# Computational Gene Expression Profiling in Non-Small Cell Lung Cancer Reveals Network-Based Immune Signatures Associated With Smoking and Overall Survival

**Rong-feng Song**
Cancer Hospital of Jiangxi Province
**Xiao-liang Cheng**
Cancer Hospital of Jiangxi Province
**Xin Zhang**
Origimed
**Jun-kai Zhang**
Affiliated zhongshan Hospital of Sun Yat-sen University
**Wen-fan Fu**
Cancer center of Guangzhou Medical University
**Kunsong Li** ( 3220993802@qq.com )
Guangdong Second Provincial General Hospital

**Research**

**Keywords:** non-small cell lung cancer, weight gene co-expression network analysis, immune, prognosis, biomarkers

**Posted Date:** August 24th, 2020

**DOI:** https://doi.org/10.21203/rs.3.rs-60971/v1

# Abstract

**Background** Lung cancer, a common risky disease, has been the leading cause of death since 2010. In this study, the prognosis of immune genes and its targeted miRNA of non-small lung cancer (NSCLC) was explored.

**Methods** Based on microarray GSE13213 (lung adenocarcinoma (LAD) patient mRNA expression profiles), microarray GSE4573 (lung squamous cell carcinoma (LSC) patient mRNA expression profiles), and microarray GSE102286 (NSCLC patient miRNA profiles), the gene co-expression network was constructed with weight gene co-expression network analysis (WGCNA). KEGG and Gene Ontology were enriched. The binding site was predicted on miRanda. The genes corresponding to immune as well as smoking and its targeted miRNA which was also related to smoking was discovered. And then the prognostic value of them was estimated and confirmed in another data cohort.

**Results** Smoking-related module from LSC had a remarkable association with immune gene set (P = 0.0001), which was not observed in LAD. KEGG and GO enrichment analysis in LSC also exhibited immunity-related pathways and functions, such as the B cell receptor signaling pathway, which was still not found in LAD. High expression of nine immune genes (CD27, CD38, CD79A, MZB1, IGHM, IGKC, IGLL3P, GUSBP11, and IGHD) from the purple module in LSC had better overall survival. In an independent LSC microarray, six of them (CD27, CD38, GUSBP11, IGKC, IGLL3P, and IGHD) were validated. IGHD related to plasma cell regulatory showed better overall survival (OS) in two datasets both. Low expression of IGHD targeted miR-29a and low infiltration of its regulated plasma cells was both associated with better OS.

**Conclusion** Our study revealed that, in LSC, smoking might trigger a more severe immune response, compared to LAD. There were six immune genes related to prognosis being discovered. And miR-29a might bind to IGHD affect OS by regulating plasma cells in LSC.


# Background

With the increasing incidence and mortality rates of cancer, lung cancer, a common but risky disease, has been the leading cause of death since 2010 [1]. In China, there are roughly 733, 000 people that are diagnosed with lung cancer per year [2]. Although clinical diagnosis techniques and chemotherapy has had considerable development, the overall 5-year survival rate still remains low [3, 4].

In line with the likelihood of metastasis and response to feasible treatment, lung cancers are clinically classified into two groups: small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC) [5], and the latter closely accounts for 85% of all lung cancers [6]. Lung adenocarcinoma (LAD) and lung squamous cell carcinoma (LSC) are the major types of NSCLC and they make up about 40% and 20 to 30% of NSCLC patients, respectively [7]. Lung cancer is induced by various factors, such as smoking history, which contribute to the high incidence of lung cancer [8]. However, the effect of cigarette smoking on different histological types of lung cancer is diverse. Previous studies discovered that smoking

cessation leading to the highest decline in the risk of LSC but the lowest reduction in LAD [9], which might suggest that LSC had more close relationship than LAD. However, the reason was still not clear. But the findings of cigarette smoking alters the immune system responses provide us some new clues. It was observed that smoking exerts profound and lasting effects on chronic inflammation and immunity [10, 11] in cancer and non-cancer lung disease, such as bladder cancer [12] and chronic obstructive pulmonary disease (COPD) [13]. Therefore, to explore the association between smoking-related gene module and immune-response signatures in lung cancer is considered desirable and valuable.

Weighted correlation network analysis (WGCNA), a kind of systematic biology method to find the gene correlation patterns based on microarray samples, is used to reveal clusters of highly correlated genes and discover the correlation of clusters with clinical traits [14]. In this study, 3 microarrays including LAD and LSC patient expression profiles, as well as clinical information, were carried out with WGCNA to cluster the modules to find genes and miRNAs which had a connection with tobacco smoking. Then, the modules that had a high correlation with smoking history were carried out with Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) pathway enrichment analysis. We investigated the prognosis related immune genes in smoking module and confirmed it in another cohort. Finally, the miRNAs associated with smoking and targeted prognosis related immune genes was screened out.

# Methods

### Data collection

The microarray GSE4573, including 130 LSC expression profiles from 129 patients, was downloaded from Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse4573). Another microarray, GSE13213 which consisted of expression profiles from 117 patients with LAD was also obtained from GEO. Moreover, 66 LSC expression profiles from microarray GSE37745 and 66 LSC miRNA expression profiles from microarray GSE102286 were downloaded from GEO. The RNA-seq miRNAs and mRNA profiles of LSC patients were from TCGA (https://portal.gdc.cancer.gov). The relative fraction of immune cell types of each LSC patient was estimated using CIBERSORT based on LM22 file [15, 16].

### WGCNA

The WGCNA packages offer R function to cluster the gene co-expression network [17]. WGCNA function “goodSamplesGenes” was used to screen out lower expressed genes, and then according to the median absolute deviation (MAD) of gene expression, the first 5000 genes with the smallest deviation were included in the following investigation (miRNAs were all involved in). Based on the expression matrix, the function “hclust” preliminarily clusters the samples in cases of obvious outliers which would result in a bias in the analysis to follow. In order to maximize scale-free topology, a soft threshold power was set with a scale free topology model fitting index $R^2 > 0.9$. The function “moduleEigengenes” calculated module eigengenes and figured out the similarity of module eigengenes to merge the parallel module with

75% similarity. Then, dynamic cutting tree was plotted by the function “plotDendroAndColors.” Clinical information, such as patient age, sex, tumor stage as well as smoking history, was also introduced to discover the relationship between modules and clinical traits. The coefficient of association between the modules and clinical traits was calculated by covariance and the significance of the coefficient of association was determined via the function “corPvalueStudent.” Finally, modules were imported into Cytoscape (Version 3.6.1) to visualize the network.

### KEGG pathways and GO enrichment

The high connectivity genes in each module were used to perform KEGG and GO enrichment analysis in Cytoscape with App ClueGO (Version 2.5.4) and CluePedia (Version 1.5.4) [18, 19]. The enriched GO in biological processes, cellular components and molecular function and KEGG network with $P < 0.05$ were visualized by Cytoscape (Version 3.6.1).

### The association analysis of immune gene set and module genes, prognostic analysis

The intersection genes of smoking module and immune gene set was screened. The correlation significance of them was calculated based on hypergeometric test. Then, the common genes of immune gene set and module genes were screened out and the prognosis value of them was estimated with log rank t test. The immune gene set consists of 547 immune related gene set and they were reported by Aaron M. Newman et al [15]. The binding site and prediction of targeted gene was performed on miRanda.

### Statistical analysis

All statistical analysis in the present study was performed in R (Version 3.5.2).

