte HaCaT cells. *Journal of ethnopharmacology* , 211 , 217-223.

[53] Guo, L., Zhao, Z. Y., Bai, J., & Han, S. Y. (2012). Preparation and treatment for oral ulcer of Quercetin and the drug-loaded chitosan composite material. In *Advanced Materials Research* (Vol. 583, pp. 44-48). Trans Tech Publications Ltd.

[54] Chen, P. N., Chu, S. C., Kuo, W. H., Chou, M. Y., Lin, J. K., & Hsieh, Y. S. (2011). Epigallocatechin-3 gallate inhibits invasion, epithelial- mesenchymal transition, and tumor growth in oral cancer cells. *Journal of agricultural and food chemistry* , 59 (8), 3836-3844.

[55] Huang, B., & Chen, H. (2015). (-)-Epigallocatechin-3-gallate inhibits matrix metalloproteinases in oral ulcers. *RSC Advances*, 5 (30), 23758-23766.

Figures

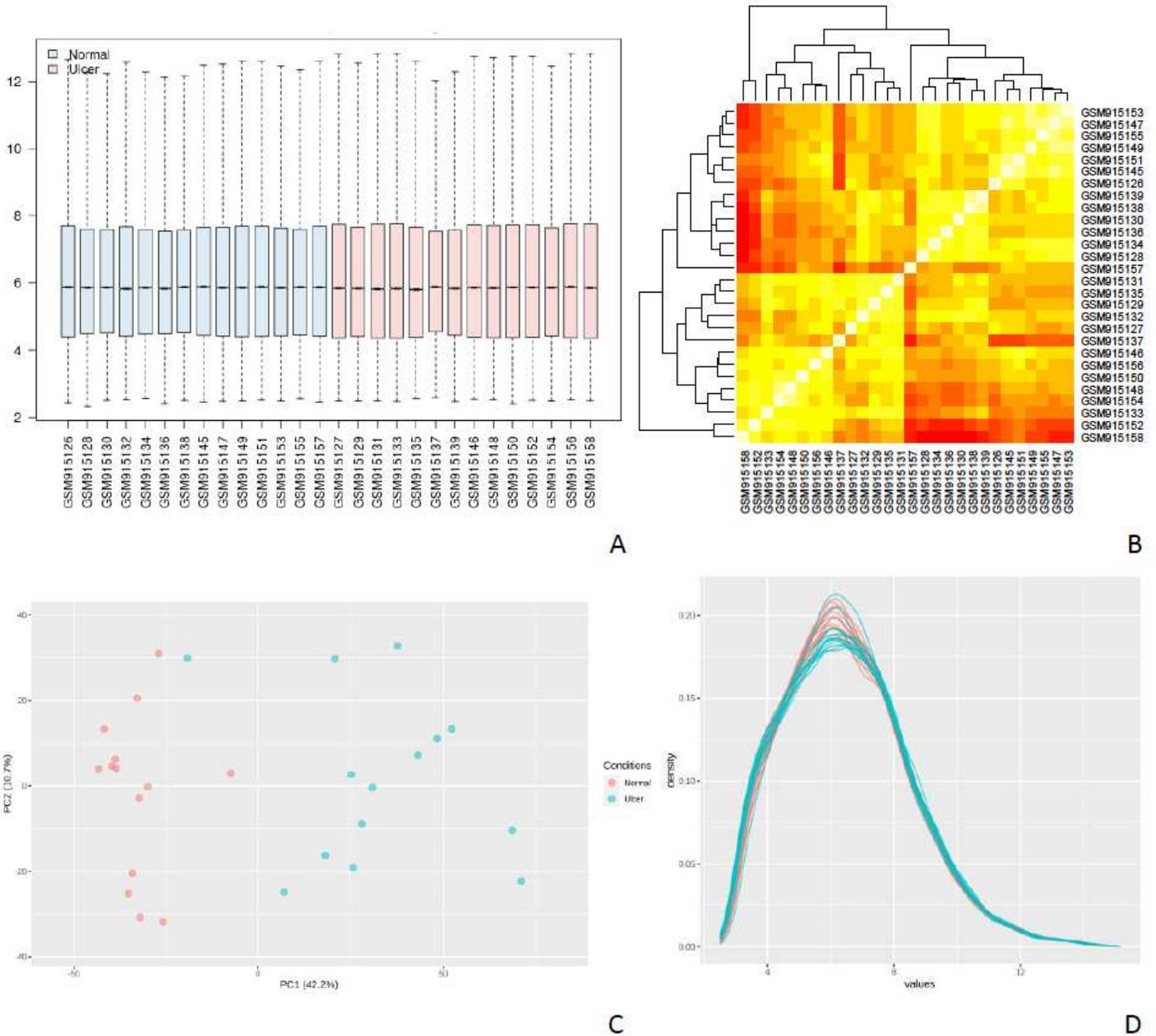


Figure 1

(A) box plot showing distribution of raw read counts in the GSE37265 dataset (B) Pearson's correlation analysis of samples from the GSE37265 dataset. (C) PCA of samples from the GSE37265 dataset. (D) The diagnostic plot

