Unveiling Cellular Dynamics: Single-Cell Profiling of Human Bone Marrow

Popov Ilia

Abstract

This CITE-seq analysis investigates the cellular landscape of human bone marrow through a dataset capturing 33,454 cells. Utilizing Scanpy, a robust computational approach was applied to perform quality control, normalization, logarithmic transformation, and identification of highly variable genes. Eight distinct cell populations were characterized, including various T cell subsets and dendritic cell types. Marker gene identification, informed by contemporary studies, facilitated the annotation of these cell groups. Notably, the analysis illuminated the differential yet overlapping transcriptional signatures of monocytes and myeloid dendritic cells, reflecting their divergent roles despite a shared progenitor in the bone marrow.

Keywords: Single cell, scRNA-seq, CITE-seq

Introduction

The bone marrow serves as a critical site for hematopoiesis, hosting a complex assortment of cells that form the backbone of the immune system[1]. Advances in single-cell sequencing, particularly Cellular Indexing of Transcriptomes and Epitopes by sequencing (CITE-seq), have unlocked the potential to dissect this complexity in detail[2]. This report delves into the transcriptional landscape of human bone marrow cells to reveal the nuanced interplay of immune cell types within this environment.

Materials and Methods

For this project dataset of CITE-seq experiment on human bone marrow cells, capturing 33,454 cells, was used[3].

Data analysis was conducted using Scanpy[4], following the methodology provided in the developers' tutorial[5]. The workflow included: quality control, normalization and logarithmic transformation of the data, identification of highly variable genes, principal component analysis, computation of the neighborhood graph, clustering of the neighborhood graph, and marker gene identification.

Quality Control parameters: Cells were filtered based on specific criteria:

- Cells with fewer than 3000 genes detected.
- Cells with a total count fewer than 9000.
- Cells with less than 10% mitochondrial gene counts.

The identification of highly variable genes utilized the 'seurat' parameter for automated parameter selection.

The neighborhood graph was computed with default parameters (n_neighbors=10, n_pcs=40). Clustering was performed with a resolution of 0.15, resulting in eight cell groups.

The identification of marker genes was done manually and was guided by data from recent publications[6].

Results

Quality Control

Table 1 presents the outcomes of the data quality control, illustrating the number of cells prior to and subsequent to filtration. Figures 1-2 provide additional visual data. The quality control process effectively eliminated data outliers.

Table 1 Results of CITE-seq dataset quality control.

Cells before	Cells after
33,454	31,382

Annotated cell types

Figure 3 revealed eight major groups of cells from the dataset which included:

- Proliferating macrophages and monocytes
- CD4+ transitional memory
- CD8+ terminally exhausted
- Monocytes
- B cells
- Plasma B cells
- Plasmacytoid dendritic cells (pDCs)
- Myeloid dendritic cells (mDCs)

Marker genes

The list of marker genes was established based on recent work[6], as shown in Table 2.

Visualization of marker genes and annotated cell types is provided in Figure 4.

Table 2 List of marker genes for each cell type

Group	Markers	Cell type
0	PCNA, TOP2A, STMN1, TUBA1B	Macro. and mono. prolif
1	RORA, CCR7	CD4 transitional memory
2	FCN1, VCAN, AIF1	Monocytes
3	CCL5, NKG7, GZMA	CD8 terminally exhausted
4	CD79A, CD79B, MS4A1	B cells
5	IRF8, TCF4, BCLL11A, SPIb	pDC
6	SPI1, CD68	mDC
7	TNFRSF17, CD38, IGHG1	Plasma B cells

Discussion

The annotation of seven out of the eight cell populations demonstrated definitive results. An exception was observed in the CD4+ transitional memory group, which showed a lower expression of genes, annotated specifically with the markers RORA and CCR7. A point of particular interest is the overlap in the gene expression profiles of Monocytes and Myeloid dendritic cells (mDCs). Monocytes were primarily identified by the expression of FCN1, VCAN, and AIF1, while exhibiting lower levels of SPI1 and CD68. In contrast, mDCs shared the expression of FCN1 and AIF1 but were distinctively marked by higher levels of SPI1 and CD68, which are more definitive for this group. This annotation approach considered the bone marrow origin of these cells, recognizing the shared lineage yet distinct transcriptional and functional profiles—monocytes being key in inflammatory processes and mDCs in antigen presentation and T cell activation[7,8].

Supplementary materials

Lab journal can be found in GitHub repository.

