TCPGdb: A Comprehensive T-cell Perturbation Genomics Database for the Identification of Critical T-cell Regulators

Chuanpeng Dong[1,2,3,4,5], Feifei Zhang[1,2,3], Kaiyuan Tang[1,2,3], Nipun Verma[1,2,3], Xinxin Zhu[4,5], Di Feng[5,6], James Cai[5,6], Hongyu Zhao[1,4,5,7], and Sidi Chen[1,2,3,4,5,8]

ABSTRACT

Large parallel genetic screens have been used to identify targets and regulators that enhance T-cell antitumor capability and persistence in the tumor microenvironment. We hypothesized that by combining the pooled screen data from multiple independent genetic screens, we could provide a systematic, comprehensive, and robust analysis of the effect of gene perturbation on T cell–based immunotherapies. After collecting data from previously published T-cell screens, including CRISPR-based and open reading frame–based screens, through the Gene Expression Omnibus, we reprocessed the gene hits summary and conducted a pathway enrichment analysis. A T-cell screen perturbation score metric was employed to quantify the impact of a gene perturbation on T-cell function. Additionally, gene expression data (both bulk RNA level and single-cell RNA level) from autoimmune disease cohorts and patients with T cell–derived cancer were incorporated to gain further insight into gene perturbations that potentially augment T-cell proliferation. We integrated all data and analysis on 35 T-cell screens into our state-of-the-art T-cell perturbation genomics database (TCPGdb), which is accessible through our web server (http://tcpgdb.sidichenlab.org/) and allows users to interactively explore the impact of query genes on T-cell function.

Introduction

T cells engineered with a chimeric antigen receptor (CAR T-cell therapy) are an adoptive cell therapy that recognizes and targets cancer cells expressing a specific target antigen. Since its first clinical application in chronic lymphoblastic leukemia in 2010, CAR T-cell therapies have shown remarkable clinical efficacy, particularly against hematologic malignancies, in numerous clinical trials (1, 2). However, CAR T-cell therapies are facing a persistence issue due to several factors, such as reduced expression of the target antigen on cancer cells or dampened function of the transferred CAR T cells. CAR T cells, which are chronically stimulated by a cancer-specific antigen, will eventually differentiate into a dysfunctional exhausted state, characterized by reduced proliferative capacity and reduced effector function (3).

High-throughput CRISPR-based genetic screens with either targeted or genome-wide libraries have become a widely utilized method for identifying key regulators that can prevent CAR T-cell exhaustion and enhance their function (4). Shifrut and colleagues (5) performed a genome-wide loss-of-function (LoF) screen in primary human T cells and identified essential regulators for T-cell receptor signaling and proliferation following stimulation. Dong and colleagues (6) performed the first *in vivo* genome-scale CRISPR LoF screen of primary murine $CD8^+$ T cells and identified regulators of $CD8^+$ T-cell tumor infiltration in a triple-negative breast cancer mouse model. Schmidt and colleagues (7) used both CRISPR activation (CRISPRa) and CRISPR interference (CRISPRi) genome-wide screens conducted in primary human T cells to identify the functional regulators of cytokine production in response to stimulation, which is often dysfunctional in autoimmune diseases and cancers. Wang and colleagues (8) conducted CRISPR LoF screens to directly identify essential regulators for cytotoxicity function in IL13Rα2-targeted CAR T cells. With the growing number of screens (9–19) identifying genes that regulate specific T-cell characteristics (e.g., survival, proliferation, or exhaustion) and functions (e.g., tumor infiltration, cytokine production, or cytotoxicity), the need for systematic databases that specifically map T-cell gene functions has become increasingly urgent. Although CRISPR databases such as DepMap (20), BioGRID (21), iCSDB (22), CRISP-view (23), and ICRAFT (24) provide valuable insights into cancer cell functions, they are more focused on CRISPR screening in cancer cells. There remains a critical need for dedicated resources to support the advancement of T cell–based therapy research and to address the growing demand in this field.

In this study, we developed a large-scale, publicly accessible data repository called the T-cell perturbation genomics database (TCPGdb). TCPGdb will serve as a comprehensive resource depicting the functional effect of individual gene perturbations in T cells, covering all previously published T-cell genome-wide screen datasets, as well as the transcriptional profiles of sorted T cells, CAR T cells, T-derived cancer cells, and autoimmune disease datasets. In summary, TCPGdb provides a one-stop platform for searching, browsing, and visualizing gene functions in T cells.

Materials and Methods

CRISPR screen data collection and processing

T-cell CRISPR pooled screen datasets were obtained from the Gene Expression Omnibus (GEO; ref. 25) and previously published

[1]Department of Genetics, Yale University School of Medicine, New Haven, Connecticut. [2]System Biology Institute, Yale University, West Haven, Connecticut. [3]Center for Cancer Systems Biology, Yale University, West Haven, Connecticut. [4]Center for Biomedical Data Science, Yale University School of Medicine, New Haven, Connecticut. [5]Yale-Boehringer Ingelheim Biomedical Data Science Fellowship Program, New Haven, Connecticut. [6]Computational Innovation, Boehringer Ingelheim Pharmaceuticals, Inc., Ridgefield, Connecticut. [7]Department of Biostatistics, Yale University School of Public Health, New Haven, Connecticut. [8]Yale Comprehensive Cancer Center, Yale University School of Medicine, New Haven, Connecticut.

Corresponding Author: Sidi Chen, Department of Genetics and System Biology Institute, Yale University School of Medicine, 850 West Campus Drive, RM320, West Haven, CT 06516. E-mail: sidi.chen@yale.edu

Cancer Immunol Res 2026;14:219–27

doi: 10.1158/2326-6066.CIR-25-0168

literature through searches of PubMed and Google Scholar with the keywords "CRISPR screen AND T cell." To combine the CRISPR screen datasets, we retrieved the raw single-guide RNA (sgRNA) count table from all CRISPR screens and then processed each dataset using MAGeCK (26) with uniform parameters. For datasets without raw data, the summarized data from original literature will be used. For those screens using mouse models, mouse genes were matched to their orthologous human genes using the biomaRt (27) package in R. Detailed information on the included CRISPR screen datasets is provided in Supplementary Table S1.

RNA sequencing and microarray data processing

For the RNA sequencing (RNA-seq) data, raw sequencing reads were downloaded and unpacked using the SRA-Toolkit (version 2.9.0-ubuntu64). Fastp was used to trim the adapter sequence and remove low-quality reads (28). The trimmed reads were then mapped to the human reference genome GRCh38 (gencode.v45) and quantified at the gene level using the --quantMode GeneCounts option in STAR (version 2.7.10b; ref. 29). The count data were normalized to transcripts per million in $\log_2(x + 1)$ format.

The R packages affy or oligo were used to process the raw CEL microarray data. Robust multichip average was applied to normalize microarray expression values. Probes were mapped to Ensembl IDs using the biomaRt package in R.

