Proceedings of the 12th World Congress on Electrical Engineering and Computer Systems and Sciences (EECSS 2026)
August 16 – 18, 2026– London, United Kingdom
Paper No. ICBES 132
DOI: 10.11159/icbes26.132

Baseline Epigenetic Programs of Antigen-Specific Naïve T Cells

Yicheng Zhao[1], Henry Hou[2,†], Celine Wu[2,†], Zoey Chen[2,†], Pracheth Vitthaladevuni[2,†], Huimin Zhang[3], Linda Shi[2]
[1]Department of Biochemistry, University of California, Riverside, Riverside, CA 92507
[2]Institute of Engineering in Medicine, University of California, San Diego
9500 Gilman Dr, La Jolla, San Diego, CA 92093, USA.
[3]Department of Molecular, Cell and Systems Biology, University of California, Riverside, Riverside, CA 92507
yicheng.zhao@email.ucr.edu; hyperbeast618@gmail.com; celinewu92127@gmail.com; bblueoceann1@gmail.com; ppsvzoom@gmail.com; huimin.zhang@ucr.edu; zshi@ucsd.edu
[†]High school students participating in IEM OPALS program

Abstract - Naïve T cells exhibit diverse transcriptional and epigenetic states prior to antigen exposure, which may influence the strength of immune responses. Understanding these baseline molecular states is essential for uncovering the mechanisms that govern immunodominance, T cell priming, and adaptive immunity. This study analyzed publicly available single-cell multimeric datasets of human peripheral blood mononuclear cells (PBMCs) from a healthy donor using integrated scRNA-seq and ATAC-seq data. Quality control filtering and Weighted Nearest Neighbor (WNN) clustering identified distinct naïve T-cell populations characterized by high expression of CCR7 and IL7R. Differential chromatin accessibility analysis revealed uniquely open regions in one naïve T cell cluster, with motif enrichment highlighting transcription factors ETV1 and EGR1 as potential regulators of naïve T cell priming. ChromVAR analysis further confirmed significant enrichment of ETV1 and EGR1 activity. These findings suggest naïve T cells possess pre-existing regulatory programs that may influence activation potential and immunodominance. This study demonstrates the value of integrating transcriptomic and epigenetic analyses to better understand adaptive immune regulation and improve future T cell-based therapies.

Keywords: Naïve T cells, Single-cell RNA sequencing (scRNA-seq), Assay for Transposase-Accessible Chromatin sequencing (ATAC-seq), Chromatin accessibility, Transcription factors, Epigenetic regulation

1. Introduction

Naïve T cells are a critical component of the adaptive immune system, and they serve as precursors to effector and memory T cells during antigen exposure. Adaptive immunity depends on the ability of naïve T cells to recognize antigens and differentiate into specialized effector populations. During antigen exposure, naïve T cells undergo extensive transcriptional and epigenetic remodeling that determines the various factors of immune responses. Recent studies have demonstrated that distinct T cell states are regulated through coordinated transcriptional and chromatin accessibility programs that influence differentiation, exhaustion, and long-term cellular fitness [1]. In particular, exhausted CD8+ T cell subsets exhibit diverse developmental trajectories controlled by underlying epigenetic landscapes and transcription factor networks [2]. These findings suggest that T cell fate decisions may be influenced by pre-existing molecular programs present even before activation occurs [3].

Recent advances in single-cell multiomic technologies have enabled simultaneous analysis of gene expression and chromatin accessibility at single-cell resolution. Bulk T cell receptor repertoire sequencing has improved understanding of inflammation-mediated diseases and immune heterogeneity [4], while single-cell RNA sequencing (scRNA-seq) provides insight into transcriptional activity [5]. Paralleling this, Assay for Transposase-Accessible Chromatin sequencing (ATAC-seq) identifies regions of open chromatin associated with active regulatory elements and transcription factor binding [6]. Integrating these modalities allows researchers to connect epigenetic regulation with downstream gene expression, providing a more comprehensive understanding of immune cell heterogeneity and functional potential. These technologies allow identification of regulatory signatures that may predispose naïve T cells toward stronger or weaker activation responses. Computational integration methods such as Weighted Nearest Neighbor

(WNN) analysis further improve the identification of distinct cellular populations by combining transcriptional and epigenetic information into one unified framework.

In this study, publicly available single-cell multiomic datasets of human peripheral blood mononuclear cells (PBMCs) from a healthy donor were analyzed to investigate regulatory differences among naïve T cell populations. Through quality control filtering, WNN clustering, and downstream chromatin accessibility analyses, distinct naïve T cell subsets were identified and characterized. Differential accessibility and motif enrichment analyses highlighted transcription factors ETV1 and EGR1 as potential regulators associated with uniquely accessible chromatin regions. By integrating scRNA-seq and ATAC-seq analyses, this study aims to establish a computational framework for understanding how pre-existing molecular states influence the transition from naïve to activated T cells and contribute to adaptive immune responsiveness.

