Briefings in Bioinformatics, 2024, **25(5)**, bbae368
https://doi.org/10.1093/bib/bbae368
Problem Solving Protocol

Identifying cell type-specific transcription factor-mediated activity immune modules reveal implications for immunotherapy and molecular classification of pan-cancer

Feng Li‡, Jingwen Wang‡, Mengyue Li‡, Xiaomeng Zhang, Yongjuan Tang, Xinyu Song, Yifang Zhang, Liying Pei, Jiaqi Liu, Chunlong Zhang, Xia Li*, Yanjun Xu*, Yunpeng Zhang*

College of Bioinformatics Science and Technology, Harbin Medical University, 157 Baojian Road, Harbin, China

*Corresponding authors. Yunpeng Zhang, College of Bioinformatics Science and Technology, Harbin Medical University, 157 Baojian Road, Harbin 150081, China. E-mail: zhangyp@hrbmu.edu.cn; Yanjun Xu, College of Bioinformatics Science and Technology, Harbin Medical University, 157 Baojian Road, Harbin 150081, China. E-mail: xuyanjun@hrbmu.edu.cn; Xia Li, College of Bioinformatics Science and Technology, Harbin Medical University, 157 Baojian Road, Harbin 150081, China. E-mail: lixia@hrbmu.edu.cn
‡The first three authors contributed equally to this work.

Abstract

Systematic investigation of tumor-infiltrating immune (TII) cells is important to the development of immunotherapies, and the clinical response prediction in cancers. There exists complex transcriptional regulation within TII cells, and different immune cell types display specific regulation patterns. To dissect transcriptional regulation in TII cells, we first integrated the gene expression profiles from single-cell datasets, and proposed a computational pipeline to identify TII cell type-specific transcription factor (TF) mediated activity immune modules (TF-AIMs). Our analysis revealed key TFs, such as BACH2 and NFKB1 play important roles in B and NK cells, respectively. We also found some of these TF-AIMs may contribute to tumor pathogenesis. Based on TII cell type-specific TF-AIMs, we identified eight CD8+ T cell subtypes. In particular, we found the PD1 + CD8+ T cell subset and its specific TF-AIMs associated with immunotherapy response. Furthermore, the TII cell type-specific TF-AIMs displayed the potential to be used as predictive markers for immunotherapy response of cancer patients. At the pan-cancer level, we also identified and characterized six molecular subtypes across 9680 samples based on the activation status of TII cell type-specific TF-AIMs. Finally, we constructed a user-friendly web interface CellTF-AIMs (http://bio-bigdata.hrbmu.edu.cn/CellTF-AIMs/) for exploring transcriptional regulatory pattern in various TII cell types. Our study provides valuable implications and a rich resource for understanding the mechanisms involved in cancer microenvironment and immunotherapy.

Keywords: single cell; transcriptional regulation; tumor microenvironment; transcription factor

Introduction

The complexity of tumor microenvironment (TME) and immune system is the key factor affecting the development of malignant tumors and the response to treatment of individuals. Therefore, it is of great significance to study the molecular characteristics and regulatory networks of tumor infiltrating immune (TII) cells for elucidating the immune-related mechanisms of malignant tumors, revealing the mechanisms related to immunotherapy response, and discovering new therapeutic targets.

TII cells in the TME may play an anti-tumor or pro-tumor role in the process of tumor occurrence and development [1, 2]. T cell infiltration in cancers is closely associated with disease progression and clinical immunotherapy response [3]. The proportion of TCF7 + CD8+ T cells in the TME of melanoma tumors can be used to predict the outcome of immunotherapy in patients [4]. Liang *et al.* [5] observed a higher proportion of TII cells in local urothelial carcinoma. Ni *et al.*[6] have revealed that infiltrating B cell in TME is associated with the progression of gastric cancer. These studies indicate that the tumor immune microenvironment is highly complex and heterogeneous, and is closely related to the occurrence and development of cancer, survival prognosis and treatment response. However, the underlying molecular mechanisms in these TII cells remain unclear. Therefore, it is important to study the gene regulation mechanism in TII cells and elucidate the influence of TII cells on cancer progression and treatment response.

Transcription factor (TF) regulates the expression of genes and are key molecules in cells. The activity of TF determines not only the function of cells, but also plays an important role in tumor immunity. For example, IRF2 drives CD8+ T cell exhaustion to restrict anti-tumor immunity [7]. FOXP3+ plays an important role in autoimmunity and cancer through regulating T cell heterogeneity and function [8]. In recent years, researchers have developed a series of databases related to TF regulation, which

Received: February 25, 2024. **Revised:** June 11, 2024. **Accepted:** July 15, 2024

provide valuable resources for revealing the complex TF regulation mechanisms in malignant tumors. For example, the TRANSFAC database collected transcriptional regulation data from eukaryotes [9]. The ChIP-X Enrichment Analysis (CHEA) database contains a collection of TF target genes screened from published low-throughput experiments [10]. The TRRUST database stores experimentally validated TF-target gene information extracted from the published literatures [11]. Based on these data resources, researchers conducted a series of studies on transcriptional regulation in tumors. Kibinge *et al.* [12] proposed a TF-centric computational approach to identify transcriptional dysregulation modules in tumors. Liu *et al.* [13] proposed a strategy for construction of individual-specific regulatory networks to identify dysregulated transcription factors (TFs) in each sample . The regulatory processes and functions of TFs in TME cells are quite different. In recent years, the rapid development of sequencing technology has allowed researchers to perform transcriptional level analysis at cell resolution [14]. Aibar *et al.* [15] proposed the 'SCENIC' method for single-cell regulatory network inference and clustering. With the accumulation of cancer single cell transcriptome data, it is necessary to use these data to mine effective markers and explore the related mechanisms of cancer immunity and immunotherapy.

Here, we proposed a computational pipeline for identifying cell type-specific TF-meditated activity immune modules (TF-AIMs) based on dual network restart random walk algorithm and random perturbation strategy. We then integrated the transcription profiles from tumor single-cell datasets and applied our method to identify TF-AIMs specific for six TII cell types. We performed a systematic analysis of these identified TII cell type-specific TF-AIMs and found that the activation frequency of these TF-AIMs can predict immunotherapy response information in tumor patients. In addition, we identified six tumor molecular subtypes based on the activation status of TII cell type-specific TF-AIMs across 9680 tumor samples of The Cancer Genome Atlas (TCGA) database. Finally, we developed CellTF-AIMs (http://bio-bigdata.hrbmu.edu.cn/CellTF-AIMs/), an online database which collected TII cell type-specific TF-AIMs.

Materials and methods

In this study, we integrated single-cell transcriptome data, transcription factor regulation data, and immune marker data to identify immune cell type-specific TF-AIMs in the TME (Fig. 1).

Data collection

Single cell gene expression data

Gene expression of scRNA-seq datasets were retrieved from Gene Expression Omnibus [16] and EMBL-the European Bioinformatics Institute (EBI). In total, 15 tumor single cells RNA sequencing datasets were obtained, which includes 14 datasets from GEO and one dataset from EMBL-EBI (https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-6149/) (Fig. S1). We collected 197 019 single cells from six cancer types (breast cancer, colorectal cancer, glioblastoma, head and neck, lung cancer and melanoma). Furthermore, cell types for 75 245 of these 197 019 single cells have been annotated in the original study (alveolar cells: $n = 1710$; B cells: $n = 6421$; endothelial cells: $n = 1696$; fibroblasts: $n = 1774$; immune cells: $n = 9941$; macrophage: $n = 420$; malignant cells: $n = 2793$; mast cells: $n = 120$; myeloid cells: $n = 9756$; NK cells: $n = 125$; T cells: $n = 40\ 489$) (See online supplementary material for a colour version of this Fig. S1). In this study, cells that have annotated cell type in original study were used for subsequently identifying TII cell type-specific TF-AIMs.

Single cell gene expression of CD4+ helper T cells and CD8+ cytotoxic T cells were downloaded from https://support.10xgenomics.com/single-cell-gene-expression/datasets [17].

Furthermore, two datasets (GSE115978 and GSE120575) contain immunotherapy information (Table S1). The therapeutic response of patient was evaluated using the Response Evaluation Criteria in Solid Tumors (RECIST) guidelines. Patients with either a complete response, partial response, or stable disease were classified as responders, whereas those with progressive disease were deemed non-responders. Specifically, in the GSE115978 and GSE120575 datasets, the post-treatment remaining tumor tissues were identified as resistant tissues.

Expression profiling of tissue sample

We also obtained gene expression datasets at tissue sample level, which contains immunotherapy information. Four datasets including GSE35640, GSE91061, GSE78220 from GEO and PRJEB23709 from European Nucleotide Archive were collected. We also obtained the immunotherapy response information of the samples in the above four datasets. The therapeutic responses were also classified according to the RECIST criteria. Patients with complete response, partial response, or stable disease were categorized as responders, whereas those with progressive disease were deemed non-responders. The detailed treatment information for datasets is shown in Table S1.

In addition, the mRNA expression data of 9680 tumor tissue samples of 32 cancer types were obtained from the TCGA database [18] (See online supplementary material for a colour version of this Fig. S2).

TF regulation and BioPax pathway data

We integrated TF-target gene regulation data from multiple data sources, including TRANSFAC [19], ENCODE [20], CHEA [10], and TRRUST [11]. Datasets were combined into a TF-targets network. We download TRANSFAC, ENCODE, and CHEA datasets from Harmonizonme (http://amp.pharm.mssm.edu/Harmonizome/) [21] and TRRUST datasets from TRRUST web sites (www.grnpedia.org/trrust). Finally, the integrated network contains 913 TFs, 23 626 target genes, and 2 022 650 transcriptional regulatory relationships (See online supplementary material for a colour version of this Fig. S3).

BioPAX [22] format integrated pathways were downloaded from https://www.pathwaycommons.org/archives/PC2/v10/. Finally, we obtained the integrated biological pathway network contains a total of 2 374 707 pairs of interactions between 32 875 gene.

Immune signatures

The Molecular Signatures Database (MSigDB) [23] collected immune signatures which represent cell types, states, and perturbations in the immune system. We downloaded 4872 immune signature gene sets from 389 published immunological studies in the MSigDB database (http://software.broadinstitute.org/gsea/msigdb/collections.jsp#C7) [24], which contained a total of 20 653 genes.

