

# A Pan-Cancer Immunogenomic Atlas for Immune Checkpoint Blockade Immunotherapy

Jing Yang<sup>1,2</sup>, Shilin Zhao<sup>1,2</sup>, Jing Wang<sup>1,2</sup>, Quanhu Sheng<sup>1,2</sup>, Qi Liu<sup>1,2</sup>, and Yu Shyr<sup>1,2</sup>



## ABSTRACT

The ability to identify robust genomic signatures that predict response to immune checkpoint blockade is restricted by limited sample sizes and ungeneralizable performance across cohorts. To address these challenges, we established Cancer-Immu (<http://bioinfo.vanderbilt.edu/database/Cancer-Immu/>), a comprehensive platform that integrates large-scale multidimensional omics data, including genetic, bulk, and single-cell transcriptomic, proteomic, and dynamic genomic profiles, with clinical phenotypes to explore consistent and rare

immunogenomic connections. Currently Cancer-Immu has incorporated data for 3,652 samples for 16 cancer types. It provides easy access to immunogenomic data and empowers researchers to translate omics datasets into biological insights and clinical applications.

**Significance:** Cancer-Immu is a comprehensive functional portal for unraveling immune-genomic connections to improve immune checkpoint blockade-based cancer immunotherapy.

## Introduction

A low response rate is the major challenge for existing immune checkpoint blockade (ICB)-based cancer immunotherapies. Fortunately, the research community has identified some factors that influence response outcomes to ICB therapy. For example, *PD-L1* expression (1), tumor mutational burden (TMB; refs. 2, 3), and tumor infiltration lymphocytes (4, 5) have been reported as potential predictive biomarkers. However, most published signatures were only evaluated on limited cohorts and showed ungeneralizable performance across different cohorts. Therefore, it is still a lack and a challenge for evaluation of efficacy of known biomarkers and the discovery of new signature(s) in a large-scale study. Here we developed Cancer-Immu (<http://bioinfo.vanderbilt.edu/database/Cancer-Immu/>) to prioritize genomic features associated with ICB response, containing 3,652 samples for 16 cancer types. Cancer-Immu provides two analysis strategies, meta-analysis and pan-cancer analysis. Meta-analysis reveals consistent signatures across multiple tumor/study cohorts, while pan-cancer analysis enhances our ability to detect and analyze rare features by aggregating samples across cohorts/tumor types. Furthermore, Cancer-Immu enables linking dynamic features with immunotherapy response. Cancer-Immu also allows users to upload and analyze their own data independently or to co-analyze with existing data simultaneously. Cancer-Immu provides a comprehensive resource for identifying predictive genomic features of

immunotherapy response, and presents an easy way for signature prioritization, known biomarker assessment, novel signature discovery, and signature-of-interest validation.

## Materials and Methods

### Data collection

We collected publicly available datasets with both genomic profiling and ICB therapy outcomes, including 3,652 samples across 16 cancer types (Supplementary Table S1). The genomic profiles were generated by either bulk DNA/RNA-sequencing, protein arrays, single-cell mass cytometry, or single-cell RNA sequencing. The definition of responder and nonresponder was obtained from original studies, if provided. Otherwise, the definition was based on the RECIST, which stratifies the treatment effect into complete response (CR), partial response (PR), stable disease (SD), and progressive disease (PD). Patients with CR and PR were considered as responders, while those with SD and PD were defined as nonresponders. Patients without RECIST were assigned as “NA.”

### Multiomics features

Cancer-Immu collected three types of omics data, genetic, transcriptomics, and single-cell data. Genetic data consist of gene mutations, TMB, and mutational signatures. Gene mutations and TMB were obtained from the original publication/study or from public databases. Mutational signatures, characteristic combinations of mutation types as defined in the literature (6), were calculated using the “deconstructSigs” R package (v1.6.0). Briefly, combinations of known mutational signatures were used to determine the observed mutational profile for each sample (7). Only somatic mutations in exome regions were included, and trinucleotide counts were normalized by the number of times each trinucleotide context was observed in the exome region.

