



# Tumor-resilient T-cell model

Peng Jiang ([peng.jiang@nih.gov](mailto:peng.jiang@nih.gov))

Cancer Data Science Lab, National Cancer Institute, USA

## Table of Content

<b>Introduction</b>	<b>1</b>
<b>Usage 1: Query genes</b>	<b>2</b>
Result Tables	2
Significance analysis.	4
Level 1: significance of group values	5
1A: query a single gene	5
1B: query a target gene versus control	5
Level 2: significance of individual values	5
2B: Tres scores	5
2C: Response	6
2D: Screen	7
Lineage expression	7
<b>Usage 2: Predict tumor-resilience of T cells</b>	<b>9</b>
Input data format	10
Result format	10
<b>Usage 3: Prioritize marker genes of tumor-resilient T cells</b>	<b>11</b>
<b>Usage 4: Evaluate the biomarker value of a gene signature</b>	<b>13</b>
4A, Response	13
4B, Survival	14
<b>Usages not available on Webapp</b>	<b>15</b>

## Introduction

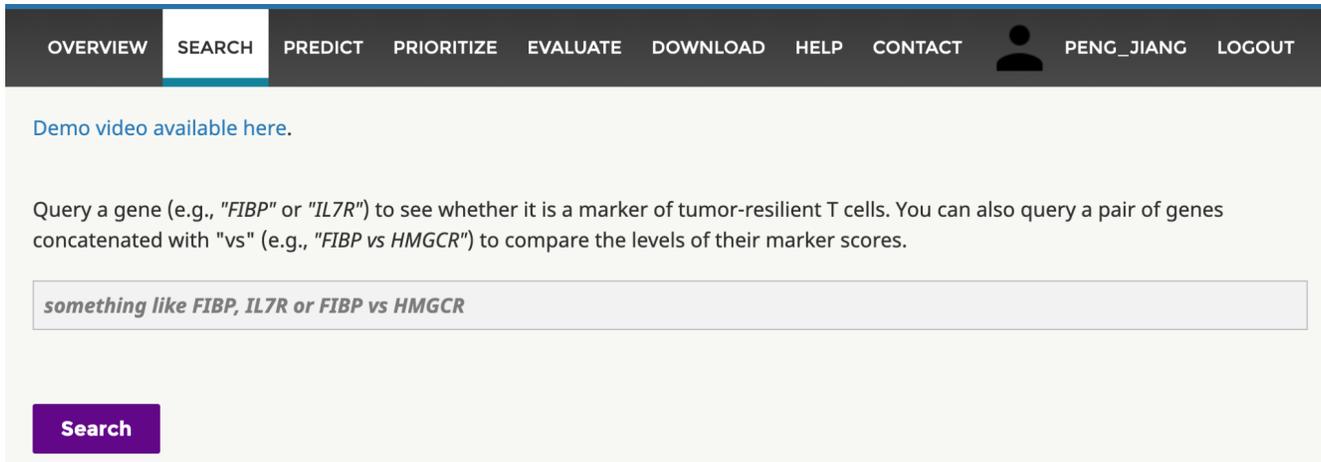
Despite breakthroughs in cancer immunotherapy, most T cells reactive to tumor targets cannot persist in immunosuppressive solid tumors. Identifying molecular programs of T cells sustaining effective antitumor immunity is the center of cancer research. We developed a computational framework named the **tumor-resilient T cell (Tres) model**. Tres utilizes single-cell transcriptomic data from solid tumors to identify signatures of T cells that are resilient to immunosuppressive signals, including TGFB1, TRAIL, and PGE2. Analyzing single-cell data cohorts, the Tres model can predict the clinical efficacies of T cells in immune checkpoint blockade and adoptive cell transfer.

The Tres webapp has four usages.

- 1:** Query whether a gene is a positive or negative marker of tumor-resilient T cells.
- 2:** Input gene expression profiles of T cells or T-cell enriched samples and predict the antitumor efficacy of T cells.
- 3:** Input a gene list and prioritize markers of tumor-resilient T cells.
- 4:** Compare the performance of an input gene signature on predicting T-cell clinical efficacies in cancer immunotherapies against a few previously published signatures.