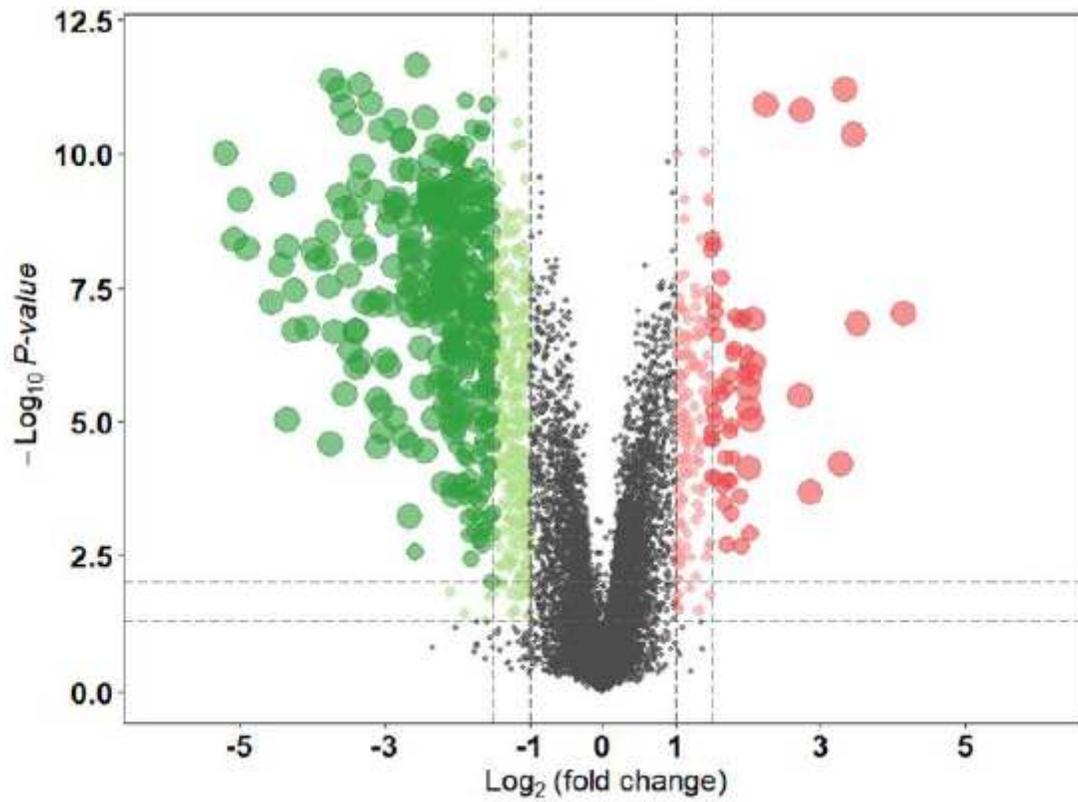


Figure 2

The volcano plot shows the up-regulated and down-regulated genes in RAS. The horizontal axis represents the fold change between health and RAS. The vertical axis represents the P value of t test for the differences between health and RAS.

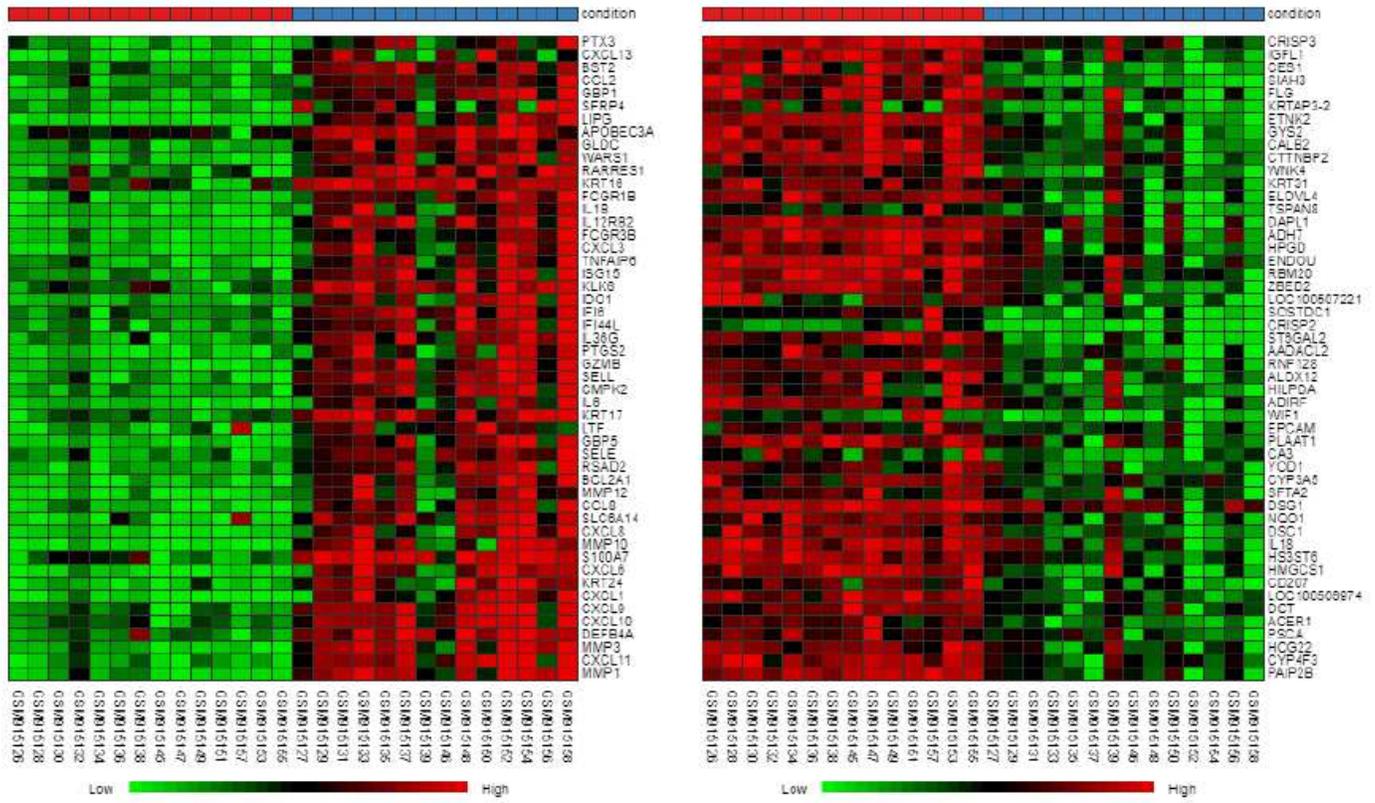


Figure 3

The heatmap shows the top 50 differential gene expression ($P < 0.05$) between health and RAS.