## Results

### The construction of gene co-expression networks

The co-expression network built by the WGCNA package was defined as undirected and weighted gene networks. In this network, the node was associated with gene expression and the edge was related to pairwise correlations of gene expression. Moreover, the patient clinical information also corresponded to the network. After removing the obvious outliers, the samples from 2 microarrays were clustered based on the first 5000 genes with the smallest deviation and both results are shown in Figure S1. In order to emphasize high correlations and understate low correlations, the absolute value of the correlation to a power (soft thresholding) was raised. The soft threshold selected with the recommended index ($R^2 > 0.9$) was 3 and 4, respectively, in the LSC and LAD dataset (Figure S2A&C). Figure S2B&D revealed the mean connectivity under different soft thresholds. For the LSC dataset, the mean connectivity was 34.6, and for the LAD dataset, the mean connectivity was 19.3. After determining the connectivity, the selected soft threshold was checked and the $R^2$ was 0.92 and 0.97, respectively, which indicated that the selected soft

threshold was available (Figure S3). Then, the expression profiles of selected genes from the LSC and LAD dataset were re-calculated and defined be module eigengenes of various modules. The parallel module was merged with 75% similarity. However, in our study, there was no module being merged. Figure 1A&B shows a cluster dendrogram and the merged module, in which each black line presents a gene and each color represents a module. The similarity of the 2 nodes in the commonality of the connected nodes is reflected by their topological overlap (Figure 1C&D).

## Identification of clinical trait related modules

One of the advantages of WGCNA is that the classified module can be linked to the clinical trait. Clinical traits such as age, sex, smoking history, survival and tumor stage, were introduced in this investigation. The correlation coefficient of clinical traits and modules was shown in Figure 2. The smoking related module was regarded as the main research object in the next study. The turquoise module (correlation coefficient = 0.32, $P$ = 7e-04) and purple module (correlation coefficient = -0.26, $P$ = 0.008) from LAD were positively and negatively correlated, respectively, with smoking history, while in the LSC dataset, the purple module (correlation coefficient = 0.17, $P$ = 0.048) and salmon module (correlation coefficient = -0.37, $P$ = 1e-05) were positively and negatively correlated, respectively, with smoking history. In Figure 3, the module that had a positive or negative correlation with smoking as well as module genes are shown, where nodes with a deeper red color and a larger size indicate that the gene had more related genes, and edges with a deeper red color as well as a thicker line represents that the gene pairwise had a stronger correlation. The external cycle demonstrated that 20 genes which had more related genes and the more edges with other genes. Gang Liu et al. found that cigarette smoke could reduce the expression of immune regarding genes [20]. Hence, the association of smoking-related module with immune gene set was observed, which discovered that only the purple module in LSC had significantly correlation with the immune gene set ($P$ = 0.0001). Nine immune genes (*IGHD*, *IGHM*, *IGKC*, *IGLL3P*, *GUSBP11*, *MZB1*, *CD27*, *CD38* and *CD79A*) in purple module had been pointed out with red arrows. The large majority of nine immune genes from smoking module (purple module) are involved in the regulatory of B cells memory (7/9) and B cell naïve (6/9) (Table S1). Moreover, the modules that had a positive correlation with age, sex, survival and tumor stage and module genes are also shown (Figure S4).

## KEGG pathways and GO enrichment analysis

In order to know more information about each module and to contribute to screen hub genes, KEGG pathways and GO enrichment analysis were performed. Figure 4&S5 exhibit the pathways and GO function enriched by the smoking corresponding modules of LSC and LAD, in which the same color indicates pathways or functions involved in the same group, and the size of the nodes reflects the enrichment significance of the terms. Most of immune gene (*IGHD*, *IGHM*, *IGKC*, *MZB1*, *CD27*, *CD38* and *CD79A*) mentioned above had been pointed out by a red arrow in the network. And they were all involved in positive regulation of B cell activation function (Figure 4B). Besides, pathways and functions enriched in LSC and LAD could also confirm the result mentioned above. In the LSC samples, there were functions and pathways regarding B cell being enriched, such as positive regulation of B cell activation and B cell

receptor signaling pathway (Figure 4B&C), while in LAD, no immune related pathways and functions were presented (Figure S5). Moreover, the enriched pathways and functions of other clinical trait related modules are shown in Figure S6-S7.

### Prognostic analysis of the immune genes

The correlation coefficient of nine immune gene expressions was calculated and the result showed that the expression of them had high association with the others (Figure 5). The prognosis value of nine immune genes corresponding to smoking (*CD27*, *CD38*, *CD79A*, *MZB1*, *IGHM*, *IGKC*, *IGLL3P*, *GUSBP11*, and *IGHD*) was estimated, which revealed that high expression of them had better overall survival rate (Figure 6, $P$ = 0.0076, 0.015, 0.0059, 0.042, 0.013, 0.04, 0.019, 0.0099, and 0.0056, respectively). The confidence interval of each gene was shown in Table 1. To confirm the result, overall survival analysis of nine genes in another microarray was performed. High expressions of two third genes (*CD27*, *CD38*, *GUSBP11*, *IGHD*, *IGKC*, and *IGLL3P*) were still present significant correlation with overall survival with $P$ value as 0.041, 0.04, 0.032, 0.043, 0.047 and 0.048, respectively (Figure 7) and the survival curves of the remaining genes with no significant $P$-value were shown in Figure S8.

### The prognosis related miRNA-mRNA pair

To explain the mechanism of the immune genes involved in prognostic outcomes, the miRNA cluster related smoking exposure was obtained. The result of WGNCA analysis showed that miRNAs in turquoise module was highly associated with tobacco smoking (P = 0.012, correlation coefficient = 0.46, Figure 8A). The targeted miRNAs of nine prognosis related immune gene was predicted on miRanda and shown in Figure 8B. The binding site between miR-29a and IGHD was presented in Figure 8C. The patients with low expression of miR-29a had better OS (Figure 8D, logrank P = 0.018), while based on TCGA dataset, people with miR-29a-5p low expression also present better OS (Figure 8G, logrank P = 0.022, HR = 0.67). Given that miRNA and mRNA had inhibitory regulatory relationship, the targeted miRNA-mRNA pairs had opposite expression in tumor. After observing, only IGHD and miR-29a-5p in SCLC tumor had inverse expression (Figure 8D-8E). Further, IGHD related plasma cell was also associated with OS, which indicated low infiltration of plasma predicted better OS (Figure 8H, logrank P = 0.03, HR = 0.73).

## Discussion

The lack of effective early detection methods and acquired drug resistance of chemotherapy are the major challenges in the treatment of lung cancer. In the past decade, the import effect of the tumor microenvironment (TME) in the progression of primary and secondary lung carcinoma has been known as a target enriched environment of anticancer agents [21–23]. Right now, approved drugs have been used in the clinic which target different biomarkers in the TME including immune check points and vascular endothelial growth factor (VEGF) [24]. Furthermore, the identification of novel and effective immunotherapeutic targets was focused on.

In this study, the genes from LSC and LAD patient tissues were classified into different modules by WGCNA packages, and then the modules that had significantly correlated with clinical traits, such as smoking history, were discovered. Cigarette smoking was the major cause of the majority of lung cancers [25]. The chemical components in cigarette can up-regulate the pro-inflammatory cytokines, such as IL-1 family and IL-18, resulting in the allergic airway disease asthma [26] and enhance ROS levels as well as reduce the DNA stability leaded to the noncancerous lung disease [27]. Repeated inflammation and impairing DNA stability could both induce the occurrence of cancer. Therefore, the smoking-related module in LSC and LAD was put more concern on. Bieber V et al. unraveled that auto-reactive B cell could be enhanced by cigarette smoke and its products [28]. After analyzing the smoking-related module, it was found that the smoking module of LSC had a significant correlation with the immune gene set, which was not found in LAD. The large majority of nine immune genes from smoking module (purple module) are involved in the regulatory of B cells memory (7/9) and B cell naïve (6/9). The reason might be LSC is more relevant to smoking history as compared to LAD [9, 29], which triggered to diverse immunity changes.