# Usage 1: Query genes

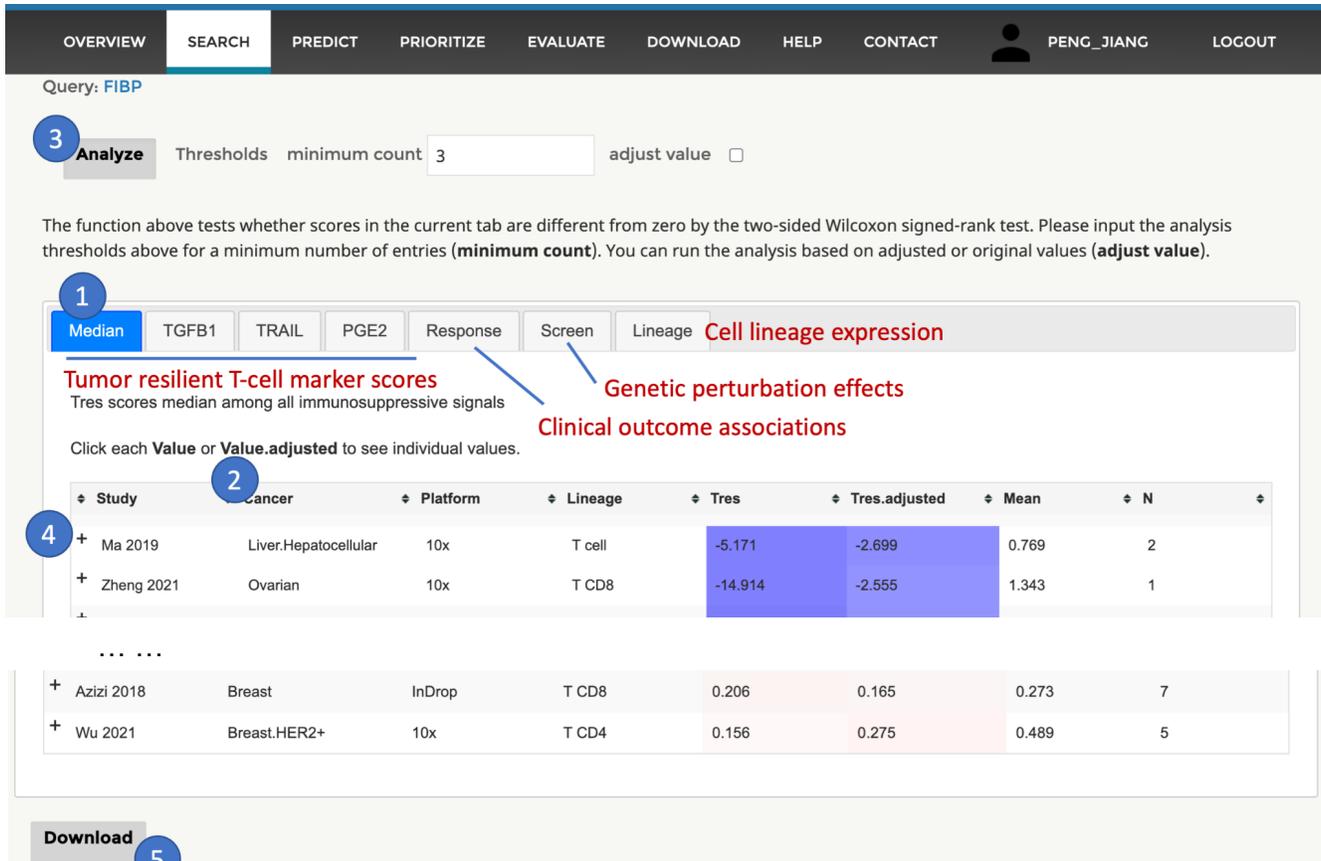
Click the “SEARCH” menu to enter this module. You can query one gene (e.g., “FIBP”) to check its potential as a T-cell efficacy marker, or query a gene pair (e.g., “FIBP vs HMGCR”) to compare their scores.



The query module will present results through a few tables with statistical significance tests and plots.

## Result Tables

For example, you can input FIBP in the search box and click Search. After submitting, you will get a result interface like this:



**1, Result type tabs.** A few result types prioritize gene markers related to the antitumor effectiveness of T cells, which will be explained in detail below.

**2, Header with multiple functions.**

- Mouse over each column name will trigger the explanation of each field.
- Mouse click on each column name will sort rows in either ascending or descending (click again) order by column values.
- Row selector will pop up once you put the mouse cursor on the space beneath the header. You may input a keyword to select a subset of rows using the column value. For example, you can input “NSCLC” in the box beneath the column “Cancer” to only analyze non-small cell lung cancer profiles. The row selector will affect the analysis modules introduced later.

Study	Cancer	Platform	Lineage	Tres	Tres.adjusted	Mean	N
	Median@Tres						
+ Ma 2019	Liver.Hepatocellular	10x	T cell	-5.171	-2.699	0.769	2
+ Zheng 2021	Ovarian	10x	T CD8	-14.914	-2.555	1.343	1
+ Zheng 2021	Esophageal	10x	T CD8	-7.227	-2.493	1.381	6
+ Zhang 2019	Liver.Hepatocellular	SmartSeq2	T CD8	-4.431	-2.309	1.858	2

The meaning of each column is available below.

- **Study:** The study citation name for each dataset in a row. Clicking each study name in a row will open the Pubmed Record page.
- **Cancer:** Cancer type
- **Platform:** Transcriptomic profiling platform
- **Lineage:** Cell lineage subset for each analysis. Please note that we only annotate “T cell” instead of detailed “T CD8” or “T CD4” for certain rows, because the related studies did not divide T cell annotations into CD4 versus CD8 sub-clusters due to the high similarity of their transcriptomics profiles.
- **Therapy (for Response tab):** therapy patients received
- **Outcome (for Response tab):** Type of patient outcome recorded in the study
- **Sample, Condition (for Screen tab):** The sample name of a model for screening and genetic perturbation conditions.
- **Value:**
  - **Median tab** presents median values among TGFB1, TRAIL, and PGE2 tabs, explained below.
  - **TGFB1, TRAIL, and PGE2 tabs:** the value column presents the median Tres gene score among all tumor samples in each dataset.
  - **Response tab:** the value column presents association scores between gene expression and favorable outcomes of cancer immunotherapies. There are several types of outcome
    - Survival (overall survival, B cell aplasia): For the right censored survival length, the association scores will be negative z-scores from the Wald test in Cox-PH regression.
    - Responder: For binary status of whether patients respond to the treatment, the association scores will be z-scores from the Wilcoxon rank-sum test.

- **Screen tab:** the value column presents the CRISPR screening phenotype score, computed as the differential guide RNA frequency between screening and control conditions.
- **Lineage tab:** gene expression in diverse cell lineages.
- **Value.adjusted:** Adjusted value through normalizing the original value by zero mean and standard deviation in each profile to make scores comparable across different profiles. For CRISPR activation or ORF overexpression screens, we reversed the sign of logFC values to ensure that the phenotypic scores have the same meaning as knockout or inhibition screens.
- **Value control, Value.adjusted control (with control gene in the query):** If users query a gene with a control (e.g., FIBP vs HMGCR), the result table will have two additional columns. Thus, scores of target and control genes can be compared together in analysis modules.

Dataset	Cancer	Platform	Tres	Tres.adjusted	Tres control	Tres.adjusted control
Breast.GSE176078.10x	Breast.ER+	10x	-4.263	-0.984	0.242	1.247
Breast.GSE176078.10x	Breast.HER2+	10x	-3.153	-1.187	2.397	2.736

- **Mean:** The mean value across all samples for the query gene. For Tres score analysis, this value will be the mean value across all single cells for the query gene.
- **N:** The sample size of the current statistical analysis.

