Proceedings of ICMMGH 2026 Symposium: Cell and Molecular Biology: Understanding Immune Cell Functions
DOI: 10.54254/2753-8818/2026.PJ30861

Single-Cell Transcriptomics Reveals T-Cell State Transitions and Their Association with Immunotherapy Response in the Melanoma Tumor Microenvironment

Shangyu Ning

School of Biological Sciences, Nanyang Technological University, Singapore, Singapore
mahoutsukai.megumin@gmail.com

Abstract. The tumor infiltrating T cells determine whether a patient would respond positively to immunotherapy. This paper has analyzed an mRNA scRNA-seq dataset of melanoma in the context of mapping heterogeneity in T-cells and the question of whether clinical features can shape cell clustering. Unsupervised clustering with k-means based on quality control followed by a description of the major T-cell groups by canonical immune markers. The identification of the differentially expressed genes that were identified was then done by a posteriori information based ranking approach. This study have created a confidence-weighted approach, which is able to merge Spearman and cosine similarity measures with marker-gene rules, to enhance cell-type annotation. Metadata analysis indicated that treatment status is an element that causes the state changes in γ/δ T cells pre- and post-checkpoint therapy. Combining graph-based Leiden clustering and k-means, this study refined the definition of the fine-grained T-cell states and proved that, despite the significant heterogeneity of T cells, the strong definition of subclass depends on the combination of several methods of clustering and the use of orthogonal knowledge of experiments. Pseudotime analysis found evidence in gradual transition between naive or early-activated states to cytotoxic or dysfunctional phenotypes that indicate intratumoral T-cell diversity captures a reflection not only of inherent developmental trajectories, but also of the effects due to therapy. Lastly, survival analysis revealed that there was a distinct relationship between clinical response and overall survival. These findings combined narrow the scope of the T-cell picture and offer understanding in the form of transcriptional programs connected with treatment outcomes. Refer to this link for supplementary materials: https://drive.google.com/drive/folders/1UYW-6EMjSUMVOwHxvuYdFSZwREqrGp89

Keywords: Single-cell RNA-seq, T cell heterogeneity, Cluster-specific signature genes, Immunotherapy response, Pseudotime analysis

1. Introduction

T cells are effective in different tumor microenvironments in a broad range as they are infiltrated, activated, and exhausted via programmes that are unique to each cancer setting. There have been very visible clinical improvements due to immunotherapy employing immune checkpoint blockade (ICB) therapies (anti-PD-1 and anti-CTLA-4) in both melanoma and lung cancer. But the vast majority of patients do not respond, and treatment-related toxicities preclude the expanded usage [1]. Tumor

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2026.PJ30861

immune evasion, as well as the resistance to the therapeutic effects of the immunotherapies, is heavily dependent on the state of functionality, the clonal organization and the phenotypic diversity of the tumor-infiltrating T cells and the disregard of these subsets has been documented in the tumor immunotherapy [2].

The microenvironment (TME) of tumors is highly heterogeneous, putting T cells in varied lineages and functional states. It consists of, but is not limited to,γδT cells, traditional 8T cells of the CD4+ and CD8+ compartments, of a double-positive and double-negative compartments, and of various memory-like and regulatory populations [3]. These T cells experience progressive functional impairment or T-cell exhaustion associated with decreased proliferative and cytolytic function with an increase in various inhibitory receptors, including PD-1, LAG-3, CD160, 2B4, TIM-3, and TIGIT under chronic antigenic stimulation [4]. This phenotype indicates a distinct program (e.g., TOX-dependent) that is epigenetically and transcriptionally regulated and conserved across chronic infection and cancer settings and which can have significant implications for immune checkpoint therapy [5]. Simultaneously, therapeutic intervention also propagates dynamic phenotypic and functional state changes [2]. Such pressures cause tumor invading T lymphocytes (TITLs) that are dynamic; that is, they migrate between tumor regions and through treatment regimens. To elucidate the responsiveness of some and non-responsiveness of others to tumors, it is required to capture this movement or rather merely to list fixed phenotypes.

RNA-seq Bulk RNA-seq groups all T-cell states on one average. Single-cell RNA-seq does not have this limitation, because it measures transcriptional programs cell by cell, which is often thousands of TITLs per specimen. Technical advancements in the technology to prepare libraries and multiomic profiling currently enable researchers to compare transcriptional, chromatin and surface-protein variation in the same cells.

The present interest in this study is to determine T-cell subpopulations and to follow the dynamics of their changing in the course of the therapy. This paper combines clustering, enrichment analyses, pseudotime reconstruction, and survival modeling, not as isolated analyses but to link molecular state dynamics with the clinical outcomes observed in patients.

2. Methods overview

This single-cell RNA-sequencing study (Figure 1) began with dataset quality control (QC), followed by supervised visualization using established canonical marker genes. Next, k-means unsupervised clustering was applied to partition immune-cell populations, providing the structural basis for subsequent differential expression (DE) analyses and metadata-associated comparisons.

Cluster-specific marker genes were then identified and interpreted based on DE and metadata associations, enabling the biological annotation and definition of each transcriptional cluster. Finally, downstream analyses—including pseudotime trajectory reconstruction of T-cell subsets and survival analyses linked to clinical metadata—were performed to characterize T-cell states and their functional relevance.

DOI: 10.54254/2753-8818/2026.PJ30861

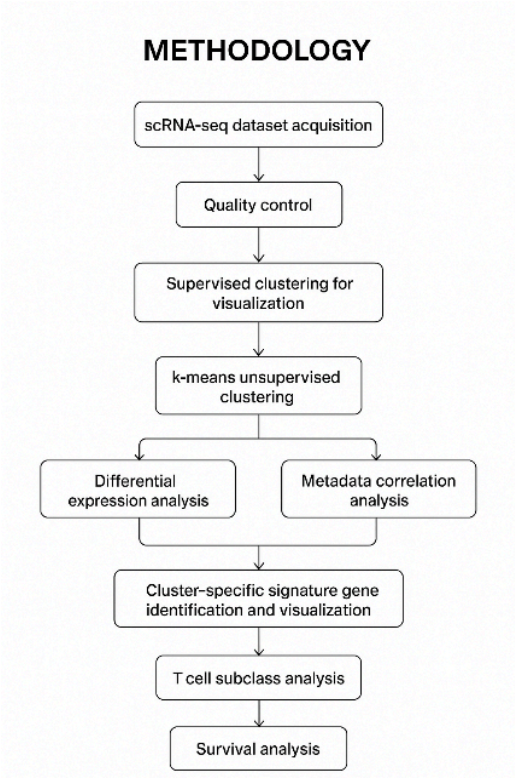

Figure 1. Flowchart of the study. (picture credit: original)

2.1. Data and preprocessing

2.1.1. Data acquisition

Data were obtained from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO), a public repository that archives and freely distributes high-throughput functional genomics datasets submitted by the research community.