Literature cited

1. Lucas, D. Structural organization of the bone marrow and its role in hematopoiesis. *Curr. Opin. Hematol.* **28**, 36–42 (2021).
2. Stoeckius, M. *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017).
3. Stuart, T. *et al.* Comprehensive integration of single-cell data. *Cell* **177**, 1888-1902.e21 (2019).
4. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
5. Scanpy tutorials — scanpy-tutorials 0.1.dev50+g02c4946 documentation. https://scanpy-tutorials.readthedocs.io/en/latest/.
6. Nirmal, A. J. *et al.* Immune Cell Gene Signatures for Profiling the Microenvironment of Solid Tumors. *Cancer Immunol. Res.* **6**, 1388–1400 (2018).
7. van Leeuwen-Kerkhoff, N. *et al.* Human Bone Marrow-Derived Myeloid Dendritic Cells Show an Immature Transcriptional and Functional Profile Compared to Their Peripheral Blood Counterparts and Separate from Slan+ Non-Classical Monocytes. *Front. Immunol.* **9**, 1619 (2018).
8. Wacleche, V. S., Tremblay, C. L., Routy, J.-P. & Ancuta, P. The Biology of Monocytes and Dendritic Cells: Contribution to HIV Pathogenesis. *Viruses* **10**, 65 (2018).

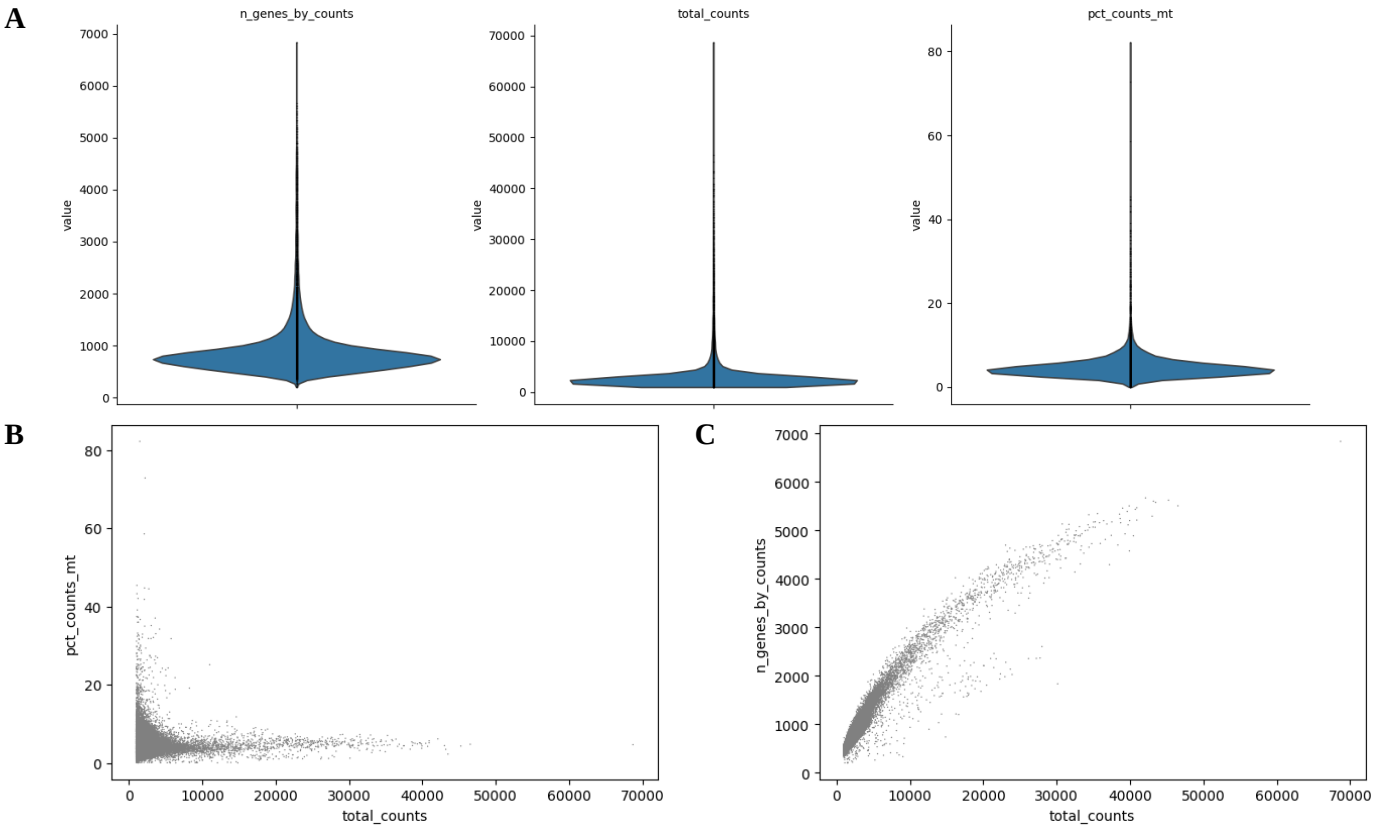


Figure 1 CITE-seq dataset metrics before cells filtration. A – violin plots of the number of genes expressed in the count matrix, the total counts per cell, the percentage of counts in mitochondrial genes; B – scatter plot of the percentage of counts in mitochondrial genes by total counts per cell; C - scatter plot of the number of genes expressed in the count matrix by total counts per cell.

