OMTO, Volume *20*

Supplemental Information

Tumor immune microenvironment-based classifications of bladder cancer for enhancing the response rate of immunotherapy

Jialin Meng, Xiaofan Lu, Yujie Zhou, Meng Zhang, Qintao Ge, Jun Zhou, Zongyao Hao, Shenglin Gao, Fangrong Yan, and Chaozhao Liang

Supplementary methods

Pattern discovery of immune expression and unsupervised analysis

In TCGA training cohort, tumour, stromal, and immune cell transcriptome profiling data were virtually microdissected employing unsupervised NMF method as previously described [1] via GenePattern [2]. The NMF algorithm, which is suitable for decomposing biological data, can factorize the gene expression matrix V (*n* genes × *m* samples) into two matrixes: gene factor matrix *W* of (*n* genes × *k* factors) and sample factor matrix *H* of (*m* samples × *k* factors) [3]. We chose $k = 9$ as the number of factors or expression patterns, given it could produce a high cophenetic coefficient [1] as well as effectively decompose the dataset in our TCGA training cohort. The identification of an immune class, as reported similarly by Sia *et al.* [4], involved the following steps. Firstly, identification of immune-related NMF factors was achieved through single-sample set enrichment analysis (GenePattern module "ssGSEA") of immune enrichment score (IES) gene signature [5]. To obtain the robust immune module, we pre-set the numbers of module as five to 10, respectively. When the total modules is nine, the first module strongly enriched the patients with a highly IES while the average IES of other factors are low, therefore, this module was then named as the "immune module" .

The top 150 weighted genes (**Table S1**) in the immune module were defined as the exemplar genes which could inflect the characteristics of the immune module, these genes were ranked according to the descending order by difference between factor loading value in first column of matrix *W* (immune factor weight) and the largest factor loading in other columns of *W*. Secondly, the top 150 exemplar genes were selected to classify into two

preliminary subgroups, immune and non-immune for the TCGA training cohort. This procedure was accomplished by supervised clustering via GenePattern module "NMFConsensus". Finally, the immune and non-immune classes were adjusted by the multidimensional scaling (MDS) random forest method, which could visualize the level of similarity of individual cases of a dataset[4]. The immune class was furthermore divided into immune-exhausted, and immune-activated subtypes by the nearest template prediction (GenePattern module 'NTP') of the activated stroma [6].

Correlation of Immune class with copy number alterations, tumour-infiltrating lymphocytes

The tumour-infiltrating lymphocytes (TIL) abundance estimated by H&E stained whole-slide images of TCGA samples were obtained from a previous study [7]. Copy number alterations (CNA) data were generated by GISTIC2.0 from GDAC Firehose (https://gdac.broadinstitute.org). We compared the differences in amplification or deletion events of both focal and arm level between Immune and non-Immune classes. The neoantigen number was accessed from a previous study by Rooney *et al.* [8]. The mutation data were retrieved from TCGA (https://tcga-data.nci.nih.gov); we calculated the number of nonsynonymous mutations per million bases to evaluate tumour mutation burden (TMB). Whats more, the mutation landscape Oncoprint was drawn by R package "maftools"[9]. The different distribution of gene mutations among immunosubtypes were evaluated by the Chi-square test.

Molecular characterization of Immune class

Hand-curated gene signatures representing various immune cell types or host anti-tumour immunity (**Table S3**) from literature and databases were used to characterize immune class in TCGA cohort. Immune-exhausted and activation subtypes was identified by using ssGSEA (GenePattern module "ssGSEA") and nearest template prediction (GenePattern module "NTP") of stroma activation[2]. The signature of stroma activation was derived from Figure 2 of Moffitt et al.'s work [6]. Overexpression or downregulation of genes in immune vs. non-immune classes was performed by "limma" package with R, genes with a false discovery rate (FDR) < 0.05 and a $\log_2$ fold change (FC) ≥ 1 were considered differentially expressed between two groups. Subsequently, gene set enrichment analysis (GSEA, http://www.broadinstitute.org/gsea/index.jsp) was performed to determine gene sets and pathways enriched in Immune vs. non-Immune classes.

Validation of immune molecular subtypes in independent external datasets

We identified top 150 upregulated genes between immune and non-immune classes (**Table S4**). Then NMF-based consensus clustering based on the immune classifier was applied to identify the three immunophenotypes in three independent external datasets (**Table 3**) using GenePattern module "NMFConsensus" with the 150 DEGs. Immune-related gene signature ssGSEA scores were calculated to feature molecular characteristics and validate the existence of abovementioned immune molecular subtypes in each dataset, and then, the immune class divided to activated and exhausted subgroups by the and nearest template prediction (NTP) module.