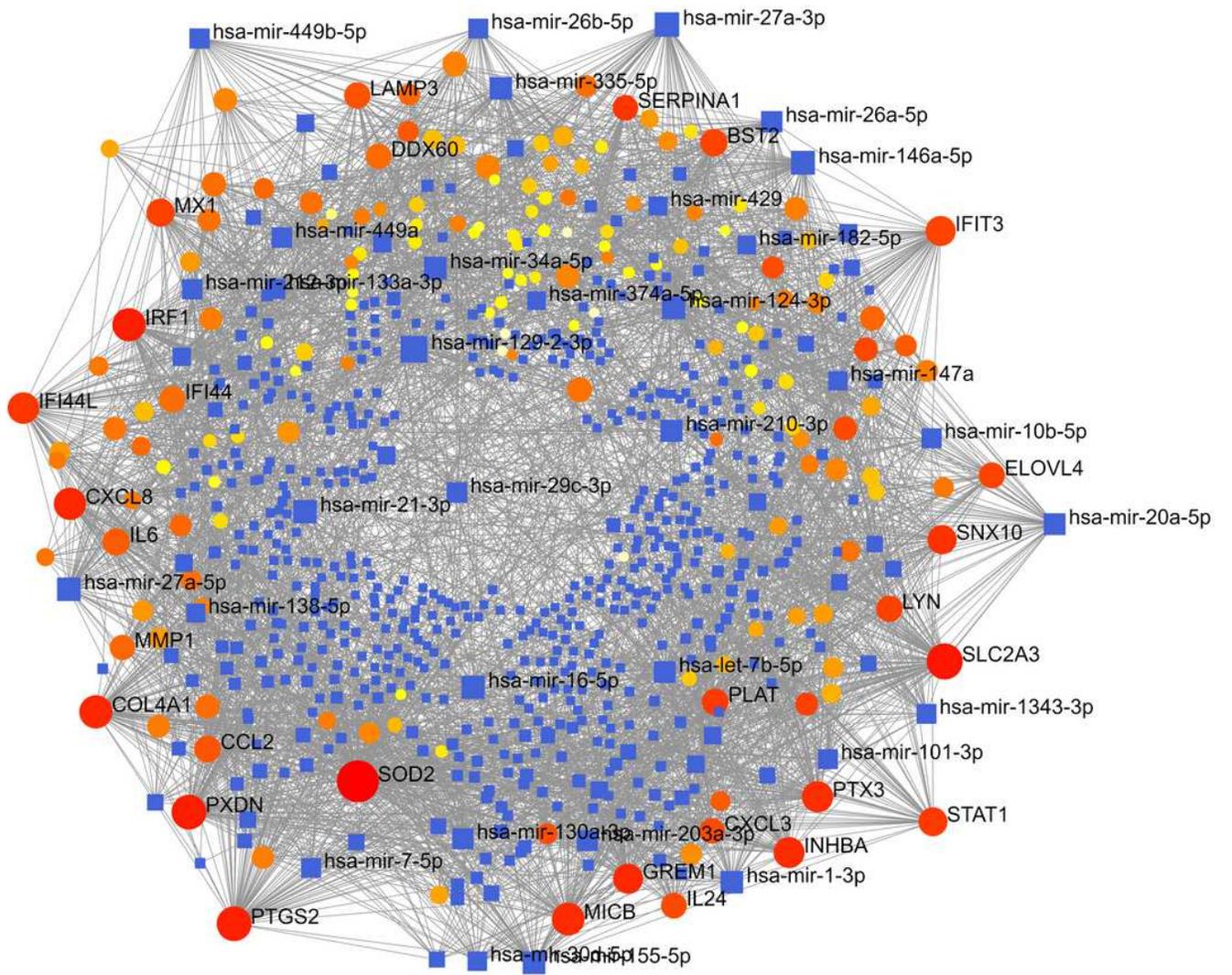


Figure 4

The miRNA-gene interaction network of RAS

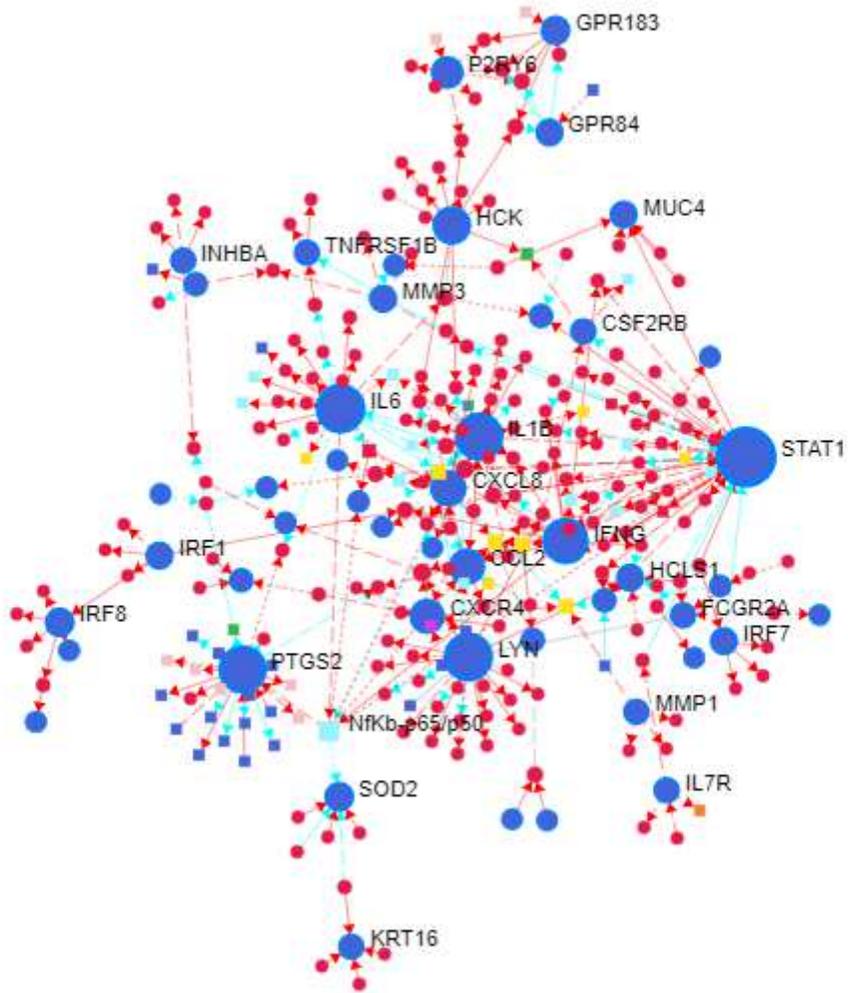
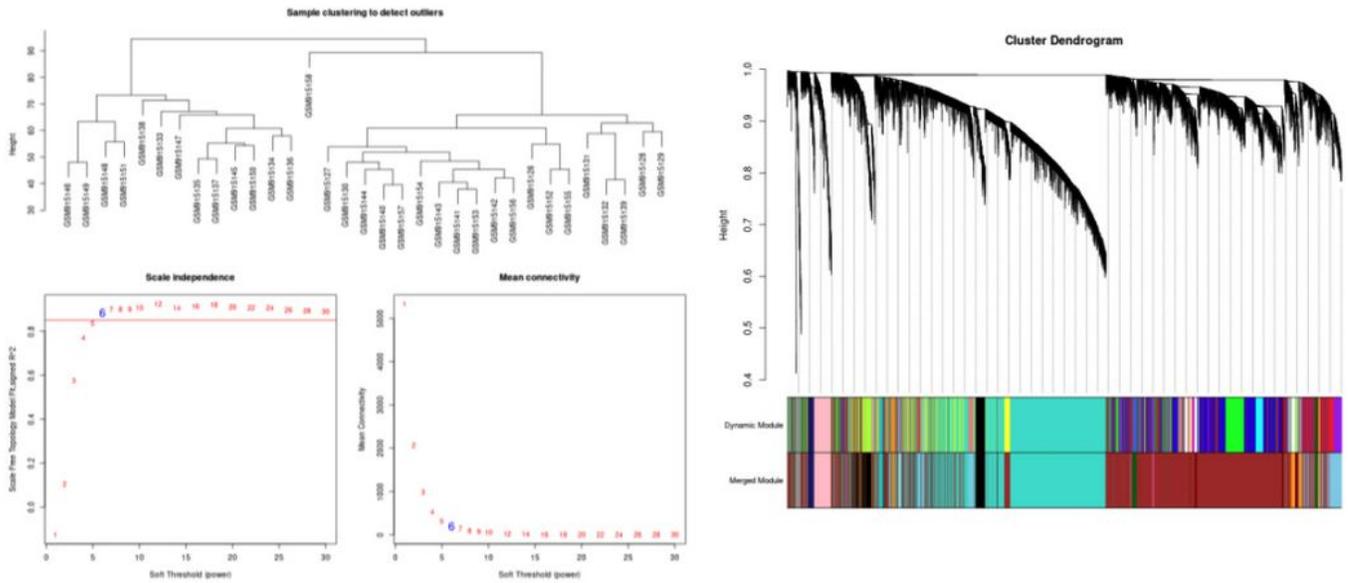


