Mining of prognosis-related genes in cervical squamous cell carcinoma immune microenvironment

Jiong Ma [Equal first author, 1], **Pu Cheng** [Equal first author, 1, 2], **Xuejun Chen** [1], **Chunxia Zhou** [1], **Wei Zheng** [Corresp. 1]

[1] Department of Gynecology, Second Affiliated Hospital, Zhejiang University School of Medicine, Hang Zhou, China

[2] Key Laboratory of Tumor Microenvironment and Immune Therapy of Zhejiang Prov-ince, Hang Zhou, China

Corresponding Author: Wei Zheng
Email address: zhengwei@zju.edu.cn

Purpose: The aim of this study was to explore the effective immune scoring method and mine the novel and potential immune microenvironment-related diagnostic and prognostic markers for cervical squamous cell carcinoma (CESC). **Materials and Methods:** The Cancer Genome Atlas (TCGA) data was downloaded and multiple data analysis approaches were initially used to search for the immune-related scoring system on the basis of ESTIMATE algorithm. Afterwards, the representative genes in the gene modules correlated with immune-related scores based on ESTIMATE algorithm were further screened using WGCNA and network topology analysis. Gene functions were mined through enrichment analysis, followed by exploration of the correlation between these genes and immune checkpoint genes. Finally, survival analysis was applied to search for genes with significant association with overall survival and external database was employed for further validation. **Results:** The immune-related scores based on ESTIMATE algorithm was closely associated with other categories of scores, the HPV infection status, prognosis and the mutation levels of multiple CESC-related genes (HLA and TP53). 18 new representative immune microenvironment-related genes were finally screened closely associated with patient prognosis and were further validated by the independent dataset GSE44001. **Conclusion:** Our present study suggested that the immune-related scores based on ESTIMATE algorithm can help to screen out novel immune-related diagnostic indicators, therapeutic targets and prognostic predictors in CESC.

Mining of Prognosis-related Genes in Cervical Squamous Cell Carcinoma Immune Microenvironment

Jiong Ma[1,*], Pu Cheng[1,2,*], Xuejun Chen[1], Chunxia Zhou[1], Wei Zheng[1,#]

[1] Department of Gynecology, Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China

[2] Key Laboratory of Tumor Microenvironment and Immune Therapy of Zhejiang Province

[*]These authors contributed equally to this work.

[#] Correspondence: Wei Zheng

E-mail address: zhengwei@zju.edn.cn

Abstract

Purpose: The aim of this study was to explore the effective immune scoring method and mine the novel and potential immune microenvironment-related diagnostic and prognostic markers for cervical squamous cell carcinoma (CESC).

Materials and Methods: The Cancer Genome Atlas (TCGA) data was downloaded and multiple data analysis approaches were initially used to search for the immune-related scoring system on the basis of ESTIMATE algorithm. Afterwards, the representative genes in the gene modules correlated with immune-related scores based on ESTIMATE algorithm were further screened using WGCNA and network topology analysis. Gene functions were mined through enrichment analysis, followed by exploration of the correlation between these genes and immune checkpoint genes. Finally, survival analysis was applied to search for genes with significant association with overall survival and external database was employed for further validation.

Results: The immune-related scores based on ESTIMATE algorithm was closely associated with other categories of scores, the HPV infection status, prognosis and the mutation levels of multiple CESC-related genes (HLA and TP53). 18 new representative immune microenvironment-related genes were finally screened closely associated with patient prognosis and were further validated by the independent dataset GSE44001.

Conclusion: Our present study suggested that the immune-related scores based on ESTIMATE algorithm can help to screen out novel immune-related diagnostic indicators, therapeutic targets and prognostic predictors in CESC.

Keywords: cervical carcinoma, TCGA, immune, prognosis

Introduction

Cervical squamous cell carcinoma (CESC) is one of the most common malignancies in female reproductive system, which severely threatens female health and life quality(1). CESC is highly prevalent in developing countries, accounting for 60-90% of global cases(2). Radical

hysterectomy is currently considered as the dominant therapy for early-stage cervical cancer(3,4). With the popularization of cervical cancer screening, the therapeutic efficacy and prognosis of early-stage patients has been greatly improved(5,6). Postoperative relapse and metastasis of CESC remain the major causes of death in clinical practice(7,8). Patients with advanced-stage CESC generally undergo adjuvant radiotherapy and/or chemotherapy, however, the therapeutic effect seems unsatisfactory(9,10). At present, the International Federation of Gynaecology and Obstetrics (FIGO) staging classification is the major criterion for the prognostic prediction of patients with CESC(11). Nevertheless, CESC patients within similar clinical stage usually show diverse prognostic outcomes. Thus, there is an urgent need to identify high-risk subgroups for individualized monitoring and optimized postoperative therapy in routine clinical practice.

Given the increasing evidence that various immune cells and inflammatory mediators are closely associated with the development of CESC, tumor microenvironment is drawing accumulating attention nowadays(12). The leukocytes, neutrophils, lymphocytes and macrophages directly contribute to the immune response, which could be easily and conveniently detected(13-16). In the last decade, various studies have investigated the relationship between the prognosis of patients with primary CESC and the immunological landscape through high-throughput quantitative measurements of cellular and molecular characteristics(17,18). These studies revealed the great heterogeneity of the inflammatory/immune response in CESC, which might determine to a large extent the final outcome of patients(19). More recently, several researchers proposed a novel classification based on the immunological status of CESC according to the ratio of different immune cells (such as monocyte/lymphocyte ratio or Th17/Treg ratio) in the tumor microenvironment, which might play a significant role in the accurate prediction of patient prognosis(20,21). Unfortunately, almost none of the previous studies have reached clinical practice because of lacking the exploration from large sample data.

On this account, multiple immune scoring methods have been exploited using the expression data of immune-related genes in TCGA database which enable us to quantify the immune microenvironment status of a specific patient(22,23). For instance, the systemic immune-

inflammation index (SII) established according to peripheral lymphocyte, neutrophil and platelet counts has been considered as a good indicator reflecting the local immune response and systemic inflammation(24). Moreover, SII has been confirmed to have remarkable association with the prognosis of numerous tumors, including non-small cell lung cancer(25), esophageal cancer(26) and colorectal cancer(27). However, there have been only limited studies concerning CESC up to now.

To this end, our present study was designed to explore the immune scoring method suitable for CESC. In addition, the gene members in the scoring system were further analyzed by a series of bioinformatic means to mine the novel and potential immune microenvironment-related diagnostic and prognostic markers.

Materials and methods

Database sources and pre-processing

The RNA-seq counts data, SNP data, and clinical follow-up information were downloaded from the TCGA database. The FPKM data of RNA-Seq were transformed into TPM expression profiles. In consistent with previous studies, 13 metagenes (shown in ImmuneScore.genes.ids.txt) corresponded to various immunocyte types, reflecting the different immune functions.

Computational methods of multiple immune scores and result determination

The scores of each sample in the 13 types of metagenes were calculated based on the log2-transformed expression of each gene member in the immune metagene (shown in immune.meta.score.txt)(28). TIMER (https://cistrome.shinyapps.io/timer/) database (immune.immu.score.txt) was utilized to calculate the scores of each sample in the immunocyte infiltration (six categories in total)(29). Moreover, the ImmuneScore, StromalScore and ESTIMATEScore of each sample (immune.est.score.txt) were calculated by ESTIMATE function of R software package(30). Finally, R software package MCPcounter was utilized for the calculation of the abundances of ten immune-related cell (eight categories of immune cells, endothelial cells and fibroblasts) populations in the tumor microenvironment

(immune.MCPcounter.score.txt).

Survival analysis

Patients were divided into several groups according to each specific parameter (including ImmuneScore, StromalScore, ESTIMATEScore and gene expression level). Afterwards, the association between the gene expression level (or level of ImmuneScore, StromalScore, ESTIMATEScore) and overall survival was analyzed by univariate Cox regression model.

The construction of immune scores-related gene modules through WGCNA

To begin with, transcripts with over 75% TPM of >1 and median absolute deviation (MAD) of >median were chosen from the expression profile data of all the obtained samples. Hierarchical clustering for cluster analysis of the samples was also adopted. Subsequently, samples with a distance of over 80000 were taken as the outlier samples for screening. Moreover, the distance between any two transcripts was calculated by Pearson correlation coefficient, the establishment of the distance between any two transcripts was performed by the R software package WGCNA(31), and the soft threshold was set as eight for the screening of the co-expression modules. The co-expression network has been suggested to conform to the scale-free network. In other words, the logarithm of node with the connectivity of k (log(k)) should be negatively correlated with the logarithm of the occurrence probability of the specific node (log(P(k)), and the correlation coefficient should be >0.85. Proper β value was selected in order to ensure the network as a scale-free network. The expression matrix was subsequently transformed into the adjacent matrix, and the latter was further transformed into the topological matrix for gene clustering based on TOM utilizing the average-linkage hierarchical clustering method in accordance with the mixed dynamic shear tree standard. In addition, the gene number of each gene network module was set at least 30. The dynamic shear method was used to determine the gene module, followed by calculation of the eigengene value of each module in succession. Afterwards, clustering analysis was performed on the modules, in which, modules close to each other were merged into a new module, with re-set appropriate height, deepSplit and minModuleSize values. Finally, the association of the acquired gene modules with ImmuneScore,

StromalScore and ESTIMATEScore were separately calculated, in order to explore the gene modules with high correlation for further research.

Establishment of the gene interaction network and functional analysis

Genes were mapped into the String database(32). The gene-gene interactions were acquired at the score threshold of >0.4, followed by visualization using Cytoscape software. Meanwhile, KEGG and GO enrichment analysis was performed by utilizing the clusterprofile R package(33) to examine the signaling pathways affected by these genes.