Sorted T-cell RNA-seq dataset

T-cell bulk RNA-seq data were obtained from two sources: the Database of Immune Cell Gene Expression, Epigenomics, and Expression Quantitative Trait Loci (DICE; ref. 30) and the Absolute Immune Signal (ABIS) deconvolution dataset (31). DICE includes 967 samples covering 11 T-cell subtypes, with gene expression data downloaded from DICE. The ABIS dataset comprises 46 samples, including 12 T-cell subtypes; the raw RNA-seq data were downloaded and processed using standard RNA-seq data processing methods as described above. T-cell gene coexpression analysis was conducted using the DICE dataset.

CAR T-cell bulk and single-cell RNA-seq dataset

We collected CAR T-cell datasets that include healthy/control samples and dysfunctional (exhausted, nonresponding/impaired) samples. A total of five datasets were gathered: two bulk RNA-seq datasets and three single-cell RNA-seq (scRNA-seq) datasets from three different studies (32–34). Detailed information on these datasets is summarized in Supplementary Table S2.

T-cell lymphoma/leukemia dataset

We obtained three T-cell lymphoma/leukemia datasets featuring gene expression prognostic analysis from the GEO: GSE19069 ($n = 147$; ref. 35), GSE58445 ($n = 193$; ref. 36), and GSE90597 ($n = 29$; ref. 37). The array datasets were downloaded and processed using the microarray data processing procedures described later. Clinical and survival information were retrieved from the original studies or requested from the corresponding authors. Detailed information on the T-cell lymphoma datasets is provided in Supplementary Table S3.

Transcriptomic profiling of autoimmune diseases

We obtained gene expression datasets profiling peripheral blood mononuclear cells (PBMC) for three types of autoimmune diseases from the GEO: four multiple sclerosis (MS) datasets (38–41), three inflammatory bowel disease (IBD) datasets (42–44), and three systemic lupus erythematosus (SLE) datasets (45–47). The microarray and RNA-seq datasets were downloaded and processed using the microarray data processing procedures described later. Detailed information on the collected autoimmune disorder datasets is summarized in Supplementary Table S4.

Gene set enrichment analysis

Gene set enrichment analysis (GSEA; ref. 48) was performed to evaluate the effect of CRISPR screen gene targets at the pathway level. The analyses were conducted using the R package clusterProfiler (49) with a curated gene set of biological process (BP) category in gene ontology (GO) (50).

Survival analysis

The survival analysis of a specified gene in the T-cell lymphoma/leukemia dataset was calculated using a Kaplan–Meier model through R survival packages. *P* values less than 0.05 were considered statistically significant.

T-cell perturbation score calculation

To overcome the bias in gene selection results from various T-cell screening datasets, we developed a T-cell perturbation scoring (TPS) metric to systematically measure the impact of specific genes on T-cell function following CRISPRa or CRISPRi. A gene is considered more impactful if it consistently ranks near the top across multiple datasets. We utilized the robust rank aggregation (RRA) algorithm (51) to compute the significance score (ρ value) for each gene by integrating multiple MAGeCK results. To interpret both LoF and gain-of-function (GoF) screens in the same manner of enhancing T-cell function, we integrated the MAGeCK results as follows: for scoring gene activation impact, we chose negative ranks toward function in knockout screens and positive ranks toward function in activation screens; conversely, for scoring gene knockout impact, we chose positive ranks toward function in knockout screens and negative ranks toward function in activation screens.

We first normalized the MAGeCK gene summary ranks into percentile ranks $U = (u_1, u_2, \ldots, u_m)$, where $u_i = r_i/m$ $(i = 1, 2, \ldots, m)$ and m represents the total number of genes in the rank. Under the null hypothesis, where the percentiles follow a uniform distribution between 0 and 1, the k-th smallest value among $u_1, u_2, \ldots, u_m$ is an order statistic that follows a beta distribution. RRA computes the *P* value for each gene based on the beta distribution in each rank. The significance score of the gene across ranks, the ρ value, is defined as $\rho = \min(p_1, p_2, \ldots, p_n)$, where *n* is the number of informative ranks. The TPS for each gene was calculated with the modified RRA ρ value as follows:

$$\mathrm{TPS} = \begin{cases} -0.2 * \log10(\rho), & \text{if } <= 1 \\ 1 & \text{if } > 1 \end{cases}$$

The TPS ranges from 0 to 1. Genes with a TPS near 1 are associated with a strong positive impact on T-cell functions. Conversely, genes with a TPS near 0 indicate minimal effect on enhancing T-cell functions. The TPS provides researchers with a quantifiable measure to evaluate the extent of a gene's impact on T-cell functions.

Least absolute shrinkage and selection operator model for CAR T-cell function prediction

Least absolute shrinkage and selection operator (LASSO), a widely used regularization method for handling multicollinearity

and small datasets, was employed to build a regression model using CAR T-cell transcriptomic datasets. We utilized the scikit-learn package in Python and applied leave-one-out cross-validation (LOOCV) to determine the optimal regularization parameter (λ).

Database development

TCPGdb was powered by the Python Flask framework (https://flask-restful.readthedocs.io/). HTML and CSS were used for the rendering and interactive operations of the front-end pages. MongoDB Atlas (https://cloud.mongodb.com/) was used for the storage of the processed data (https://www.mongodb.com/). The charts were created using ECharts and Matplotlib in Python. Finally, the bioinformatics analyses were performed using R scripts. The R scripts used are included as supplementary material. TCPGdb is hosted using Heroku and redirected to a custom domain.

Results

Data summary of T-cell screen datasets used to generate TCPGdb

The TCPGdb was generated from T-cell CRISPR screen datasets in 17 published studies. These 17 studies included 6 CRISPR-activation screens, 2 open reading frame (ORF) screens, 25 CRISPR knockout screens, and 2 CRISPRi screens. Within the 35 included screens, 22 of them were genome-wide (**Fig. 1A**). The cell types targeted by the CRISPR screens included $CD8^+$ T cells ($n = 13$), $CD4^+$ T cells ($n = 12$), regulatory T cells (Treg; $n = 6$), CAR T cells ($n = 3$), and mixed T cells ($n = 1$). Fourteen of the 35 screens used mouse cells, and the remainder ($n = 21$) used human cells (**Fig. 1B**). The biological functions studied varied, including T-cell killing, proliferation, survival, and the regulation of *IL2*, *IFNG*, and *TNF* expression. Most screens were conducted under *in vitro* conditions using genome-scale libraries (**Fig. 1B**).

