Macrophage correlates with immunophenotype and predicts anti-PD-L1 response of urothelial cancer

Dongqiang Zeng†, Zilan Ye†, Jiani Wu¹, Rui Zhou¹, Xinxian Fan², Gaofeng Wang³, Yiqiang Huang⁴, Jianhua Wu¹, Huiying Sun¹, Miaohong Wang¹, Jianping Bin⁵, Yulin Liao⁵, Nailin Li⁶, Min Shi¹, Wangjun Liao¹*

¹Department of Oncology, Nanfang Hospital, Southern Medical University, Guangzhou, Guangdong, P. R. China; ²Department of Urology, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, China; ³Department of Dermatology, Johns Hopkins School of Medicine, Baltimore, Maryland, USA; ⁴Department of Urology, Daping Hospital, Army Medical University, Chongqing, P.R. China; ⁵Department of Cardiology, State Key Laboratory of Organ Failure Research, Nanfang Hospital, Southern Medical University, Guangzhou, China; ⁶Karolinska Institutet Department of Medicine-Solna, Clinical Pharmacology Group, Karolinska University Hospital-Solna, 171 76, Stockholm, Sweden.

† These authors contributed equally to this work.
* Correspondence author: Department of Oncology, Nanfang Hospital, Southern Medical University, 1838 North Guangzhou Avenue, Guangzhou 510515, China. E-mail: nfyyliaowj@163.com (W. Liao)
Supplementary Methods

1. Inference of Immune Cell Infiltration

- **CIBERSORT** [1]: To quantify the proportions of immune cells in the tumor samples, we used the CIBERSORT algorithm and the LM22 gene signature, which allows for highly sensitive and specific discrimination of 22 human immune cell phenotypes including B cells, T cells, NK cells, macrophages, dendritic cells (DCs), and myeloid subsets. CIBERSORT is a deconvolution algorithm that uses a set of reference gene expression values (a signature with 547 genes) considered a minimal representation for each cell type and, based on those values, infers cell type proportions in data from bulk tumor samples with mixed cell types using support vector regression. Gene expression profiles were prepared using standard annotation files and data were uploaded to the CIBERSORT web portal (http://cibersort.stanford.edu/), with the algorithm run using the LM22 signature at 1,000 permutations.

- **MCP-counter** [2]: Proportions of stromal cell were also applied Microenvironment Cell Populations-counter (MCP-counter) method, which allows the robust quantification of the absolute abundance of eight immune and two stromal cell populations in heterogeneous tissues from transcriptomic data.

- **TIMER** [3]: Allows explorations of the disease-specific clinical impact of different immune infiltrates in the tumor microenvironment. TIMER was developed to estimate the abundance of six tumor-infiltrating immune cell types (B cells, CD4 T cells, CD8 T cells, neutrophils, macrophages, and dendritic cells) to study 23 cancer types in The Cancer Genome Atlas (TCGA). This tool was validated thanks to Monte Carlo simulations, orthogonal estimates from DNA methylation-based inferences, as well as pathological assessment.

- **EPIC** [4]: EPIC package uses a constrained least square minimization to estimate the proportion of each cell type with a reference profile and another uncharacterized cell type in bulk gene expression samples.

- **xCell** [5]: Performs cell type enrichment analysis from gene expression data for 64 immune and stroma cell types. xCell is a gene signatures-based method learned from thousands of pure cell types from various sources. xCell applies a technique for reducing associations between closely related cell types. xCell signatures were validated using extensive in-silico simulations and cytometry immunophenotyping, and were shown to outperform previous methods.
2. Tumor Purity Assessment
Tumor purity was assessed computationally in all longitudinal samples using estimates derived from RNA-seq data using ESTIMATE algorithm [6] that uses gene expression signatures to infer the fraction of stromal and immune cells in tumor samples.

3. BLCA molecular subtyping
Molecular subtypes of tumor samples were estimated using the approach of BLCAsubtyping [7], an R package which integrates six published molecular classifications.

4. Gene expression profile obtained from GEO and preprocessing
Raw data from the microarray datasets generated using Affymetrix® and Illumina® were downloaded from the Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/). The raw data for the dataset from Affymetrix® were processed using the RMA algorithm for background adjustment using the Affy package [8]. RMA was used to perform background adjustment, quantile normalization, and final summarization of oligonucleotides per transcript using the median polish algorithm. The raw data for the dataset from Illumina® were processed using the lumi package. The ComBat algorithm [9] was applied to reduce the likelihood of batch effects from non-biological technical biases.

5. Calculating signature score of gene sets using PCA algorithm
For gene expression (normalized by RMA or TPM methods) matrix, the expression of each gene in a signature was standardized so that its mean expression was zero, and standard deviation was 1 across samples. Then, a principal component analysis (PCA) was performed, and principal component 1 was extracted to serve as the gene signature score. This approach has the advantage of focusing the score on the set with the largest block of well-correlated (or anti-correlated) genes in the set, while down-weighting contributions from genes that do not track with other set members [10, 11].

