**tigeR: Tumor Immunotherapy Gene Expression Data Analysis R package**

Yihao Chen1,2,4,#, Li-Na He2,#, Yuanzhe Zhang2, Yuelong Shu1, Di Zhang3,\*, Guangchuang Yu4,\*, Zhixiang Zuo2,\*

1Key Laboratory of Pathogen infection prevention and control (Peking Union Medical College, Ministry of Education), State Key Laboratory of Respiratory Health and Multimorbidity, National Institute of Pathogen Biology of Chinese Academy of Medical Science (CAMS)/ Peking Union Medical College (PUMC), Beijing 100730, China

2State Key Laboratory of Oncology in South China, Guangdong Provincial Clinical Research Center for Cancer, Collaborative Innovation Center for Cancer Medicine, Sun Yat-sen University Cancer Center, Guangzhou, China.

3Department of Gastrointestinal Surgery, Guangdong Provincial Key Laboratory of Colorectal and Pelvic Floor Diseases, Guangdong Institute of Gastroenterology, The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, China

4Department of Bioinformatics, School of Basic Medical Sciences, Southern Medical University, Guangzhou, China

#These authors contributed equally to this work

\*Correspondence should be addressed to Prof. Zhixiang Zuo (zuozhx@sysucc.org.cn) or Prof. Guangchuang Yu ([gcyu1@smu.edu.cn](mailto:gcyu1@smu.edu.cn)) or Prof. Di Zhang (zhangdi@mail2.sysu.edu.cn)

**Abstract**

Immunotherapy shows great promise for treating advanced cancers, but its effectiveness varies widely among different patients and cancer types. Identifying biomarkers and developing robust predictive models to discern which patients are most likely to benefit from immunotherapy is of great importance. In this context, we have developed the Tumor Immunotherapy Gene Expression R package (tigeR) to address the increasing need for effective tools to explore biomarkers and construct predictive models. tigeR encompasses four distinct yet closely interconnected modules. The Biomarker Evaluation module enables researchers to evaluate whether the biomarkers of interest are associated with immunotherapy response via built-in or custom immunotherapy gene expression data. The Tumor Microenvironment Deconvolution module integrates 10 open-source algorithms to obtain the proportions of different cell types within the tumor microenvironment, facilitating the investigation of the association between immune cell populations and immunotherapy response. The Prediction Model Construction module equips users with the ability to construct sophisticated prediction models using a range of built-in machine learning algorithms. The Response Prediction module predict the immunotherapy response for the patients from gene expression data using our pre-trained machine learning models or public gene expression signatures. By providing these diverse functionalities, tigeR aims to simplify the process of analyzing immunotherapy gene expression data, thus making it accessible to researchers without advanced programming skills. The source code and example for the tigeR project can be accessed at <http://github.com/YuLab-SMU/tigeR> or <http://github.com/canceromics/tigeR>.

Keywords: Tumor Immunotherapy, R package, Machine Learning, Gene Expression, Tumor Microenvironment

**Introduction**

Immunotherapy has emerged as a promising approach in the treatment of cancer patients[1-4], sparking considerable optimism in the medical community by harnessing the body's immune system to target and eliminate tumors[5, 6]. However, the efficacy of immunotherapy remains variable among different individuals and cancer types due to the heterogeneous nature of both cancer types and individual patient[7, 8]. Thus, precise identification of suitable candidates for immunotherapy is crucial in the area of precision medicine, highlighting the critical importance of robust diagnostic models.

Advancements in high-throughput sequencing technologies have revolutionized our ability to characterize the molecular landscape of tumors with unprecedented detail[9, 10]. High-throughput sequencing, particularly in the context of transcriptomics, facilitates the comprehensive analysis of gene expression patterns, identification of tumor biomarkers, and evaluation of the tumor microenvironment[11, 12]. These insights not only aid in unraveling the underlying mechanisms of tumorigenesis and immune evasion but also hold immense potential in guiding the development of personalized diagnostic and therapeutic strategies. By integrating high-throughput sequencing data into precision medicine initiatives, researchers can construct robust diagnostic models that account for the heterogeneity of cancer and pave the way for precision immunotherapy[13-17].

There are numerous web servers, including TIDE[18], TIMER 2.0[19], TIRSF[20], and TIGER[21], that are dedicated to integrating high-throughput expression data to assist users in exploring molecular biomarkers relevant to immunotherapy. While these tools offer convenience, they come with several limitations. Some of these constraints include limited flexibility in customizing analyses, potential concerns regarding data security and privacy when uploading sensitive information to external servers, processing constraints due to reliance on server-side computational resources, as well as the learning curve required to navigate specific interfaces and workflows. These aspects emphasize the importance for users to carefully consider alternative solutions, such as R packages, which do not exhibit the aforementioned limitations.