Results

The immune-related scores based on ESTIMATE algorithm is the most suitable immune scoring method for CESC

To be specific, we retrieved CESC samples from the TCGA database and analyzed their scores in 23 types of scoring systems, including 13 types of metagenes scores, six types of immunocyte infiltration scores, three types of immune-related scores according to ESTIMATE algorithm (ImmuneScore, StromalScore and ESTIMATEScore) and 10 types of abundances of immune-related cell. In addition, Spearman's correlation coefficient was used to calculate the correlations among these scoring systems (shown in Fig.1). As shown in Fig.1A, the average correlation between different types of immune-related scores was greater than 0.4. The first three scoring systems with most obvious correlation with others including ImmuneScore (R=0.59), Co_inhibition(R=0.59) and LCK(R=0.62), indicating that the consistency among the immune scores calculated by different algorithms to a certain extent. The clustering heat maps of various types of scoring systems were shown in Fig.1B, suggesting the great correlation among the scoring systems MHC1, MHC2, Monocytic lineage, Dendritic, Macrophages, ESTIMATEScore, ImmuneScore, Tfh, LCK, Co_stimulation, Co_inhibition, Mete_ImmuneScore, Neutrophil and STAT1. We further investigated the average correlation among immune scores according to four different algorithms. As shown in Fig.1C, the immune-related scores calculated by the ESTIMATE algorithm harbored the highest average correlation with the other three algorithms,

which is greater than 0.52 on average. These findings implicated that the immune-related scores based on ESTIMATE algorithm were the most representative immune scoring methods for CESC.

It is widely accepted that HPV infection has a significant association with the occurrence and progression of CESC(34). Therefore, we separately analyzed the ImmuneScore, StromalScore and ESTIMATEScore distribution among CESC patients with or without HPV infection. As shown in Fig.2A-C, the three immune-related scores in CESC with HPV infection were significantly higher than those without HPV infection. It should be noted that ImmuneScore was most significantly correlated with the infection status of HPV ($p<0.05$).

Subsequently, in order to investigate the association between the above three immune-related scores and prognosis, samples were sorted based on the median of scores of all samples. And then, prognostic difference was analyzed by Kaplan-Meier method (shown in Fig.3). As a result, the prognosis of samples in different groups was significantly different. And the five-year survival rate of samples with high ImmuneScore and ESTIMATEScore were significantly superior in comparison with those with low scores, suggesting that the three immune-related scores on the basis of ESTIMATE algorithm could be accepted as promising novel prognostic markers for CESC.

A large number of somatic mutations of HLA genes have been reported in CESC, strongly indicating that loss of function due to HLA mutations is tightly correlated with the immune escape of cancer cells(35). It is of great significance for us to analyze the changes of HLA gene sequence in tumor patients. In addition, the mutation of TP53, a tumor suppressor gene, can induce unlimited proliferation and apoptosis resistance of tumor cells(36,37). Next, we focused on analyzing the associations of three immune-related scores with mutations of HLA and TP53. To this end, we extracted the mutation data of HLA-A, HLA-B, HLA-C and TP53 from the mutect-processed SNP database and then calculated the three immune-related scores based on ESTIMATE algorithm in HLA-A, HLA-B, HLA-C and TP53 mutation and non-mutation groups. As shown in Fig.4, there was higher level of ImmuneScore in HLA-A and HLA-B mutation

groups compared with wild-type groups, while there was also higher level of ESTIMATEScore in HLA-B mutation groups but lower level in TP53 mutation groups comparison with that in wild-type groups.

In summary, we demonstrated that the immune-related scores on the basis of ESTIMATE algorithm were the most proper immune scoring method for CESC. Additionally, the co-expressed genes with remarkable correlation with these three immune-related scores might be considered as the representative genes in CESC immune microenvironment, which could be further validated as potential prognostic markers and novel therapeutic targets of CESC.

Screening of the representative genes in the immune scores-related gene modules

In this section, clustering analysis was first conducted through hierarchical clustering. As shown Fig.5A, a total of 296 samples were finally screened out among all the outlier samples, which had a distance of larger than 80000. Subsequently, the weight co-expression network was constructed by WGCNA with $\beta=8$ to guarantee the scale-free network (Fig.5B, C). Afterwards, dynamic shear method was utilized to determine the gene modules, and clustering analysis was performed on these modules. Additionally, modules with close distance were further merged into the new module, having height, deepSplit and minModuleSize set to 0.25, 2 and 30, respectively. Finally, a total of 30 modules were acquired (Fig.5D). Of note, the grey module indicated gene sets that could not be clustered into other modules. The transcripts of each module were counted and displayed in Table 1. In total, 6679 transcripts were allocated to 29 co-expression modules. The correlations of the eigenvectors of these 30 modules with ImmuneScore, StromalScore and ESTIMATEScore were subsequently calculated, respectively. As shown in Fig.5E, the yellow module obviously harbored extremely high association with these three immune-related scores based on ESTIMATE algorithm containing 422 genes.

The gene functions in the yellow module were subsequently analyzed. Meanwhile, KEGG and GO enrichment analysis was also conducted using the clusterProfiler of R software package, with FDR set as <0.05. The detailed enrichment results were shown in yellow enrich.txt. As a result, the genes in the yellow module were enriched into 50 KEGG pathways, 670 GO

biological processes, 85 GO cell compositions and 74 molecular functions. The most significant top 20 KEGG pathways and GO terms were shown in Fig.6. The enriched pathways mainly included Th1 and Th2 cell differentiation, cytokine-cytokine receptor interaction and so on. And the enriched biological processes primarily included T cell activation, leukocyte cell-cell adhesion and so on. The enriched cell components mainly included MHC class II protein complex and T cell receptor complex, and so on. The enriched molecular functions mainly included cytokine receptor activity and MHC class II receptor activity, and the rest. Intriguingly, these enriched pathways and GO term have previously been reported to have close association with CESC and its immune microenvironment(38-41).

Finally, to further mine the immune scores-related genes, the weight co-expression relationship between genes in the yellow modules was calculated, with the weight threshold greater than 0.2. Cytoscape software was used for derivation and visualization of the co-expression network of these genes (as shown in Fig.7A). Afterwards, we further analyzed the topological properties of the network, which contained 244 nodes and 4083 edges, indicating that genes with greater association with modules had more close correlation with other genes in the network. As shown in Fig.7B, the degree distribution of the network was further analyzed, suggesting that the degree of the majority of nodes was extremely small, while the degree of a few nodes was rather large, which was consistent with the characteristics of biological network. The correlation between the gene and the module was further calculated. As shown in Fig.7C, the correlation between most genes and the module was over 0.6, suggesting a high expression similarity between the genes in the module. Moreover, a total of 26 genes (Table 2 and lst.genes.txt) with a correlation over 0.9 and a degree over 50 in the network were selected, with seven members of LCK Metagenes, and one member of Co_inhibition Metagenes. Thus, 18 new representative immune microenvironment-related genes were finally screened.

Function analysis of 18 novel representative immune microenvironment-related genes in CESC patients

Firstly, to further analyze the functions of these 18 novel representative immune

microenvironment-related genes, the R software package clusterProfiler was utilized for KEGG and GO enrichment analysis, with the significance FDR set at <0.05. The detailed results were summarized in lst enrich.txt. In brief, these 18 genes were enriched into 11 KEGG pathways, 202 GO biological processes, 8 GO cell components, 19 molecular functions. The most significant 20 KEGG pathways and GO terms were shown in Fig.8, the majority of which were involved in the proliferation, growth and differentiation of T cells. Intriguingly, LAPTM5, EVI2A and MS4A6A were not enriched in any signaling pathways and GO term, indicating that the functions of these three genes remained completely unclear, which is the focus of our further studies.

Secondly, to further investigate the potential roles of the 18 novel representative immune microenvironment-related genes in clinical practice, the R package corrgram was utilized for the calculation of the association between these genes and immune checkpoints (PDCD1、CD274、PDCD1LG2、CTLA4、CD86、CD80、CD276、VTCN1). As shown in Fig.9, apart from CD276 and VTCN1, the other six immune checkpoints were significantly related to these 18 genes, with an average correlation coefficient over 0.5, which indicated that these immune microenvironment-related genes might be promising targets for immunotherapy.

Finally, the prognostic significance of 18 novel representative immune microenvironment-related genes was assessed. According to the median of gene expression, samples were categorized into high and low expression groups. And then the differences of prognosis between these groups were analyzed. As shown in Fig.10, high expression of 13 genes were significantly associated with better overall survival according to the threshold of $p<0.05$, suggesting that these genes might be closely associated with patient prognosis.

Validation of the correlations of 18 immune microenvironment-related genes with ImmuneScore for CESC patients by using external dataset

External database was used for further validation of the correlations of 18 immune microenvironment-related genes with the immune-related scores according to ESTIMATE algorithm for CESC patients. MergeExpro contrib1-GPL14951.txt was downloaded from an independent dataset GSE44001(42) from GEO to extract the standardized expression matrix. R

packages hgu133plus2.db was utilized to map a probe for gene to extract the expression profiles of these 18 genes, followed by the calculation of the ImmuneScore for each sample using R software package ESTIMATE. Subsequently, the Pearson correlation was calculated between expression of these genes and the level of ImmuneScore for every CESC sample in this dataset. As shown in Fig.11, apart from CCR5, the other 17 genes were significantly associated with the ImmuneScore, which was consistent with our previous findings.

Discussion

Great attention has been paid to the association of the immune system with the pathogenesis and progression of tumor in recent years, which has shed light on CESC therapy, promoting the continuous development of anti-cancer therapy(43,44). The external anti-CESC approaches are frequently applied in previous clinical practice, including surgical resection and chemotherapy. However, the effect of surgical resection is generally restricted due to the invasion into adjacent tissues by cancer cells or distant metastasis. In addition, the application of chemotherapy is limited due to its toxicity to normal tissues(45). Thus, conventional therapies would exert great burden on the body while providing therapeutic benefits. To this end, it has been widely accepted as a novel direction of anti-cancer therapy by starting from the tumor origin, in other words, the immune system of human body, to control and even kill tumor cells via the modulation of the immune system and enhancement of the anti-tumor immunity in the tumor microenvironment(46).

The tumor microenvironment, mainly composed of immune cells, inflammatory cells, mesenchymal cells, tumor cells, stromal cells, inflammatory mediators and cytokines, provides support for tumor biological behavior including the pathogenesis, progression, invasion and metastasis(12,47,48). Therefore, it is of great significance to discover novel and meaningful immune microenvironment-related genes in CESC as prognostic predictor and therapeutic targets. In this study, the TCGA database was used to search for the immune microenvironment markers related to the survival time of CESC patients. And 18 genes were finally detected having

remarkable correlation with the prognosis of patients, which was further validated in the GEO database.