Transcriptomic data comprise four types of features: gene expression, expression sum of gene sets, gene expression relation pairs, and immune cell components. Gene expression values were obtained from the original publication/study or public databases. Expression sum focuses on the sum of a list of genes, which is a powerful and simplified way to explore gene sets or pathways of interest. Cancer-Immu allows users to choose predefined gene sets or to define their own gene sets. Gene expression relation pairs is a summed score of relative

<sup>1</sup>Center for Quantitative Sciences, Vanderbilt University Medical Center, Nashville, Tennessee. <sup>2</sup>Department of Biostatistics, Vanderbilt University Medical Center, Nashville, Tennessee.

**Note:** Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org/>).

**Corresponding Authors:** Qi Liu, Department of Biostatistics and Center for Quantitative Sciences, Vanderbilt University Medical Center, Nashville, TN 37232. Phone: 615-322-6618; E-mail: qi.liu@vanderbilt.edu; and Yu Shyr, yu.shyr@vanderbilt.edu

Cancer Res 2022;82:539–42

doi: 10.1158/0008-5472.CAN-21-2335

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 International (CC BY-NC-ND).

©2021 The Authors; Published by the American Association for Cancer Research

comparison of gene pairs, which was reported to be predictive of ICB response (8). For  $m$  gene pairs,  $G_1$  and  $G_1'$ ,  $G_2$  and  $G_2'$ , ...,  $G_m$  and  $G_m'$ , the score is calculated by comparing and summing the expression of each pair. If  $G_1 > G_1'$ , this gene pair returns a score of 1, otherwise a score of 0. Immune cell components were generated from gene expression profiles using CIBERSORT (9, 10), which estimated a relative percentage for each immune cell type in each sample.

Single-cell data support two types of features, gene-cell expression and cell populations. Gene expression in each cell was downloaded from the original studies. Single-cell data were processed and cell populations were identified by "Seurat" R package v3.0.1 (11).