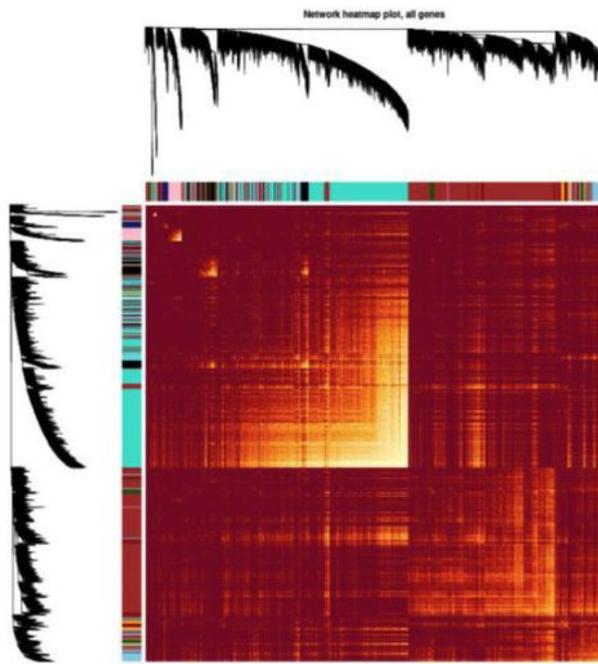
Figure 5

The signaling information network based on SIGNOR

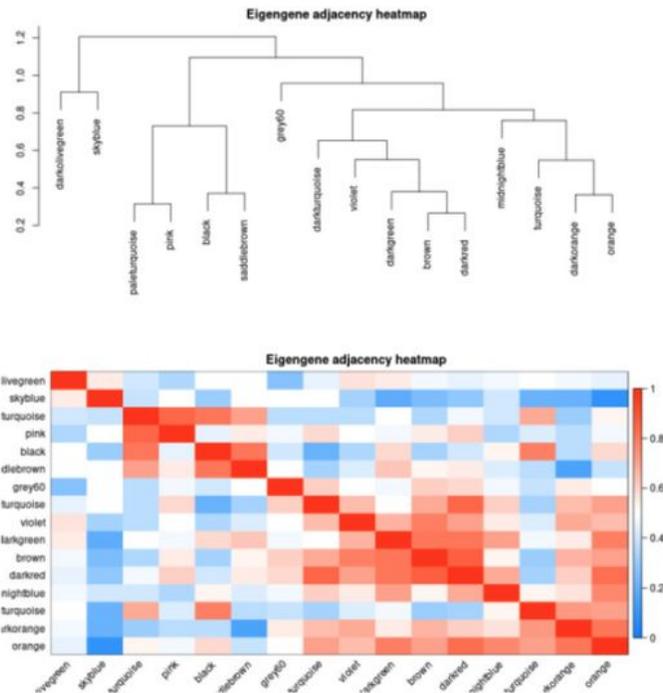


A

B



C



D

Figure 6

The result of WGCNA analysis A: Sample clustering to detect outliers; B: Analysis of network topology for a set of soft-thresholding powers. Scale independence and mean connectivity of various soft-thresholding values (β). The left picture displays the scale free fit index (y-axis) as a function of the soft-thresholding power (x-axis). The right picture shows the mean connectivity (degree, y-axis) as a function of the soft- thresholding power (x-axis); C: Clustering dendrograms of genes, with dissimilarity based on

topological overlap, together with assigned module colors. Cluster dendrogram of all filtered genes enriched based on the dissimilarity measure and the cluster module colors; D: The heatmap plot describes the Topological Overlap Matrix (TOM) among DEGs in the analysis; E: The eigengene dendrogram and heatmap identify groups of correlated eigengenes termed meta modules.

