Machine learning prediction models based on hub genes related to immune phenotypes in muscle-invasive bladder cancer treated with immune checkpoint blockade

Peiheng Li (lipeiheng@pumc.edu.cn)
Peking Union Medical College Hospital https://orcid.org/0000-0002-5642-6778
Zhi-Xin Chen
Peking Union Medical College School of Basic Medicine: Chinese Academy of Medical Sciences and Peking Union Medical College Institute of Basic Medical Sciences
Dong Wang
Peking Union Medical College Hospital
Zhi Zheng
Peking Union Medical College School of Basic Medicine: Chinese Academy of Medical Sciences and Peking Union Medical College Institute of Basic Medical Sciences
Zhi-Gang Ji
Peking Union Medical College Hospital

Research article

Keywords: Muscle-invasive bladder cancer (MIBC), Immune checkpoint blockade (ICB), Immune microenvironment, Metabolism, Multi-omic analysis, Weighted Gene Co-expression Network Analyses (WGCNA), Classification and regression tree (CART)

Posted Date: September 22nd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-63592/v2

Abstract

Background

Bladder cancer is one of the most frequent cancer in the world. Muscle-invasive bladder cancer (MIBC) is a more aggressive subtype with higher morbidity and mortality. ICB therapy has shown potential for treating MIBC, but is limited due to the lack of predictive biomarkers.

Methods

1601 MIBC transcriptomic profiles were obtained from 10 datasets. Unsupervised clustering of immune phenotypes in MIBC were performed based on immune-related signature genes. We analyzed the characteristics including immune microenvironments, metabolic pathways, and survival rates in different phenotypes. Multi-omic analysis and WGCNA were performed to identify hub genes distinguishing phenotypes related to prognosis. A model was established and CART was employed to predict the response of patients treated with ICB.

Results

Of various immune phenotypes, cluster 3C was the most “inflamed” subcluster with the best prognosis, while cluster 1A was associated with “non-inflammation” and worse prognosis. WGCNA analysis identified 5 hub genes related to the survival rate of patients, IFNG, CXCR6, IL2RB, LCK, and PSMB10, which were utilized for prognostic score model and decision tree analysis. The areas under the curve (AUC) of the ROC curves predicting 5-year endpoint generated from the risk-based prediction model were 0.652. The mean accuracy, sensitivity, and specificity of CART for predicting stable disease/progressive disease were 70.1%, 70.0% and 71.7%.

Conclusions

The 5 hub genes and generated models showed the potential for predicting the prognosis for patients receiving ICB therapy. The molecular mechanisms regulating the expression of the hub genes require further studies in the future.

Background

Bladder cancer is the tenth most frequent cancer worldwide in both sexes combined, resulting in an estimated 549000 new cases and 200000 deaths worldwide in 2018[1]. Most bladder cancers are urothelial carcinomas, which can be divided into non-muscle-invasive bladder cancer (NMIBC) and muscle-invasive bladder cancer (MIBC) according to their distinct prognosis and management. Compared with NMIBC, MIBC is more aggressive with higher morbidity and mortality, which leads to a poor 5-year survival of 40–60%[2]. Treatment for patients with non-metastatic MIBC included cystectomy, extended lymph node dissection, and perioperative chemotherapy. However, for patients with metastatic disease, the response rate of first-line treatment, cisplatin-based chemotherapy, is only 46%-49%[2]. Immune checkpoint blockade (ICB) therapy targeting programmed death ligand-1 (PD-L1)/PD-1 has shown its potential for MIBC patients who are not sensitive to chemotherapy. In 2017, FDA approved 2 ICBs, atezolizumab (a PD-L1 inhibitor) and pembrolizumab (a PD-1 inhibitor), for first-line treatment of cisplatin-ineligible patients[3]. However, the response rate of ICB is also limited due to the tumor intrinsic heterogeneity and the lack of predictive biomarkers.

Gene expression profiling has been used to identify molecular heterogeneity in bladder cancer, which could be used to predict responses to different therapies. There is a consensus that bladder cancer can be divided into basal and luminal

subtypes, which show different responses to chemotherapy[4]. A recent study defines a molecular classification of 6 subtypes, stratifying patients for prognosis or response to treatment[5]. However, these studies focus mainly on malignant cells. In addition, stromal cells and immune cells make up a tumor microenvironment, which is essential to predict the efficacy of ICB therapy. The cancer immunogram involving infiltrated cells and immune related signaling has been shown with the potential for personalized immunotherapy in bladder cancer[6]. The different immune phenotypes of MIBC were likely to be closely related to various responses to immunotherapy.

The major objective of our study was to construct prediction models to distinguish immune phenotypes in MIBC and to improve prognosis prediction of ICB therapy. At first, we performed unsupervised clustering of immune phenotypes of MIBC patients/samples based on immune-related signature genes. We then explored the microenvironment including infiltrated immune cells and stromal cells to evaluate the correlated immune signaling and survival rate in each phenotype. We also analyzed the difference of metabolic pathways among various immune phenotypes in MIBC. The concept of immunometablism has become widespread because metabolism regulation has been found to be essential for immune cells activation[7, 8]. Next, Multi-omics data analysis was applied to highlight molecular principles of differentially expressed genes (DEGs) in corresponding phenotypes. Besides, somatic copy-number alterations were shown to be associated with tumor immune phenotype[9]. We also utilized weighted gene coexpression network analysis (WGCNA) to identify hub genes to distinguish different immune phenotypes. At last, in order to establish models for predicting prognostic status, the hub genes underwent lasso regression analysis to eliminate collinearity in a cohort of patients receiving ICB therapy. In addition, classification and regression tree (CART) algorithm was employed to predict the response of patients treated with ICB.

Methods

Sample Collection

We conducted a systematic search of the electronic databases, including PubMed, GEO profile, TCGA and EGA datasets and used 1601 MIBC transcriptomic profiles from 10 datasets. Details of datasets, including their respective normalizations, are given in Supplementary Table 1.

Data reprocessing

Raw data normalization was carried out in different platforms with default setting. (Supplementary Table 1)[5, 10–18].

13,832 common genes in this study were extracted from the normalized data for six different platforms, and the combat function in the sva package was applied to remove the batch effects of these datasets[19].

Unsupervised immune clusters in MIBC

785 immune genes for 28 immune cell types were obtained from Charoentong et al. [20] Angiogenesis marker genes and antigen MHC I and II presenting machinery expression signature involved in antigen presenting machinery (APM) were obtained from Şenbabaoğlu and Rody et al [21, 22]. A signature of Anti-inflammatory and Pro inflammatory genes were obtained from Azizi et al[23]. IFN signature genes were obtained from Rody et al[22]. All immune signature genes are listed in Supplementary Table 2. The unsupervised clustering for 1601 tumor samples were performed with hierarchical clustering, Ward linkage, and Euclidean distance based on immune signature genes.

Infiltration levels for immune cell types and activity levels for angiogenesis, APM, Anti-inflammatory, Pro inflammatory and IFN were quantified using the ssGSEA[24] implementation in R package gsva[25]. Normalized RNA-Seq datasets mentioned above were provided as input without further processing. The ssGSEA score was normalized to unity distribution, for which zero is the minimal and one is the maximal score for each gene panels. The output for each

signature is a near-Gaussian list of decimals that can be used in visualization/statistical analysis without further processing.

ssGSEA analysis ranked the list of genes(Supplementary Table 3) to identify immune pathways[14, 21, 22] and metabolic pathways[23] enriched in genes with highest variability for different clusters[24]. Expression levels for checkpoints were also evaluated. Immune scores and stromal scores were calculated by applying the ESTIMATE algorithm to the downloaded database [26].

Somatic mutation analysis

Somatic mutation data were obtained from the publicly available TCGA database via the GDC data portal (https://portal.gdc.cancer.gov/). From the four subtypes of data files, we selected the “Masked Somatic Mutation” data and processed it based on the VarScan software. We prepared the Mutation Annotation Format (MAF) of somatic variants and implemented the “maftools” [27] R package which provides a multiple of analysis modules to perform the visualization process.

Copy-number variation analysis

The TCGA copy-number variation (CNV) data were downloaded from https://tcga.xenahubs.net/download/TCGA.BLCA.sampleMap/Gistic2_CopyNumber_Gistic2_all_thresholded.by_genes.gz. Circos plots were performed by the R package “Rcircos” [28].

DNA methylation analysis

TCGA-HNSC DNA methylation data (Illumina Human Methylation 450 k) was downloaded from https://tcga.xenahubs.net/download/TCGA.BLCA.sampleMap/HumanMethylation450.gz. Methylation data were normalized with the β-mixture quantile normalization method (BMIQ) [29] was utilized to preprocess β-values. ChAMP was used to perform quality control, standardization, and calculation of methylation sites and regions [30]. By using the ChAMP package (parameters: adjusted P-value < 0.05), differentially methylated probes (DMPs) between cluster 1A and cluster 3B were identified with bumphunter method. [31] And we use 95% quantile as the cut off value for the candidate regions, which contain at leaste 7 CpG probe. The heatmap was created in R (v3.4.1) using the pheatmap package on the quantile normalised methylation (beta) values.

Differentially Expressed Genes Screening

According to unsupervised clustering, samples were divided into three groups: cluster 1, cluster 2, cluster 3. Then, the DESeq2 function in the R software package [32] was used to analyze the genetic differences between cluster 1A and cluster 3C for TCGA and IMvigor210 cohorts. Then, we screened the DEGs to obtain those with the adjusted P-value < 0.05 and |log2(Foldchange)|>1. Statistical analysis was carried out for each dataset, and the intersecting part of DEGs(2705 genes)was identified using the Venn diagram webtool.

