



OPEN

Bioinformatic profiling identifies prognosis-related genes in the immune microenvironment of endometrial carcinoma

Pu Cheng^{1,2}✉, Jiong Ma¹, Xia Zheng¹, Chunxia Zhou¹ & Xuejun Chen¹

Endometrial carcinoma (EC) is a common malignancy of female genital system which exhibits a unique immune profile. It is a promising strategy to quantify immune patterns of EC for predicting prognosis and therapeutic efficiency. Here, we attempted to identify the possible immune microenvironment-related prognostic markers of EC. We obtained the RNA sequencing and corresponding clinical data of EC from TCGA database. Then, 3 immune scores based on the Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data (ESTIMATE) algorithm were computed. Correlation between above ESTIMATE scores and other immune-related scores, molecular subtypes, prognosis, and gene mutation status (including BRCA and TP53) were further analyzed. Afterwards, gene modules associated with the ESTIMATE scores were screened out through hierarchical clustering analysis and weighted gene co-expression network analysis (WGCNA). Differentially expressed analysis was performed and genes shared by the most relevant modules were found out. KEGG pathway enrichment analysis was conducted to explore the biological functions of those genes. Survival analysis was carried out to identify prognostic immune-related genes and GSE17025 database was further used to confirm the correlation between immune-related genes and the ImmuneScore. The immune-related scores based on ESTIMATE algorithm was closely related to the immune microenvironment of EC. 3 gene modules that had the closest correlations with 3 ESTIMATE scores were obtained. 109 immune-related genes were preliminarily found out and 29 pathways were significantly enriched, most of which were associated with immune response. Univariate survival analysis revealed that there were 14 genes positively associated with both OS and PFS. Among which, 11 genes showed marked correlations with ImmuneScore values in GSE17025 database. Our current study profiled the immune status and identified 14 novel immune-related prognostic biomarkers for EC. Our findings may help to investigate the complicated tumor microenvironment and develop novel individualized therapeutic targets for EC.

Abbreviations

EC	Endometrial carcinoma
ESTIMATE	Malignant tumor tissues using expression data
FIGO	The international federation of gynecology and obstetrics
TME	Tumor microenvironment
TCGA	The cancer genome atlas
WGCNA	Weighted gene co-expression network analysis
GEO	Gene expression omnibus
TPM	Transcripts PerKilobase million
OS	Overall survival
PFS	Progression-free survival
TMB	Tumor mutation burden
MAD	Median absolute deviation
TOM	Topological overlap matrix

¹Department of Gynecology, Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China. ²Key Laboratory of Tumor Microenvironment and Immune Therapy of Zhejiang Province, Hangzhou, China. ✉email: drchengpu@zju.edu.cn

KEGG Kyoto encyclopedia of genes and genomes
DEGs Differentially expressed genes

The global morbidity and mortality of endometrial carcinoma (EC) shows an increasing trend in recent years^{1,2}. In China, EC ranks the second place in terms of morbidity among all malignant tumors of female genital system, for whom the 5-year survival rate is 55.1%³. At present, FIGO staging and histological classification are still the chief factors applied to patient stratification and prognosis prediction in EC^{4,5}. Over the past few past decades, great individual differences have been found in the outcomes of EC treatments due to the tumor heterogeneity, which is partially dependent on the molecular biological features of the primary lesion⁶. Moreover, quite a few patients show distinct responses to adjuvant therapy even though they are at the same clinical stage⁷. Thus, more effective approaches are needed in order to hierarchically classify patients into high- or low- risk subgroups for monitoring and optimizing the treatment of EC⁸.

As in many other types of cancer, immunotherapy is currently recognized as a novel promising therapeutic option in EC⁹. Interactions between different infiltrating immune cells and EC cells in tumor microenvironment (TME) significantly affect tumor progression and recurrence¹⁰. Besides, EC exhibits a unique immune profile and can be used to construct suitable models for exploring molecular crosstalk between immune and tumor cells¹¹. Great progresses have been attained in the past few years, making it possible to recognize novel molecular therapeutic targets within the microenvironment of EC¹². At the same time, several immune-related factors have been identified to predict patient prognosis, which emphasizes the significance of certain immune status on EC outcomes¹³. Nonetheless, the vast majority of existing studies are pre-clinical basic experiments or have limited available clinical information, while study with a large sample size has not been carried out so far¹⁴.

Consequently, several immune scoring systems were developed based on the immune-related gene expression patterns through integrated analysis of The Cancer Genome Atlas (TCGA) database to explore the relationships between tumor cells and immunocytes in TME, as well as to quantify the immune microenvironment for each individual cancer case. Of them, the inflammation-based index was reported to be associated with the local immune response and prognosis in various cancers, including pancreatic cancer, colorectal cancer¹⁵, non-small cell lung cancer¹⁶ and tongue cancer¹⁷ and so on. However, limited studies have been designed in an attempt to develop an immune-related prognostic signature for EC.

In the current study, we obtained the RNA sequencing (RNA-seq) and corresponding clinical data of EC from TCGA database. Then, we calculated 3 immune scores for each EC sample based on the Estimation of STromal and immune cells in MAlignant Tumor tissues using Expression data (ESTIMATE) algorithm and analyzed the correlation between ESTIMATE immune score and molecular features of EC. Thereafter, gene modules associated with the immune scores were identified using weighted gene co-expression network analysis (WGCNA), and finally, 14 novel prognostic immune-related genes were screened out which were further explored in the GSE17025 database.

Materials and methods

Data collection and immune score calculation. All the data we used in our study are publicly accessible at TCGA and NCBI GEO (accession number: GSE17025) database. Firstly, GDC application programming interface was utilized to download clinical follow-up, RNA-seq and SNP data. RNA-seq FPKM values were subsequently converted into Transcripts PerKilobase Million (TPM) files. The expression levels of 13 previously published immune metagenes¹⁸ were calculated as the median of log₂-transformed expression levels of clustered genes for each sample. Meanwhile, the infiltration of 6 immune cells (B cells, CD4 + T cells, CD8 + T cells, neutrophils, macrophages and dendritic cells) were calculated and downloaded from the TIMER¹⁹ (<https://cistrome.shinyapps.io/timer/>) database. In addition, the ESTIMATEScore, StromalScore and ImmuneScore values for each sample were computed by the ESTIMATE function of R package.

Assessment of the correlation between ESTIMATE scores and immune status. Correlation analysis was carried out to assess the relevance between 3 ESTIMATE immune scores, expression of 13 immune metagenes and the infiltration status of 6 immune cells. Moreover, the correlation of 3 ESTIMATE immune scores, 13 immune metagenes expression and the infiltration of 6 immune cells were further analyzed among the 4 previously reported EC subtypes respectively.

The correlation between ESTIMATE scores and patient prognosis. We firstly conducted survival analysis using Kaplan–Meier method with survival function of R package to explore the overall survival (OS) of above mentioned 4 EC subtypes respectively. Afterwards, patients were divided into high- and low-score groups according to the median of 3 ESTIMATE immune scores. Then, the differences of OS between these groups were examined through Kaplan–Meier method with survival package under R environment.

Exploration of the association between ESTIMATE scores and gene mutation. The mutation data of BRCA2, BRCA1 and TP53 were extracted from TCGA derived SNP dataset and processed with Mutect. Then, the correlation between different mutation status of these genes and ESTIMATE immune scores were analyzed using Wilcox.test Package under R environment. Furthermore, patients were divided into high- and low- tumor mutation burden (TMB) groups according to the median of TMB value. The correlation between TMB status and ESTIMATE immune scores were assessed using Wilcox.test Package under R environment.