To be specific, firstly, multiple methods of data analysis were utilized to search for the three immune-related scores on the basis of ESTIMATE algorithm, showing high correlations with diverse other immune-related scores, patients prognosis, HPV infection status and the mutation levels of multiple well-defined CESC-related genes (HLA and TP53). Secondly, the representative genes in the gene modules associated with immune-related scores according to ESTIMATE algorithm were further searched using WGCNA and network topology analysis. Thirdly, we mined the gene functions through enrichment analysis, followed by the exploration of the association between these genes and immune checkpoint genes. Finally, survival analysis was employed to search for the genes with evident correlation with OS. In addition, external database was employed for further validation of the association of these immune microenvironment-related genes with ImmuneScore for CESC patients. In total, we successfully mined 18 novel potential immune microenvironment-related diagnostic and prognostic indicators or therapeutic targets.

Of note, 11 out of these 18 genes (IL10RA, CD4, HAVCR2, CD2, CCR5, CD3E, BTK, etc) have previously been demonstrated to participate in the pathogenesis, progression, malignant transformation, and pathological process of immune microenvironment of CESC, which are also significantly associated with patient survival, prognosis and diagnosis(18,49-51). These above-described observations validates the great reliability and accuracy of the bioinformatic mining results in our present study, in which, we combined TCGA database screening with GEO database for verification. However, the correlations of two genes (LAPTM5 and EVI2A) with CESC have never been confirmed by any basic or clinical studies, which we are most interested in. LAPTM5, Laptm5, a lysosomal transmembrane protein enhancing the degradation of several targets involved in immune signaling (such as ubiquitin-editing enzyme A20), has been validated to be participate in the modulation of the lethal T cell alloreactivity mediated by dendritic cells and immunoreactions in multiple inflammatory disease, such as GVHD(52,53). On the other

hand, EVI2A has been confirmed to be involved in lymphocyte proliferation and viability, which is a well-defined immune-specific tumor suppressor in head and neck cancer(54).

At present, accumulating studies focus on the mining of the association of numerous genes expression with the survival of CESC patients, however, the majority of previous studies are only performed in animal model, *in vitro* cell model or small sample samples of tumor patients. Thus, more comprehensive, large-scale population studies are required due to the complexity of CESC microenvironment. Fortunately, the rapid development of genome-wide sequencing renders the free utilization of high-throughput tumor databases, such as TCGA, making it possible to apply the bioinformatic big data for the large-scale CESC population.

In the present study, we mainly studied the CESC immune microenvironment-related gene characteristics. Consequently, these genes are involved in the pathogenesis, progression and malignant transformation of CESC, affecting OS of CESC patients. Our present findings can offer more information to decode the complex tumor-tumor interactions in CESC microenvironment. In addition, these findings will help to mine the novel immune-related diagnostic indicators, therapeutic targets and prognostic predictors in CESC.

Author Contribution

Wei Zheng: Project development, administration and supervision

Jiong Ma: Methodology development, manuscript wrtiing

Pu Cheng: Data collection and analysis, manuscript review and editing

Xuejun Chen and Chunxia Zhou: data collection, figure organization

Funding

We gratefully acknowledge the financial support from the Natural Science Fonudation of Zhejiang Province (LY15H040007) and Chinese National Natural Science Foundation (81902629).

335 ## Compliance with ethical standards

336 ### Conflict of interest

337 The authors declare no conflicts of interest in this work.

338

339 ### Ethical approval

340 This article does not contain any studies with human participants or animals performed by any of

341 the authors.

342

References

1. **Marth, C.*, et al.* (2018) Cervical cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol*, 29, iv262.**
2. **Chen, S.B.*, et al.* (2017) Clinicopathological features and surgical treatment of cervical oesophageal cancer. *Sci Rep*, 7, 3272.**
3. **Uppal, S.*, et al.* (2019) Trends and comparative effectiveness of inpatient radical hysterectomy for cervical cancer in the United States (2012-2015). *Gynecol Oncol*, 152, 133-138.**
4. **Gil-Moreno, A.*, et al.* (2019) Minimally Invasive or Abdominal Radical Hysterectomy for Cervical Cancer. *N Engl J Med*, 380, 794.**
5. **Altobelli, E.*, et al.* (2019) HPV-vaccination and cancer cervical screening in 53 WHO European Countries: An update on prevention programs according to income level. *Cancer Med*.**
6. **Ngo-Metzger, Q.*, et al.* (2019) Screening for Cervical Cancer. *Am Fam Physician*, 99, 253-254.**
7. **Nanthamongkolkul, K.*, et al.* (2018) Predictive Factors of Pelvic Lymph Node Metastasis in Early-Stage Cervical Cancer. *Oncol Res Treat*, 41, 194-198.**
8. **Alvarado-Cabrero, I.*, et al.* (2017) Factors Predicting Pelvic Lymph Node Metastasis, Relapse, and Disease Outcome in Pattern C Endocervical Adenocarcinomas. *Int J Gynecol Pathol*, 36, 476-485.**
9. **Bosque, M.A.S.*, et al.* (2018) Clinical and dosimetric factors associated with the development of hematologic toxicity in locally advanced cervical cancer treated with chemotherapy and 3D conformal radiotherapy. *Rep Pract Oncol Radiother*, 23, 392-397.**
10. **Angeles, M.A.*, et al.* (2019) Magnetic resonance imaging after external beam radiotherapy and concurrent chemotherapy for locally advanced cervical cancer helps to identify patients at risk of recurrence. *Int J Gynecol Cancer*, 29, 480-486.**
11. **Matsuo, K.*, et al.* (2019) Validation of the 2018 FIGO cervical cancer staging system. *Gynecol Oncol*, 152, 87-93.**
12. **Piersma, S.J. (2011) Immunosuppressive tumor microenvironment in cervical cancer patients. *Cancer Microenviron*, 4, 361-75.**
13. **Rangel-Corona, R.*, et al.* (2011) Cationic liposomes bearing IL-2 on their external surface induced**

mice leukocytes to kill human cervical cancer cells in vitro, and significantly reduced tumor burden in immunodepressed mice. *J Drug Target*, 19, 79-85.

14. Heintzelman, D.L., *et al.* (2000) Characterization of the autofluorescence of polymorphonuclear leukocytes, mononuclear leukocytes and cervical epithelial cancer cells for improved spectroscopic discrimination of inflammation from dysplasia. *Photochem Photobiol*, 71, 327-32.
15. Lu, X.Y., *et al.* (2018) Remifentanil on T lymphocytes, cognitive function and inflammatory cytokines of patients undergoing radical surgery for cervical cancer. *Eur Rev Med Pharmacol Sci*, 22, 2854-2859.
16. Chen, X.J., *et al.* (2019) The role of the hypoxia-Nrp-1 axis in the activation of M2-like tumor-associated macrophages in the tumor microenvironment of cervical cancer. *Mol Carcinog*, 58, 388-397.
17. Minion, L.E., *et al.* (2018) Cervical cancer - State of the science: From angiogenesis blockade to checkpoint inhibition. *Gynecol Oncol*, 148, 609-621.
18. Punt, S., *et al.* (2015) Correlations between immune response and vascularization qRT-PCR gene expression clusters in squamous cervical cancer. *Mol Cancer*, 14, 71.
19. Bachtiary, B., *et al.* (2006) Gene expression profiling in cervical cancer: an exploration of intratumor heterogeneity. *Clin Cancer Res*, 12, 5632-40.
20. Zhang, Y., *et al.* (2011) The imbalance of Th17/Treg in patients with uterine cervical cancer. *Clin Chim Acta*, 412, 894-900.
21. Huang, H., *et al.* (2019) Prognostic Value of Preoperative Systemic Immune-Inflammation Index in Patients with Cervical Cancer. *Sci Rep*, 9, 3284.
22. Shen, L., *et al.* (2018) A ten-long non-coding RNA signature for predicting prognosis of patients with cervical cancer. *Onco Targets Ther*, 11, 6317-6326.
23. Lee, Y.Y., *et al.* (2012) Clinical significance of changes in peripheral lymphocyte count after surgery in early cervical cancer. *Gynecol Oncol*, 127, 107-13.
24. Fest, J., *et al.* (2019) The systemic immune-inflammation index is associated with an increased risk of incident cancer-A population-based cohort study. *Int J Cancer*.

25. **Guo, W.*, et al.* (2019) Systemic immune-inflammation index (SII) is useful to predict survival outcomes in patients with surgically resected non-small cell lung cancer. *Thorac Cancer*, 10, 761-768.**

26. **Ishibashi, Y.*, et al.* (2018) Prognostic Value of Preoperative Systemic Immunoinflammatory Measures in Patients with Esophageal Cancer. *Ann Surg Oncol*, 25, 3288-3299.**

27. **Xie, Q.K.*, et al.* (2018) The systemic immune-inflammation index is an independent predictor of survival for metastatic colorectal cancer and its association with the lymphocytic response to the tumor. *J Transl Med*, 16, 273.**

28. **Safonov, A.*, et al.* (2017) Immune Gene Expression Is Associated with Genomic Aberrations in Breast Cancer. *Cancer Res*, 77, 3317-3324.**

29. **Li, B.*, et al.* (2016) Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. *Genome Biol*, 17, 174.**

30. **Yoshihara, K.*, et al.* (2013) Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun*, 4, 2612.**

31. **Xia, W.X.*, et al.* (2019) Identification of four hub genes associated with adrenocortical carcinoma progression by WGCNA. *PeerJ*, 7, e6555.**

32. **Szklarczyk, D.*, et al.* (2019) STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*, 47, D607-D613.**

33. **Xu, Z.*, et al.* (2019) Characterization of mRNA Expression and Endogenous RNA Profiles in Bladder Cancer Based on The Cancer Genome Atlas (TCGA) Database. *Med Sci Monit*, 25, 3041-3060.**

34. **Ding, L.*, et al.* (2019) Association of estradiol and HPV/HPV16 infection with the occurrence of cervical squamous cell carcinoma. *Oncol Lett*, 17, 3548-3554.**

35. **Xiao, X.*, et al.* (2013) HLA-A, HLA-B, HLA-DRB1 polymorphisms and risk of cervical squamous epithelial cell carcinoma: a population study in China. *Asian Pac J Cancer Prev*, 14, 4427-33.**

36. **Li, B.*, et al.* (2015) TP53 codon 72 polymorphism and susceptibility to cervical cancer in the Chinese population: an update meta-analysis. *Int J Clin Exp Med*, 8, 9055-62.**

37. **Laprano, T.D.*, et al.* (2014) Association of TP53 codon 72 and intron 3 16-bp Ins/Del polymorphisms

with cervical cancer risk. *Tumour Biol*, 35, 7435-40.