The dataset used in this study [6] consists of 48 tumor specimens collected from 32 melanoma patients across multiple hospitals in the United States. Single-cell transcriptomes were profiled from a total of 16,291 cells using second-generation sequencing technologies. The patients varied in age and sex, and all had received at least one immune checkpoint therapy—either anti-CTLA-4 or anti-PD-1—with tumor biopsies obtained before and/or after treatment. The dataset provides log-transformed TPM expression matrices, generated following alignment of raw FASTQ reads to the NCBI human reference genome hg19.

2.1.2. Quality control

Cell-level QC was performed and removed cells that do not meet any of the following requirements: (1) cells lacking expression of CD45 and CD3E; (2) cells with fewer than 1,000 detected genes; and (3) cells in which the average expression of housekeeping genes [7] showed $\log_2(\text{TPM}+1) < 2.5$.

After filtering, a gene-by-cell expression matrix of ~16k × 16k was obtained. Kernel density estimation (KDE) plots comparing gene-expression distributions before and after QC are shown in Figure 2.

DOI: 10.54254/2753-8818/2026.PJ30861

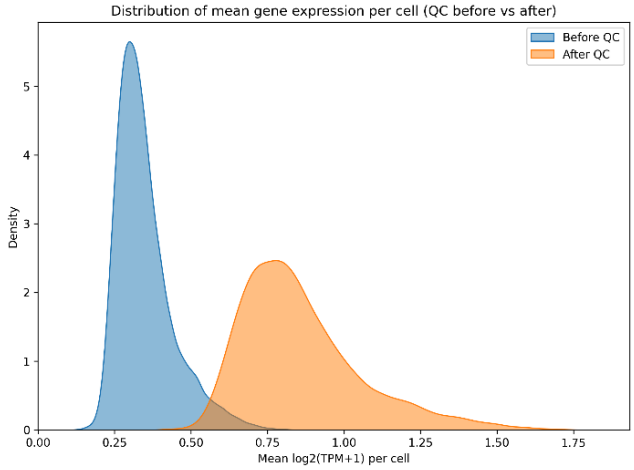

Figure 2. Kernel density plots showing the distribution of mean gene expression before and after quality control. (picture credit: original)

2.1.3. Supervised marker-based visualization

A supervised, marker-based annotation using a panel of well-established immune cell markers [8] (see Supplementary Excel Table) was performed. Importantly, this annotation was applied for visualization purposes only and did not determine downstream cluster assignments. This approach labelled roughly 40% of cells on average, providing a useful preview of the major cellular groupings in the dataset.

Nonlinear embedding dimension-reduction methods (t-distributed Stochastic Neighbor Embedding, t-SNE and Uniform Manifold Approximation and Projection, UMAP) were adopted to represent the structure of the high-dimensional space for visualization; representative projections are shown in Figure 3.

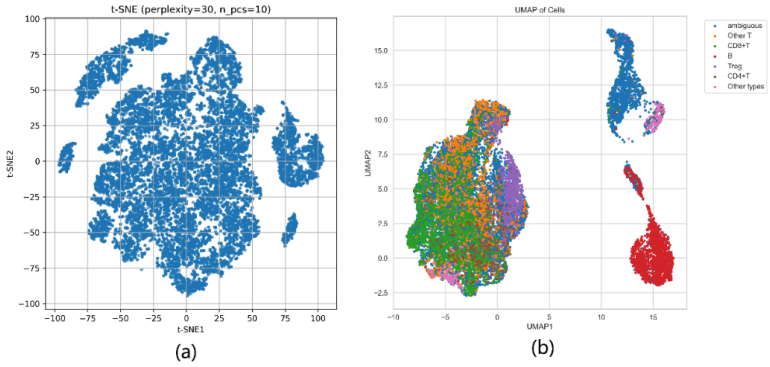

Figure 3. Visualization of cells after quality control. (a) t-SNE with perplexity=30, n_pcs=10, (b) UMAP with default parameters. (picture credit: original)

2.2. Unsupervised clustering and metadata integration

2.2.1. k-means unsupervised clustering

The k-means clustering algorithm (Equations (1), (2), (3)) [9] was applied to the quality-controlled dataset for unsupervised classification. First, the most variable genes across the dataset were picked, selecting 4,157 genes with a variance greater than 6 for clustering.

DOI: 10.54254/2753-8818/2026.PJ30861

A linear correlation matrix R_{ij} between cells was calculated and then converted into a distance matrix D_{ij} . The V value is computed based on the sum of all pairwise distances between cells in different clusters ($D_{\text{is_between}}$). Therefore, a larger V indicates greater separation between clusters and smaller variation within clusters, reflecting a better clustering outcome. As a complement, the silhouette coefficient was simultaneously computed to evaluate local cluster cohesion and separation.

Then, the elbow method was used to select the optimal k: the point at which increases in clustering effectiveness begin to plateau was chosen as the best cluster number, balancing cluster complexity with variance explained.

Pearson correlation matrix between cells:

$$R_{ij} = Rearsoncorrelationbetweencelliandcellj \tag{1}$$

Distance matrix:

$$D_{ij} = 1 - R_{ij} \tag{2}$$

Ratio of between-cluster to total distance:

$$V = \frac{D_{is_{between}}}{D_{total}} = \frac{\sum_{k=1}^{K} \sum_{i \in C_k, j \notin C_k} D_{ij}}{\sum_{ij} D_{ij}} \tag{3}$$

2.2.2. Metadata variable effects on cluster composition

For categorical metadata, including treatment timepoint (Pre vs. Post), clinical response, therapy type, gender, and survival status, associations were evaluated using χ^2 tests, followed by Cramér's V to estimate effect size. For continuous metadata, including age and overall survival, one-way ANOVA was applied and calculated the eta-squared (η^2) as a measure of effect size.