MAGeCK reanalysis of CRISPR screen datasets used in TCPGdb

For 13 screens, we were able to obtain the raw sgRNA data. For these screens, we reanalyzed the dataset using the MAGeCK pipeline with uniform parameters. Detailed process metrics can be found in Supplementary Table S1. GSEA was used to evaluate the pathway enrichment based on gene-level results from MAGeCk, including targets identified through both positive selection and negative selection. Our reanalysis of all datasets revealed a large variability in the number of significant gene hits and pathways between screens, highlighting the impact of differences in experimental design, even among the same T-cell types (**Fig. 2A**). It should be noted that gene hits were filtered by log fold change and *P* value, whereas pathway enrichment used the full gene list ranked by log fold change. In some cases, such as Schmidt_CRISPRa_Prim_CD8_IFNG, nonparametric methods identified few enriched pathways despite many gene hits. Additionally, correlation analysis of the screening datasets revealed that among the 13 $CD8^+$ T-cell screen datasets, 15% of the comparisons between individual screens showed significantly positive correlations at the dataset level (12 out of 78 comparisons, $P < 0.05$ and coefficients >0.1). In contrast, only two significant negative correlations were identified (Supplementary Fig. S1A; Supplementary Table S5). For the $CD8^+$ T-cell proliferation screen datasets, in particular, we observed only positive correlation coefficients between screens (**Fig. 2B**), indicating the underlying similarities despite heterogeneity between the datasets. Furthermore, we observed shared target hits between different screen datasets (**Fig. 2C**). For instance, among the six $CD8^+$ T-cell

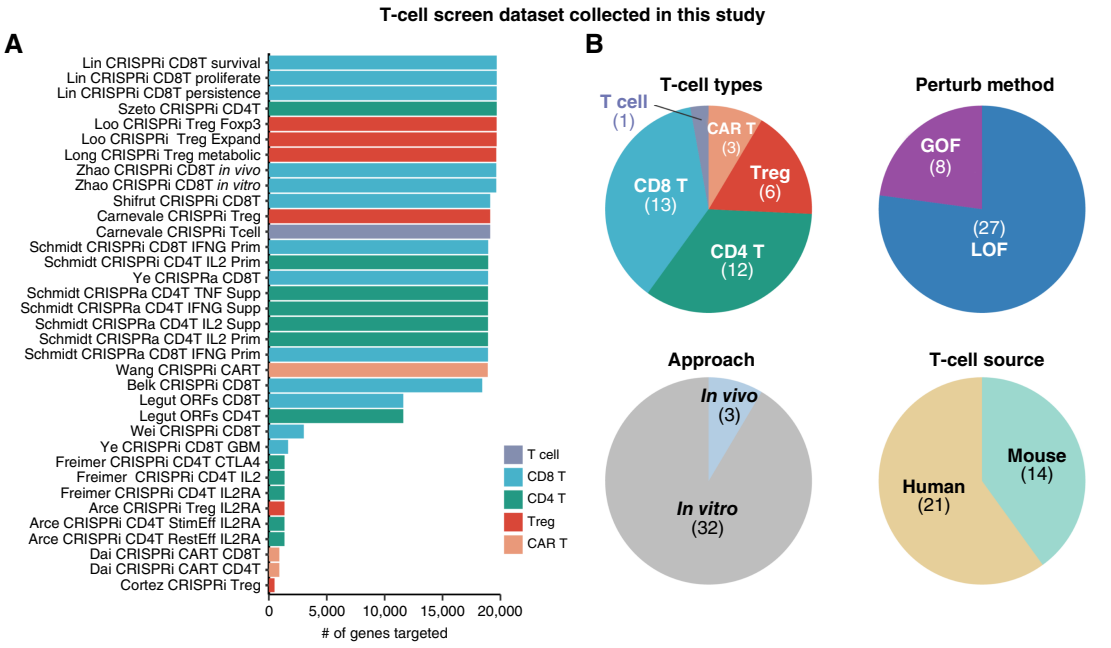

Figure 1.

Summary of T-cell screen datasets used to generate TCPGdb. **A,** Bar plot of target gene library size in each screen experiment. **B,** Pie chart showing the number of screen experiments belonging to different subcategories.

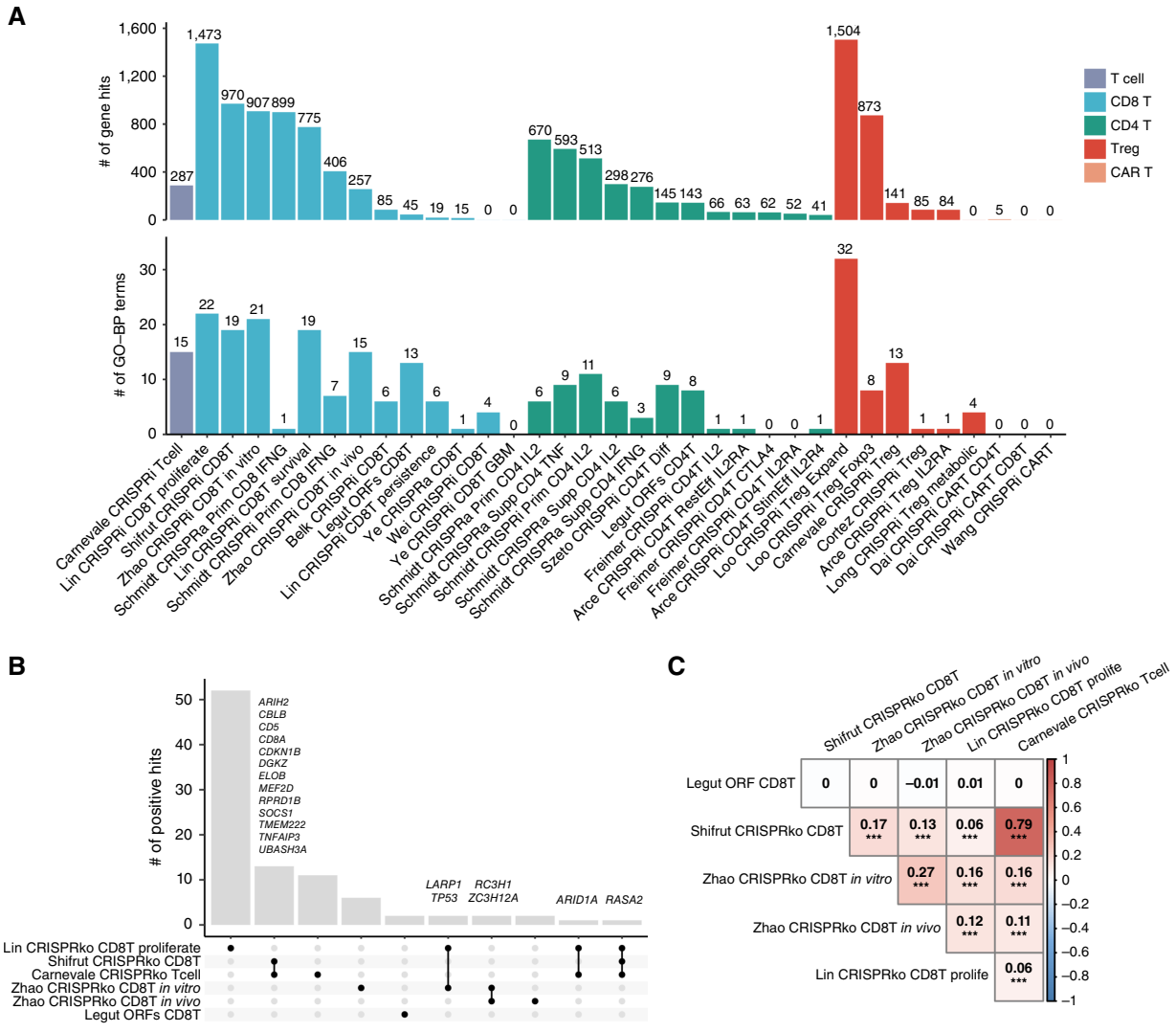

Figure 2.