KEGG and GO function analysis (Fig. 4) also confirmed it. In LSC, positive regulation of B cell activation and B cell receptor signaling pathway was enriched, whereas there was no function regarding immunity being enriched. Besides, there were other pathways or functions involved in cancer occurrence and progression being enriched. Unfold protein response (UPR) was enriched in smoking positively related module (purple module). Yadav et al. had confirmed that UPR play a role in oncogenic transformation and involved in the network of oncogene and tumor suppressor genes to affect the progression of lung cancer [30]. Protein misfolding took place in endoplasmic reticulum (ER) can trigger ER stress and result in UPR to ER, which was took part in the development of various cancers [31]. KEGG and GO function analysis for smoking module also exhibited pathway or function about ER, for example protein processing in ER, retrograde protein transport, ER to cytosol, and ER chaperone complex, which might indicate tobacco smoking could promote the development of cancer by affect UPR and ER associated process.

Beside, nine immune genes (*CD27*, *CD38*, *CD79A*, *MZB1*, *IGHM*, *IGKC*, *IGLL3P*, *GUSBP11*, and *IGHD*) from smoking related module in LSC were found out and patients with high expression of them all showed better OS. Six of them (*CD27*, *CD38*, *GUSBP11*, *IGKC*, *IGLL3P*, and *IGHD*) presented similar results in validation microarray. Among them, *CD38* was found being up-regulated in life-long smokers [32] and highly expressed in lung cancers [33]. However, the remaining genes had few reports about connecting their expression with smoking. But most of them including *IGHD*, *CD27*, *CD38*, *GUSBP11*, *MZB1*, *IGHM*, and *IGKC*) had been revealed being connected with prognosis in various cancers, such as lung cancer and hepatocellular carcinoma [34–39]. *IGHD* served as tumor suppressor and correlated with recurrence in three negative breast cancer [38]. As the targeted miRNA of *IGHD*, low expression of miR-29a was also found to suggest better OS in NSCLC [40]. Besides, it was found that LSC patient with low infiltration of plasma cells showed longer OS.

Our research tried to identify prognostic immune genes of LAD and LSC, but had limitations. Firstly, the sample size of LAD and LSC from GEO was not big. Secondly, the microarray dataset used in present study lacked of normal samples so that the expression level of the chosen immune genes in healthy tissues could not be observed. Both these disadvantages could be investigate in the following research.

# Conclusions

Our study revealed that, in LSC, smoking might trigger more severe immune response, compared to LAD. And miR-29a might bind to *IGHD* affect OS by regulating plasma cell in LSC.

# List Of Abbreviations

NSCLC: non-small lung cancer; LAD: lung adenocarcinoma; LSC: lung squamous cell carcinoma; WGCNA: weight gene co-expression network analysis; OS: overall survival; SCLC: small cell lung cancer; COPD: chronic obstructive pulmonary disease; KEGG: Kyoto Encyclopedia of Genes and Genomes; GO: Gene Ontology; UPR: Unfold protein response; TME: tumor microenvironment; VEGF: vascular endothelial growth factor

# Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Availability of data and materials

The datasets analysed during the current study are available in the gene expression omnibus (GEO, https://www.ncbi.nlm.nih.gov/gds/) repository with accession number of GSE4573, GSE13213, GSE37745, GSE102286, and Xena (https://xena.ucsc.edu/)

### Competing interests

Xin Zhang is an employee of Origimed. The remaining authors declare that they have no competing interests.

### Funding

No funding was received.

### Authors' contributions

RS, XC, WF, and KL participated in the design of this study. XZ and JZ drafted the manuscript. RS, XC, XZ, and JZ revised the manuscript. RS and XC performed the statistical analysis. RS, XC, XZ, JZ, WF, and KL contributed to the retrieval of database. All authors read and approved the final manuscript.

**Acknowledgements**

None.

# References

1. Ding M, Li F, Wang B, Chi G, Liu H: A comprehensive analysis of WGCNA and serum metabolomics manifests the lung cancer-associated disordered glucose metabolism. J Cell Biochem 2019, 120:10855-10863.
2. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, Jemal A, Yu XQ, He J: Cancer statistics in China, 2015. CA Cancer J Clin 2016, 66:115-132.
3. Huang T, Yang J, Cai YD: Novel candidate key drivers in the integrative network of genes, microRNAs, methylations, and copy number variations in squamous cell lung carcinoma. Biomed Res Int 2015, 2015:358125.
4. Valdmanis PN, Roy-Chaudhuri B, Kim HK, Sayles LC, Zheng Y, Chuang CH, Caswell DR, Chu K, Zhang Y, Winslow MM, et al: Upregulation of the microRNA cluster at the Dlk1-Dio3 locus in lung adenocarcinoma. Oncogene 2015, 34:94-103.
5. Kwon MS, Shin SH, Yim SH, Lee KY, Kang HM, Kim TM, Chung YJ: CD63 as a biomarker for predicting the clinical outcomes in adenocarcinoma of lung. Lung Cancer 2007, 57:46-53.
6. Pang Y, Liu J, Li X, Xiao G, Wang H, Yang G, Li Y, Tang SC, Qin S, Du N, et al: MYC and DNMT3A-mediated DNA methylation represses microRNA-200b in triple negative breast cancer. J Cell Mol Med 2018, 22:6262-6274.
7. Ling DJ, Chen ZS, Liao QD, Feng JX, Zhang XY, Yin TY: Differential effects of MTSS1 on invasion and proliferation in subtypes of non-small cell lung cancer cells. Exp Ther Med 2016, 12:1225-1231.
8. Torre LA, Siegel RL, Jemal A: Lung Cancer Statistics. Adv Exp Med Biol 2016, 893:1-19.
9. Khuder SA, Mutgi AB: Effect of smoking cessation on major histologic types of lung cancer. Chest 2001, 120:1577-1583.
10. Lee J, Taneja V, Vassallo R: Cigarette smoking and inflammation: cellular and molecular mechanisms. J Dent Res 2012, 91:142-149.
11. Perricone C, Versini M, Ben-Ami D, Gertel S, Watad A, Segel MJ, Ceccarelli F, Conti F, Cantarini L, Bogdanos DP, et al: Smoke and autoimmunity: The fire behind the disease. Autoimmun Rev 2016, 15:354-374.
12. J MC, Duan F, Srivastava PK: Smoking-induced immune deviation contributes to progression of bladder and other cancers. Oncoimmunology 2015, 4:e1019199.