2.3. Differential expression and functional interpretation

2.3.1. Differential expression analysis

Single-cell RNA-seq data were preprocessed using Python Scanpy (v1.11.5), integrating metadata including patient information and k-means cluster assignments. A pseudobulk approach was applied by averaging gene expression across all cells within each (cluster × patient) combination.

DE analysis was performed for each cluster versus all others using Welch's t-test, with log2 fold changes (logFC) and p-values calculated per gene. Bonferroni correction was applied to obtain adjusted p-values (adj_pval). Following Scanpy community defaults, genes with |logFC| > 0.25 and adj_pval < 0.05 were retained for downstream analysis [10].

To enable comparability of top DE genes across clusters, an a posteriori information fusion approach [11] was applied and defined a composite score:

$$score = -log_{10}(adj_pval + 1 \times 10^{-300}) \times |\ logFC\ | \tag{4}$$

Gene biotype annotations were sourced from the HUMAN GENCODE 49 (September 2025) release [12]. This dataset provides comprehensive classifications of human genes into categories such as protein-coding, lncRNA, pseudogene, mitochondrial genes, among others, and was used to assign gene biotypes in subsequent analyses.

DOI: 10.54254/2753-8818/2026.PJ30861

2.3.2. Functional enrichment analysis

Cluster-specific gene lists were analyzed using the Enrichr database via the gseapy (v1.1.10) Python package. Enrichment significance was determined using an adjusted p-value < 0.05. The resulting enriched pathways and Gene Ontology (GO) terms were visualized for each cluster.

2.3.3. Rule-based cell type annotation

Cell-type identities for transcriptional clusters were assigned using a rule-based marker evaluation framework combined with a quantitative confidence scoring procedure. Marker definitions were curated from canonical immunology literature used in the prior supervised clustering session (Methods 2.4), as well as widely adopted single-cell references (see Supplementary Excel Table). Markers were organized into four categories: mandatory markers (must_have), alternative markers (must_have_any), negative markers (must_not), and functional enrichment markers (rich_in) for activated, memory, and exhausted T-cell states.

A composite confidence metric was developed, integrating several complementary signals commonly used in reference-based classifiers such as SingleR. Rank-based agreement using Spearman correlation was calculated and directional alignment of standardized expression vectors via cosine similarity was measured.

This confidence framework is motivated by established principles in state-of-the-art cell type annotation tools: marker concordance (match rate and marker weighting) follows the rule-based strategy of Garnett [13]; rank-based correlation is the primary metric in SingleR [14]; and cosine similarity has been widely applied for evaluating transcriptomic signature alignment in scRNA-seq and multimodal frameworks (e.g., scVI/scANVI) [15]. Consistent with scClassify [16], these complementary measures are integrated through a weighted ensemble formulation. The final confidence score was calculated as

$$\text{Confidence} = 0.40 \times \text{match_rate} + 0.25 \times \text{scaled Spearman} + 0.25 \times \text{cosine similarity} + 0.10 \times \text{marker weight facto}$$

and constrained within [0,1]. For each cluster, the cell type with the highest confidence score was assigned as the final annotation.

2.4. T-cell dynamics and clinical outcome analysis

2.4.1. Metadata impact analysis of T cells

Patient and cell-level metadata were integrated with the expression matrix using cell identifiers. From the total of 11 clusters, six clusters corresponding to canonical T-cell populations were selected for analysis.

2.4.2. Trajectory inference and pseudotime computation

Highly variable genes (n = 2,000; Seurat flavor) were selected, and expression values were scaled to a maximum of 10 to limit the impact of outliers. Principal component analysis (PCA) was performed using the top 30 principal components, and a k-nearest neighbor (kNN) graph (n_neighbors = 15) was constructed to model cell–cell relationships.

A low-resolution Leiden clustering (resolution = 0.5) was applied to assist in selecting the root cell for diffusion pseudotime (DPT). Partition-based graph abstraction (PAGA) was computed to summarize cluster connectivity, followed by diffusion map calculation and DPT to assign pseudotime values to

DOI: 10.54254/2753-8818/2026.PJ30861

individual cells. Pseudotime distributions were visualized on UMAP embeddings to examine the global trajectory structure and the progression of T-cell states.

2.4.3. Survival analysis

Kaplan–Meier (KM) survival analysis was conducted to assess the influence of key clinical factors on patient outcomes. Grouping variables included clinical response, gender, and therapy type.

3. Results

3.1. Identification of 11 optimal clusters

K-means clustering by calculating the silhouette coefficient shows the V ratio increased steadily with k from 7 to 16. The silhouette score reached a clear maximum at k = 11 (V = 0.8764, silhouette score = 0.1604) (Figure 4). An 11-cluster solution optimally balances the capture of global data complexity with strong internal cluster consistency.

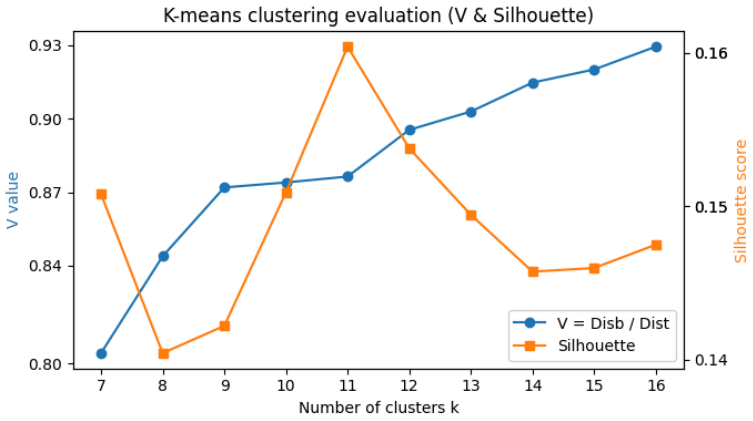

Figure 4. V value and silhouette score determination of optimal cluster number for unsupervised analysis. (picture credit: original)

t-SNE and UMAP revealed a heterogeneous landscape of immune cells (Figure 5). The distribution of cells across clusters was not uniform, with certain populations forming distinct, well-separated groups (e.g., Clusters 7 and 9), while others occupied more contiguous regions in the low-dimensional space (e.g., Clusters 0, 3, 4, 8, and 10).