Summary of MAGeCK analysis results for screen datasets used to generate TCPGdb. **A,** Bar chart showing the number of differentially enriched genes (above) and GO terms enriched among identified genes (bottom) with significant statistical differences. The color of the bar indicates different T-cell types. The level of statistical significance is as follows: DEGs (adjusted *P* value < 0.05) and GSEA BP terms (*P* value < 0.05). **B,** Correlation plot showing the correlation of $CD8^+$ T-cell proliferation screen datasets by positive log fold change from MAGeCK analysis. The displayed values represent the Spearman correlation coefficient, with asterisks indicating the significance levels. **C,** Upset plot showing the positively selected genes shared by $CD8^+$ T-cell proliferation screen datasets from MAGeCK analysis. *, *P* < 0.05; **, *P* < 0.01; ***, *P* < 0.001. ko, knockout.

proliferation CRISPR knockout screen datasets, *ARIH2*, *CBLB*, *CD5*, *CD8A*, *CDKN1B*, *DGKZ*, *ELOB*, *MEF2D*, *RPRD1B*, *SOCS1*, *TMEM222*, *TNFAIP3*, and *UBASH3A* were positively selected in Shifrut and colleagues and Carnevele and colleagues; *LARP1* and *TP53* were both identified in Lin and colleagues and Zhao and colleagues; *RC3H1* and *ZC3H12A* were enriched in the *in vitro* and *in vivo* datasets in Zhao and colleagues; *ARID1A* was identified in both Lin and colleagues and Carnevele and colleagues; and finally, *RASA2* was found to enhance T-cell proliferation in three datasets (Lin and colleagues, Shifrut and colleagues, Carnevale and colleagues; **Fig. 2C**). We did not observe shared signaling pathway enrichment among the $CD8^+$ T-cell screen datasets.

Evaluation and characterization of genes with top T-cell gene perturbation scores from activation and knockout CRISPR screens

After integrating the CRISPR screen datasets, we used a TPS metric to comprehensively characterize the effect of each gene perturbation on T-cell function. The TPS is calculated by summarizing multiple selection rank results obtained from MAGeCK, which allows for a more intuitive understanding of the effect of each gene perturbation (both activation and inactivation) on augmenting T-cell function (see the "Materials and Methods" section for details). The calculated TPS ranges from 0 to 1, with a higher TPS indicating that perturbing the gene has a greater impact on enhancing T-cell functions, whereas a lower TPS is associated with less impact. We

applied the TPS model to calculate the gene score under activation and knockout screen designs using CD8$^+$ T-cell, CD4$^+$ T-cell, Treg, and CAR T-cell screening datasets. We observed positive correlations between the average of *z*-scored log fold change of genes and their TPSs across each T-cell type (Supplementary Fig. S1B). We further analyzed the *z*-scored log fold change for the top 50 genes ranked by the TPS for both activation/GoF (**Fig. 3A**) and knockout/LoF (**Fig. 3B**) perturbations in CD8$^+$ T cells. Intriguingly, 17 of the top 20 ranked LoF genes have already been functionally validated through CRISPR knockout studies, as summarized in Supplementary Table S6. In addition to the overall impact on T-cell function, we also used the TPS model to evaluate the gene potential in

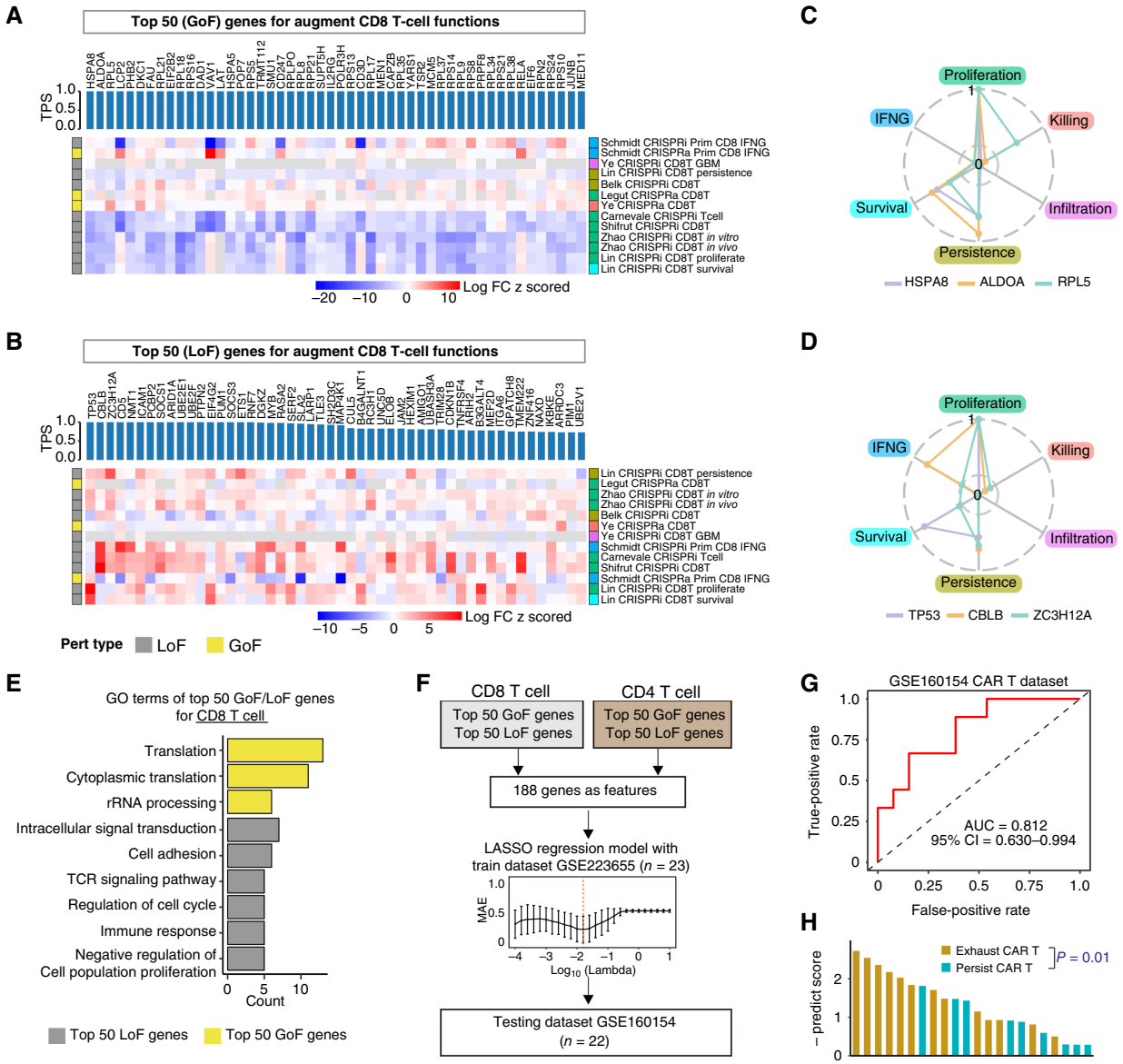

Figure 3.