13. Brandsma CA, Hylkema MN, Geerlings M, van Geffen WH, Postma DS, Timens W, Kerstjens HA: Increased levels of (class switched) memory B cells in peripheral blood of current smokers. Respir Res 2009, 10:108.
14. Langfelder P, Horvath S: WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 2008, 9:559.
15. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA: Robust enumeration of cell subsets from tissue expression profiles. Nat Methods 2015, 12:453-457.
16. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, Porta-Pardo E, Gao GF, Plaisier CL, Eddy JA, et al: The Immune Landscape of Cancer. Immunity 2018, 48:812-830 e814.
17. Mao Q, Zhang L, Zhang Y, Dong G, Yang Y, Xia W, Chen B, Ma W, Hu J, Jiang F, Xu L: A network-based signature to predict the survival of non-smoking lung adenocarcinoma. Cancer Manag Res 2018, 10:2683-2693.
18. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, Fridman WH, Pages F, Trajanoski Z, Galon J: ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics 2009, 25:1091-1093.
19. Bindea G, Galon J, Mlecnik B: CluePedia Cytoscape plugin: pathway insights using integrated experimental and in silico data. Bioinformatics 2013, 29:661-663.
20. Li R, Yang YE, Jin J, Zhang MY, Liu X, Liu XX, Yin YH, Qu YQ: Identification of lncRNA biomarkers in lung squamous cell carcinoma using comprehensive analysis of lncRNA mediated ceRNA network. Artif Cells Nanomed Biotechnol 2019, 47:3246-3258.
21. Chen Z, Fillmore CM, Hammerman PS, Kim CF, Wong KK: Non-small-cell lung cancers: a heterogeneous set of diseases. Nat Rev Cancer 2014, 14:535-546.
22. Hanahan D, Coussens LM: Accessories to the crime: functions of cells recruited to the tumor microenvironment. Cancer Cell 2012, 21:309-322.
23. Peinado H, Zhang H, Matei IR, Costa-Silva B, Hoshino A, Rodrigues G, Psaila B, Kaplan RN, Bromberg JF, Kang Y, et al: Pre-metastatic niches: organ-specific homes for metastases. Nat Rev Cancer 2017, 17:302-317.
24. Herbst RS, Morgensztern D, Boshoff C: The biology and management of non-small cell lung cancer. Nature 2018, 553:446-454.
25. Clark SJ, Molloy PL: Smoke-Induced Changes to the Epigenome Provide Fertile Ground for Oncogenic Mutation. Cancer Cell 2017, 32:278-280.
26. Kang MJ, Homer RJ, Gallo A, Lee CG, Crothers KA, Cho SJ, Rochester C, Cain H, Chupp G, Yoon HJ, Elias JA: IL-18 is induced and IL-18 receptor alpha plays a critical role in the pathogenesis of cigarette smoke-induced pulmonary emphysema and inflammation. J Immunol 2007, 178:1948-1959.
27. Morse D, Rosas IO: Tobacco smoke-induced lung fibrosis and emphysema. Annu Rev Physiol 2014, 76:493-513.

28. Bieber V, Cohen AD, Freud T, Agmon-Levin N, Gertel S, Amital H: Autoimmune smoke and fire--coexisting rheumatoid arthritis and chronic obstructive pulmonary disease: a cross-sectional analysis. Immunol Res 2013, 56:261-266.
29. Yang CP, Gallagher RP, Weiss NS, Band PR, Thomas DB, Russell DA: Differences in incidence rates of cancers of the respiratory tract by anatomic subsite and histologic type: an etiologic implication. J Natl Cancer Inst 1989, 81:1828-1831.
30. Yadav RK, Chae SW, Kim HR, Chae HJ: Endoplasmic reticulum stress and cancer. J Cancer Prev 2014, 19:75-88.
31. Xing YH, Zhang JL, Lu L, Li DG, Wang YY, Huang S, Li CC, Zhang ZB, Li JG, Xu GS, Meng AM: Identification of specific gene modules in mouse lung tissue exposed to cigarette smoke. Asian Pac J Cancer Prev 2015, 16:4251-4256.
32. Wylam ME, Sathish V, VanOosten SK, Freeman M, Burkholder D, Thompson MA, Pabelick CM, Prakash YS: Mechanisms of Cigarette Smoke Effects on Human Airway Smooth Muscle. PLoS One 2015, 10:e0128778.
33. Ehlerding EB, England CG, Jiang D, Graves SA, Kang L, Lacognata S, Barnhart TE, Cai W: CD38 as a PET Imaging Target in Lung Cancer. Mol Pharm 2017, 14:2400-2406.
34. Kashima J, Okuma Y, Hosomi Y, Hishima T: High Serum Soluble CD27 Level Correlates with Poor Performance Status and Reduced Survival in Patients with Advanced Lung Cancer. Oncology 2019:1-8.
35. Lam JH, Ng HHM, Lim CJ, Sim XN, Malavasi F, Li H, Loh JJH, Sabai K, Kim JK, Ong CCH, et al: Expression of CD38 on Macrophages Predicts Improved Prognosis in Hepatocellular Carcinoma. Front Immunol 2019, 10:2093.
36. Cao W, Liu JN, Liu Z, Wang X, Han ZG, Ji T, Chen WT, Zou X: A three-lncRNA signature derived from the Atlas of ncRNA in cancer (TANRIC) database predicts the survival of patients with head and neck squamous cell carcinoma. Oral Oncol 2017, 65:94-101.
37. Herold T, Mulaw MA, Jurinovic V, Seiler T, Metzeler KH, Dufour A, Schneider S, Kakadia PM, Spiekermann K, Mansmann U, et al: High expression of MZB1 predicts adverse prognosis in chronic lymphocytic leukemia, follicular lymphoma and diffuse large B-cell lymphoma and is associated with a unique gene expression signature. Leuk Lymphoma 2013, 54:1652-1657.
38. Hsu HM, Chu CM, Chang YJ, Yu JC, Chen CT, Jian CE, Lee CY, Chiang YT, Chang CW, Chang YT: Six novel immunoglobulin genes as biomarkers for better prognosis in triple-negative breast cancer by gene co-expression network analysis. Sci Rep 2019, 9:4484.
39. Lohr M, Edlund K, Botling J, Hammad S, Hellwig B, Othman A, Berglund A, Lambe M, Holmberg L, Ekman S, et al: The prognostic relevance of tumour-infiltrating plasma cells and immunoglobulin kappa C indicates an important role of the humoral immune response in non-small cell lung cancer. Cancer Lett 2013, 333:222-228.
40. Joerger M, Baty F, Fruh M, Droege C, Stahel RA, Betticher DC, von Moos R, Ochsenbein A, Pless M, Gautschi O, et al: Circulating microRNA profiling in patients with advanced non-squamous NSCLC

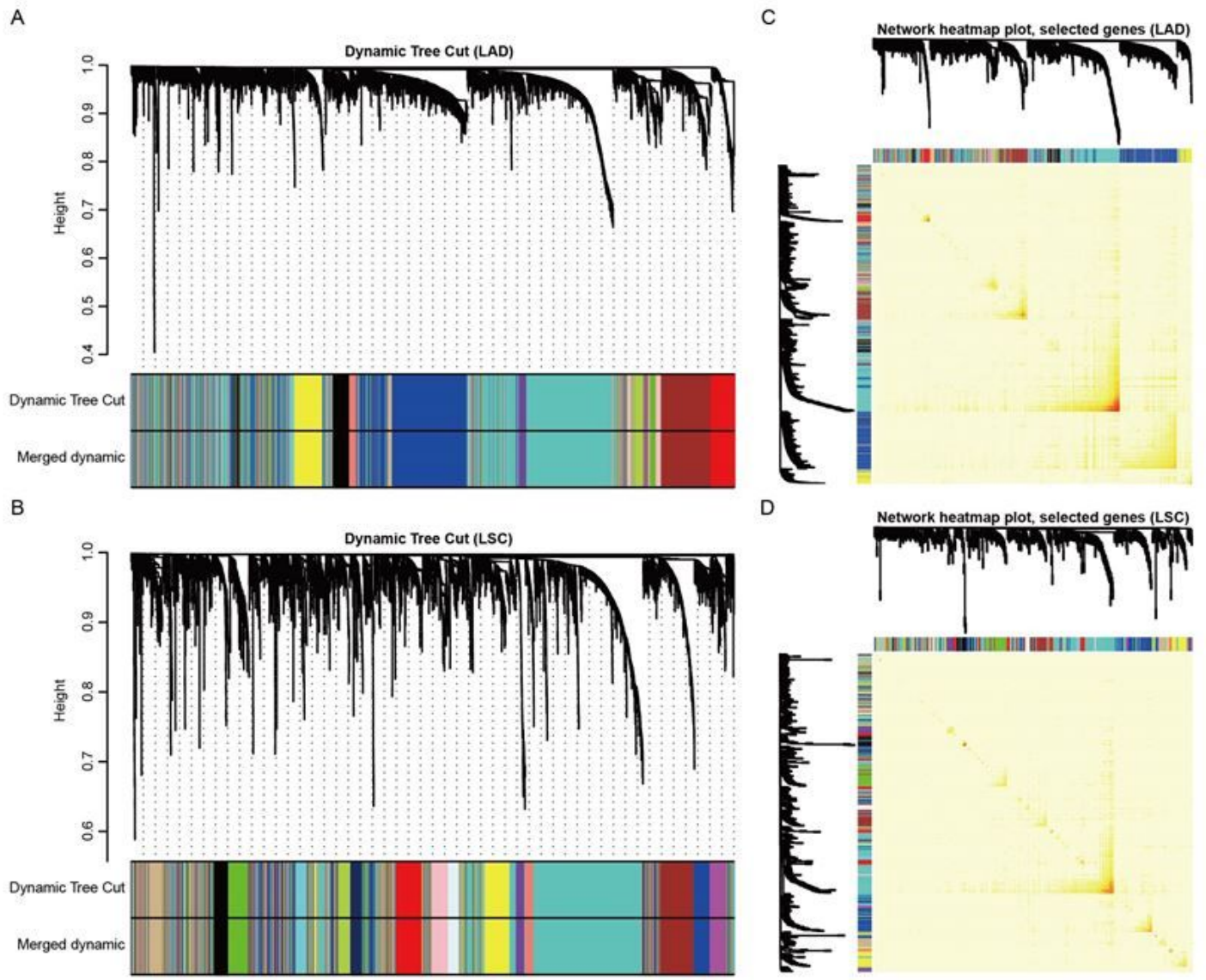
receiving bevacizumab/erlotinib followed by platinum-based chemotherapy at progression (SAKK 19/05). Lung Cancer 2014, 85:306-313.