Gene Ontology Annotations and Kyoto Encyclopedia of Genes and Genomes Pathway Analyses

Gene Ontology (GO) analysis was performed to show the unique biological significance based on differentially expressed genes. The Kyoto Encyclopedia of Genes and Genomes (KEGG) database was carried out to find out the important pathway. The ClusterProfiler packages [33] in R was applied to analysis and demonstrated GO annotation and KEGG pathway.

Gene Set Enrichment Analysis

Expression dataset from SRP049695 was conducted Pathway enrichment analysis of a ranked gene list using Gene Set Enrichment Analysis (GSEA) [32]. The following operations were carried out according to the protocol . (http://software.broadinstitute.org/gsea/)

Co-expression Network Construction

WGCNA package of R software was applied to uncover the correlation among genes. [34] The soft-thresholding power β was calculated in the construction of each module using the pickSoftThreshold function of the WGCNA. The power of β was set at 10 to ensure a scale-free network. The minimum number of module genes was set at 30. The hierarchical clustering dendrogram summarized the Gene modules with different colors. Heat map and topological overlap matrix (TOM) plot were used to visualize the module structure. The relationships between modules and clinic traits were assessed using the correlations between the module eigengene and the clinic trait, allowing the identification of hub modules that are correlated with external traits and searching for the most significant associations. For hub modules, the quantitative measure of module membership was defined as the correlation of the module eigengene and the gene expression profile; gene significance was defined as the absolute value of the correlation between the gene and the clinic trait. We set the ModuleMembership > 0.8 and the Gene-Significance > 0.2 for candidate hub genes.

Selection and Validation of Hub Genes related to prognosis

We selected candidate hub genes in TCGA dataset and used the Search Tool for the Retrieval of Interacting Genes (STRING; https:// string-db.org/) database to construct PPI network. The confidence score was set at 0.95. [35] Gene network files exported from STRING were input into Cytoscape software. The plug-in cytoHubba of Cytoscape was used for screen out hub genes with the intersecting top 50 genes based on 12 kinds of Global-based methods. Hub Genes related to prognosis were screened to plot the Kaplan–Meier survival curve in ggplot2 of R software. Consistent with TCGA results, we enrolled survival data and expression level of hub genes related to prognosis from an independent cohort IMvigor210.

Survival analysis

We used cohort IMvigor210 to establish the prognostic model. The 5 hub genes underwent lasso regression analysis to eliminate collinearity between genes. After performing 1000 10-fold cross-validations, the lambda value with minimized error was selected as the optimum lambda parameter value. We used multivariate Cox proportional hazards regression analysis to find key genes involved in the model. The disease risk score, which was used as predictors of prognostic status, was defined by parameter β of the multivariate cox proportional hazards regression analysis and the expression of each selected gene. The risk score for each patients was calculated, categorized into high or low-risk. The predictive performance of this model at 5-year endpoint was evaluated with a time-dependent receiver operating characteristic (ROC) curve.

CART analysis

CART analysis was a nonparametric, supervised statistical learning technique combining variable values involving the 5 hub genes which best discriminated between classes, in our case stable disease (SD) and progressive disease (PD). The optimal combination of variables and possible cutoff values for classification was determined by an exhaustive search of all possibilities by the CART algorithm[36]. The Gini criterion was applied to minimize node impurity after splitting. we also performed 10-fold cross validation to assess the classified efficiency of the decision tree.

Results

Molecular immune phenotypes in MIBC patients

Unsupervised clustering of 1601 MIBC patients using immune signature genes was performed. Three main clusters were identified as cluster 1, cluster 2, and cluster 3 (Fig. 1A). ssGSEA scores indicated that Cluster 3 was infiltrated with high levels of innate and adaptive immune cells. Cluster 2 was heterogeneously infiltrated with immune cells while cluster 1 was non-infiltrated. Cluster 3 showed significantly high expression of signatures of CD8 T cells, macrophages, Th1 cells, NK cells, and Tregs, followed by cluster 2 and cluster 1 (Fig. 1B). Immune-related pathways, involving interferon (IFN) signaling, TGF-beta signaling, Antigen presenting mechanism (APM), angiogenesis, epithelial-mesenchymal transition (EMT) were mostly enhanced in cluster 3, followed by cluster 2 and cluster 1. On the other hand, DNA damage response (DDR) and FGFR3 signaling were mostly enhanced in cluster 3, followed by cluster 2 and cluster 1. Cluster 3 was divided into three subclusters 3A, 3B and 3C and cluster 1 was divided into two subclusters 1A and 1B. Cluster 3C was the most “inflamed” subcluster with high levels of infiltrated immune cells and activation of immune-related pathways. On the contrary, cluster 1A was associated with “non-inflammation” due to the low levels of infiltrated immune cells and inhibition of immune cells activation. We used RNA-based ESTIMATE algorithm to confirm that cluster 3C displayed the highest stromal score and immune score, which was the least pure subcluster. The analysis of ssGSEA scores in metabolic pathways showed that hypoxia regulated genes and glucose deprivation related genes were increased in cluster 3, with a potential to activate immune cells through immunometabolic mechanisms. Cluster 3 also had the best overall survival compared to the other two clusters (Fig. 1C). The expression levels of IFNG, TGFβ, immune checkpoints PD1 and PD-L1 were most enhanced in cluster 3, while the expression level of FGFR3 was most enhanced in cluster 1 (Fig. 1D).

Differences in somatic mutations related to the immune phenotype

After applying unsupervised clustering in TCGA cohort, mutation information of each gene in each sample was exhibited in waterfall plot in different subclusters, as cluster 1A, 1B, 2, 3A, 3B, and 3C (Fig. 2A, 2B, 2C, 2D, 2E, 2F). Various colors with annotations at the bottom represented different mutation types. Detailed information on somatic mutations in different subclusters was shown in supplementary Fig. 1. The top 5 mutated genes in subcluster 1A were TP53 (46%), TTN (39%), ARID1A (29%), KDM6A (29%), and SYNE1 (29%), while the most mutated genes in subcluster 3C were TP53 (53%), RB1 (27%), TTN (27%), FLG (24%), and ARID1A (22%).

DEGs between cluster 3C and cluster 1A

After applying unsupervised clustering in TCGA and IMvigor210 cohorts, DEGs between cluster 3C and cluster 1A were analyzed. 3790 significantly up-regulated DEGs and 4794 down-regulated DEGs were identified in TCGA dataset. In IMvigor210 cohort, 2950 up-regulated genes and 1067 down-regulated genes are identified. The intersection includes 2043 significantly up-regulated and 662 down-regulated genes (Fig. 3A). The volcano plots of TCGA cohort and IMvigor 210 cohort were shown (Fig. 3B). The GO analysis of upregulated DEGs in cluster 3C were enriched in immune cell activation involving T cell activation, leukocyte activation, phagocytosis, and leukocyte migration (Fig. 3C). KEGG analysis showed that the downregulated genes in cluster 3C were enriched in various metabolic pathways involving amino acid metabolism and fatty acid degradation (Fig. 3D).

Difference in genomic CNVs and methylation in DEGs between cluster 3C and cluster 1A

We then performed an integrated analysis of Multi-omics data on TCGA cohorts. To evaluate whether copy number variations (CNVs) affect transcription of affected genes, analysis of genomic alteration of DEGs indicated regions of copy-number gains (chromosomes 1,2,3,4,10,12,17,19,21) or loss (chromosomes 2,6,9,10,11,17) as characteristic features of cluster 1A compare with cluster 3C in the TCGA cohort (Fig. 4A). 159 upregulated DEGs in cluster 3C were coded by regions with a higher frequency for copy-number deletions in cluster 1A, while 62 upregulated DEGs in cluster 1A were encoded by regions with more frequent gains in the cluster (Fig. 4B). Global methylation data available for the TCGA cohort were analyzed. In total, differentially methylated probes of 349 genes were identified comparing cluster 1A and

cluster 3C. Probes with significantly higher beta values in cluster 3C were located in the proximal promoter of 25 DEGs with higher expression in cluster 1A, while Probes with significantly higher beta values in cluster 1A were located in the proximal promoter of 16 DEGs with higher expression in cluster 3C (Fig. 4C).

WGCNA analysis

The intersected 2705 DEGs were used to perform subsequent WGCNA analysis. The power of β was set at 10 to ensure a scale-free network (Fig. 5C). Gene modules were calculated. The gray module identified the genes that can't be classified to other modules. 12 gene modules were identified by the hierarchical clustering dendrogram (Fig. 5A). The relationships between modules and clinic traits were assessed. Many modules are correlated with survival time. The correlation coefficients of the red, yellow, blue, and pink modules were greatest, at 0.28 (0.001), 0.28 (0.001), 0.25 (0.005), and 0.23 (0.009), respectively (Fig. 5B). MM and GS value in the upper quartile of genes in the module were considered as key genes of this module (The MM and GS cut-off of these modules were 0.8 and 0.2, respectively). The gene distribution of red, yellow, blue and pink modules were shown (Fig. 5D). A total of 415 hub genes were investigated from 4 modules, which were analyzed using matascape for pathway and process enrichment analysis. GO analysis indicate these genes are closely related to cytokine-mediated signaling, regulation of cell activation, regulation of cytokine production and leukocyte differentiation (Fig. 5E).