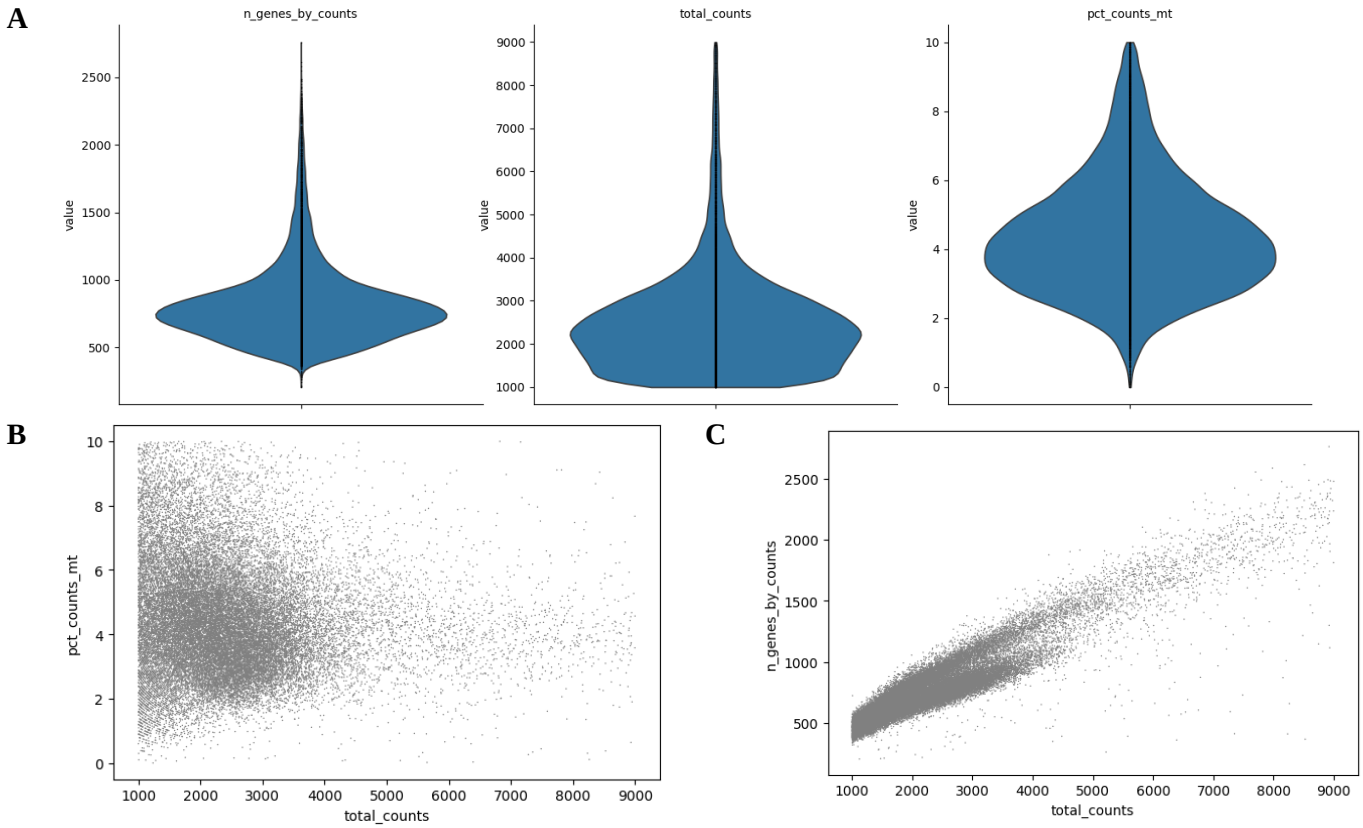


Figure 2 CITE-seq dataset metrics after cells filtration. A – violin plots of the number of genes expressed in the count matrix, the total counts per cell, the percentage of counts in mitochondrial genes; B – scatter plot of the percentage of counts in mitochondrial genes by total counts per cell; C - scatter plot of the number of genes expressed in the count matrix by total counts per cell.