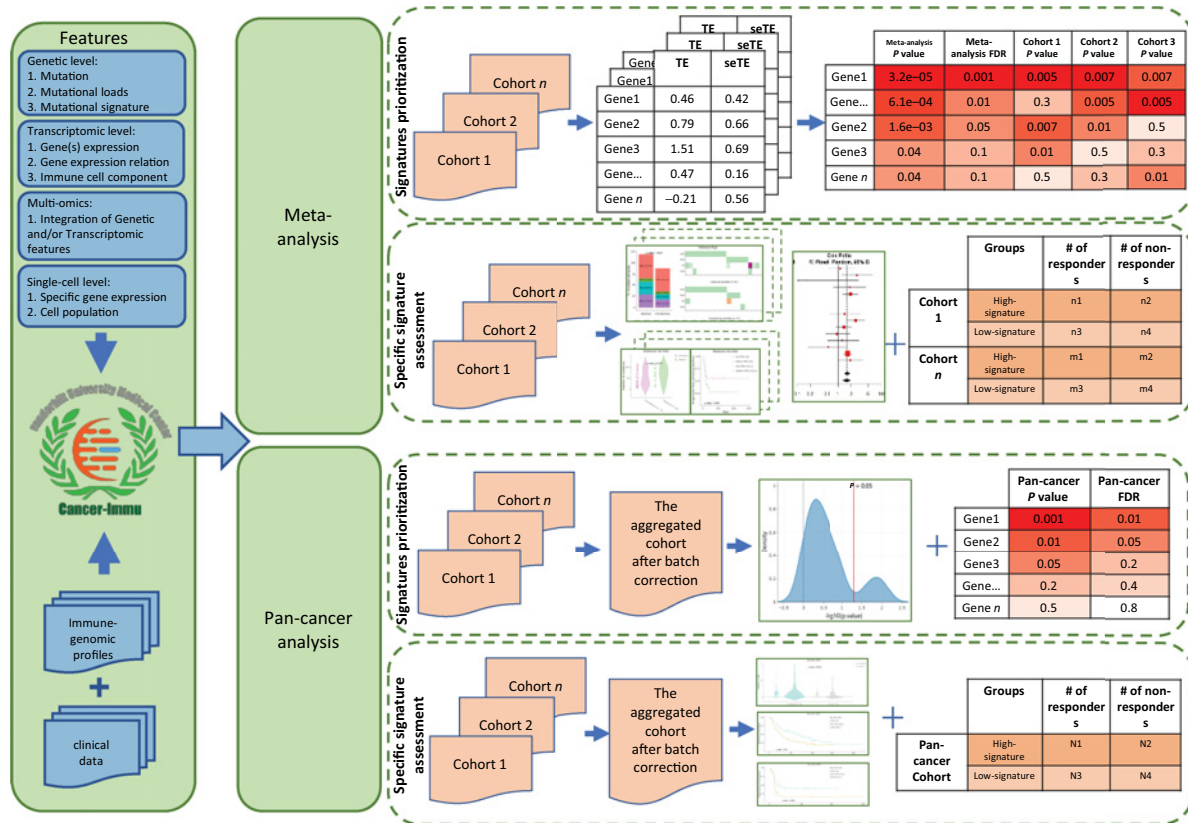
**Meta-analysis**

A random effects meta-analysis model was used to combine results from multiple studies. Each feature was first evaluated individually in each study, where logistic regression was used to estimate its association with ICB response (12, 13). Then the effect size of each signature, measured as the  $\log_2$  OR for responder versus nonresponder, and SEs were combined through random effects meta-analysis to generate an overall effect size and  $P$  value.  $P$  values were adjusted by the Benjamini and Hochberg procedure. To make all signatures with varying measurement scales to be compared equivalently, signature

values in each individual cohort were converted to standard z-scores with mean of zero and SD of one. Meta-analysis was performed using the "meta" R package.

**Pan-cancer analysis**

Compared with meta-analysis combining results from multiple studies, pan-cancer analysis pools multiple datasets into one large dataset and then evaluates the association. Genomic features, such as mutations, mutation signatures and immune cell components, have negligible batch effects across datasets, therefore they were combined directly without any further processing. For features with strong batch effects across datasets, such as gene expression, batch effect correction was performed before aggregation. Batch effects were removed by the removeBatchEffect in the "limma" R package along with a design matrix that preserved the response effect. We demonstrated that batch-corrected data were appropriate for ICB studies, which not only removed batch effects successfully, but also recapitulated ICB-related signatures that were identified in individual datasets (Supplementary Fig. S1). To be noted, because the batch correction method only took the response effect into account, additional cautions should be paid when applied to other studies where unconsidered covariates might violate the assumption, leading to undercorrection or overcorrection.



**Figure 1.**

Framework of Cancer-Immu. Cancer-Immu explores the associations of 10 types of features with clinical outcome (ICB responsiveness, overall survival, and progression-free survival) via meta-analysis and pan-cancer analysis. Each analysis includes two functions, signatures prioritization and specific signature assessment. Signatures prioritization screens all features and ranks them based on statistical significance, while specific signature assessment provides a detailed view into one specific feature. In meta-analysis, signatures are ranked on the basis of the consensus of statistical significance of the associations between signatures and ICB response across multiple cohorts. Pan-cancer analysis first aggregates samples into one dataset, then signatures are ranked on the basis of the statistical significance of the associations in the aggregated dataset.

After aggregation, the association of each individual feature with ICB response was tested in the pooled dataset and *P* values were adjusted by the Benjamini and Hochberg procedure.

### Uploading function

Cancer-Immu allows users to upload their own datasets, which can be analyzed independently or co-analyzed with existing datasets. Uploaded data require clinical outcome and at least one type of omics data: genetic, transcriptomic, or single-cell profiles. Genetic data that describe somatic mutations should include chromosome, mutational position, reference base, alternative base, mutational type (such as Missense\_Mutation, Nonsense\_Mutation, and so on), and gene symbol. Mutational position should be reported based on the UCSC hg19 assembly (i.e., GRCh37). Transcriptomic data are normalized and log<sub>2</sub> transformed. Single-cell data require a gene-cell matrix with raw counts. After uploading, mutation loads/burden are calculated and mutational signatures are detected from mutation data by the “deconstructSigs” R package. Immune cell components are derived from transcriptomic data using CIBERSORT. Batch correction is performed for the pan-cancer analysis. Single-cell data are normalized and cell population is identified by the “Seurat” R package.

### Data availability statement

All the data can be browsed and downloaded on <http://bioinfo.vanderbilt.edu/database/Cancer-Immu/>. All the code can be downloaded from <https://github.com/JingYangSciBio/Cancer-Immu>.