Methods

Identifying candidate immune signature gene sets regulated by TFs

Hypergeometric tests were used to calculate the *P*-value of significance for the regulated associations between TF and immune

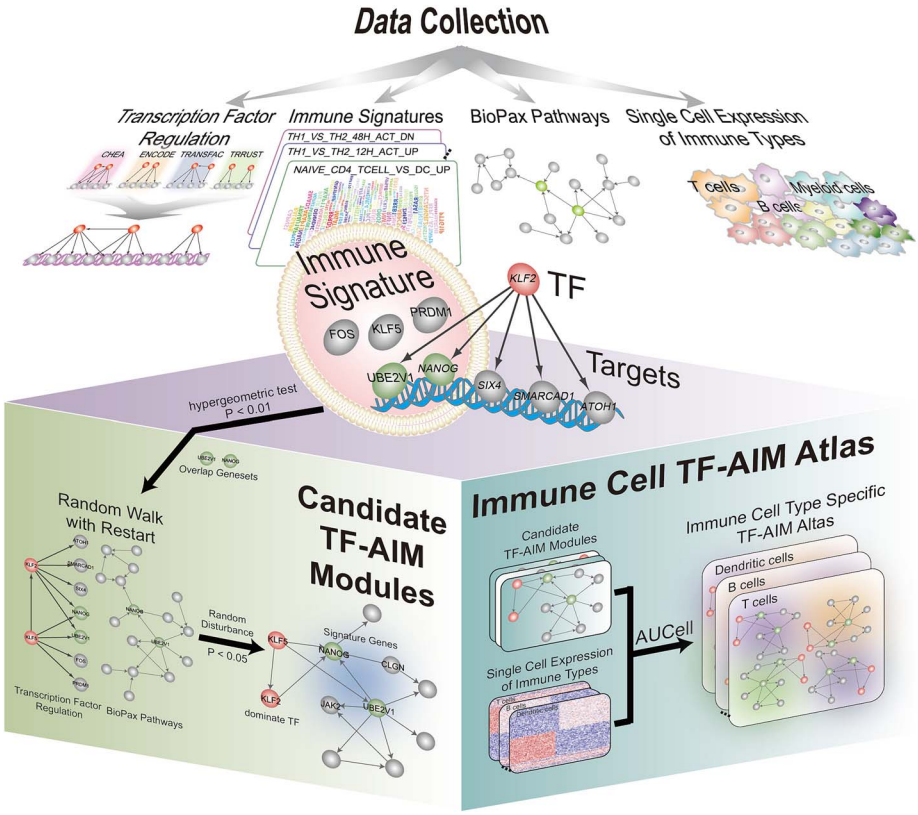

Figure 1. Flowchart for identifying TII cell type-specific TF-AIMs.

signature gene set. For example, if an immune signature gene set is denoted as A, and all target genes of a TF are denoted as set B, then the *P*-value of regulatory significance between the TF and the immune signature gene set is calculated as follows:

$$P(X \geq x) = 1 - \sum_{k=0}^{x-1} \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}$$

Here, x represents the number of overlapping genes of A and B; *N* is the number of background genes (background genes include: i. all target genes in TF regulation data; ii. the sum of genes in all immune signature sets); *n* represents the size of set B; *M* represents the size of set A. For any pair of TF and immune signature set, if the P-value of their hypergeometric enrichment analysis is <.01, the TF is considered to significantly regulate the immune signature set. We extracted their (A and B) overlapping genes for further analysis. This strategy uses target gene enrichment to establish TF-immune module candidate regulation, which can reduce the influence of TF-target relation redundancy and false positive to some extent.

Mining TF-mediated immune modules based on dual network restart random walk algorithm

For each pair of TF and immune signature regulatory relationship with significant hypergeometric test ($P < .01$), overlapping genes were taken as seed nodes, and the dual network restart random walk algorithm was used to score genes in the constructed dual network, and the random perturbation strategy was used to calculate the *P*-value of gene random walk scores in the dual network, so as to identify TF-meditated immune modules.

Firstly, integration of TF regulation and BioPax pathway networks to construct dual network. The dual network contains two networks (TF regulation and BioPax pathway networks) which include the same gene nodes. To construct dual network, if the gene in the transcriptional regulatory (pathway) network is not in the pathway (transcriptional regulatory) network, the gene is added to the pathway (transcriptional regulatory) network, but does not add its connection relationship, thus obtaining two networks with identical gene nodes but different connection relationships.

The genes on the network were scored based on dual network random walk algorithm [25]. For each pair of TF and immune signature, overlapping genes were taken as seed nodes, and then the dual network restart random walk algorithm [25] was used to evaluate the importance scores of all genes on the network except the seed genes.

Next, the significance of gene importance scores was evaluated based on random perturbation strategy. Specifically, (i) 200 networks with the same node degree are randomly generated; (ii) for each pair of TF (A) and immune signature (B), dual network restart random walks are performed on 200 random networks using the same seed nodes (overlapping genes of A and B), and thus each gene has a true score S_g and 200 random scores S_i ($i = 1, \ldots, 200$); (iii) the significance of gene importance scores was calculated as: $P_g = num/200$, *num* represents the number of S_i which is greater than or equal to S_g.

We extracted TF mediated immune modules as follows: (i) all genes on the network are sorted in descending order according to random walk importance score to generate a list of genes $[g_1, g_2, \ldots, g_n]$ (g_1 has the highest score); (ii) according to the significance P-value of gene importance score, the significance list of gene importance score $[I_1, I_2, \ldots, I_n]$ corresponding to the list of genes $[g_1, g_2, \ldots, g_n]$ was generated ($I = 1$ for genes with $P_g < .05$, otherwise $I = 0$); (iii) assuming that the position of TF (A)

in the gene list is m, if $I_m = 0$, the subnetwork composed of TF (A) and seed genes (overlapping genes of A and B) is identified as TF-mediated immune module; if $I_m = 1$ and $\sum_{k=1}^{m} I_k < m$, the TF-mediated immune module is a subnetwork composed of genes $[g_1, g_2, \ldots, g_m]$ and seed genes; if $I_m = 1$ and $\sum_{k=1}^{m} I_k = m$, the TF-mediated immune module is a subnetwork composed of genes $[g_1, g_2, \ldots, g_q]$ and seed genes (q satisfied $I_q = 1$, $I_{q+1} = 0$ and $\sum_{k=1}^{q} I_k = q$).

Based on the above processes, we can extract the TF-mediated immune modules corresponding to each TF. Here, we identified ~150 000 TF-mediated immune modules regulated by 742 TFs (See online supplementary material for a colour version of this Fig. S4).

Evaluating the activity of TF-mediated immune modules in TII cells

The activity score of the TF-mediated immune module in all cells is calculated using the AUCell algorithm [15]. AUCell is a gene-based expression ranking method that uses 'area under the curve' (AUC) to calculate whether a critical subset of the input gene set is enriched in the expressed genes of each cell [15]. It is used to identify cells with active gene signatures in scRNA-seq data. AUCell algorithm explored the threshold of activity score for each TF-mediated immune module, evaluated their activation state in each cell, and obtained the binary activation spectrum of TF-mediated immune modules in cells. In this study, the activity of TF-AIMs was evaluated for all of these 197 019 cells at single-cell resolution. Then, cells that have annotated cell type information from the original study were used for subsequent analysis (See online supplementary material for a colour version of this Fig. S1).

Identifying TII cell type specific TF-AIMs

On the constructed TF-mediated immune module activation profile, there are six TII cell types including B cells, macrophages, mast cells, myeloid cells, NK cells, and T cells. In the case of T cells, for each TF-mediated immune module, determine whether it is T cell-specific: (i) the proportion of TF-mediated immune module activated in T cells is calculated, if the proportion value >30%, the next step is taken; (ii) 100 T cells were randomly selected, and 100 cells were also randomly selected from each of the other cell types respectively; then, Fisher's exact test (unilateral test) was used to check whether the TF-mediated immune module tend to activate in T cells compared with other cell types. In this step, the P values of T cells and other five cell types are obtained. If $P < .05$ for at least four cell types, then variable $x = 1$; otherwise variable $x = 0$; (iii) repeat step (ii) 1000 times, and if the variable $x = 1$ is >800 times, the TF-mediated immune module is considered T cell type-specific TF-AIM.

In addition, we used similar approach to identify CD4+ and CD8+ T cell type-specific TF-AIMs. Due to CD4+ and CD8+ T cells are T cell subsets, more stringent criteria are used here: requiring Fisher's exact test P value <.001 in step (ii); the frequency of variable $x = 1$ was 1000 in step (iii).

Results and discussion

The landscape of TII cell type-specific TF-AIMs

Integration of single cell expression profile, TF-target gene regulation and molecular network, we identified TII cell type-specific TF-AIMs and constructed the landscape of TII cell type-specific TF-AIMs of TME. Cells that have annotated cell type in, see online supplementary material for a colour version of this, Fig. S1 were used for identifying cell type-specific TF-AIMs. As a result, TF-AIMs specific for six TII cell types (B cells, CD4+/CD8+ T cells, myeloid cells, macrophage cells, natural killer [NK] cells, and mast cells) were identified. These TII cell type-specific TF-AIMs were merged to construct global landscape of TII cell type-specific TF-AIMs network in TME (See online supplementary material for a colour version of this Fig. S5). The number of TFs ranged from 7 to 77 across different TII cell type-specific networks.