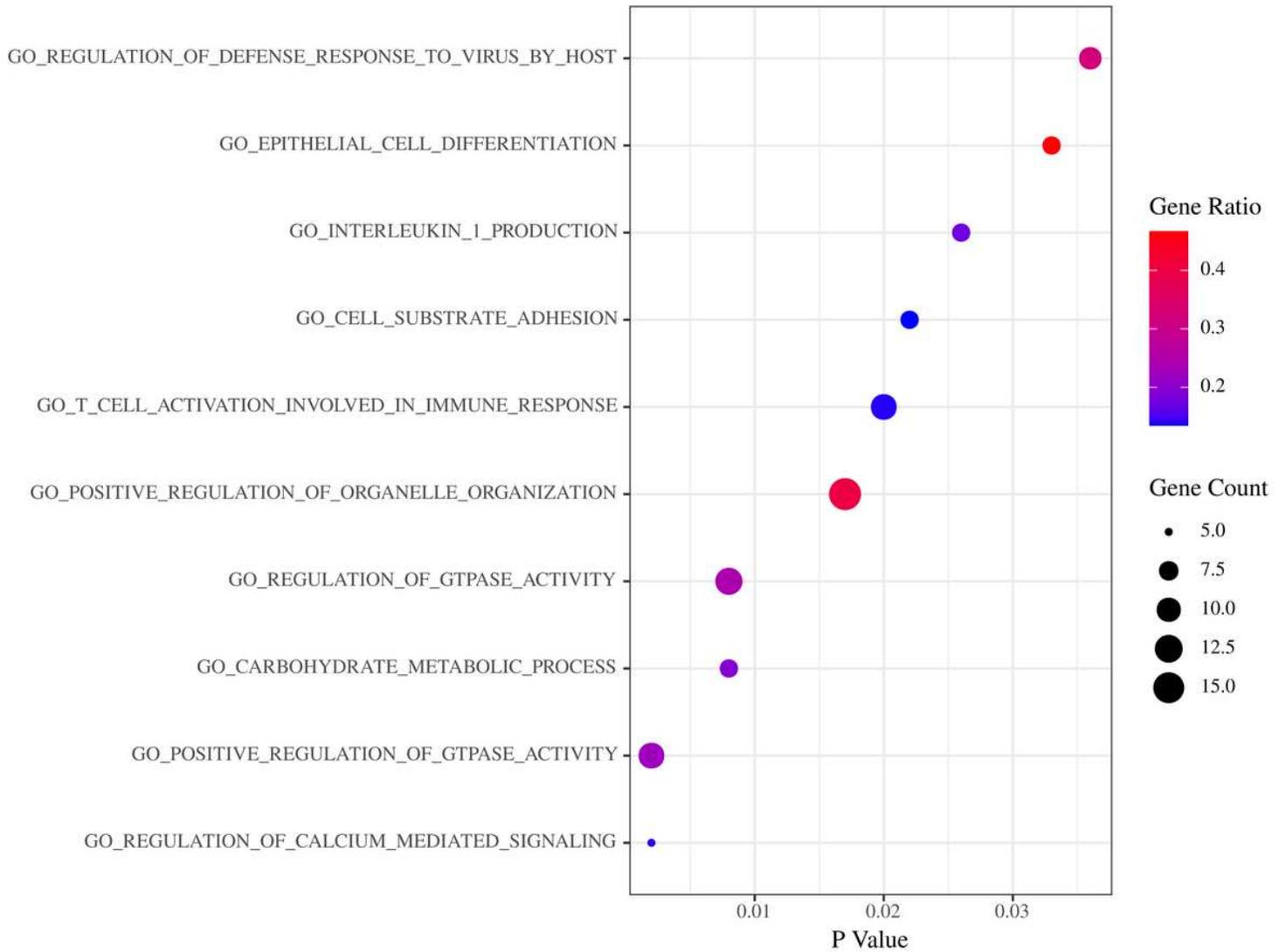


Figure 7

The result of GSEA enrichment analysis based on Go (biological process)