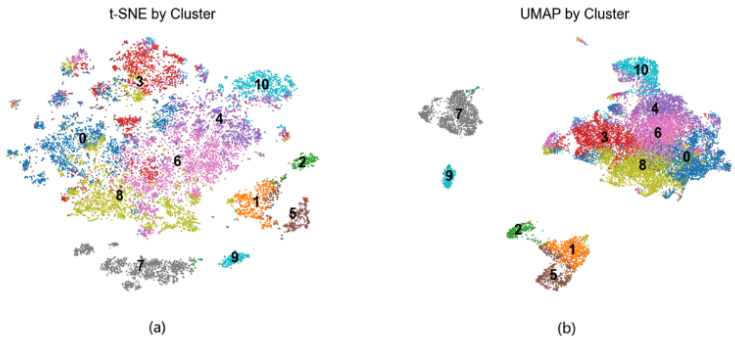

Figure 5. t-SNE(a) and UMAP(b) plot mapping with unsupervised cell clustering. (picture credit: original)

DOI: 10.54254/2753-8818/2026.PJ30861

3.2. Metadata effects on T-cell cluster composition: strong influence of treatment timepoint and clinical response

Among the categorical metadata, treatment timepoint displayed the strongest non-random association with T-cell clusters (Cramér's V = 0.41), indicating that pre- and post-treatment cells largely occupied distinct transcriptional states rather than being evenly distributed across clusters. Clinical response also showed a substantial effect (Cramér's V = 0.39), with responder-derived cells preferentially enriched in specific clusters, suggesting that response status is reflected at the population structure level (Figure 6a). Other categorical variables, including gender and therapy subtype, had minimal or negligible influence on clustering (Cramér's V < 0.15).

For continuous metadata, overall survival exhibited a modest association with cluster identity (η^2 = 0.083) (Figure 6b), whereas patient age showed little to no effect.

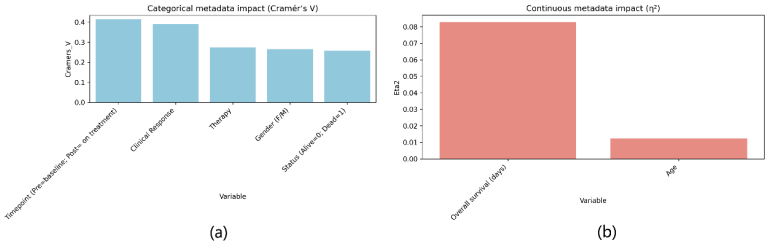

Figure 6. Metadata influence on T-cell clustering. (a) Categorical variables were assessed using χ^2 tests with Cramér's V to estimate effect size. (b) Continuous variables were evaluated via one-way ANOVA with η^2. (picture credit: original)

3.3. Identification of cluster-specific transcriptional signatures

DE analysis revealed that each cluster harbored a set of both up- and downregulated genes (Figure 7). Using the scoring method described in Methods 2.7, the most representative cluster-specific genes across the entire dataset were identified (shown in Table 1). The top-ranked cluster-specific DE genes were selected to generate heatmaps (Figure 8).

Table 1. Top 3-ranked genes in each k-means cluster

Cluster#	Gene	logFC	P_val	adj_P_val	Score	Rank in Cluster
0	TNFAIP3	4.181627	6.06E-32	3.38E-27	110.6922	1
	TRBC2	4.182324	1.54E-29	8.57E-25	100.6561	2
	IFI30	-3.51725	8.97E-34	5.00E-29	99.54181	3
1	TRAC	-5.72088	2.76E-67	1.54E-62	353.6216	1
	TRBC2	-6.90582	1.78E-54	9.94E-50	338.4032	2
	CST3	8.65351	7.96E-44	4.43E-39	331.8932	3
2	HLA-DRA	5.880879	8.75E-58	4.88E-53	307.6381	1
	TRBC2	-6.61677	1.43E-49	7.96E-45	291.7936	2
	CD3D	-4.90465	3.29E-61	1.83E-56	273.3732	3
3	TRAC	5.325947	1.44E-55	8.03E-51	266.8048	1
	TRBC2	5.226594	1.22E-48	6.79E-44	225.6223	2
	IL32	5.19586	6.01E-37	3.35E-32	163.5395	3
4	CD3G	4.787857	1.84E-51	1.03E-46	220.18	1
	CD2	4.698981	8.08E-51	4.50E-46	213.0837	2
	CCL5	5.824128	1.92E-38	1.07E-33	192.0251	3
5	GRN	7.345731	2.41E-83	1.34E-78	572.0334	1

DOI: 10.54254/2753-8818/2026.PJ30861

Table 1. (continued)

	GPX1	6.021401	3.19E-69	1.78E-64	383.8618	2
	HLA-DRA	6.89696	1.25E-60	6.99E-56	380.4054	3
6	CCL5	7.734873	3.48E-91	1.94E-86	662.973	1
	NKG7	7.550167	7.46E-71	4.16E-66	493.6368	2
	GZMA	6.15829	6.60E-47	3.68E-42	255.1635	3
7	FYB	-4.97398	2.06E-111	1.15E-106	526.9395	1
	IL32	-5.77733	9.27E-64	5.17E-59	336.7406	2
	ANXA1	-4.33808	1.20E-75	6.67E-71	304.4285	3
8	HLA-DRA	-6.30585	1.83E-59	1.02E-54	340.4615	1
	HLA-DPA1	-5.58157	1.12E-48	6.25E-44	241.1468	2
	HLA-DRB5	-4.1147	7.92E-52	4.41E-47	190.7394	3
9	TRBC2	-6.92903	1.84E-68	1.02E-63	436.4694	1
	IL32	-5.65968	3.87E-60	2.16E-55	309.3897	2
	CD2	-4.69068	4.02E-64	2.24E-59	275.1074	3
10	SNRPB	4.67716	1.81E-60	1.01E-55	257.2236	1
	SLC25A5	3.940426	5.43E-69	3.03E-64	250.2902	2
	PGAM1	4.890744	3.59E-56	2.00E-51	247.9557	3