References:

1. Brunet, J.P., Tamayo, P., Golub, T.R., and Mesirov, J.P. (2004). Metagenes and molecular pattern discovery using matrix factorization. Proc Natl Acad Sci U S A *101,* 4164-4169.
2. Reich, M., Liefeld, T., Gould, J., Lerner, J., Tamayo, P., and Mesirov, J.P. (2006). GenePattern 2.0. Nat Genet *38,* 500-501.
3. Lee, D.D., and Seung, H.S. (1999). Learning the parts of objects by non-negative matrix factorization. Nature *401,* 788-791.
4. Sia, D., Jiao, Y., Martinez-Quetglas, I., Kuchuk, O., Villacorta-Martin, C., Castro de Moura, M., Putra, J., Camprecios, G., Bassaganyas, L., Akers, N., et al. (2017). Identification of an Immune-specific Class of Hepatocellular Carcinoma, Based on Molecular Features. Gastroenterology *153,* 812-826.
5. Yoshihara, K., Shahmoradgoli, M., Martinez, E., Vegesna, R., Kim, H., Torres-Garcia, W., Trevino, V., Shen, H., Laird, P.W., Levine, D.A., et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun *4,* 2612.
6. Moffitt, R.A., Marayati, R., Flate, E.L., Volmar, K.E., Loeza, S.G., Hoadley, K.A., Rashid, N.U., Williams, L.A., Eaton, S.C., Chung, A.H., et al. (2015). Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. Nat Genet *47,* 1168-1178.
7. Saltz, J., Gupta, R., Hou, L., Kurc, T., Singh, P., Nguyen, V., Samaras, D., Shroyer, K.R., Zhao, T., Batiste, R., et al. (2018). Spatial Organization and Molecular Correlation of Tumor-Infiltrating Lymphocytes Using Deep Learning on Pathology Images. Cell Rep *23,* 181-193 e187.

8. Rooney, M.S., Shukla, S.A., Wu, C.J., Getz, G., and Hacohen, N. (2015). Molecular and genetic properties of tumors associated with local immune cytolytic activity. Cell *160*, 48-61.

9. Mayakonda, A., Lin, D.C., Assenov, Y., Plass, C., and Koeffler, H.P. (2018). Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res. *28*, 1747-1756.

Supplementary Tables

Table S1. Top 150 weighted genes.

Table S2. Top 150 exemplar genes of the immune module enriched in the immune associated ontology biological process and KEGG pathways.

ID	Description	P adjust
Ontology Biological Process		
GO:0042110	T cell activation	5.48E-49
GO:0019882	antigen processing and presentation	8.12E-10
GO:0042113	B cell activation	1.37E-09
GO:0038110	interleukin-2-mediated signaling pathway	4.30E-03
GO:2000316	regulation of T-helper 17 type immune response	1.91E-01
GO:0070098	chemokine-mediated signaling pathway	3.33E-19
KEGG Pathway		
hsa04658	Th1 and Th2 cell differentiation	1.80E-10
hsa04060	Cytokine-cytokine receptor interaction	1.56E-21
hsa04650	Natural killer cell mediated cytotoxicity	1.03E-06
hsa04660	T cell receptor signaling pathway	3.65E-07
hsa04662	B cell receptor signaling pathway	3.48E-04
hsa05235	PD-L1 expression and PD-1 checkpoint pathway in cancer	5.89E-04

Table S3. Immune associated gene signatures used in this study.

Signature Name	Reference
Immune enrichment score	Yoshihara *et al.* Nat Commun. 2013;4:2612
Stromal enrichment score	Yoshihara *et al.* Nat Commun. 2013;4:2612
Immune signalling molecules	Cancer Genome Atlas Network. Cell. 2015;161:1681-96
13 T-cell signature	Spranger *et al.* Proc Natl Acad Sci U S A. 2016;113(48):E7759-E7768.
T cells	Bindea *et al.* Immunity. 2013;39:782-95
CD8 T cells	Bindea *et al.* Immunity. 2013;39:782-95
Treg cells	Angelova *et al.* Genome Biol. 2015;16:64
TITR signature	Magnuson *et al.* PANS. 2018;115(45):E10672-e81
MDSC	Angelova *et al.* Genome Biol. 2015;16:64
T.NK. metagene	Alistar *et al.* Genome Med. 2014;6:80
B-cell cluster	Iglesia *et al.* Clin Cancer Res. 2014;20(14):3818–3829.
B.P. metagene	Alistar *et al.* Genome Med. 2014;6:80
Macrophages	Bindea *et al.* Immunity. 2013;39:782-95
TLS	Finkin *et al.* Nat Immunol. 2015;16:1235-44
6-gene IFN signature	Chow *et al.* J Clin Oncol. 34, (suppl; abstr 6010) 2016
CYT	Iglesia *et al.* Clin Cancer Res. 2014;20(14):3818–3829.
WNT/TGF-β signature	Lachenmayer *et al.* Clin Cancer Res. 2012;18:4997-5007
TGF-β1 activated	Ingenuity Pathway Analysis
C-ECM signature	Chakravarthy *et al.* Nature Communications. 2018;9(1)
Six immune subtypes of Pan-Cancer Atlas	Thorsson *et al.* Immunity. 2018;5(5):489-500
PAM50 pan-cancer	Zhao *et al.* Clin Cancer Res. 2019;25(8):2450-2457

Abbreviations: TITR, tumour-infiltrating Tregs; MDSC, myeloid-derived suppressor cell; IFN: interferon; TLS, tertiary lymphoid structure; CYT, cytolytic activity score; C-ECM, cancer-associated extracellular matrix.