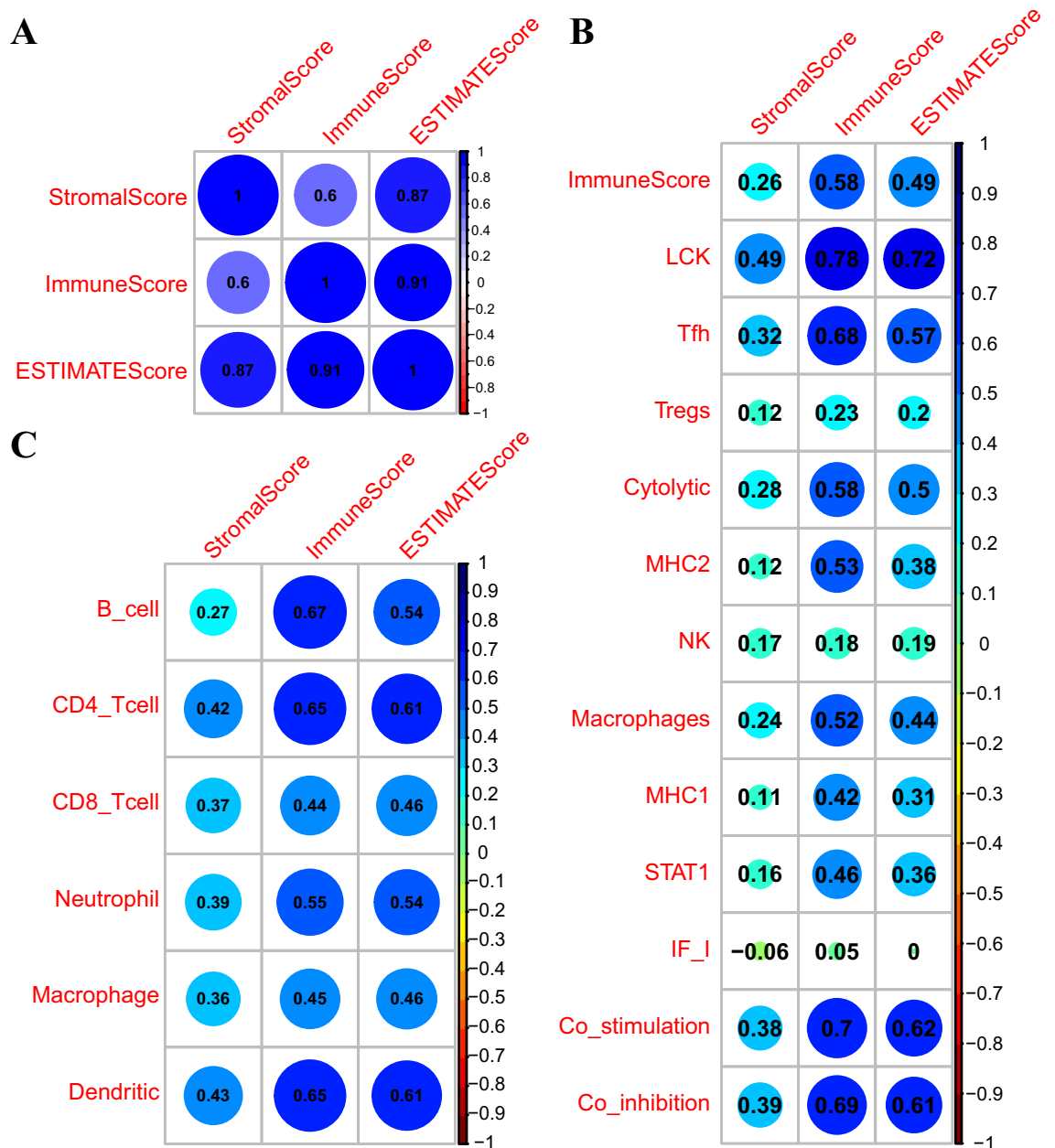


Figure 1. Correlations between ESTIMATE immune score and other immune scores for EC patients. (A) Correlations among the 3 ESTIMATE immune scores. (B) Correlations between 3 ESTIMATE immune scores and 13 metagenes scores for EC patients. (C) Correlations between 3 ESTIMATE immune scores and 6 immunocyte infiltration scores for EC patients. Coefficients of Spearman correlation were displayed color-coded, so as to demonstrate the negative (red) or positive (blue) association.

Identifying immune score-related gene modules through WGCNA. Firstly, transcripts that had at least 75% TPM greater than 1 and median absolute deviation (MAD) greater than the median were selected for following analysis according to the expression patterns. Then, samples were clustered through hierarchical clustering method with the distance > 80,000 considered as cut-off value of outlier sample. The distance between each 2 transcripts was computed according to Pearson correlation coefficient, then the R package WGCNA²⁰ was used to establish the weight co-expression network, and co-expression modules were selected at the soft threshold of 10 in order to ensure the constructed co-expression network conformed to the scale-free network. That was to say, the node/k connectivity ($\log(k)$) logarithm was negatively correlated with logarithm in terms of the occurrence probability of node ($\log(P(k))$), and the correlation coefficient was > 0.8. The appropriate β value was also selected to ensure the scale-free network. Then, the expression matrix was converted into the adjacent one, followed by further conversion into the topological one for gene clustering on the basis of topological overlap matrix (TOM), in accordance with the average linkage hierarchical cluster approach following the mixed dynamic shear tree standards. In addition, more than 30 genes had been selected for every gene network module. The dynamic shear approach was also utilized to determine gene modules, and eigengenes values of all modules

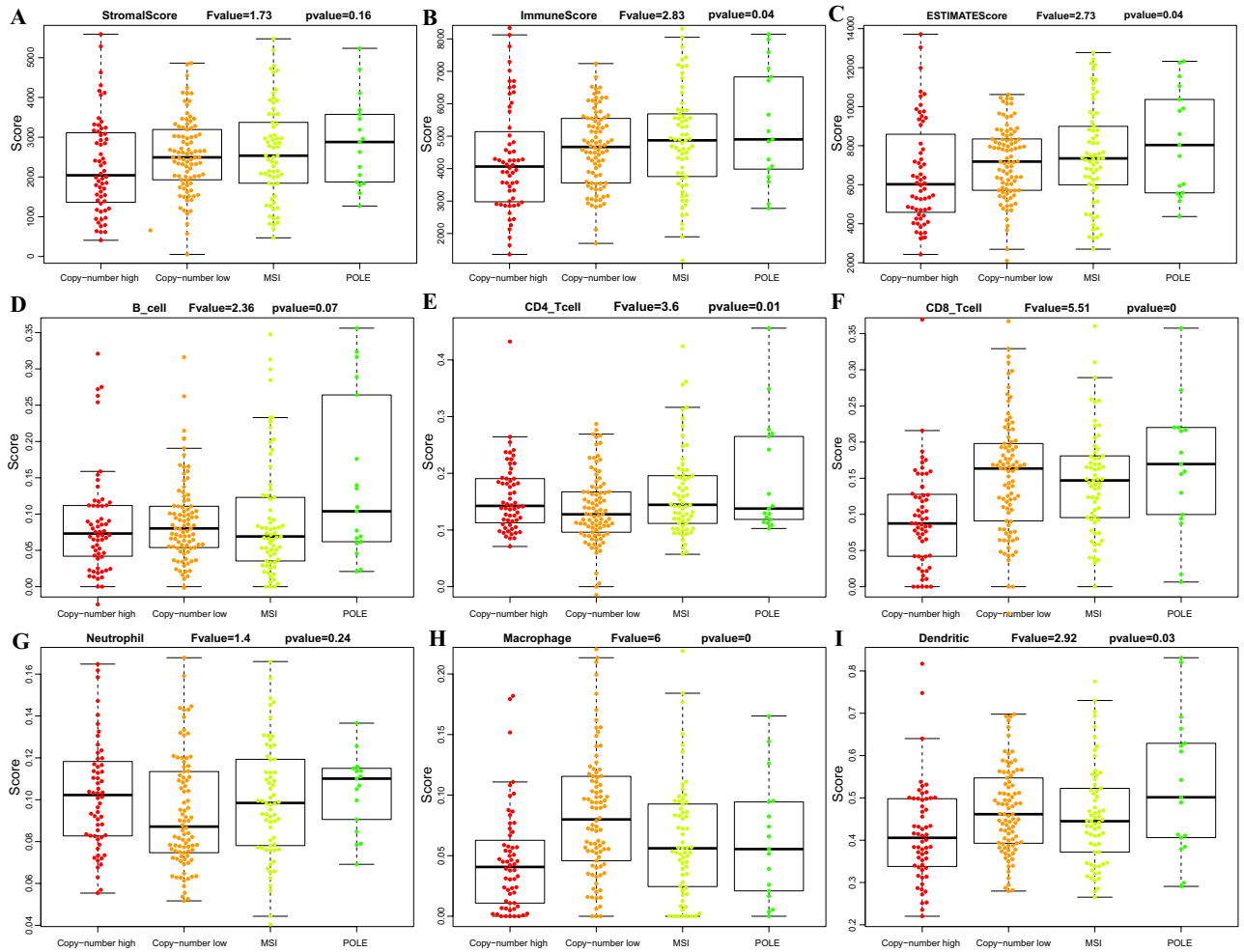


Figure 2. Distribution of StromalScore (A), ImmuneScore (B), ESTIMATEScore (C) and B cell (D), CD4 + T cell (E), CD8 + T cell (F), Neutrophil (G), Macrophage (H), Dendritic cell (I) infiltration among 4 different molecular subtypes (Copy – number high, Copy – number low, MSI and POLE).

were calculated. Cluster analysis was then carried out on the modules, adjacent modules were fused together to obtain a novel one, and appropriate minModuleSize, deepSplit, and height values were assigned. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis was performed to explore the potential biological functions of genes within these 3 modules using the ClusterProfiler package under R environment, with the significant FDR level of < 0.05 . Then, the associations between the recognized gene modules and ESTIMATEScore, StromalScore as well as ImmuneScore values were calculated, respectively, in order to mine the gene modules having highest correlation.