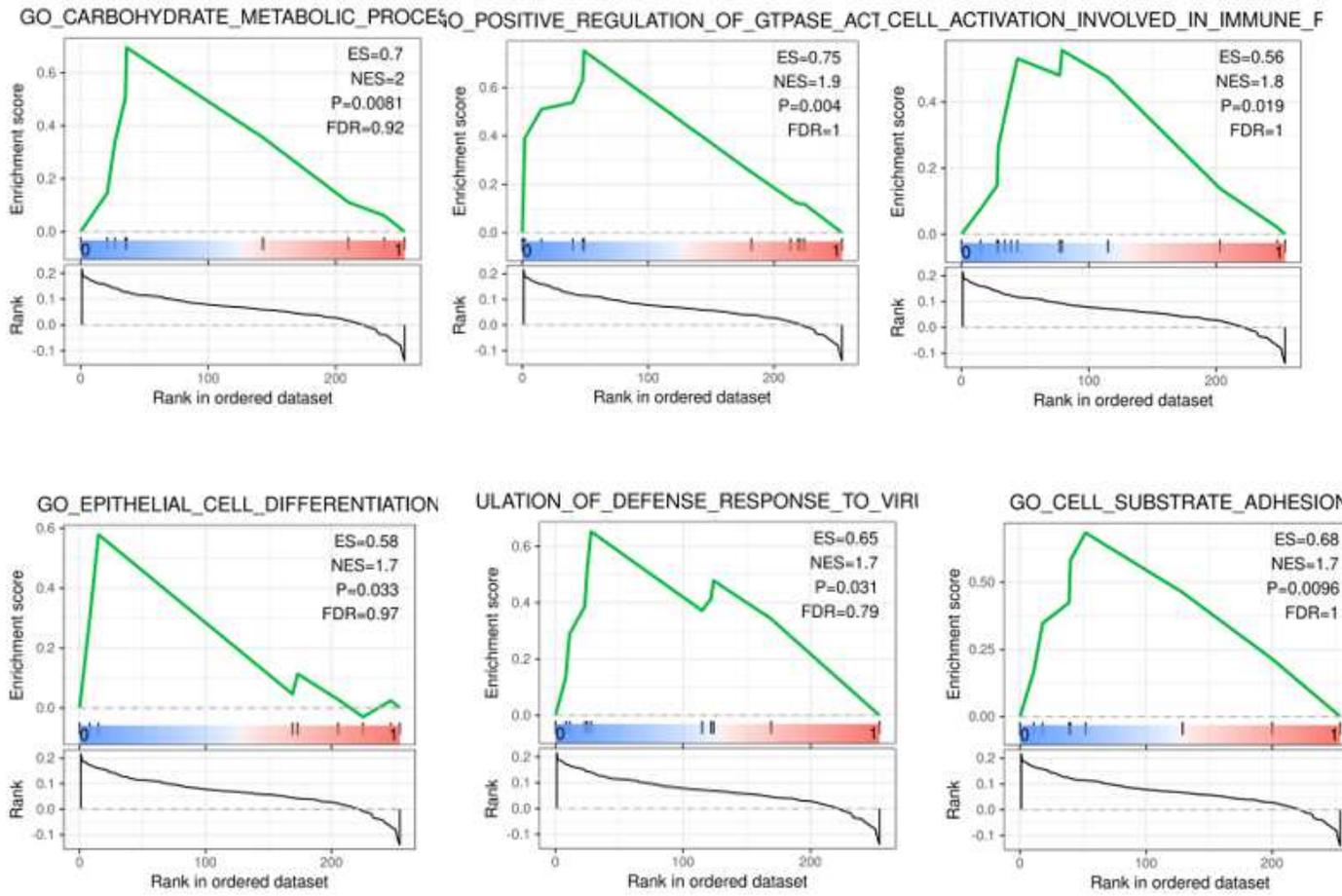


Figure 8

The Go (biological process) pathways highly related to RAS

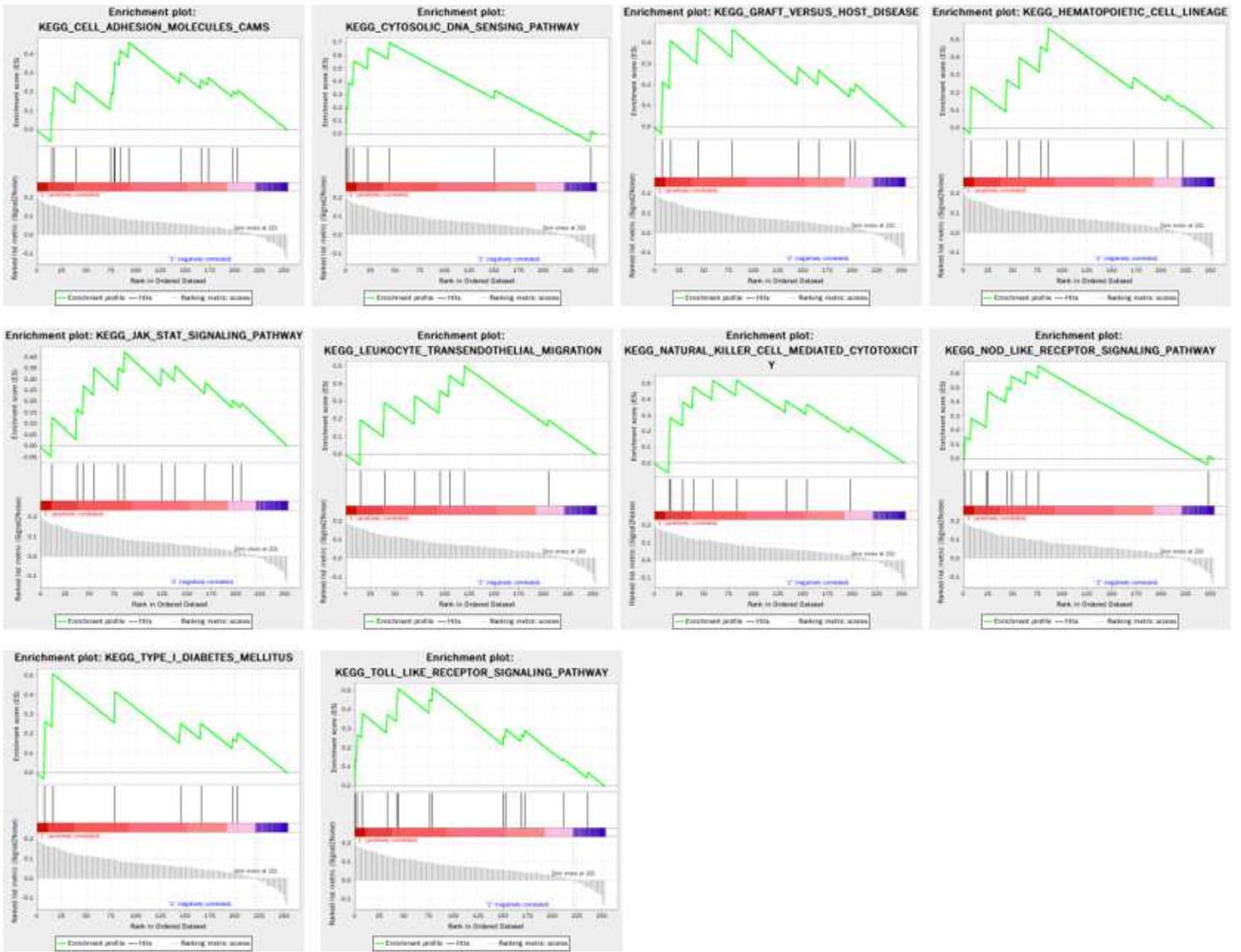


Figure 9

The result of GSEA enrichment analysis based on KEGG

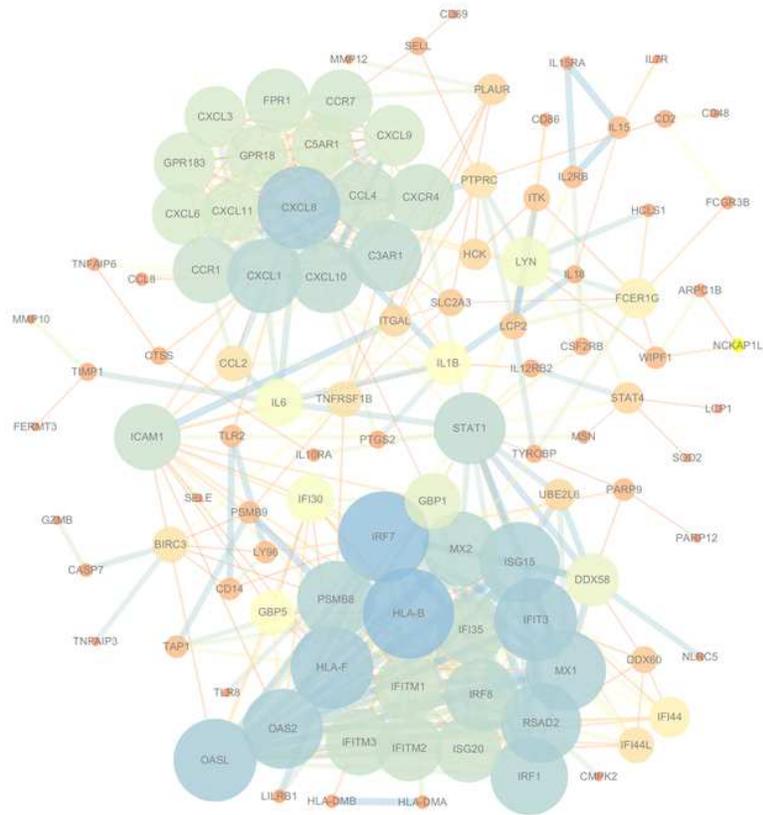


Figure 10