Table S4. Top 150 difference genes in immune class compare with non-immune class in TCGA training cohort.

Table S5. List of the 21 genes selected form random forest algorithm.

Gene name

BCL2A1	GZMB
C1QA	GZMH
C1QB	HAVCR2
C1QC	HK3
CCL3	LAG3
CD2	LILRB2
CD8A	NKG7
CTLA4	PRF1
CXCL9	SLAMF8
FCER1G	TNFSF13B
GBP5	

Supplementary Figures

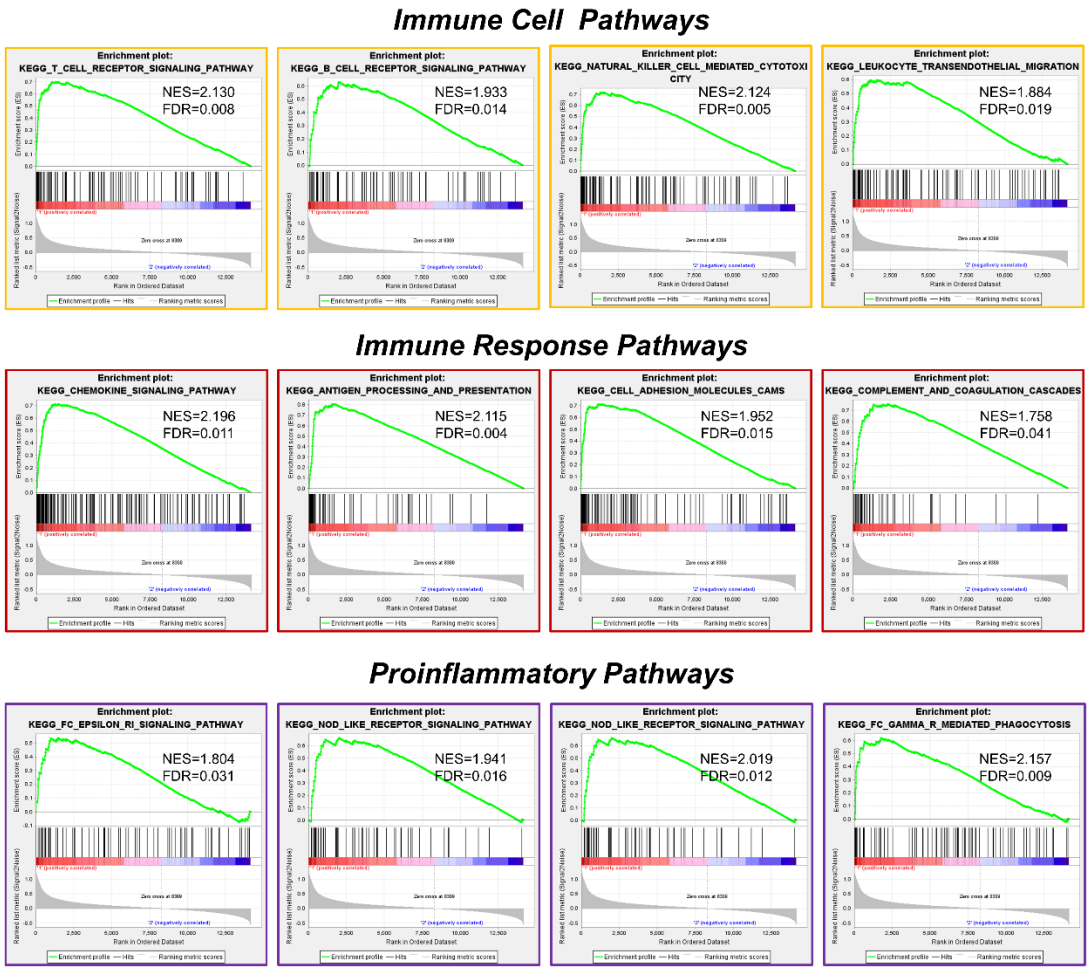

Figure S1

Figure S1. GSEA results showing the activated signaling pathways in the immune class. NES, normalized enrichment score; FDR, false discovery rate; FDR less than 0.05 indicates statistical significance.