6. Feature Engineering
To avoid the shortcoming of overfitting of lasso cox regression model developed by the training dataset, feature engineering was conducted. Firstly, all features (N = 7556) were standardized across all samples. Then, univariate Log-rank (Mantel-Cox) test was conducted to seek out signatures that were associated with anti-PD-L1 survival outcome with a cutoff of P = 0.01. Moreover, Mann-Whitney U test was applied to find out features related to treatment response (CR/PR vs SD/PD) with a cutoff of p = 0.05. As a result, we got 780 features for training a LASSO cox regression model.
7. Bootstrap sampling to obtain robustness of predictors
LASSO Cox regression [12] was performed for the signatures (N = 780) passing the initial filtering of feature engineering using 80 percent of the training set, randomly selected. This procedure was repeated 10,000 times and we got 10,000 cox regression model which was used to calculate the frequency of each feature that was enrolled in anti-PD-L1 immunotherapy predictive model.

8. Assessment of prognostic and predictive value of biomarkers
The sensitivity and specificity of the survival prediction based on the risk model and signature scores were depicted by the time-dependent receiver operating characteristic curve (ROC) and quantified by the area under the ROC (AUC) using timeROC and survivalROC package [13]. “roc.test” function of pROC package [14] was used to compares the AUC or partial AUC of two correlated or uncorrelated ROC curves.

9. Consensus clustering of TME infiltrating pattern
Cell fraction matrix deconvoluted by CIBERSORT [1] was scaled, and unsupervised clustering methods (K-means) [15] for analysis of dataset was applied to identify immune cell infiltrating pattern and classify patients for further analysis. This procedure using ConsensusClusterPlus R package [16] was repeated 1000 times to ensure the stability of classification.

10. Gene Set Enrichment Analysis and visualization
Gene sets were downloaded from the MSigDB database (v.6.2) of Broad Institute [17], then HALLMARK [18] and KEGG [19] gene sets were selected to perform over-representation hypergeometric test [17]. Enrichment P values were based on 1,000 permutations and subsequently adjusted for multiple testing using the Benjamini-Hochberg procedure to control the FDR [20]. The R package enrichplot (https://github.com/GuangchuangYu/enrichplot), implements several visualization methods to help interpreting enrichment results and was adopted to visualize GSEA result.

11. Other gene signatures enrolled in this study
To characterize the metabolism, immune microenvironment and other prevalent gene signatures activation in each tumor samples, PCA algorithm was apply to determine the pathway activity using gene sets (see Supplementary Table S11) curated by Mariathasan et al. [10], Cristescu et al. [21], Rooney et al. [22], Rosario et al. [23] and Zeng et al.[24]. We thereby obtained, for each signature, an enrichment score per sample that indicated the extent of upregulation or downregulation of the associated genes. A minimum overlap of two genes was required.
References


Supplementary Figures
Figure S1

Flowchart of LASSO cox regression model construction and validation.

Deconvolution of signatures ($N = 7556$)
Gene sets: GO, KEGG, REACTOME, HALLMARK, Other signatures;
Cell types: CIBERSORT, TIMER, EPIC, xCell, MCP-counter, ESTIMATE;

Feature Engineering: ($N = 780$)
① Response relevant features; (Wilcoxon test, adjust p value <= 0.05);
② Features associated with outcome (Log rank test p value <= 0.01);
③ Standardized features;

Randomly assign samples ($N = 348$)
Training cohort vs Validation cohort: (6 : 4)

LASSO cox regression model construction
① 10 time cross-validations;
② minimal lambda;
③ using the coefficients of features to construct risk score model;

Validation of risk score model
① using the same coefficients of features;
② using the same cutoff of risk score to validate the prognostic value;

Bootstrap sampling to obtain robustness of predictors
① Randomly choose 80% of patients;
② Using LASSO cox regression algorithm to select combination of signatures;
③ Calculate the frequency of features.
Correlation and comparison of M1 macrophage with reported biomarkers.

(A-C) The predictive capacity of M1 macrophage (A) in estimating overall survival was superior to TMB (B), TNB (C). (M1: 12-month AUC = 0.647, 24-month AUC = 0.707; TMB: 12-month AUC = 0.644, 24-month AUC = 0.666; TNB: 12-month AUC = 0.647, 24-month AUC = 0.704.) TMB: tumor mutation burden; TNB: tumor neoantigen burden.

(D) M1 macrophage was statistically associated with favorable survival outcome independent of TNB (Kaplan-Meier survival analyses, \( p < 0.0001 \)).

(E-F) Distribution of M1 macrophage varied among UNC subtypes (F) and Consensus Class (G) respectively (Mann Whitney U test, \( p = 1.2e-04, p = 1.3e-10 \); respectively).

(G) Multi-variates regression analysis of hazard ratio among M1 macrophage and prior reported biomarkers. M1 macrophage collaborated with TMB were protect factors in resistance to tumor progression. Additional statistical analysis of biomarkers refer to Table S6.
M1 macrophage reproductively correlates with immunophenotype in TCGA.