## Results

### Architecture

Cancer-Immu explores the associations of multiomics features with immunotherapy response, which are derived from genetic, transcriptomic, and single cell (Fig. 1, left). Cancer-Immu provides meta-analysis and pan-cancer analysis for signature prioritization and specific signature assessment (Fig. 1, right). In addition, an uploading module is included for analysing user’s own datasets.

### Comparison with existing databases

Compared with existing databases with ICB response outcome, Cancer-Immu covers the greatest number of datasets and omics data types and supports a variety of functions (Table 1). In addition,

Cancer-Immu implements two unique functions. One is to explore those complicated signatures reported to be associated with ICB response, such as dynamic expression changes before and after treatment, integrative genetic and/or transcriptomic features, expression sum of pathways or user-defined gene sets and etc. The other is to provide systematic analysis across multiple studies for signatures prioritization.

### Signatures prioritization in the meta-analysis module

In the meta-analysis module, signatures are prioritized on the basis of their overall associations with immunotherapy response across multiple cohorts using random-effects meta-analysis. The OR of each genomic feature predictive of ICB response is evaluated in each individual cohort and subsequently combined to derive an overall OR and value across cohorts. The current analysis provided the signature prioritization for all 3,652 samples among 16 tumor types to investigate the robustness and generality of known and novel response-related biomarkers (Supplementary Fig. S2; Supplementary Table S2).

To analyze the immunogenomic data stratified by drug type or treatment time, Cancer-Immu provided meta-analysis results across samples with anti-PD-1/L1 treatment, samples with anti-CTLA-4 treatment, samples before treatment, and samples after treatment.

### Specific signature assessment in the meta-analysis module

Compared with signature prioritization, which screens and ranks all features simultaneously, specific signature assessment provides a more detailed view of each feature individually. In addition to immunotherapy response, the associations with overall survival and progression-free survival are also evaluated in the module. We used TMB as an example to show its associations with ICB responsiveness, overall survival, and progression-free survival (Supplementary Fig. S3).

### Signatures prioritization and specific signature assessment in the pan-cancer analysis module

When there are limited samples and few events, for example, rare mutations, it is difficult to conduct meaningful association studies using meta-analysis. Pan-cancer analysis aggregates multiple datasets into one, which enhances detection and analysis of rare features. Whereas the meta-analysis across all 3,652 samples failed to detect any significant gene mutations, the pan-analysis detected 182 genes whose mutations were significantly associated with ICB response

**Table 1.** Comparison between Cancer-Immu and existing databases.

		Cancer-Immu	TIDE	TCIA	CRI-iAtlas	cBioportal
Data	No. of patients with ICB	3,652	1,247	328 <sup>a</sup>	636	2,040
	Genetic data	✓	—	✓	—	✓
	Transcriptomic data	✓	✓	✓	✓	✓
	Single-cell data	✓	—	—	—	—
	Users’ data	✓	✓	✓	—	—
Function	ICB response association	✓	—	—	✓	✓
	ICB response prediction	—	✓	✓	—	—
	ICB-related features <sup>b</sup>	✓	—	—	—	—
	Survival association	✓	✓	✓	✓	✓
	Meta-analysis	✓	—	—	—	—
	Pan-cancer analysis	✓	—	—	—	✓

<sup>a</sup>Melanoma only.

<sup>b</sup>Complicated features that have been reported to be associated with ICB response.

(Supplementary Table S3). We used *DGKZ* mutations as an example to show how pan-cancer analysis revealed signatures that were masked by cohort-dependent noise and thus could not be identified in small scale studies (Supplementary Fig. S4; Supplementary Table S4).

### Exploration of dynamic features

Recent studies have highlighted that “time-frozen snapshots” of biomarkers cannot reflect dynamic immune reactivity and thus have low diagnostic potential (14). Compared with static signature, the dynamic profiles of signature have more potential to optimize patient selection and extend the diagnostic utility. Cancer-Immu not only enables users to check dynamic expression for a single gene but also provides an interactive volcano plot to view all genes or all immune cell types at once. Using altered expression of *MAPK4* and *MET*, which promote tumor progression by stimulating PI3K/AKT pathway as examples (Supplementary Fig. S5), we demonstrated that dynamic features provide a different and powerful way to explore the tumor-immune interactions. With the ability of exploring dynamic features from data generated in multiple timepoints, Cancer-Immu is expected to stimulate interest in dynamic interactions between immune and cancer cells.