Construction of gene interaction network and functional analyses. We separated patients into two groups equally based on the ImmuneScore and ESTIMATEScore. The DESeq2 function of R package was utilized to screen differentially expressed genes (DEGs) between these groups. The cut-off criterion for DEGs was set as $p < 0.05$ and $|\log_2(\text{Foldchange})| > 1$. Genes shared by the most relevant module, DEGs of ImmuneScore and ESTIMATEScore groups were finally screened out. All genes were subsequently mapped into the String database²¹ separately using the STRINGdb package under R environment, with the threshold of > 0.4 to obtain the gene–gene interaction. Cytoscape was used for visualization. At the same time, the R package clusterprofile²² was utilized to carry out KEGG enrichment analyses for visualizing the signaling pathways affected by genes. The prognostic value of each immune-related gene was calculated using survival package under R environment. The correlations between immune-related genes and ImmuneScore were further explored with an independent external GEO dataset (GSE17025) through Pearson correlation analysis.

Results

The immune scoring system based on the ESTIMATE algorithm was closely related to the immune microenvironment of EC. The expression levels of 13 immune metagenes, infiltration status of 6 types of immunocytes, and 3 immune-related scores based on ESTIMATE algorithm (ESTIMATEScore,

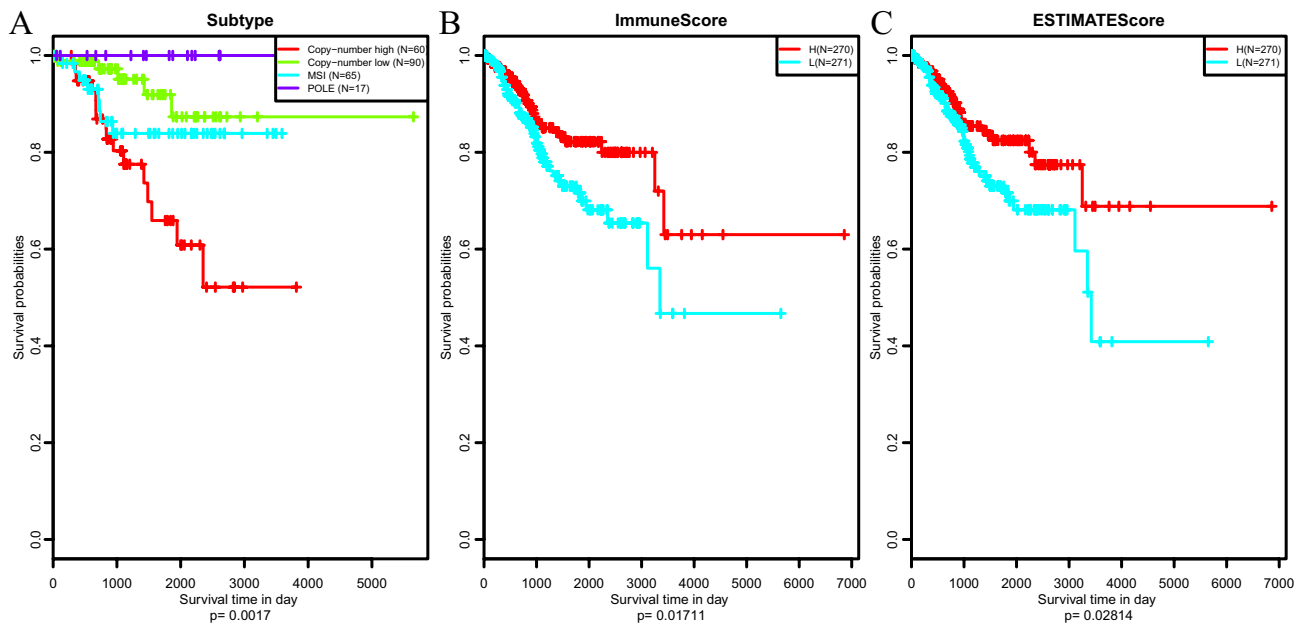


Figure 3. Prognosis of EC patients with different molecular subtypes (A), ImmuneScores (B) and ESTIMATEScores (C). H: high-risk; L: low-risk.

StromalScore, and ImmuneScore) were computed. Additionally, Spearman correlation coefficient was employed for quantifying the relationships between these scoring systems (Fig. 1). Our results suggested that, immune-related scores calculated using the ESTIMATE algorithm had an average internal correlation >0.8 and also high correlations with the other 2 algorithms. Above findings indicated that immune-related scores computed based on the ESTIMATE algorithm was closely related to the immune microenvironment of EC.

Besides, ESTIMATEScore, ImmuneScore, and StromalScore among the 4 recognized subtypes²³ had also been examined (Fig. 2A–C). It was clear that, differences in ESTIMATEScore and ImmuneScore levels among different molecular subtypes of EC were statistically significant. Moreover, distribution of 6 types of immunocytes infiltration (Fig. 2D–I) and 13 metagenes expression (Fig. S1) among these 4 subtypes was also analyzed. Statistical significances were observed in 6 out of 13 metagenes expression and 5 out of 6 types of immunocytes infiltration.

To assess the relationship between 3 ESTIMATE scores and patient prognosis, the prognosis of above 4 subtypes was firstly examined. As shown in Fig. 3A, difference in patient prognosis across those 4 subtypes was statistically significant, among which, the copy-number high subtype had the poorest prognosis. Next, all samples were classified according to median of ESTIMATE scores followed by survival analysis, which was consistent with previous report²⁴ (Fig. 3B–C). Obviously, the prognosis for samples with the high ESTIMATEScore and ImmuneScore was much better than that with low scores, suggesting that ESTIMATE immune scoring system might be used as novel promising markers to predict the prognosis for EC.

Afterwards, correlations between those 3 ESTIMATE scores and 3 independent prognostic gene mutations (BRCA2, BRCA1 and TP53) were explored^{25,26}. As a result, 3 immune-related scores of different mutant groups were generally higher when compared with those of wild-type groups, especially in TP53 and BRCA1 subgroups (Fig. 4A–I). Subsequently, the TMB of each sample was computed, and the relationships between 3 ESTIMATE scores and TMB level were analyzed. As presented in Fig. 4J–L, the ESTIMATEScore and ImmuneScore showed significant positive correlation with TMB.

To sum up, the immune-related scores based on ESTIMATE algorithm was closely related to the immune microenvironment of EC and could be used to be the optimal immune scoring methods for prognosis prediction.

Identification of immune score-related gene modules by WGCNA. Samples were firstly selected to performed hierarchical clustering analysis (Fig. 5A). After excluding outlier data, 579 samples were finally used to construct a weight co-expression network through WGCNA (Fig. 5B,C). Then, gene modules were examined by dynamic shear, and all the recognized modules were further clustered. Next, adjacent gene modules were fused to form a new one, with minModuleSize, deepSplit and height set at 30, 2 and 0.25, respectively. 17 modules were ultimately obtained (Fig. 5D,E), and 5362 transcripts were assigned to 16 co-expression modules except the grey module. Correlations between eigenvectors of these 16 modules and 3 ESTIMATE scores were then calculated (Fig. 5E), from which we could see that the red (170 genes), purple (55 genes) and tan (49 genes) modules had the closest correlations with 3 ESTIMATE scores, with the average correlation coefficients >0.5 .

As following, we performed KEGG pathway enrichment analysis to explore the potential biological functions of the 274 genes within these 3 modules. According to the results, there were 23 and 54 pathways enriched to the purple module and red module, respectively (Fig. 6). It was easy to find that the genes were primary enriched to immune-related pathways, including T cell differentiation, primary immunodeficiency, chemokine signaling pathway, cytokine-cytokine receptor interaction, B cell receptor signaling pathway and so on. Thus, it can be