38. Yasmeen, A., *et al.* (2010) Locking Src/Abl Tyrosine Kinase Activities Regulate Cell Differentiation and Invasion of Human Cervical Cancer Cells Expressing E6/E7 Oncoproteins of High-Risk HPV. *J Oncol*, 2010.

39. Wang, X., *et al.* (2017) Overexpression of dendritic cell-specific intercellular adhesion molecule-3-grabbing nonintegrin-related protein in cervical cancer and correlation with squamous cell carcinoma antigen. *Oncol Lett*, 14, 2813-2821.

40. Zehbe, I., *et al.* (2005) Differential MHC class II component expression in HPV-positive cervical cancer cells: implication for immune surveillance. *Int J Cancer*, 117, 807-15.

41. Roca, A.M., *et al.* (2019) T-cell receptor V and J usage paired with specific HLA alleles associates with distinct cervical cancer survival rates. *Hum Immunol*, 80, 237-242.

42. Lee, Y.Y., *et al.* (2013) Genetic profiling to predict recurrence of early cervical cancer. *Gynecol Oncol*, 131, 650-4.

43. Dyer, B.A., *et al.* (2019) Role of Immunotherapy in the Management of Locally Advanced and Recurrent/Metastatic Cervical Cancer. *J Natl Compr Canc Netw*, 17, 91-97.

44. Orbegoso, C., *et al.* (2018) The current status of immunotherapy for cervical cancer. *Rep Pract Oncol Radiother*, 23, 580-588.

45. Menderes, G., *et al.* (2016) Immunotherapy and targeted therapy for cervical cancer: an update. *Expert Rev Anticancer Ther*, 16, 83-98.

46. Ring, K.L., *et al.* (2017) Potential immunotherapy targets in recurrent cervical cancer. *Gynecol Oncol*, 145, 462-468.

47. Qi, D., *et al.* (2019) Cancer prognosis: Considering tumor and its microenvironment as a whole. *EBioMedicine*.

48. Tuccitto, A., *et al.* (2019) Immunosuppressive circuits in tumor microenvironment and their influence on cancer treatment efficacy. *Virchows Arch*, 474, 407-420.

49. Hussain, S.K., *et al.* (2013) Nucleotide variation in IL-10 and IL-12 and their receptors and cervical and vulvar cancer risk: a hybrid case-parent triad and case-control study. *Int J Cancer*, 133, 201-13.

50. **Cao, Y.*, et al.* (2013) Tim-3 expression in cervical cancer promotes tumor metastasis. *PLoS One*, 8, e53834.**

51. **Che, L.F.*, et al.* (2016) Downregulation of CCR5 inhibits the proliferation and invasion of cervical cancer cells and is regulated by microRNA-107. *Exp Ther Med*, 11, 503-509.**

52. **Glowacka, W.K.*, et al.* (2012) LAPTM5 protein is a positive regulator of proinflammatory signaling pathways in macrophages. *J Biol Chem*, 287, 27691-702.**

53. **Hubbard-Lucey, V.M.*, et al.* (2014) Autophagy gene Atg16L1 prevents lethal T cell alloreactivity mediated by dendritic cells. *Immunity*, 41, 579-91.**

54. **Li, X.W.*, et al.* (2014) New insights into the DT40 B cell receptor cluster using a proteomic proximity labeling assay. *J Biol Chem*, 289, 14434-47.**

Figure legends

Fig.1. The correlations of immune-related scoring system based on ESTIMATE algorithm with other categories of scores among CESC samples. (A) The correlations of various immune scoring systems among CESC samples. Spearman correlation coefficients are shown color-coded to illustrate positive (red) or negative (green) associations. (B) The clustering heat maps of various types of scoring systems. (C) The relationships among immune scores according to four different algorithms.

Fig.2. StromalScore (A), ImmuneScore (B) and ESTIMATEScore (C) distribution among CESC patients with or without HPV infection.

Fig.3. The relationships between levels of StromalScore (A), ImmuneScore (B) or ESTIMATEScore (C) and prognosis for CESC patients.

Fig.4. The correlations of immune-related scores based on ESTIMATE algorithm with gene mutations. The StromalScore (A, D, G, J), ImmuneScore (B, E, H, K) and ESTIMATEScore (C, F, I, L) were calculated respectively in HLA-A (A, B, C), HLA-B (D, E, F), HLA-C (G, H, I) and TP53 (J, K, L) mutation and non-mutation groups. Green represents the mutant group and red represents the wild type.

Fig.5. Immune scores-related gene modules mined through WGCNA. (A) Sample clustering analysis. (B,C) Analysis of network topology for various soft-thresholding powers. (D) Gene dendrogram and module colors. (E) Correlation between each module and three immune-related scores.

Fig.6 The KEGG pathway and GO enrichment analysis of the genes in yellow module. (A) Top20 KEGG pathways enriched by the genes in yellow module. (B) Top20 GO BP terms enriched by the genes in yellow module. (C) Top20 GO CC terms enriched by the genes in yellow module. (D) Top20 GO MF terms enriched by the genes in yellow module.

Fig.7 Construction of co-expression network of yellow module-related genes. (A) Co-expression network of weights between genes in yellow module. (B) The degree distribution of nodes in yellow module. (C) The correlation of genes and module in the network.

Fig.8 The KEGG pathway and GO enrichment analysis of 18 novel representative immune microenvironment-related genes for CESC patients. (A) Top20 KEGG pathways enriched by 18 novel representative immune microenvironment-related genes. (B) Top20 GO BP terms enriched by 18 novel representative immune microenvironment-related genes. (C) Top20 GO CC terms enriched by 18 novel representative immune microenvironment-related genes. (D) Top20 GO MF terms enriched by 18 novel representative immune microenvironment-related genes.

Fig.9 The association between 18 novel representative immune microenvironment-related genes for CESC patients and immune checkpoints.

Fig.10. The relationship between 18 novel representative immune microenvironment-related genes and prognosis.

Fig.11 The correlations of 18 immune microenvironment-related genes with ImmuneScore for CESC patients in independent dataset.

Figure 1

The correlations of immune-related scoring system based on ESTIMATE algorithm with other categories of scores among CESC samples

(A) The correlations of various immune scoring systems among CESC samples. Spearman correlation coefficients are shown color-coded to illustrate positive (red) or negative (green) associations. (B) The clustering heat maps of various types of scoring systems. (C) The relationships among immune scores according to four different algorithms.

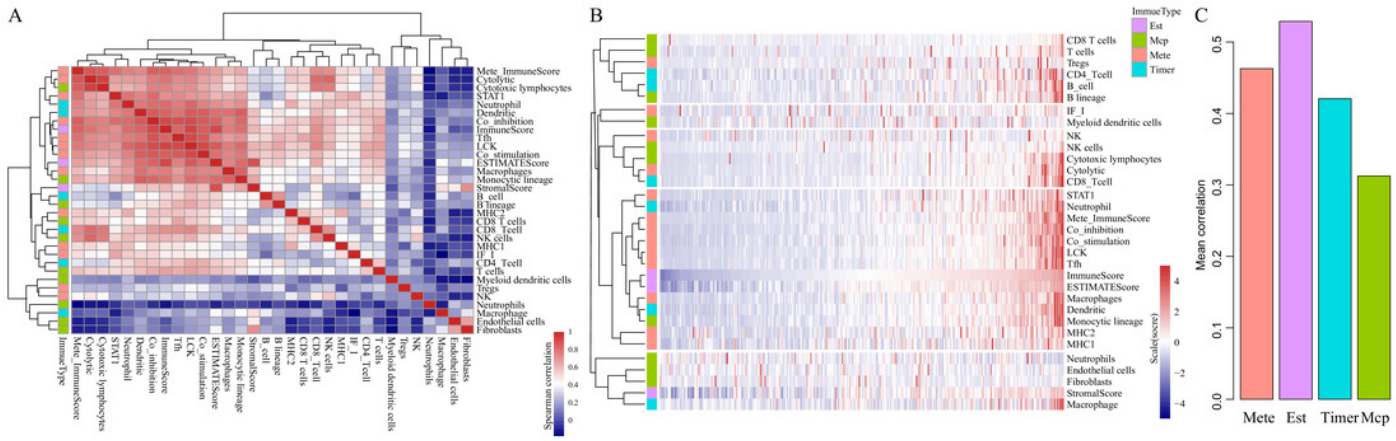

Figure 2

StromalScore (A), ImmuneScore (B) and ESTIMATEScore (C) distribution among CESC patients with or without HPV infection

