

The Negative Correlation Between TNBC Mutation Frequency and Immune Function: Is Immune Involvement in Regulating Mutations?

Bo Wang

Xinjiang Medical University Affiliated First Hospital

Mengyan Li

Xinjiang Medical University Affiliated First Hospital

Chao Li

Xinjiang Medical University Affiliated First Hospital

Anna Su

Urumqi First People's Hospital, Xinjiang

Hui Wang

Xinjiang Medical University Affiliated First Hospital

Wenying Liu

Xinjiang Medical University Affiliated First Hospital

Yongmei Gao

Xinjiang Medical University Affiliated First Hospital

Yan Shi

Xinjiang Medical University Affiliated First Hospital

Yuqing Ma (✉ yuqingm0928@126.com)

Xinjiang Medical University Affiliated First Hospital <https://orcid.org/0000-0001-8009-4358>

Research article

Keywords: Triple-negative breast cancer, Gene mutation frequency, Immune cell infiltration, Prognostic model, Immune checkpoint

Posted Date: April 8th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-393949/v1>

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

[Read Full License](#)

Abstract

Introduction: Triple-negative breast cancer (TNBC) is a highly specific disease. It has been reported that TNBC has a higher gene mutation frequency than non-triple-negative breast cancer (nTNBC). Most TNBC gene mutation studies focus on BRCA1, BRCA2, TP53 and other highly mutated genes. However, few studies on extensive somatic mutations of TNBC have been conducted. Exploring the differences of TNBC mutations among individuals and the regulatory factors of gene mutations will help to find targeted treatment directions.

Method: Using bioinformatics methods, we analyzed the gene expression in patients with TNBC, clinical pathologic features and mutation data. Mutation frequency analysis was performed on each TNBC sample, and differentially expressed genes were obtained. Unsupervised cluster classified TNBC patients into high mutation subtypes and low mutation subtypes. Subsequently, the enrichment pathways among different subtypes were evaluated by KEGG enrichment analysis. The TNBC data from TCGA were used as the train group to construct the prognosis model through Univariate Cox survival analysis, Lasso regression, multivariate Cox survival analysis and stepAIC. The GEO database TNBC dataset GSE58812 was used as the test group, and the ROC curve was used to verify its accuracy.

Results: The frequency of gene mutation in TNBC samples was significantly higher than that in nTNBC samples, and there were significant differences in the frequency of mutations within TNBC. We compared the immune function among the different subtypes, and the results showed that there were significant differences in the level of immune cell infiltration and the expression of immune checkpoint markers (PD-L1, CTLA4) between the mutant subtypes. Therefore, the low mutant subtypes were more suitable for immunotherapy. Eventually, a prognostic risk model consisting of 6 mRNA (SEMA4B, IL22RA2, GPR25, TCF7, BACH2 and CLND5) is constructed. The verification is based on the calculated risk score by time-dependent receiver operating characteristic (ROC) curve analysis. The area under the curve (AUC) of the train group and test group was greater than 0.7. A survival analysis of SEMA4B using an online database showed that high expression of SEMA4B was associated with a worse prognosis ($P < 0.01$). We found that the relationship between SEMA4B expression and prognosis was inconsistent in different tumors, but all of them were positively correlated with mutation. Therefore, through extensive differences in somatic mutations, we found the key genes involved in the regulation of mutations and proved the relationship between mutations and immune cell infiltration, providing a basis for TNBC immunotherapy.

Conclusion: In this study, we investigated the pattern and prognostic role of gene mutation frequency and immune cell infiltration in TNBC, which provided a new perspective for tumor microenvironment and immunotherapy, and developed a new prognostic model for TNBC.

Introduction

TNBC, a specific subtype of breast cancer, accounts for about 10–20% of breast cancers [1,2] and is characterized by lack of estrogen receptor (ER) expression, lack of progesterone receptor (PR) expression,

and deletion of ERBB2 gene amplification [1]. TNBC is a highly heterogeneous disease [3] with higher recurrence and metastasis rates [4,5]. Standard treatment for TNBC is not targeted, so identifying biomarkers for TNBC and targeting treatment is the top priority [6–8].

Gene mutations are the basis of malignant tumors [9–11], among which mutations in TP53, BRCA1 and BRCA2 have been proven to be the leading cause of tumor progression and metastasis [12–15]. It has been reported that BRCA1, BRCA2, and TP53 gene mutation rates are higher in tnbc than in other subtypes, but the low mutation rate of the PIK3CA gene is lower than in nTNBC [16,17]. Most of the studies on mutant genes focus on the above several highly mutated genes, while there are few reports on the extensive somatic mutations of TNBC. To explore the regulatory factors and markers of somatic mutations in TNBC may be the way to distinguish TNBC subtypes further and find their therapeutic targets.

In this study, we attempted to combine the expression profile of TNBC in the TCGA database with the mutation data to find the regulatory factors of gene mutations. By further cluster analysis, TNBC was divided into different subtypes, and the differences of immune levels among different subtypes were compared. Finally, a computational framework based on the hypothesis of mutant gene regulation was developed to verify its prognostic efficacy.

Methods

Data standardization and preprocessing

Clinical features, RNA-seq expression data and somatic mutation information of female patients with breast tumors were collected from The Cancer Genome Atlas (TCGA) database (<https://portal.gdc.cancer.gov/>). We screened 173 TNBC cases with complete clinical data. Another one independent TNBC validation sets GSE58812 with 107 TNBC patients was obtained from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE58812>).

Differential expression analysis

The whole project design and data analysis process is shown in Figure 1. Somatic mutation data downloaded from TCGA database (VarScan2) [18]. We accumulated somatic mutation events in each sample and ranked the samples according to gene mutation frequency. The Limma R package was used to screen the differential genes of the 10% samples with the highest mutation frequency and the 10% samples with the lowest mutation frequency, the screening criteria were $|\log FC| > 1$, $FDR < 0.05$.

Gene Functional Enrichment Analysis

To explore the functions of 558 genes that were differentially expressed, we performed a Gene Ontology (GO) function analysis using R software with the “org.Hs.eg.db,” “clusterProfiler,” “enrichplot,” and “ggplot2” packages.

Gene mutation clusters of TNBC based on the prognostic genes

To explore the function of TNBC' prognostic related mutant regulatory gene, we classified TNBC into various clusters, use "ConsensusClusterPlus" [19] (R package, v1.54.0, 50 iterations, 80% resampling rate Euclidean distance, <http://www.bioconductor.org/>).

Gene Set Enrichment Analysis (GSEA)

To further investigate the enrichment of gene set pathways between the two TNBC subtype, GSEA was performed based on the JAVA8 platform. We selected the "c2.cp.kegg.v7.0.symbols.gmt" gene sets as reference sets, which were obtained from the MSigDB database [20].

Immune Infiltration Analysis Between Low and High gene mutation subtype

The CIBERSORT algorithm [21] was used to estimate the immune infiltration of each TNBC sample. Quantification of each immunocyte subtype among TNBC samples was based on the gene expression signatures of 22 different subtypes of immunocytes. Samples with a CIBERSORT output of $P < 0.05$ were considered to be eligible for analysis, use "barplot" R package to visualize.

Immune scores and stromal scores Between Low and High gene mutation subtype

To calculate immune scores and stromal scores, a novel ESTIMATE algorithm implemented as an "ESTIMATE" package in R was used [22]. ESTIMATE is a tool used to predict stromal/immune cell infiltration. The ESTIMATE algorithm was constructed based on single-sample gene set enrichment analysis(ssGSEA) and generated a stromal score and an immune score.

Construction and Verification of Prognostic Models

Univariate cox survival analysis was performed for differentially expressed genes, use "Survival" (R package, v3.1-12, $p < 0.01$). LASSO regression, multivariate Cox survival analysis, and stepAIC were conducted to further screening genes to reduce the number in the risk model. The final selected genes were those of the prognostic model. The computational formula of the prognostic risk model was as follows:

$$\text{RiskScore} = \sum_{i=1}^n \text{coef}(i) \times \text{gene}(i)$$

Evaluation of the prognostic signature

The TNBC patients were classified into high-risk or low-risk groups based on their prognostic risk score by using the median risk score as a cut-off point. The receiver-operating characteristic (ROC) curves were applied to evaluate the diagnostic efficacy of each clinicopathological characteristic and the prognostic signature. Univariate and multivariate Cox regression analyses were performed to evaluate whether the risk score and gene mutation frequency were independent of other clinical variables such as age, AJCC

stage, T stage and N stage in determining the prognosis of the TNBC patients. M stage was not analyzed because the data was missing for several patients. $P < 0.05$ was considered statistically significant. The Kaplan Meier survival curve and two-sided log-rank test was used to compare the overall survival (OS) at the different subtypes and groups. Stratified survival analysis was performed to examine the accuracy of the prognostic signature in predicting patient survival outcomes. Furthermore, as validation of the external dataset, we used Kaplan-Meier plotter (<http://kmplot.com/analysis/>), an online database including gene expression data and clinical data, to evaluate the prognostic significance of the gene expression of prognostic model in TNBC.

Results

TNBC gene mutations are associated with patient outcomes

The TCGA database obtained the prognosis data of 1103 breast cancer patients. Among 192 TNBC patients, 16.67% had an OS outcome ($n = 32$), with an average OS duration of 1473.25 days, 22.92% had a PFS outcome ($n = 44$) average PFS time was 993.36 days. Among 911 nTNBC patients, 13.39% had OS outcome ($n = 122$), and the mean OS time was 1708.95 days, 17.89% had PFS outcome ($n = 163$), and the mean time of PFS was 1430.38 days. We found that patients with TNBC have a worse prognosis than those with nTNBC. Is this related to genetic mutations? We further analyzed the differences in gene mutation frequency between TNBC and nTNBC patients. Breast cancer mutation data were downloaded from the TCGA database, including 173 TNBC patients and 804 nTNBC patients. The mutation frequency of the TNBC gene was significantly higher than that of nTNBC, $P < 0.001$ (Fig. 2c). The TNBC patients were ranked according to mutation frequency, and the top 10% ($n = 17$) and the last 10% ($n = 17$) of the TNBC patients had the most significant differences, with 888.82 and 20.88 mutations/sample, respectively. Therefore, we believed that these two parts of patients could better represent all TNBC patients with High and Low mutations. We used the High mutation group (HMG) and Low mutation group (LMG) to annotate them, and survival analysis showed that LMG had a better prognosis ($P = 0.028$, Fig. 1). We demonstrated that the mutation frequency of the TNBC gene was higher than that of the nTNBC gene, which was associated with prognosis. Moreover, there are also significant differences in gene mutation frequency among TNBC patients, and high mutation frequency brings a worse prognosis.