# Tables

Table 1
Confidence interval and HR value of different group

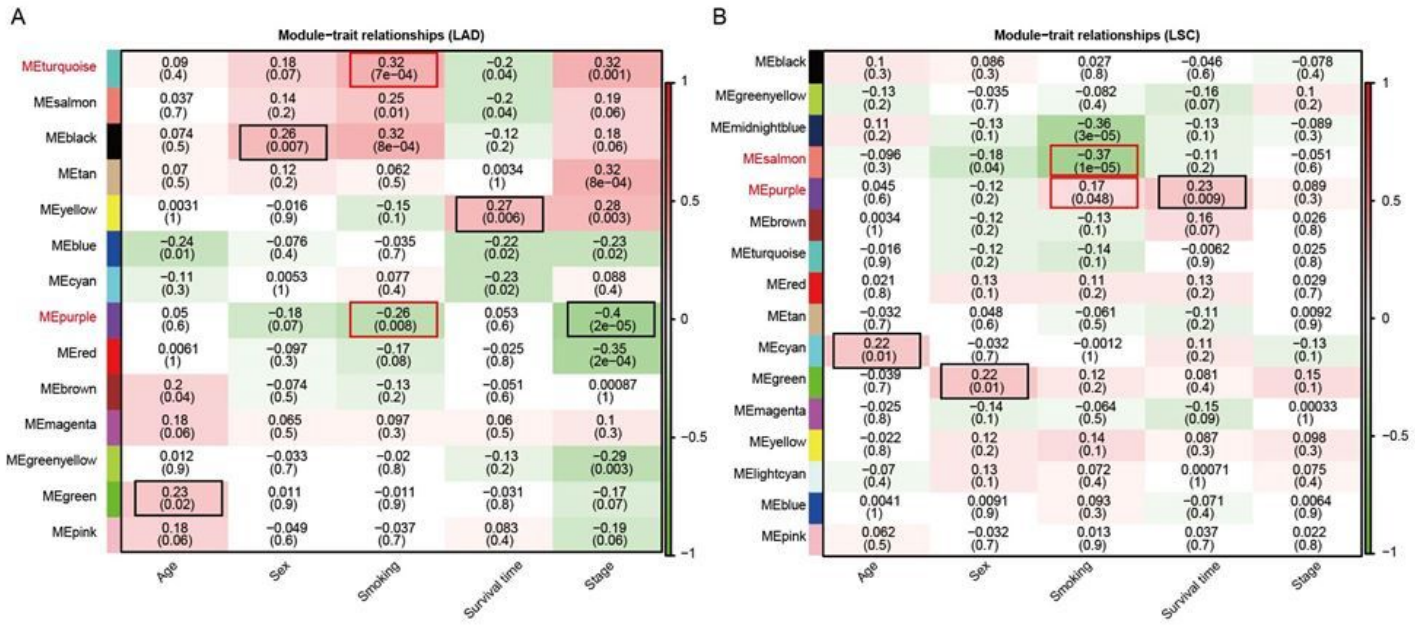
| Group | Hazard ratio(HR) | lower.95 | upper.95 |
|---|---|---|---|
| high expression of CD27 | 0.5144 | 0.3136 | 0.8439 |
| high expression of CD38 | 0.4859 | 0.268 | 0.8804 |
| high expression of CD79A | 0.5046 | 0.3077 | 0.8273 |
| high expression of IGLL3P | 0.5595 | 0.3429 | 0.9128 |
| high expression of MZB1 | 0.6065 | 0.3731 | 0.9857 |
| high expression of GUSBP11 | 0.5778 | 0.3536 | 0.9439 |
| high expression of IGHM | 0.5172 | 0.3105 | 0.8614 |
| high expression of IGKC | 0.6038 | 0.3723 | 0.9794 |
| high expression of IGHD | 0.5017 | 0.3055 | 0.824 |
| low expression of miR-29a-5p | 0.6741 | 0.4802 | 0.9463 |
| low infiltration of plasma cells | 0.7354 | 0.5563 | 0.9723 |