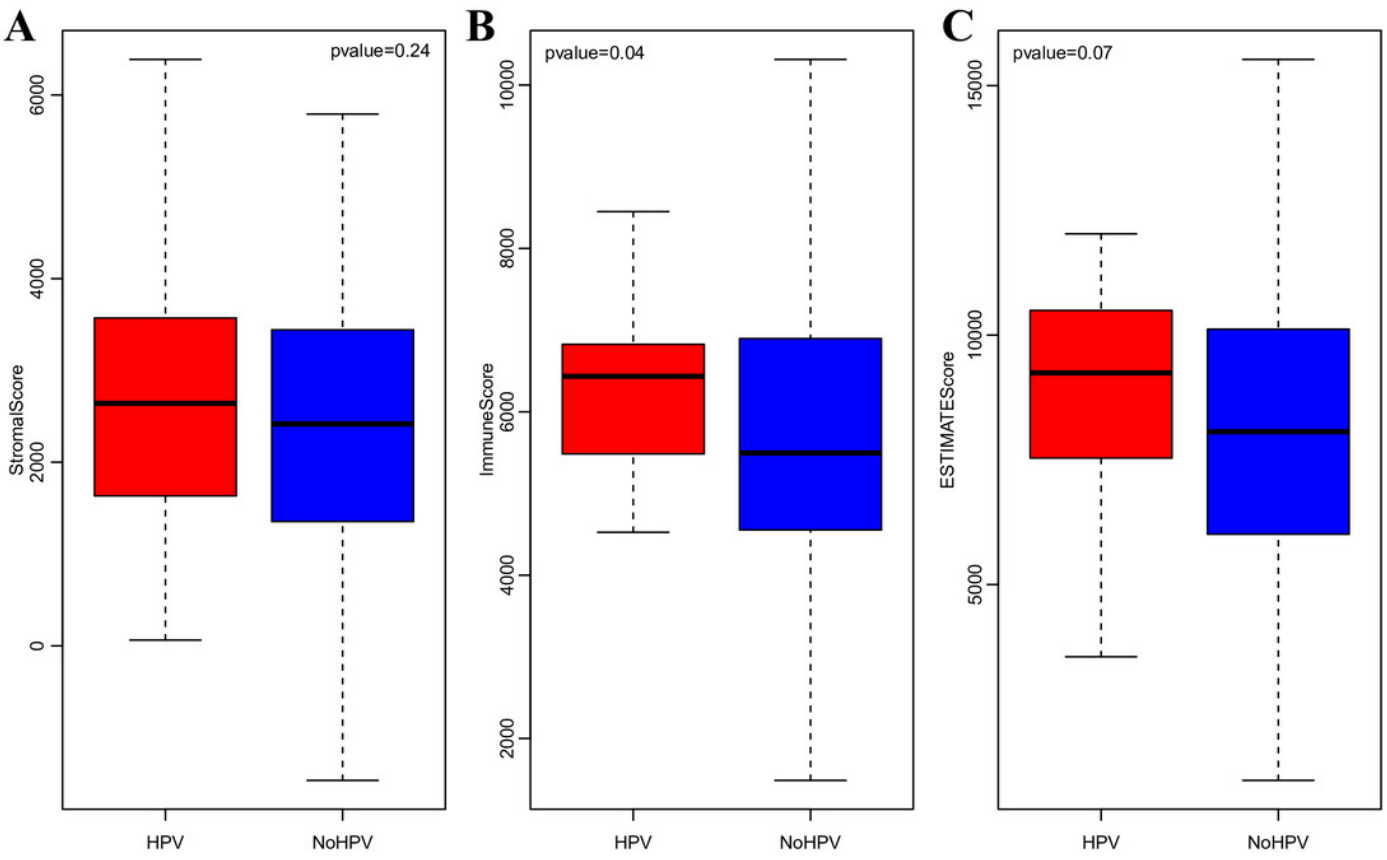

Figure 3

he relationships between levels of StromalScore (A), ImmuneScore (B) or ESTIMATEScore (C) and prognosis for CESC patients

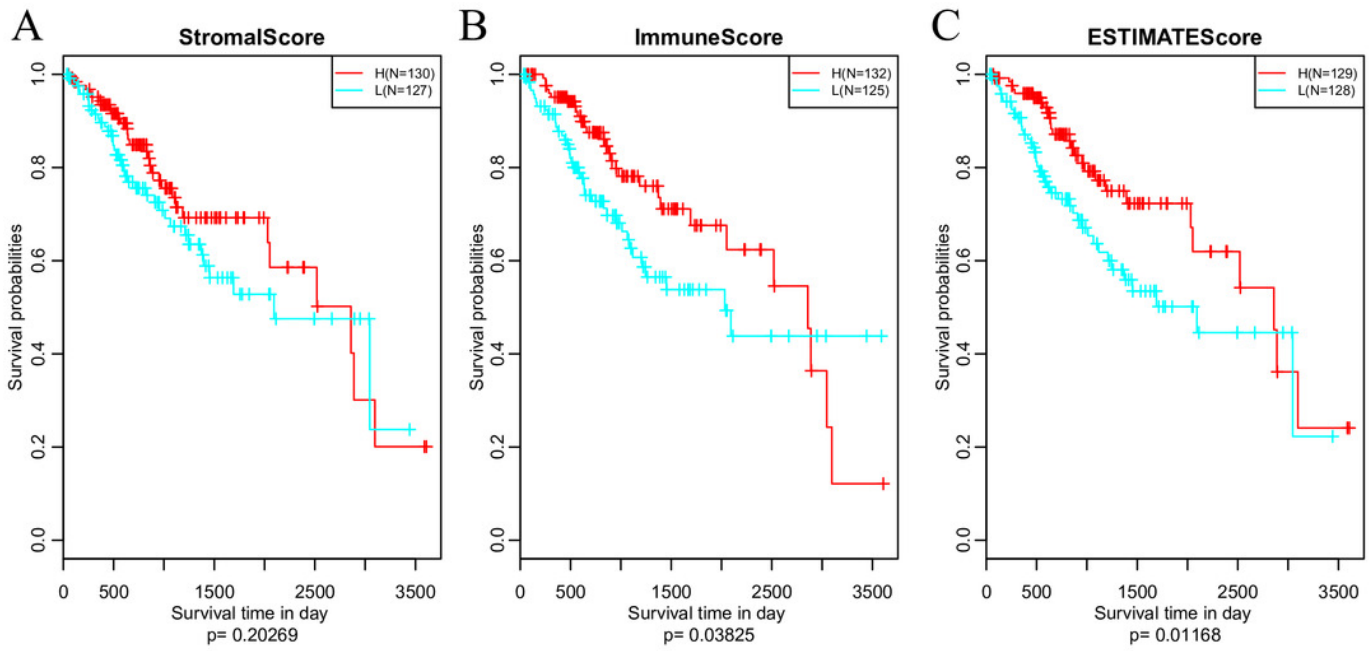

Figure 4

The correlations of immune-related scores based on ESTIMATE algorithm with gene mutations.

The StromalScore (A, D, G, J), ImmuneScore (B, E, H, K) and ESTIMATEScore (C, F, I, L) were calculated respectively in HLA-A (A, B, C), HLA-B (D, E, F), HLA-C (G, H, I) and TP53 (J, K, L) mutation and non-mutation groups. Green represents the mutant group and red represents the wild type.

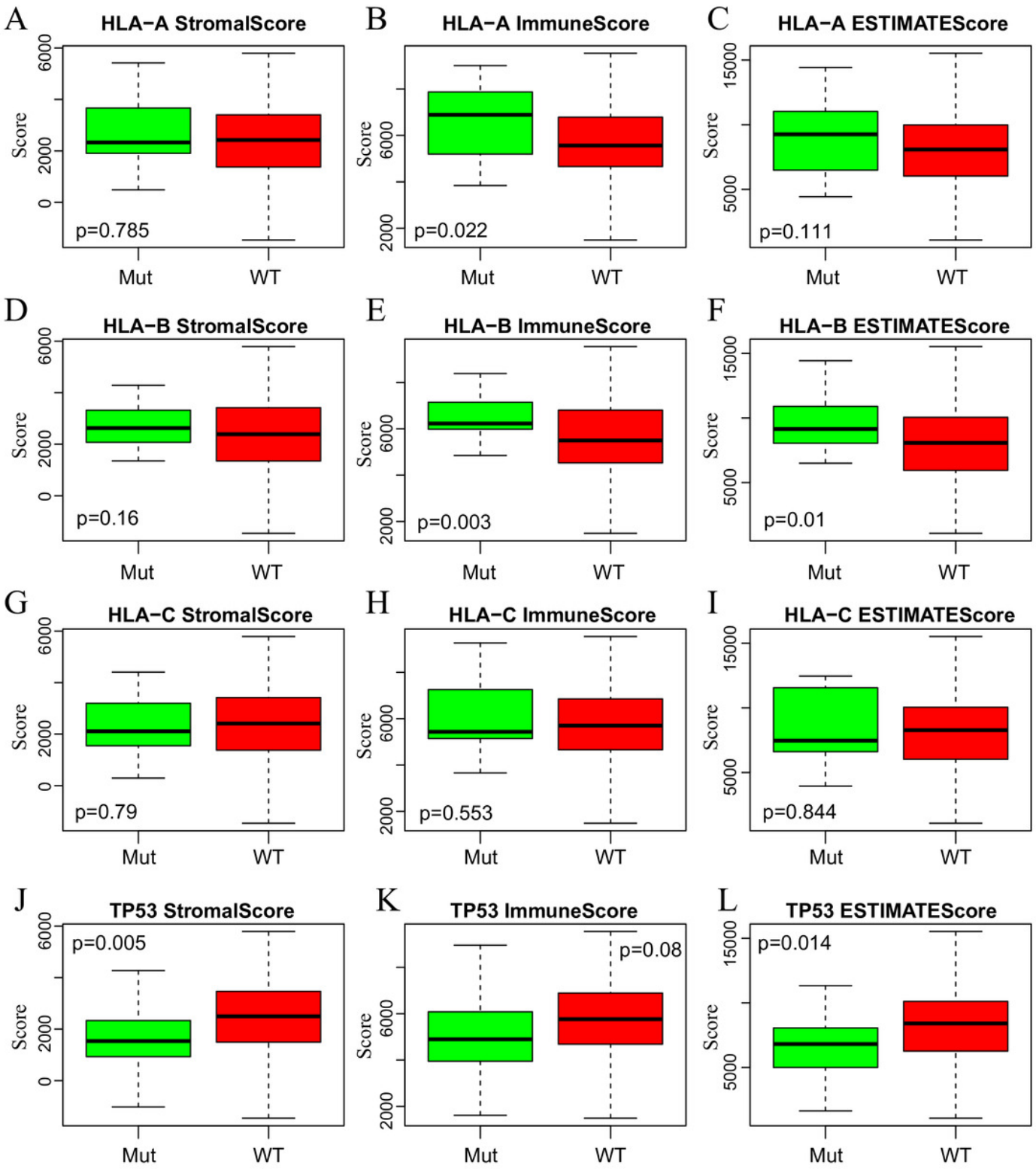
A
HLA−A StromalScore
Score
p=0.785
Mut
WT
B
HLA−A ImmuneScore
p=0.022
C
HLA−A ESTIMATEScore
p=0.111
D
HLA−B StromalScore
p=0.16
E
HLA−B ImmuneScore
p=0.003
F
HLA−B ESTIMATEScore
p=0.01
G
HLA−C StromalScore
p=0.79
H
HLA−C ImmuneScore
p=0.553
I
HLA−C ESTIMATEScore
p=0.844
J
TP53 StromalScore
p=0.005
K
TP53 ImmuneScore
p=0.08
L
TP53 ESTIMATEScore
p=0.014

Figure 5

Immune scores-related gene modules mined through WGCNA

(A) Sample clustering analysis. (B,C) Analysis of network topology for various soft-thresholding powers. (D) Gene dendrogram and module colors. (E) Correlation between each module and three immune-related scores.