- **The LM subtype of TNBC has a better prognosis and is associated with immune function**

Subsequently, differential gene analysis was performed on these two groups of patients, and 558 regulatory genes thought to be involved in TNBC gene mutations were screened out for GO enrichment analysis. The GO analysis results show that in biological processes (BP), differential genes are mainly enriched in T cell activation, lymphocyte proliferation and mononuclear cell proliferation. In terms of cellular composition (CC), the differential genes are primarily located in the “external side of plasma membrane” and “plasma membrane signaling receptor complex.” As for the molecular function (MF), “immune receptor activity” and “cytokine receptor activity” are the most important mode (Fig. 2a, b).

Univariate Cox proportional risk regression was used to analyze the relationship between differential genes and prognosis. Twenty mutant regulatory genes were found to be significantly associated with prognosis in patients with TNBC ($P < 0.01$, Fig. 5d). The unsupervised cluster was used to divide TNBC samples into two subtypes: Cluster1 and Cluster2 (Fig. 2d, e, f). Comparison of gene mutation frequency between the two subtypes showed that Cluster1 was significantly lower than Cluster2 ($P = 0.0023$), as shown in Fig. 3a. Therefore, we believed that patients with Cluster1 were low mutation (LM) subtype, and the prognosis was better than Cluster2, as high mutation (HM) subtype ($P = 0.026$, Fig. 3b). The pathway enrichment analysis of the two subtypes (Fig. 3c, d) showing that the main enrichment pathway of HM subtype was aminoacyl tRNA biosynthesis, basal transcription factors, base excision, DNA replication, glycosylphosphatidylinositol GPI anchor biosynthesis, homologous recombination, mismatch repair, nucleotide excision repair, spliceosome and terpenoid backbone biosynthesis, these pathways are closely related to gene mutation and repair. While the LM subtype is mainly in the B cell receptor signalling pathway, Cell adhesion molecules CAMs, chemokine signalling pathway, cytokine-cytokine receptor interaction, Fc epsilon RI signalling pathway, Fc gamma R-mediated phagocytosis, JAK-STAT signalling pathway, Leukocyte transendothelial migration, Natural killer cell mediated cytotoxicity and T cell receptor signaling pathway, which are related to immune function. The results of GO and KEGG enrichment analysis verified the accuracy of cluster analysis of mutant regulatory genes, and different mutant subtypes were found to be related to immune function.

- **The LM subtype has higher immune response and immune checkpoints**

We analyzed the immune cell infiltration of normal samples ($n = 111$), HM subtypes ($n = 52$) and LM subtypes ($n = 121$) (Fig. 4a). The results showed that the immune infiltration level in normal samples was higher than that in TNBC samples. By comparing the immune cell infiltration of the HM and LM subtypes, we found that CD4T cells, CD8T cells and macrophage M1 had a higher proportion in the LM subtype. In contrast, macrophage M0, macrophage M2, and dendritic cell activated had a higher proportion in the HM subtype (Fig. 4b). Immune and matrix scores were performed for both subtypes, and the results showed that the LM subtype score was higher than that of the HM subtype (Fig. 4c, d). We further compared the expression of the immune checkpoint gene PD-L1 in the two subtypes, and the results showed that the expression of the LM subtype PD-L1 was significantly higher than that of the HM subtype (Fig. 5a). Interestingly, CTLA4, another major immune checkpoint gene, is one of the 20 prognostic genes and is significantly overexpressed in the LM subtype (Fig. 5b). Correlation analysis was conducted between 20 prognostic mutant regulatory genes and PD-L1 (Fig. 5c), and the results showed that PD-L1 was positively correlated with most of its genes. Subsequently, we analyzed the relationship between gene expression levels in different subtypes and clinicopathological data. And, the results showed that there was no significant statistical difference between AJCC stage, AGE, T stage, N stage and other clinical indicators and mutation types (Fig. 5d).

The Prognostic Model Was Constructed By Mutant Regulator Genes

We performed Lasso regression on 20 mutagenesis regulatory genes to exclude multicollinearity among genes (Fig. 6a, b). By Cox factor analysis found six more independent prognostic genes, CLDN5, GPR25, BACH2, TCF7, IL22RA2 and SEMA4B. Each TNBC patient was given a risk score using the following formula for the gene mutation prognostic model signature (GMSig):

$$\text{Risk score} = (0.797645 * \text{CLDN5}) - (0.89621 * \text{GPR25}) - (1.17745 * \text{BACH2}) - (0.4336 * \text{TCF7}) - (0.68928 * \text{IL22RA2}) + (0.541952 * \text{SEMA4B})$$
, and patients were ranked according to risk score (Fig. 6c). As a training group, 173 TNBC patients were divided into high risk (n = 87) and low risk (n = 86) groups. As the risk score increased, the number of deaths increased, and survival time gradually decreased (Fig. 6e). Survival analysis showed that the low-risk group had a better prognosis (Fig. 6d). We used an external data set, GEO database TNBC data set GSE58812, as the test group to verify the model. Survival analysis showed a significant difference in prognosis between the high-risk and low-risk groups (Fig. 6f). The ROC curve was used to evaluate the accuracy of the model, and the results showed that the AUC values of the samples from the train group and the test group in 1, 3, 4, 8, and 10 years were all higher than 0.7, indicating good prediction efficiency (Fig. 6g, h). We further compared the predictive performance of GMSig with two TNBC risk signatures: an immune signature from the Wang study (hereinafter referred to as WangSig) [23] and a 6-mRNA signature from the Lv study (hereinafter referred to as LvSig) [24], using the same cohort of TCGA patients. The AUC of GMSig in 3 years was 0.745, higher than that of LvSig (AUC = 0.707) and WangSig (AUC = 0.649) (Fig. 6i). These results suggest that GMSig has better prognostic performance in predicting survival compared to the other two mRNA signatures.

The Risk Score Is An Independent Prognostic Factor

Next, we performed a univariate and multivariate Cox regression analysis to determine if the prognostic signature was an independent prognostic factor for patients with TNBC. Univariate analyses showed that AJCC stage (P = 0.002), N stage (P = 0.03) and prognostic risk score (P < 0.001) were significantly associated with OS (Fig. 7a). Multivariate analyses showed that the AJCC stage (P = 0.002) followed by risk score (P < 0.001) were significantly associated with OS (Fig. 7b). As shown in Fig. 7c, the ROC curve analysis demonstrated that the AUC value for the prognostic risk score was 0.80, which was higher than the AUC values for age (AUC = 0.516), AJCC stage (AUC = 0.637), T stage (AUC = 0.568) and N stage (AUC = 0.665). These data demonstrate that the mutation regulator prognostic signature is an independent prognostic factor for TNBC patients. Risk grouping, gene expression, AJCC stage, T stage, N stage, age, immune score and the distribution of mutant subtypes were demonstrated by heatmap (Fig. 7f). The results showed that mutant subtypes and immune score were closely related to risk grouping. LM subtypes and high immune scores were associated with lower risk scores (Fig. 7d, e). We also performed a correlation analysis between the risk score and immune cell content to assess whether our prediction model could indicate the status of the tumor's immune microenvironment. Macrophages M0, Macrophages M2, and mast cells resting were positively correlated with the risk score. CD4 + T cells, CD8 + T cells, NK cells resting and Macrophages M1 were negatively correlated with the risk score (Fig. 8a-i).

Stratified Survival Analysis And Signature Gene Survival Analysis

Prognostic significance of risk scores based on six mutation regulator signatures in the TNBC patients stratified by different clinicopathological parameters. We stratified TNBC patients according to age, AJCCstage, T stage and N stage. Kaplan–Meier survival curve analysis showed that the OS was significantly shorter for the TNBC patients in the high-risk group than the low-risk group TNBC patients (Fig. 9a-h). Subsequently, six mutant regulatory genes in the prognostic model signature were analyzed through an online database. Sema4b was used as a risk factor in the model. Patients with high expression (n = 207) had lower OS than those with low expression (n = 71) (Fig. 10a). IL22ra2, GPR25, TCF7, and BACH2 were used as protective factors, and high expression led to a better prognosis (Fig. 10.b-e). CLND5 was unable to assess prognosis without an independent probe. These results indicate that these six genes' signature can predict TNBC patients' prognosis without considering clinical parameters.

Discussion

The occurrence, development, and treatment of breast cancer have been extensively studied in the past 20 years (25). The classification based on traditional clinicopathological staging and biomarkers can make breast cancer treatment more accurate and effective (26,27). However, as a special subtype of TNBC, the standard treatment is still single, and there is no targeted treatment plan (28). Therefore, finding biomarkers with prognostic value for TNBC is an urgent requirement for precision treatment.

Gene mutations are the basis for the occurrence and development of tumors, and also play an essential role in the process of metastasis and drug resistance of tumors [9–11]. At present, the research on the mutant genes of TNBC mainly focuses on the highly mutated pathogenic genes such as TP53, BRCA1 and BRCA2 [16,17], while there are few studies on the genes with low mutation frequency. Studies have shown that TNBC is a disease with high mutation heterogeneity. Tumors of some patients have a small number of hidden pathways and a few mutations, while other patients' tumors contain extensive mutation burden and multi-pathway involvement [29]. Therefore, we believe that single gene mutation and its regulatory factors may not be enough to reflect the impact on tumor prognosis. We hope to determine the factors and markers that regulate mutations through the different probability of gene mutation among TNBC samples.

The data of breast cancer mutations were downloaded from the TCGA database, and the frequency of mutated genes was counted for each sample. Our results showed that the mutation frequency of TNBC samples was significantly higher than that of nTNBC samples (Fig. 2c), and there were also significant differences in the mutation frequency between TNBC samples (Fig. 3a). We identified 558 differentially expressed genes that are believed to be regulatory signatures closely related to gene mutations. We used these differential genes for GO analysis, and the results showed that the enrichment in BP, CC and MF was closely related to immunity. By combining with survival data, we screened 20 genes for an unsupervised cluster. TNBC patients were divided into two subtypes, and there were significant

differences in mutation frequency between the two subtypes ($P < 0.01$). Therefore, we named them as HM subtype and LM subtype, respectively, and demonstrated that the LM subtype prognosis is better than that of the HM subtype ($P < 0.05$). Through KEGG enrichment analysis of these two subtypes, we found that the HM subtype's enrichment pathway is related to gene mutation and gene repair. In contrast, the enrichment pathway of the LM subtype is closely related to immunity. According to the enrichment results of GO and KEGG pathways, we found that there was a close relationship between mutant differences and immunity. Can we assume that immunity is an essential factor in regulating gene mutations?

We then compared the levels of immune cell infiltration between the two subtypes. Our results showed that the proportion of macrophage M1, Tregs, CD8 T cells and CD4 T cells memory activated were relatively high in the LM subtype. The abundance of macrophage M0 and macrophage M2 was relatively high in the HM subtype. The role of macrophages in tumors is complex and bidirectional. Macrophage M1 initiates cytokines' production in the tumor microenvironment and promotes tumor cell destruction [30]. At the same time, macrophage M2, especially tumor-associated macrophages, plays a vital role in tumor growth and metastasis [31,32]. Tregs maintain immune homeostasis by suppressing the immune response and antitumor effects in the tumor microenvironment [33]. Thus, anti-immunotherapy with Tregs, such as Tregs' elimination, can improve immunotherapy efficacy [34]. Activation of CD4 + T cell memory can promote the proliferation of effector T cells and enhance the anti-tumor immune response [35]. CD8 + T cells are an essential component of the tumour immune response and play a key role in killing tumor cells [36]. These results fully indicate that the LM subtype of TNBC has a more robust tumor killing function, and more Tregs may increase the chance of immune escape. The HM subtype appears to have greater immunodeficiency, leading to faster tumor progression and metastasis.