**3, Significance analysis.** This function module will report the statistical significance test (in detail below). Users can set the minimum number of values and whether to use the original or adjusted value. If a row selector is triggered to filter rows (function 2 explained above), this module will only analyze active rows.

**4, Expandable marker (if available).** Show sub-row results if available.

- **Tres (Median, TGFB1, TRAIL, PGE2):** Each row shows the median scores across all tumors in a dataset. Thus, the sub-row values are the Tres scores in individual tumors.
- **Response:** Each row shows the association on the gene level. Thus, the sub-row values are associations for individual transcript isoforms if the row RNA-seq data is available.
- **Screen:** Each row shows the median log2FC among all guide RNAs. Thus, the sub-row values are log2FC for individual guide RNAs.

**5, Download.** Download the table in CSV. Only major row values will be included, but not sub-row values (function 4 above). If you are interested in further details of our scores, please download our complete results in the Download menu.

## Significance analysis.

The Tres query module provides two types of statistical analysis. The level-1 analysis shows the significance of values in the entire table. The level-2 analysis shows the significance of each value in table rows.

## Level 1: significance of group values

Above the result table in each tab, clicking the Analyze button will trigger statistical analysis and plots.

### 1A: query a single gene

If a single gene is a query, the statistical test will compare values (sample median in each dataset) with zero through the two-sided Wilcoxon signed-rank test.

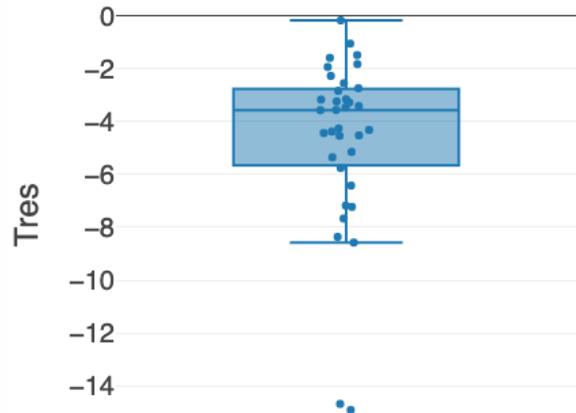
#### two-sided Wilcoxon signed-rank test comparing FIBP vs 0

**Median = -3.57**

**Std = 3.26**

**P = 2.48e-7**

**N = 35**



### 1B: query a target gene versus control

If a gene pair is a query, the statistical test will compare values of the target gene against the control gene through the two-sided Wilcoxon signed-rank test. In the box plot, two values of the same dataset in the target and control groups are connected with grey lines.

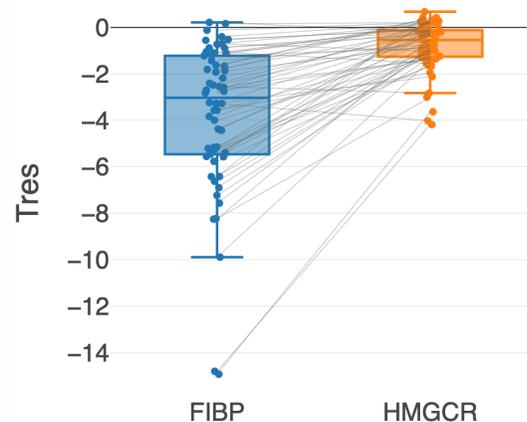
#### two-sided Wilcoxon signed-rank test comparing FIBP vs HMGCR

**Median difference = -2.27**

**Std difference = 2.74**

**P = 2.01e-10**

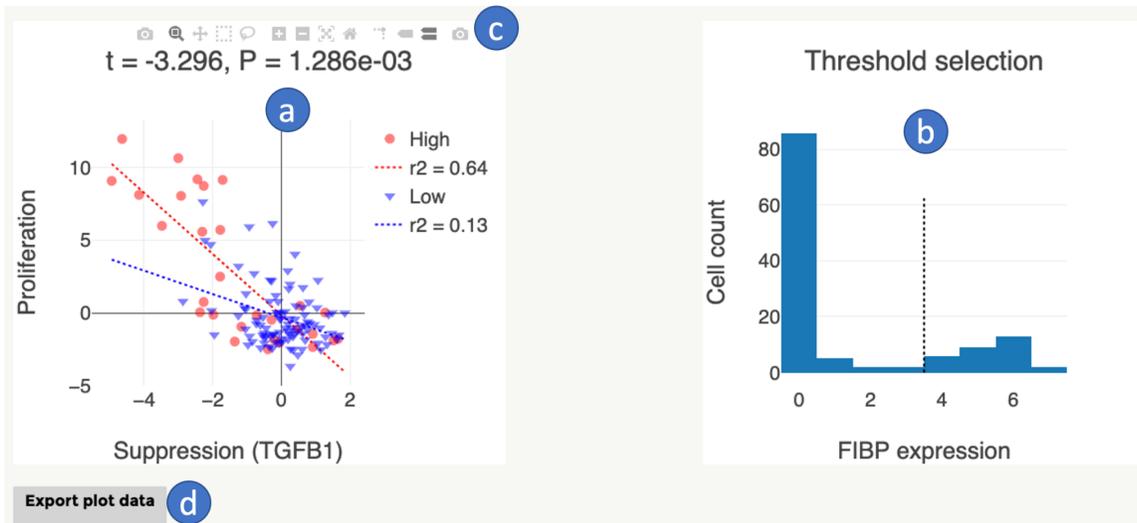
**N = 58**



## Level 2: significance of individual values

### 2B: Tres scores

For TGFB1, TRAIL, and PGE2 tabs, the web server can generate an interaction plot for the Tres model if users click the value numbers for each sample. Please note that the interaction plot is unavailable for the Median tab, which simply shows median scores among TGFB1, TRAIL, and PGE2 tabs, and thus does not have its interaction plots.