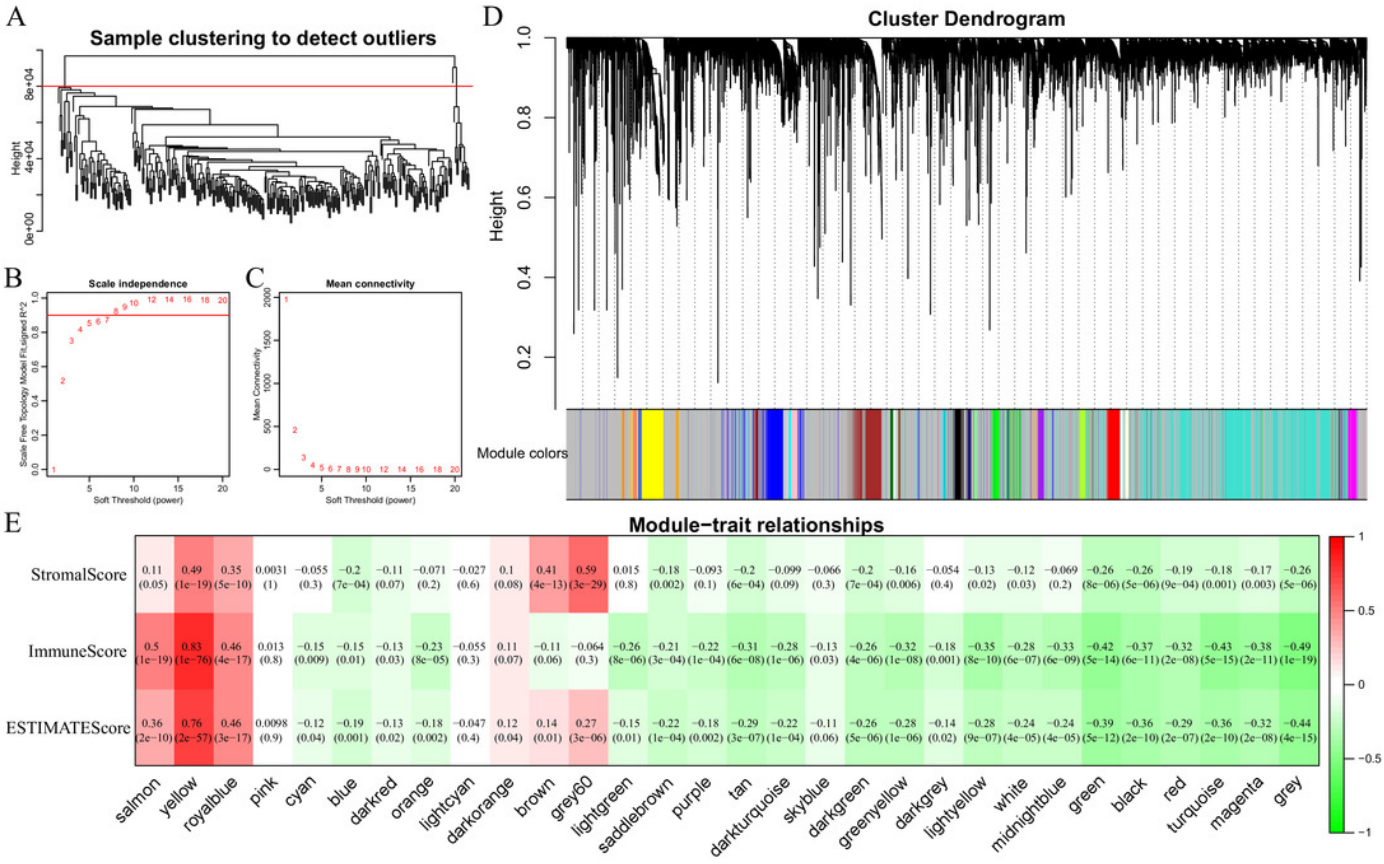

Figure 6

The KEGG pathway and GO enrichment analysis of the genes in yellow module.

(A) Top20 KEGG pathways enriched by the genes in yellow module. (B) Top20 GO BP terms enriched by the genes in yellow module. (C) Top20 GO CC terms enriched by the genes in yellow module. (D) Top20 GO MF terms enriched by the genes in yellow module.

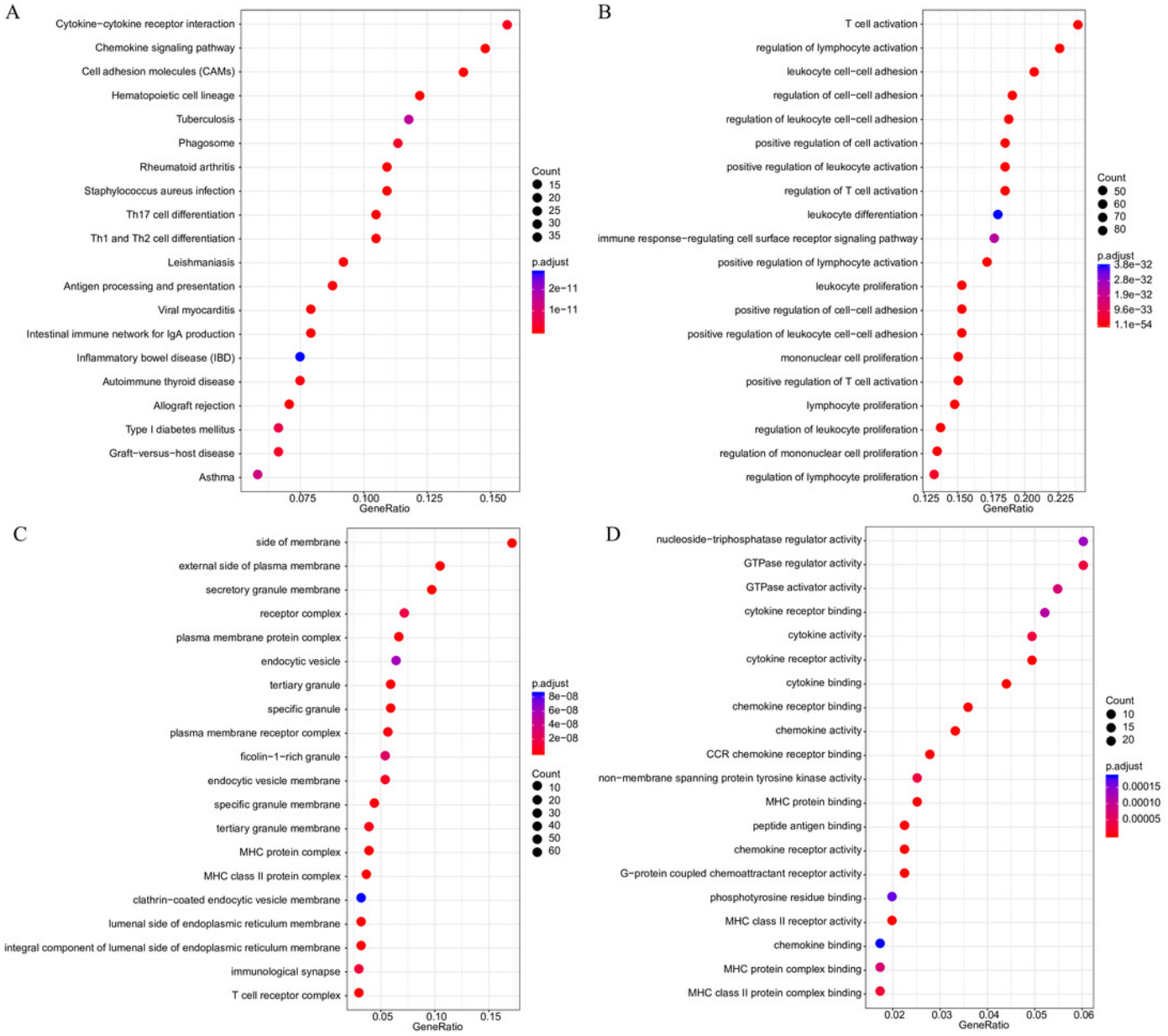

Figure 7

Construction of co-expression network of yellow module-related genes.

(A) Co-expression network of weights between genes in yellow module. (B) The degree distribution of nodes in yellow module. (C) The correlation of genes and module in the network.