(A) Unsupervised consensus clustering analysis of TCGA based on the TME-cell signatures inclined to divide into two TME clusters. The plot of two and three clusters were both displayed.

(B) Heatmap of (red = high expression; blue = low expression) TME pattern of TCGA data with two TME clusters A (blue) and B (yellow). Rows of the heatmap show expression of TME-infiltrating cell signatures (Z scores) calculated by CIBERSORT.

(C) TME clusters A statistically associated with better survival (Kaplan-Meier survival analysis, $p = 3.4e-02$) in TCGA BLCA dataset.

(D) TCGA BLCA dataset validated the significant correlation between TME cluster A and high M1 infiltration (Mann Whitney U test, $p < 1.2e-04$).

(E) Immunophenotype-determine capacity of M1 macrophage was validated in analysis of TCGA (M1: AUC = 0.61; monocyte: AUC = 0.497; CD8+T: AUC = 0.723).
Combination with macrophage M2 did not raise the predictive value of predictive capacity of M1 alone.

(A) Lower M2 frequency was statistically correlated with more favorable overall survival ($p = 0.011$)

(B) M2 quantification discrepancy in different response subgroups didn’t reach statistical significance ($p = 0.19$).

(C) Bare statistical correlation was observed between macrophage M1 and M2 frequency ($p = 0.916; r = 0.006$).

(D) High M1/M2 ratio was significantly associated with better response to anti-PD-L1 therapy ($p = 4.6e-06$).

(E) ROC curve suggested that M2 exerted inferior predictive sensitivity to anti-PD-L1 response. M1/M2 ratio didn’t elevate the predictive capacity of M1 alone. (M1: AUC = 0.701; M2: 0.552; M1/M2 ratio: AUC = 0.653). Corresponding $p$ values are exhibited in Table S4.
**Figure S5**

**A** Tumor microenvironment relevant signatures expression in IMvigor210.

Expression of Tumor microenvironment related signatures were elevated in high (yellow) macrophage infiltration versus the low (blue) in IMvigor210. All statistics in the figure use two-sided Mann Whitney U test. *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001. ns, not significant compared to isotype group.

**B** Metabolic signatures expression were significantly higher in low (blue) macrophage infiltration than those in high (yellow) in IMvigor210. All statistics in the figure use two-sided Mann Whitney U test. *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001. ns, not significant compared to isotype group.
Transcriptomic and metabolic programs are reproducible in TCGA. (A-B) Gene ontology (GO) (A) and KEGG pathways (B) were significantly associated with M1 macrophage infiltration in bladder TCGA data. Tumor with M1-deficient subtype have dramatically higher activation in steroid metabolism, xenobiotics metabolism whereas tumor with high-M1 infiltration embraced elevated immune activation. The top ten genes per set are shown (ranked by single-gene P value, GO: red: high, blue: low; KEGG: blue: high, green: low). Complete lists are given in Table S8. (C) GSEA analyses of TCGA validated the key pathways enriched in high (up) and low (down) M1 subset. (light blue: metabolism of xenobiotics by cytochrome P450; green: drug metabolism by cytochrome P450; blue violet: steroid hormone biosynthesis; scarlet: cytokine-cytokine receptor interaction; brick red: natural killer cell mediated cytotoxicity; dark violet: Th1 and Th2 cell differentiation; navy: antigen processing and presentation). Complete information and similar results of hallmark analyses are demonstrated in Table S8. (D) Heatmap of unsupervised clustering different expressing gene signatures is consistent in (red: high expression; blue: low expression) in TCGA. Binary M1-macrophage infiltration was show as annotation on the top (red: high; green: low). Comprehensive information is displayed in Table S9.
Validation of Tumor microenvironment and metabolic signatures profiles in TCGA.

(A) TCGA dataset externally validated that expression of Tumor microenvironment related signatures were elevated in high (yellow) macrophage infiltration versus the low (blue). All statistics in the figure use two-sided Mann Whitney U test. *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001. ns, not significant compared to isotype group.

(B) External TCGA verified that Metabolic signatures expression were significantly higher in low (blue) macrophage infiltration than those in high (yellow). All statistics in the figure use two-sided Mann Whitney U test. *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001. ns, not significant compared to isotype group.
FGFR related signatures landscape and verification of genome in TCGA.

(A-B) Signatures upregulated in FGFR mutation deficient subset (A) and mutated group (B) in IMvigor210 respectively. Steroid metabolism signatures are inclined to upregulated in FGFR mutated setting while T cell inflamed signatures and immune check point signatures in mutation-deficient fraction. Corresponding p values are shown.

(C-D) FGFR mutated group was linked with relative better response to anti-PD-L1 therapy. CR: complete response; PR: partial response; SD: stable disease; PD: progressive disease.

(E-G) External validation of M1 macrophage related DEGs mutation versus wild type in TCGA. FGFR3 (E), TP53 (F), FBXW7 (G), (Mann Whitney U test, p = 1.6e-07, p = 1.5e-06; p = 5e-02, respectively).