To explore the regulation of TF in TII cells, we dissected TF-AIMs which contained cell type-related immune signatures that recorded in the MsigDB database [26, 27]. First, we screened the immune-signature gene sets associated with different cell types. For example, B cell associated immune-signature gene sets (immune-signatures) refer that genes in sets may be up/down regulated in B cell by comparing with control (recorded in immunologic signature gene sets C7 of MsigDB). Then, TII cell type-specific TF-AIMs which contain immune-signatures that associated with the corresponding cell type were selected. Finally, these screened TF-AIMs were integrated to form immune signature-specific regulatory networks in different TII cell types (Fig. 2). Dissecting these networks found that TF BACH2 participates in the regulatory network formed by B cell type-specific TF-AIMs. BACH2 is mainly expressed in B and T lymphocytes and controls terminal differentiation and maturation of B lymphocytes [28]. BACH2 directly regulates JUN, BLHM, and SPATS2 genes in the network of B cell type-specific TF-AIMs. The NK cell-associated network is regulated by TF ETV7 and NFKB1 (Fig. 2). NFKB1 functions as an important regulator to mediate cell maturation and effectors of human NK cells [29]. The loss of NFKB1 promotes the expression of tumor necrosis factor (TNF) and activates the TF STAT1, thereby promoting the occurrence of gastric cancer [30]. These evidences confirm that NFKB1 is an important TF in NK cells and has multiple functions. The TF ETV7 was significantly positively correlated with CD8+ T cell infiltration in melanoma, which indicates the regulatory role of ETV7 in TME [31]. However, the regulatory mechanism of ETV7 in tumor NK cells remains unclear. In this study, we found that among the NK cell type-specific TF-AIMs regulated by ETV7, ETV7 directly regulates genes such as JUNB, DGKA, and MAP4K2, which provides important guidance for revealing the role and function of ETV7 in NK cell-mediated natural immune response. In summary, the above analysis results not only demonstrate the reliability of our proposed method for identifying cell type-specific TF-AIMs, but also suggest these cell type-specific TF-AIMs have important significance for revealing the regulatory function of TF in tumor immune microenvironment and to studying the mechanisms of tumorigenesis.

TII cell type-specific TF-AIMs contribute to tumor pathogenesis

To further capture key transcriptional regulation in different cell types, we next screened TII cell type-representative TF-AIMs. First, we counted the activation ratio of TII cell type-specific TF-AIMs in different cell types (Fig. 3A). We found that some TF-AIMs activated with a high frequency in multiple TII cell types. Then, TII cell type- representative TF-AIMs were defined as follows (e.g. myeloid cells): (i) TF-AIMs should be activated in at least 30% of myeloid cells; (ii) the activation ratio of TF-AIMs in myeloid cells are higher (≥two-fold) than at least other four TII cell types. Representative TF-AIMs of myeloid cells and B cells recognized by the above rules are shown in Fig. 3B. TFs such as SP1, MAF, and HSF2 were involved in the representative TF-AIMs of myeloid cells. HSF2 directly regulates GDPD3 gene which displayed key roles in the maintenance of chronic myelogenous leukemia stem cells [32]. Deletion of c-MAF may enhance antitumor T cell immunity via reducing tumor burden [33]. The results further support

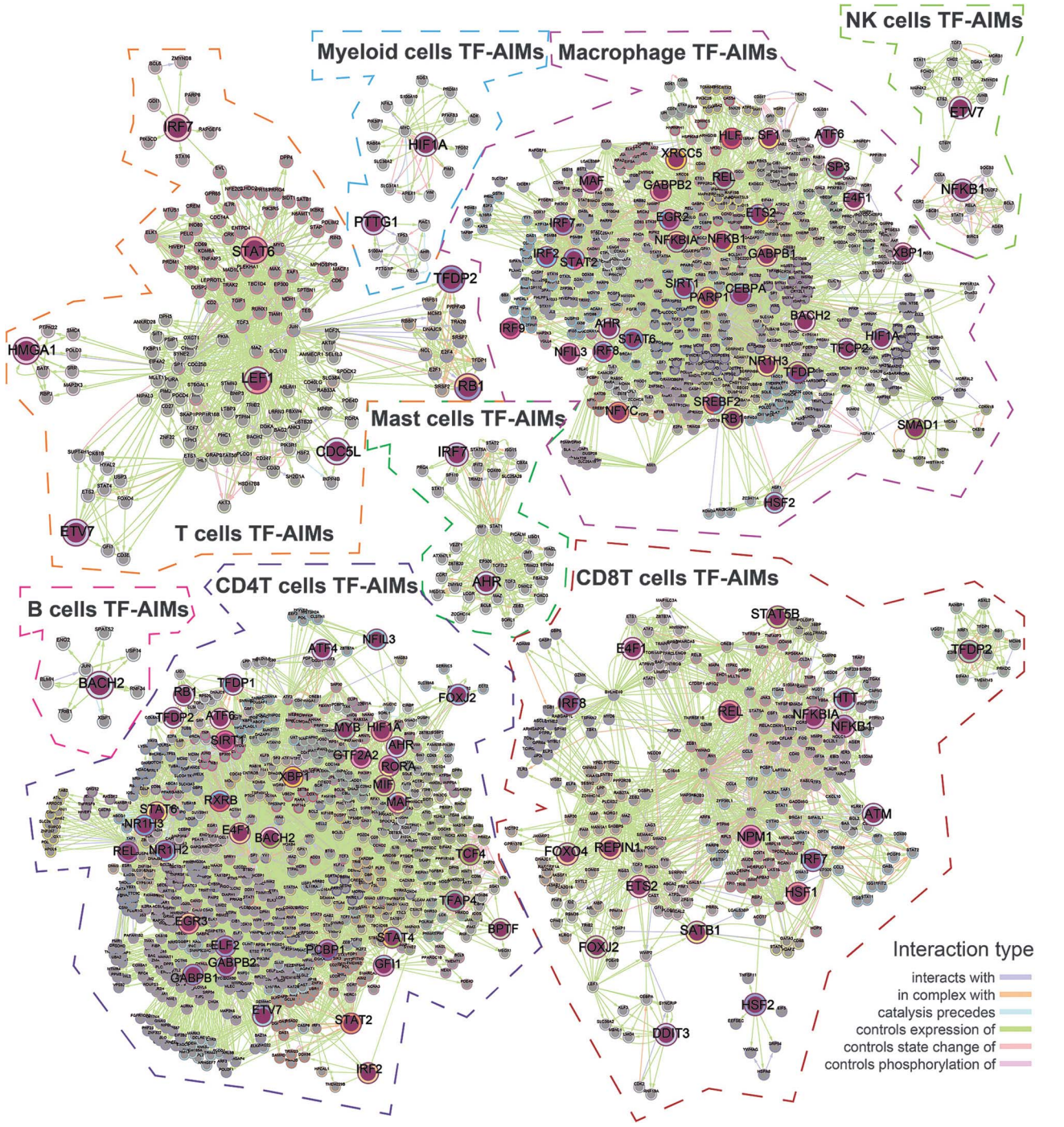

Figure 2. TII cell type-specific TF-AIMs that related with immune features of the corresponding cell type. Big nodes represent TFs; smaller nodes represent other genes in TF-AIMs.

the ability of these identified TF-AIMs to help reveal regulatory mechanisms in tumor immune microenvironment cells. Modules regulated by SP3, SMAD1, ELF2, GABPB1, and MAF are interconnected to form a crosstalk transcription modulation. SP1 is the key gene that links these TF-AIMs. This gene plays an important role in the immune system or anti-tumor immune response. For example, SP1 binds to and activates the promoters of a number of important myeloid genes [34]. We found that SP1 acts as a 'bridge' in myeloid cell type- representative TF-AIMs (Fig. 3B), revealing its central role in the transcriptional regulation of myeloid cells in TME. In the representative TF-AIMs of B cell, BACH2 regulates key genes of tumorigenesis and prognosis, such as TRIB1, SPATS2, and JUN. Especially, SPATS2 related with cell cycle progression and immune cells infiltration in hepatocellular carcinoma [35]. Zhang *et al*. [36] revealed that c-Jun plays a critical role in liver metastasis in human breast cancer model. Kim *et al*. [37] have demonstrated that TRIB1 regulates tumor growth and is associated with breast cancer survival and treatment response. In addition, BACH2 also directly regulates TF XBP1, which can reduce the proliferation and stemness of cancer cells [38]. These two key TFs BACH2 and XBP1 in the B cell type-representative TF-AIMs reveal transcriptional regulatory pathways against tumor proliferation.

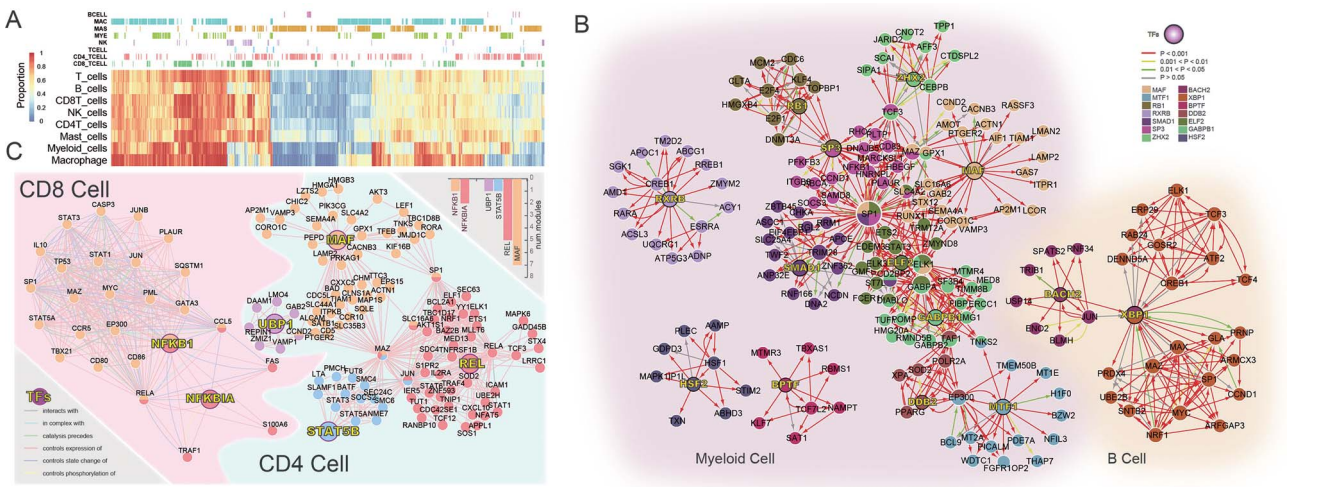

Figure 3. Analysis of TII cell type-representative TF-AIMs. (A) Heatmap of activation status of TII cell type-specific TF-AIMs in different cell types. Each column represents a TF-AIMs, each row represents a type of immune cell, and the heatmap shows the proportion of cells in which the corresponding TF-AIM is activate. (B) the representative TF-AIMs network for myeloid cell and B cell. P values of edges represent significance of co-expression among genes. (C) CD4+/CD8+ T cell type-representative TF-AIMs. Histogram of the number of TF-AIMs regulated by TFs (upper-right).