Identification and validation of hub genes

Protein-protein interactions (PPI) were applied to identify 8 genes as hub genes (LCK, HLA-E, IRF1, IL2RB, IFNG, CCL13, CXCR6, PSMB10) (Fig. 6A). To validate our finding of hub genes, we examined TCGA and IMvigor210 cohorts. The KM curves for genes with $P < 0.05$ in TCGA cohort were shown in supplementary Fig. 2. IFNG, CXCR6, IL2RB, LCK, and PSMB10 were found to be closely related to prognosis in both cohorts (Fig. 7A, 7B, 7C, 7D, and 7E). We constructed an immune infiltration interaction network based on the 5 hub genes using the STRING dataset. The results were imported into Cytoscape for visualization (Fig. 6B)

In order to establish a model for predicting prognostic status, 5 genes were used to perform a lasso regression, with 10-fold cross-validation with 1000 repeats. λ value with the smallest Partial Likelyhood Deviance was shown in 2 of the 5 genes had coefficients that were not zero. A prognostic score model involving LCK and PSMB10 was established.

Risk Score=(-0.1973 × LCK) + (-0.0091 × PSMB10)

By predicting survival of patients at 5 years, the areas under the curve (AUC) of the ROC curves generated from the risk-based prediction model in the test data were 0.652 (Fig. 6F). The test data were obtained from IMvigor210 cohorts, consisting of patients receiving ICB therapy.

The decision tree illustrated decision rules, among all the 5 hub genes that were entered in the analysis, 4 were selected by the program for the classification tree (Fig. 8). The four parameters were LCK, IFNG, PSMB10, and CXCR6. LCK was the most essential determining factor, which was the first-level split of two initial branches of the classification tree. The mean accuracy, sensitivity, and specificity for predicting SD/PD were 70.1%, 70.0% and 71.7%.

Discussion

Unsupervised clustering of 1601 MIBC patients using immune signature genes has divided our cohorts into three clusters, Cluster 1, 2, and 3. Cluster 3 was termed as most "inflammed" while cluster 1 was identified as "non-inflammation" according to the difference in infiltrated immune cells and immune-related signaling. Cluster 3C showed the highest expression level of CD8 T cells, which are primed and activated to cytotoxic T lymphocytes (CTLs). CTLs are the key immune cells for killing malignant cells. CD8 T cells priming requires IFN-γ secreted by Th1 cells and NK cells and TNF-α

secreted by Macrophages[37]. The expression level of Th1, NK cells, and Macrophages are also the highest in cluster 3. ssGESA score of immune-related signaling showed high level of IFN signaling and TGF-β signaling in cluster 3. IFN upregulates APM, which was the highest level in cluster 3. Upregulation of APM could presents more antigen and increases the number of CD8 + T cells. IFN-γ also plays an important role in the regulation of PD-L1 expression[38]. Tumors with high IFN-γ levels are more likely to respond to anti-PD-1 therapy[39]. IFN-γ was suggested to have a protective response limiting the process of bladder cancer[40]. TGF-β is the effector molecule of macrophages and it is associated with stromal activation, angiogenesis and epithelial-mesenchymal transition (EMT), which are induced in cluster 3, with protumorigenic effects[41]. Besides, cluster 3 displayed lowest expression in genes involved in DNA damage response (DDR), which increased tumor mutation burden, likely inducing immune cells activation and increasing response to ICB therapy[42]. Cluster 1 showed the highest level of FGFR3 signaling. FGFR is associated with a lack of T-cell inflammation[43], likely lead to “non-inflammation” phenotype. Metabolic regulation is essential for immune cells activation. Hypoxia regulated genes are significantly increased in cluster 3, which might lead to increased activation of fatty acid catabolism and CD8 + T cell activation[44]. Glucose deprivation related genes are also induced in cluster 3, increasing PD-1 expression on CD8 + T cells. Metabolically, Glycolysis is reduced while fatty acid metabolism is induced in immune cells[45]. However, direct evidence on metabolism regulate immune cells in tumor microenvironment needs further studies.

The analysis of somatic mutation, genomic CNVs and DNA methylation in our study didn’t show direct relation to the 5 hub genes. The molecular mechanisms regulating the expression of the hub genes require further studies in the future.

In order to distinguish different immune subtypes in MIBC patients, we performed a WGCNA analysis, 5 hub genes, IFNG, CXCR6, IL2RB, LCK, and PSMB10, which correlated with survival time were identified by us. A prognostic score model involving LCK and PSMB10 was established to predict patients undergoing ICB therapy. In addition, the decision tree selected 4 genes including LCK, PSMB10, IFNG, and CXCR6 by the program. In our study, higher expression of CXCR6 was correlated with higher survival rate. However, CXCR6 has been shown to predict poor prognosis in gastric cancer, breast cancer, prostate cancer and bladder cancer[46–49]. On contrary, CXCR6 was also found to be necessary for promoting NKT and CD4 T cells to remove senescent hepatocytes to prevent hepatocarcinogenesis[50], which might indicate a similar mechanism in MIBC. Interleukin-2 (IL-2) was one of the first FDA-approved immunotherapy drugs for melanoma and renal cell cancer. IL-2 receptor (IL-2R) complex is important to control immune response[51, 52]. There are three IL-2R subunits: IL-2Rα, IL-2Rβ, and IL-2γ. The application of IL-2 treatment was limited because it induced proliferation of Tregs through IL-2 Rα, which is preferentially expressed on Tregs[53]. Increased Tregs are correlated with worse prognosis[54, 55]. On contrary, CD8 T cells and NK cells express IL-2Rβ and IL-2γ, which lack IL-2 Rα. The activation of CD8 T cells and NK cells contribute to better prognosis. Strategies targeting on IL-2Rβ have been shown effective tumor control through activating CD8 T cells and NK cells[56, 57]. Lymphocyte-specific protein tyrosine kinase (LCK) was essential for TCR signaling, associated with Th1, Th2, and Th17 cell differentiation[58]. As a result, LCK has become a druggable target to improve the efficacy of chimeric antigen receptors and to potentiate T-cell response in immunotherapy[59]. Another study indicated that LCK was associated with bladder cancer prognosis, which is consistent with our results[60]. Proteasome subunit beta type-10 (PSMB10) is one of the immunoproteasome subunits. Interferon-γ induces the expression of the catalytic immunoproteasomes[61]. PSMB10 plays a pivotal role in antigen generation/processing, which is crucial for cytotoxic T lymphocytes to eliminate cancer cells[62, 63]. Upregulation of immunoproteosomes in breast cancer shown abundance of tumor infiltrating lymphocytes and a good prognosis[64].

Conclusions

The 5 hub genes, IFNG, CXCR6, IL2RB, LCK, and PSMB10, and generated models showed the potential for predicting the prognosis for patients receiving ICB therapy. The molecular mechanisms regulating the expression of the hub genes require further studies in the future.

Abbreviations

MIBC Muscle-invasive bladder cancer

AUC areas under the curve

ICB Immune checkpoint blockade

WGCNA Weighted Gene Co-expression Network Analyses

CART Classification and regression tree

NMIBC non-muscle-invasive bladder cancer

PD-L1 programmed death ligand-1

DEGs differentially expressed genes

APM antigen presenting machinery

CNV copy-number variation

GO Gene Ontology

KEGG The Kyoto Encyclopedia of Genes and Genomes

GSEA Gene Set Enrichment Analysis

ROC receiver operating characteristic

SD stable disease

PD progressive disease

IFN interferon

EMT epithelial-mesenchymal transition

DDR DNA damage response

PPI Protein-protein interactions

CTL cytotoxic T lymphocytes

IL-2 Interleukin-2

IL-2R IL-2 receptor

LCK Lymphocyte-specific protein tyrosine kinase

PSMB10 Proteasome subunit beta type-10

Declarations

Ethic approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

None.

Author's contributions

Pei-Heng Li and Zhi-xin Chen designed the study and collected the data. Zhi-Xin Chen analyzed the data. Pei-Heng Li interpreted the results. Dong Wang, Zhi Zheng, and Zhi-Gang Ji corrected the results and interpretation of data. All authors read and approved the final manuscript.

Acknowledgements

None.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. Cancer J Clin. 2018;68(6):394–424.
2. Alfred Witjes J, Lebret T, Comperat EM, Cowan NC, De Santis M, Bruins HM, Hernandez V, Espinos EL, Dunn J, Rouanne M, et al. Updated 2016 EAU Guidelines on Muscle-invasive and Metastatic Bladder Cancer. Eur Urol. 2017;71(3):462–75.
3. Flaig TW, Spiess PE, Agarwal N, Bangs R, Boorjian SA, Buyyounouski MK, Downs TM, Efstathiou JA, Friedlander T, Greenberg RE, et al: **NCCN Guidelines Insights**: **Bladder Cancer**, **Version 5**.2018. *J Natl Compr Canc Netw* 2018, **16**(9):1041–1053.
4. Dadhania V, Zhang M, Zhang L, Bondaruk J, Majewski T, Siefker-Radtke A, Guo CC, Dinney C, Cogdell DE, Zhang S, et al. Meta-Analysis of the Luminal and Basal Subtypes of Bladder Cancer and the Identification of Signature Immunohistochemical Markers for Clinical Use. EBioMedicine. 2016;12:105–17.
5. Kamoun A, de Reynies A, Allory Y, Sjodahl G, Robertson AG, Seiler R, Hoadley KA, Groeneveld CS, Al-Ahmadie H, Choi W, et al. A Consensus Molecular Classification of Muscle-invasive Bladder Cancer. Eur Urol. 2020;77(4):420–33.