We therefore developed an R package named tigeR (Tumor Immunotherapy Gene Expression R package), as illustrated in **Figure 1**, to enable users to explore biomarkers associated with immunotherapy response based on built-in or custom immunotherapy gene expression data, investigate the interplay between immune cell populations and treatment response, equip users with the ability to construct sophisticated prediction models using a range of built-in machine learning algorithms, and predict the immunotherapy response for the patients from gene expression data using our pre-trained machine learning models or public gene expression signatures.

**tigeR Workflow**

tigeR comprises four distinct yet interconnected functional modules, as depicted in **Figure 2**. Users have the flexibility to load built-in gene expression data with immunotherapy outcome information or to utilize their own data for subsequent analysis. The Biomarker Evaluation module serves to assess the correlation between biomarkers and immunotherapy outcomes. Furthermore, the Tumor Microenvironment Deconvolution module enables the derivation of cell type proportions within the tumor microenvironment using 10 open-source algorithms[22-31]. This module also provides functionality for evaluating the interplay between fractions of tumor microenvironment cells and immunotherapy outcomes. Subsequently, based on the features selected from these two modules, users can leverage the Prediction Model Construction module, which incorporates a range of machine learning models, to train a model for predicting immunotherapy response using transcriptome gene expression data. Users can use the Response Prediction module to predict the immunotherapy response for the patients from gene expression data using our pre-trained machine learning models or public gene expression signatures.

**Utility and Discussion**

**Evaluating biomarkers associated with immunotherapy response**

Subsequently, we demonstrated the utilization of tigeR for the assessment of biomarkers linked to immunotherapy response. Users have the option to choose one or more immunotherapy datasets from the preloaded datasets or employ their own dataset for the evaluation of specific biomarkers. These biomarkers may encompass a single gene or a set of genes. For a single gene, taking CXCL13 as an example, a well-established biomarker signifying enhanced response to immune checkpoint blockade (ICB) therapy[32-34]. By employing the 'diff\_biomk' function within the Biomarker Evaluation module, we discerned significantly higher expression of CXCL13 in responder samples compared to non-responder samples across datasets such as Melanoma-PRJEB23709[35] and Melanoma-GSE93157[36], which is visualized using boxplots (**Figure 3A, B**). The performance of CXCL13 differentiating the responders and non-responders can also be evaluated using ‘roc\_biomk’ function and visualized using ROC curve (**Figure 3C, D**). Additionally, the 'diff\_biomk' function facilitated the display of differential expression between pre-treatment and post-treatment samples, revealing a marked elevation in CXCL13 expression in post-treatment samples relative to pre-treatment samples in various datasets, which is visualized using boxplots (**Figure 3E, F**). Furthermore, leveraging the 'surv\_biomk' function within the Biomarker Evaluation module, we established an association between elevated CXCL13 expression and improved overall survival in patients undergoing ICB therapy, as evidenced in multiple datasets, which is visualized using Kaplan-Meier curve (**Figure 3G-I**). Notably, these findings align with existing literature, substantiating CXCL13 as a robust biomarker indicative of favorable response to ICB therapy.

For a gene set, users can firstly use function ‘score\_biomk’ with various algorithms such as “average mean”, “weighted mean” and “GSVA” to calculate the gene set score. Then, users can evaluate the correlation between the gene set of interest and immunotherapy response using the above functions for a single gene. For instance, we evaluated the performance of the gene set score of tertiary lymphoid structure (TLS) signature[37], known to be associated with favorable immunotherapy response, on the predicting immunotherapy response using public dataset GSE145996[38], and found TLS gene set score could differentiate between responder group and non-responder group well (**Figure 3J, K**), and patients with higher TLS gene set score have a better overall survival (**Figure 3L**).

Moreover, we have collected 23 immunotherapy response related biomarkers from literature and integrated these biomarkers into our R package (**Supplemental Table 1**). Users can compare the biomarkers of interest with these existed biomarkers using function ‘compare\_biomk’ to get a rough sense of the performance of the new biomarkers (**Figure 3M**).