There are few studies on the correlation between somatic mutations of TNBC and immune cell infiltration. In other tumor studies, most results suggest that tumor mutation load is positively correlated with immune cell infiltration and serves as a predictor of immunotherapy [37–39]. However, our results showed a significant negative correlation between somatic mutations and immune cell infiltration in TNBC, which was similar to the effects of Anton Safonov et al [40].

According to the hypothesis of three stages between tumor and immunity [41], this theory can well explain our results. Different subtypes of TNBC are in different immune stages. The LM subtype has higher immune enrichment and immune checkpoint markers expression and has a better prognosis. In this subtype, tumor and immunity are in close balance. Cancer with an HM subtype lacking immune invasion, which has escaped immune surveillance and is no longer cleared by immune cell clones, has a worse prognosis.

These results suggest that immune checkpoint inhibitors may be an effective means of immune-monitoring in patients with immune-rich TNBC in equilibrium tumors. For TNBC patients with little or no immune infiltration, more sophisticated immunotherapy strategies may be required to reactivate the immune response against a population with clono-diverse tumors.

Through multivariate analysis, we identified mRNA signatures of gene mutation regulator (GMsig) containing six genes (Sema4b, IL22ra2, GPR25, TCF7, BACH2 and CLND5): Sema4b and CLND5 were the risk factors for TNBC; the other four mRNAs were the protective factors for TNBC. Through a literature search, no relevant studies of these six genes in TNBC were found. Sema4b, as a risk factor for TNBC, has not been widely studied in tumors. It has been reported to inhibit tumor progression in non-small cell lung cancer (NSCLC) [42]. It has also been shown that lack of Sema4b leads to reduced astrocyte proliferation [43], similar to our results. Because SEMA4B is a gene we found that may have a mutation-regulating effect, and its expression is significantly associated with prognosis in patients with TNBC (Fig. 10a). Further literature review showed that gene mutations were positively correlated with prognosis in NSCLC [44], while high mutations in astrocytoma had poor prognosis [45, 46], similar to our results. The above evidence proves that, regardless of the relationship between mutation and prognosis, SEMA4B positively correlates with mutations. Therefore, we believe that SEMA4B is involved in regulating gene mutations in various malignant tumors. The effect of SEMA4B on prognosis is consistent with the impact of mutations on the prognosis of tumors.

We established a TNBC mutation regulation model using TCGA database, verified its accuracy using the ROC curve, and verified it using external data sets. However, the current research still has some limitations. In further studies, a large sample of the clinical cohort is still needed to validate the model. In addition, basic experiments are required to verify the association between tumor mutations and immune cell infiltration in TNBC. There are few studies on SEMA4B, and we have listed some evidence to prove that SEMA4B is involved in the regulation of mutation. The mechanism of regulating gene mutation is still needed to be further studied.

Conclusion

Based on the somatic mutation frequency, we used a novel algorithm to divide TNBC into high mutation subtype and low mutation subtype, and found that there were significant differences in immune function among different subtypes, which were closely related to the prognosis. Specifically, the low mutation subtype has higher immune cell infiltration and immune checkpoint expression, which not only has a better prognosis, but also is more suitable for immunotherapy. By modeling the prognosis of gene mutations, we found the signature composed of six mRNAs, among which Sema4b is a poorly studied risk gene. Our study demonstrates that it may be involved in regulation and may be a potential target for the treatment of highly mutated subtypes.

Declarations

Acknowledgements Thanks to the public database, we can access and further study breast cancer.

Authors' contributions BW and YM participated in the design of the study. BW, ML and CL take part in the data processing. BW, AS, HW and WL were involved in the drawing. BW and YG participated in the collection of relevant literature. BW and YS participated in the writing of the paper.

Funding Not applicable.

Availability of supporting data Our data is from the TCGA database and supports shared manuscript related data.

Ethical Approval and Consent to participate

Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare that they have no competing interests.

References

1. Brown M, Tsodikov A, Bauer KR, Parise CA, Caggiano V. The role of human epidermal growth factor receptor 2 in the survival of women with estrogen and progesterone receptor-negative, invasive breast cancer: the California Cancer Registry, 1999-2004. *Cancer*. 2008 Feb 15;112(4):737-47. doi: 10.1002/cncr.23243. PMID: 18189290.
2. Acheampong T, Kehm RD, Terry MB, Argov EL, Tehranifar P. Incidence Trends of Breast Cancer Molecular Subtypes by Age and Race/Ethnicity in the US From 2010 to 2016. *JAMA Netw Open*. 2020 Aug 3;3(8):e2013226. doi: 10.1001/jamanetworkopen.2020.13226. PMID: 32804214; PMCID: PMC7431997.
3. Bianchini G, Balko JM, Mayer IA, Sanders ME, Gianni L. Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease. *Nat Rev Clin Oncol*. 2016 Nov;13(11):674-690. doi: 10.1038/nrclinonc.2016.66. Epub 2016 May 17. PMID: 27184417; PMCID: PMC5461122.
4. Dent R, Trudeau M, Pritchard KI, Hanna WM, Kahn HK, Sawka CA, Lickley LA, Rawlinson E, Sun P, Narod SA. Triple-negative breast cancer: clinical features and patterns of recurrence. *Clin Cancer Res*. 2007 Aug 1;13(15 Pt 1):4429-34. doi: 10.1158/1078-0432.CCR-06-3045. PMID: 17671126.
5. Tseng LM, Hsu NC, Chen SC, Lu YS, Lin CH, Chang DY, Li H, Lin YC, Chang HK, Chao TC, Ouyang F, Hou MF. Distant metastasis in triple-negative breast cancer. *Neoplasma*. 2013;60(3):290-4. doi: 10.4149/neo_2013_038. PMID: 23373998.
6. Jia H, Truica CI, Wang B, Wang Y, Ren X, Harvey HA, Song J, Yang JM. Immunotherapy for triple-negative breast cancer: Existing challenges and exciting prospects. *Drug Resist Updat*. 2017 May;32:1-15. doi: 10.1016/j.drug.2017.07.002. Epub 2017 Aug 19. PMID: 29145974.
7. Jin J, Tao Z, Cao J, Li T, Hu X. DNA damage response inhibitors: An avenue for TNBC treatment. *Biochim Biophys Acta Rev Cancer*. 2021 Feb 5;1875(2):188521. doi: 10.1016/j.bbcan.2021.188521. Epub ahead of print. PMID: 33556453.
8. Malone MK, Smrekar K, Park S, Blakely B, Walter A, Nasta N, Park J, Considine M, Danilova LV, Pandey NB, Fertig EJ, Popel AS, Jin K. Cytokines secreted by stromal cells in TNBC microenvironment

- as potential targets for cancer therapy. *Cancer Biol Ther*. 2020 Jun 2;21(6):560-569. doi: 10.1080/15384047.2020.1739484. Epub 2020 Mar 26. PMID: 32213106; PMCID: PMC7515526.
9. Martincorena I, Campbell PJ. Somatic mutation in cancer and normal cells. *Science*. 2015 Sep 25;349(6255):1483-9. doi: 10.1126/science.aab4082. Epub 2015 Sep 24. Erratum in: *Science*. 2016 Mar 4;351(6277). pii: aaf5401. doi: 10.1126/science.aaf5401. PMID: 26404825.
 10. Vijg J. Somatic mutations, genome mosaicism, cancer and aging. *Curr Opin Genet Dev*. 2014 Jun;26:141-9. doi: 10.1016/j.gde.2014.04.002. Epub 2014 Oct 2. PMID: 25282114; PMCID: PMC5480293.
 11. Fox EJ, Salk JJ, Loeb LA. Exploring the implications of distinct mutational signatures and mutation rates in aging and cancer. *Genome Med*. 2016 Mar 17;8(1):30. doi: 10.1186/s13073-016-0286-z. PMID: 26987311; PMCID: PMC4797182.
 12. Morton JP, Timpson P, Karim SA, Ridgway RA, Athineos D, Doyle B, Jamieson NB, Oien KA, Lowy AM, Brunton VG, Frame MC, Evans TR, Sansom OJ. Mutant p53 drives metastasis and overcomes growth arrest/senescence in pancreatic cancer. *Proc Natl Acad Sci U S A*. 2010 Jan 5;107(1):246-51. doi: 10.1073/pnas.0908428107. Epub 2009 Dec 14. PMID: 20018721; PMCID: PMC2806749.
 13. Doyle B, Morton JP, Delaney DW, Ridgway RA, Wilkins JA, Sansom OJ. p53 mutation and loss have different effects on tumourigenesis in a novel mouse model of pleomorphic rhabdomyosarcoma. *J Pathol*. 2010 Oct;222(2):129-37. doi: 10.1002/path.2748. PMID: 20662002.
 14. Deng M, Chen HH, Zhu X, Luo M, Zhang K, Xu CJ, Hu KM, Cheng P, Zhou JJ, Zheng S, Chen YD. Prevalence and clinical outcomes of germline mutations in BRCA1/2 and PALB2 genes in 2769 unselected breast cancer patients in China. *Int J Cancer*. 2019 Sep 15;145(6):1517-1528. doi: 10.1002/ijc.32184. Epub 2019 Feb 22. PMID: 30720863.
 15. Vincent-Salomon A, Bataillon G, Djerroudi L. Les cancers héréditaires du sein vus par le pathologiste [Hereditary breast carcinomas pathologist's perspective]. *Ann Pathol*. 2020 Apr;40(2):78-84. French. doi: 10.1016/j.annpat.2020.02.023. Epub 2020 Mar 30. PMID: 32241645.
 16. Stevens KN, Vachon CM, Couch FJ. Genetic susceptibility to triple-negative breast cancer. *Cancer Res*. 2013 Apr 1;73(7):2025-30. doi: 10.1158/0008-5472.CAN-12-1699. Epub 2013 Mar 27. PMID: 23536562; PMCID: PMC3654815.
 17. Millis SZ, Gatalica Z, Winkler J, Vranic S, Kimbrough J, Reddy S, O'Shaughnessy JA. Predictive Biomarker Profiling of > 6000 Breast Cancer Patients Shows Heterogeneity in TNBC, With Treatment Implications. *Clin Breast Cancer*. 2015 Dec;15(6):473-481.e3. doi: 10.1016/j.clbc.2015.04.008. Epub 2015 Apr 28. PMID: 26051240.
 18. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*. 2012 Mar;22(3):568-76. doi: 10.1101/gr.129684.111. Epub 2012 Feb 2. PMID: 22300766; PMCID: PMC3290792.
 19. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics*. 2010 Jun 15;26(12):1572-3. doi:

- 10.1093/bioinformatics/btq170. Epub 2010 Apr 28. PMID: 20427518; PMCID: PMC2881355.
20. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics*. 2011 Jun 15;27(12):1739-40. doi: 10.1093/bioinformatics/btr260. Epub 2011 May 5. PMID: 21546393; PMCID: PMC3106198.
21. 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 May;12(5):453-7. doi: 10.1038/nmeth.3337. Epub 2015 Mar 30. PMID: 25822800; PMCID: PMC4739640.
22. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, Treviño V, Shen H, Laird PW, Levine DA, Carter SL, Getz G, Stemke-Hale K, Mills GB, Verhaak RG. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun*. 2013;4:2612. doi: 10.1038/ncomms3612. PMID: 24113773; PMCID: PMC3826632.
23. Wang P, Fu Y, Chen Y, Li Q, Hong Y, Liu T, Ding Z. Nomogram Personalizes and Visualizes the Overall Survival of Patients with Triple-Negative Breast Cancer Based on the Immune Genome. *Biomed Res Int*. 2020 Nov 24;2020:4029062. doi: 10.1155/2020/4029062. PMID: 33299869; PMCID: PMC7709499.
24. Lv X, He M, Zhao Y, Zhang L, Zhu W, Jiang L, Yan Y, Fan Y, Zhao H, Zhou S, Ma H, Sun Y, Li X, Xu H, Wei M. Identification of potential key genes and pathways predicting pathogenesis and prognosis for triple-negative breast cancer. *Cancer Cell Int*. 2019 Jun 28;19:172. doi: 10.1186/s12935-019-0884-0. PMID: 31297036; PMCID: PMC6599314.
25. Cardoso JS, Silva W, Cardoso MJ. Evolution, current challenges, and future possibilities in the objective assessment of aesthetic outcome of breast cancer locoregional treatment. *Breast*. 2020 Feb;49:123-130. doi: 10.1016/j.breast.2019.11.006. Epub 2019 Nov 21. PMID: 31790958; PMCID: PMC7375658.
26. Hurvitz SA, Martin M, Symmans WF, Jung KH, Huang CS, Thompson AM, Harbeck N, Valero V, Stroyakovskiy D, Wildiers H, Campone M, Boileau JF, Beckmann MW, Afenjar K, Fresco R, Helms HJ, Xu J, Lin YG, Sparano J, Slamon D. Neoadjuvant trastuzumab, pertuzumab, and chemotherapy versus trastuzumab emtansine plus pertuzumab in patients with HER2-positive breast cancer (KRISTINE): a randomised, open-label, multicentre, phase 3 trial. *Lancet Oncol*. 2018 Jan;19(1):115-126. doi: 10.1016/S1470-2045(17)30716-7. Epub 2017 Nov 23. PMID: 29175149.
27. Prat A, Guarneri V, Paré L, Griguolo G, Pascual T, Dieci MV, Chic N, González-Farré B, Frassoldati A, Sanfeliu E, Cejalvo JM, Muñoz M, Bisagni G, Brasó-Maristany F, Urso L, Vidal M, Brandes AA, Adamo B, Musolino A, Miglietta F, Conte B, Oliveira M, Saura C, Pernas S, Alarcón J, Llombart-Cussac A, Cortés J, Manso L, López R, Ciruelos E, Schettini F, Villagrasa P, Carey LA, Perou CM, Piacentini F, D'Amico R, Tagliafico E, Parker JS, Conte P. A multivariable prognostic score to guide systemic therapy in early-stage HER2-positive breast cancer: a retrospective study with an external evaluation. *Lancet Oncol*. 2020 Nov;21(11):1455-1464. doi: 10.1016/S1470-2045(20)30450-2. PMID: 33152285.
28. Wang X, Wang SS, Huang H, Cai L, Zhao L, Peng RJ, Lin Y, Tang J, Zeng J, Zhang LH, Ke YL, Wang XM, Liu XM, Chen QJ, Zhang AQ, Xu F, Bi XW, Huang JJ, Li JB, Pang DM, Xue C, Shi YX, He ZY, Lin HX,

- An X, Xia W, Cao Y, Guo Y, Su YH, Hua X, Wang XY, Hong RX, Jiang KK, Song CG, Huang ZZ, Shi W, Zhong YY, Yuan ZY; South China Breast Cancer Group (SCBCG). Effect of Capecitabine Maintenance Therapy Using Lower Dosage and Higher Frequency vs Observation on Disease-Free Survival Among Patients With Early-Stage Triple-Negative Breast Cancer Who Had Received Standard Treatment: The SYSUCC-001 Randomized Clinical Trial. *JAMA*. 2021 Jan 5;325(1):50-58. doi: 10.1001/jama.2020.23370. PMID: 33300950; PMCID: PMC7729589.
29. Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao Y, Turashvili G, Ding J, Tse K, Haffari G, Bashashati A, Prentice LM, Khattra J, Burleigh A, Yap D, Bernard V, McPherson A, Shumansky K, Crisan A, Giuliany R, Heravi-Moussavi A, Rosner J, Lai D, Birol I, Varhol R, Tam A, Dhalla N, Zeng T, Ma K, Chan SK, Griffith M, Moradian A, Cheng SW, Morin GB, Watson P, Gelmon K, Chia S, Chin SF, Curtis C, Rueda OM, Pharoah PD, Damaraju S, Mackey J, Hoon K, Harkins T, Tadigotla V, Sigaroudinia M, Gascard P, Tlsty T, Costello JF, Meyer IM, Eaves CJ, Wasserman WW, Jones S, Huntsman D, Hirst M, Caldas C, Marra MA, Aparicio S. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature*. 2012 Apr 4;486(7403):395-9. doi: 10.1038/nature10933. PMID: 22495314; PMCID: PMC3863681.
30. Ngambenjawong C, Gustafson HH, Pun SH. Progress in tumor-associated macrophage (TAM)-targeted therapeutics. *Adv Drug Deliv Rev*. 2017 May 15;114:206-221. doi: 10.1016/j.addr.2017.04.010. Epub 2017 Apr 25. PMID: 28449873; PMCID: PMC5581987.
31. Owen JL, Mohamadzadeh M. Macrophages and chemokines as mediators of angiogenesis. *Front Physiol*. 2013 Jul 5;4:159. doi: 10.3389/fphys.2013.00159. PMID: 23847541; PMCID: PMC3701799.
32. Baay M, Brouwer A, Pauwels P, Peeters M, Lardon F. Tumor cells and tumor-associated macrophages: secreted proteins as potential targets for therapy. *Clin Dev Immunol*. 2011;2011:565187. doi: 10.1155/2011/565187. Epub 2011 Nov 17. PMID: 22162712; PMCID: PMC3227419.
33. Lee GR. Phenotypic and Functional Properties of Tumor-Infiltrating Regulatory T Cells. *Mediators Inflamm*. 2017;2017:5458178. doi: 10.1155/2017/5458178. Epub 2017 Dec 31. PMID: 29463952; PMCID: PMC5804416.
34. Sharma N, Vacher J, Allison JP. TLR1/2 ligand enhances antitumor efficacy of CTLA-4 blockade by increasing intratumoral Treg depletion. *Proc Natl Acad Sci U S A*. 2019 May 21;116(21):10453-10462. doi: 10.1073/pnas.1819004116. Epub 2019 May 10. PMID: 31076558; PMCID: PMC6534983.
35. Vahidi Y, Faghieh Z, Talei AR, Doroudchi M, Ghaderi A. Memory CD4+ T cell subsets in tumor draining lymph nodes of breast cancer patients: A focus on T stem cell memory cells. *Cell Oncol (Dordr)*. 2018 Feb;41(1):1-11. doi: 10.1007/s13402-017-0352-6. Epub 2017 Oct 9. PMID: 28994018.
36. Durgeau A, Virk Y, Corgnac S, Mami-Chouaib F. Recent Advances in Targeting CD8 T-Cell Immunity for More Effective Cancer Immunotherapy. *Front Immunol*. 2018 Jan 22;9:14. doi: 10.3389/fimmu.2018.00014. PMID: 29403496; PMCID: PMC5786548.
37. Linxweiler M, Kuo F, Katabi N, Lee M, Nadeem Z, Dalin MG, Makarov V, Chowell D, Dogan S, Ganly I, Hakimi AA, Wong RJ, Riaz N, Ho AL, Chan TA, Morris LGT. The Immune Microenvironment and Neoantigen Landscape of Aggressive Salivary Gland Carcinomas Differ by Subtype. *Clin Cancer Res*.

- 2020 Jun 15;26(12):2859-2870. doi: 10.1158/1078-0432.CCR-19-3758. Epub 2020 Feb 14. PMID: 32060100; PMCID: PMC7918996.
38. Wu Z, Wang M, Liu Q, Liu Y, Zhu K, Chen L, Guo H, Li Y, Shi B. Identification of gene expression profiles and immune cell infiltration signatures between low and high tumor mutation burden groups in bladder cancer. *Int J Med Sci.* 2020 Jan 1;17(1):89-96. doi: 10.7150/ijms.39056. PMID: 31929742; PMCID: PMC6945555.
39. Zhu J, Zhang T, Li J, Lin J, Liang W, Huang W, Wan N, Jiang J. Association Between Tumor Mutation Burden (TMB) and Outcomes of Cancer Patients Treated With PD-1/PD-L1 Inhibitions: A Meta-Analysis. *Front Pharmacol.* 2019 Jun 14;10:673. doi: 10.3389/fphar.2019.00673. PMID: 31258479; PMCID: PMC6587434.
40. Safonov A, Jiang T, Bianchini G, Györfy B, Karn T, Hatzis C, Pusztai L. Immune Gene Expression Is Associated with Genomic Aberrations in Breast Cancer. *Cancer Res.* 2017 Jun 15;77(12):3317-3324. doi: 10.1158/0008-5472.CAN-16-3478. Epub 2017 Apr 20. PMID: 28428277.
41. Mittal D, Gubin MM, Schreiber RD, Smyth MJ. New insights into cancer immunoediting and its three component phases—elimination, equilibrium and escape. *Curr Opin Immunol.* 2014 Apr;27:16-25. doi: 10.1016/j.coi.2014.01.004. Epub 2014 Feb 14. PMID: 24531241; PMCID: PMC4388310.
42. Jian H, Zhao Y, Liu B, Lu S. SEMA4B inhibits growth of non-small cell lung cancer in vitro and in vivo. *Cell Signal.* 2015 Jun;27(6):1208-13. doi: 10.1016/j.cellsig.2015.02.027. Epub 2015 Mar 4. PMID: 25746385.
43. Ben-Gigi L, Sweetat S, Besser E, Fellig Y, Wiederhold T, Polakiewicz RD, Behar O. Astrogliosis Induced by Brain Injury Is Regulated by Sema4B Phosphorylation. *eNeuro.* 2015 May 25;2(3):ENEURO.0078-14.2015. doi: 10.1523/ENEURO.0078-14.2015. PMID: 26464987; PMCID: PMC4586933.
44. Hurkmans DP, Kuipers ME, Smit J, van Marion R, Mathijssen RHJ, Postmus PE, Hiemstra PS, Aerts JGJV, von der Thüsen JH, van der Burg SH. Tumor mutational load, CD8+ T cells, expression of PD-L1 and HLA class I to guide immunotherapy decisions in NSCLC patients. *Cancer Immunol Immunother.* 2020 May;69(5):771-777. doi: 10.1007/s00262-020-02506-x. Epub 2020 Feb 12. PMID: 32047958; PMCID: PMC7183487.
45. Mirchia K, Sathe AA, Walker JM, Fudym Y, Galbraith K, Viapiano MS, Corona RJ, Snuderl M, Xing C, Hatanpaa KJ, Richardson TE. Total copy number variation as a prognostic factor in adult astrocytoma subtypes. *Acta Neuropathol Commun.* 2019 Jun 10;7(1):92. doi: 10.1186/s40478-019-0746-y. Erratum in: *Acta Neuropathol Commun.* 2019 Aug 14;7(1):131. PMID: 31177992; PMCID: PMC6556960.
46. Richardson TE, Sathe AA, Kanchwala M, Jia G, Habib AA, Xiao G, Snuderl M, Xing C, Hatanpaa KJ. Genetic and Epigenetic Features of Rapidly Progressing IDH-Mutant Astrocytomas. *J Neuropathol Exp Neurol.* 2018 Jul 1;77(7):542-548. doi: 10.1093/jnen/nly026. PMID: 29741737; PMCID: PMC6005148.