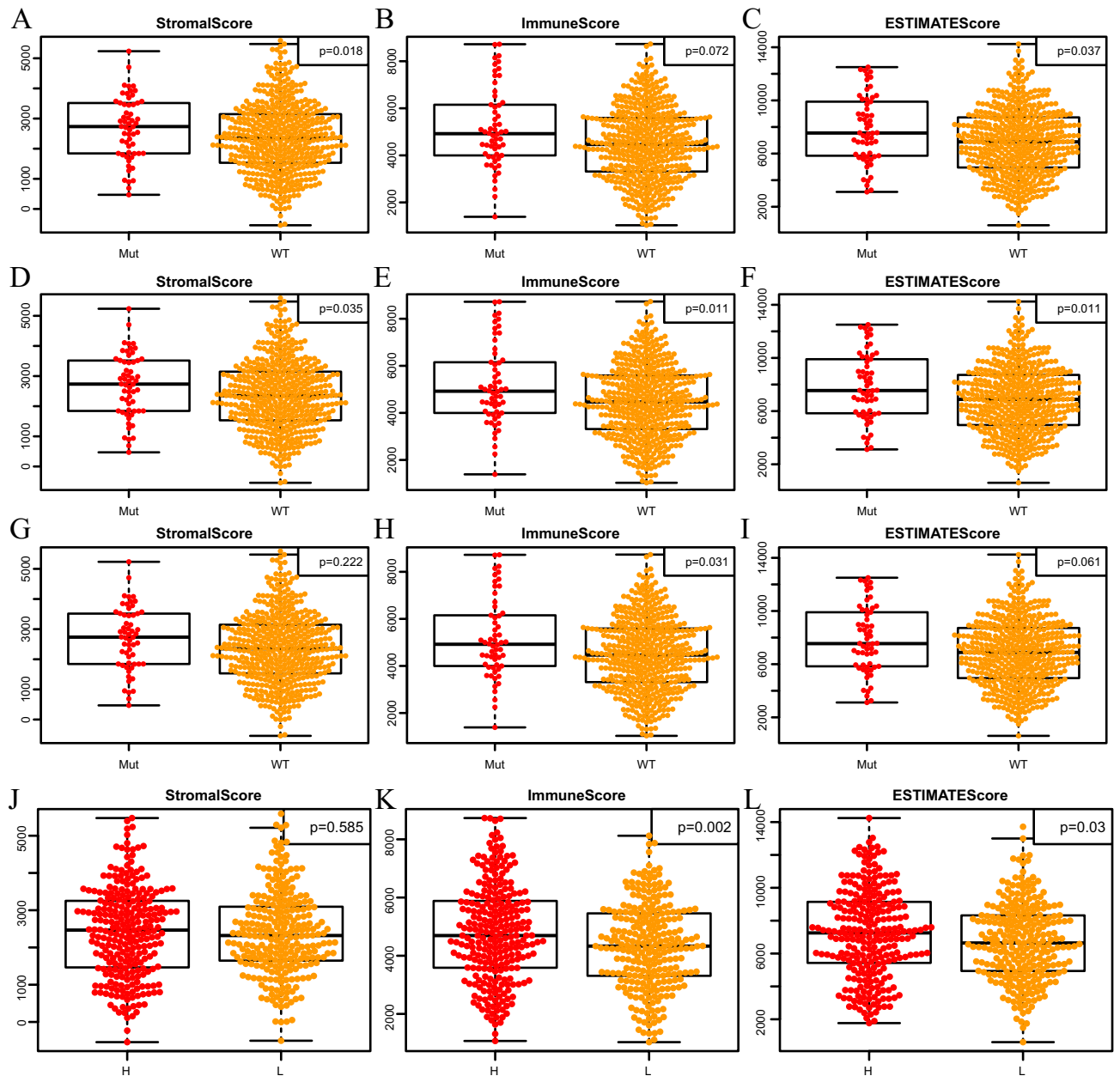


Figure 4. Correlations between ESTIMATE immune scores and gene mutations. The StromalScore in TP53 (A), BRCA1 (D) and BRCA2 (G) non-mutation and mutation groups. ImmuneScore in TP53 (B), BRCA1 (E) and BRCA2 (H) non-mutation and mutation groups. ESTIMATEScore in TP53 (C), BRCA1 (F) and BRCA2 (I) non-mutation and mutation groups. StromalScore (J), ImmuneScore (K) and ESTIMATEScore (L) in TMB-high and -low groups.

inferred that the ESTIMATE score-related genes may closely associate with the immune microenvironment of EC.

Identification of immune microenvironment-related prognostic genes. To seek out the most immune-related genes, correlations between previously obtained ESTIMATE score-related modules and genes were computed. As illustrated in Fig. 7, the correlation coefficients were distributed in a bimodal manner, with the intersection point value of 0.77. Based on this, 136 genes with the maximum correlation coefficients with those 3 modules greater than 0.77 were screened out.

According to the previously described screening method, 379 DEGs were identified among high- and low-ESTIMATEScore subgroups (Fig. 8A,B) using the R package DESeq2²⁷. At the meantime, 526 DEGs were acquired among ImmuneScore-high and -low subgroups (Fig. 8C,D). The results were presented in a volcano plot and a heatmap for each subgroup. Obviously, the DEGs showed distinct expression patterns in low-ESTIMATEScore

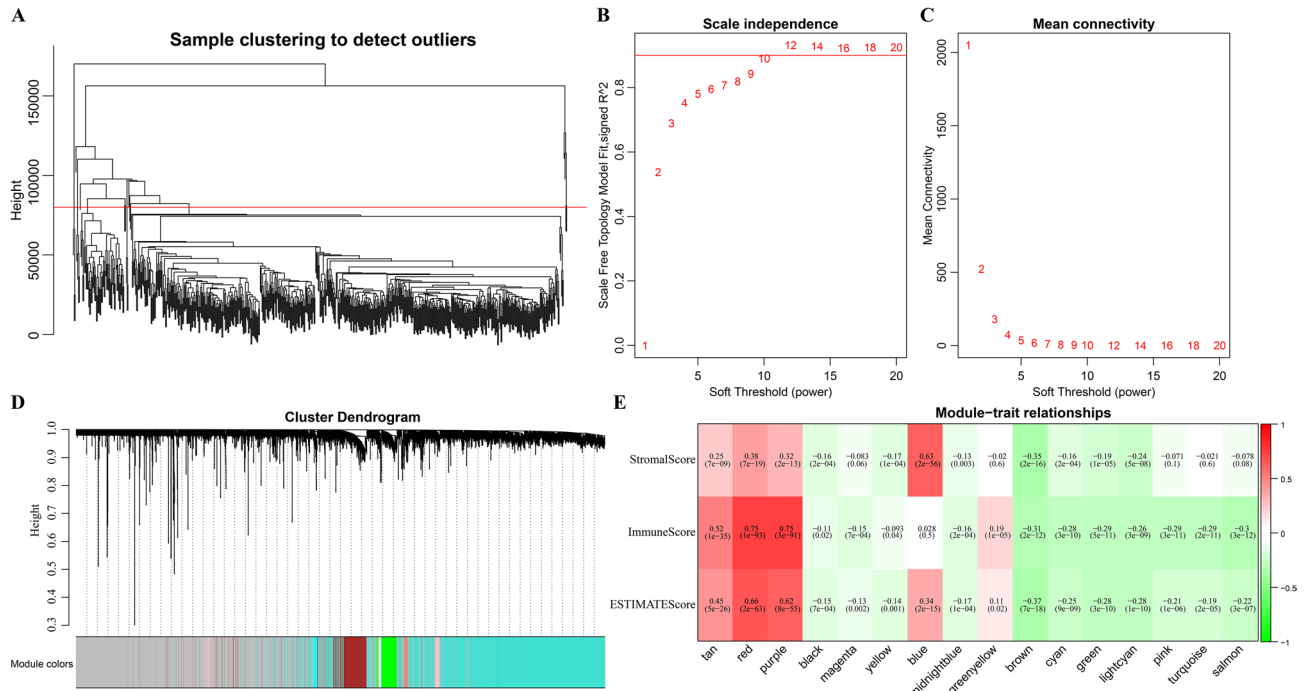


Figure 5. Immune score-related gene modules identified via WGCNA. (A) Cluster analysis of samples. (B,C) Network topological analysis for different soft-thresholding powers. (D) Module colors and gene dendrogram. (E) Correlations between different gene modules and 3 ESTIMATE Scores. The intensity of red and green colors indicates the value of correlation coefficient.