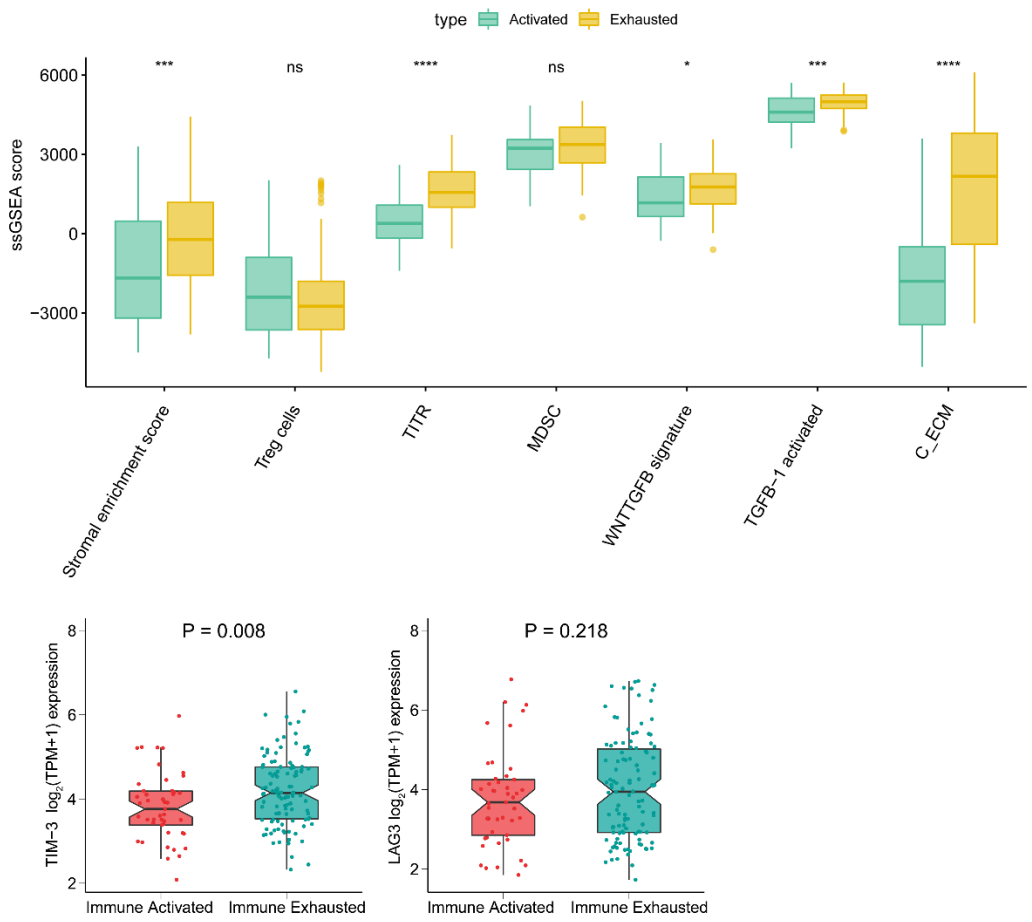

Figure S2

Figure S2. Stromal representative signatures and markers between immune-activated and immune-exhausted subgroups. ****, P<0.0001; ***, P<0.001; **, P<0.01; *, P<0.05; ns, no significance.

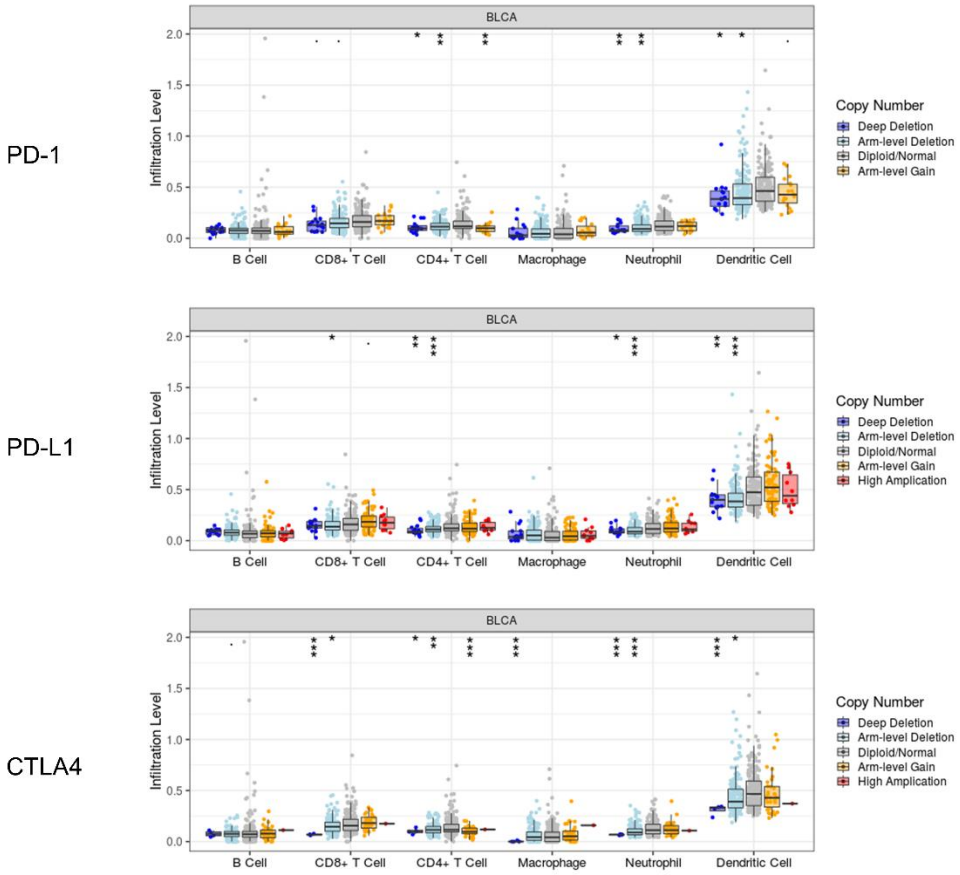

Figure S3

Figure S3. The association between copy number alteration of immune checkpoints and immunocyte infiltration. ***, P<0.001; **, P<0.01; *, P<0.05.