Figures

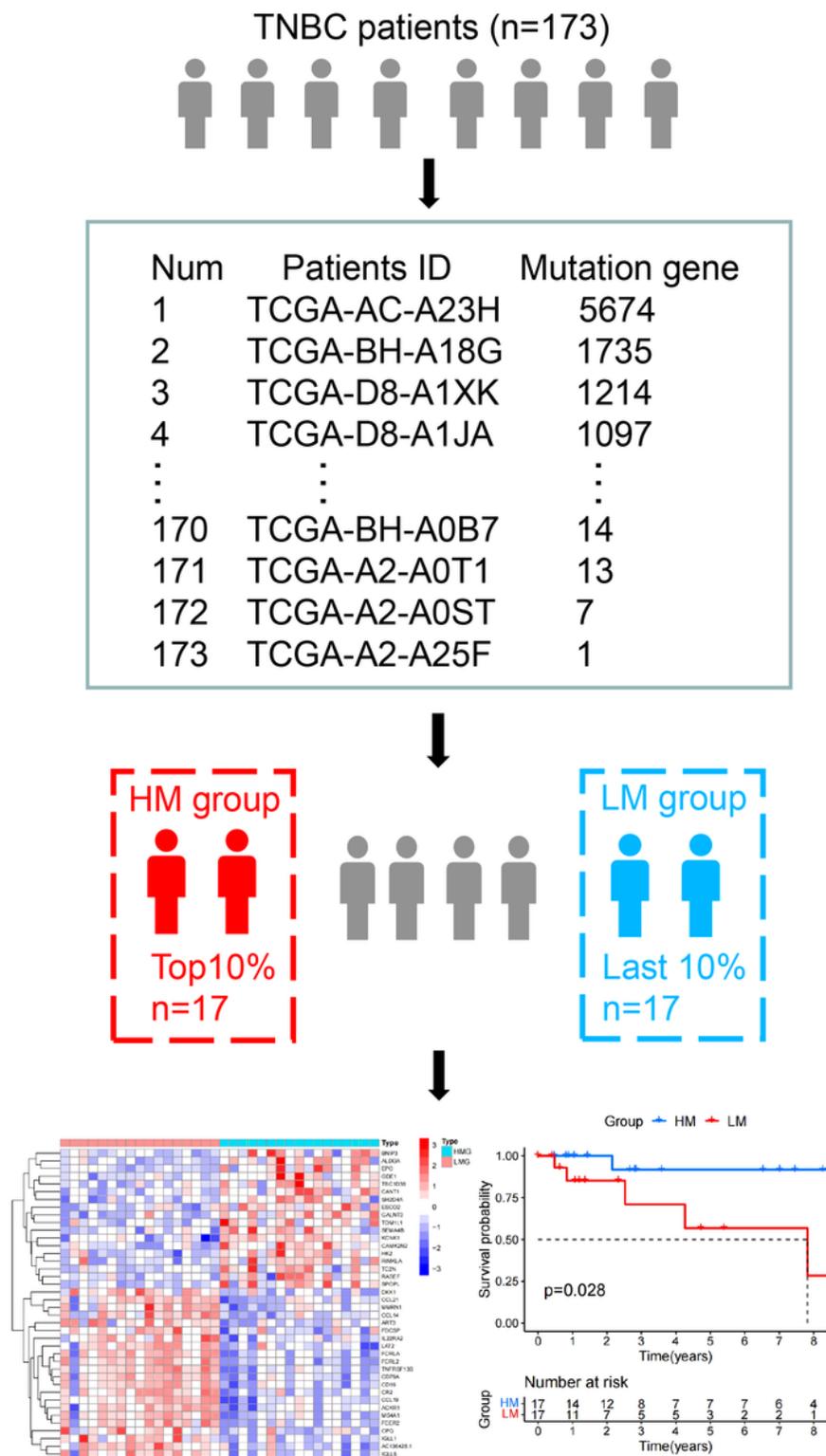


Figure 1

Computational overview of mutations regulate differential genes detection. Patients were ranked according to the frequency of genetic mutations. Samples were divided into two groups HM group (patients' mutator phenotype ranked in the top 10%) and LM group (patients' mutator phenotype ranked in the last 10%), according to their mutator phenotype. Gene mutation regulator mRNAs were detected by comparing the gene expression profile between HM group and LM group.

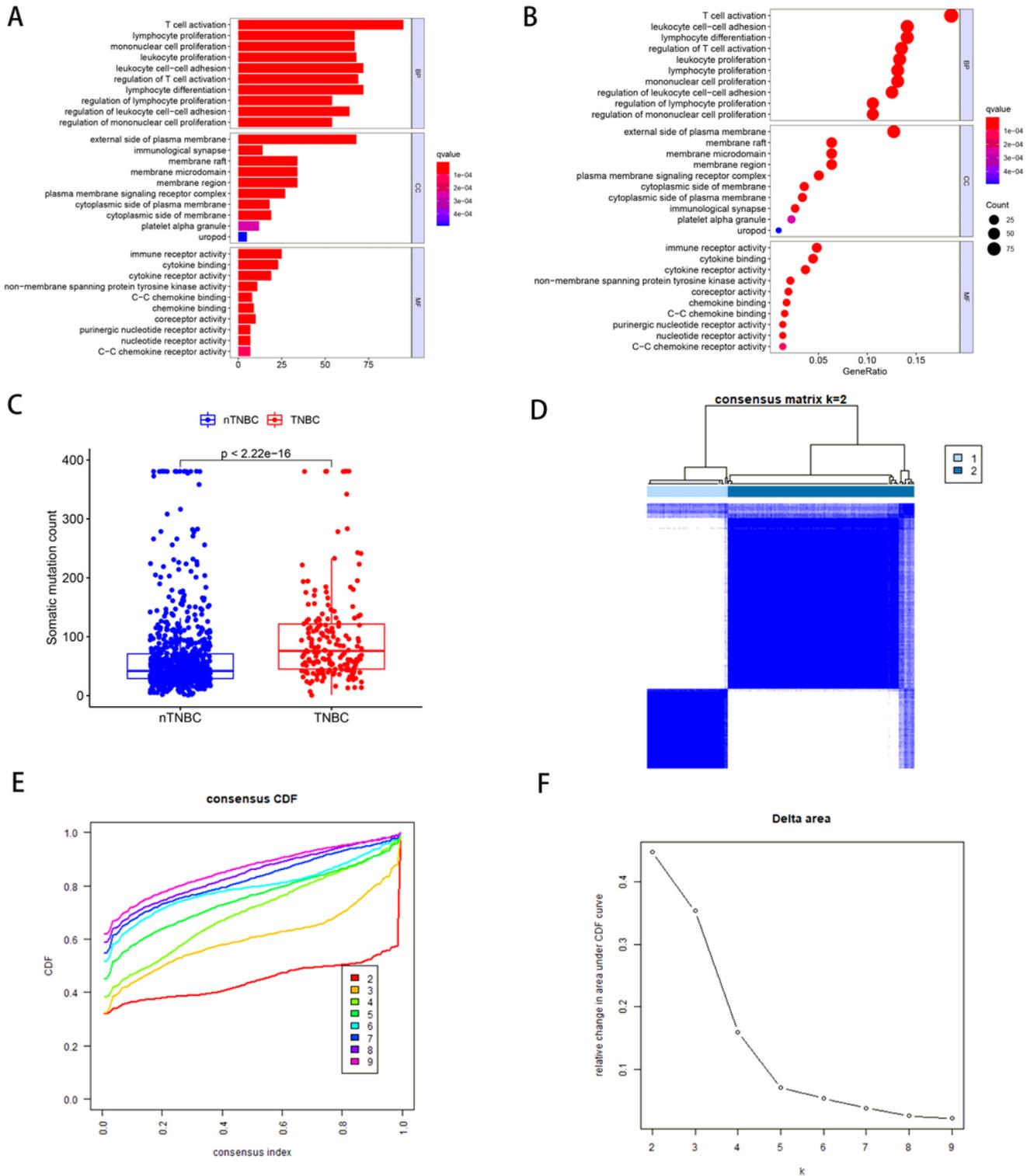


Figure 2

Functional enrichment analysis and unsupervised clustering of TNBC were performed by differential genes. a, b Gene Ontology (GO) analysis results show the enriched at biological processes, cell components and molecular functions associated with the gene mutation differential mRNA. c Difference of gene mutation frequency between TNBC and nTNBC in TCGA. d Consensus matrix when k = 2. Both

the rows and columns of the matrix represent samples. From white to dark blue, the value of the consistency matrix is from 0 to 1. e Consensus CDF. f Delta area for validation of the clustering results.