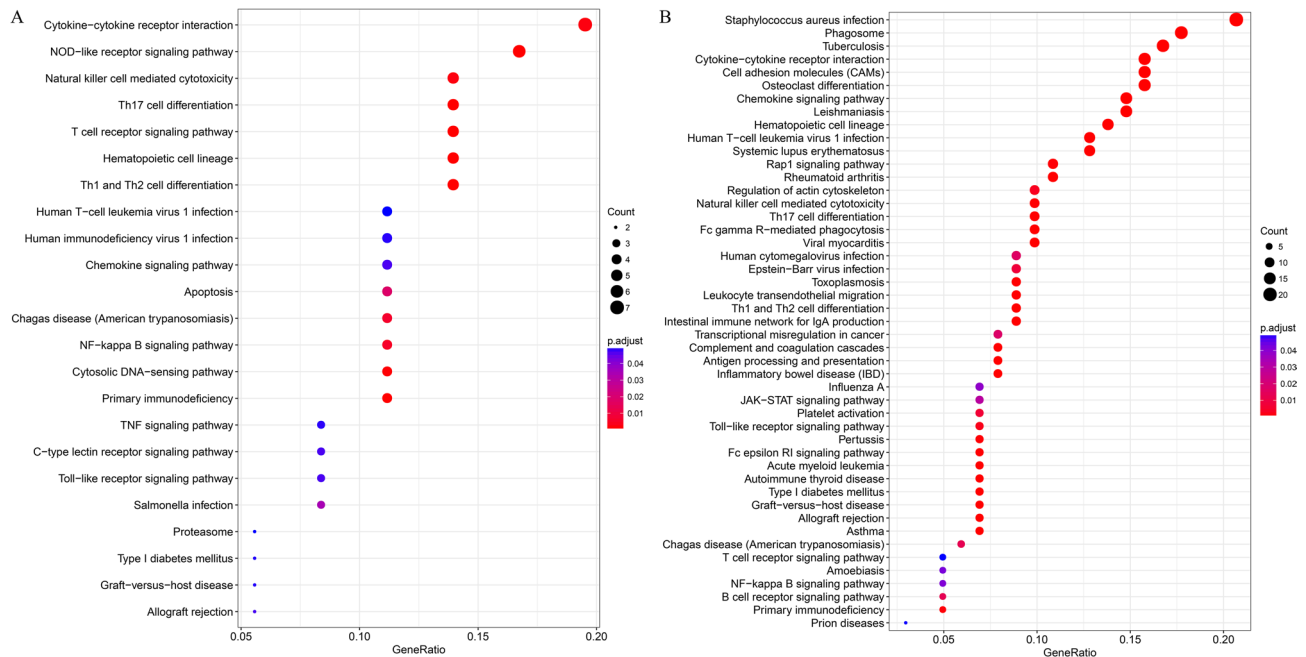


Figure 6. The KEGG pathway enrichment analysis of the genes in purple (A) and red (B) module.

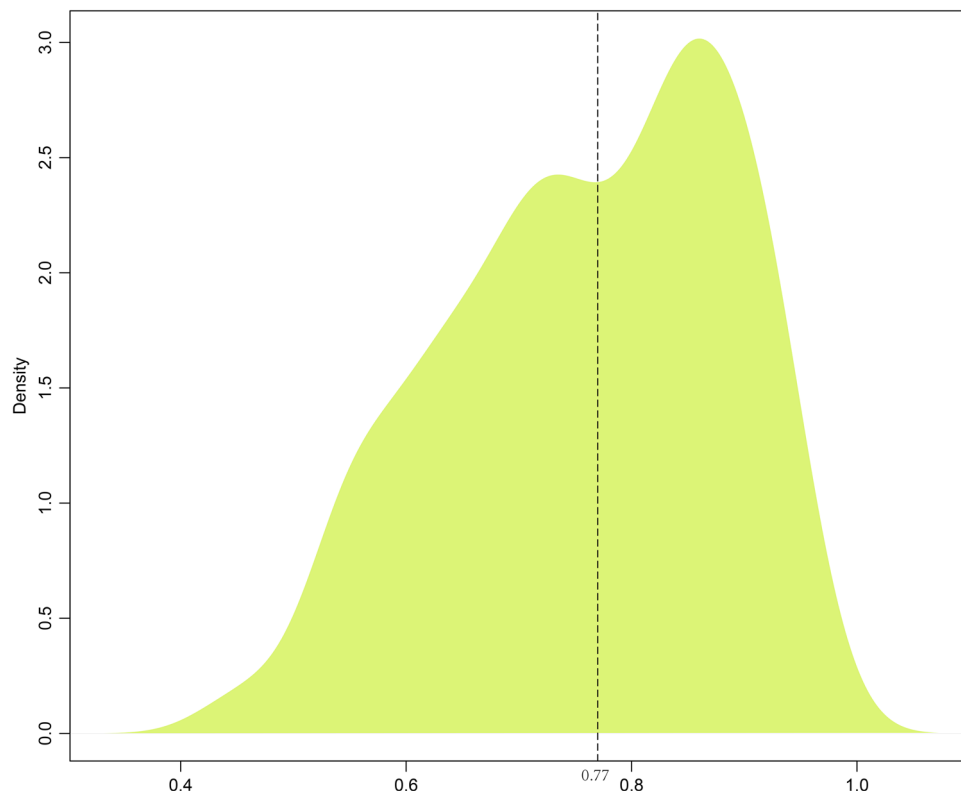


Figure 7. The distribution of maximum correlation coefficients of 274 module-related genes.

subgroup compared with high-ESTIMATEScore subgroup. Similar results were also observed in low- and high-ImmuneScore subgroups.

To further investigate the immune-related genes, we integrated above 3 gene sets (136 genes with the maximum correlation coefficients with 3 modules, 379 DEGs in high- and low-ESTIMATEScore subgroups and 526 DEGs in high- and low-ImmuneScore subgroups). After intersection, 133 common genes were obtained and 24 genes belonged to 13 metagenes were eliminated subsequently. Finally, 109 immune-related genes were preliminarily found out (Fig. 9A).

Subsequently, KEGG enrichment analyses was conducted with clusterProfiler package under R environment. As shown in Fig. 9B, 29 pathways were significantly enriched, most of which were associated with immune response. Afterwards, the R STRINGdb package²⁸ was utilized to analyze the protein–protein interaction network of those 109 genes. All these genes were mapped to String database in order to acquire the relationship network and 67 nodes were finally obtained (Fig. 9C). As is depicted in Fig. 9D, the degree value of each node was high (5.49 on average), which demonstrated the close association between those immune-related genes.

Then, univariate survival analysis was performed to identify the immune microenvironment-related prognostic markers using the survival and expression data of above 109 genes. The results revealed that there were 34 and 19 genes associated with OS and PFS, respectively. Among which, 14 genes were found to be positively correlated with both OS and PFS (Table 1, Fig. 10).

Further exploration of the correlations between 14 immune-related prognostic genes and ImmuneScore using an external database. The independent dataset GSE17025²⁹ was chosen and normalized expression matrix was downloaded. Then, the R package ESTIMATE was used to calculate the sample ImmuneScore values. Pearson correlation coefficients between expression levels of 12 genes (2/14 genes were excluded due to unavailable expression data) and ImmuneScore values were further determined. As illustrated in Fig. 11, 11 genes (except S1PR4) showed marked correlations with ImmuneScore values.

Discussion

Until recently, the 5-year survival of advanced or recurrent EC is still not optimistic. Even though postoperative adjuvant therapy (chemotherapy and/or radiotherapy) can improve patient outcome, great individual differences in therapeutic effect exist due to the biological heterogeneity of EC cells. Hence, there is a critical need for reliable prognostic biomarkers to evaluate the risk of cancer progression and develop the patient-tailored treatment strategy.

With the rapid development of next generation sequencing techniques, numbers of novel molecular biomarkers have been identified, which were of great help to the personalized treatment of EC^{8,30}. Nevertheless, most studies are conducted on animal models, surgical cancer tissues samples or cell lines under in vitro conditions.