In this study, T cell type contains the largest number of cells, and CD4+/CD8+ T cells are the dominant cell types of T cell. Using the same strategy, we identified 14 and 3 representative TF-AIMs for CD4+ and CD8+ T cell types, respectively (Fig. 3C). In the CD4+ T cell type-representative TF-AIMs, MAZ, SP1 and JUN genes act as 'bridges' to connect key TFs. MAZ is closely related to tumor proliferation and metastasis [39]. MAF, a TF regulates MAZ, induces CD4+ T cells to produce the anti-inflammatory cytokine IL-10 *in vitro* [40]. Moreover, TF STAT5B, which regulates MAF, is closely related to CD4+ T cells in immune cells [41]. These results not only verified the accuracy of these TII cell type-specific TF-AIMs, but also indicated that CD4+ T cell type-representative TF-AIMs dominated by STAT5 → MAF → MAZ played an important role in the tumor immune microenvironment. In addition, CD4+ T cell type-representative TF-AIMs also contain other well-known cancer-related genes, such as MAPK6, AKT1S1, AKT3 and CDC5L. The CD8+ T cell type-representative TF-AIMs are regulated by two TFs (NFKB1 and NFKBIA). Further analysis found a key regulatory axis: NFKB1 → SP1 → IL10. Rivas *et al.* [42] revealed that suppression of Interleukin-10 (IL10) can enhance T-cell antitumor immunity and responses to immunotherapy in chronic lymphocytic leukemia. Furthermore, cancer genes such as MYC, SP1, and JUN that involved in TF-AIMs of CD4+ T cell also exist in CD8+ T cell type-representative TF-AIMs. The above results suggest that TF could mediate anti-tumor immune response by regulating cancer immune-related factors such as IL10, and also directly regulate cancer-related genes, thus affecting tumorigenesis.

CD4+/CD8+ T cell type-specific TF-AIMs participate in consistent immune pathway

We further performed functional enrichment analysis for CD4+ and CD8 + T cell type-representative TF-AIMs, and found that genes in these TF-AIMs were enriched in important immune-related pathways such as IL27 pathway (M36), IL2 pathway (M122), IL2/STAT5 pathway (M234), and T cell activation (GO:0042110) regulation of inflammatory response (GO:0050727) and leukocyte cell–cell adhesion (GO:0007159) (Fig. 4A). It may indicate that CD4+ and CD8+ T cell type-representative TF-AIMs share some tumor - or immune-related genes that are widely expressed in T cells. In addition, CD8+ T cell type-representative TF-AIMs were more highly enriched in functions such as signaling by interleukins (R-HSA-449147), AP1 pathway (M167), and negative regulation of cell proliferation (GO:0008285). We further analyzed the functional annotation of genes in CD4+/CD8+ T cell type-representative TF-AIMs, and calculated the differential expression of these genes (Fig. 4B and C). Most of genes in CD4+/CD8+ T cell type-representative TF-AIMs were differential expression between normal CD4+ helper T cells and CD8+ toxic T cells (Fig. 4C). This suggests that these CD4+/CD8+ T cell type-representative TF-AIMs is indeed cell type-specific. Furthermore, we found that genes in CD4+ and CD8+ T cell type-representative TF-AIMs tend to be complementary (Fig. 4C). The above results suggest that CD4+ and CD8+ T cells in TME activate distinct subregions of consistent immunomodulation-related functions.

TII cell type-specific TF-AIMs reveal CD8+ T cell subtypes associated with immunotherapy

CD8 + T lymphocytes in the tumor immune microenvironment drive the anti-cancer immune response [43–45]. Immunotherapy, which activates the anti-tumor immune response of T cells, has been widely studied in cancers. Therefore, we further identified CD8+ T cell subtypes based on CD8+ T cell type-specific TF-AIM. Firstly, based on the activation status of 201 CD8+ T cell type-specific TF-AIMs in each cell, we used UMAP for dimensionality reduction. UMAP is a nonlinear dimensionality reduction technique for analyzing high-dimensional data. It is able to discern subtle cell population differences while preserving local and global structure in the data [46]. Then, all CD8+ T cells (mainly from GSE111894, GSE115978, GSE120575 and GSE123139 datasets) were unsupervised clustered by k-medoids algorithm. As a result, eight CD8+ T cell subtypes were identified (Fig. 5A). It was found that the distribution of CD8+ T cell subtypes in tissue samples was heterogeneous (See online supplementary material for a colour version of this Fig. S6). Figure 5B shows the normalized expression rank scores of T cell marker genes in each subtype. According to the expression of T cell marker gene in different cell subtypes, we found that the expression rank of PDCD1 (PD1), CD8A and CD8B in G2, G3, G5, G7, and G8 was high, and cells of these subtypes were defined as PD-1 + CD8+

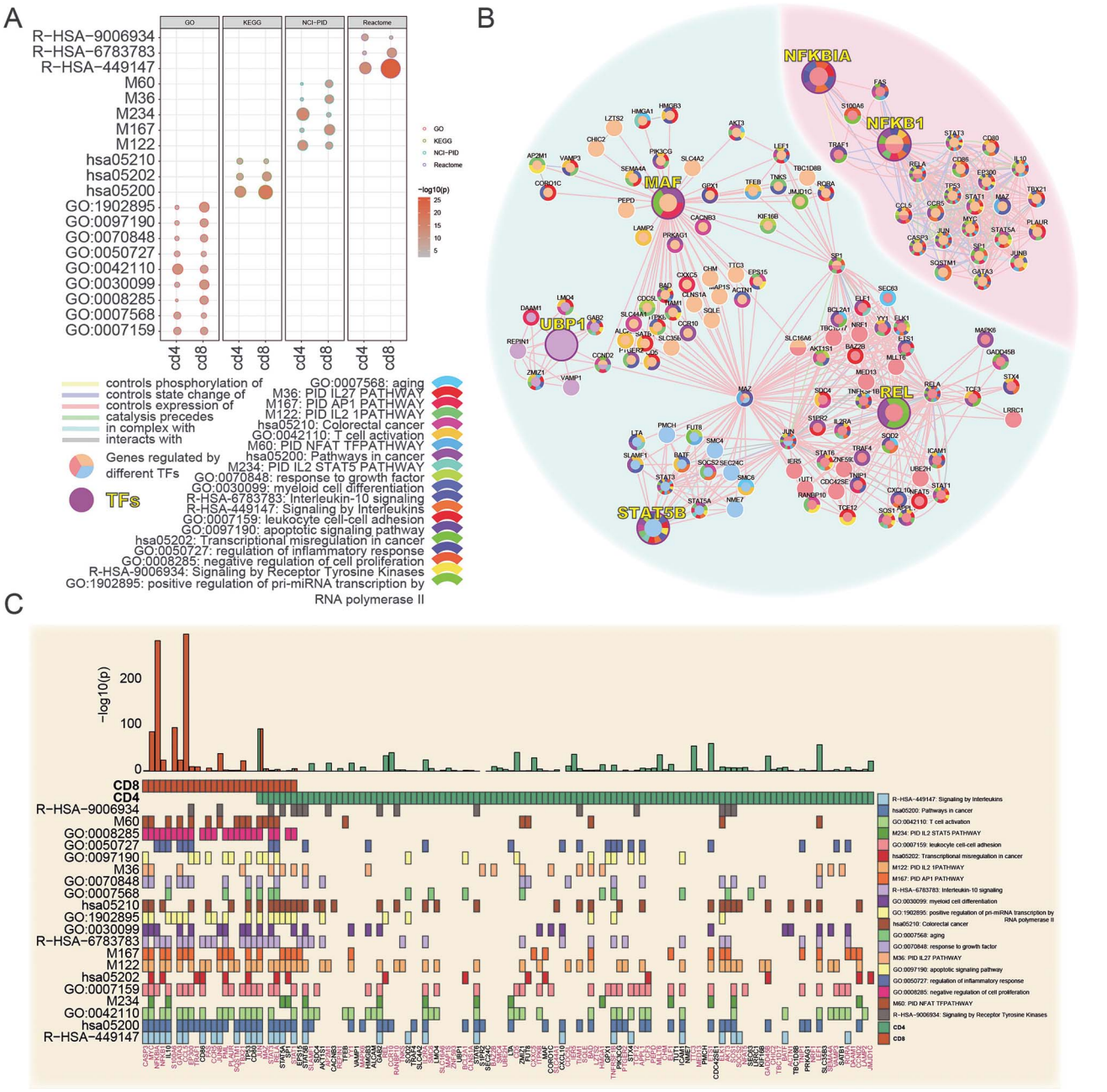

Figure 4. CD4 +/CD8 + T cell type-representative TF-AIMs function analysis. (A) Common enrichment pathways for CD4+ and CD8+ T cell type-representative TF-AIMs. Bubble size represents negative logarithm of function enrichment *P* value, enrichment function data source: GO; KEGG (Kyoto Encyclopedia genes and genomes); NCI-PID (the NCI-nature pathway interaction) and Reactome. (B) Networks of CD4+ and CD8+ T cell type-representative TF-AIMs. (C) Function annotation of genes in CD4 + T cells and CD8+ T cell representative modules. The upper bar chart shows the negative logarithm of *P* value of the differential expression of genes between CD4+ helper T cells and CD8+ toxic T cells. The font of significantly differentially expressed genes (Wilcoxon sign rank test $P < 0.05$) was purple.

T cell (Fig. 5B). Studies have shown that PD-1 + CD8+ T cells proliferation in peripheral blood of tumor patients after receiving PD-1 targeted therapy [47], and PD-1 + CD8+ T cells have predictive potential for immunotherapy effect in PD-1 block therapy samples [48, 49]. These results suggest that PD-1 + CD8+ T cell subtype-specific TF-AIMs may relate to the effect of immunotherapy. We calculated the proportion of CD8+ T cell type-specific TF-AIMs activated in each CD8+ T cell subtype (Fig. 5C). In PD-1 + CD8+ T cell subtypes, TF-AIMs regulated by IRF8, NR1H3, IRF7, and NFKBIA had a very high activation frequency (Fig. 5C), while in other cell subtypes, TF-AIMs regulated by these TFs had a low activation frequency.