# Figures



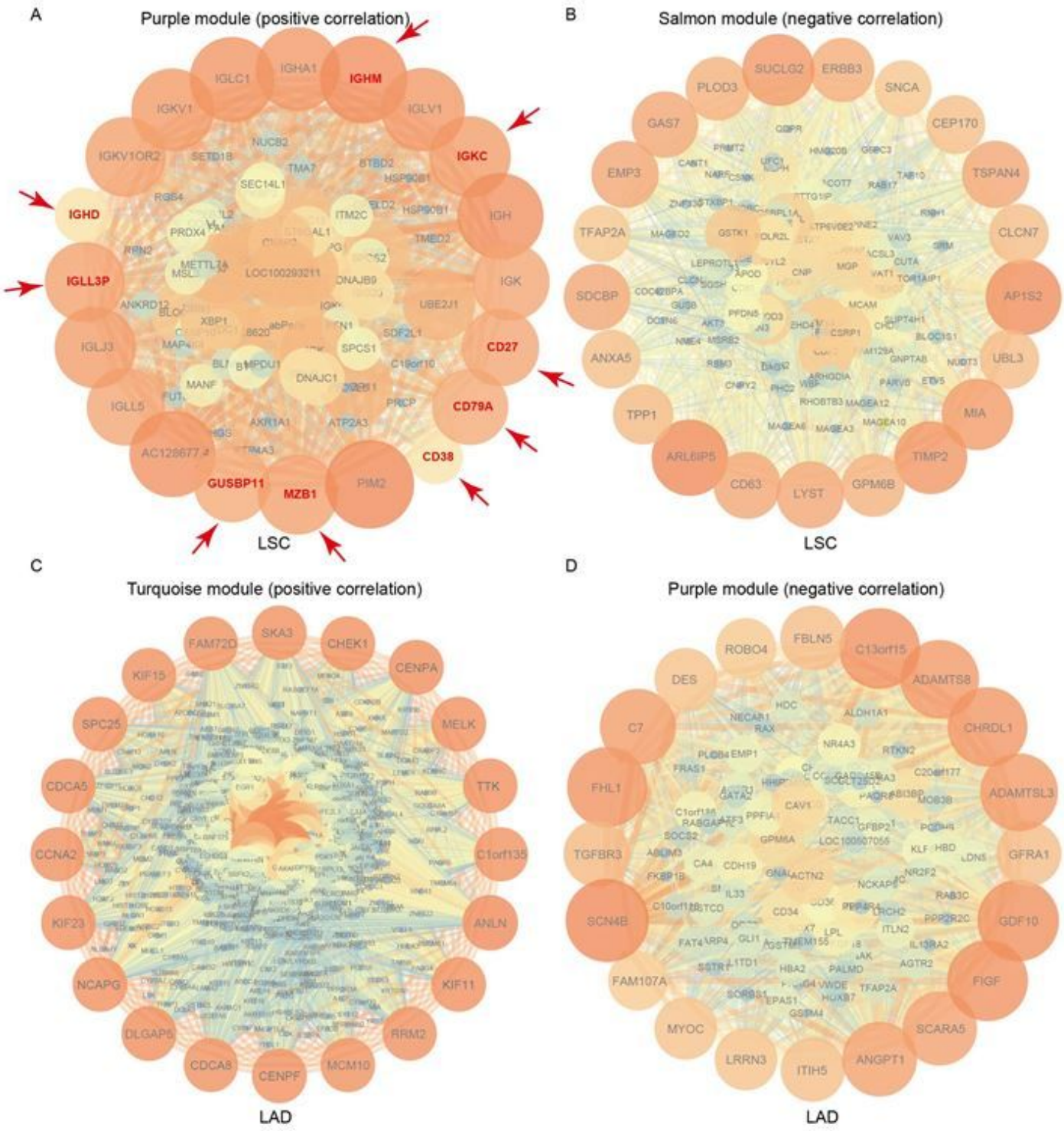

## Figure 1

The construction of gene co-expression networks. (A&B) Cluster dendrogram and merged module from the LAD and LSC dataset; (C&D) the topological overlap of various modules and samples in LAD and LSC.



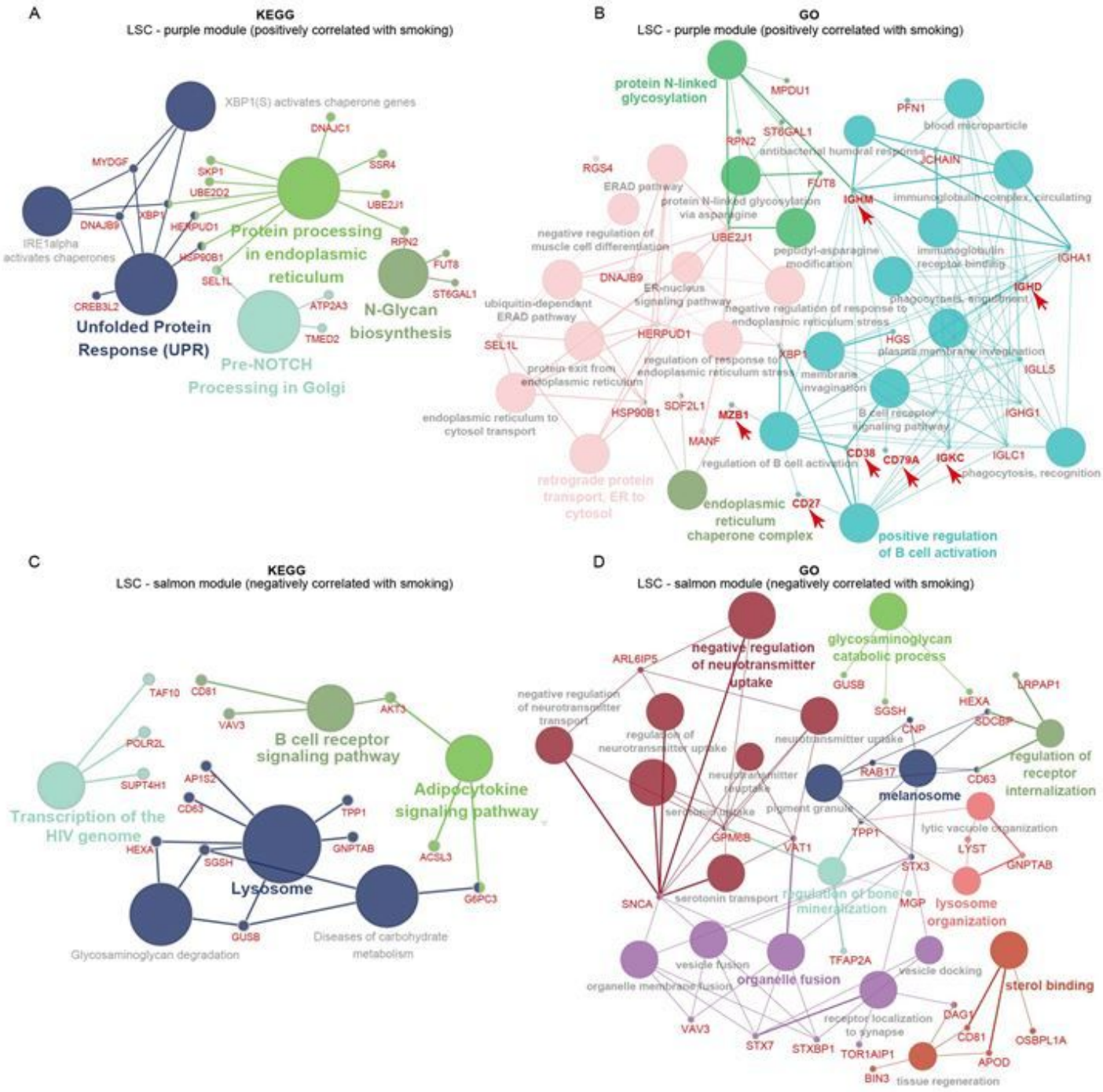

## Figure 2

Module-trait relationships in LAD (A) and LSC (B).