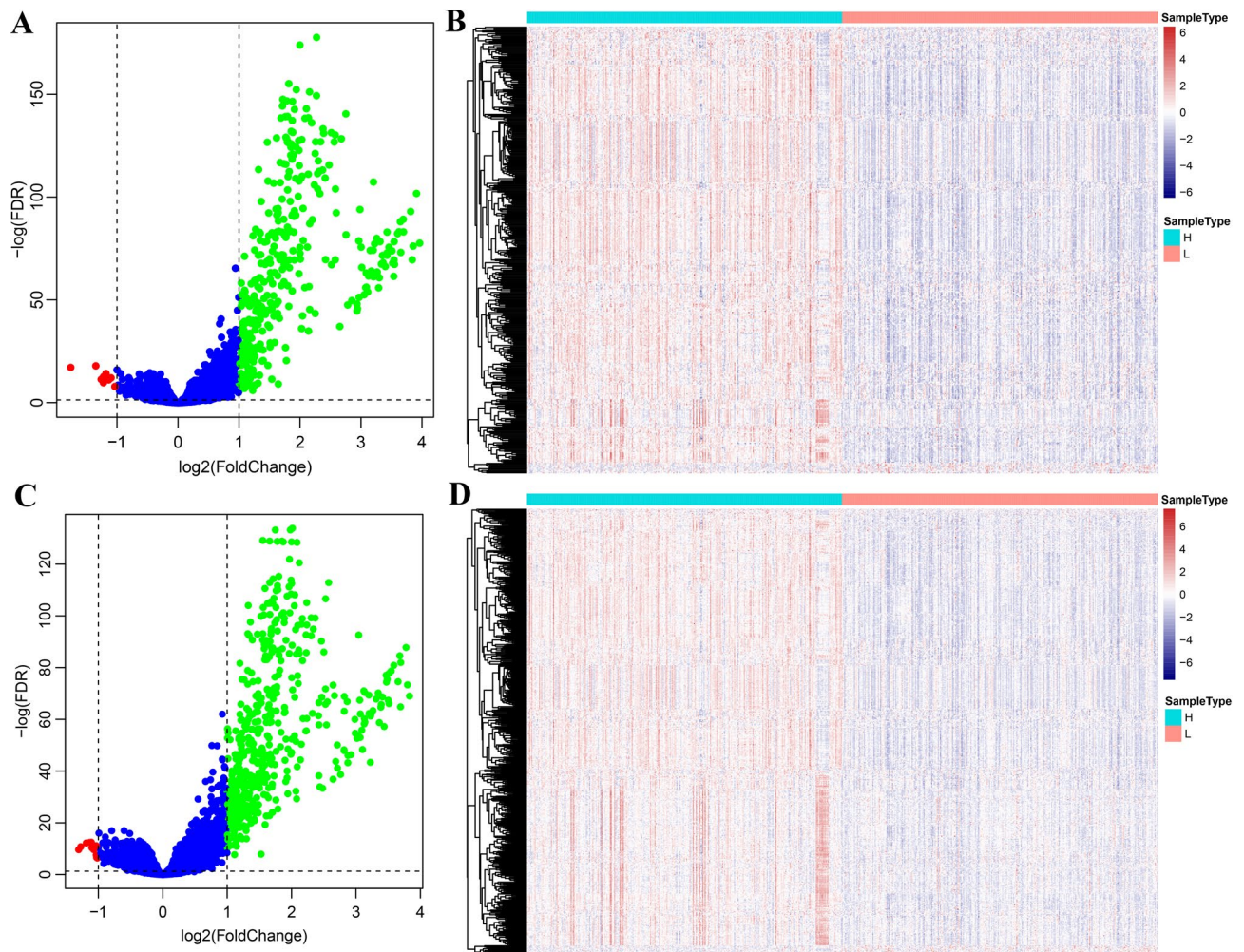


Figure 8. Gene expression profile of high- and low-ImmuneScore /ESTIMATEScore groups. (A,C) Volcanic maps. (B,D) Heatmap.

Tumor microenvironment, which is comprised of tumor cells, stromal cells and the secreted inflammatory mediators and cytokines, is a complex system and supports various tumor biological behaviors, including tumor genesis, progression, metastasis and so on. Stromal cells are mainly made up of immunocytes, fibroblasts, mesenchymal cells and tumor-associated endothelial cells. The abnormal infiltration of stromal cells, especially for immunocytes (such as neutrophils, monocytes and lymphocytes), has been verified in numerous studies^{11,31}. For now, great attention has been paid to the association between immune system and tumor biology³². A growing number of studies have not only revealed the interaction between EC cells and the host immune system, but also enhanced the efficacy of immunotherapies³³. As mentioned before, EC has been widely accepted as a kind of immunogenic malignancy. The treatment of EC has reached a new milestone through artificially manipulating the tumor immune microenvironment. Therefore, it is of great importance for us to explore the immune-related prognostic and therapeutic biomarkers for EC³⁴.

In current study, we used the RNA-seq data downloaded from TCGA database to calculate 3 immune-related scores based on ESTIMATE algorithm, which showed marked correlation with the immune status, survival time, prognosis-related gene mutations, and molecular subtypes of EC. Next, ESTIMATE immune score-related gene modules were obtained by means of WGCNA. 109 immune-related genes were then screened out using differential expression analysis and their functions were examined through enrichment analysis. Survival analysis was then performed and 14 novel immune-related prognostic genes were finally screened out, among which, 12 genes were further confirmed having marked correlations with ImmuneScore values in GSE17025 dataset.

Our newly discovered 14 immune-related prognostic markers include WAS, GZMH, CD7, NKG7, LINC01871, TRAC, CD8A, TRBC2, CD3E, IGSF6, RASAL3, ITGAL, S1PR4 and CCL4. Unfortunately, only a few studies have explored the roles of these genes in EC. Among these, 4 genes (CD8A, CD3E, CCL4 and ITGAL) were reported to have close relationship with tumor immune microenvironment and to be involved in various pathological processes of EC^{35–37}. TRBC2 encodes a specific region of the T-cell receptor beta-2 chain³⁸ and has been identified as a promising biomarker for the distinction of multiple cancer types, including breast cancer, colorectal cancer, glioblastoma, hepatobiliary cancer, lung cancer and pancreatic cancer and so on³⁹. It is worth mentioning that, our findings demonstrated that TRBC2 had the most significant correlation with

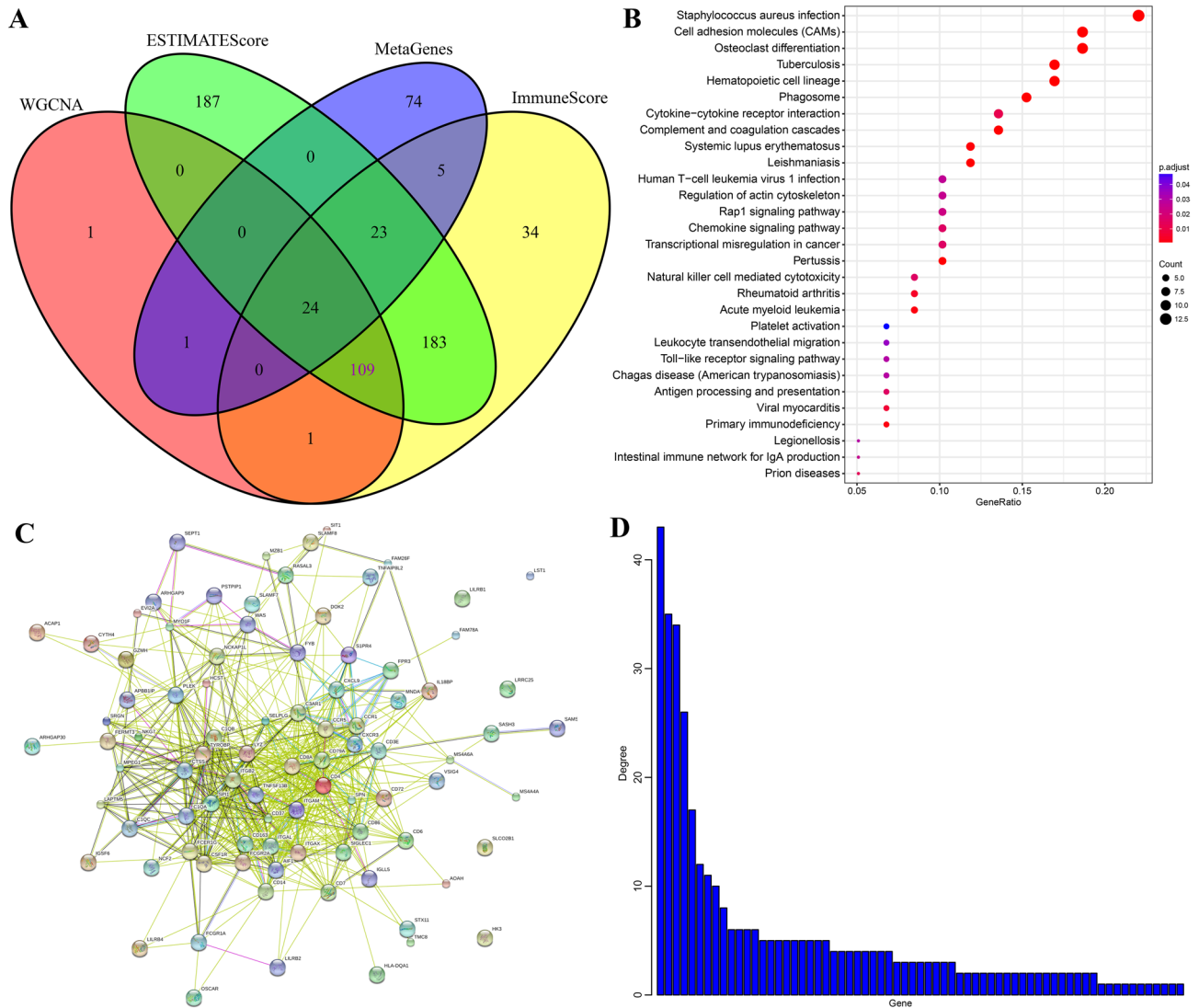


Figure 9. Mining of the immune-related genes. **(A)** The co-expressed genes that were markedly correlated with ImmuneScore/ESTIMATEScore. **(B)** KEGG enrichment analysis of the 109 genes. **(C)** Protein-protein interaction analysis of the 109 genes. **(D)** The degree value of each node within protein-protein interaction network.

Gene symbol	Gene name	p value (OS)	p value (PFS)
WAS	Wiskott-Aldrich syndrome	0.005103711	0.033465715
GZMH	Granzyme H	0.008033252	0.006877271
CD7	CD7 molecule	0.010859632	0.006642321
NKG7	Natural killer cell granule protein 7	0.000412089	0.030792653
LINC01871	Long intergenic non-protein coding RNA 1871	0.000140799	0.007739673
TRAC	T cell receptor alpha constant	4.40E-05	0.009042146
CD8A	CD8a molecule	0.003341004	0.012404753
TRBC2	T cell receptor beta constant	1.03E-05	0.001855756
CD3E	CD3e molecule	0.000106431	0.002676012
IGSF6	Immunoglobulin superfamily member 6	0.00173721	0.043427798
RASAL3	RAS protein activator like 3	0.006739461	0.039618006
ITGAL	Inegrin subunit alpha L	0.004923864	0.041623366
S1PR4	Sphingosine-1-phosphate receptor 4	0.007261307	0.029160931
CCL4	C-C motif chemokine ligand 4	0.001976119	0.034895726

Table 1. 14 immune-related genes positively correlated with both OS and PFS of EC.