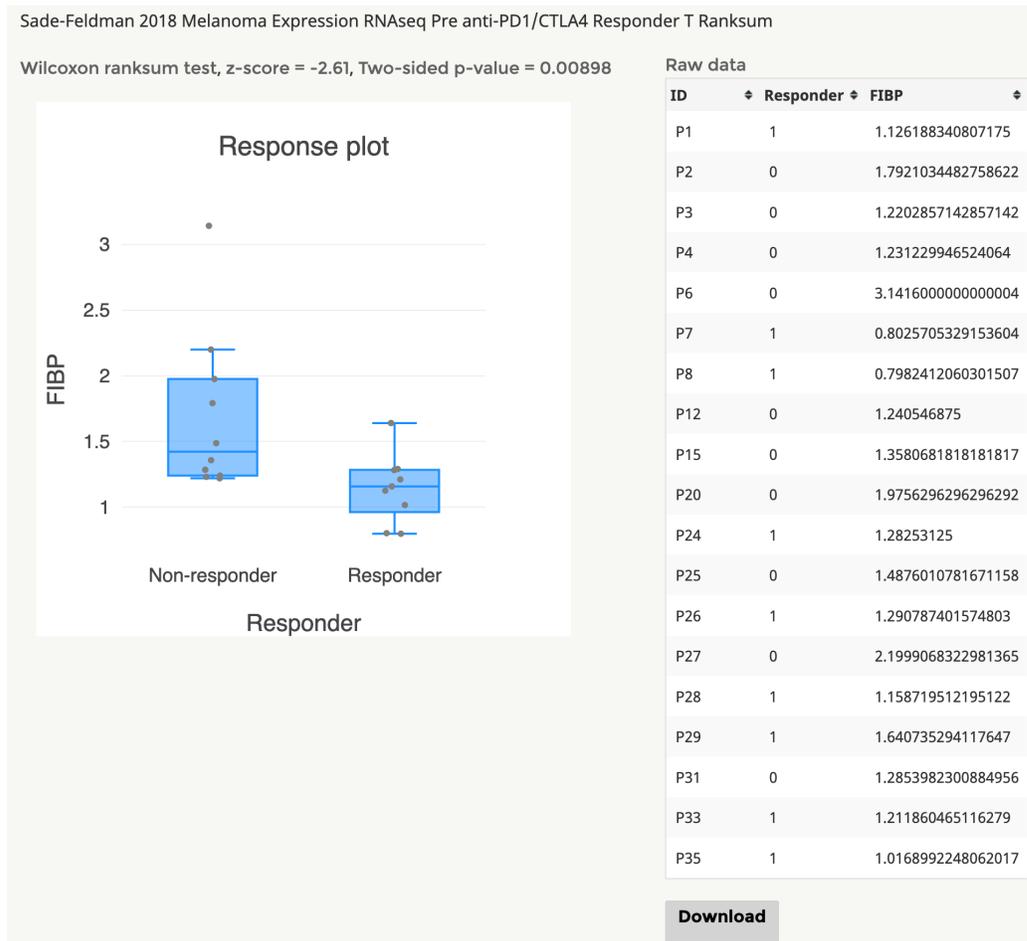


The left panel (a) presents the correlation between immunosuppression from each signal and the T-cell proliferation score. Each data point represents one single T cell in the sample. T cells are divided into high or low groups according to the query gene expression with the threshold selected in the right panel (b). The users can download the image in SVG or PNG by clicking the top corn of the left panel (c) and export all data values in the plot into an excel file (d).

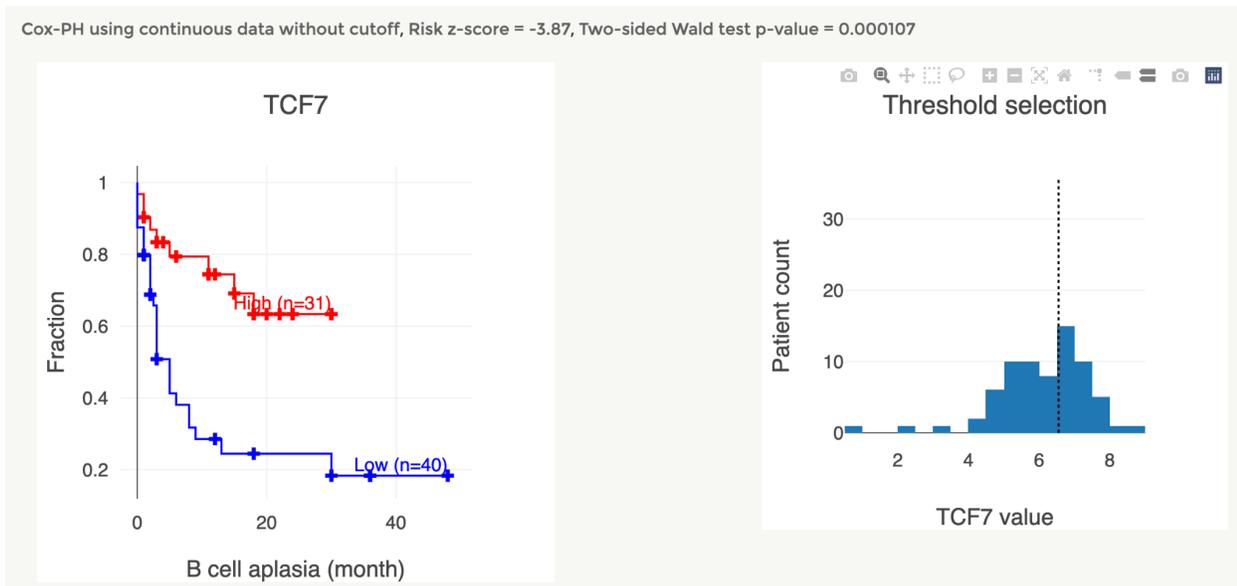
## 2C: Response

In the Tab for therapy response, clicking a value will trigger a box plot (if the outcome type is responder status) or Kaplan-Meier (KM) (if the outcome type is survival) plot, showing associations between gene expression and clinical outcomes.