Next, we analyzed the association of the above PD-1 + CD8+ T cell subtype-specific TF-AIMs with immunotherapy based on immunotherapy response/non-response information of tumor patients from GSE115978 and GSE120575. Patients receiving immunotherapy can be divided into three groups: immunotherapy-responsive (R), non-responsive (NR), and resistance groups (see materials and methods). We compared the activation ratio of IRF8, NR1H3, IRF7, and NFKBIA-regulated TF-AIMs in PD-1 + CD8+ T cell subtypes of different groups of patients (Fig. 5D). The results showed that the proportion of IRF7 and NFKBIA-regulated TF-AIMs activation was higher in the immunotherapy-resistant samples and NFKBIA-regulated

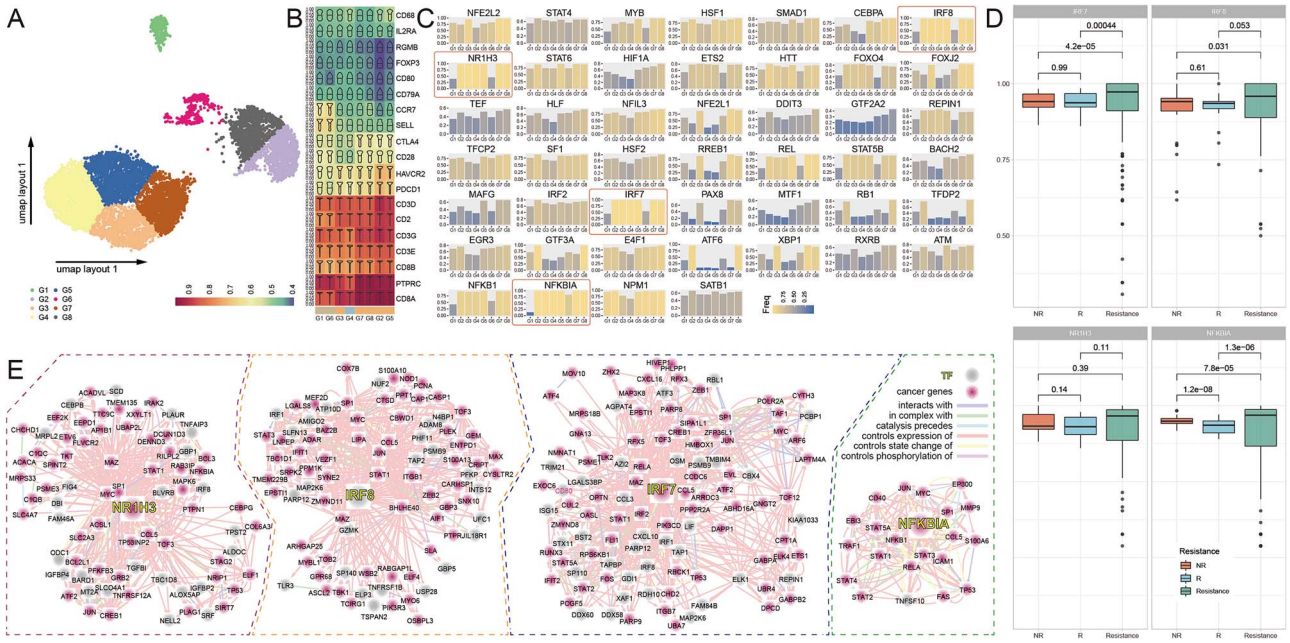

Figure 5. Analysis of CD8+ T cell subtype-specific TF-AIMs. (A) the UMAP view of CD8+ T cell subtypes. (B) Rank normalized expression of T cell markers in eight subtypes. (C) Histogram of activation ratios of CD8+ T cell type-specific TF-AIMs in different subtypes. (D) Proportion of TF-AIMs activation in PD-1 + CD8+ T cell subtypes in immunotherapy patients. Rank sum test was used to evaluate the significance P-value. (E) TF-AIMs network for IRF8, NR1H3, IRF7 and NFKBIA in PD-1+ CD8+ T cell subtypes.

TF-AIMs activation was higher in the immunotherapy NR samples compared with these response (R) samples. TF-AIMs commonly activated in PD-1 + CD8+ T cells contain many tumor driver genes (Fig. 5E). For example, TFs in these TF-AIMs all regulate widely expressed cancer related genes such as CCL5, SP1, JUN, and MYC. We also found that genes in these TF-AIMs also play important roles in cancer immunotherapy. In the IRF7 related TF-AIMs, the IRF7-STAT1-CXCL10 axis may relate with CD8+ T cell infiltration and immune evasion in cancer [50]. CCL5, which is the gene exist in all these four TF-AIMs has been nominated as biomarkers for immunotherapy of cancer [51]. These results suggest that PD-1 + CD8+ T cell subtype-specific TF-AIMs have the potential to be used as cancer immunotherapy-related markers, and also can help to reveal the relevant mechanisms of cancer immunotherapy.

TII cell type-specific TF-AIMs predict immunotherapy response of cancer

In order to systematically analyze the relevance of TII cell type-specific TF-AIMs in immunotherapy, we calculated the activation frequency of TF-AIMs in cells of immunotherapy patients from GSE115978 and GSE120575. Immunotherapy response-related TF-AIMs were defined as the activation frequency of which were significantly differed between the immunotherapy responsive group (R) and the non-responsive group (NR and resistance) (Fig. 6A). Figure 6A shows TII cell type-specific TF-AIMs with activation frequency changed by more than 1.5 times. These immunotherapy response-associated TF-AIMs are regulated by TFs such as XBP1, REL, MAFG, STAT5B, NFKB, and IRF7. TF-AIMs regulated by TFs such as XBP1 and NFKB1 tended to be activated in the immunotherapy non-responsive group, while TF-AIMs regulated by TFs such as REL and STAT5B tended to be activated in the immunotherapy responsive group. To further verify these immunotherapy related TF-AIMs, we collected four immunotherapy related datasets (GSE91061, GSE78220, GSE35640, and PRJEB23709), in which samples were classified as responders and non-responder (see materials and methods). We calculated the differential expression of genes in these immunotherapy-associated TF-AIMs between immunotherapy responder and non-responder groups (Fig. 6B). The results showed that genes in immunotherapy-associated TF-AIMs were differentially expressed in independent datasets at the tumor tissue/sample level, further confirming the association between these TII cell type-specific TF-AIMs and immunotherapy.

Next, we dissected the regulatory associations in these immunotherapy related TF-AIMs (Fig. 6C). We found that multiple TF-AIMs regulated by TFs such as IRF7, STAT5B, and NFKBIA are relevant to immunotherapy, and these TF-AIMs can crosstalk with each other. For example, the two TF-AIMs regulated by NFKBIA both contain the regulatory relationship between NFKBIA and genes such as CCL5, STAT1, and NFKB1. Therefore, we constructed a TF-AIMs crosstalk network related to immunotherapy response (Fig. 6D). Functional analysis showed that these immunotherapy related TF-AIMs are generally enriched in response to interferon-gamma (GO:0034341), myeloid differentiation (GO:0030099:) and leukocyte differentiation (GO:0002521). In addition, genes of STAT6 related TF-AIMs specifically enriched in the IL4 pathway (M28), and genes in NFKB1 regulated TF-AIMs specifically enriched in the IL12/STAT4 pathway (M290) (Fig. 6E and See online supplementary material for a colour version of this Fig. S7).

We then evaluated the predictive power of TII cell type-specific TF-AIMs in Fig. 6A for cancer patients respond to immunotherapy based on activation frequency in TME cells (Fig. 6F). We found that the frequency of activation of IRF7-AIM2, IRF7-AIM3, and NFKBIA-AIM was a good predictor for patient's immunotherapy response (AUC values >0.8). We further used a cross-validation approach to evaluate the predictive efficacy of TF-AIMs for immunotherapy in independent tissue sample expression datasets. The logistic regression coefficient of genes in each TF-AIM to immunotherapy response was calculated in the training set, and then the coefficients were used to predict immunotherapy response of individuals in validation datasets. The predictive efficacy of