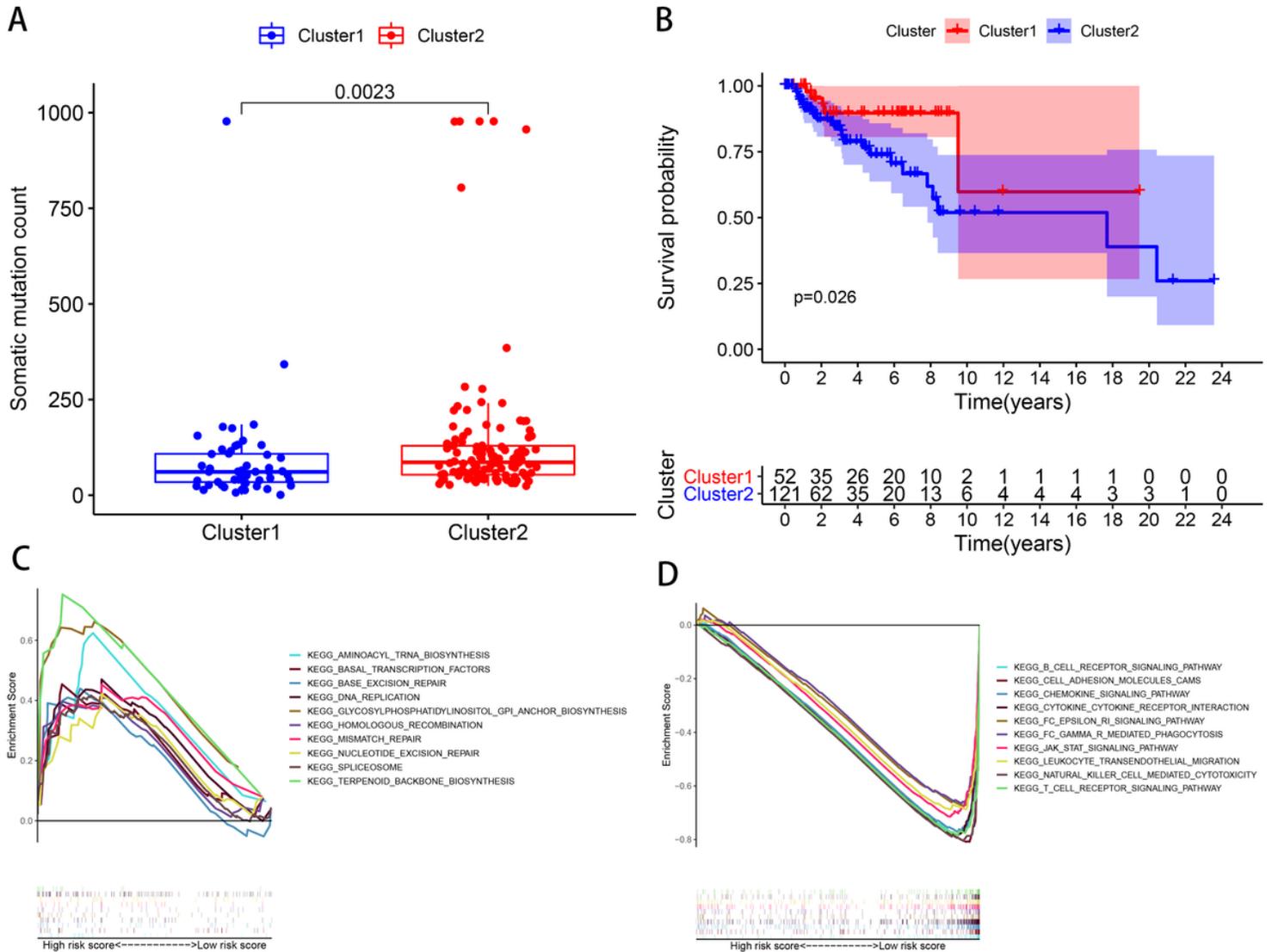


Figure 3

Mutation frequency, enriched pathway and prognostic difference among different subtypes were analyzed. a The variation of mutation frequency among different subtypes. b Prognostic differences among different subtypes. c, d The HM subtype and LM subtype mainly enriched pathways.

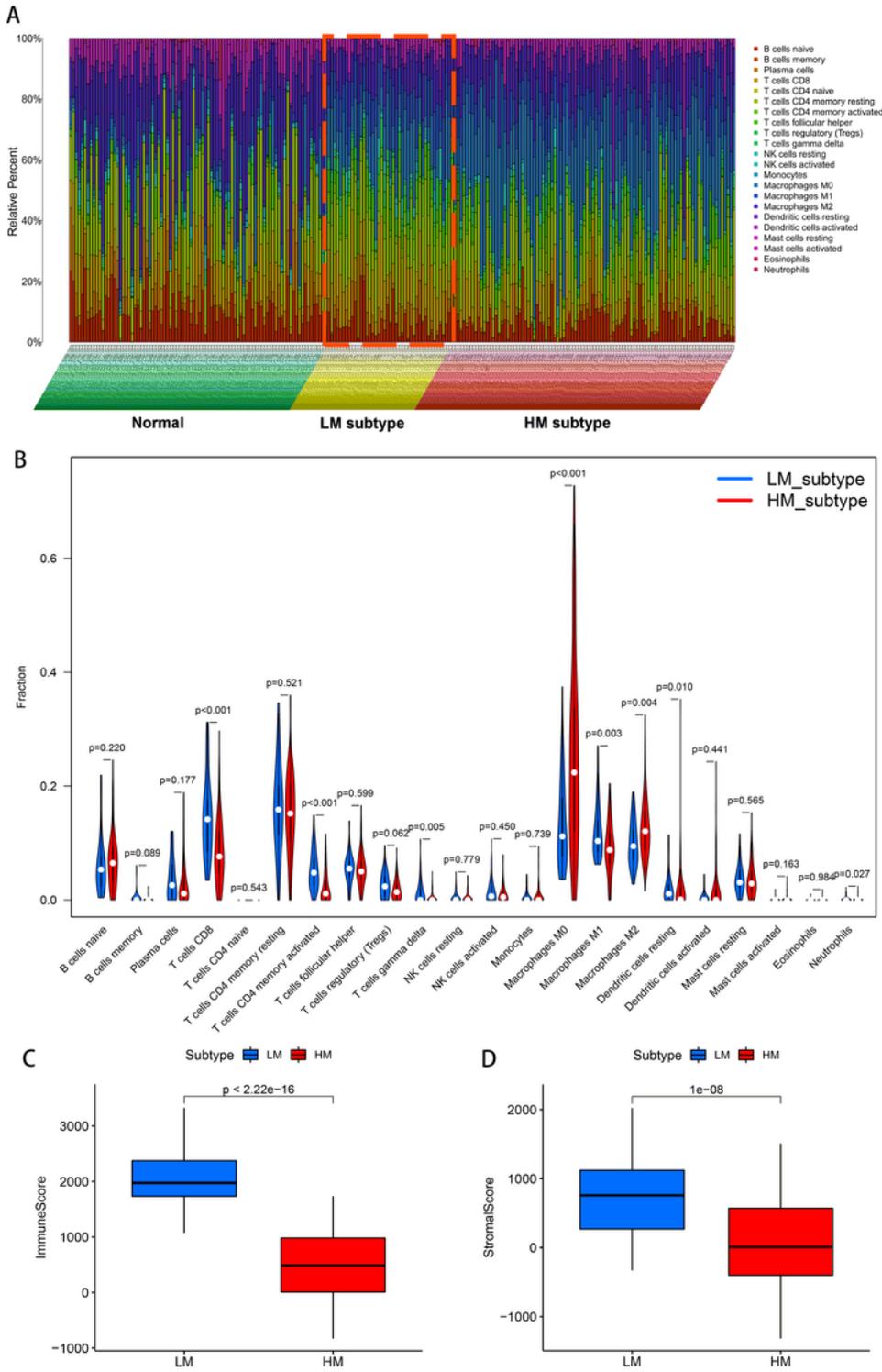


Figure 4

Difference of immune cell infiltration and immune score among different subtypes. a Levels of immune cell infiltration in normal, TNBC and NTNBC samples were compared. b Comparison of the differences of 22 immune cells in different subtypes. c,d Differences in immune and matrix scores between different subtypes.

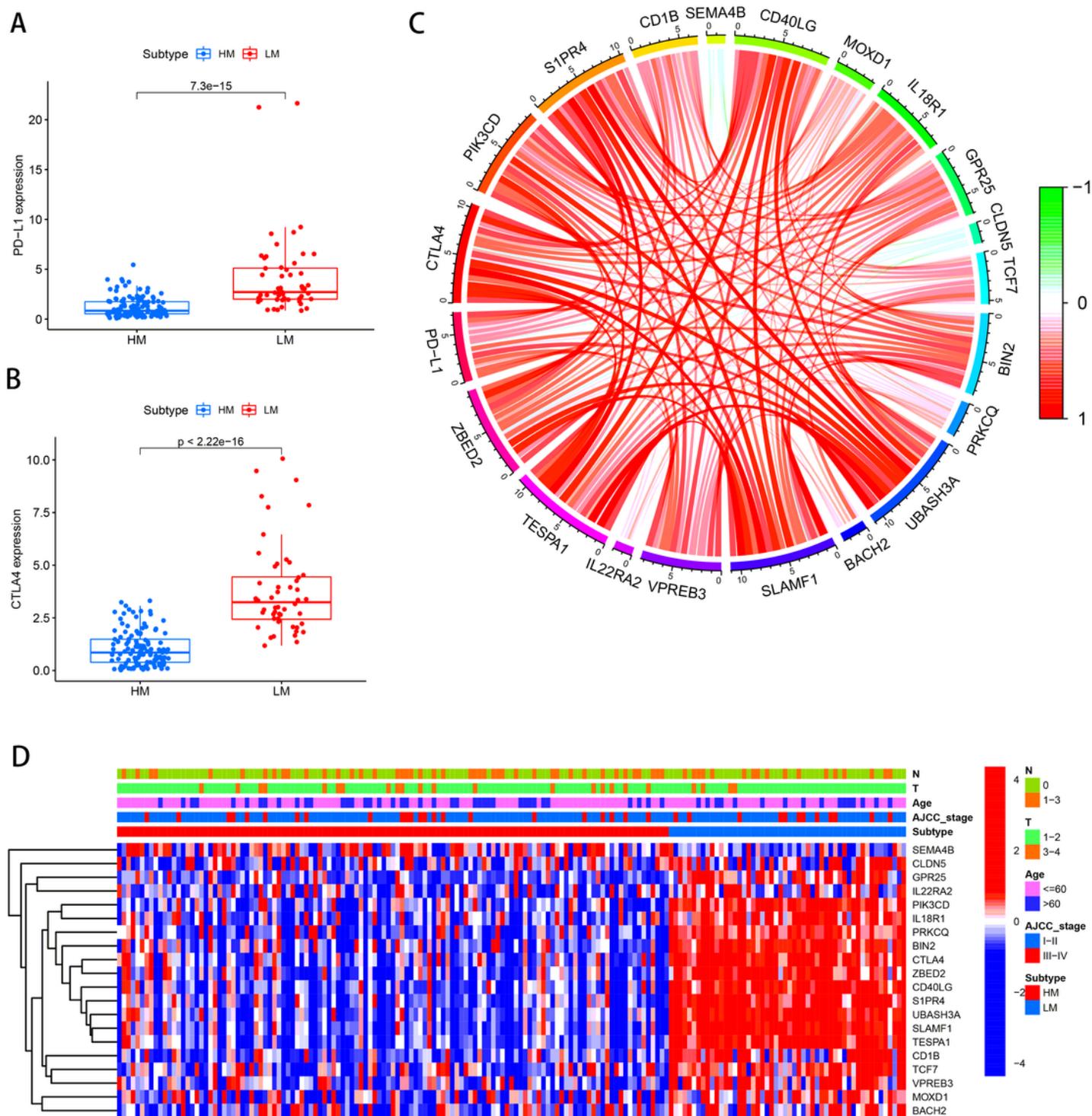


Figure 5

Correlation analysis of prognostic differential genes immune checkpoint genes. a, b The expression of immune checkpoint markers PD-L1 and CTLA4 differed among different subtypes. c Associations between 20 prognostic genes including CTLA4 and with PD-L1. d The association of 20 prognostic genes with age, AJCC stage, T stage, N stage and different subtypes.