**Identifying tumor microenvironment compositions associated with immunotherapy response**

Tumor microenvironment analysis is crucial for understanding tumor immune evasion mechanisms and predicting the efficacy of immunotherapy. It provides insights into the interactions between tumor cells, immune cells, and stromal components that influence the anti-tumor immune response and treatment outcomes. We implemented 10 open source tumor microenvironment deconvolution methods including CIBERSORT[22], TIMER[23], ESTIMATE[24], IPS[25], xCell[26], EPIC[27], ConsensusTME[28], ABIS[29], quanTIseq[30], and MCPCounter[31] in tigeR. Users can use function ‘deconv\_TME’ to derive the proportions of different tumor microenvironment cell types from gene expression data with these tools. The cell type proportions derived from melanoma dataset MEL\_GSE78220[39] using function ‘deconv\_TME’ were shown as an example (**Figure 4A**). Users can further calculate the correlations between TME cell types and immunotherapy response using functions ‘roc\_biomk’ and 'surv\_biomk' (**Figure 4B-K**). We found that in melanoma dataset MEL\_GSE91061[40], cell type such as naïve B cell was significantly positively correlated with overall survival (**Figure 4L**), and cell type such as Monocytes was significantly negatively correlated with overall survival (**Figure 4M**).

**Constructing immunotherapy response prediction model**

Automating the construction of machine learning models can accelerate the discovery of personalized tumor immunotherapy strategies by empowering clinical experts without programming skills to efficiently handle complex biological data. tigeR enable users to build machine learning models with a simple function ‘build\_Model’. This function allow user to build response prediction model based on gene expression data with various machine learning algorithms such as Naive Bayes, Random Forest, Support Vector Machine, CancerClass[41], Adaboost, Logitboost, and Logistics Regression. Furthermore, users can evaluate the trained model using test gene expression dataset by ‘test\_Model’ function.

We then proceeded with the workflow outlined in **Figure 5A** to demonstrate the utilization of functions in the Construct Prediction Model module for constructing a robust immunotherapy response prediction model. Gene expression data from 275 pre-treatment samples across three Melanoma cohorts treated with anti-PD-1 or anti-CTLA-4 antibody (MEL\_GSE91061, MEL\_phs000452, MEL\_Nathanson\_2017)[42-44] was utilized. Subsequently, the samples were randomly divided into a training set (n=187 and a testing set (n=88). The TIMER algorithm[23] was employed to deconvolute the proportions of TME cell types. Features used for training the machine learning models included 72 genes with variance larger than 0.08, 23 existing biomarkers, and the TME cell type proportions. Six different machine learning algorithms implemented in the 'build\_Model' function were used to train the prediction models (**Figure 5B**). The performance of each model was evaluated using the 'test\_Model' function with the testing set, revealing that the Random Forest model achieved the highest AUC value of 0.98 (**Figure 5C**). We next applied this Random Forest model to one independent melanoma dataset MEL\_GSE78220[39], which demonstrates promising performance (AUC=0.7944) (**Figure 5D**). Additionally, the Kaplan-Meier plot indicated significantly different overall survival between patients categorized as "High Risk" and "Low Risk" by the Random Forest model (**Figure 5E**, *P* = 0.022). We implemented a function ‘pred\_response’ to allow users predict whether a patient will respond to immunotherapy based on our newly constructed prediction model and public signatures, as demonstrated using another independent melanoma data MEL\_GSE93157[36] (**Figure 5F**).

**Conclusion**

In conclusion, we presented an R package, tigeR (Tumor Immunotherapy Gene Expression R package) to advance the field of cancer immunotherapy research. With its comprehensive suite of functionalities, tigeR empowers users to delve into the intricacies of immunotherapy response by exploring biomarkers, dissecting the dynamics between immune cell populations and treatment outcomes, and constructing predictive models using state-of-the-art machine learning algorithms. By providing access to both built-in and customizable immunotherapy gene expression data, tigeR facilitates a deeper understanding of the molecular mechanisms underlying treatment response. tigeR not only streamlines the analysis process but also catalyzes discoveries in the realm of tumor immunotherapy, ultimately contributing to advancements in patient care and outcomes. As the field continues to evolve, tigeR stands poised to remain at the forefront, driving innovation and insights in the pursuit of enhanced cancer treatments.