## Discussion

Cancer-Immu is by far the largest ICB-related data portal to explore consistent and rare immunogenomic connections. Equipped with two analysis modules, Cancer-Immu not only makes it easy to reproduce and validate previous findings, such as TMB and *PD-L1* expression,

but also greatly facilitates the discovery of novel signatures, such as *DGKZ* mutations and dynamic expression of stimulators of PI3K/AKT pathway. As immunogenomic data become increasingly available, Cancer-Immu has the potential to develop a predictive model by integrating multiple biomarkers. One limitation of this study is potential existence of confounding factors that we were unable to control due to availability of limited clinical data.

### Authors' Disclosures

Y. Shyr reports grants from NIH/NCI during the conduct of the study. No disclosures were reported by the other authors.

### Authors' Contributions

**J. Yang:** Conceptualization, resources, data curation, software, formal analysis, visualization, methodology, writing—original draft. **S. Zhao:** Software. **J. Wang:** Software. **Q. Sheng:** Software. **Q. Liu:** Conceptualization, supervision, validation, investigation, methodology, writing—review and editing. **Y. Shyr:** Conceptualization, resources, supervision, funding acquisition, methodology, writing—review and editing.

### Acknowledgments

This work was supported by the NCI (SPORE in Gastrointestinal Cancer 5P50CA236733-02 to Y. Shyr; SPORE in Breast Cancer 5P50CA098131-18 to Y. Shyr) and Cancer Center Support Grant (2P30 CA068485-24 to Y. Shyr). Funding for open access charge: NCI (2P30 CA068485-24 to Y. Shyr).

We acknowledge all the users and reviewers for their testing the web and giving valuable feedbacks. We thank Bryan Helm for correcting English language.

Received July 19, 2021; revised October 19, 2021; accepted December 2, 2021; published first December 13, 2021.

## References

1. Khunger M, Rakshit S, Pasupuleti V, Hernandez AV, Mazzone P, Stevenson J, et al. Incidence of pneumonitis with use of programmed death 1 and programmed death-ligand 1 inhibitors in non-small cell lung cancer: a systematic review and meta-analysis of trials. *Chest* 2017;152:271–81.
2. Cristescu R, Mogg R, Ayers M, Albright A, Murphy E, Yearley J, et al. Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy. *Science* 2018;362:eaar3593.
3. Samstein RM, Lee CH, Shoushtari AN, Hellmann MD, Shen R, Janjigian YY, et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat Genet* 2019;51:202–6.
4. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, et al. TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res* 2017;77:e108–e10.
5. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med* 2018;24:1550–8.
6. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. *Nature* 2013;500:415–21.
7. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol* 2016;17:31.
8. Auslander N, Zhang G, Lee JS, Frederick DT, Miao B, Moll T, et al. Robust prediction of response to immune checkpoint blockade therapy in metastatic melanoma. *Nat Med* 2018;24:1545–9.
9. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* 2015;12:453–7.
10. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling tumor infiltrating immune cells with CIBERSORT. *Methods Mol Biol* 2018;1711:243–59.
11. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* 2018;36:411–20.
12. Litchfield K, Reading JL, Puttick C, Thakkar K, Abbosh C, Bentham R, et al. Meta-analysis of tumor- and T cell-intrinsic mechanisms of sensitization to checkpoint inhibition. *Cell* 2021;184:596–614.
13. Vokes NI, Liu D, Ricciuti B, Jimenez-Aguilar E, Rizvi H, Dietlein F, et al. Harmonization of tumor mutational burden quantification and association with response to immune checkpoint blockade in non-small-cell lung cancer. *JCO Precis Oncol* 2019;3:PO.19.00171.
14. Liu C, He H, Li X, Su MA, Cao Y. Dynamic metrics-based biomarkers to predict responders to anti-PD-1 immunotherapy. *Br J Cancer* 2019;120:346–55.