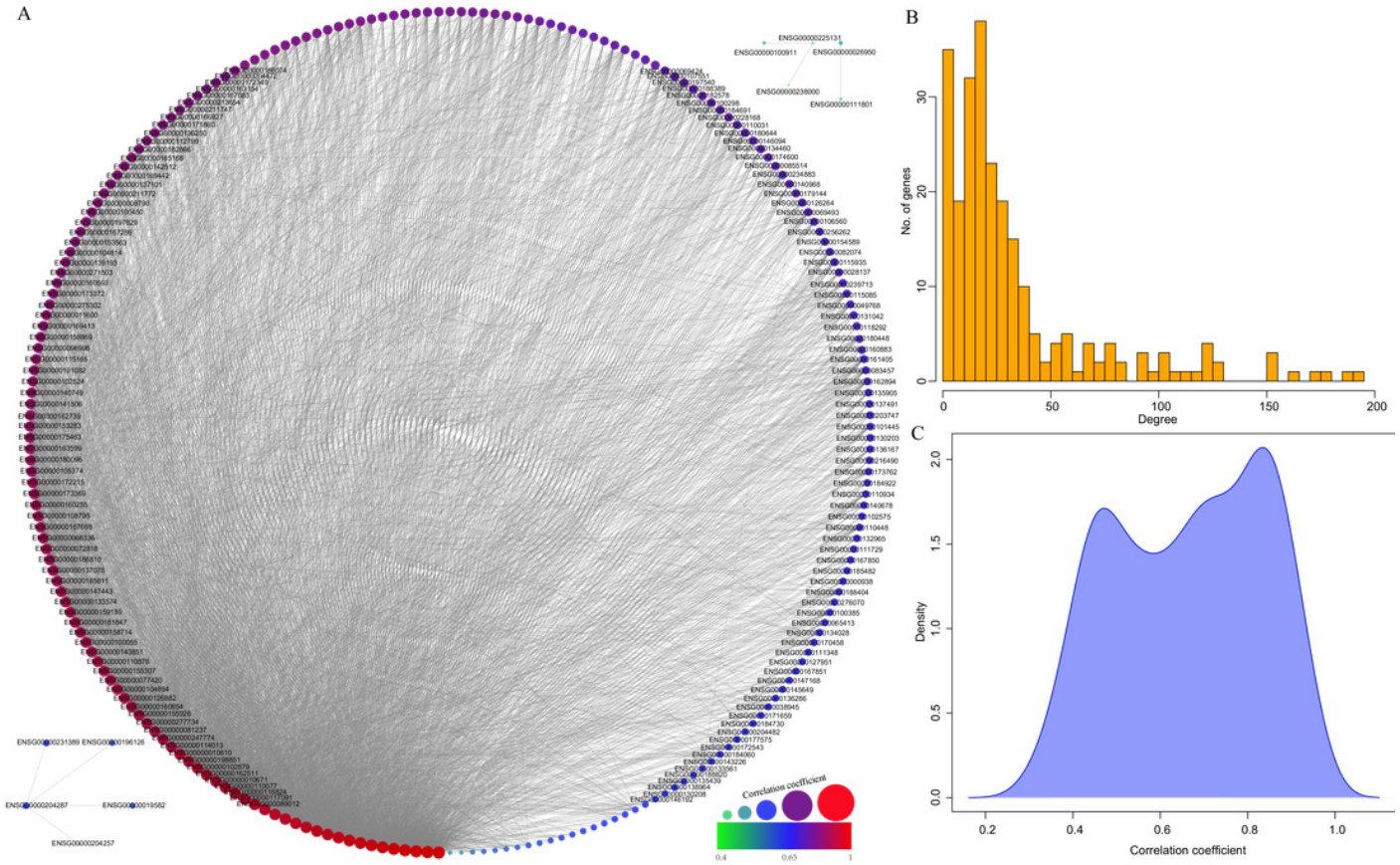

Figure 8

The KEGG pathway and GO enrichment analysis of 18 novel representative immune microenvironment-related genes for CESC patients

(A) Top20 KEGG pathways enriched by 18 novel representative immune microenvironment-related genes. (B) Top20 GO BP terms enriched by 18 novel representative immune microenvironment-related genes. (C) Top20 GO CC terms enriched by 18 novel representative immune microenvironment-related genes. (D) Top20 GO MF terms enriched by 18 novel representative immune microenvironment-related genes.

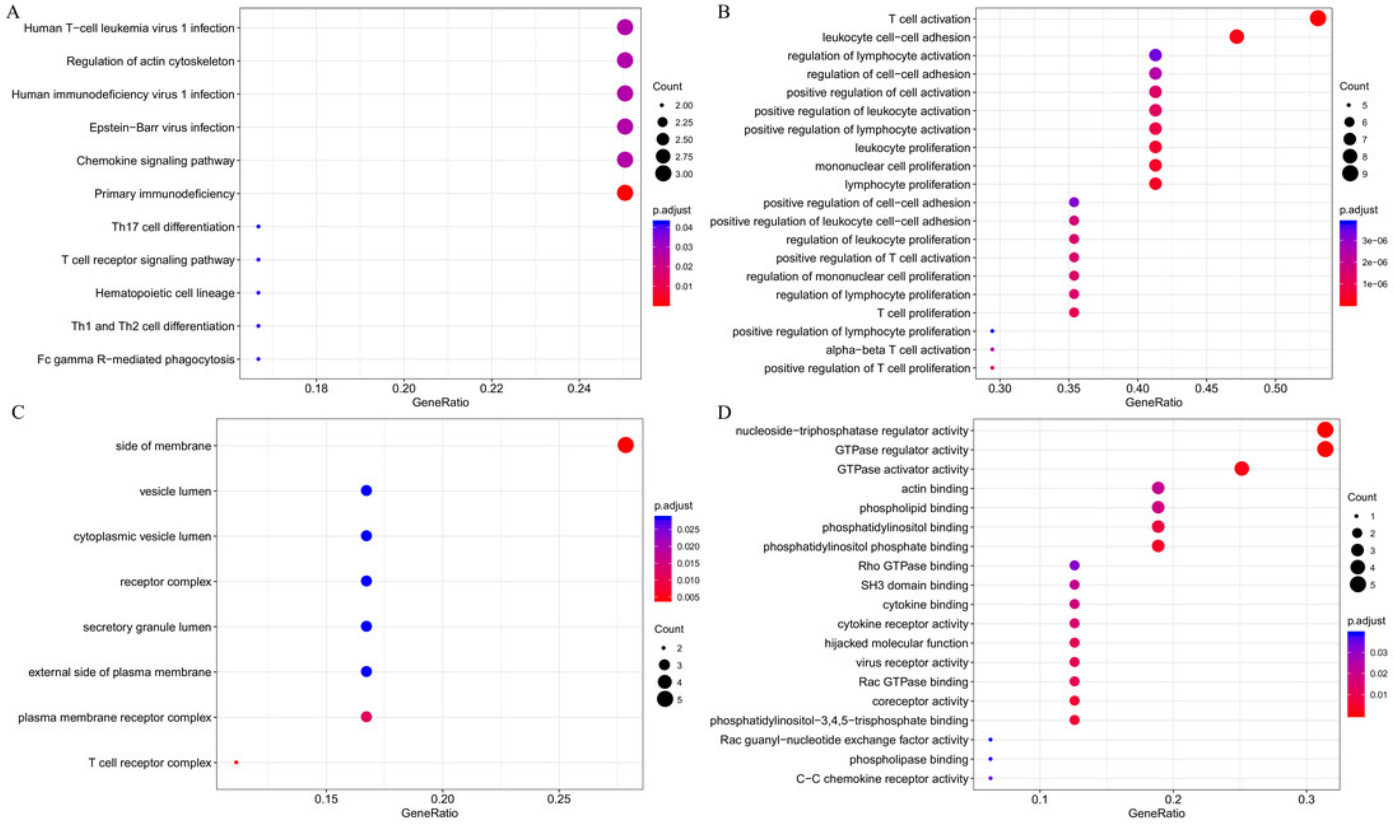

Figure 9

The association between 18 novel representative immune microenvironment-related genes for CESC patients and immune checkpoints

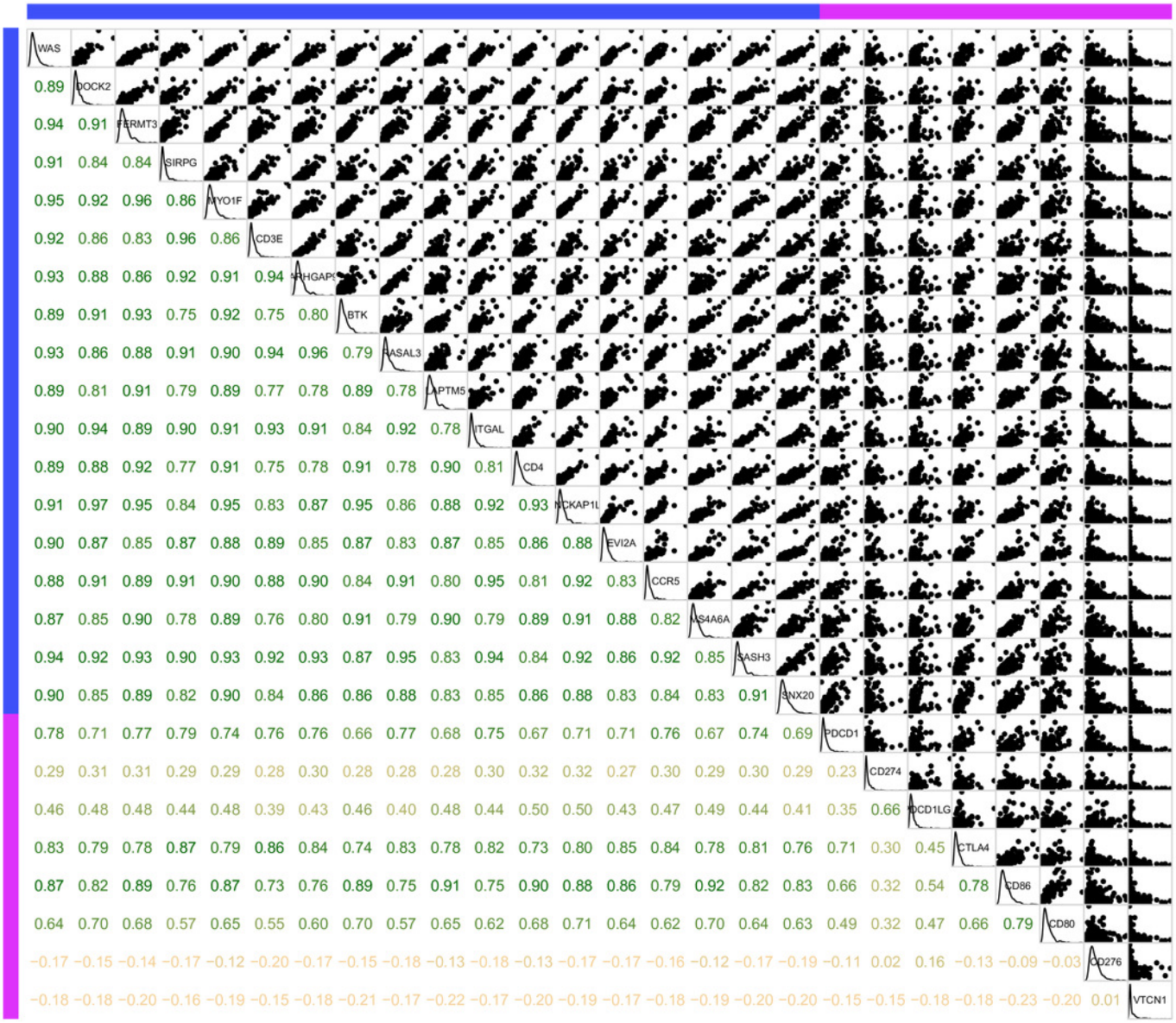

Figure 10

The relationship between 18 novel representative immune microenvironment-related genes and prognosis