**Data and code Availability**

The tigeR package is open source and freely available on github (<http://github.com/YuLab-SMU/tigeR> or <http://github.com/canceromics/tigeR>). The codes that are used to generate the figures in the main text are available in the Supplementary Information file. The tigeR package includes 1060 samples with immunotherapy clinical information from a total of 11 melanoma datasets, 3 lung cancer datasets, 2 kidney cancer datasets, 1 gastric cancer dataset, 1 low-grade glioma dataset, 1 glioblastoma dataset and 1 Head and Neck Squamous data set (**Supplementary Table 2**). These datasets were organized into R language ‘SummarizedExperiment’ objects[45] and can be downloaded directly into R environment using ‘Dataloader’ function in tigeR. Users can choose to download the internal data set in tigeR from either the bioConductor ExperimentHub (https://github.com/Bioconductor/ExperimentHub) or the TIGER web server[21] (http://tiger.canceromics.org) using the Dataloader function in the tigeR package. The built-in datasets in the tigeR package are constructed in the form of ‘SummarizedExperiment’ objects, generated with the ‘SummarizedExperiment’ package in R[45]. These objects contain expression matrix data, which has been normalized to fragments per kilobase of transcript per million (FPKM) values, as well as clinical information specifically related to tumor immunotherapy patients.

**Competing interests**

The authors declare no competing interests.

**Funding**

This study was supported by the National Key R&D Program of China [2023YFF1204600]; Guangdong Basic and Applied Basic Research Foundation [2021B1515020108]; Guangdong Esophageal Cancer Institute Science and Technology Program [M202206];

**Author Contributions**

Z.Z., G.Y. and D.Z. conceptualized and supervised the research. Y.C., L.N.H. and Y.Z. curated the database and performed data analysis. Y.C. developed the web interface. Y.C. and Y.S. prepared the figures. Y.C. and L.N.H. drafted the manuscript. Z.Z. reviewed the manuscript. All authors read and approved the final manuscript.

**References**

1. Wang, Y., et al., *Advancing to the era of cancer immunotherapy.* Cancer Commun (Lond), 2021. **41**(9): p. 803-829.

2. Eggermont, A.M.M., et al., *Adjuvant Pembrolizumab versus Placebo in Resected Stage III Melanoma.* N Engl J Med, 2018. **378**(19): p. 1789-1801.

3. Hodi, F.S., et al., *Improved survival with ipilimumab in patients with metastatic melanoma.* N Engl J Med, 2010. **363**(8): p. 711-23.

4. Vaddepally, R.K., et al., *Review of Indications of FDA-Approved Immune Checkpoint Inhibitors per NCCN Guidelines with the Level of Evidence.* Cancers (Basel), 2020. **12**(3).

5. Wang, Y., et al., *Advancing to the era of cancer immunotherapy.* Cancer Communications, 2021. **41**(9): p. 803-829.

6. van Weverwijk, A. and K.E. de Visser, *Mechanisms driving the immunoregulatory function of cancer cells.* Nat Rev Cancer, 2023. **23**(4): p. 193-215.

7. Sharma, P., et al., *The Next Decade of Immune Checkpoint Therapy.* Cancer Discovery, 2021. **11**(4): p. 838-857.

8. Jia, Q., et al., *Heterogeneity of the tumor immune microenvironment and its clinical relevance.* Exp Hematol Oncol, 2022. **11**(1): p. 24.

9. Reuter, J.A., D.V. Spacek, and M.P. Snyder, *High-throughput sequencing technologies.* Mol Cell, 2015. **58**(4): p. 586-97.

10. Stark, R., M. Grzelak, and J. Hadfield, *RNA sequencing: the teenage years.* Nat Rev Genet, 2019. **20**(11): p. 631-656.

11. Hong, M., et al., *RNA sequencing: new technologies and applications in cancer research.* J Hematol Oncol, 2020. **13**(1): p. 166.

12. Gentles, A.J., et al., *The prognostic landscape of genes and infiltrating immune cells across human cancers.* Nat Med, 2015. **21**(8): p. 938-945.

13. Li, B., et al., *Multiomics identifies metabolic subtypes based on fatty acid degradation allocating personalized treatment in hepatocellular carcinoma.* Hepatology, 2024. **79**(2): p. 289-306.

14. Liu, W., et al., *Large-scale and high-resolution mass spectrometry-based proteomics profiling defines molecular subtypes of esophageal cancer for therapeutic targeting.* Nat Commun, 2021. **12**(1): p. 4961.

15. Jin, X., et al., *Molecular classification of hormone receptor-positive HER2-negative breast cancer.* Nat Genet, 2023. **55**(10): p. 1696-1708.