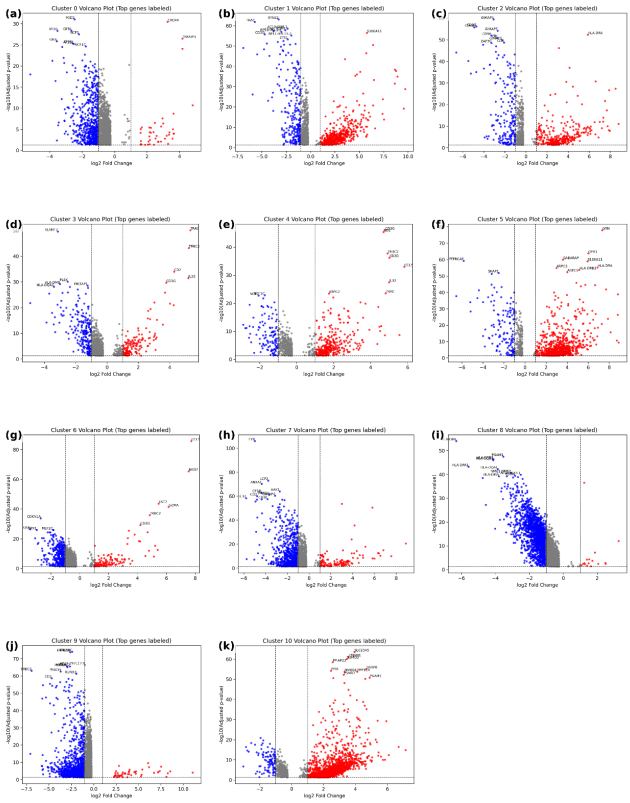

Figure 7. Volcano plots of differentially expressed genes for each cluster. Panels (a)–(k) correspond to clusters 0–10, with Cluster 8 (panel i) uniquely showing all significant genes with negative logFC values following filtering. (picture credit: original)

DOI: 10.54254/2753-8818/2026.PJ30861

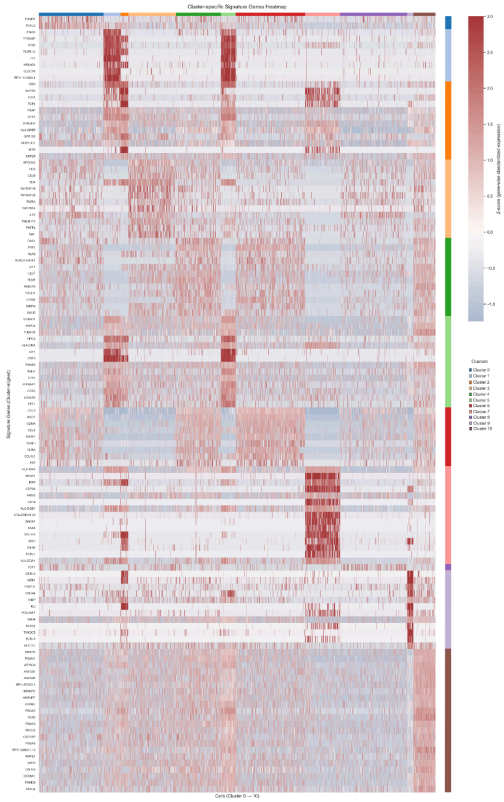

Figure 8. Heatmap of cluster-specific differentially expressed genes. (picture credit: original; high resolution file stored at supplementary web link)

3.4. Enrichment analysis

The enriched terms on the top 50 differentially expressed genes (DEGs) per cluster using GO Biological Process (2021) and KEGG 2021 Human databases are stored at the supplementary web link. For instance, cluster 0 was enriched for neutrophil-related pathways such as neutrophil activation and degranulation, indicating an innate immune granulocytic population; cluster 1 showed strong enrichment of T-cell activation, TCR signaling, and interferon-γ pathways, consistent with an activated T-cell compartment; cluster 2 was dominated by metabolic and translational processes including RNA metabolism and ribosome biogenesis, suggesting a high-biosynthetic, tumor-like cell population.

3.5. Evidence-based annotation identified six T-cell–related clusters

Based on DE and functional enrichment analyses combined with marker evaluation, the predicted identities of the 11 clusters are summarized in Table 2. Among these, Clusters 0, 3, 4, 6, 8, and 10 were strongly associated with T-cell populations, exhibiting clear separation from other clusters.

Table 2. Predicted cell identities for each cluster

Cluster	Predicted Identity	n_cells	n_patients
0	Activated T cells; Memory T cells; Neutrophils; Regulatory T cells; CD8 T cells	2631	29
1	Monocyte Immature; Macrophage; cDCs dendritic cells	701	29
2	pDCs	315	27
3	Regulatory T cells; Activated T cells; CD4 T cells; Neutrophils; NKT cells; Memory T cells	1934	32

DOI: 10.54254/2753-8818/2026.PJ30861

Table 2. (continued)

4	Activated/Exhausted T cells; cDCs dendritic cells; Neutrophils; Regulatory T cells; CD8 T cells; Myeloid cells	1838	29
5	Monocyte Immature; Macrophage; cDCs dendritic cells	592	23
6	Activated/Exhausted T cells; Neutrophils; Regulatory T cells; CD8 T cells	2832	31
7	B cells	1420	30
8	Memory T cells	2734	32
9	Monocyte Immature; Myeloid cells	248	26
10	Activated/Exhausted T cells; Memory T cells; cDCs dendritic cells; Myeloid cells; Regulatory T cells; Neutrophils	905	32

Notes: (1) Predicted identities are ranked by probability, with higher-ranking entries reflecting greater likelihood. (2) For most clusters, identities with >70% probability or the top six most likely cell types were reported; Cluster 9 is an exception due to limited cell number (~40% probability). (3) Listing "Activated T cells" or "Exhausted T cells" alone indicates predominance of that activation state, whereas "Activated/Exhausted T cells" indicates roughly equal probabilities for both states (difference≤3%). (4) n_cells represents the number of cells assigned to the cluster, and n_patients indicates the number of patients contributing cells to that cluster. (5) pDCs: plasmacytoid dendritic cells; cDCs: conventional (or classical) dendritic cells.