Summary of top genes and pathways ranked by TPS from 13 CD8$^+$ T-cell screen datasets. **A,** Heatmap of *z* score fold change for top 50 GoF genes ranked by TPS from 13 CD8$^+$ T-cell screen datasets. **B,** Heatmap of *z* score fold change for top 50 LoF genes ranked by TPS from 13 CD8$^+$ T-cell screen datasets. **C,** Radar plots showing the top 3 GoF genes and TPS for gene potential in enhancing specific functional characteristics (proliferation, killing, infiltration, persistence, survival, and *IFNG* expression). **D,** Radar plots showing the top 3 LoF genes and TPS for gene potential in enhancing specific functional characteristics (proliferation, killing, infiltration, persistence, survival, and *IFNG* expression). **E,** GO terms enriched from the top 50 GoF and top 50 LoF in CD8$^+$ T cells. The count represents the number of genes in the enriched terms. **F,** Overview of the LASSO model training and testing procedure: a LASSO multivariate model was trained on T-cell genes with top TPSs to predict CAR T-cell putative function. The model used leave-one-out validation to identify the optimized hyperparameters before being applied to an independent CAR T-cell dataset (**G**) ROC curve analysis of LASSO model performance in prediction using the independent CAR T-cell dataset GSE160164. **H,** Bar plot of prediction scores in the testing dataset, categorized by true labels. Statistical significance was assessed using the Wilcoxon rank-sum test. MAE, mean absolute error; TCR, T-cell receptor.

enhancing specific functions such as proliferation, cytotoxic killing, tumor infiltration, persistence, survival, and *IFNG* expression. The top 3 activation genes, *HSPA8*, *ALDOA*, and *RPL5*, all showed a high TPS for proliferation. *HSPA8* and *ALDOA* had a high TPS in proliferation, persistence, and survival, whereas *RPL5* also showed enhanced cytotoxic killing (**Fig. 3C**). For the top three knockout genes, *TP53*, *CBLB*, and *ZC3H12A*, inactivation of all three genes was associated with enhanced proliferation and persistence. Additionally, *TP53* was associated with survival, and *CBLB* was associated with *IFNG* expression (**Fig. 3D**). Pathway enrichment analysis of the top 50 TPS activation genes showed that the top activation genes were mainly enriched in RNA processing– and translation–related pathways (**Fig. 3E**). In contrast, the top knockout genes were mainly enriched in immune signaling and cell proliferation pathways, including immune response, T-cell receptor signaling pathway, cell adhesion, intracellular signal transduction, regulation of the cell cycle, and negative regulation of cell proliferation (**Fig. 3E**). Similarly, we applied the TPS model to $CD4^+$ T-cell screening data, and pathway enrichment analysis revealed that both top activation and knockout genes were enriched in T-cell receptor signaling and immune response pathways (Supplementary Fig. S2A–S2C).

We noticed that some genes showed a high TPS in both $CD8^+$ and $CD4^+$ T cells. The top 50 TPS activation–associated genes included *CD247*, *LAT*, *LCP2*, *RELA*, *RPS5*, and *VAV1* (Supplementary Fig. S2D). The top 50 TPS inactivation–associated genes included *CBLB*, *MAP4K1*, *SLA2*, and *UBASH3A*. Permutation testing confirmed that overlaps involving two or more genes occur at a significantly higher frequency than expected by random chance. Furthermore, we tested the overlapping genes in two CAR T-cell bulk RNA-seq datasets, GSE160154 (CAR T-cell exhaustion) and GSE223655 (CAR T-cell response), and observed that the shared activation gene *RPS5* was significantly lower in exhausted CAR T cells compared with controls, whereas the other five activation genes also showed a trend of decreased expression in exhausted CAR T cells [Supplementary Fig. S2E (top)]. Additionally, *LAT* and *RELA* showed a statistically significant increase in CAR T cells from responders compared with nonresponders [Supplementary Fig. S2E (bottom)]. Among the shared inactivation-associated genes, none were significant in exhausted CAR T cells, but *CBLB* expression was increased in responder CAR T cells (Supplementary Fig. S2F).

To facilitate the prediction of CAR T-cell therapy outcomes, we developed a multivariate model on the top 50 GoF and LoF genes with the highest TPSs from $CD8^+$ and $CD4^+$ T-cell screens. Specifically, we trained a LASSO logistic regression model using the GSE223655 dataset, which includes data from 12 responders and 11 nonresponders to CAR T-cell therapy. We optimized the model parameters using LOOCV (**Fig. 3F**). We then tested the model's performance on an independent test dataset, GSE160154, which consists of 9 persistent and 13 exhausted CAR T-cell samples. The regression model demonstrated robust power in predicting CAR T-cell functionality with an AUC of 0.812 in the testing dataset (**Fig. 3G**). A Wilcoxon test revealed a significant difference in prediction scores between exhausted and persistent CAR T-cell samples (**Fig. 3H**). The successful validation of this TPS gene-based model underscores the important roles these genes play in CAR T-cell function despite variations in CAR design.

TCPGdb web interface and functional modules

The datasets and analyses described above were integrated into the TCPGdb web interface. The website consists of three main functional components: T-cell CRISPR screen view, TPS, and T-cell gene function explorer (T-GEXP). The CRISPR screen module includes summaries of gene hits from 35 T-cell screen datasets and GSEA BP pathway term enrichments (**Fig. 4A**). The TPS module provides a scoring system for evaluating the impact of specific gene perturbations on both overall and specific T-cell functions by integrating results from multiple CRISPR screen datasets (**Fig. 4B**). We also provide a radar chart indicating $CD8^+$ T-cell functions in different aspects such as proliferation, survival, killing ability, infiltration, persistence, and expression of *IFNG* or *IL2* (**Fig. 4B**).

To provide a comprehensive view of gene function specifically for T cells, we present the T-GEXP module for exploring gene expression within different T-cell subtypes, network analysis of coexpressed genes and pathways, and links to T cell and CAR T cell–related pathologies using existing clinical and experimental data (**Fig. 4C**). The query will return five main characteristics of the gene, including (i) query gene expression: expression profiles across 16 T-cell subtypes using sorted T-cell transcriptomic data from DICE (30) and ABIS (31); (ii) T cell–specific coexpression (network): coexpressed gene and pathway analysis using DICE T-cell gene expression data ($n = 967$); (iii) gene and CAR T-cell functions: analysis of gene impact on exhaustion and therapy responses using 12 CAR T-cell datasets (Supplementary Table S2), including four scRNA-seq and eight bulk RNA-seq datasets; (iv) gene and T-cell cancer prognosis: prognostic analysis for T cell–derived cancers using three T-cell lymphoma datasets such as GSE19069, GSE58445, and GSE90597 that reflect the potential of the query gene in enhancing T-cell vigor and aggressiveness, inspired by Garcia and colleagues (52), who discovered that naturally occurring T-cell mutations can enhance engineered T-cell functions; and (v) gene and autoimmune disorders: analysis using transcriptomic profiles from PBMCs in three autoimmune diseases (MS, IBD, and SLE) to explore associations between specific genes and autoimmune diseases, addressing potential side effects of T-cell therapies. The T-GEXP search function is accessible via the T-GEXP search page or the quick search box located at the top left corner of each page, making it convenient for users to retrieve information related to T-cell therapies. Our toolkit is designed to support biologists in effectively exploring the functional profiles of target genes across heterogeneous datasets, as demonstrated by the visualization of the top 50 TPS-ranked genes from CAR T-cell datasets and their relevance to T-cell lymphoma outcomes (Supplementary Fig. S3A–S3D).