16. Katipally, R.R., et al., *Integrated Clinical-Molecular Classification of Colorectal Liver Metastases: A Biomarker Analysis of the Phase 3 New EPOC Randomized Clinical Trial.* JAMA Oncol, 2023. **9**(9): p. 1245-1254.

17. Lin, Y., et al., *Immune cell infiltration signatures identified molecular subtypes and underlying mechanisms in gastric cancer.* NPJ Genom Med, 2021. **6**(1): p. 83.

18. Jiang, P., et al., *Signatures of T-cell dysfunction and exclusion predict cancer immunotherapy response.* Cancer Immunology Research, 2019. **7**(2): p. 2.

19. Li, T.W., et al., *TIMER2.0 for analysis of tumor-infiltrating immune cells.* Nucleic Acids Research, 2020. **48**(W1): p. W509-W514.

20. Chen, L., et al., *TIRSF: a web server for screening gene signatures to predict Tumor immunotherapy response.* Nucleic Acids Research, 2022. **50**(W1): p. W761-W767.

21. Chen, Z., et al., *TIGER: A Web Portal of Tumor Immunotherapy Gene Expression Resource.* Genomics, Proteomics & Bioinformatics, 2023. **21**(2): p. 337-348.

22. Newman, A.M., et al., *Robust enumeration of cell subsets from tissue expression profiles.* Nature Methods, 2015. **12**(5): p. 453-+.

23. Li, B., J.S. Liu, and X.S. Liu, *Revisit linear regression-based deconvolution methods for tumor gene expression data.* Genome Biology, 2017. **18**: p. 5.

24. Yoshihara, K., et al., *Inferring tumour purity and stromal and immune cell admixture from expression data.* Nature Communications, 2013. **4**: p. 11.

25. Charoentong, P., et al., *Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade.* Cell Reports, 2017. **18**(1): p. 248-262.

26. Aran, D., Z.C. Hu, and A.J. Butte, *xCell: digitally portraying the tissue cellular heterogeneity landscape.* Genome Biology, 2017. **18**: p. 14.

27. Racle, J., et al., *Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data.* Elife, 2017. **6**: p. 25.

28. Jiménez-Sánchez, A., O. Cast, and M.L. Miller, *Comprehensive Benchmarking and Integration of Tumor Microenvironment Cell Estimation Methods.* Cancer Research, 2019. **79**(24): p. 6238-6246.

29. Monaco, G., et al., *RNA-Seq Signatures Normalized by mRNA Abundance Allow Absolute Deconvolution of Human Immune Cell Types.* Cell Reports, 2019. **26**(6): p. 1627-+.

30. Finotello, F., et al., *Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data (vol 11, 34, 2019).* Genome Medicine, 2019. **11**: p. 3.

31. Becht, E., et al., *Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression (vol 17, 218, 2016).* Genome Biology, 2016. **17**: p. 2.

32. Liu, B., et al., *Single-cell meta-analyses reveal responses of tumor-reactive CXCL13(+) T cells to immune-checkpoint blockade.* Nat Cancer, 2022. **3**(9): p. 1123-1136.

33. Yang, M., et al., *CXCL13 shapes immunoactive tumor microenvironment and enhances the efficacy of PD-1 checkpoint blockade in high-grade serous ovarian cancer.* J Immunother Cancer, 2021. **9**(1).

34. Cohen, M., et al., *The interaction of CD4(+) helper T cells with dendritic cells shapes the tumor microenvironment and immune checkpoint blockade response.* Nat Cancer, 2022. **3**(3): p. 303-317.

35. Gide, T.N., et al., *Distinct Immune Cell Populations Define Response to Anti-PD-1 Monotherapy and Anti-PD-1/Anti-CTLA-4 Combined Therapy.* Cancer Cell, 2019. **35**(2): p. 238-255.e6.

36. Prat, A., et al., *Immune-Related Gene Expression Profiling After PD-1 Blockade in Non-Small Cell Lung Carcinoma, Head and Neck Squamous Cell Carcinoma, and Melanoma.* Cancer Res, 2017. **77**(13): p. 3540-3550.

37. Cabrita, R., et al., *Tertiary lymphoid structures improve immunotherapy and survival in melanoma (vol 577, pg 561, 2020).* Nature, 2020. **580**(7801): p. E1-E1.

38. Amato, C.M., et al., *Pre-Treatment Mutational and Transcriptomic Landscape of Responding Metastatic Melanoma Patients to Anti-PD1 Immunotherapy.* Cancers (Basel), 2020. **12**(7).