3.6. Transcriptional heterogeneity and metadata effects within TITLs

Given the weak T cell intrinsic clustering, the analysis focused on whether sample-level or clinical variables contributed to the observed dispersion. Treatment Timepoint and Clinical Response, which showed strong metadata–cluster associations in the Results 3.2 session, were prioritized in downstream analyses to assess their impact on T-cell state distribution.

DE analysis of the full immune population and the T-cell subset revealed that metadata variables drive both shared and cell type-specific transcriptional changes. Volcano plots (Figure 9) highlight, for example, upregulation of TRDV1 and TRDC along treatment progression, suggesting enhanced activity or proportion of γ/δ T cells. Conversely, CD38 and IFI6 were downregulated in non-responders, indicating reduced T-cell activation or interferon-response pathways in treatment-failure samples. While many DEGs were common across all cells and T cells, a notable fraction was cell type-specific, and the magnitude of DE varied across subsets, reflecting context-dependent transcriptional responsiveness.

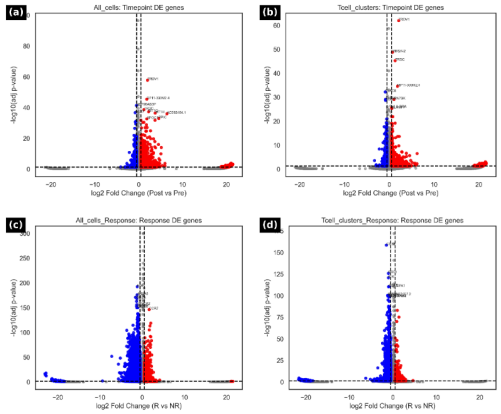

Figure 9. Volcano plot of DE genes influenced by timepoint and clinical response. (picture credit: original)

Apart from protein coding genes, T cell gene set enrichment analysis of initial DEGs was dominated by lncRNAs, pseudogenes, and mitochondrial genes (Figure 10, gene biotype annotation file stored at the supplementary web link), producing noisy and ambiguous pathway signals. Restricting the analysis

DOI: 10.54254/2753-8818/2026.PJ30861

to protein-coding genes revealed enrichment in pathways such as Ribosome, Oxidative phosphorylation, Antigen processing and presentation, and Th1/Th2/Th17 differentiation. These suggest that Treatment Timepoint primarily influences immune activation and metabolic pathways, whereas Clinical Response predominantly affects ribosomal function and antigen processing in T cells. Importantly, enrichment patterns differed between T cells and the full immune population, highlighting that metadata-driven transcriptional changes manifest differently across cellular contexts.

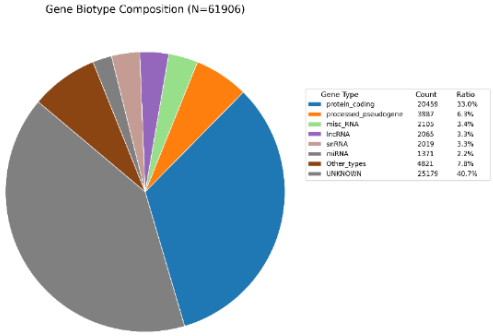

Figure 10. Gene biotype composition mapped with GENCODE 49 (September 2025). (picture credit: original)

3.7. Trajectory inference and state transitions of TITLs

DPT analysis coupled with PAGA-based trajectory inference was applied on TITLs (k-means clusters 0, 3, 4, 6, 8 and 10). Leiden clustering resolved 16 transcriptionally coherent subpopulations, which were used as nodes in the PAGA graph (Figure 11a). Leiden clusters 0 and 1 in early pseudotime states contained substantially more cells than the others. In the pseudotime UMAP plot (Figure 11b), color encodes mean pseudotime, where darker shades correspond to earlier states (Leiden clusters 0, 1, 5, 7, 8, 10, 11, 12, and 13). TITLs in early pseudotime states show strong transcriptional connectivity; for example, Leiden cluster 0 shows strong links to neighboring communities such as clusters 11, 12, 5, 8, 10, and 13. Combined with nodes and pseudotime, the trajectory of T cell development is proposed (Figure 11c).

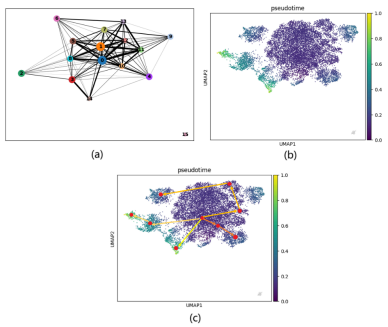

Figure 11. Pseudotime trajectory analysis of tumor-infiltrating T cells. (a) PAGA graph constructed from Leiden clustering, showing 16 transcriptionally coherent communities. Node size reflects the number of cells per community, and edge thickness represents the degree of connectivity between clusters. (b) UMAP embedding colored by diffusion pseudotime, illustrating the continuous progression of T-cell states. (c) UMAP embedding with inferred T-cell transition trajectories, where arrows of different colors indicate alternative potential cell-state fates. (picture credit: original)

DOI: 10.54254/2753-8818/2026.PJ30861

The PAGA graph indicates that T-cell states form a continuous trajectory rather than a set of discrete, well-separated groups. Early-pseudotime clusters (e.g., Leiden clusters 0 and 1) were positioned centrally in the UMAP embedding, whereas later-pseudotime clusters extended toward the periphery, consistent with a gradual shift in transcriptional states (Figure 11c). Intermediate clusters (5, 7, 8, 10, 11, 12, and 13) occupied positions along a continuum connecting early and late states rather than forming isolated "islands." This pattern supports the interpretation that the k-means clusters represent distinct segments along a shared differentiation or activation trajectory rather than fundamentally separate cell types. A similar structure is visible in the UMAP density maps of T-cell clusters across treatment timepoints (Figure 12).