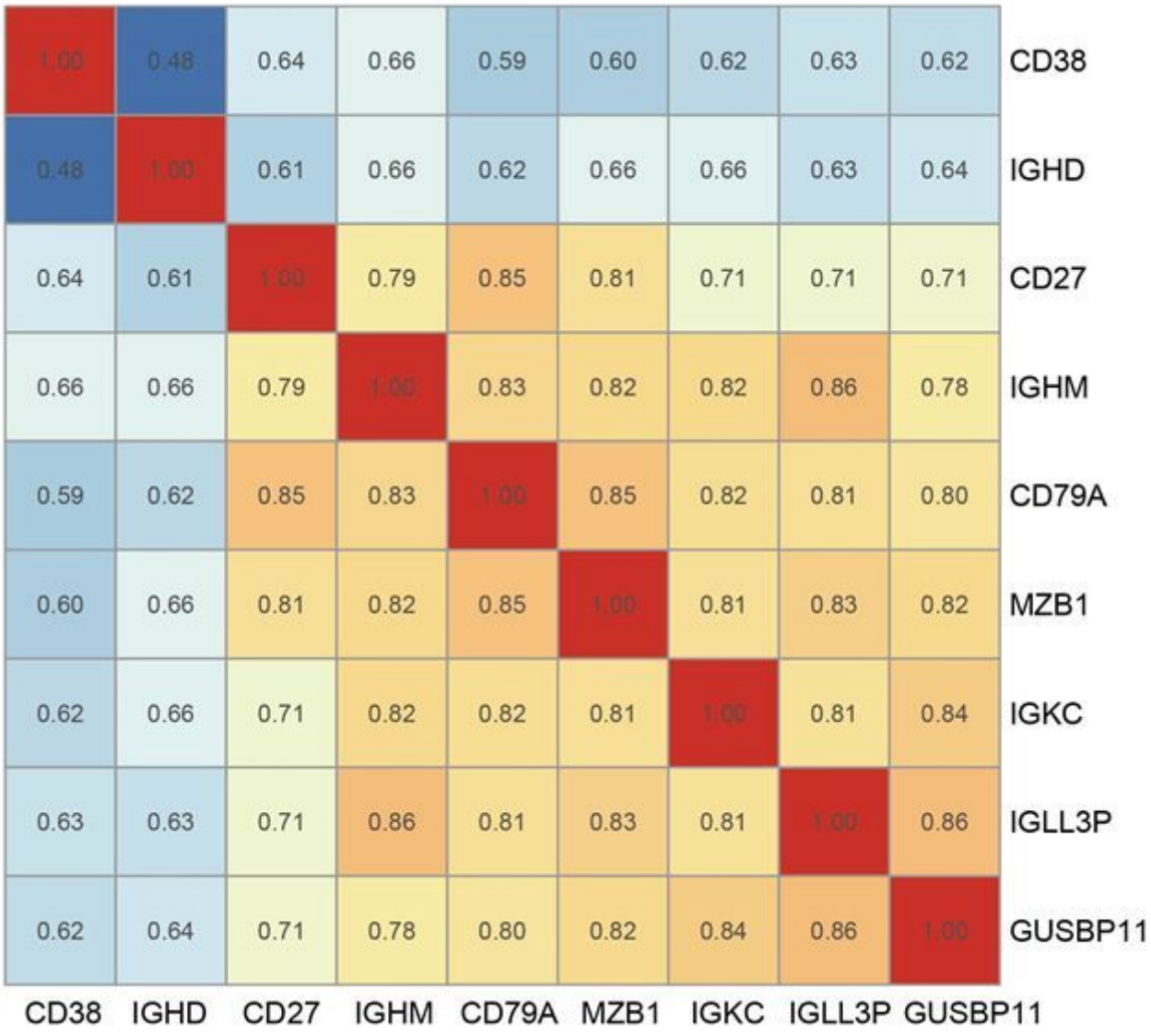

## Figure 3

The network of smoking-related modules. (A&C) The network of the purple module (LSC) and turquoise module (LAD) that were positively correlated with smoking history; (B&D) The visualization of the salmon module and the purple module that negatively correlated with smoking history.



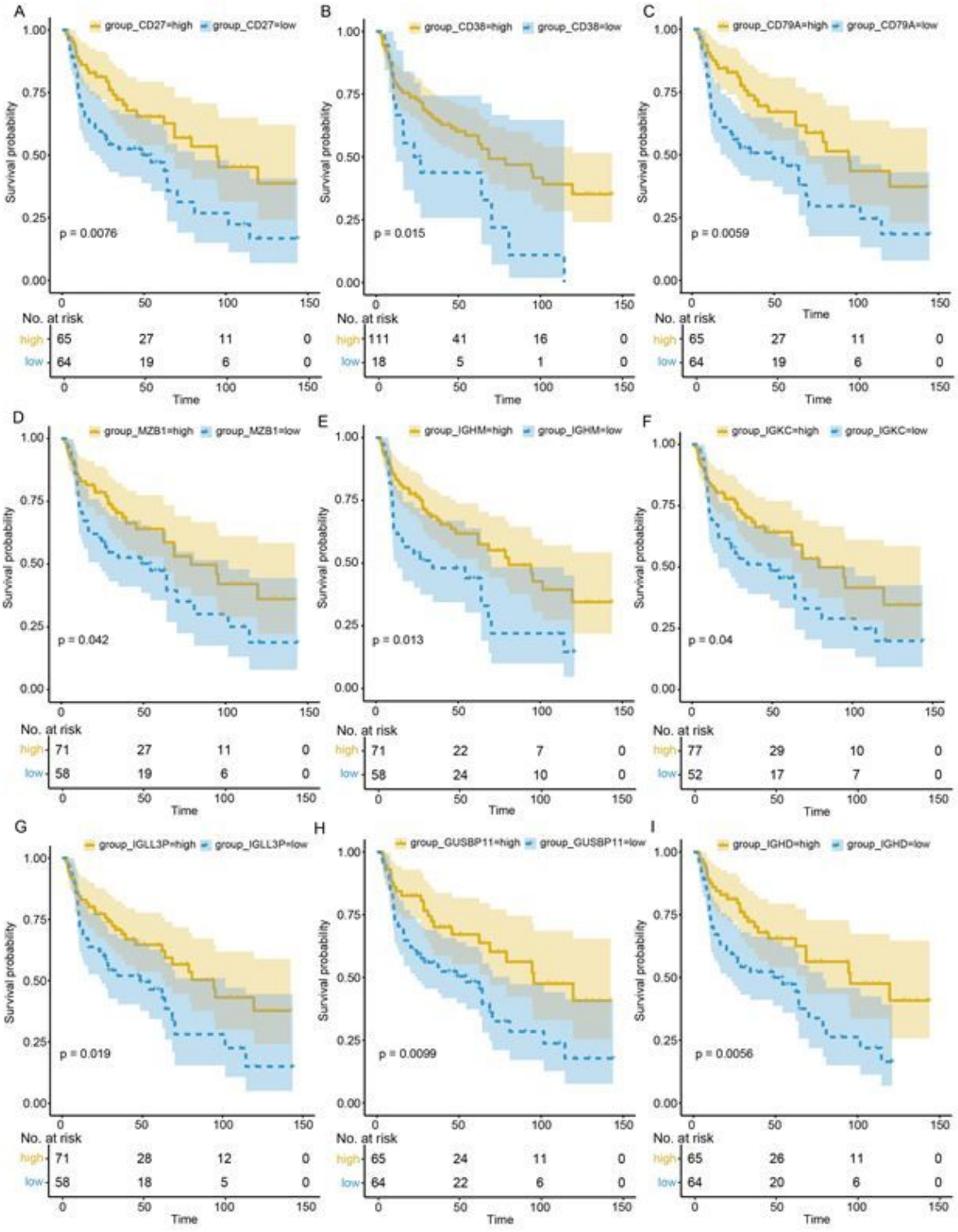

## Figure 4

KEGG and GO enrichment analysis of smoking-corresponding module in LSC. (A&C) KEGG pathways enrichment analysis was performed for genes of smoking-related modules (purple and salmon module); (B&D) GO function enrichment analysis was carried out on genes in smoking-related modules (purple and salmon module).