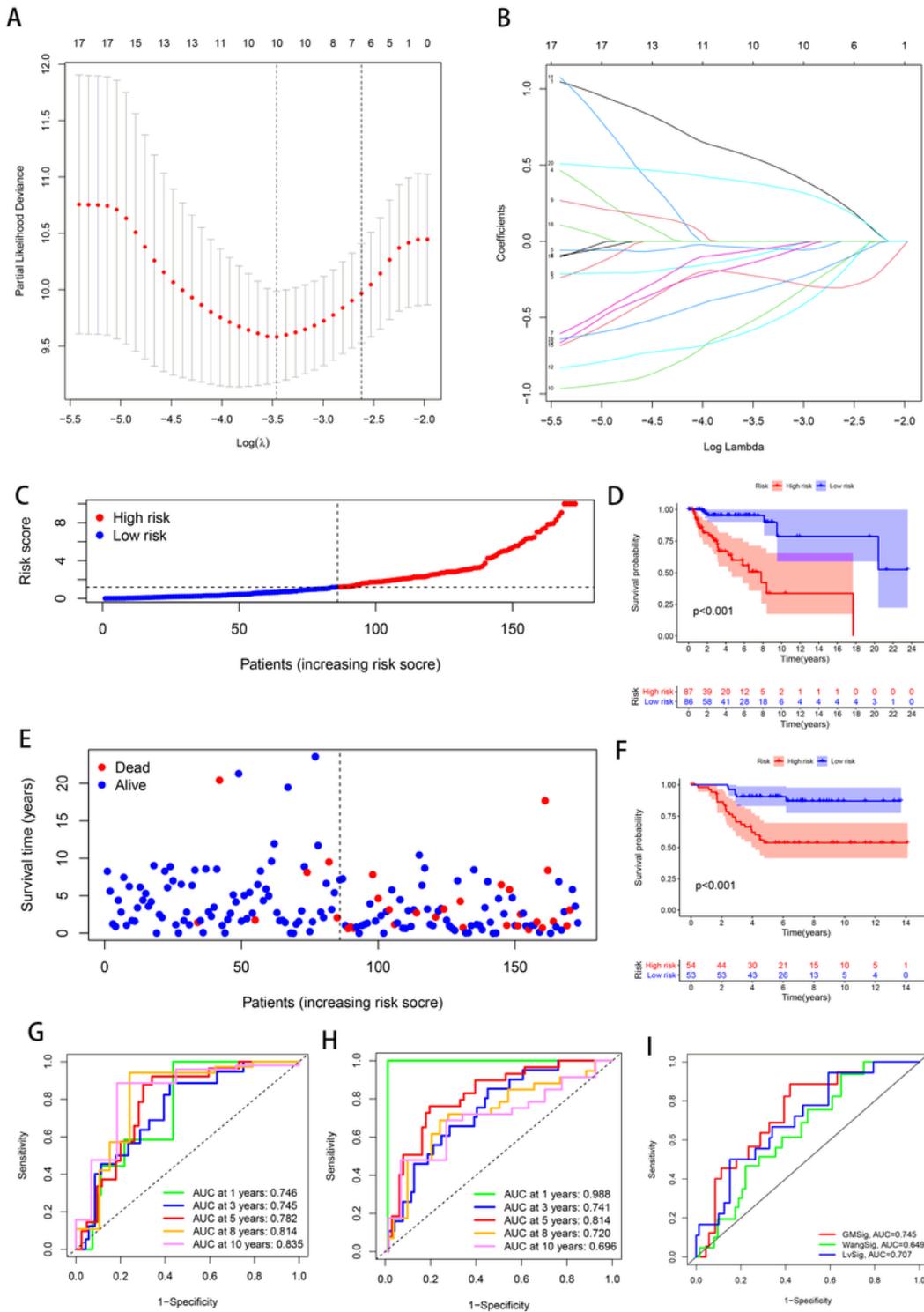


Figure 6

Construction of a gene mutation signature for prediction of prognosis in TNBC. a LASSO regression with 10-fold cross-verification. b LASSO coefficient profiles of 20 gene mutation regulator mRNAs. c distribution of risk scores of high and low-risk TNBC patients based on the gene mutation prognostic signature. e These patients died more and more as the risk score increased. d, f Prognosis analysis of patients in Train group and Text group at high and low risk. g, h The ROC curve depicts the AUC values of

Train and Test groups at 1,3,5,8,10 years respectively. i The ROC analysis at 3 years of overall survival for the GMSig, WangSig and LvSig.

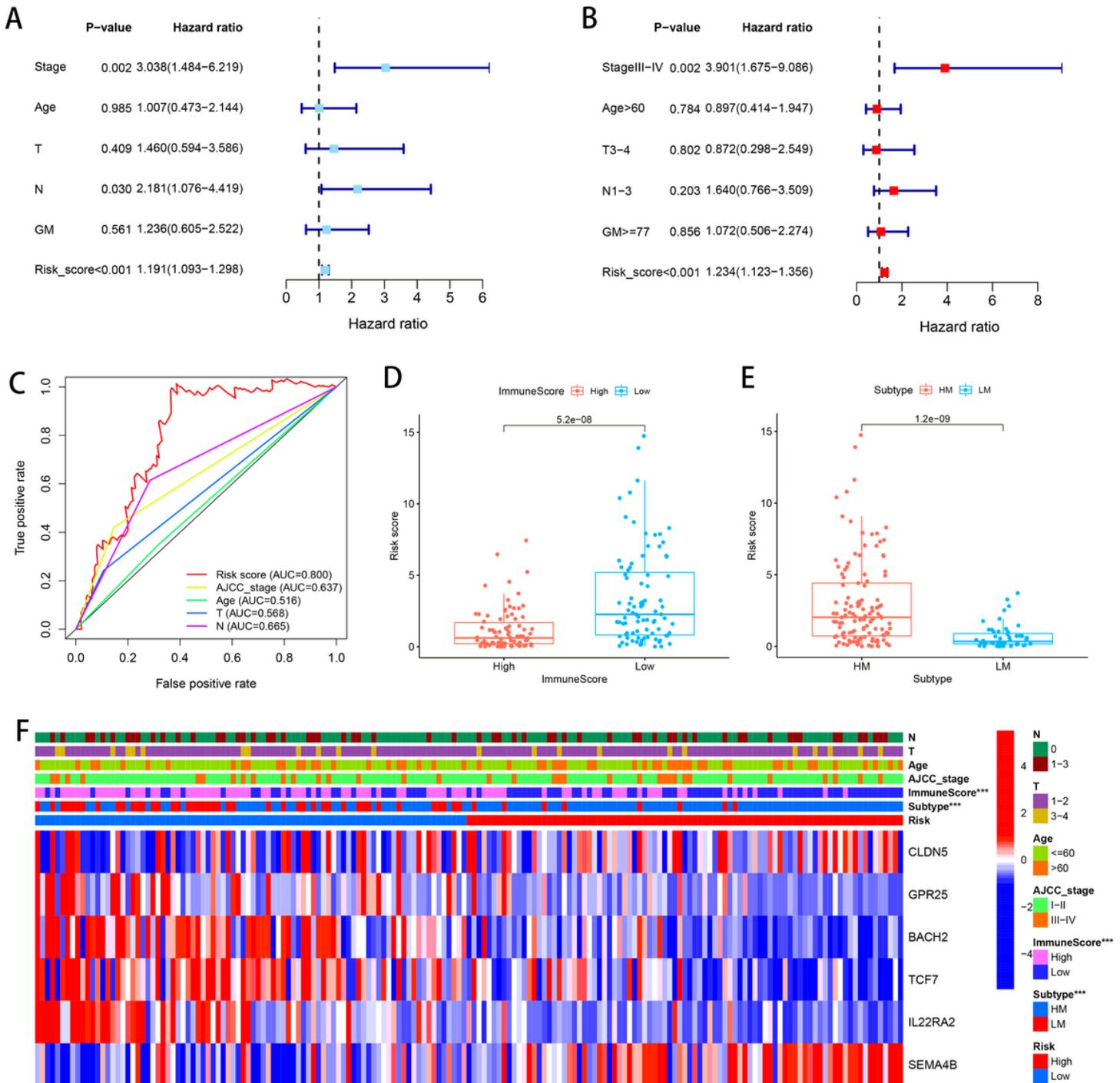


Figure 7

Evaluation of the stability and reliability of the signature for prediction of prognosis in TNBC. a Univariate and b multivariate cox regression analysis of age, AJCC stage, T stage, N stage, gene mutation frequency and risk score in the training group. c ROC curve analysis shows the prognostic accuracy of clinicopathological parameters such as age, AJCC stage, T stage, N stage and risk score. f Heatmap of

association between risk group and mutation subtypes, immune score, Age, AJCC stage, T stage, N stage. Among them, immunization score and different subtypes are closely related to risk score. Adjusted P values were showed as: ns, not significant; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$. Among them, d immunization score and e different subtypes are closely related to risk score.