6. van Dijk N, Funt SA, Blank CU, Powles T, Rosenberg JE, van der Heijden MS. The Cancer Immunogram as a Framework for Personalized Immunotherapy in Urothelial Cancer. Eur Urol. 2019;75(3):435–44.
7. Jung J, Zeng H, Horng T. Metabolism as a guiding force for immunity. Nat Cell Biol. 2019;21(1):85–93.
8. O'Neill LA, Kishton RJ, Rathmell J. A guide to immunometabolism for immunologists. Nature reviews Immunology. 2016;16(9):553–65.
9. Davoli T, Uno H, Wooten EC, Elledge SJ. **Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy**. *Science (New York, NY)* 2017, 355(6322).
10. Lindgren D, Frigyesi A, Gudjonsson S, Sjodahl G, Hallden C, Chebil G, Veerla S, Ryden T, Mansson W, Liedberg F, et al. Combined gene expression and genomic profiling define two intrinsic molecular subtypes of urothelial carcinoma and gene signatures for molecular grading and outcome. Cancer Res. 2010;70(9):3463–72.
11. Riester M, Taylor JM, Feifer A, Koppie T, Rosenberg JE, Downey RJ, Bochner BH, Michor F. Combination of a novel gene expression signature with a clinical nomogram improves the prediction of survival in high-risk bladder cancer. Clin Cancer Res. 2012;18(5):1323–33.
12. Choi W, Porten S, Kim S, Willis D, Plimack ER, Hoffman-Censits J, Roth B, Cheng T, Tran M, Lee IL, et al. Identification of distinct basal and luminal subtypes of muscle-invasive bladder cancer with different sensitivities to frontline chemotherapy. Cancer Cell. 2014;25(2):152–65.
13. Seiler R, Ashab HAD, Erho N, van Rhijn BWG, Winters B, Douglas J, Van Kessel KE, Fransen van de Putte EE, Sommerlad M, Wang NQ, et al. Impact of Molecular Subtypes in Muscle-invasive Bladder Cancer on Predicting Response and Survival after Neoadjuvant Chemotherapy. Eur Urol. 2017;72(4):544–54.
14. Mariathasan S, Turley SJ, Nickles D, Castiglioni A, Yuen K, Wang Y, Kadel EE III, Koeppen H, Astarita JL, Cubas R, et al. TGFbeta attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. Nature. 2018;554(7693):544–8.
15. de Jong JJ, Liu Y, Robertson AG, Seiler R, Groeneveld CS, van der Heijden MS, Wright JL, Douglas J, Dall'Era M, Crabb SJ, et al. Long non-coding RNAs identify a subset of luminal muscle-invasive bladder cancer patients with favorable prognosis. Genome Med. 2019;11(1):60.
16. Efstathiou JA, Mouw KW, Gibb EA, Liu Y, Wu CL, Drumm MR, da Costa JB, du Plessis M, Wang NQ, Davicioni E, et al. Impact of Immune and Stromal Infiltration on Outcomes Following Bladder-Sparing Trimodality Therapy for Muscle-Invasive Bladder Cancer. Eur Urol. 2019;76(1):59–68.
17. Seiler R, Gibb EA, Wang NQ, Oo HZ, Lam HM, van Kessel KE, Voskuilen CS, Winters B, Erho N, Takhar MM, et al. Divergent Biological Response to Neoadjuvant Chemotherapy in Muscle-invasive Bladder Cancer. Clin Cancer Res. 2019;25(16):5082–93.
18. Song BN, Kim SK, Mun JY, Choi YD, Leem SH, Chu IS. Identification of an immunotherapy-responsive molecular subtype of bladder cancer. EBioMedicine. 2019;50:238–45.
19. Xia L, Su X, Shen J, Meng Q, Yan J, Zhang C, Chen Y, Wang H, Xu M. ANLN functions as a key candidate gene in cervical cancer as determined by integrated bioinformatic analysis. Cancer Manag Res. 2018;10:663–70.
20. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, Hackl H, Trajanoski Z. Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. Cell Rep. 2017;18(1):248–62.
21. Şenbabaoğlu Y, Gejman RS, Winer AG, Liu M, Van Allen EM, de Velasco G, Miao D, Ostrovnaya I, Drill E, Luna A, et al. Tumor immune microenvironment characterization in clear cell renal cell carcinoma identifies prognostic and immunotherapeutically relevant messenger RNA signatures. Genome Biol. 2016;17(1):231.
22. Rody A, Karn T, Liedtke C, Pusztai L, Ruckhaeberle E, Hanker L, Gaetje R, Solbach C, Ahr A, Metzler D, et al. A clinically relevant gene signature in triple negative and basal-like breast cancer. Breast Cancer Res. 2011;13(5):R97.

23. Azizi E, Carr AJ, Plitas G, Cornish AE, Konopacki C, Prabhakaran S, Nainys J, Wu K, Kiseliovas V, Setty M, et al. Single-Cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment. Cell. 2018;174(5):1293–308.e1236.
24. Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, Schinzel AC, Sandy P, Meylan E, Scholl C, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature. 2009;462(7269):108–12.
25. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics. 2013;14:7.
26. Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, Trevino V, Shen H, Laird PW, Levine DA, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun. 2013;4:2612.
27. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res. 2018;28(11):1747–56.
28. Ayers M, Lunceford J, Nebozhyn M, Murphy E, Loboda A, Kaufman DR, Albright A, Cheng JD, Kang SP, Shankaran V, et al. IFN-gamma-related mRNA profile predicts clinical response to PD-1 blockade. J Clin Invest. 2017;127(8):2930–40.
29. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, Beck S. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. Bioinformatics. 2013;29(2):189–96.
30. Tian Y, Morris TJ, Webster AP, Yang Z, Beck S, Feber A, Teschendorff AE. ChAMP: updated methylation analysis pipeline for Illumina BeadChips. Bioinformatics. 2017;33(24):3982–4.
31. Williams LA, Mills L, Hooten AJ, Langer E, Roesler M, Frazier AL, Krailo M, Nelson HH, Bestrashniy J, Amatruda JF, et al. Differences in DNA methylation profiles by histologic subtype of paediatric germ cell tumours: a report from the Children's Oncology Group. Br J Cancer. 2018;119(7):864–72.
32. Reimand J, Isserlin R, Voisin V, Kucera M, Tannus-Lopes C, Rostamianfar A, Wadi L, Meyer M, Wong J, Xu C, et al. Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. Nat Protoc. 2019;14(2):482–517.
33. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16(5):284–7.
34. Liu Z, Li M, Fang X, Shen L, Yao W, Fang Z, Chen J, Feng X, Hu, Zeng Z, et al. Identification of surrogate prognostic biomarkers for allergic asthma in nasal epithelial brushing samples by WGCNA. J Cell Biochem. 2019;120(4):5137–50.
35. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019;47(D1):D607–13.
36. Lemon SC, Roy J, Clark MA, Friedmann PD, Rakowski W. Classification and regression tree analysis in public health: methodological review and comparison with logistic regression. Annals of behavioral medicine: a publication of the Society of Behavioral Medicine. 2003;26(3):172–81.
37. Farhood B, Najafi M, Mortezaee K. CD8(+) cytotoxic T lymphocytes in cancer immunotherapy: A review. J Cell Physiol. 2019;234(6):8509–21.
38. Garcia-Diaz A, Shin DS, Moreno BH, Saco J, Escuin-Ordinas H, Rodriguez GA, Zaretsky JM, Sun L, Hugo W, Wang X, et al. Interferon Receptor Signaling Pathways Regulating PD-L1 and PD-L2 Expression. Cell reports. 2017;19(6):1189–201.
39. Maleki Vareki S. High and low mutational burden tumors versus immunologically hot and cold tumors and response to immune checkpoint inhibitors. J immunother Cancer. 2018;6(1):157.