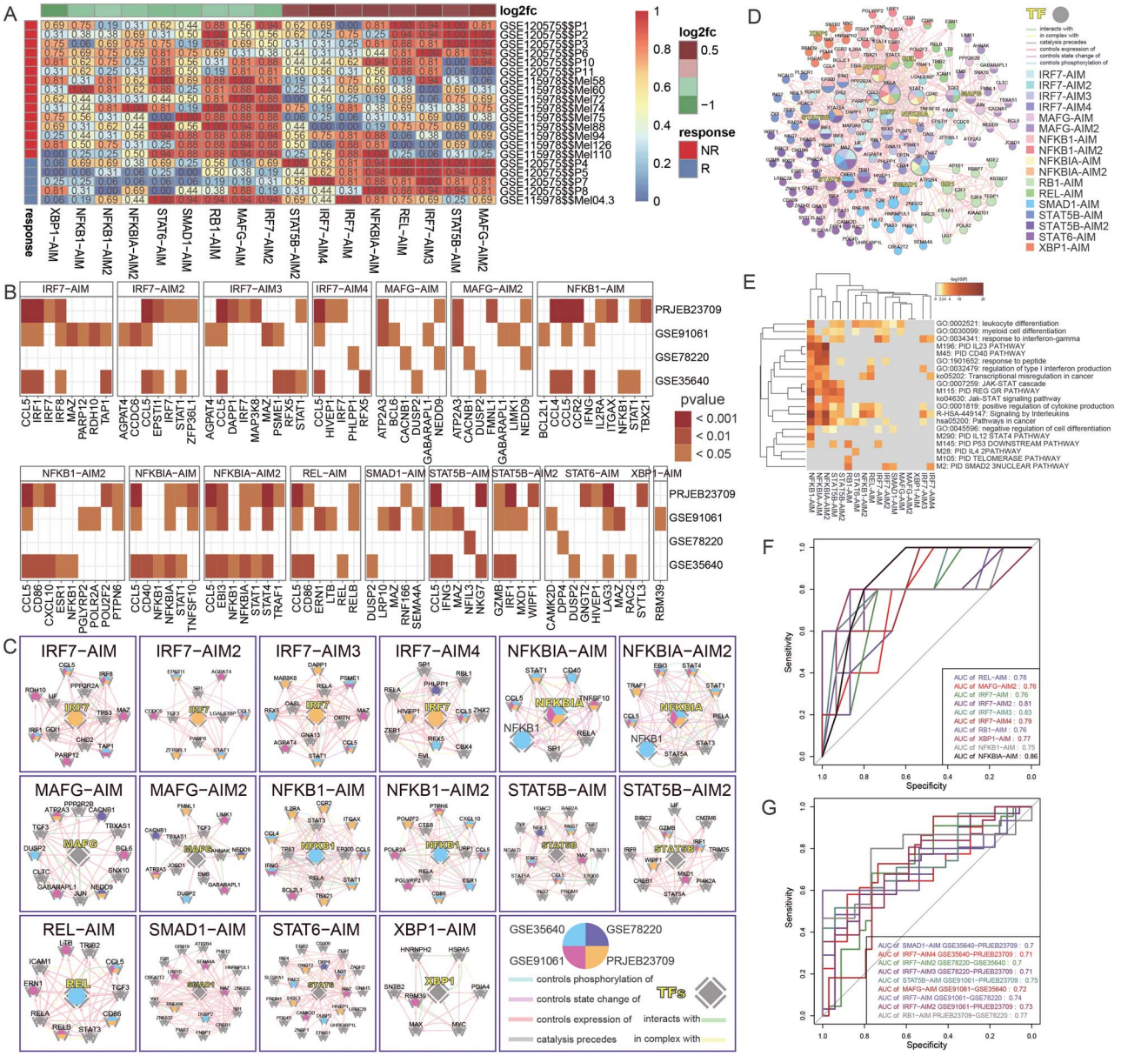

Figure 6. Association analysis of TII cell type-specific TF-AIMs and immunotherapy response. (A) Heatmap of differences for the frequency of TF-AIMs activation in patient cells in the immunotherapy response/non-response group. Rows are patients and columns are TF-AIMs. Only TF-AIMs with activation frequency differences greater than 1.5 times are shown. (B) Heatmaps of differential significance (Wilcoxon symbol rank test p value is < 0.05, 0.01, and 0.001, respectively) of gene expression in TF-AIMs in independent samples. Row: Genes in TF-AIMs; column: The accession of the dataset. (C) The regulatory networks of TF-AIMs associated with immunotherapy response. Pie charts of each gene: Differentially expressed in independently validated datasets. (D) the network of TF-AIMs associated with immunotherapy. (E) Functional enrichment heatmap of immunotherapy associated TF-AIMs. (F) Predictive efficacy of activation frequency of TF-AIMs in TME cells to immunotherapy response. (G) Predictive efficacy of immunotherapy response-related TF-AIMs in validation sets.

immunotherapy response-related TF-AIMs in validation sets are shown in Fig. 6G and See online supplementary material for a colour version of this Fig. S8A. Furthermore, we also evaluated the expression of genes in immunotherapy-associated TF-AIMs for predicting immunotherapy response (See online supplementary material for a colour version of this Fig. S8B). We found that TFs such as IFR7, STAT5B, and NFKB related TF-AIMs exhibited good predictive performance in validation sets (AUC > 0.7) (Fig. 6F and G). These results further demonstrate the potential of TII cell type-specific TF-AIMs as a marker of cancer immunotherapy and facilitate the understanding of the relevant mechanisms of immunotherapy. These TII cell type-specific TF-AIMs complement the current panel of immunotherapy response markers.

TII cell type-specific TF-AIMs contribute to re-classification of pan-cancer

TII cell type-specific TF-AIMs can reflect the regulation and function of TFs in TME. We next dissected the activation status of TII cell type-specific TF-AIMs in TCGA tumor tissue samples. We used the AUCell algorithm to evaluate the activation of TII cell type-specific TF-AIMs in 9680 tumor tissue samples from 32 cancers in TCGA. We found that some TII cell type-specific TF-AIMs were generally activated in tumor tissue samples. TII cell type-specific

TF-AIMs activities in 80% of TCGA tumor samples is shown in, See online supplementary material for a colour version of this, Fig. S9A. The universal activation network contains TF-AIMs specific to multiple TII cell types. TF STAT2 is expressed in CD4+ T cells, myeloid cells, and macrophages, and regulates cell type-specific TF-AIMs. We performed functional enrichment analysis of cell type-specific TF-AIMs in the universal activation network. As a result, TF-AIMs which are generally activated in tumors, are generally enriched in functions such as leukocyte differentiation (GO:0002521), transcriptional dysregulation in cancer (hsa05202), negative regulation of cell proliferation (GO:0008285), B cell activation (GO:0042113), and cellular response to drug (GO:0035690) (See online supplementary material for a colour version of this Fig. S9B and C).

By analyzing the activation status of TII cell type-specific TF-AIMs in each tumor tissue sample, it was found that many TF-AIMs were only activated in some samples. Therefore, we inferred that the activation status of TF-AIMs has the potential to distinguish tumor molecular subtypes. Based on the activation profile of TF-AIMs in TCGA samples of 32 cancer types, we used consistent clustering to identify six TCGA sample subclusters (Fig. 7A). The distribution of clinical characteristics such as age, sex and tumor stage of these six sample clusters is shown in, See online supplementary material for a colour version of this, Fig. S10. Then, the distribution of cancer samples in different clusters was analyzed (Fig. 7B). Breast cancer was further divided into different molecular subtypes (C1, C2). In addition, Pheochromocytoma and Paraganglioma (PCPG), low-grade glioma of the brain (LGG) and Glioblastoma Multiforme (GBM) mainly distributed in the C6 subtype, this suggests that PCPG is similar to brain tumor in immune-related transcriptional regulatory modules. The results showed that the survival time of samples in different subtypes was significantly different (Log rank test: $P < .0001$) (Fig. 7C). The C6 subtype (mainly contains LGG, GBM, and PCPG tumor samples) has the worst survival prognosis, which is consistent with previous reports [52, 53]. The above results indicate that these identified molecular subtypes have biological and clinical significance.

To further characterize these pan-cancer molecular subtypes, we calculated the activation frequency of each TII cell type-specific TF-AIMs (shown in Fig. 7A) in these six molecular subtypes (Fig. 7D). We next identified TCGA subtype-specific TF-AIMs (Fig. 7E). TF-AIMs whose activation frequency is 1.5 times higher in a given subtype than in all other subtypes are defined as subtype-specific. We found macrophage, CD4 + T cell and mast cell type-specific TF-AIMs are universally activated in C2, C3, and C5 subtypes. The TCGA pan-cancer tumor molecular subtype-specific TF-AIMs crosstalk networks are shown in Fig. S11. For each TF, the number of subtype-specific TF-AIMs was counted. It was found that some TFs such as IRF7, ETV7, GABPB1, GABPB2, and REL regulated TF-AIMs in different subtypes (Fig. 7F). These TFs may have differential regulatory mechanisms across different subtypes. Then, we focused on TF-AIMs that are activated in only one pan-cancer molecular subtype (See online supplementary material for a colour version of this Fig. S12). Previous studies have shown that the occurrence of LGG and GBM tumors is generally accompanied by up-regulation of heat shock family genes, and the expression of heat shock family genes is positively correlated with the invasion ability of LGG and GBM [54, 55]. In addition, interference with HSF2/HSPH1 pathway may play an auxiliary role in the efficacy of glioma antitumor drugs [56]. In this study, the C6 subtype mainly contained LGG and GBM tumor tissue samples, and we found highly specific TF-AIMs in the C6 subtype, namely C6-HSF2 (Fig. 7G). In this TF-AIMs, the heat shock TF HSF2 regulates the heat shock family gene HSPH1 and the heat shock TF HSF1(Fig. 7G). These results indicate that these TF-AIMs can help to characterize the molecular characteristics of the corresponding subtypes and further understand the relevant regulatory mechanisms.

CellTF-AIMs: A web interface for exploring TII cell type-specific TF-AIMs

To facilitate the usage of these identified TF-AIMs, we developed CellTF-AIMs (http://bio-bigdata.hrbmu.edu.cn/CellTF-AIMs/), which stores TF-AIMs specific for different TII cell types including B cell, macrophage, mast cell, myeloid cell, NK cell, T cell, CD4+, and CD8+ T cell. In total, CellTF-AIMs documents entries of associations between 2114 TF-AIMs and 8 TII cell types/subtypes. TF-AIMs provide user-friendly interfaces to search or browse these associations. The results of search/browse can be freely obtained. Users can explore TF-gene target relations and the biology function of these TF-AIMs through the network visualization function. Furthermore, we used 'SingleR' package [57] to annotate cells for datasets containing only 'unknown' cell type (See online supplementary material for a colour version of this Fig. S1). CellTF-AIMs collects this annotation information and provides 'analysis' function, users can analyze and visualize the activity of TF-AIMs across all of these 197 019 cells in TME.

Conclusion

In this study, we provided a computational pipeline for identifying TII cell type-specific TF-AIMs. We constructed a transcriptional regulation map of TII cells and systematically analyzed these TII cell type-specific TF-AIMs. Our analysis revealed important roles of BACH2 and NFKB1 in tumor infiltrating B cells and NK cells, respectively. In addition, we also found some of these TF-AIMs may contribute to tumor pathogenesis.

We evaluated the association of TF-AIMs with cancer immunotherapy response and found the activation frequency of these TF-AIMs can predict immunotherapy responses of tumor patients. This suggests the potential of TII cell type-specific TF-AIMs as a predictive marker of cancer immunotherapy response. We then identified six tumor molecular subtypes based on activation states of TII cell type-specific TF-AIMs across 32 cancers. The above indicates that these TF-AIMs we identified have application significance. Finally, in order to further promote the application of these TII cell type-specific TF-AIMs, we build a CellTF-AIMs database (http://bio-bigdata.hrbmu.edu.cn/CellTF-AIMs/).

In summary, we proposed a computational approach to identify and analyze cell type-specific TF-AIMs in TME, and screen out TF-AIMs that are predictive of immunotherapy response. Finally, we identified novel tumor molecular subtypes based on the activation status of TII cell type-specific TF-AIMs in TCGA tumors. The research of this study has important theoretical significance and application value for understanding the transcriptional regulation mechanism of tumor-infiltrating immune cells and realizing precision treatment of tumor patients.