Discussion

TCPGdb represents a pioneering and comprehensive resource designed for the analysis and visualization of T-cell perturbation genomics within a single platform. We expect this resource will be beneficial for the scientific community to easily access and explore prior screen-based knowledge on the T-cell function–related gene targets. TCPGdb incorporates a wide array of data, including both GoF and LoF screens conducted *in vitro* and *in vivo*, encompassing multiple T-cell subtypes, engineered CAR T-cell screen datasets, and gene expression data from pure T cells and T lymphocyte–derived cancers. Through its database modules, researchers gain an informative gateway to explore the functional profiles of their target genes across various T-cell perturbomics dataset collections.

Although we have endeavored through TCPGdb to construct a comprehensive collection of T-cell screen data encompassing various methodologies and phenotypes, we acknowledge the limitations arising from the cohort sizes of the datasets, which may introduce

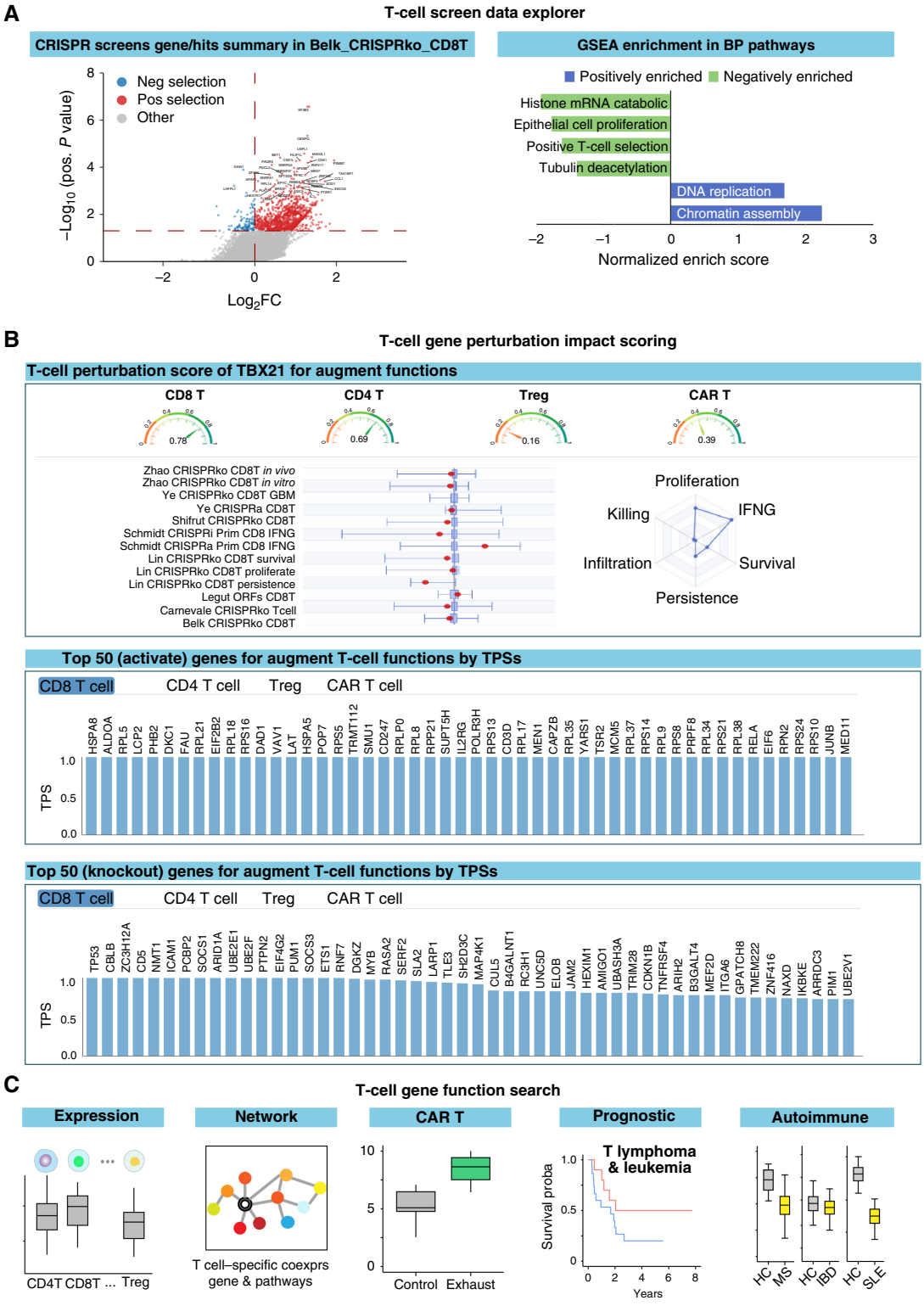

Figure 4.

Overview of the TCPGdb web interface functional modules. **A,** Gene summary and GSEA pathway analysis of T-cell screen data included in this study. **B,** Representative TPS page scoring system for an example gene. **C,** T-GEXP search module in TCPGdb, including expression and coexpression, CAR T-cell performance association, T cell–derived cancer prognosis, and autoimmune disease association. FC, fold change; Neg, negative; Pos, positive.

biases stemming from experimental conditions. We are aware of the challenges inherent in data heterogeneity, particularly when attempting to uniformly process and integrate datasets originating from diverse platforms, experimental conditions, or species. Despite these concerns, we do not expect the data processing approach to introduce substantial bias, as it is based on log fold change values, which are generally robust to variations in processing strategies. Nevertheless, log fold changes can still be affected by factors such as library design, sgRNA efficiency, and the number of sgRNAs targeting each gene. These factors, in turn, can impact the resulting TPSs. Additionally, for the general function TPS, we lack a standardized approach to summarizing scores that effectively represent the overall impact of specific gene perturbations while integrating diverse T-cell fitness phenotypes. We also acknowledge that the limited availability of CAR T-cell datasets with both clinical outcomes and transcriptomic profiles presents a significant constraint for predictive model training and further limits its applicability.