For KM plot, a gene expression cutoff will be selected through the best separation criteria. Users can change this value by clicking in the threshold

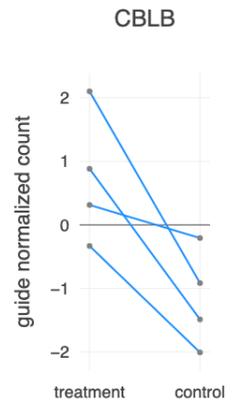


selection panel. This cutoff selection will not affect the statistical significance test result, using continuous values without any cutoffs.



## 2D: Screen

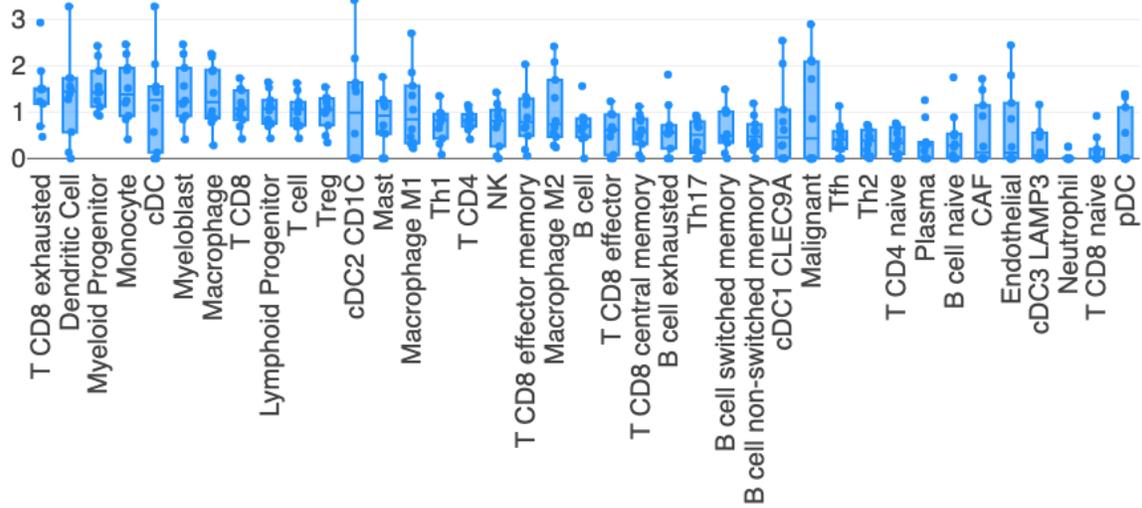
Clicking a log<sub>2</sub>FC value in the Screen tab will show the guide RNA (gRNA) frequency in each study between two screening conditions. The normalized gRNA count is defined as the  $\log_2(\text{count} + 1) - \text{mean}(\log_2(\text{count} + 1))$  across all control gRNAs (typically non-targeting).



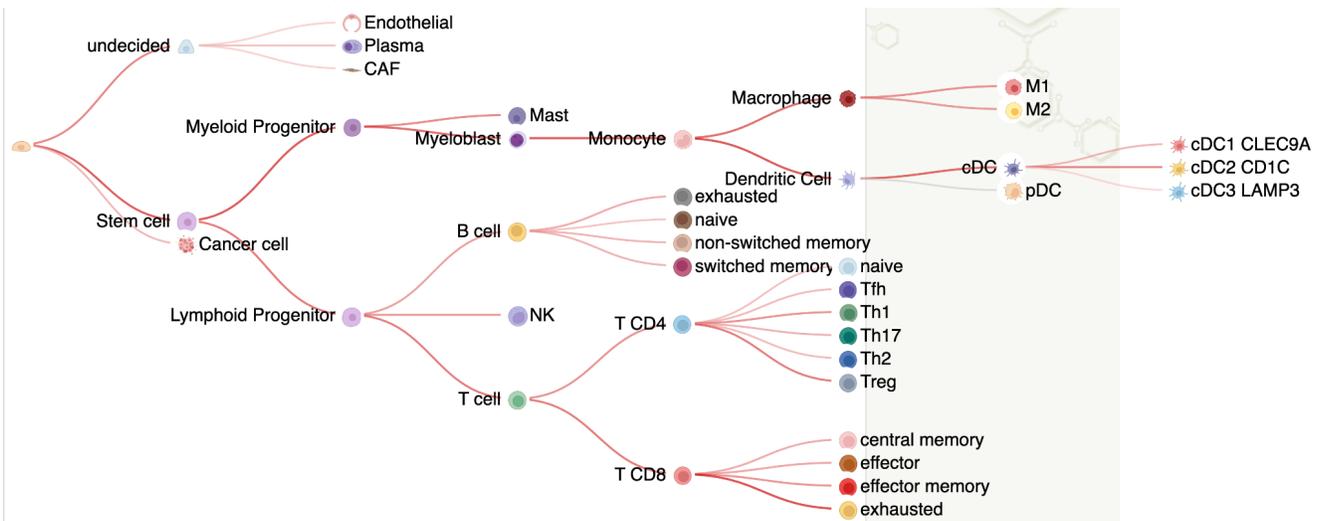
## Lineage expression

Besides all the results discussed above, we also have a lineage tab presenting the expression levels of the query gene in diverse cell lineages, as defined in our SpaCET study (Pubmed ID 36732531). The first plot is a boxplot. Each dot represents the average expression among all tumors and single cells in a study.

Normalized expression



The second plot is a tree-lineage plot, with each edge representing the average expression of the query gene across all datasets. You can click on each cell type node to shrink or expand the sub-branch. Please note that for a gene pair query (e.g., FIBP vs HMGCR), the edge color will present the difference between expression values (FIBP - HMGCR).