2. Methods

Using datasets from 10x Genomics, data analysis was performed using R4.5.2 and R Studio. This study aimed to compare chromatin accessibility and transcriptional programs of the antigen-specific naïve T cells. Using quality control and analysis, external factors were removed to only target the marker genes of the T cells.

Initially, quality control was administered on the single-cell RNA sequencing data (scRNAseq) and single-cell Assay for Transposase-Accessible Chromatin sequencing (ATACseq) to eliminate dead, stressed, or damaged cells. Cells with high mitochondrial RNA (>15%), indicative of dead or stressed cells, were excluded to get relevant results. Potential cell doublets were excluded from the analysis to avoid technical errors.

The ATAC and RNA datasets included many cell-specific features. To develop a unified single representation, the scRNAseq and scATACseq were integrated with Weighted Nearest Neighbor (WNN) analysis in Seurat. WNN assigns different weights to data sets rather than treating them equally. The algorithms calculated the weights of each cell using RNA specificity under the Kernel Function. The WNN function calculated the individual cell-specific modality weights for each cell and combined them into a single multimodal graph.

To isolate the T cells, external factors, including B Cells and NK cells, were subsetted from the dataset computationally using marker identification. Marker genes, including IL7R and CCR7, were targeted while identifying naïve T Cells. The exact naïve T cell clusters were then extracted through UMAP.

Differential analysis was then performed to locate epigenetically favorable regions. Differentially Accessible Regions (DAR) of DNA are displayed where the chromatin accessibility was significantly changed. The downstream analysis of the DAR clusters showed that the epigenetically favored regions modulated transcriptional outcomes during T cell differentiation.

In the open DNA sequences of Cluster 4, the transcription factors bind to a specific sequence. To determine which sequence is enriched, ChromVAR analysis software was used to predict the transcription factors controlling the cells. This algorithm computes a deviation score and determines the openness and the overall accessibility of a motif. The significance of the deviation score was set at $P<0.05$, and it determines which transcription factors control T cell differentiation.

3. Results

To characterize regulatory variation within naïve T cells, integrated RNA expression and chromatin accessibility data was examined across PBMC populations. Distinct cell clusters were identified, enabling the isolation of naïve T cell subsets. Comparison of subsets revealed differences in chromatin accessibility and transcription factor activity, indicating regulatory heterogeneity prior to antigen exposure.

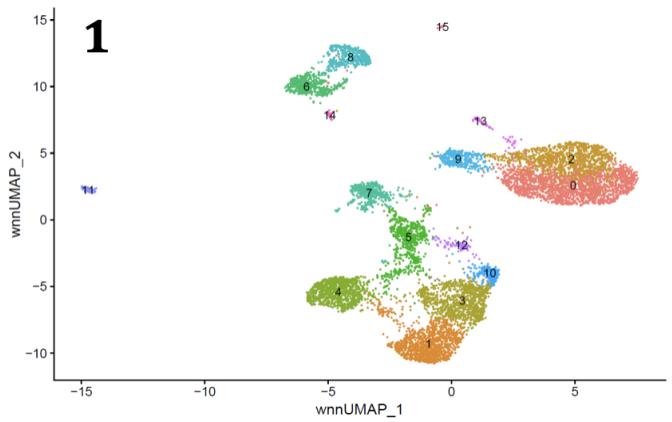

Fig. 1: UMAP clustering of the single cell PBMC data into sixteen distinct cell groups

The integrated UMAP visualization of single-cell PBMC data following Weighted Nearest Neighbor (WNN) analysis of RNA and ATAC modalities is shown in Figure 1. Dimensional reduction separated the cells into sixteen different clusters, representing heterogeneous immune cell populations with unique transcriptional and chromatin accessibility profiles. The separation between clusters indicates successful integration of the multiomic datasets and provided the bases for subsequent identification and isolation of naïve T cell populations for downstream analysis.