The protein-protein interaction (PPI) network of enriched DEGs

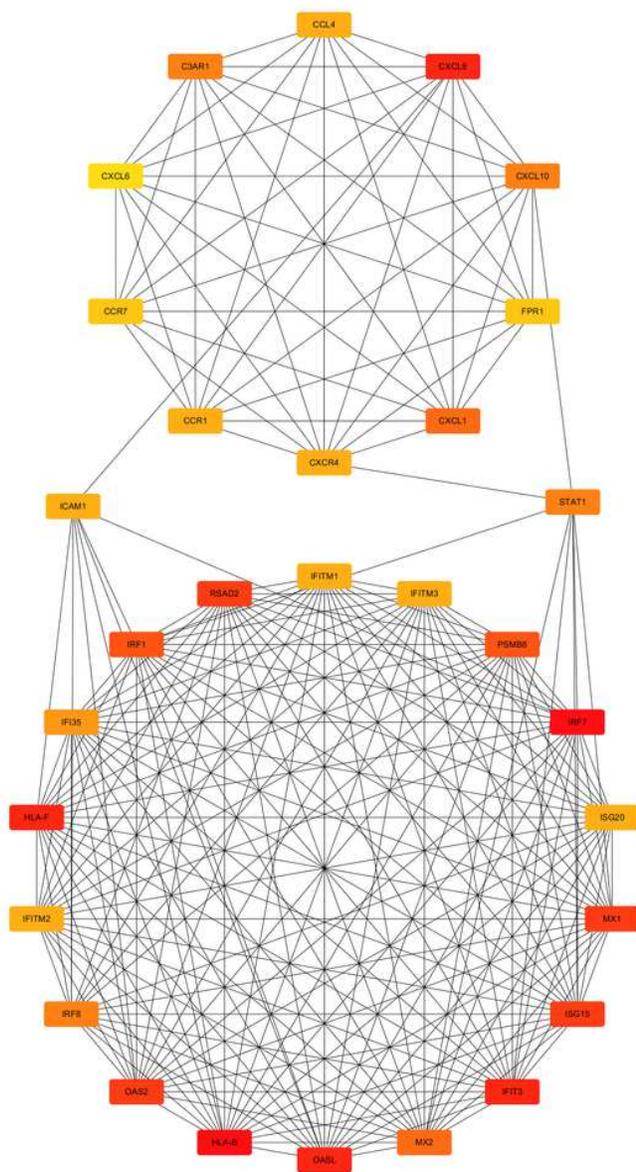


Figure 11

The Clusters of PPI network determined by MCODE module

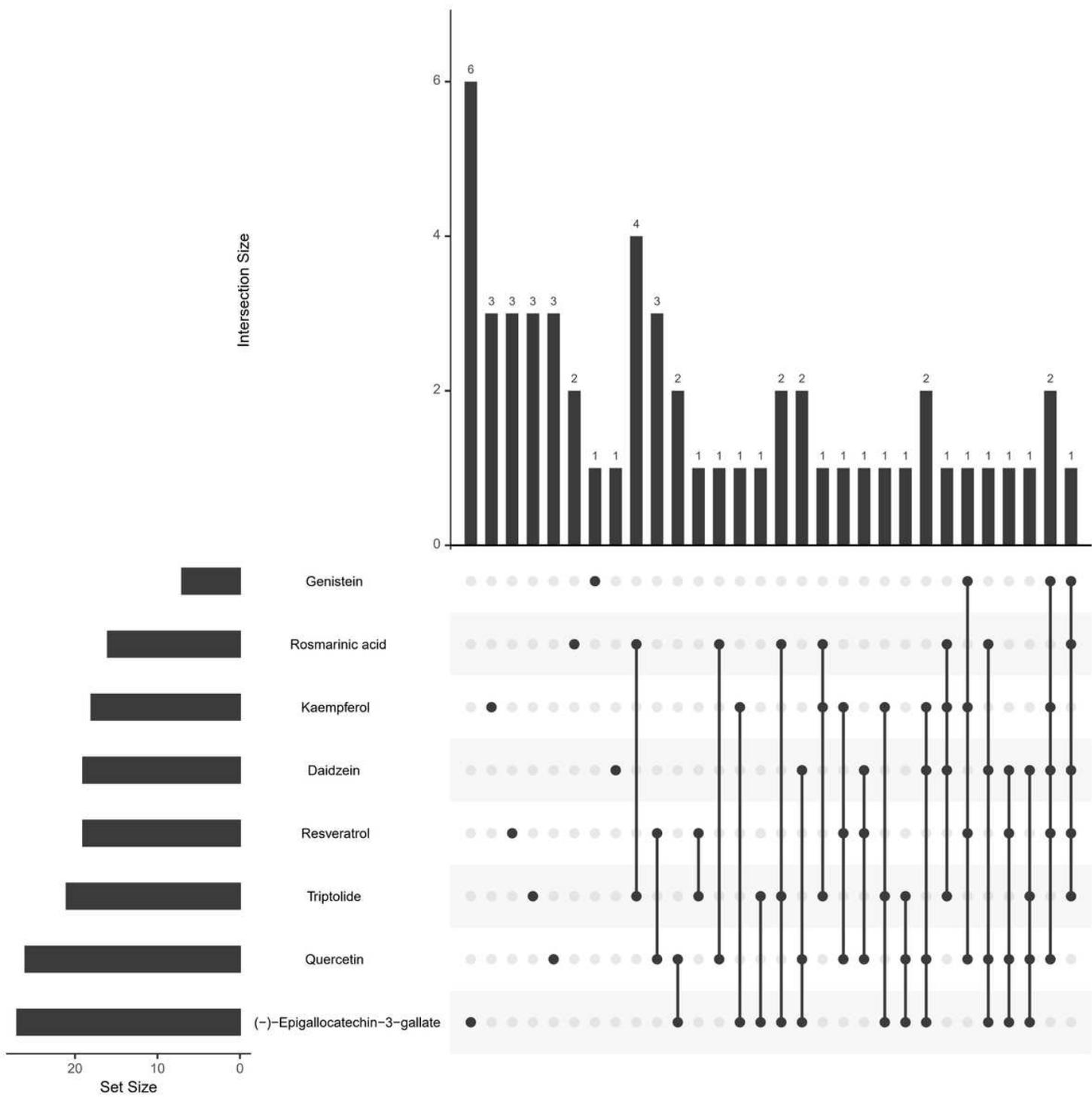


Figure 12

The intersection of the potential targets of 8 compounds