As the number of CRISPR screens conducted to investigate the effect of gene perturbation on T-cell function continues to grow, we are also committed to continuously updating TCPGdb with new datasets. Our plan includes biannual updates to our database to incorporate all newly available datasets. Further datasets will increase the predictive accuracy of TCPGdb and potentially allow analysis of additional phenotypic characteristics beyond proliferation, survival, persistence, infiltration, killing, and *IFNG* expression. Our web interface is publicly accessible and allows researchers with a limited bioinformatics background to easily search for their gene of interest. Furthermore, we will introduce additional modules to facilitate seamless querying of T-cell gene functionalities. TCPGdb stands poised as a valuable resource for studying T-cell perturbation genomics and providing support to the rapidly evolving field of adoptive immune cell therapy.

Data Availability

The data analyzed in this study were obtained from previously published articles. We have included references for each of the datasets used, and all datasets are publicly available. The sources of all datasets are shown in Supplementary Tables. The source code for the web server, as well as the processed data and modeling scripts, is available on GitHub (https://github.com/cpdong/TCPGdb).

Authors' Disclosures

D. Feng reports personal fees from Boehringer Ingelheim during the conduct of the study as well as personal fees from Boehringer Ingelheim outside the submitted work. D. Feng and J. Cai report employment with Boehringer Ingelheim Pharmaceuticals. S. Chen is a (co)founder of EvolveImmune, Cellinfinity, MagicTime, and Chen Consulting, unrelated to this study. No disclosures were reported by the other authors.

Authors' Contributions

C. Dong: Conceptualization, data curation, software, formal analysis, validation, investigation, visualization, writing–original draft. **F. Zhang:** Data curation, formal analysis, writing–review and editing. **K. Tang:** Software, writing–review and editing. **N. Verma:** Validation. **X. Zhu:** Supervision, project administration, writing–review and editing. **D. Feng:** Resources, supervision, funding acquisition, methodology, project administration, writing–review and editing. **J. Cai:** Supervision, writing–review and editing. **H. Zhao:** Resources, supervision, methodology, project administration, writing–review and editing. **S. Chen:** Conceptualization, resources, supervision, funding acquisition, project administration, writing–review and editing.

Acknowledgments

We thank all members of the Chen laboratory, as well as various colleagues in Yale Genetics, Systems Biology Institute, Immunobiology, Yale Cancer Center, and Biomedical Informatics & Data Science for assistance and/or discussions. TCPGdb was built by integrating public CRISPR screen, RNA-seq, and microarray datasets. We sincerely acknowledge all researchers and participants for sharing their data. Special thanks to Dr. Javeed Iqbal for providing pathologic classification and survival information for GSE19069. All the contributors are listed on our website. S. Chen is supported by NIH (R33CA281702), Department of Defense (W81XWH-21-1-0514, HT94252310472), Cancer Research Institute Lloyd J. Old STAR Award (CRI4964), and Sontag Foundation (DSA). C. Dong is supported by the Boehringer Ingelheim Biomedical Data Science Fellowship. N. Verma is supported by the American Board of Radiology's B. Leonard Holman Research Pathway Fellowship and ASTRO Seed Grant.

Note

Supplementary data for this article are available at Cancer Immunology Research Online (http://cancerimmunolres.aacrjournals.org/).

Received February 8, 2025; revised June 11, 2025; accepted November 19, 2025; posted first November 21, 2025.

References

1. Zhang X, Zhang H, Lan H, Wu J, Xiao Y. CAR-T cell therapy in multiple myeloma: current limitations and potential strategies. Front Immunol 2023;14:1101495.
2. Sterner RC, Sterner RM. CAR-T cell therapy: current limitations and potential strategies. Blood Cancer J 2021;11:69.
3. Gumber D, Wang LD. Improving CAR-T immunotherapy: overcoming the challenges of T cell exhaustion. EBioMedicine 2022;77:103941.
4. Cheng J, Lin G, Wang T, Wang Y, Guo W, Liao J, et al. Massively parallel CRISPR-based genetic perturbation screening at single-cell resolution. Adv Sci 2023;10:2204484.
5. Shifrut E, Carnevale J, Tobin V, Roth TL, Woo JM, Bui CT, et al. Genome-wide CRISPR screens in primary human T cells reveal key regulators of immune function. Cell 2018;175:1958–71.e1915.
6. Dong MB, Wang G, Chow RD, Ye L, Zhu L, Dai X, et al. Systematic immunotherapy target discovery using genome-scale in vivo CRISPR screens in CD8 T cells. Cell 2019;178:1189–204.e1123.
7. Schmidt R, Steinhart Z, Layeghi M, Freimer JW, Bueno R, Nguyen VQ, et al. CRISPR activation and interference screens decode stimulation responses in primary human T cells. Science 2022;375:eabj4008.
8. Wang D, Prager BC, Gimple RC, Aguilar B, Alizadeh D, Tang H, et al. CRISPR screening of CAR T cells and cancer stem cells reveals critical dependencies for cell-based therapies. Cancer Discov 2021;11:1192–211.
9. Loo C-S, Gatchalian J, Liang Y, Leblanc M, Xie M, Ho J, et al. A genome-wide CRISPR screen reveals a role for the non-canonical nucleosome-remodeling BAF complex in Foxp3 expression and regulatory T cell function. Immunity 2020;53:143–57.e148.
10. Dai X, Park JJ, Du Y, Na Z, Lam SZ, Chow RD, et al. Massively parallel knock-in engineering of human T cells. Nat Biotechnol 2023;41:1239–55.
11. Lin C-P, Levy PL, Alflen A, Apriamashvili G, Ligtenberg MA, Vredevoogd DW, et al. Multimodal stimulation screens reveal unique and shared genes limiting T cell fitness. Cancer Cell 2024;42:623–45.e610.
12. Zhao H, Liu Y, Wang L, Jin G, Zhao X, Xu J, et al. Genome-wide fitness gene identification reveals Roquin as a potent suppressor of CD8 T cell expansion and anti-tumor immunity. Cell Rep 2021;37:110083.
13. Szeto AC, Ferreira AC, Mannion J, Clark PA, Sivasubramaniam M, Heycock MW, et al. An αvβ3 integrin checkpoint is critical for efficient Th2 cell cytokine polarization and potentiation of antigen-specific immunity. Nat Immunol 2023;24:123–35.
14. Ye L, Park JJ, Peng L, Yang Q, Chow RD, Dong MB, et al. A genome-scale gain-of-function CRISPR screen in CD8 T cells identifies proline metabolism as a means to enhance CAR-T therapy. Cell Metab 2022;34:595–614.e514.
15. Arce MM, Umhoefer JM, Arang N, Kasinathan S, Freimer JW, Steinhart Z, et al. Central control of dynamic gene circuits governs T cell rest and activation. Nature 2025;637:930–9.
16. Freimer JW, Shaked O, Naqvi S, Sinnott-Armstrong N, Kathiria A, Garrido CM, et al. Systematic discovery and perturbation of regulatory genes in human T cells reveals the architecture of immune networks. Nat Genet 2022;54:1133–44.