40. Ahlen Bergman E, Hartana CA, Johansson M, Linton LB, Berglund S, Hyllienmark M, Lundgren C, Holmstrom B, Palmqvist K, Hansson J, et al. Increased CD4(+) T cell lineage commitment determined by CpG methylation correlates with better prognosis in urinary bladder cancer patients. Clin Epigenetics. 2018;10(1):102.
41. Akhurst RJ, Hata A. Targeting the TGFβ signalling pathway in disease. Nature reviews Drug discovery. 2012;11(10):790–811.
42. Teo MY, Seier K, Ostrovnaya I, Regazzi AM, Kania BE, Moran MM, Cipolla CK, Bluth MJ, Chaim J, Al-Ahmadie H, et al. Alterations in DNA Damage Response and Repair Genes as Potential Marker of Clinical Benefit From PD-1/PD-L1 Blockade in Advanced Urothelial Cancers. Journal of clinical oncology: official journal of the American Society of Clinical Oncology. 2018;36(17):1685–94.
43. Sweis RF, Spranger S, Bao R, Paner GP, Stadler WM, Steinberg G, Gajewski TF. Molecular Drivers of the Non-T-cell-Inflamed Tumor Microenvironment in Urothelial Bladder Cancer. Cancer immunology research. 2016;4(7):563–8.
44. Maimela NR, Liu S, Zhang Y. Fates of CD8 + T cells in Tumor Microenvironment. Comput Struct Biotechnol J. 2019;17:1–13.
45. Chang CH, Curtis JD, Maggi LB Jr, Faubert B, Villarino AV, O'Sullivan D, Huang SC, van der Windt GJ, Blagih J, Qiu J, et al. Posttranscriptional control of T cell effector function by aerobic glycolysis. Cell. 2013;153(6):1239–51.
46. Lee JT, Lee SD, Lee JZ, Chung MK, Ha HK. Expression analysis and clinical significance of CXCL16/CXCR6 in patients with bladder cancer. Oncol Lett. 2013;5(1):229–35.
47. Xiao G, Wang X, Wang J, Zu L, Cheng G, Hao M, Sun X, Xue Y, Lu J, Wang J: **CXCL16/CXCR6 chemokine signaling mediates breast cancer progression by pERK1/2-dependent mechanisms**. *Oncotarget* 2015, **6**(16):14165–14178.
48. Singh R, Kapur N, Mir H, Singh N, Lillard JW Jr, Singh S. CXCR6-CXCL16 axis promotes prostate cancer by mediating cytoskeleton rearrangement via Ezrin activation and αvβ3 integrin clustering. Oncotarget. 2016;7(6):7343–53.
49. Jin JJ, Dai FX, Long ZW, Cai H, Liu XW, Zhou Y, Hong Q, Dong QZ, Wang YN, Huang H. CXCR6 predicts poor prognosis in gastric cancer and promotes tumor metastasis through epithelial-mesenchymal transition. Oncol Rep. 2017;37(6):3279–86.
50. Mossanen JC, Kohlhepp M, Wehr A, Krenkel O, Liepelt A, Roeth AA, Mockel D, Heymann F, Lammers T, Gassler N, et al. CXCR6 Inhibits Hepatocarcinogenesis by Promoting Natural Killer T- and CD4(+) T-Cell-Dependent Control of Senescence. Gastroenterology. 2019;156(6):1877–89 e1874.
51. Rosenberg SA, Yang JC, White DE, Steinberg SM. Durability of complete responses in patients with metastatic cancer treated with high-dose interleukin-2: identification of the antigens mediating response. Annals of surgery. 1998;228(3):307–19.
52. Rosenberg SA. IL-2: the first effective immunotherapy for human cancer. Journal of immunology (Baltimore Md: 1950). 2014;192(12):5451–8.
53. Ahmadzadeh M, Rosenberg SA. IL-2 administration increases CD4 + CD25(hi) Foxp3 + regulatory T cells in cancer patients. Blood. 2006;107(6):2409–14.
54. Barth SD, Schulze JJ, Kühn T, Raschke E, Hüsing A, Johnson T, Kaaks R, Olek S. **Treg-Mediated Immune Tolerance and the Risk of Solid Cancers: Findings From EPIC-Heidelberg**. *Journal of the National Cancer Institute* 2015, 107(11).
55. Imai H, Saio M, Nonaka K, Suwa T, Umemura N, Ouyang GF, Nakagawa J, Tomita H, Osada S, Sugiyama Y, et al. Depletion of CD4 + CD25 + regulatory T cells enhances interleukin-2-induced antitumor immunity in a mouse model of colon adenocarcinoma. Cancer Sci. 2007;98(3):416–23.
56. Sun Z, Ren Z, Yang K, Liu Z, Cao S, Deng S, Xu L, Liang Y, Guo J, Bian Y, et al. A next-generation tumor-targeting IL-2 preferentially promotes tumor-infiltrating CD8(+) T-cell response and effective tumor control. Nat Commun. 2019;10(1):3874.

57. Jounaidi Y, Cotten JF, Miller KW, Forman SA. Tethering IL2 to Its Receptor IL2Rbeta Enhances Antitumor Activity and Expansion of Natural Killer NK92 Cells. Cancer Res. 2017;77(21):5938–51.
58. Nika K, Soldani C, Salek M, Paster W, Gray A, Etzensperger R, Fugger L, Polzella P, Cerundolo V, Dushek O, et al. Constitutively active Lck kinase in T cells drives antigen receptor signal transduction. Immunity. 2010;32(6):766–77.
59. Bommhardt U, Schraven B, Simeoni L. **Beyond TCR Signaling: Emerging Functions of Lck in Cancer and Immunotherapy**. *Int J Mol Sci* 2019, 20(14).
60. Zhang Y, Ou DH, Zhuang DW, Zheng ZF, Lin ME. In silico analysis of the immune microenvironment in bladder cancer. BMC Cancer. 2020;20(1):265.
61. Goldberg AL, Gaczynska M, Grant E, Michalek M, Rock KL. Functions of the proteasome in antigen presentation. Cold Spring Harb Symp Quant Biol. 1995;60:479–90.
62. de Verteuil D, Muratore-Schroeder TL, Granados DP, Fortier MH, Hardy MP, Bramoullé A, Caron E, Vincent K, Mader S, Lemieux S, et al. Deletion of immunoproteasome subunits imprints on the transcriptome and has a broad impact on peptides presented by major histocompatibility complex I molecules. Molecular cellular proteomics: MCP. 2010;9(9):2034–47.
63. Aptsiauri N, Cabrera T, Garcia-Lora A, Lopez-Nevot MA, Ruiz-Cabello F, Garrido F. MHC class I antigens and immune surveillance in transformed cells. Int Rev Cytol. 2007;256:139–89.
64. Rouette A, Trofimov A, Haberl D, Boucher G, Lavallee VP, D'Angelo G, Hebert J, Sauvageau G, Lemieux S, Perreault C. Expression of immunoproteasome genes is regulated by cell-intrinsic and -extrinsic factors in human cancers. Sci Rep. 2016;6:34019.

Figures

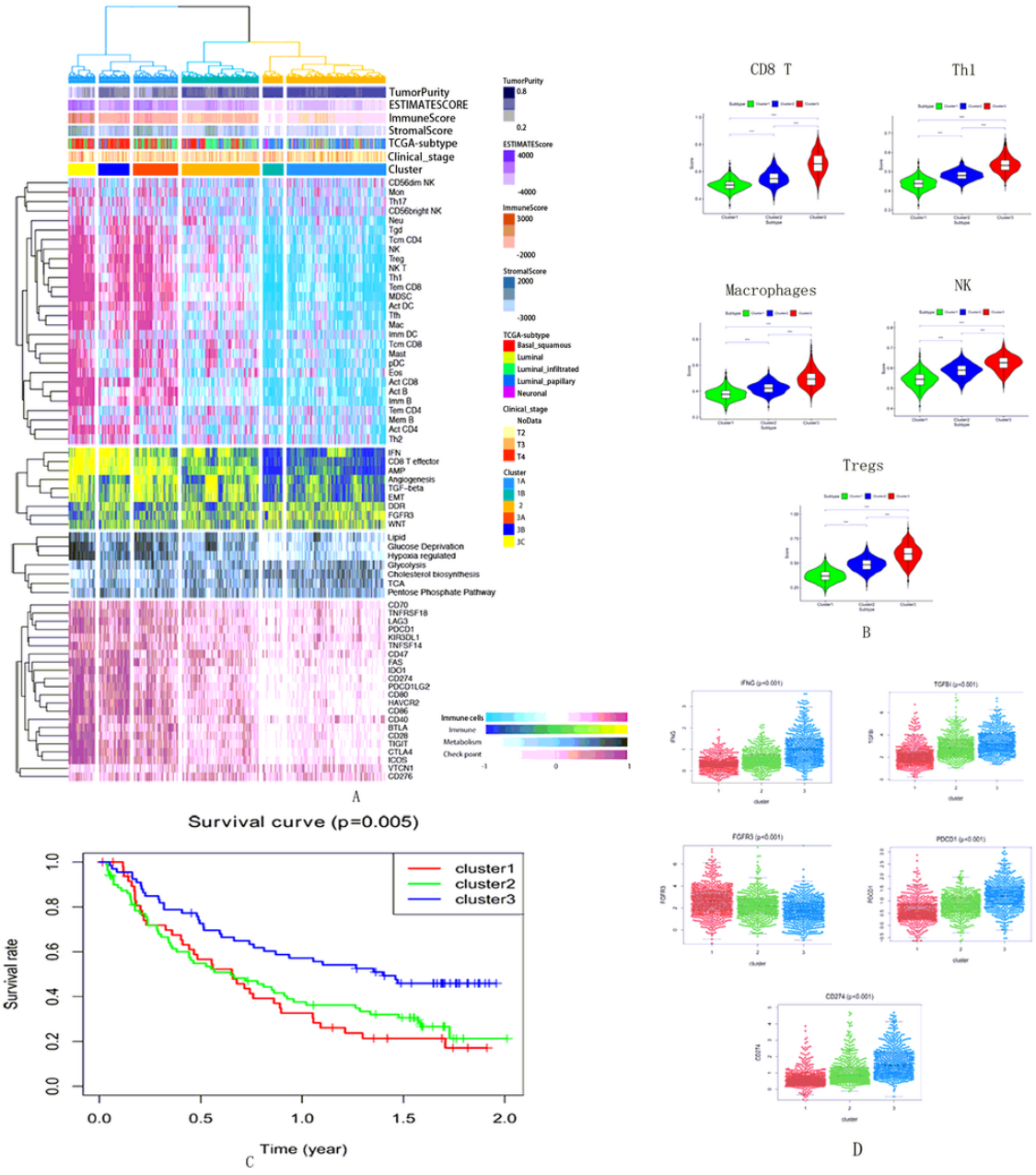

Figure 1

Molecular immune subtypes in MIBC patients. A Unsupervised clustering of 1601 MIBC patients using ssGSEA scores from immune cell types and immune-related pathways. Hierarchical clustering was performed with Euclidean distance and Ward Linkage. Three main clusters were identified as cluster 1, cluster 2 and cluster 3. Of the three clusters, subcluster 3C was the most “inflamed” while subcluster 1A was “non-inflammation”. ssGSEA analysis of metabolic pathways in different clusters was also shown. B The level of infiltrated immune cells, CD8 T cells, Th1 cells, macrophages, NK cells, and Tregs in the three clusters. Cluster 3 was infiltrated with highest level of innate and adaptive immune cells, followed by cluster 2 and cluster 1. C Survival curves of the three clusters. Cluster 3 was correlated with the best overall survival, followed by cluster 2 and cluster 1 (p=0.005). D The expression levels of IFNG, TGFβ, FGFR3, immune checkpoints PD1 and PD-L1 in the three clusters.