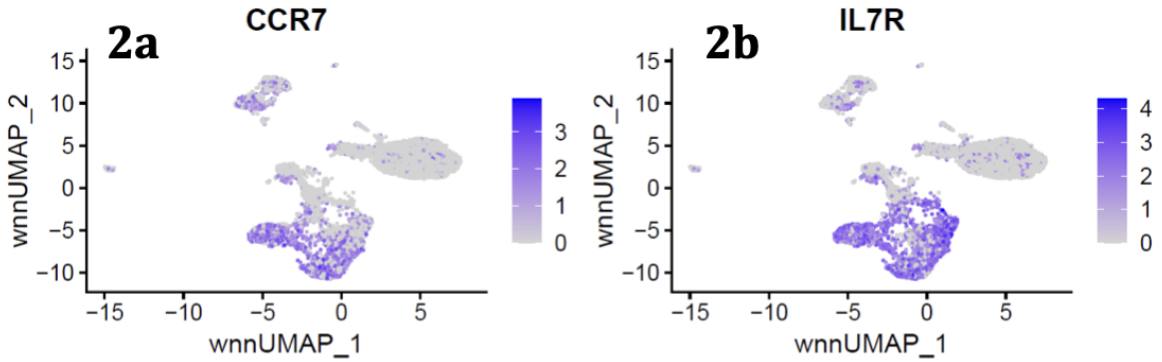

Fig. 2: Naïve T cell marker gene expression in UMAP plot (a) CCR7 (b) IL7R

In Figure 2, we look at naïve T cell-specific marker genes CCR7 (2a) and IL7R (2b), identifying clusters 1 and 4 as primary naïve T cell groups because of high marker gene expression. Differential chromatin accessibility analysis showed uniquely accessible chromatin regions in cluster 4 compared to cluster 1, suggesting distinct epigenetic states between naïve T cell populations. Motif enrichment analysis revealed elevated activity of transcription factors ETV1 and EGR1 across both clusters. ETV1 with statistically significant activity enrichment (Wilcoxon P = 0.009) and EGR1 with the most significant statistical enrichment of P = 0.004.

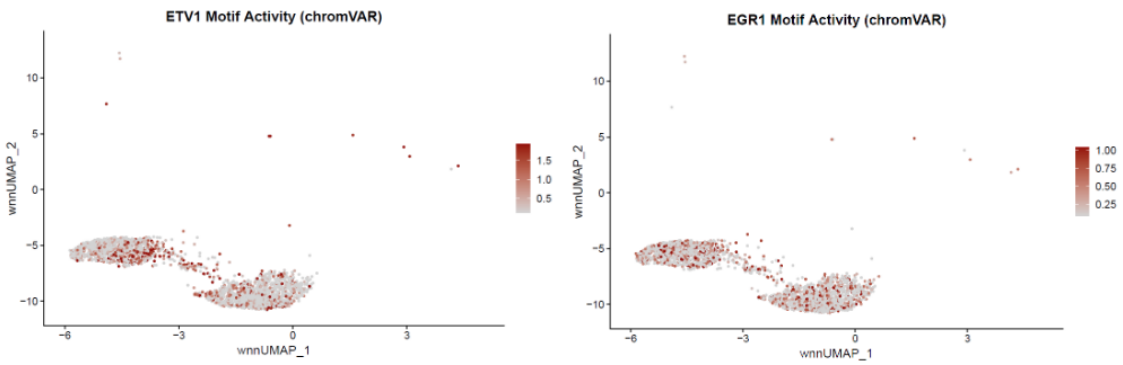

Fig. 3: Plotted are deviation scores assigned by chromVAR for transcription factor motif activity in subsetted naïve T cell clusters for (a) ETV1 (b) EGR1

Through ChromVAR analysis, Figure 3 supports strong ETV1 and EGR1 activity among isolated naïve T cell clusters 1 and 4 because of high motif activity. ETV1 and EGR1 emerge as candidate regulators associated with chromatin accessibility differences among naïve T-cell populations. Results suggest that unique pre-existing regulatory programs may contribute to differences in T cell differentiation potential.

4. Discussion

This study identifies measurable differences among naïve T cell populations within a single-cell PBMC dataset. Clusters 1 and 4 were identified as naïve T cell groups based on high CCR7 and IL7R expression. Although both clusters showed marker expression consistent with naïve T cells, differential chromatin accessibility analysis revealed accessibility differences between them, suggesting that naïve T cells in this dataset are not completely uniform at the chromatin level.

The differences between clusters 1 and 4 suggest that cells classified under the same broad naïve T cell identity can show variation in chromatin accessibility within this dataset. Since both clusters express CCR7 and IL7R, the observed differences may reflect heterogeneity within naïve T cell populations rather than a comparison between clearly separate immune cell types. However, this analysis does not determine whether these chromatin accessibility differences correspond to functional differences in activation, proliferation, or differentiation after antigen exposure.

Motif enrichment analysis identified ETV1 and EGR1 motifs as enriched features associated with these chromatin accessibility differences. ETV1 showed statistically significant enrichment by Wilcoxon test ($P = 0.009$), while EGR1 showed the strongest statistical enrichment among the analyzed motifs ($P = 0.004$). These results suggest that ETV1- and EGR1-associated motifs are present within accessible regions that distinguish the naïve T cell clusters. However, motif enrichment does not prove that ETV1 or EGR1 directly bind to these regions or regulate naïve T cell function.