## Usage 2: Predict tumor-resilience of T cells

The users can predict whether T cells are resilient to immunosuppressive signals in tumors using the correlation between the median Tres signature (trained using an extensive collection of single-cell RNA-seq data) and the transcriptomic profiles of T cells (or T-cell enriched samples). Such correlation values may predict immunotherapy response using samples from pretreatment tumors for immune checkpoint inhibitors, infusion products, or pre-manufacture samples for cellular immunotherapies. This web app only allows running small data. If users want to analyze a large dataset, please download the Tres source code.

OVERVIEW SEARCH **PREDICT** PRIORITIZE EVALUATE DOWNLOAD HELP CONTACT PENG\_JIANG LOGOUT

[Demo video available here.](#)

Please upload a matrix file with columns as samples (T cells or T-cell rich samples) and rows as genes. The platform will predict whether the current T cell (or T-cell enriched sample) are tumor-resilient T cells.

**Note:** If you want to run the Tres model to compute signature genes of tumor-resilient T cells using your single-cell RNA-Seq data, please [download the source code](#) and run locally.

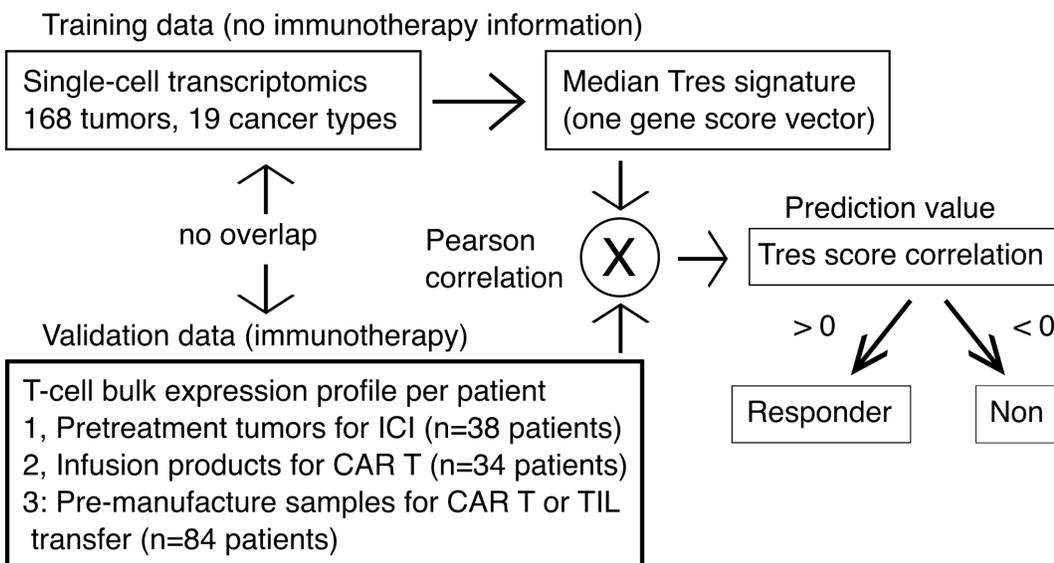
The input value separator should be the tab space, and the gene names should follow the standard gene symbols. The platform accepts text or Gzip files as input. Please see a [sample tab file](#). Alternatively, you can upload Excel files ([sample CSV](#) or [sample XLSX](#)).

The file size should be smaller than 20MB, and the number of samples should be less than 40. If you need to analyze more samples, please [download](#) our software and run it locally.

The expression values, from either RNASeq or MicroArray, should be transformed by  $\log_2(x+1)$ .  $x$  could be FPKM, RPKM, or TPM for RNASeq. For single-cell RNASeq data, we used  $\log_2(\text{TPM}/10 + 1)$ . We also recommend quantile-normalization across conditions. Some software package, such as RMA or DESeq, will automatically include all normalizations. We recommend input differential profiles between the two conditions. If data is from a sample collection without pairs, please mean-centralize the value of each gene across all samples.

Task\_file:  
 No file chosen

Here is an illustration of the prediction procedure from our original publication. You can submit your T-cell transcriptomic profiles (as validation data) to compute your T-cell samples' potential antitumor efficacy.



## Input data format

The input data should be an expression matrix with columns as samples and rows as genes. A few sample data files are available on the page. The input value separator should be the tab space, and the gene names should follow the standard gene symbols. The platform accepts text, excel, or Gzip files as input. The file size should be smaller than 10MB, and the number of samples should be less than 40. If you need to analyze more samples, please download the Tres source code and run it locally.

The expression values, from either RNASeq or MicroArray, should be transformed by  $\log_2(x+1)$ .  $x$  could be FPKM, RPKM, or TPM for RNASeq. For single-cell RNASeq data, we used  $\log_2(\text{TPM}/10 + 1)$ . Please mean-centralize the value of each gene across all samples or input differential expression profiles. The Tres model does not work on absolute gene expression values.

If a user wants a single sample prediction, we suggest the user create a reference background gene expression profile, using a pooled average from many samples previously generated from your genomics workflow. Then, you can subtract your single-sample transcriptomic profile from the pooled background profile as normalization. This strategy is similar to the pooled-tumor reference sample utilized in the FDA-approved biomarker MammaPrint.

## Result format

After uploading the sample file by clicking the “Choose File” button and submitting it, you will see correlation values ( $r$ ) and their statistical significance ( $p$ , two-sided t-test) for each sample. T cells (or T-cell enriched samples) with significant positive correlations are considered as tumor-resilient T cells (Tres) (or samples containing sufficient among of Tres).

## Usage 3: Prioritize marker genes of tumor-resilient T cells

Many users may generate a gene list from their studies. This module can prioritize the likelihood of genes as markers of effective antitumor T cells across many curated datasets. Users simply input a gene list, separated by “,” or new lines, and click the prioritize button.

Here is a gene list example:

“CBLB, TMEM222, CDKN1B, SOCS1, DGKA, CD8A, RASA2, DGKZ, UBASH3A, TCEB2”, which are the top ten genes ranked by log<sub>2</sub>FC from a CRISPR screen of T-cell proliferation upon TCR stimulation (Pubmed ID 30449619). After clicking the Prioritize button, the result includes a data table and its associated heatmap.