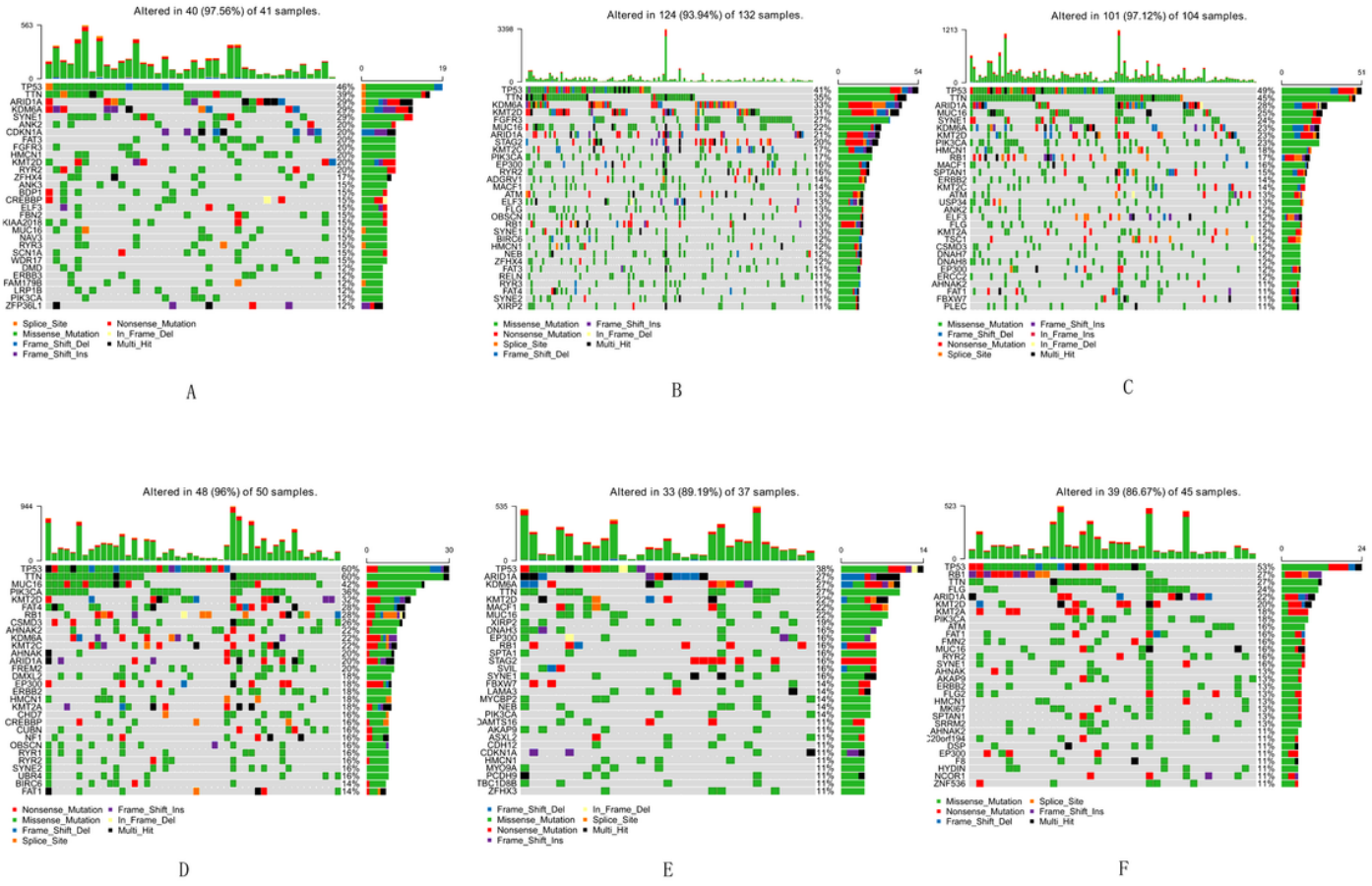

Figure 2

Landscape of somatic mutation profiles in TCGA cohorts. Mutation information of each gene in each sample of six subclusters was shown in the waterfall plot. A, B, C, D, E, F represented cluster 1A, 1B, 2, 3A, 3B, 3C, respectively. Various colors with annotations at the bottom represented the different mutation types. The barplot above the legend exhibited the mutation burden.

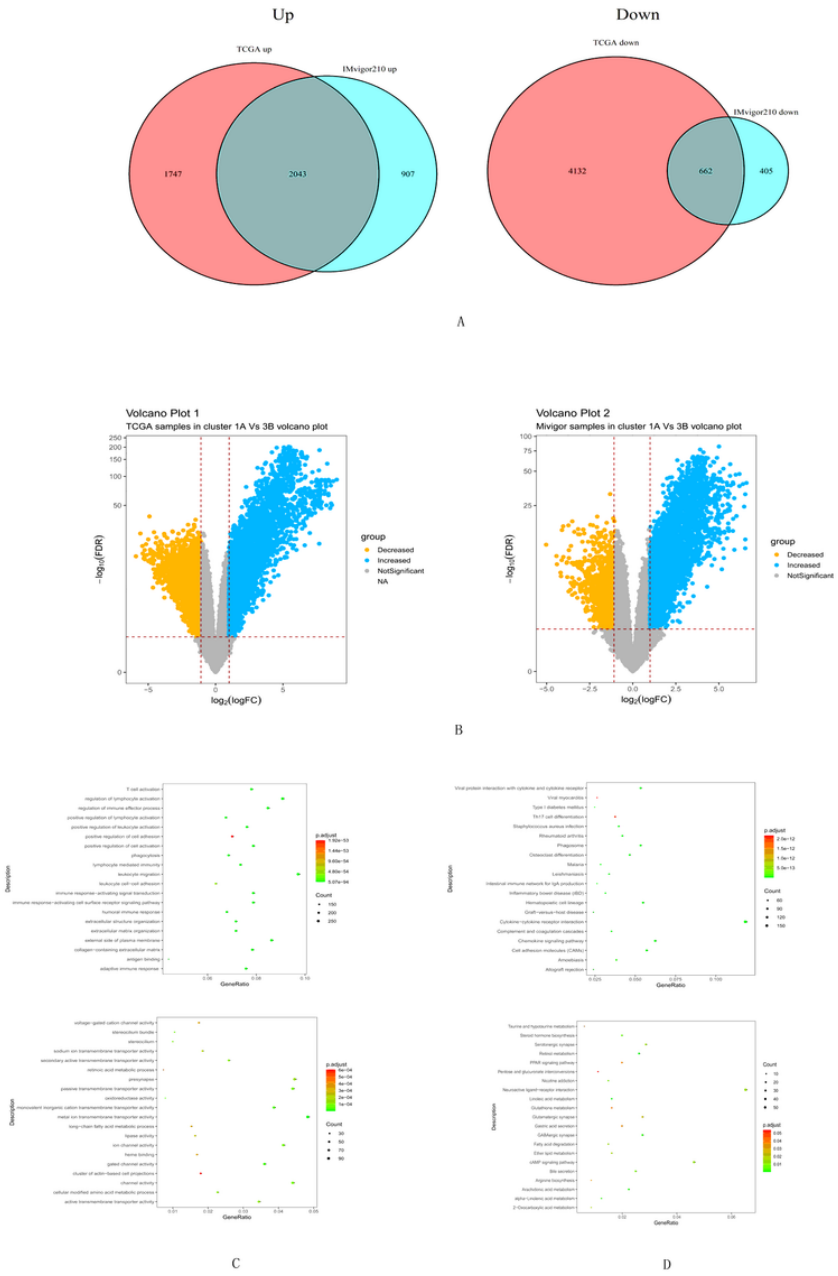

Figure 3

DEGs between cluster 1A and cluster 3C. A Venn diagram of upregulated and downregulated DEGs of the TCGA cohort and IMvigor210 cohort. B Volcano plots of DEGs in TCGA and IMvigor 210 cohort. C The GO analysis of upregulated and downregulated DEGs in cluster 3C. D KEGG analysis of upregulated and downregulated genes in cluster 3C.