Limitations, uniques and future prospects

Aibar *et al.* [15] provides SCENIC method to infer single cell gene regulatory network. This method has been widely used. Then, the updated version of SCENIC+ incorporates single-cell ATAC-seq (scATAC-seq) data for constructing TF regulatory networks [58]. In the current version of our pipeline, the gene activity matrix

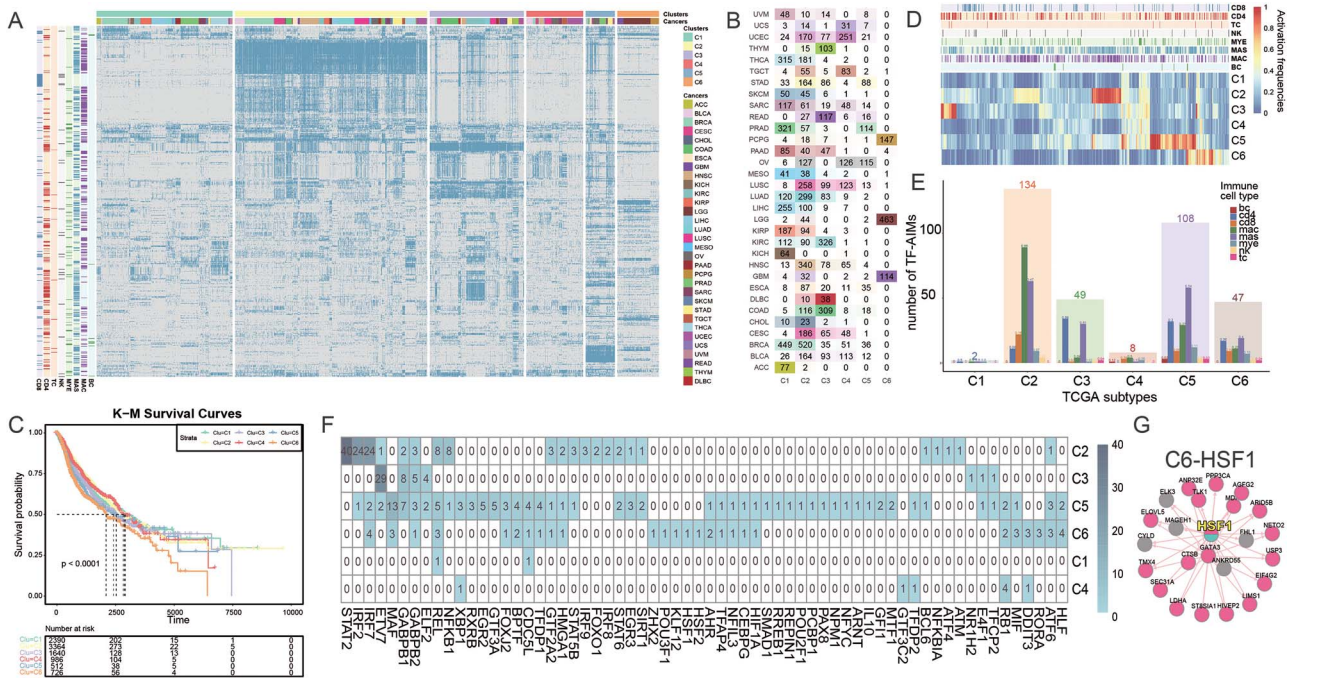

Figure 7. Analysis of pan-cancer molecular subtypes based on TF-AIMs activation states. (A) Heatmap of activation state of TII cell type-specific TF-AIMs with activation frequency between 10% and 50% in samples of six TCGA pan-cancer subtypes. The above annotation bars are the new tumor molecular subtypes and the tissue origin of the sample; note on the left is cell type, indicating cell type-specific TF-AIMs. (B) Distribution of six molecular subtypes in different cancer types of TCGA. (C) Kaplan–Meier survival curves of samples from six molecular subtypes of pan-cancer. (D) Heatmap of activation frequencies of TII cell type-specific TF-AIMs in six molecular subtypes. Column: TF-AIMs; row: Six molecular subtypes. (E) the number of TF-AIMs activated in different molecular subtypes of pan-cancer. (F) the number of TF-AIMs regulated by TFs in different molecular subtypes of pan-cancer. (G) HSF1 related TF-AIM for C6 subtype.

transformed from scATAC-seq can also be as input to identify TF-AIMs specifically for different cell types. In addition, with the increasing number of scATAC-seq data, our approach will consider further integrating scATAC-seq data to enhance the capability of method.

Furthermore, there are two unique aspects for our work: (i) in addition to the direct target relationship of TF-target gene, the regulatory modules identified by our method also contain interaction between downstream genes, which may be conducive to further molecular mechanism analysis; (ii) we mainly focused on TF-regulated immune modules, constructed and analyzed a variety of immune cell-specific TF regulatory networks in TME. We have applied our pipeline to identify activity TF-AIMs across cells of multiple datasets which mainly from lung cancer and melanoma. Our analysis involved multiple TME cell types such as T cells, B cells, NK and myeloid cells which are important components of the tumor microenvironment. However, there are still limitations in terms of data coverage such as the range of cancer types and TME. With the continuous accumulation of cancer single cell data, we will further widely integrate and analyze single cell data of various cancer types in future study. Meanwhile, both our pipeline and the TF-AIM candidates are accessible from our web interface (http://bio-bigdata.hrbmu.edu.cn/CellTF-AIMs/), it can also further promote the broad application of our pipeline in different cancer types.

Key Points

- We integrated the transcriptional profiles from single-cell datasets, and proposed a computational pipeline to identify TII cell type-specific transcription factor (TF) mediated activity immune modules (TF-AIMs).
- A comprehensive characterization and analysis of these TF-AIMs revealed key TFs in TII cells and the potential of TF-AIMs to be used as predictive markers for immunotherapy response of cancer.
- We identified and characterized six molecular subtypes of pan-cancer based on the activation status of TII cell type-specific TF-AIMs across 9680 tumor tissue samples from TCGA.
- A user-friendly web interface CellTF-AIMs (http://bio-bigdata.hrbmu.edu.cn/CellTF-AIMs/) for exploring transcriptional regulatory in various TII cell types was constructed.

Abbreviations

TII Tumor infiltrating immune
TF Transcription factor
TF-AIMs TF mediated activity immune modules
TME Tumor microenvironment
MSigDB Molecular Signatures Database
LGG Low-grade glioma of the brain
GBM Glioblastoma Multiforme

Supplementary data

Supplementary data is available at *Briefings in Bioinformatics* online.

Funding

This work was supported by the National Natural Science Foundation of China (Grant Nos. 62172131, U23A20166, 62101164 and 32070673), the China Brain Project (2021ZD0202403), the Heilongjiang Touyan Innovation Team Program, the Natural Science Foundation of Heilongjiang Province (Grant No. LH2022F042), the China Postdoctoral Science Special Foundation (Grant No. 2020 T130162) and the China Postdoctoral Science Foundation (Grant No. 2019 M661295).

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

References

1. Li T, Fu J, Zeng Z. *et al.* TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res* 2020;**48**:W509–14. https://doi.org/10.1093/nar/gkaa407.
2. Liu Q, Li J, Zheng H. *et al.* Adoptive cellular immunotherapy for solid neoplasms beyond CAR-T. *Mol Cancer* 2023;**22**:28. https://doi.org/10.1186/s12943-023-01735-9.
3. van der Leun AM, Thommen DS, Schumacher TN. CD8(+) T cell states in human cancer: insights from single-cell analysis. *Nat Rev Cancer* 2020;**20**:218–32. https://doi.org/10.1038/s41568-019-0235-4.
4. Sade-Feldman M, Yizhak K, Bjorgaard SL. *et al.* Defining T cell states associated with response to checkpoint immunotherapy in melanoma. *Cell* 2018;**175**:998–1013.e20. https://doi.org/10.1016/j.cell.2018.10.038.
5. Liang Y, Tan Y, Guan B. *et al.* Single-cell atlases link macrophages and CD8(+) T-cell subpopulations to disease progression and immunotherapy response in urothelial carcinoma. *Theranostics* 2022;**12**:7745–59. https://doi.org/10.7150/thno.77281.
6. Ni Z, Xing D, Zhang T. *et al.* Tumor-infiltrating B cell is associated with the control of progression of gastric cancer. *Immunol Res* 2021;**69**:43–52. https://doi.org/10.1007/s12026-020-09167-z.
7. Lukhele S, Rabbo DA, Guo M. *et al.* The transcription factor IRF2 drives interferon-mediated CD8(+) T cell exhaustion to restrict anti-tumor immunity. *Immunity* 2022;**55**:2369–2385.e10. https://doi.org/10.1016/j.immuni.2022.10.020.
8. Wing JB, Tanaka A, Sakaguchi S. Human FOXP3(+) regulatory T cell heterogeneity and function in autoimmunity and cancer. *Immunity* 2019;**50**:302–16. https://doi.org/10.1016/j.immuni.2019.01.020.
9. Wingender E, Dietze P, Karas H. *et al.* TRANSFAC: a database on transcription factors and their DNA binding sites. *Nucleic Acids Res* 1996;**24**:238–41. https://doi.org/10.1093/nar/24.1.238.
10. Lachmann A, Xu H, Krishnan J. *et al.* ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. *Bioinformatics* 2010;**26**:2438–44. https://doi.org/10.1093/bioinformatics/btq466.
11. Han H, Cho JW, Lee S. *et al.* TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Res* 2018;**46**:D380–6. https://doi.org/10.1093/nar/gkx1013.
12. Kibinge N, Ono N, Horie M. *et al.* Integrated pathway-based transcription regulation network mining and visualization based on gene expression profiles. *J Biomed Inform* 2016;**61**:194–202. https://doi.org/10.1016/j.jbi.2016.04.002.
13. Liu X, Wang Y, Ji H. *et al.* Personalized characterization of diseases using sample-specific networks. *Nucleic Acids Res* 2016;**44**:e164. https://doi.org/10.1093/nar/gkw772.
14. Navin N, Kendall J, Troge J. *et al.* Tumour evolution inferred by single-cell sequencing. *Nature* 2011;**472**:90–4. https://doi.org/10.1038/nature09807.
15. Aibar S, Gonzalez-Blas CB, Moerman T. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* 2017;**14**:1083–6. https://doi.org/10.1038/nmeth.4463.
16. Barrett T, Wilhite SE, Ledoux P. *et al.* NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Res* 2013;**41**:D991–5. https://doi.org/10.1093/nar/gks1193.
17. Zheng GX, Terry JM, Belgrader P. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat Commun* 2017;**8**:14049. https://doi.org/10.1038/ncomms14049.
18. Cancer Genome Atlas Research, N, Weinstein JN, Collisson EA. *et al.* The cancer genome atlas pan-cancer analysis project. *Nat Genet* 2013;**45**:1113–20. https://doi.org/10.1038/ng.2764.
19. Matys V, Fricke E, Geffers R. *et al.* TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res* 2003;**31**:374–8. https://doi.org/10.1093/nar/gkg108.
20. Consortium, E.P. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol* 2011;**9**:e1001046.
21. Rouillard AD, Gundersen GW, Fernandez NF. *et al.* The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database (Oxford)* 2016;**2016**:2016. https://doi.org/10.1093/database/baw100.
22. Demir E, Cary MP, Paley S. *et al.* The BioPAX community standard for pathway data sharing. *Nat Biotechnol* 2010;**28**:935–42. https://doi.org/10.1038/nbt.1666.
23. Liberzon A, Birger C, Thorvaldsdottir H. *et al.* The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst* 2015;**1**:417–25. https://doi.org/10.1016/j.cels.2015.12.004.
24. Godec J, Tan Y, Liberzon A. *et al.* Compendium of immune signatures identifies conserved and species-specific biology in response to inflammation. *Immunity* 2016;**44**:194–206. https://doi.org/10.1016/j.immuni.2015.12.006.
25. Valdeolivas A, Tichit L, Navarro C. *et al.* Random walk with restart on multiplex and heterogeneous biological networks. *Bioinformatics* 2019;**35**:497–505. https://doi.org/10.1093/bioinformatics/bty637.
26. Liberzon A, Subramanian A, Pinchback R. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 2011;**27**:1739–40. https://doi.org/10.1093/bioinformatics/btr260.
27. Subramanian A, Tamayo P, Mootha VK. *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. https://doi.org/10.1073/pnas.0506580102.
28. Yang L, Chen S, Zhao Q. *et al.* The critical role of Bach2 in shaping the balance between CD4(+) T cell subsets in immune-mediated diseases. *Mediators Inflamm* 2019;**2019**:1–9. https://doi.org/10.1155/2019/2609737.
29. Lougaris V, Patrizi O, Baronio M. *et al.* NFKB1 regulates human NK cell maturation and effector functions. *Clin Immunol* 2017;**175**:99–108. https://doi.org/10.1016/j.clim.2016.11.012.