**Gene prioritization results**

Gene	Weighted Average				Tres					
	All	Tres	Response	Screen	Azizi 2018 Breast InDrop, T CD4	Azizi 2018 Breast InDrop, T CD8	Durante 2020 Uveal 10x, T CD8	Guo 2018 NSCLC SmartSeq2, T CD4	Guo 2018 NSCLC SmartSeq2, T CD8	JerbyArnon 2018 Mela SmartSeq2, Post, Prog ICI, T CD8
ELOB	-0.679	-0.814	-0.367	-0.856	-1.402	-0.943	-0.757	-1.208	-0.426	-1.361
TMEM222	-0.593	-0.404	0.172	-1.55	-0.061	0.351	-0.673	-1.054	0.325	-0.720
CBLB	-0.282	0.651	0.400	-1.90	0.279	-0.792	0.218	1.003	0.125	1.289
DGKZ	-0.239	-0.0216	0.202	-0.899	-0.575	0.588	0.017	1.194	-0.179	-0.116
UBASH3A	-0.186	0.123	0.175	-0.855	-0.296	0.407	0.637	0.338	0.280	1.085
RASA2	-0.0865	0.174	0.410	-0.844	1.467	-0.758	0.026	0.536	0.287	0.544
DGKA	0.0385	0.253	0.682	-0.819	-1.703	0.435	0.321	-0.630	0.614	0.205
CD8A	0.0965	0.338	0.487	-0.535	-2.283	0.734	2.559	-2.631	0.621	1.458

**Download**

Part 1: Scores of different analysis categories, including Tres, Response, and Screen. Clicking each

column will sort the table rows by scores in that column. Please move the scroll bar on the bottom to view the whole data table.

All scores listed in this table are adjusted values, which are the original values normalized by the mean and standard deviation in each profile. A positive Tres or response score indicates the gene as a positive marker of effective antitumor T cells. In the Response group, the signs of scores from survival analyses are reversed because Cox-PH regression gives risk scores, indicating the association between gene level and death risks. The sign of CRISPR activation or ORF overexpression screen scores are reversed; thus, their scores have the same directionality as CRISPR inhibition or knockout screens.

Part 2: Weight for merging scores across different analysis categories. The All column represents the average across all categories. The signs of CRISPR scores are reversed when computing the average to have consistent directionalities with Tres and Response scores, where higher scores mean the gene activity is associated with favorable antitumor outcomes.

The Webapp also generates a heatmap for each prioritization table. You can download png or svg files by clicking buttons on the top right corners.

## Usage 4: Evaluate the biomarker value of a gene signature

Many users may create a gene signature of T-cell antitumor efficacy, thus want to evaluate whether the signature can predict cancer immunotherapy outcomes and the relative performance compared to previously published signatures. This module has integrated many T-cell transcriptomics data from outcome studies of immune checkpoint blockade, adoptive cell transfer, and CAR T. A user can input gene signatures of positive and negative gene sets (Option 1) or gene weights (Option 2). This module will compute score sums (Option 1) or correlations (Option 2) for patients in each dataset and present the prediction performance.

OVERVIEW SEARCH PREDICT PRIORITIZE **EVALUATE** DOWNLOAD HELP CONTACT PENG\_JIANG LOGOUT

[Demo video available here.](#)

This module will evaluate whether the input signature can predict immunotherapy outcome in comparison with other T-cell specific signatures. If you need to analyze many signatures, please [download](#) our sourcecode and run it locally.

Option 1: Weighted gene signature:  
 No file chosen

Upload a signature of T-cell fitness with the first column as gene symbols and the second column as gene weights. The first row should be the signature name. Positive weight should indicate positive markers of T-cell clinical efficacy and negative weights indicate negative markers. The input value separator should be the tab space, and the gene names should follow the standard gene symbols. The platform accepts text or Gzip files as input. Please see sample files ([Tab format](#), [CSV format](#), [Excel XLSX format](#)). The file size should be smaller than 20MB.

Option 2: Positive marker genes:

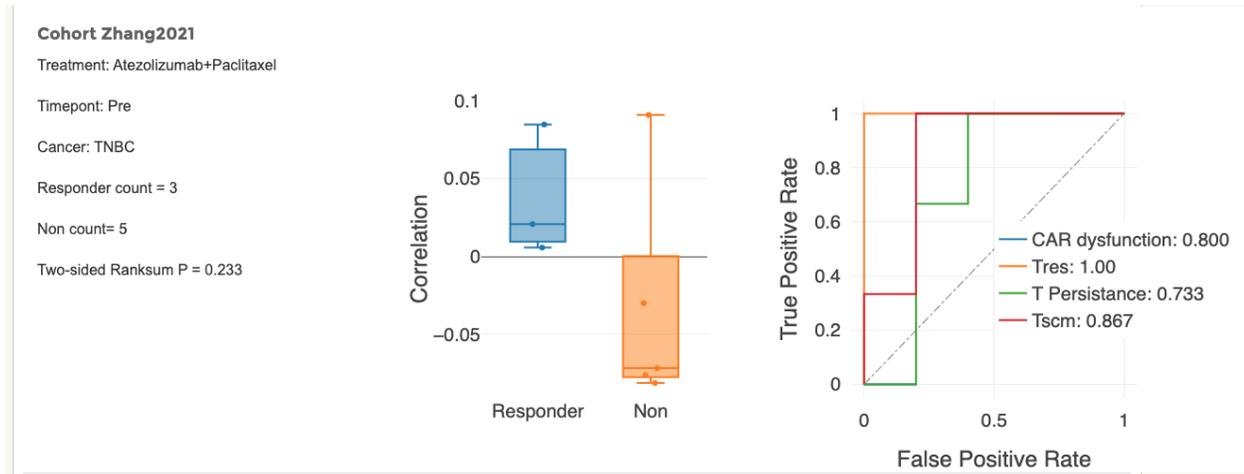
Input positive and/or negative marker genes, separated by , or new lines.