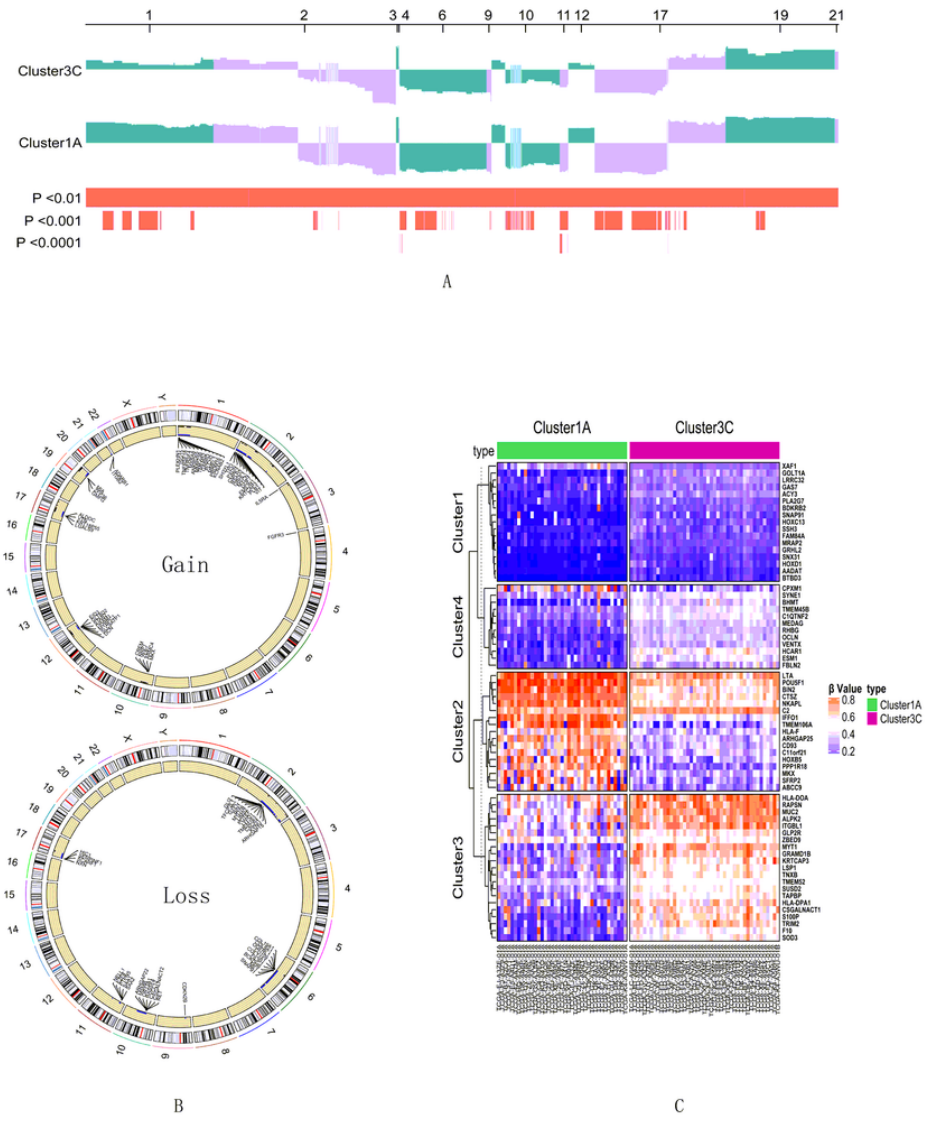

Figure 4

Differences in CNVs and DNA methylation among cluster 1A and cluster 3C. A CNV plot shows relative frequency of copy-number gains and deletions among DEGs between cluster 1A and 3C. 159 DEGs were upregulated in cluster 3C in regions with a higher frequency for copy-number deletions in cluster 1A, while 62 DEGs were upregulated in cluster 1A encoded by regions with more frequent gains in the cluster. B Circos plot display chromosomal postitions with statistically significant differences in gains or losses between cluster 1A and cluster 3C. C 349 differentially methylated probes were extracted comparing cluster 1A and cluster 3C. Heatmap shows 41 differentially methylated probes annotated for promoters of DEGs among cluster 1A versus cluster 3C.

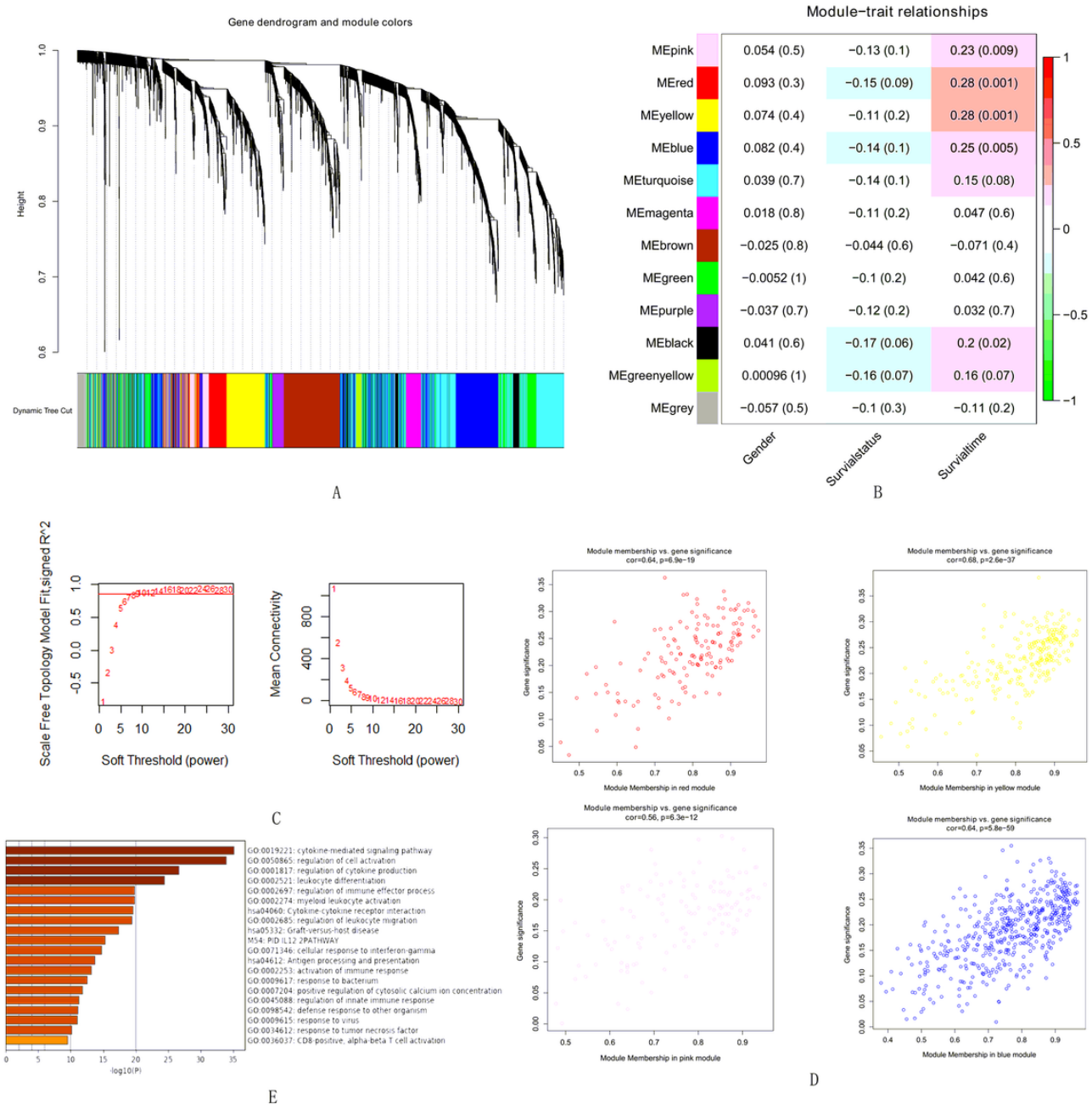

Figure 5

WGCNA analysis of DEGs between cluster 1A and cluster 3C. A Clustering dendrograms showing 12 modules that contain a group of highly connected genes. Each designated color represents a certain gene module. B Heatmap shows correlations of module epigengenes with survival time. C Determination of soft threshold in WGCNA algorithm. The power of β was set at 10 to ensure a scale-free network. D Gene correlation scatter plots of the red, yellow, pink, and blue module. E GO analysis of the four modules.

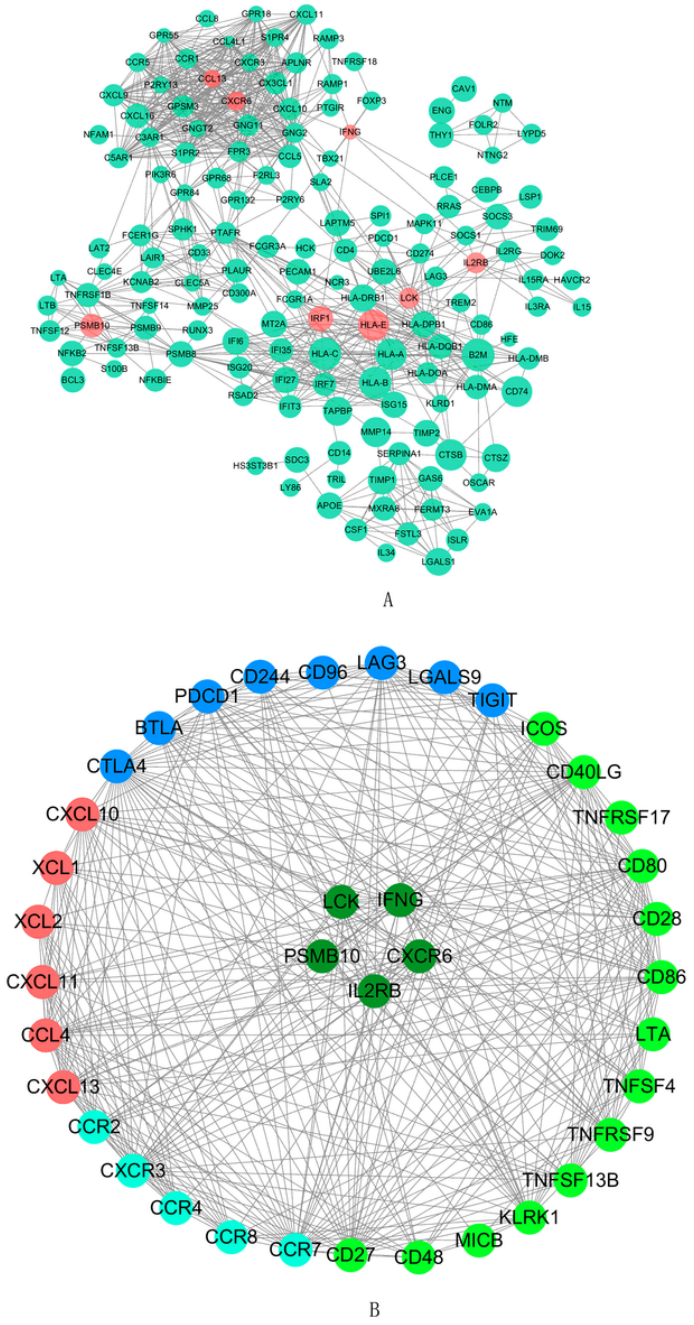

Figure 6

PPI. A The visualization of modules with hub genes. The thicker line represents higher connection strengths. B PPI map of the 5 hub genes.