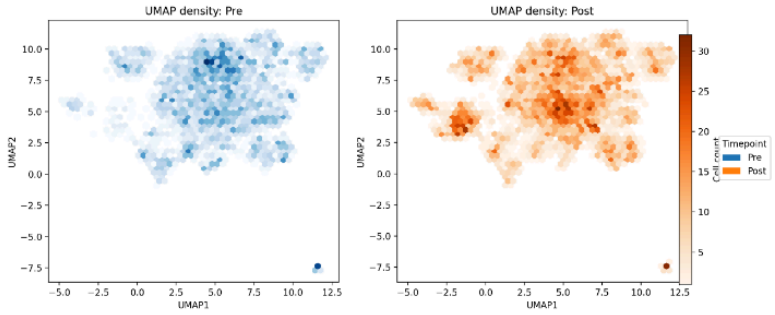

Figure 12. UMAP density plot for T cell clusters at different treatment timepoint. (picture credit: original)

Both k-means and Leiden clustering produced broadly comparable high-level T-cell groupings. When applied specifically to the T-cell subset, Leiden generated a finer-grained partition that better reflected the intrinsic topology of the underlying cell–cell graph (Figure 13). This higher topological fidelity makes Leiden-derived communities more appropriate for downstream trajectory-based analyses, including PAGA and pseudotime inference, both of which depend on an accurate representation of connectivity between transcriptional states.

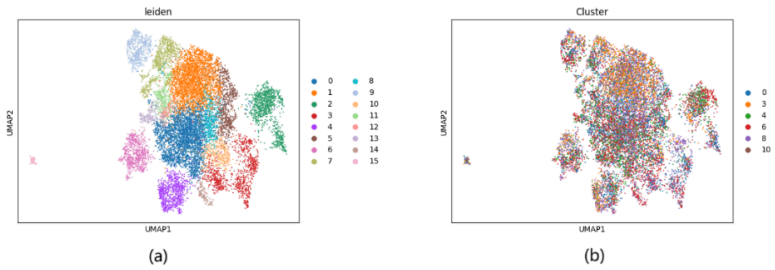

Figure 13. UMAP comparison of clustering methods applied to T cells. (a) Leiden clustering. (b) k-means clustering. (picture credit: original)

3.8. Survival analysis identifies clinical response as a major determinant of patient survival

Clinical response is strongly associated with overall patient survival (Figure 14), indicating that the therapeutic efficacy of the immunological intervention is a key driver of long-term outcomes.

DOI: 10.54254/2753-8818/2026.PJ30861

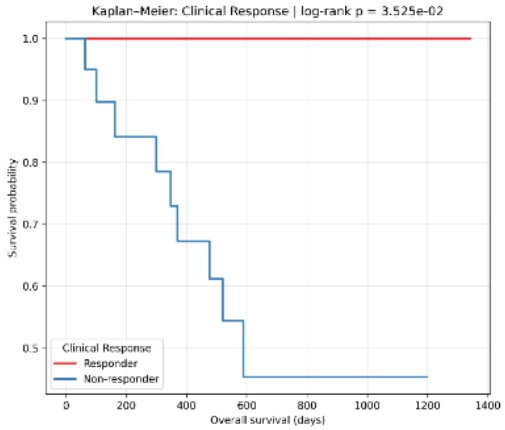

Figure 14. Kaplan–Meier survival curves stratified by clinical response. (picture credit: original)

4. Discussion

The profile of immune cells infiltrating the tumor by single cells showed a heterogeneity of the T-cell compartment. A correlation-based approach of k-means identified 11 immune groups, of which six were T-cell-associated groups. T cell subclustering formed nominally distinct sets of markers but the groups so formed became mixed up in low dimensional space. This trend indicates that the T-cell phenotype can be defined as a continuum of states, as opposed to highly distinct phenotypes; this finding is in line with other recent studies that have shown that the varied activation, dysfunction, and memory programs occur in continuous transcriptional gradients [2,17]. Both T-cell cluster composition and gene expression were influenced by complementation actions of the k-means and Clinical metadata, in particular, treatment timepoint (pre- vs. post-treatment) and patient response. The samples of post-treatment demonstrated the expression of the γδ T-cell-related genes (e.g., TRDV1, TRDC) in accordance with the therapy-related transformation of the γδ T-cell number or activation. This is in line with the prior reports that checkpoint blockade has the potential for recovery of γδ T-cell function in TME, which increases their proliferation, cytotoxicity, and antitumor activities [18]. Conversely, non-responders had decreased activation and interferon-response genes (e.g., CD38, IFI6), which means that the T-cell effector programs are disrupted. CD38 is an ecto-enzyme that is connected to nicotinamide adenine dinucleotide (NAD+) breakdown and is very much expressed in drained-out CD8+ T cells in melanoma and is related to ICB resistance [19]. Increased frequency of activated CD4+CD38+HLADR+ T cells is associated with adverse events involving the immune system [20]. Although the contribution of IFI6 to clinical response in the current studies has not been definitively defined, the validation can be done through follow-up, which may involve independent patient groups.

The increase in ribosomal and oxidative phosphorylation pathways, especially in DE genes of clinical response, indicates that translational potential and metabolic restructuring are closely interconnected with T-cell functional phenotypes that are pertinent to treatment. In addition to protein-coding genes, there is an open abundance of lncRNAs, miRNAs and other noncoding factors involved in enrichment noise and heterogeneity [21]; however, their dysregulation can impact tumorigenesis, apoptosis or cancer progression implying that their noncoding programs cannot be ignored [22,23]. Increased ribosomal and mitochondrial signatures have been linked to growing or active immune cells whereas regulation disruptions are frequently linked to the drained or maladaptive condition of immune cells. This is consistent with current findings that chronic stimulation of antigens results in T cell exhaustion with an increase in Reactive Oxygen Species (ROS), inhibited ADP-linked oxidative phosphorylation and compensatory enlargement of glycolysis [24]. As a result, the changes in these axes as observed are

DOI: 10.54254/2753-8818/2026.PJ30861

probably due to functional reprogramming that lies lying beneath patient response or resistance to a checkpoint blockade.