39. Hugo, W., et al., *Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma.* Cell, 2016. **165**(1): p. 35-44.

40. Riaz, N., et al., *Tumor and Microenvironment Evolution during Immunotherapy with Nivolumab.* Cell, 2017. **171**(4): p. 934-949.e16.

41. Budczies, J., et al., *cancerclass: An R Package for Development and Validation of Diagnostic Tests from High-Dimensional Molecular Data.* Journal of Statistical Software, 2014. **59**(1): p. 1-19.

42. Riaz, N., et al., *Tumor and Microenvironment Evolution during Immunotherapy with Nivolumab.* Cell, 2017. **171**(4): p. 934-+.

43. Van Allen, E.M., et al., *Genomic correlates of response to CTLA-4 blockade in metastatic melanoma.* Science, 2015. **350**(6257): p. 207-211.

44. Nathanson, T., et al., *Somatic Mutations and Neoepitope Homology in Melanomas Treated with CTLA-4 Blockade.* Cancer Immunology Research, 2017. **5**(1): p. 84-91.

45. Huber, W., et al., *Orchestrating high-throughput genomic analysis with Bioconductor.* Nature Methods, 2015. **12**(2): p. 115-121.

**Figure legends**

**Figure 1 Overall design of tigeR package.** The tigeR R package contains four functional modules, namely ‘Biomarker Evaluation Module’, ‘Tumor Microenvironment Deconvolution Module’, ‘Prediction Model Construction Module’ and ‘Response Prediction Module’.

**Figure 2 The workflow diagram illustrates the four distinct yet interconnected functional modules within tigeR package.**

**Figure 3** **The utility of Biomarker Evaluation Module. (A, B)** The CXCL13 gene expression difference between responder and non-responder in Melanoma-PRJEB23709 dataset (A) and Melanoma-GSE93157 dataset (B). **(C, D)** The ROC plots indicating the performance of CXCL13 gene expression on differentiating responder and non-responder in Melanoma-PRJEB23709 dataset (C) and Melanoma-GSE93157 dataset (D). **(E, F)** The CXCL13 gene expression difference between pre-treatment and post-treatment in Melanoma-GSE115821 dataset (E) and Melanoma-GSE91061 dataset (F). **(G-I)** The Kaplan‑Meier plots showing the survival difference between patients with high and low CXCL13 gene expression in MEL-PRJEB23709 dataset (G), MEL-Nathanson\_2017 dataset (H), and RCC-Braun\_2020 dataset (I). **(J)** The TLS signature difference between responder and non-responder in Melanoma-PRJEB23709 dataset. **(K)** The ROC plot indicating the performance of TLS signature on differentiating the responder and non-responder in Melanoma-PRJEB23709 dataset. **(L)** The Kaplan‑Meier plot showing the survival difference between patients with high and low TLS signature score in MEL-PRJEB23709 dataset. **(M)** The heatmap displaying the AUC of 23 built-in signatures and the custom signature in classifying responders and non-responders in 10 built-in immunotherapy datasets.

**Figure 4 Tumor microenvironment compositions associated with immunotherapy response. (A)** The deconvolution of tumor microenvironment compositions from gene expression data using TIMER, CIBERSORT, MCPCounter, xCell, EPIC, ABIS, ConsensusTME, quanTIseq, ESTIMATE, and IPS, respectively. **(B-K)** The association between tumor microenvironment compositions derived from the above 10 tools and immunotherapy response. Left panel is a barplot showing the AUC classifying responders and non-responders for each cell type. Right panel is showing the survival correlations for each cell type. **(L)** The Kaplan‑Meier plot showing the survival difference between patients with high and low fractions of naïve B cells. **(M)** The Kaplan‑Meier plot showing the survival difference between patients with high and low fractions of Monocytes cells.

**Figure 5** **Construction and utility of immunotherapy response prediction model. (A)** Workflow for training immunotherapy response prediction model using built-in functions in tigeR. **(B)** Comparison of performance of six machine learning algorithms on predicting immunotherapy response in training sets. **(C)** Comparison of performance of six machine learning algorithms on predicting immunotherapy response in testing sets. **(D)** ROC plot showing the performance of our newly constructed Random Forest model in an independent dataset MEL-GSE78220. **(E)** TheKaplan‑Meier plot showing the survival difference between patients in High and Low groups according to the Random Forest model. (**F**) Heatmap showing the results of using ‘Response Prediction Moduel’ to predict the immunotherapy response for the patients from gene expression data using our pre-trained machine learning model or public gene expression signatures.