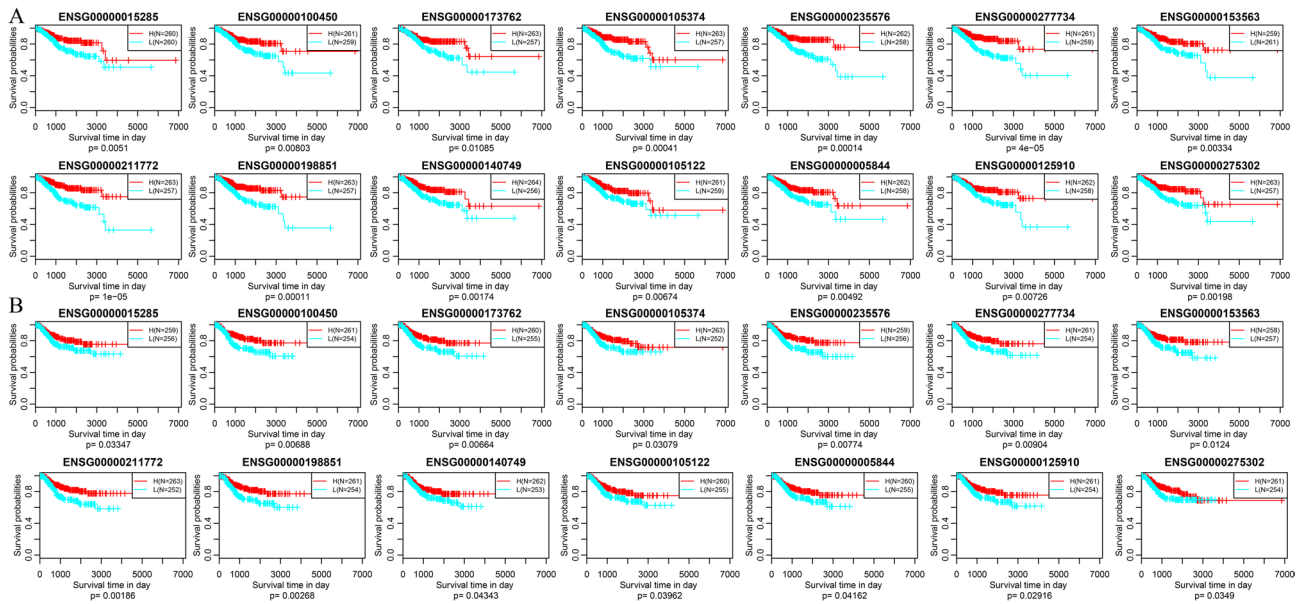


Figure 10. Correlations between 14 novel immune-related genes and OS (A) and PFS (B) of EC patients. *H* high-expression, *L* low-expression.

both OS and PFS of EC patients. Nevertheless, no studies have validated the specific role of TRBC2 in EC yet. Thus, it really deserves further study to elucidate the clinical importance and underlying molecular mechanism of TRBC2 in EC.

Conclusion

Briefly, our current study focuses on gene features associated with EC immune microenvironment. According to our findings, these genes participate in the progression of EC, and affect patient prognosis. Our work helps to investigate the complicated interactions in EC microenvironment. At the same time, our work may help to develop novel potential prognostic biomarkers and therapeutic targets for EC.

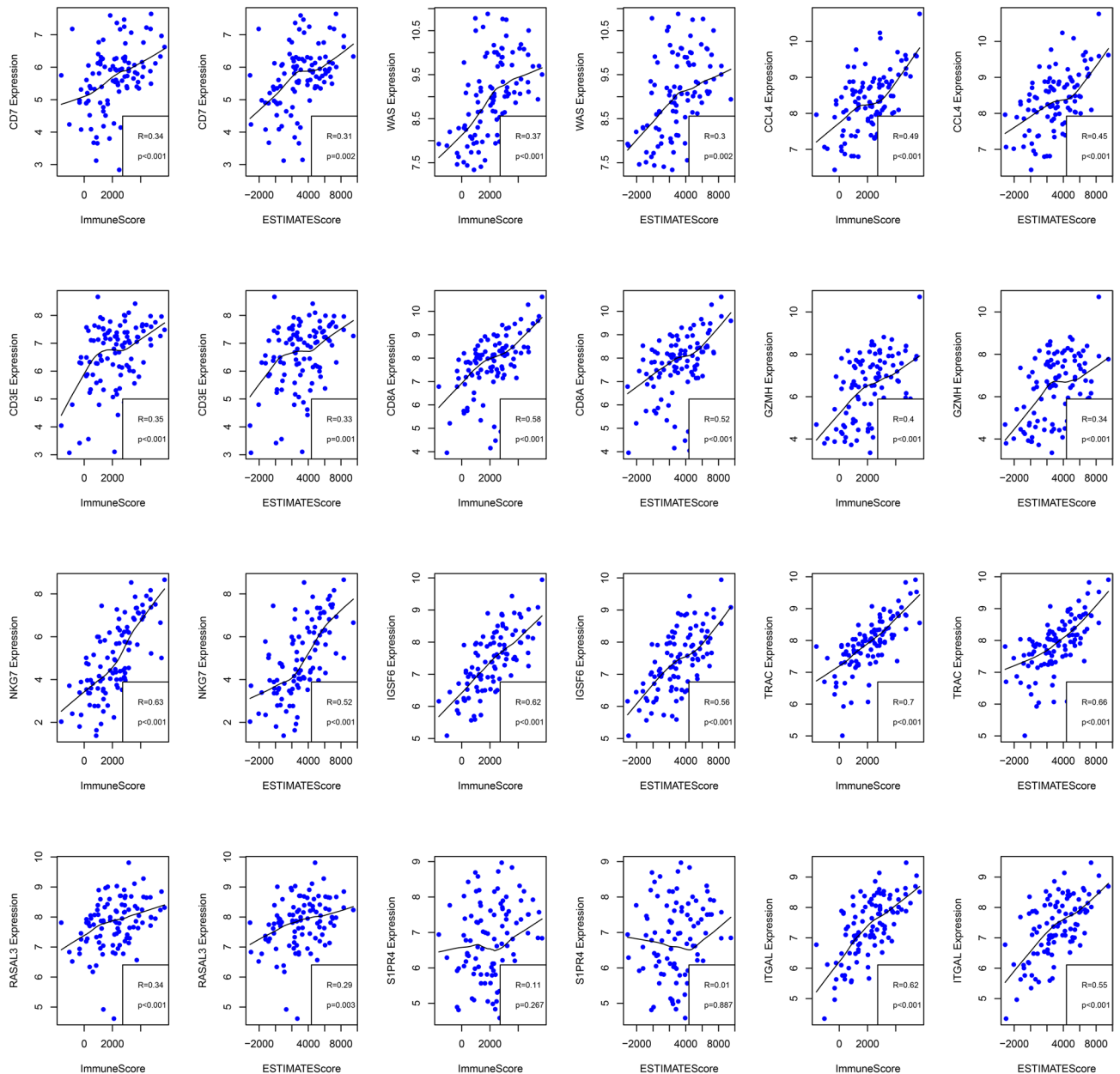


Figure 11. Correlations between 12 immune-related prognostic genes and ImmuneScore/ESTIMATEScore within GSE17025 dataset.

Data availability

The data generated are included in the manuscript and supplementary data. All the data we used in our study are publicly accessible at TCGA and NCBI GEO database.

Received: 24 September 2020; Accepted: 3 June 2021

Published online: 15 June 2021

References

- Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **68**, 394–424. <https://doi.org/10.3322/caac.21492> (2018).
- Sheikh, M. A. *et al.* USA endometrial cancer projections to 2030: Should we be concerned?. *Fut. Oncol.* **10**, 2561–2568. <https://doi.org/10.2217/fon.14.192> (2014).
- Jiang, X., Tang, H. & Chen, T. Epidemiology of gynecologic cancers in China. *J. Gynecol. Oncol.* **29**, e7. <https://doi.org/10.3802/jgo.2018.29.e7> (2018).
- Rutgers, J. K. Update on pathology, staging and molecular pathology of endometrial (uterine corpus) adenocarcinoma. *Fut. Oncol.* **11**, 3207–3218. <https://doi.org/10.2217/fon.15.262> (2015).
- Lortet-Tieulent, J., Ferlay, J., Bray, F. & Jemal, A. International patterns and trends in endometrial cancer incidence, 1978–2013. *J. Natl Cancer Inst.* **110**, 354–361. <https://doi.org/10.1093/jnci/djx214> (2018).