17. Legut M, Gajic Z, Guarino M, Daniloski Z, Rahman JA, Xue X, et al. A genome-scale screen for synthetic drivers of T cell proliferation. Nature 2022; 603:728–35.
18. Wei J, Long L, Zheng W, Dhungana Y, Lim SA, Guy C, et al. Targeting REGNASE-1 programs long-lived effector T cells for cancer therapy. Nature 2019;576:471–6.
19. Long L, Wei J, Lim SA, Raynor JL, Shi H, Connelly JP, et al. CRISPR screens unveil signal hubs for nutrient licensing of T cell immunity. Nature 2021;600: 308–13.
20. Tsherniak A, Vazquez F, Montgomery PG, Weir BA, Kryukov G, Cowley GS, et al. Defining a cancer dependency map. Cell 2017;170:564–76.e516.
21. Oughtred R, Stark C, Breitkreutz B-J, Rust J, Boucher L, Chang C, et al. The BioGRID interaction database: 2019 update. Nucleic Acids Res 2019;47: D529–41.
22. Choi A, Jang I, Han H, Kim M-S, Choi J, Lee J, et al. iCSDB: an integrated database of CRISPR screens. Nucleic Acids Res 2021;49:D956–61.
23. Cui Y, Cheng X, Chen Q, Song B, Chiu A, Gao Y, et al. CRISP-view: a database of functional genetic screens spanning multiple phenotypes. Nucleic Acids Res 2021;49:D848–54.
24. Luo C, Zhang R, Guo R, Wu L, Xue T, He Y, et al. Integrated computational analysis identifies therapeutic targets with dual action in cancer cells and T cells. Immunity 2025;58:745–65.e749.
25. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 2002;30: 207–10.
26. Li W, Xu H, Xiao T, Cong L, Love MI, Zhang F, et al. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. Genome Biol 2014;15:1–12.
27. Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, et al. BioMart and bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics 2005;21:3439–40.
28. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 2018;34:i884–90.
29. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013;29:15–21.
30. Schmiedel BJ, Singh D, Madrigal A, Valdovino-Gonzalez AG, White BM, Zapardiel-Gonzalo J, et al. Impact of genetic polymorphisms on human immune cell gene expression. Cell 2018;175:1701–15.e1716.
31. Monaco G, Lee B, Xu W, Mustafah S, Hwang YY, Carré C, et al. RNA-seq signatures normalized by mRNA abundance allow absolute deconvolution of human immune cell types. Cell Rep 2019;26:1627–40.e1627.
32. Deng Q, Han G, Puebla-Osorio N, Ma MCJ, Strati P, Chasen B, et al. Characteristics of anti-CD19 CAR T cell infusion products associated with efficacy and toxicity in patients with large B cell lymphomas. Nat Med 2020;26:1878–87.
33. Good CR, Aznar MA, Kuramitsu S, Samareh P, Agarwal S, Donahue G, et al. An NK-like CAR T cell transition in CAR T cell dysfunction. Cell 2021;184: 6081–100.e6026.
34. Wang Y, Tong C, Lu Y, Wu Z, Guo Y, Liu Y, et al. Characteristics of pre-manufacture CD8+ T cells determine CAR-T efficacy in patients with diffuse large B-cell lymphoma. Signal Transduct Target Ther 2023;8:409.
35. Iqbal J, Weisenburger DD, Greiner TC, Vose JM, McKeithan T, Kucuk C, et al. Molecular signatures to improve diagnosis in peripheral T-cell lymphoma and prognostication in angioimmunoblastic T-cell lymphoma. Blood J Am Soc Hematol 2010;115:1026–36.
36. Iqbal J, Wright G, Wang C, Rosenwald A, Gascoyne RD, Weisenburger DD, et al. Gene expression signatures delineate biological and prognostic subgroups in peripheral T-cell lymphoma. Blood J Am Soc Hematol 2014;123:2915–23.
37. Ng S-B, Chung T-H, Kato S, Nakamura S, Takahashi E, Ko Y-H, et al. Epstein-Barr virus-associated primary nodal T/NK-cell lymphoma shows a distinct molecular signature and copy number changes. Haematologica 2018; 103:278–87.
38. Jernås M, Malmeström C, Axelsson M, Nookaew I, Wadenvik H, Lycke J, et al. MicroRNA regulate immune pathways in T-cells in multiple sclerosis (MS). BMC Immunol 2013;14:1–11.
39. Kemppinen A, Kaprio J, Palotie A, Saarela J. Systematic review of genome-wide expression studies in multiple sclerosis. BMJ Open 2011;1: e000053.
40. Gandhi KS, McKay FC, Cox M, Riveros C, Armstrong N, Heard RN, et al. The multiple sclerosis whole blood mRNA transcriptome and genetic associations indicate dysregulation of specific T cell pathways in pathogenesis. Hum Mol Genet 2010;19:2134–43.
41. Srinivasan S, Di Dario M, Russo A, Menon R, Brini E, Romeo M, et al. Dysregulation of MS risk genes and pathways at distinct stages of disease. Neurol Neuroimmunol Neuroinflamm 2017;4:e337.
42. Palmer NP, Silvester JA, Lee JJ, Beam AL, Fried I, Valtchinov VI, et al. Concordance between gene expression in peripheral whole blood and colonic tissue in children with inflammatory bowel disease. PLoS One 2019;14: e0222952.
43. Burczynski ME, Peterson RL, Twine NC, Zuberek KA, Brodeur BJ, Casciotti L, et al. Molecular classification of Crohn's disease and ulcerative colitis patients using transcriptional profiles in peripheral blood mononuclear cells. J Mol Diagn 2006;8:51–61.
44. Mo A, Marigorta UM, Arafat D, Chan LHK, Ponder L, Jang SR, et al. Disease-specific regulation of gene expression in a comparative analysis of juvenile idiopathic arthritis and inflammatory bowel disease. Genome Med 2018;10:1–16.
45. Hung T, Pratt G, Sundararaman B, Townsend M, Chaivorapol C, Bhangale T, et al. The Ro60 autoantigen binds endogenous retroelements and regulates inflammatory gene expression. Science 2015;350:455–9.
46. Kennedy WP, Maciuca R, Wolslegel K, Tew W, Abbas AR, Chaivorapol C, et al. Association of the interferon signature metric with serological disease manifestations but not global activity scores in multiple cohorts of patients with SLE. Lupus Sci Med 2015;2:e000080.
47. Carpintero MF, Martinez L, Fernandez I, Romero ACG, Mejia C, Zang Y, et al. Diagnosis and risk stratification in patients with anti-RNP autoimmunity. Lupus 2015;24:1057–66.
48. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102: 15545–50.
49. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 2012;16:284–7.
50. Gene Ontology Consortium. The gene ontology knowledgebase in 2023. Genetics 2023;224:iyad031.
51. Kolde R, Laur S, Adler P, Vilo J. Robust rank aggregation for gene list integration and meta-analysis. Bioinformatics 2012;28:573–80.
52. Garcia J, Daniels J, Lee Y, Zhu I, Cheng K, Liu Q, et al. Naturally occurring T cell mutations enhance engineered T cell therapies. Nature 2024;626:1–9.