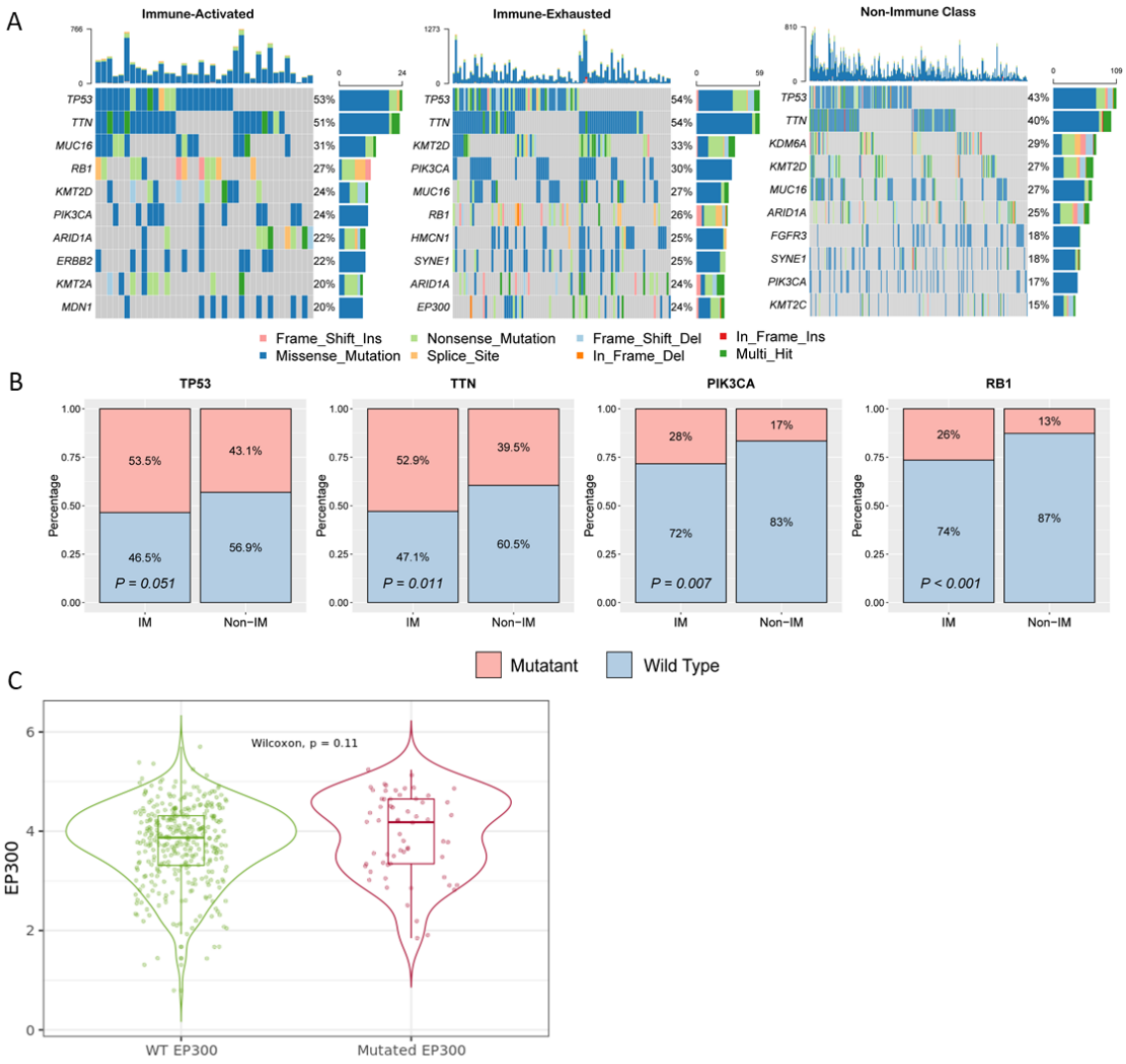

Figure S4

Figure S4. The top mutant genes of non-immune class, immune-activated subgroup, and immune-exhausted subgroup.

(A) Different distribution of mutant genes in three immunophenotypes. (B) TP53, TTN, PIK3CA and RB1 are the specific mutant genes in immune class compared with non-immune class. (C) The expression of EP300 in EP300wild type and mutated patients.

Figure S5. Reappearing the diverse immune characteristics among three immunophenotypes in E-MTAB-4321, GSE32894, GSE83586, GSE87304, GSE128702, GSE13507, GSE129871, GSE120736, GSE39016 cohorts.