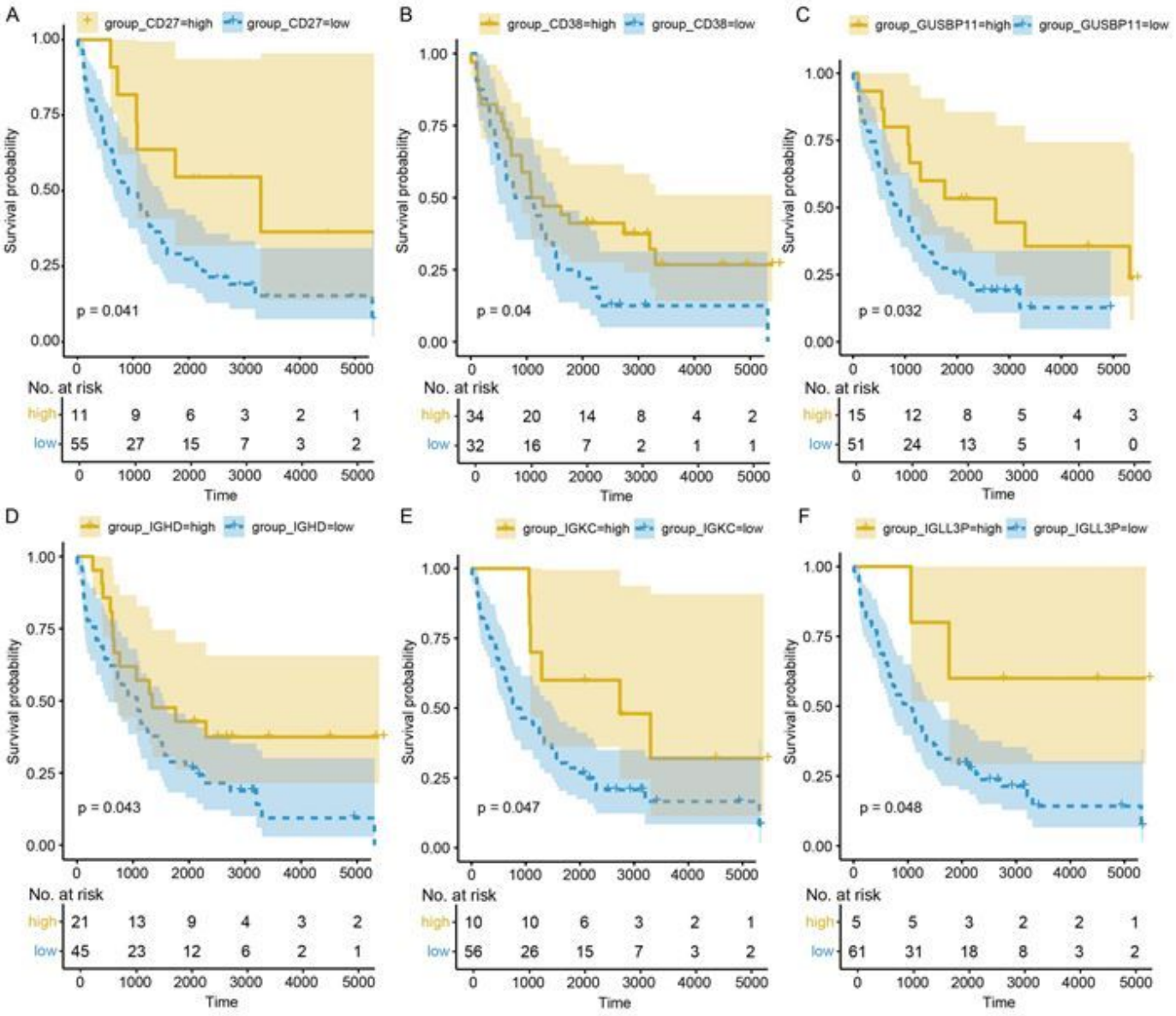

## Figure 5

Heatmap of correlation coefficient of 9 immune genes.



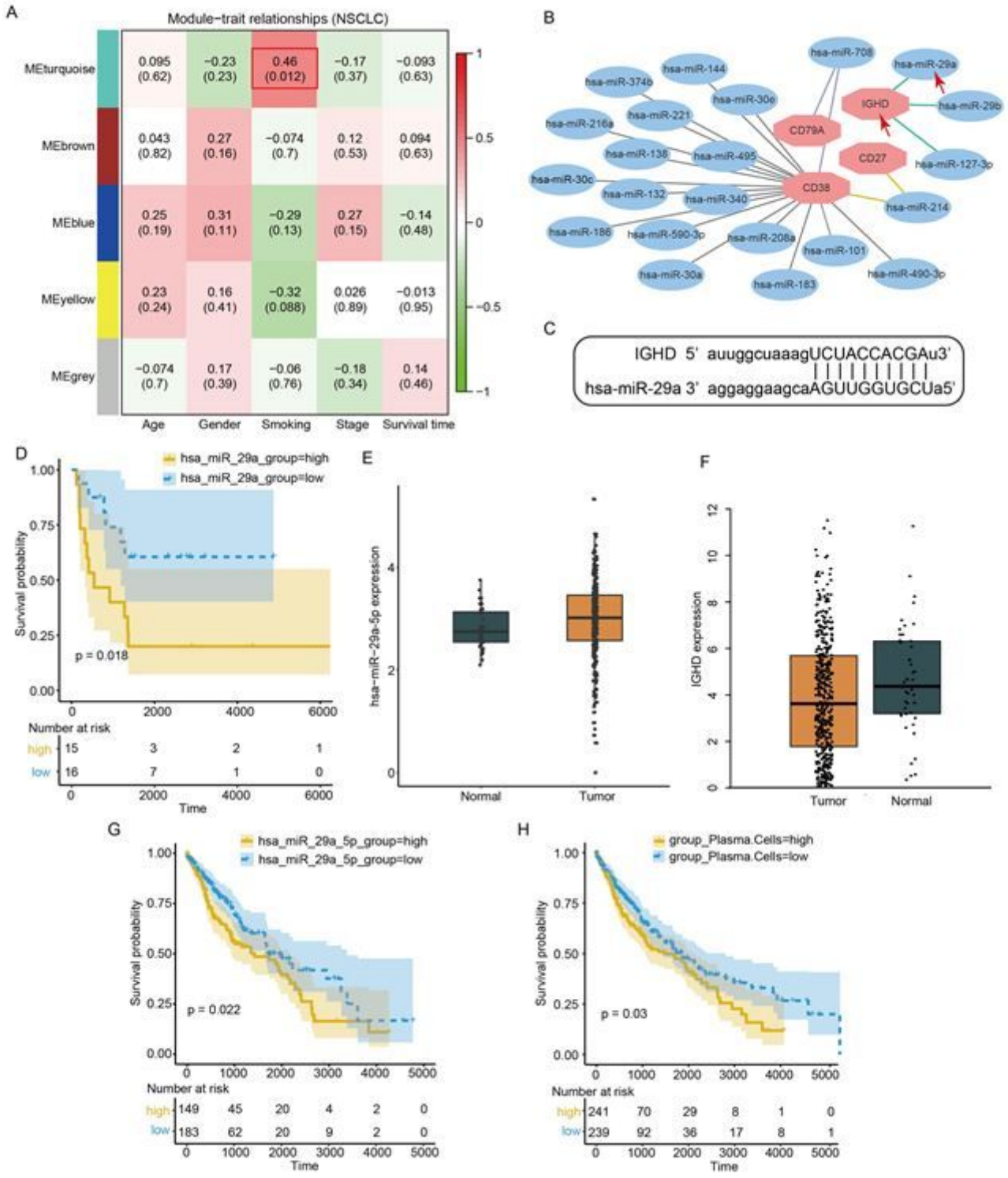

## Figure 6

Overall survival analyses of 9 immune genes in purple of microarray GSE4573.




## Figure 7

Overall survival analyses of 6 immune genes based on microarray GSE 37745.




## Figure 8

The prognosis related miRNA-mRNA pair (A) The module-trait relationships in NSCLC (based on GSE102286). (B) The prediction of miRNAs from smoking module in GSE102286 and prognosis immune gene. (C) The binding site of miR-29a and IGHD was predicated on miRanda. (D) The prognostic value of miR-29a was analyzed and low expression of it indicated better OS (based on GSE102286). (E) The expression of miR-29a-5p in LSC tumor and normal was shown (based on TCGA). (F) The expression of

IGHD in LSC tumor and normal was presented (based on TCGA). (G) The prognostic value of miR-29a-5p was evaluated and low expression of it suggested better OS (based on TCGA). (H) The prognostic value of plasma cell infiltration was assessed and low levels of it showed better OS (based on TCGA).

## Supplementary Files

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

- Supplementarymaterials.docx