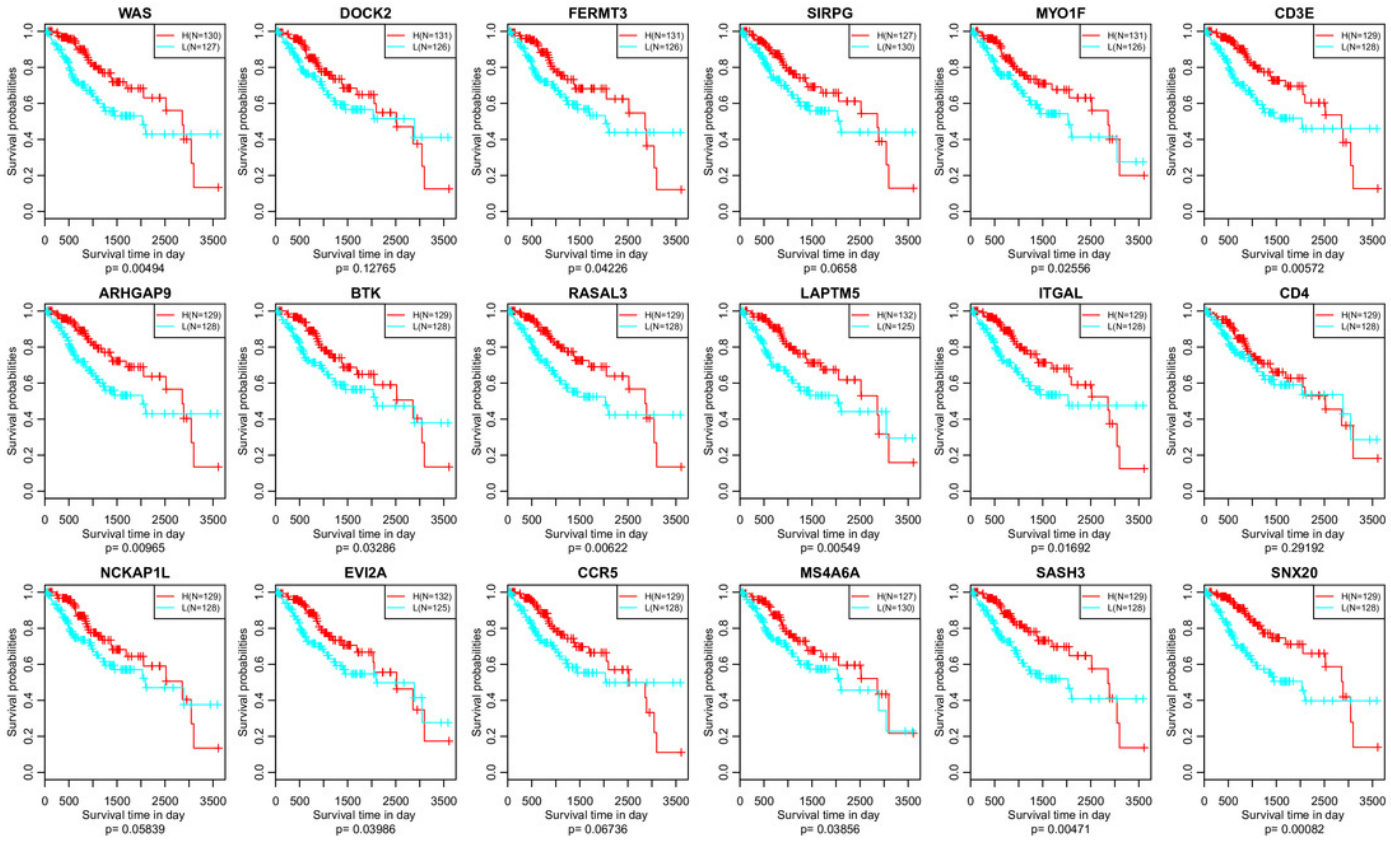

Figure 11

The correlations of 18 immune microenvironment-related genes with ImmuneScore for CESC patients in independent dataset

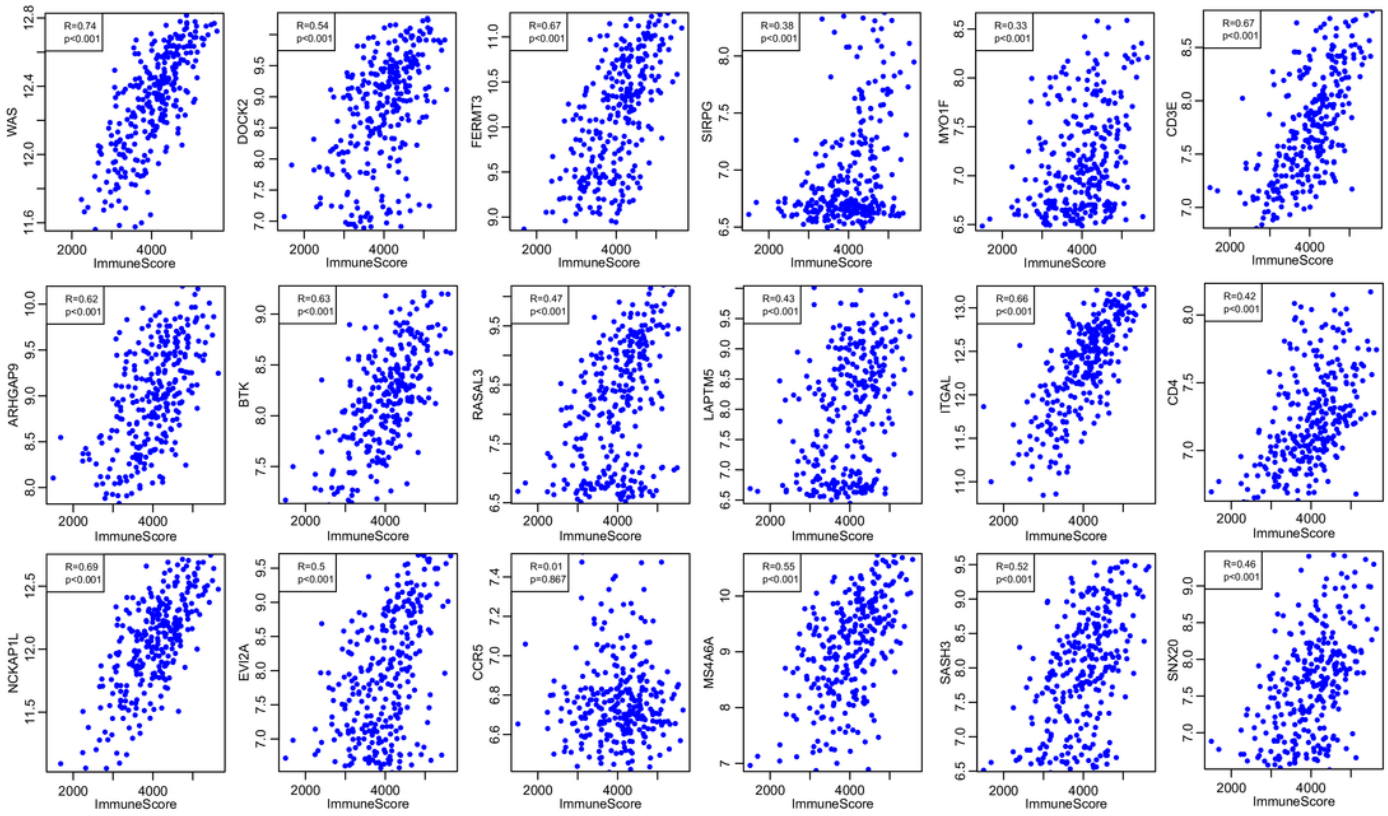

Table 1(on next page)

Number of transcripts in each module

1 Table 1 Number of transcripts in each module

Modules	Genes
Black	232
Blue	676
Brown	469
Cyan	82
Darkgreen	53
Darkgrey	44
Darkorange	38
Darkred	54
Darkturquoise	47
Green	276
Greenyellow	110
Grey	7417
Grey 60	67
Lightcyan	68
Lightgreen	66
Lightyellow	65
Magenta	181
Midnightblue	78
Orange	40
Pink	232
Purple	116
Red	261
Royalblue	64
Saddlebrown	31
Salmon	97
Skyblue	33
Tan	98
Turquoise	2642

White	37
Yellow	422

2

Table 2(on next page)

Genes with a correlation over 0.9 and a degree over 50 in the network

1 Table 2 Genes with a correlation over 0.9 and a degree over 50 in the network

ENSG	Symbol	corr.R	Degree	MeteGene
ENSG00000015285	WAS	0.964019	188	
ENSG00000110324	IL10RA	0.944217	154	LCK
ENSG00000134516	DOCK2	0.932541	113	
ENSG00000149781	FERMT3	0.957826	171	
ENSG00000043462	LCP2	0.941048	102	CLK
ENSG00000185862	EVI2B	0.94047	153	LCK
ENSG00000117091	CD48	0.918649	107	LCK
ENSG00000089012	SIRPG	0.918974	119	
ENSG00000135077	HAVCR2	0.932432	95	Co_inhibition
ENSG00000116824	CD2	0.915954	124	LCK
ENSG00000142347	MYO1F	0.962694	193	
ENSG00000198851	CD3E	0.90917	130	
ENSG00000123329	ARHGAP9	0.925285	126	
ENSG00000010671	BTK	0.913087	85	
ENSG00000105122	RASAL3	0.92036	124	
ENSG00000162511	LAPTM5	0.912211	72	
ENSG00000005844	ITGL	0.92691	125	
ENSG00000010610	CD48	0.908066	57	
ENSG00000123338	NCKAP1L	0.953232	162	
ENSG00000102879	CORO1A	0.909449	94	LCK
ENSG00000126860	EVI2A	0.923238	70	
ENSG00000143119	CD53	0.957049	178	LCK
ENSG00000160791	CCR5	0.926682	104	
ENSG00000110077	MS4A6A	0.915621	57	
ENSG00000122122	SASH3	0.949558	153	
ENSG00000167208	SNX20	0.942276	123	□

2