CYT, cytolytic activity score; TITR, tumor-infiltrating Tregs; MDSC, myeloid-derived suppressor cell; TLS, tertiary lymphoid structure; C-ECM, cancer-associated extracellular matrix.

Figure S6. Reappearing the diverse immune characteristics among three immunophenotypes in GSE128701, GSE124035, GSE86411, GSE48276, GSE128192, GSE31684, GSE134292, GSE93257, E-MTAB-1803, GSE69795.

CYT, cytolytic activity score; TITR, tumor-infiltrating Tregs; MDSC, myeloid-derived suppressor cell; TLS, tertiary lymphoid structure; C-ECM, cancer-associated extracellular matrix.

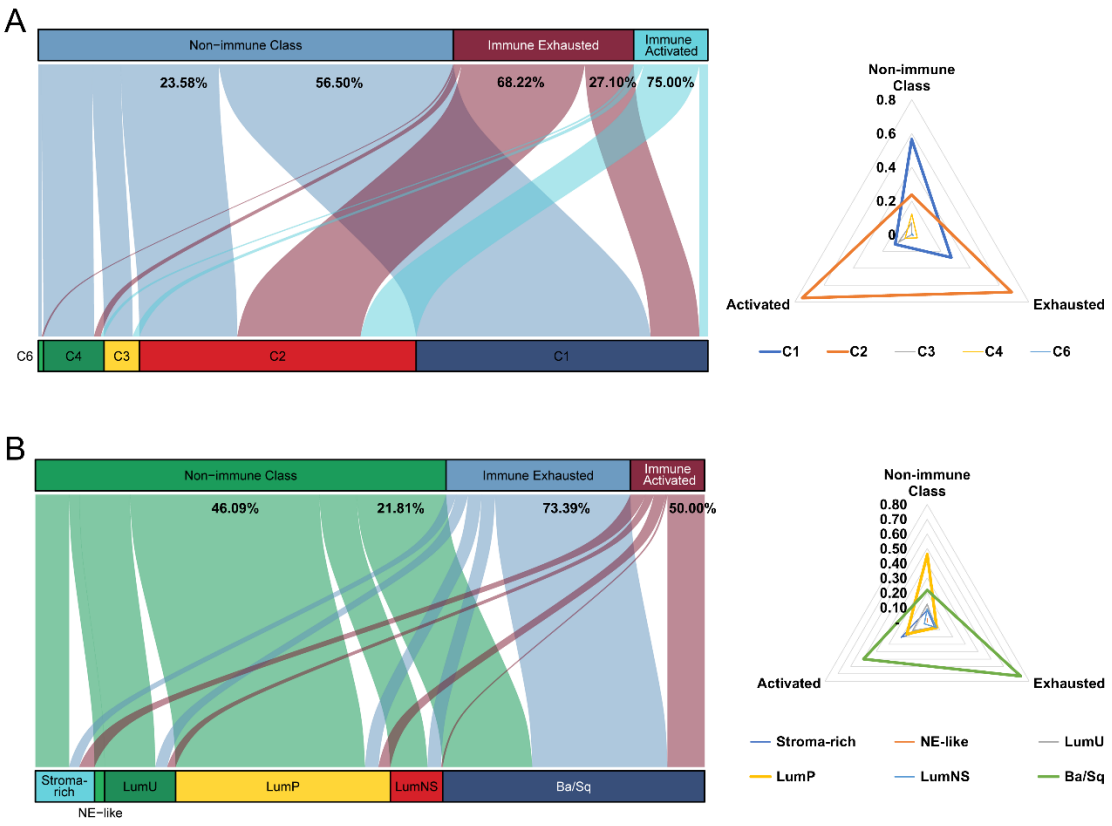

Figure S7

Figure S7. Correlate the three immunophenotypes with proposed molecular subtypes.

(A) Association with Thorsson et al. generated pan-cancer six immune molecular features;

(B) Association with Kamoun et al. identified the consensus set of six molecular classes.