Figure 13

The result of KEGG enrichment

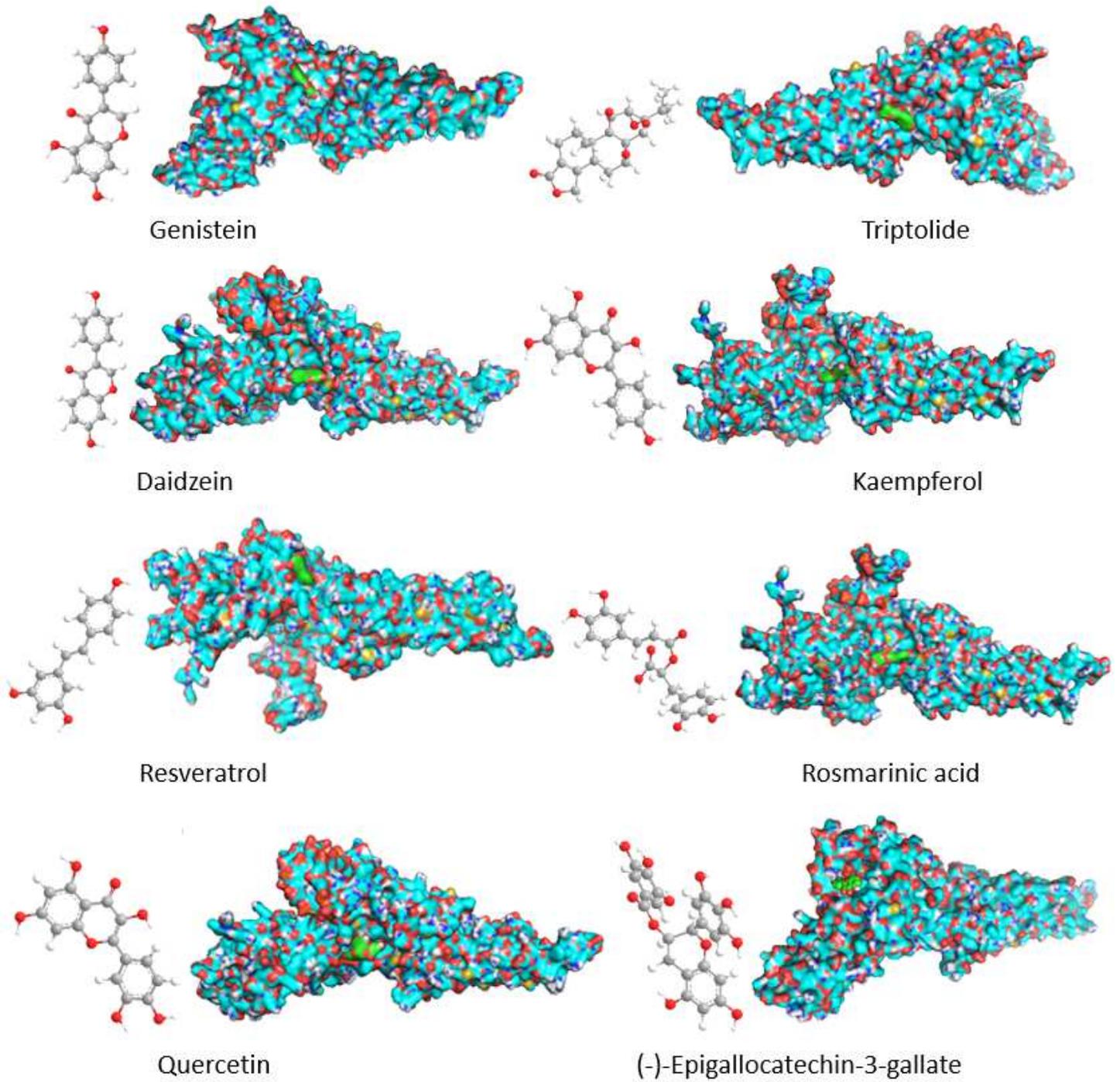


Figure 14

Interaction between STAT1 and inhibitors (genistein, daidzein, kaempferol, resveratrol, rosmarinic acid, triptolide, quercetin, and (-)-epigallocatechin-3-gallate).