Option 2: Negative marker genes:

Please submit our sample file of gene weights adapted from a CAR T dysfunction study (Pubmed ID 34861191). The gene weight file can only have one column for evaluating one biomarker. If users want to assess multiple biomarkers, please download the [Github package for reproducing our main results](#). Then, users can adapt our source code. After submitting, you will see a result interface for two data types.

### 4A, Response

Results for datasets with binary clinical outcomes. The first figure panel presents prediction metrics using the input biomarker for responders and non-responders. The second figure panel shows the ROC curve comparing prediction performance of the user input versus a few published signatures of T-cell antitumor efficacy. A better biomarker should stay above other biomarkers. Also, the web app presented the area under the ROC curve (AUC), a quantitative metric of overall prediction performance between 0 (worst) and 1 (best).

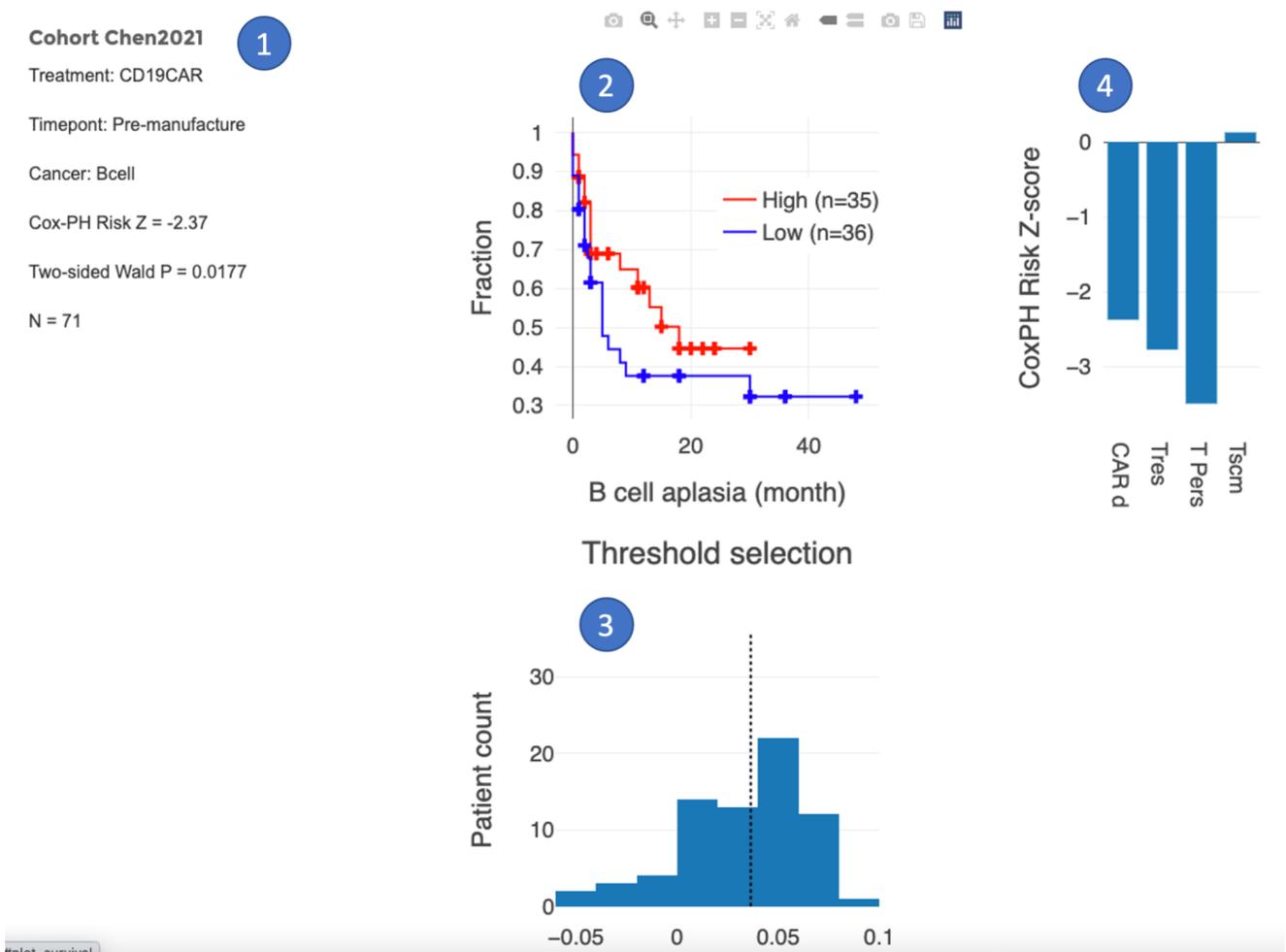


In each figure panel, users can move the mouse cursor over the top to trigger the download bar, providing a png or SVG file for image editing or source data in CSV.



## 4B, Survival

Results for datasets with survival durations.



#plot\_survival

This panel has four parts labeled above:

1. Two-sided Wald test results through Cox-PH regression, using continuous values without any cutoffs.
2. Kaplan-Meier survival plot, using a threshold for cutting prediction metrics determined in the third part explained below.
3. Cutoff selection for visualization purposes in the KM plot. The cutoff will not affect statistical test results using continuous values in part 1.
4. Wald test risk z-scores, comparing the user input and previously published signatures. A promising biomarker should have significantly negative z-scores, suggesting that biomarker-positive patients have low death risks.

In each figure panel, users can move the mouse cursor over the top to trigger the download bar, providing a png or SVG file for image editing or source data in CSV.

## Usages not available on Webapp

### **Computing signature of tumor-resilient T cells on single-cell RNA-Seq data**

Instead of using our precomputed Tres scores, users may want to compute Tres gene signatures using their single-cell RNA-Seq data in a specific study setting. Since this function will involve intensive computing resources, we do not provide modules on our web server. Instead, please go to the Download page and install Tres locally.