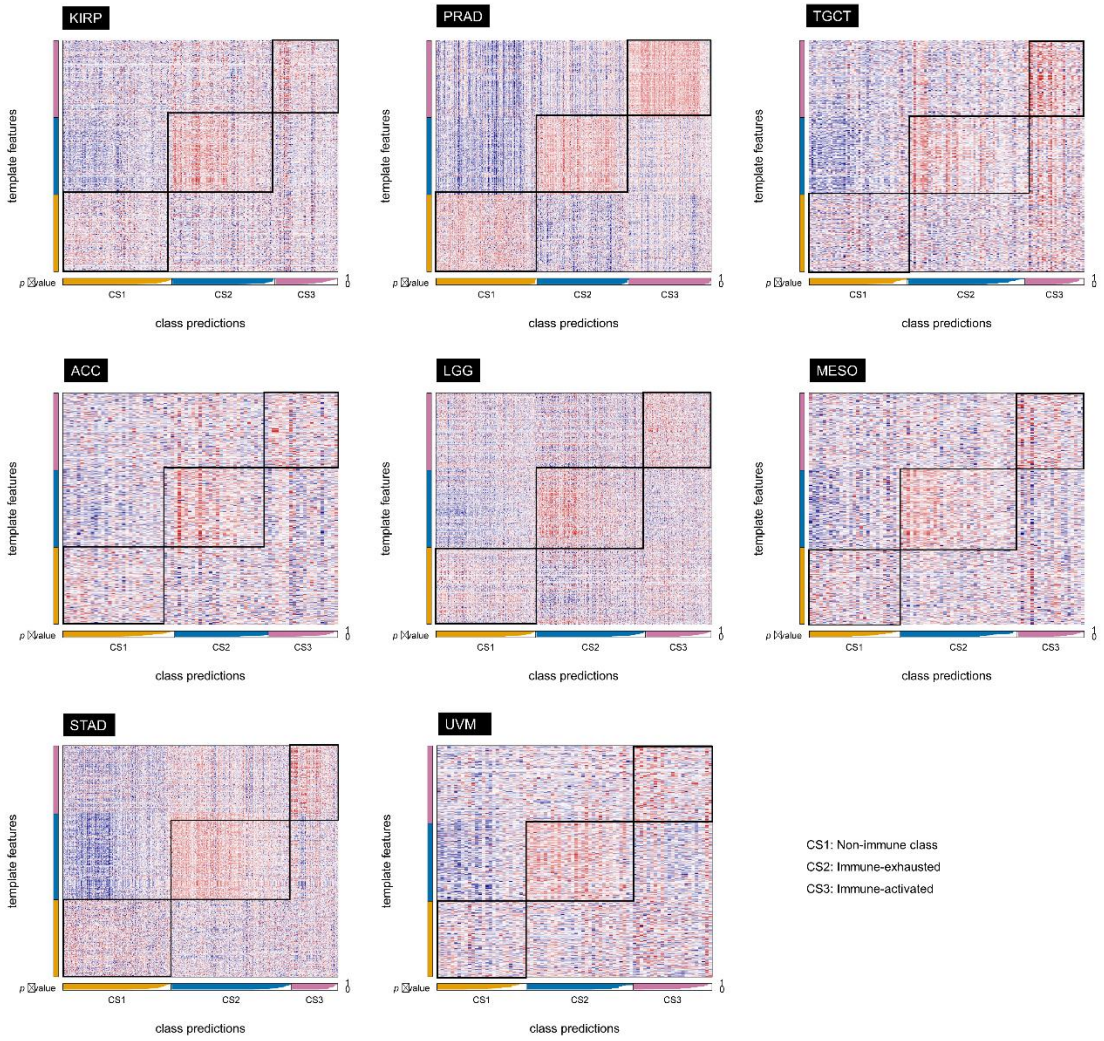

Figure S8

Figure S8. Verify the three immunophenotypes in pan-cancer.

KIRP, papillary renal cell carcinoma; PRAD: prostate cancer; TGCT: testicular germ cell tumor; ACC, adrenocortical carcinoma; LGG, brain lower grade glioma; MESO, mesothelioma; STAD, stomach adenocarcinoma; UVM, uveal melanoma.