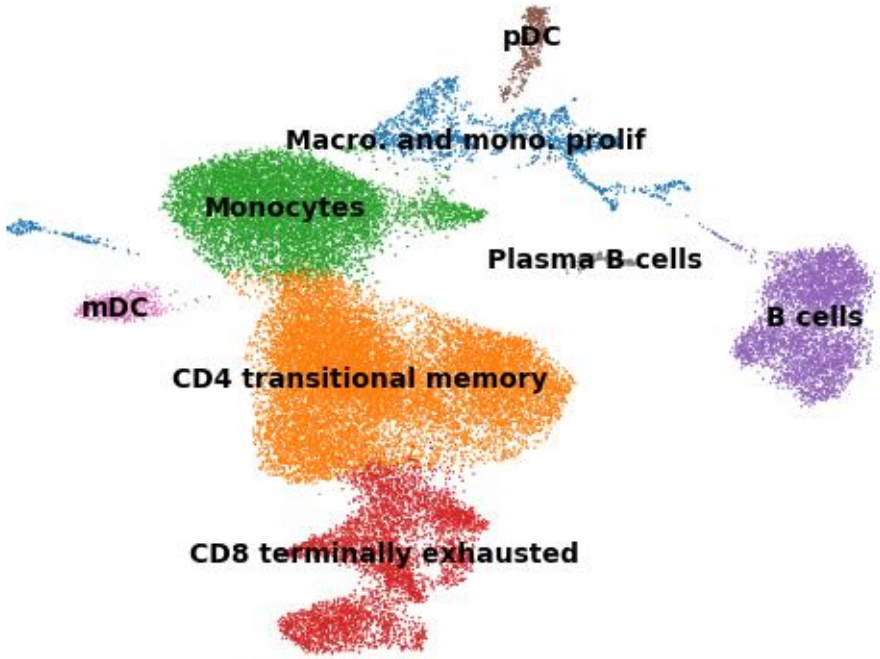
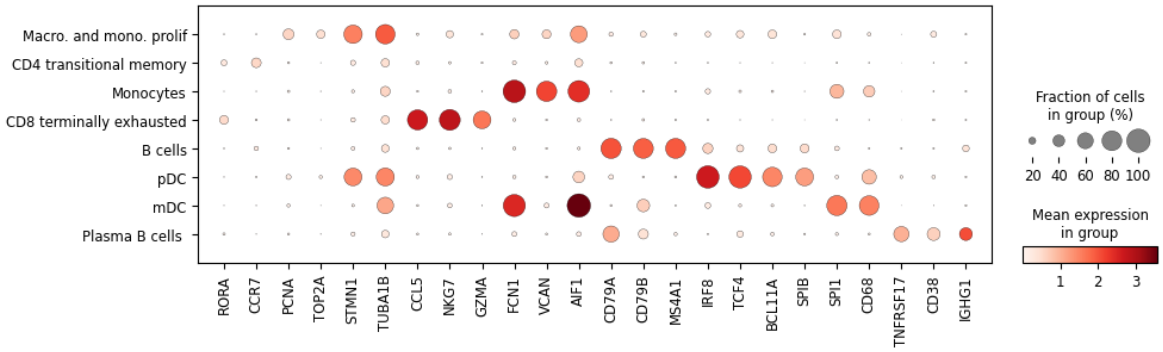


Figure 3 T-distributed Stochastic Neighbor Embedding (t-SNE) Clustering of Annotated Cell Types.

A



B

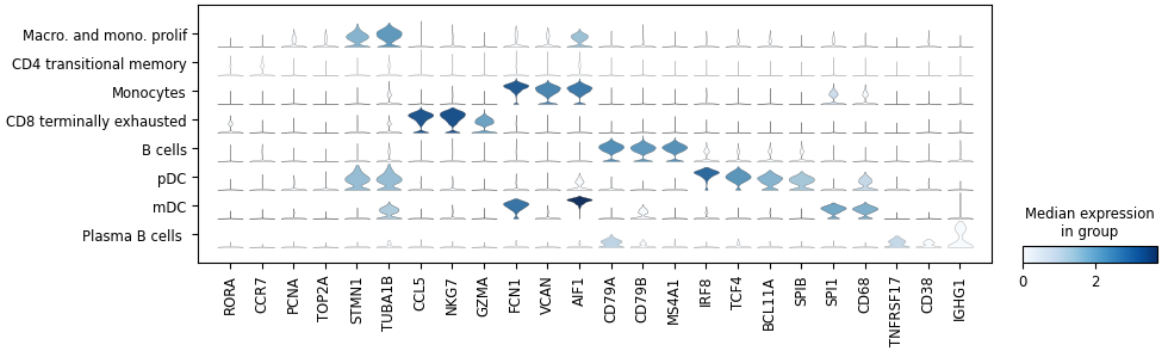


Figure 4 Marker genes visualization. A – Dot Plot Visualization of Marker Gene Expression Across Cell Types; B – Violin Plot Distribution of Marker Gene Expression Across Cell Types.