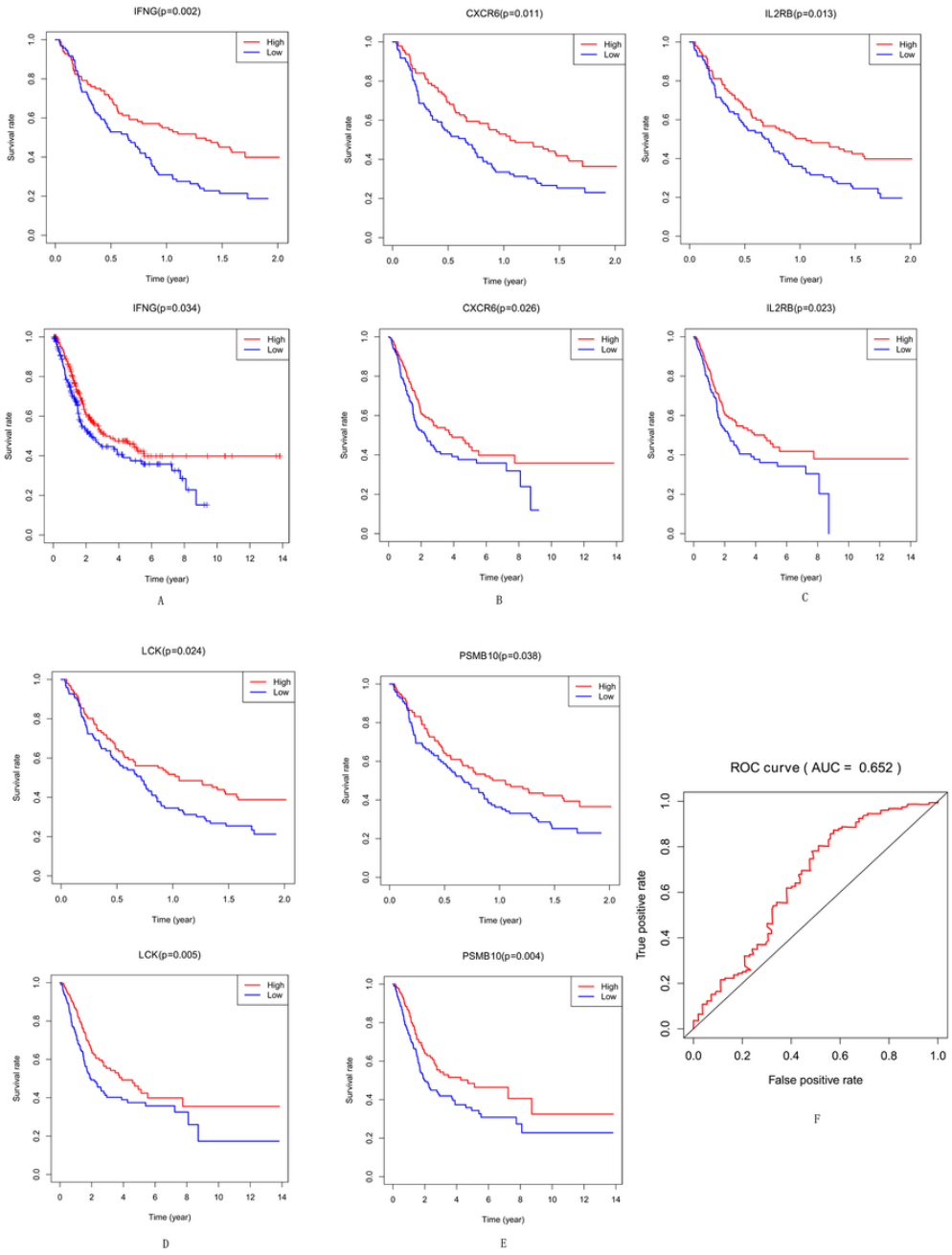

Figure 7

Selection and validation of hub genes related to prognosis. Survival curves for testing hub genes, IFNG, CXCR6, IL2RB, LCK, and PSMB10 in TCGA cohort and IMvigor210 cohort were shown in A, B, C, D, and E. ROC curve for the model representing 5-year prediction was shown in F.

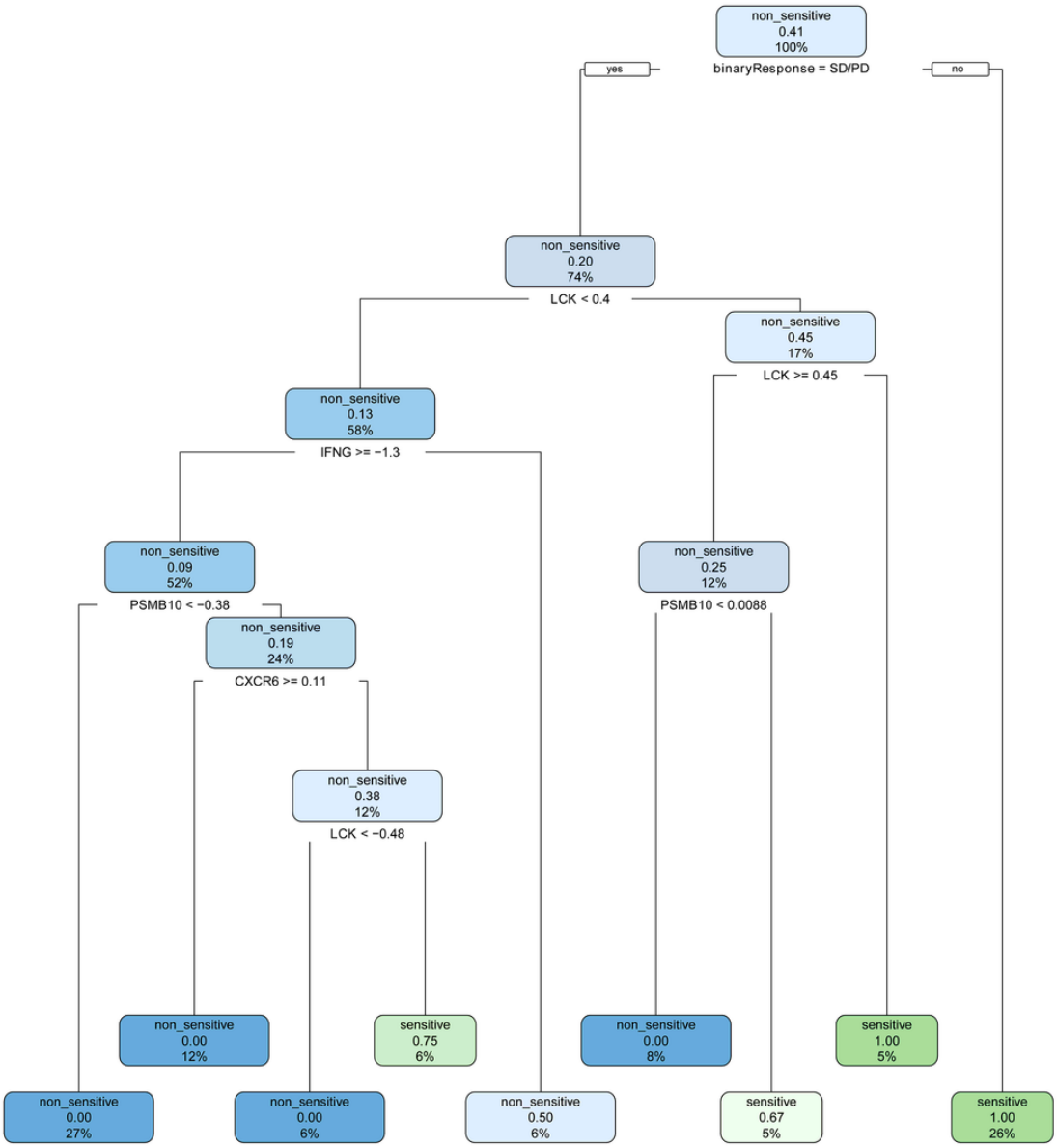

Figure 8

Decision tree. Decision tree calculated by CART to classify SD and PD patients receiving ICB therapy.

Supplementary Files

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

- supplementaryfigure2.tif
- supplementaryfigure1.tif
- SupplementaryTable3.xlsx
- SupplementaryTable2.xlsx
- SupplementaryTable1.xlsx