5. Conclusion

This study identifies transcriptional and chromatin accessibility differences among naïve T cell populations prior to antigen exposure. By integrating scRNA-seq and ATAC-seq data through WNN analysis, we identified naïve T cell clusters characterized by differences in CCR7 and IL7R expression. Motif enrichment and chromVAR analyses further identified ETV1 and EGR1 motif activity as features associated with differences in chromatin accessibility between naïve T cell clusters. These findings support the presence of measurable regulatory heterogeneity among naïve T cells and highlight the value of single-cell multiomic analysis for characterizing baseline cellular states within adaptive immune populations. Further experimental validation will be required to determine whether these observed regulatory differences have functional effects on T cell activation, immune responsiveness, or differentiation.

6. Limitations

Analysis was conducted using a single PBMC dataset. While this dataset provided sufficient resolution to identify transcriptional and chromatin accessibility patterns within naïve T cells, additional donors may exhibit different chromatin accessibility patterns. The ratio of chromatin amounts in different donors varies because cells tightly regulate

DNA transcription based on genetics, age, and environment. Therefore, the findings should be interpreted as patterns observed within this dataset rather than as conclusions that necessarily apply to all naïve T cell populations.

Additionally, only computational analysis was performed. Pre-existing data was used, and no new datasets were created. Future experiments such as flow cytometry and ChIP-seq would be required to validate the proposed regulatory roles of ETV1 and EGR1. Because no experimental validation was performed, the results identify associations between chromatin accessibility, motif enrichment, and naïve T cell clusters rather than confirmed biological mechanisms.

While motif enrichment analyses identify statistical associations between motifs and accessible chromatin regions they do not establish direct regulatory activity or causality. Therefore, while the enrichment of ETV1 and EGR1 motifs suggest the potential role of these transcription factors, further studies should be conducted to determine direct regulatory function. Experimental approaches such as transcription factor binding assays or perturbation studies would be needed to test whether ETV1 and EGR1 directly influence gene regulation or T cell behavior.

Acknowledgments

This work was supported by a generous gift from Beckman Laser Institute, Inc. to the UC San Diego Biophotonics Technology Center. We also extend our sincere gratitude to the donors of the UCSD OPALS Program, including Leighton Foundation, Inc., Ming-Wei Wu, Inc., Dr. Shu Chien (UCSD Bioengineering), Dr. Lizhu Chen (CorDx Inc.), Dr. Xinhua Zheng, David and Leslie Lee, and Mingwei Hu and Wen Shi, for their invaluable support and commitment to advancing research and education at the interface of biophysics and biomedical engineering.

References

[1] J. C. Beltra, S. Manne, M. S. Abdel-Hakeem, M. Kurachi, J. R. Giles, Z. Chen, V. Casella, S. F. Ngiow, O. Khan, Y. J. Huang, P. Yan, K. Nzingha, W. Xu, R. K. Amaravadi, X. Xu, G. C. Karakousis, T. C. Mitchell, L. M. Schuchter, A. C. Huang, and E. J. Wherry, "Developmental Relationships of Four Exhausted CD8+ T Cell Subsets Reveals Underlying Transcriptional and Epigenetic Landscape Control Mechanisms," Immunity, vol. 52, no. 5, pp. 825-841.e8, May 19, 2020.

[2] B. Daniel, K. E. Yost, S. Hsiung, K. Sandor, Y. Xia, Y. Qi, K. J. Hiam-Galvez, M. Black, C. J. Raposo, Q. Shi, S. L. Meier, J. A. Belk, J. R. Giles, E. J. Wherry, H. Y. Chang, T. Egawa, and A. T. Satpathy, "Divergent clonal differentiation trajectories of T cell exhaustion," Nature Immunology, vol. 23, no. 11, pp. 1614-1627, Nov. 2022.

[3] R. Zander and W. Cui, "Exhausted CD8+ T cells face a developmental fork in the road," Trends in Immunology, vol. 44, no. 4, pp. 276-286, 2023.

[4] A. K. H. Mahdy, E. Lokes, V. Schöpfel, V. Kriukova, O. V. Britanova, T. A. Steiert, A. Franke, and H. ElAbd, "Bulk T cell repertoire sequencing (TCR-Seq) is a powerful technology for understanding inflammation-mediated diseases," Journal of Autoimmunity, vol. 149, p. 103337, 2024.

[5] D. Jovic, X. Liang, H. Zeng, L. Lin, F. Xu, and Y. Luo, "Single-cell RNA sequencing technologies and applications: A brief overview," Clinical and Translational Medicine, vol. 12, no. 3, e694, 2022.

[6] J. D. Buenrostro, B. Wu, H. Y. Chang, and W. J. Greenleaf, "ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide," Current Protocols in Molecular Biology, vol. 109, pp. 21.29.1-21.29.9, 2015.