30. Low JT, Christie M, Ernst M. *et al.* Loss of NFKB1 results in expression of tumor necrosis factor and activation of signal transducer and activator of transcription 1 to promote gastric tumorigenesis in mice. *Gastroenterology* 2020;**159**:e1415, 1444, 1458.e15. https://doi.org/10.1053/j.gastro.2020.06.039.
31. Qu H, Zhao H, Zhang X. *et al.* Integrated analysis of the ETS family in melanoma reveals a regulatory role of ETV7 in the immune microenvironment. *Front Immunol* 2020;**11**:612784:1-16. https://doi.org/10.3389/fimmu.2020.612784.
32. Naka K, Ochiai R, Matsubara E. *et al.* The lysophospholipase D enzyme Gdpd3 is required to maintain chronic myelogenous leukaemia stem cells. *Nat Commun* 2020;**11**:4681. https://doi.org/10.1038/s41467-020-18491-9.
33. Liu M, Tong Z, Ding C. *et al.* Transcription factor c-Maf is a checkpoint that programs macrophages in lung cancer. *J Clin Invest* 2020;**130**:2081–96. https://doi.org/10.1172/JCI131335.
34. Resendes KK, Rosmarin AG. Sp1 control of gene expression in myeloid cells. *Crit Rev Eukaryot Gene Expr* 2004;**14**:171–82. https://doi.org/10.1615/CritRevEukaryotGeneExpr.v14.i3.20.
35. Lin J, Yan J, Deng XL. *et al.* SPATS2 is correlated with cell cycle progression and immune cells infiltration in hepatocellular carcinoma. *BMC Gastroenterol* 2023;**23**:8. https://doi.org/10.1186/s12876-022-02633-y.
36. Zhang Y, Pu X, Shi M. *et al.* Critical role of c-Jun overexpression in liver metastasis of human breast cancer xenograft model. *BMC Cancer* 2007;**7**:145. https://doi.org/10.1186/1471-2407-7-145.
37. Kim T, Johnston J, Castillo-Lluva S. *et al.* TRIB1 regulates tumor growth via controlling tumor-associated macrophage phenotypes and is associated with breast cancer survival and treatment response. *Theranostics* 2022;**12**:3584–600. https://doi.org/10.7150/thno.72192.
38. Spaan CN, Smit WL, van Lidth de Jeude JF. *et al.* Expression of UPR effector proteins ATF6 and XBP1 reduce colorectal cancer cell proliferation and stemness by activating PERK signaling. *Cell Death Dis* 2019;**10**:490. https://doi.org/10.1038/s41419-019-1729-4.
39. Yu ZH, Lun SM, He R. *et al.* Dual function of MAZ mediated by FOXF2 in basal-like breast cancer: promotion of proliferation and suppression of progression. *Cancer Lett* 2017;**402**:142–52. https://doi.org/10.1016/j.canlet.2017.05.020.
40. Gabrysova L, Alvarez-Martinez M, Luisier R. *et al.* c-Maf controls immune responses by regulating disease-specific gene networks and repressing IL-2 in CD4(+) T cells. *Nat Immunol* 2018;**19**:497–507. https://doi.org/10.1038/s41590-018-0083-5.
41. Pelham SJ, Caldirola MS, Avery DT. *et al.* STAT5B restrains human B-cell differentiation to maintain humoral immune homeostasis. *J Allergy Clin Immunol* 2022;**150**:931–46. https://doi.org/10.1016/j.jaci.2022.04.011.
42. Rivas JR, Liu Y, Alhakeem SS. *et al.* Interleukin-10 suppression enhances T-cell antitumor immunity and responses to checkpoint blockade in chronic lymphocytic leukemia. *Leukemia* 2021;**35**:3188–200. https://doi.org/10.1038/s41375-021-01217-1.
43. St Paul M, Ohashi PS. The roles of CD8(+) T cell subsets in antitumor immunity. *Trends Cell Biol* 2020;**30**:695–704. https://doi.org/10.1016/j.tcb.2020.06.003.
44. Park J, Hsueh PC, Li Z. *et al.* Microenvironment-driven metabolic adaptations guiding CD8(+) T cell anti-tumor immunity. *Immunity* 2023;**56**:32–42. https://doi.org/10.1016/j.immuni.2022.12.008.
45. Tanoue T, Morita S, Plichta DR. *et al.* A defined commensal consortium elicits CD8 T cells and anti-cancer immunity. *Nature* 2019;**565**:600–5. https://doi.org/10.1038/s41586-019-0878-z.
46. Becht E, McInnes L, Healy J. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol* 2018;**37**:38–44. https://doi.org/10.1038/nbt.4314.
47. Kamphorst AO, Pillai RN, Yang S. *et al.* Proliferation of PD-1+ CD8 T cells in peripheral blood after PD-1-targeted therapy in lung cancer patients. *Proc Natl Acad Sci U S A* 2017;**114**:4993–8. https://doi.org/10.1073/pnas.1705327114.
48. Siddiqui I, Schaeuble K, Chennupati V. *et al.* Intratumoral Tcf1(+)PD-1(+)CD8(+) T cells with stem-like properties promote tumor control in response to vaccination and checkpoint blockade immunotherapy. *Immunity* 2019;**50**:195–211.e10. https://doi.org/10.1016/j.immuni.2018.12.021.
49. Thommen DS, Koelzer VH, Herzig P. *et al.* A transcriptionally and functionally distinct PD-1(+) CD8(+) T cell pool with predictive potential in non-small-cell lung cancer treated with PD-1 blockade. *Nat Med* 2018;**24**:994–1004. https://doi.org/10.1038/s41591-018-0057-z.
50. Shang S, Yang YW, Chen F. *et al.* TRIB3 reduces CD8(+) T cell infiltration and induces immune evasion by repressing the STAT1-CXCL10 axis in colorectal cancer. *Sci Transl Med* 2022;**14**:eabf0992. https://doi.org/10.1126/scitranslmed.abf0992.
51. Han Y, Guo Z, Jiang L. *et al.* CXCL10 and CCL5 as feasible biomarkers for immunotherapy of homologous recombination deficient ovarian cancer. *Am J Cancer Res* 2023;**13**:1904–22.
52. Eskilsson E, Rosland GV, Solecki G. *et al.* EGFR heterogeneity and implications for therapeutic intervention in glioblastoma. *Neuro Oncol* 2018;**20**:743–52. https://doi.org/10.1093/neuonc/nox191.
53. Liu Y, Pang Y, Zhu B. *et al.* Therapeutic targeting of SDHB-mutated Pheochromocytoma/Paraganglioma with pharmacologic ascorbic acid. *Clin Cancer Res* 2020;**26**:3868–80. https://doi.org/10.1158/1078-0432.CCR-19-2335.
54. Holmes B, Benavides-Serrato A, Freeman RS. *et al.* mTORC2/AKT/HSF1/HuR constitute a feed-forward loop regulating Rictor expression and tumor growth in glioblastoma. *Oncogene* 2018;**37**:732–43. https://doi.org/10.1038/onc.2017.360.
55. Mustafa DA, Sieuwerts AM, Zheng PP. *et al.* Overexpression of Colligin 2 in glioma vasculature is associated with overexpression of heat shock factor 2. *Gene Regul Syst Bio* 2010;**4**:103–7. https://doi.org/10.4137/GRSB.S4546.
56. Antonietti P, Linder B, Hehlgans S. *et al.* Interference with the HSF1/HSP70/BAG3 pathway primes glioma cells to matrix detachment and BH3 mimetic-induced apoptosis. *Mol Cancer Ther* 2017;**16**:156–68. https://doi.org/10.1158/1535-7163.MCT-16-0262.
57. Aran D, Looney AP, Liu L. *et al.* Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol* 2019;**20**:163–72. https://doi.org/10.1038/s41590-018-0276-y.
58. Bravo Gonzalez-Blas C, De Winter S, Hulselmans G. *et al.* SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. *Nat Methods* 2023;**20**:1355–67. https://doi.org/10.1038/s41592-023-01938-4.