6. Gottwald, L. *et al.* Long-term survival of endometrioid endometrial cancer patients. *Arch. Med. Sci. AMS* **6**, 937–944. <https://doi.org/10.5114/aoms.2010.19305> (2010).
7. Chaudhry, P. & Asselin, E. Resistance to chemotherapy and hormone therapy in endometrial cancer. *Endocr. Relat. Cancer* **16**, 363–380. <https://doi.org/10.1677/ERC-08-0266> (2009).
8. Zhou, M., Zhang, Z., Zhao, H., Bao, S. & Sun, J. A novel lncRNA-focus expression signature for survival prediction in endometrial carcinoma. *BMC Cancer* **18**, 39. <https://doi.org/10.1186/s12885-017-3983-0> (2018).
9. Di Tucci, C. *et al.* Immunotherapy in endometrial cancer: New scenarios on the horizon. *J. Gynecol. Oncol.* **30**, e46. <https://doi.org/10.3802/jgo.2019.30.e46> (2019).
10. Ventriglia, J. *et al.* Immunotherapy in ovarian, endometrial and cervical cancer: State of the art and future perspectives. *Cancer Treat. Rev.* **59**, 109–116. <https://doi.org/10.1016/j.ctrv.2017.07.008> (2017).
11. Cai, Y., Chang, Y. & Liu, Y. Multi-omics profiling reveals distinct microenvironment characterization of endometrial cancer. *Biomed. Pharmacother.* **118**, 109244. <https://doi.org/10.1016/j.biopha.2019.109244> (2019).
12. Sungu, N. *et al.* Expression of immunomodulatory molecules PD-1, PD-L1, and PD-L2, and their relationship with clinicopathologic characteristics in endometrial cancer. *Int. J. Gynecol. Pathol.* **38**, 404–413. <https://doi.org/10.1097/PGP.0000000000000543> (2019).
13. Ghanekar, Y. & Sadasivam, S. In silico analysis reveals a shared immune signature in CASP8-mutated carcinomas with varying correlations to prognosis. *PeerJ* **7**, e6402. <https://doi.org/10.7717/peerj.6402> (2019).
14. Walentowicz-Sadlecka, M. *et al.* The analysis of human leukocyte antigen-G level in patients with endometrial cancer by Western blot technique. *Am. J. Reprod. Immunol.* **81**, e13070. <https://doi.org/10.1111/aji.13070> (2019).
15. Yang, J., Guo, X., Wu, T., Niu, K. & Ma, X. Prognostic significance of inflammation-based indexes in patients with stage III/IV colorectal cancer after adjuvant chemoradiotherapy. *Medicine* **98**, e14420. <https://doi.org/10.1097/MD.00000000000014420> (2019).
16. Liu, J. *et al.* Systemic immune-inflammation index, neutrophil-to-lymphocyte ratio, platelet-to-lymphocyte ratio can predict clinical outcomes in patients with metastatic non-small-cell lung cancer treated with nivolumab. *J. Clin. Lab. Anal.* <https://doi.org/10.1002/jcla.22964> (2019).
17. Huang, H. *et al.* Prognostic value of preoperative systemic immune-inflammation index in patients with cervical cancer. *Sci. Rep.* **9**, 3284. <https://doi.org/10.1038/s41598-019-39150-0> (2019).
18. Safonov, A. *et al.* Immune gene expression is associated with genomic aberrations in breast cancer. *Can. Res.* **77**, 3317–3324. <https://doi.org/10.1158/0008-5472.CAN-16-3478> (2017).
19. Li, T. *et al.* TIMER: A web server for comprehensive analysis of tumor-infiltrating immune cells. *Can. Res.* **77**, e108–e110. <https://doi.org/10.1158/0008-5472.CAN-17-0307> (2017).
20. Langfelder, P. & Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **9**, 559. <https://doi.org/10.1186/1471-2105-9-559> (2008).
21. Szklarczyk, D. *et al.* The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* **45**, D362–D368. <https://doi.org/10.1093/nar/gkw937> (2017).
22. Jiang, T. *et al.* Identification of potential prostate cancer-related pseudogenes based on competitive endogenous RNA network hypothesis. *Med. Sci. Monit.* **24**, 4213–4239. <https://doi.org/10.12659/MSM.910886> (2018).
23. Cancer Genome Atlas Research *et al.* Integrated genomic characterization of endometrial carcinoma. *Nature* **497**, 67–73. <https://doi.org/10.1038/nature12113> (2013).
24. Liu, J. *et al.* Exploration of a novel prognostic risk signatures and immune checkpoint molecules in endometrial carcinoma microenvironment. *Genomics* **112**, 3117–3134. <https://doi.org/10.1016/j.ygeno.2020.05.022> (2020).
25. Levine, D. A. *et al.* Risk of endometrial carcinoma associated with BRCA mutation. *Gynecol. Oncol.* **80**, 395–398. <https://doi.org/10.1006/gyno.2000.6082> (2001).
26. Liu, Z. *et al.* A novel loss-of-function mutation in TP53 in an endometrial cancer cell line and uterine papillary serous carcinoma model. *Mol. Cell Biochem.* **297**, 179–187. <https://doi.org/10.1007/s11010-006-9345-x> (2007).
27. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550. <https://doi.org/10.1186/s13059-014-0550-8> (2014).
28. Jeanquartier, F., Jean-Quartier, C. & Holzinger, A. Integrated web visualizations for protein-protein interaction databases. *BMC Bioinform.* **16**, 195. <https://doi.org/10.1186/s12859-015-0615-z> (2015).
29. Day, R. S. *et al.* Identifier mapping performance for integrating transcriptomics and proteomics experimental results. *BMC Bioinform.* **12**, 213. <https://doi.org/10.1186/1471-2105-12-213> (2011).
30. Wang, Y., Xu, M. & Yang, Q. A six-microRNA signature predicts survival of patients with uterine corpus endometrial carcinoma. *Curr. Probl. Cancer* **43**, 167–176. <https://doi.org/10.1016/j.currproblcancer.2018.02.002> (2019).
31. Sahoo, S. S., Zhang, X. D., Hondermarck, H. & Tanwar, P. S. The emerging role of the microenvironment in endometrial cancer. *Cancers* <https://doi.org/10.3390/cancers10110408> (2018).
32. Liu, Y. Immune response characterization of endometrial cancer. *Oncotarget* **10**, 982–992. <https://doi.org/10.18632/oncotarget.26630> (2019).
33. Arora, E. *et al.* Role of immune checkpoint inhibitors in advanced or recurrent endometrial cancer. *Cureus* **10**, e2521. <https://doi.org/10.7759/cureus.2521> (2018).
34. Vanderstraeten, A., Tuyaerts, S. & Amant, F. The immune system in the normal endometrium and implications for endometrial cancer development. *J. Reprod. Immunol.* **109**, 7–16. <https://doi.org/10.1016/j.jri.2014.12.006> (2015).
35. Suemori, T. *et al.* Intratumoral CD8+ lymphocyte infiltration as a prognostic factor and its relationship with cyclooxygenase 2 expression and microsatellite instability in endometrial cancer. *Int. J. Gynecol. Cancer* **25**, 1165–1172. <https://doi.org/10.1097/IGC.0000000000000482> (2015).
36. Jedryka, M. *et al.* Matrix metalloproteinase (MMP)-2 and MMP-9 expression in tumor infiltrating CD3 lymphocytes from women with endometrial cancer. *Int. J. Gynecol. Cancer* **22**, 1303–1309. <https://doi.org/10.1097/IGC.0b013e318269e27b> (2012).
37. Jozwik, M. *et al.* Surface antigen expression on peripheral blood monocytes in women with gynecologic malignancies. *BMC Cancer* **15**, 129. <https://doi.org/10.1186/s12885-015-1136-x> (2015).
38. Li, L., Dong, M. & Wang, X. G. The implication and significance of beta 2 Microglobulin: A conservative multifunctional regulator. *Chin. Med. J.* **129**, 448–455. <https://doi.org/10.4103/0366-6999.176084> (2016).
39. Zhang, Y. H. *et al.* Identifying and analyzing different cancer subtypes using RNA-seq data of blood platelets. *Oncotarget* **8**, 87494–87511. <https://doi.org/10.18632/oncotarget.20903> (2017).

Authors' Contributions

P.C. and J.M. conceived, designed the study, analyzed the data and prepared the manuscript, X.Z. and C.Z. collected the data and organized the figures, X.C. reviewed and edited the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by Chinese National Natural Science Foundation (81902629).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-92091-5>.

Correspondence and requests for materials should be addressed to P.C.

Reprints and permissions information is available at www.nature.com/reprints.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021