Findings may be restricted by a number of different factors. The Smart-seq2 data (GSE120575) has a relatively small number of cells per patient and the coverage of clusters is very uneven and may affect the downstream analysis. Although methods like pseudobulk or cluster-by-patient averaging can be used to minimize the noise in single cells, there can still be remaining technical variation and uncertainties in annotation, like pseudogene mapping, that could impact the technique's results of differential expression and enrichment. There are dynamical inferences to the T-cell states without establishing the lineage, i.e., trajectory or pseudotime inference. Such orthogonal approaches as TCR clonotype tracing, temporal sampling, or experimental lineage tracing would then have to be used to validate this. Moreover, spatial transcriptomics may have an additional benefit of revealing more information about the spatiotemporal dynamics of these populations and serve to understand how these populations are structured into the tumor microenvironment.

5. Conclusion

Profiling of single-cells has shown that tumor-infiltrating T cells do not constitute discrete subsets but represent a spectrum of activation, exhaustion and functional modalities. Time point of treatment and response of the patient had a strong influence on T-cell composition and transcriptional programs, with responders demonstrating increased γδ T-cell-mediated activity (TRDV1, TRDC) and non-responders demonstrating suppression of activation and interferon-response markers (CD38, IFI6). Increased translation and metabolic enrichment of additional related T-cell functional states to therapy outcomes. Although such limitations as uneven sampling and the indirect is caused by pseudotime inference are still present, the results indicate the dynamic plasticity of T cells in melanoma and provide a list of TRDV1, TRDC, CD38, and IFI6 as potential biomarkers to predict immunotherapy effectiveness.

References

[1] Tang, T., Huang, X., Zhang, G., & others. (2021). Advantages of targeting the tumor immune microenvironment over blocking immune checkpoint in cancer immunotherapy. Signal Transduction and Targeted Therapy, 6(1), 72.
[2] Yost, K. E., Satpathy, A. T., Wells, D. K., & others. (2019). Clonal replacement of tumor-specific T cells following PD-1 blockade. Nature Medicine, 25(8), 1251–1259.
[3] Paijens, S. T., Vledder, A., de Bruyn, M., & others. (2021). Tumor-infiltrating lymphocytes in the immunotherapy era. Cellular & Molecular Immunology, 18(4), 842–859.
[4] Thommen, D. S., & Schumacher, T. N. (2018). T cell dysfunction in cancer. Cancer Cell, 33(4), 547–562.
[5] Alfei, F., Kanev, K., Hofmann, M., & others. (2019). TOX reinforces the phenotype and longevity of exhausted T cells in chronic viral infection. Nature, 571(7764), 265–269.
[6] Sade-Feldman, M., Yizhak, K., Bjorgaard, S. L., & others. (2018). Defining T cell states associated with response to checkpoint immunotherapy in melanoma. Cell, 175(4), 998–1013.e20.
[7] Eisenberg, E., & Levanon, E. Y. (2013). Human housekeeping genes, revisited. Trends in Genetics, 29(10), 569–574.
[8] Menon, A. P., Moreno, B., Meraviglia-Crivelli, D., & others. (2023). Modulating T cell responses by targeting CD3. Cancers, 15(4), 1189.
[9] Lloyd, S. (1982). Least squares quantization in PCM. IEEE Transactions on Information Theory, 28(2), 129–137. https://doi.org/10.1109/TIT.1982.1056489
[10] Hao, Y., Hao, S., Andersen-Nissen, E., & others. (2021). Integrated analysis of multimodal single-cell data. Cell, 184(13), 3573–3587.e29.
[11] Xiao, Y., Hsiao, T. H., Suresh, U., & others. (2014). A novel significance score for gene selection and ranking. Bioinformatics, 30(6), 801–807.
[12] Frankish, A., Diekhans, M., Ferreira, A. M., & others. (2019). GENCODE reference annotation for the human and mouse genomes. Nucleic Acids Research, 47(D1), D766–D773.
[13] Pliner, H. A., Shendure, J., & Trapnell, C. (2019). Supervised classification enables rapid annotation of cell atlases. Nature Methods, 16(10), 983–986.

DOI: 10.54254/2753-8818/2026.PJ30861

[14] Aran, D., Looney, A. P., Liu, L., & others. (2019). Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nature Immunology, 20(2), 163–172.
[15] Mullan, K. A., Valkiers, S., de Vrij, N., & others. (2025). Where single-cell transcriptomics fails T cells: The misuse of unsupervised clustering for T-cell annotation. ImmunoInformatics, 100063.
[16] Lin, Y., Cao, Y., Kim, H. J., & others. (2020). scClassify: Sample size estimation and multiscale classification of cells using single and multiple reference. Molecular Systems Biology, 16(6), e9389.
[17] Enciso, J., Pelayo, R., & Villarreal, C. (2019). From discrete to continuous modeling of lymphocyte development and plasticity in chronic diseases. Frontiers in Immunology, 10, 1927.
[18] Gao, Z., Bai, Y., Lin, A., & others. (2023). Gamma delta T-cell-based immune checkpoint therapy: Attractive candidate for antitumor treatment. Molecular Cancer, 22(1), 31.
[19] Revach, O. Y., Cicerchia, A. M., Shorer, O., Petrova, B., Anderson, S., Park, J., ... & Jenkins, R. W. (2024). Disrupting CD38-driven T cell dysfunction restores sensitivity to cancer immunotherapy. bioRxiv.
[20] Reschke, R., Gussek, P., Boldt, A., & others. (2021). Distinct immune signatures indicative of treatment response and immune-related adverse events in melanoma patients under immune checkpoint inhibitor therapy. International Journal of Molecular Sciences, 22(15), 8017.
[21] Statello, L., Guo, C. J., Chen, L. L., & Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. Nature Reviews Molecular Cell Biology, 22(2), 96–118.
[22] Schmitt, A. M., & Chang, H. Y. (2016). Long noncoding RNAs in cancer pathways. Cancer Cell, 29(4), 452–463.
[23] Zhu, S., Li, W., Liu, J., & others. (2016). Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR–Cas9 library. Nature Biotechnology, 34(12), 1279–1286.
[24] Vardhana, S. A., Hwee, M. A., Berisa, M., & others. (2020). Impaired mitochondrial oxidative phosphorylation limits the self-renewal of T cells exposed to persistent antigen. Nature Immunology, 21(9), 1022–1033.