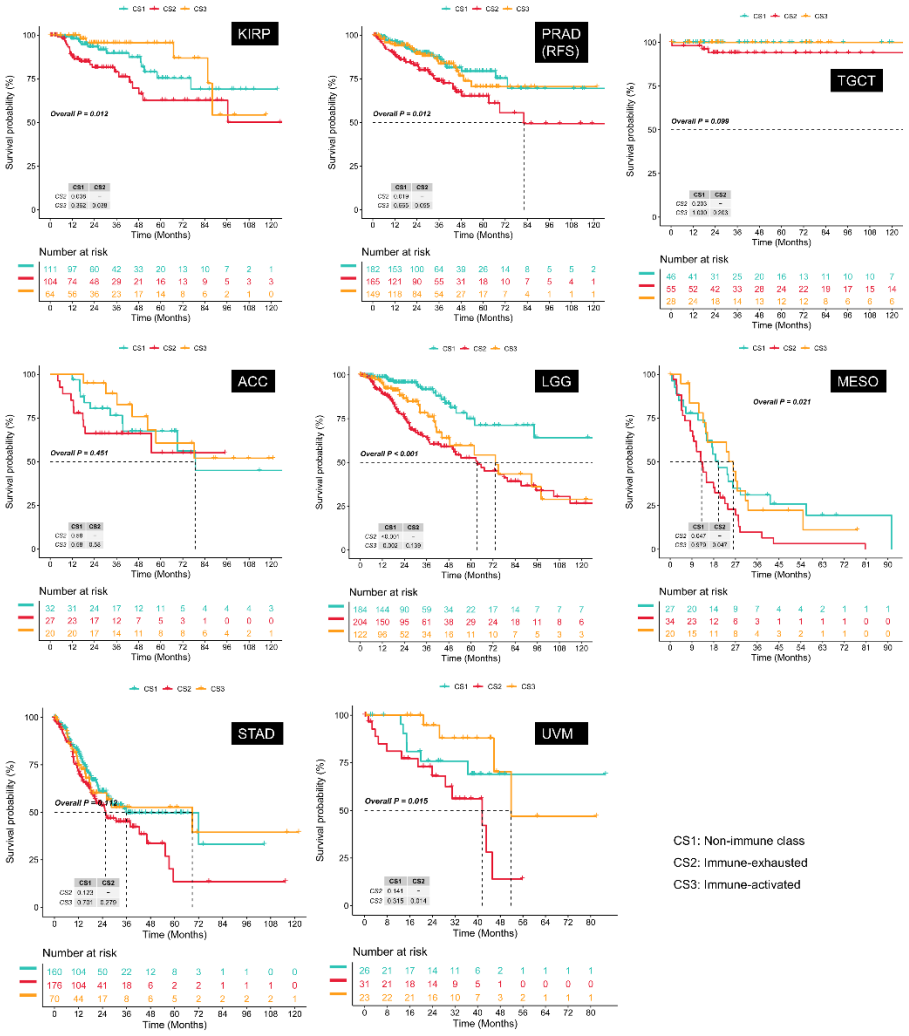

Figure S9

Figure S9. Different overall survival outcome of the three immunophenotypes in pan-cancer.

KIRP, papillary renal cell carcinoma; PRAD: prostate cancer; TGCT: testicular germ cell tumor; ACC, adrenocortical carcinoma; LGG, brain lower grade glioma; MESO, mesothelioma; STAD, stomach adenocarcinoma; UVM, uveal melanoma.

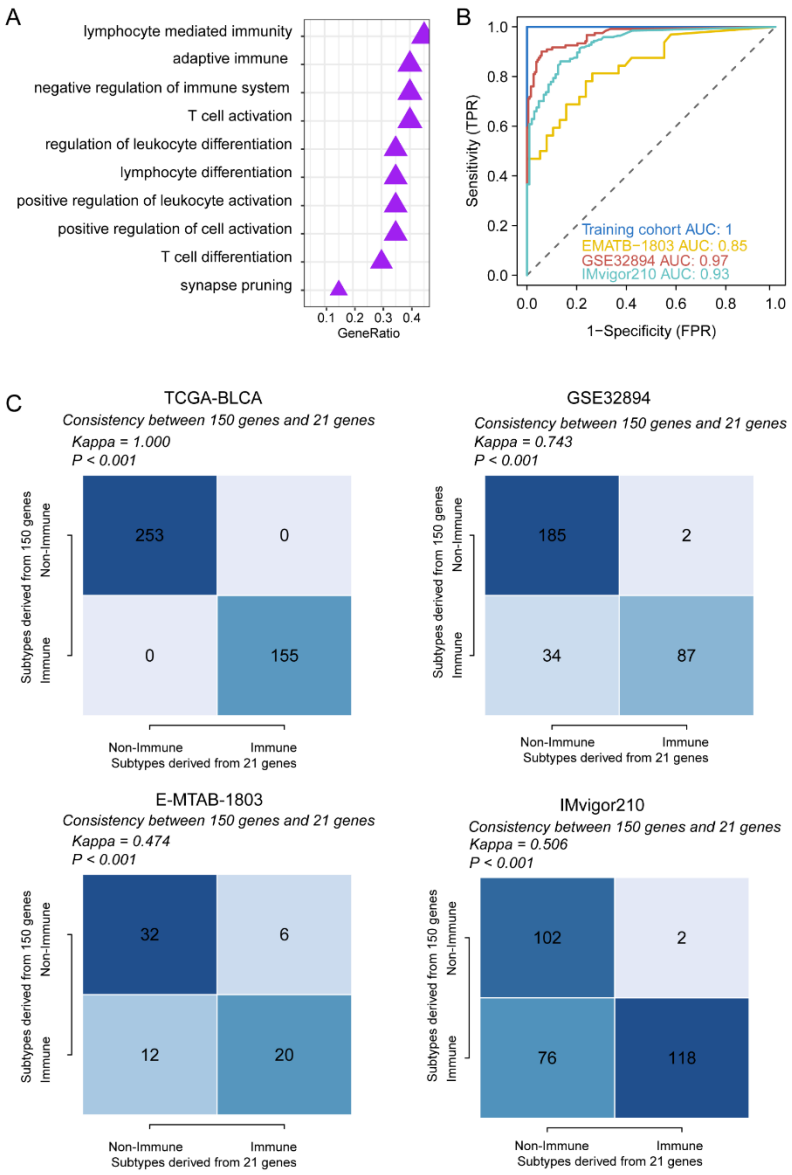

Figure S10. Dimensionality reduction of the 150 DEGs for the distinguishment of immune and non-immune classes

(A) Biological pathway enrichment of 21 genes; (B) ROC curve showing the distinguish value of the 21 genes in different cohort; (C) The consistency between 150 genes and 21 genes defined immune and non-immune classes.