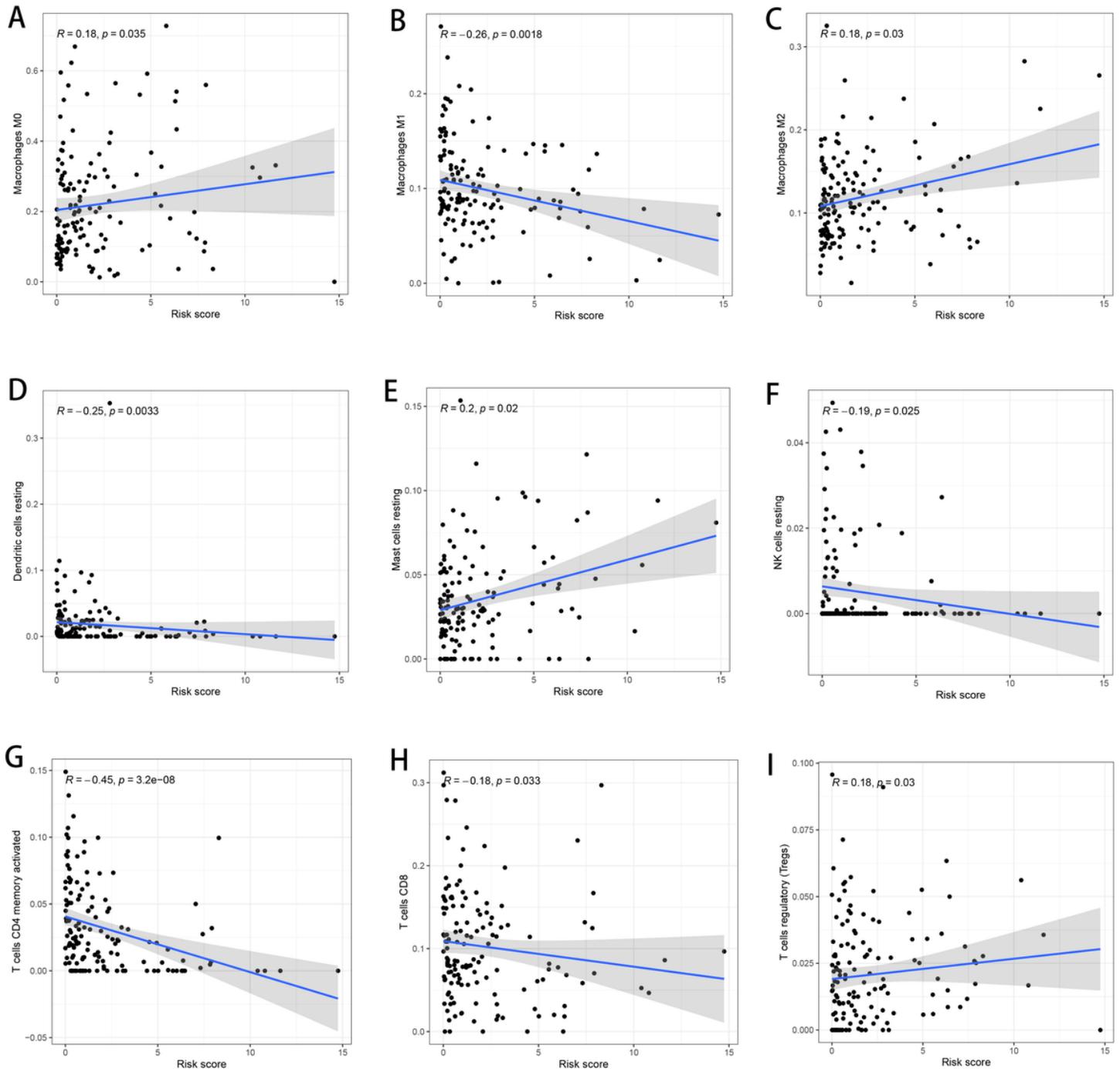


Figure 8

Correlation analysis between risk score and immune cells. a-i Represents an association between risk score and 9 different immune cells, $P < 0.05$.

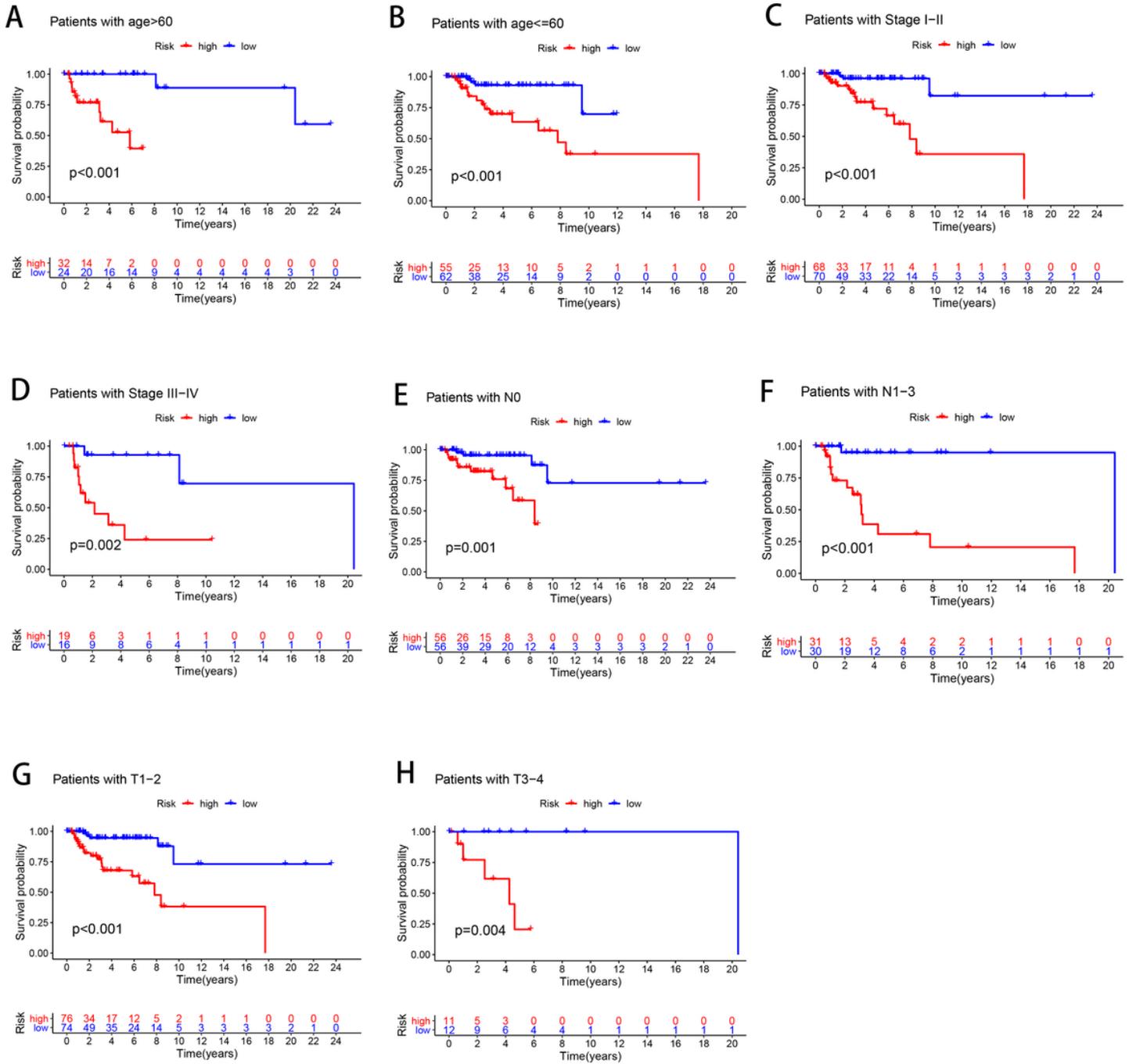


Figure 9

Kaplan–Meier survival curves analysis stratified by different clinical parameters. a Age > 60; b Age <=60; c Stage I-II; d Stage III-IV; e N stage 0; f N stage1-3. g T stage1-2; h T stage3-4.

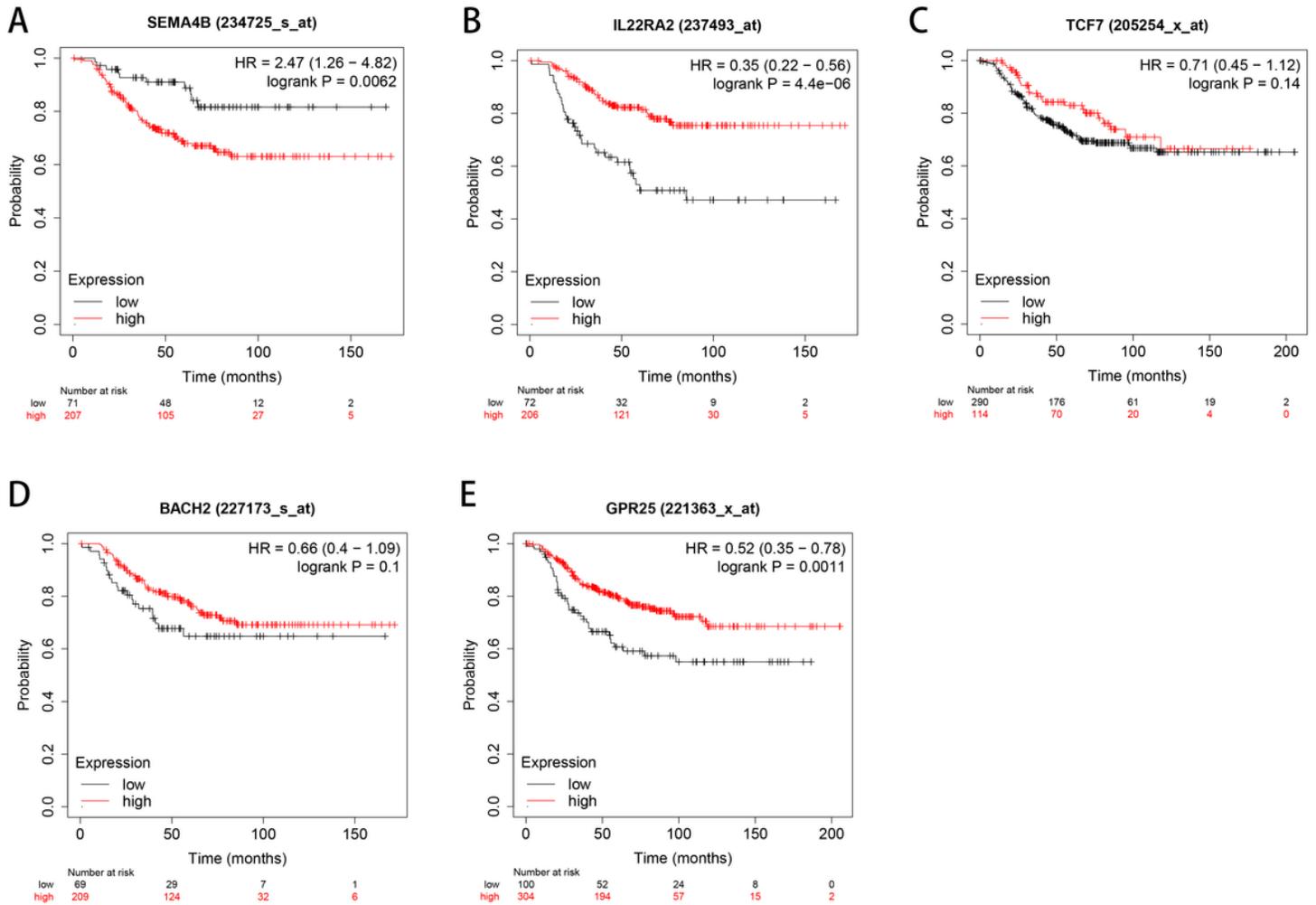


Figure 10

The effect of genes in the model on the prognosis of TNBC was validated in an online database. a sema4b, b il22RA2, c tcf7, d bach2, e gpr25.

Supplementary Files

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

- [Differentialgenes.xlsx](#)
- [Prognosticrelatedgenes.xlsx](#)
- [SixmRNAsprognosticmodel.xlsx](#)