Single-Cell RNA Sequencing of Peripheral Blood Mononuclear Cells

Gain comprehensive insight into gene expression with the Illumina® Bio-Rad® SureCell™ WTA 3′ Library Prep Kit for the ddSEQ™ System.

Introduction

The immune system consists of cells, tissues, and secreted molecules that mediate recognition and removal of foreign or "nonself" material, pathogens, cancer cells, and graft transplantations.[1–4] Failure of the immune system to recognize tissue or cells as "self" can lead to autoimmune diseases that have significant impact on human health.[5] To delineate the mechanisms that drive immune function in both health and disease states, researchers often perform gene expression studies on immune cells isolated from peripheral blood, including peripheral blood mononuclear cells (PBMCs). PBMCs are characterized by a round, uniform nucleus and include lymphocytes (B- and T-cells), dendritic cells, macrophages, and NK-cells.[1] RNA Sequencing (RNA-Seq) provides an accurate method to measure gene expression of the whole transcriptome without prior knowledge of the genes expressed. However, traditional RNA-Seq is done on cells processed in bulk, which averages gene expression across a heterogeneous cell population.[6] Single-cell RNA-Seq enables in-depth gene expression analysis that can provide insight into how individual cells contributes to the function of a complex tissue.[7]

The Illumina Bio-Rad Single-Cell Sequencing Solution combines the innovative Bio-Rad Droplet Digital™ technology with Illumina next-generation sequencing (NGS) library preparation, sequencing, and analysis. This solution provides a comprehensive, user-friendly workflow for single-cell RNA-Seq that enables controlled experiments with multiple samples, treatment conditions, and time points. Built and supported in collaboration between technology leaders, the Illumina Bio-Rad Single-Cell Sequencing Solution enables transcriptome analysis of hundreds to thousands of single cells in a single experiment. The simple push-button analysis for alignment, cell decoding, and library quality control (QC) in the SureCell RNA Single-Cell App in BaseSpace™ Sequence Hub, combined with data reduction and population identification tools in SeqGeq Software, can use gene expression profiles to resolve heterogeneous cell populations and identify subpopulations of interest.

To demonstrate the high-quality single-cell RNA-Seq data achieved with the Illumina Bio-Rad Single-Cell Sequencing Solution, experiments were performed with fresh and frozen PBMCs.

Methods

Cell Isolation and Library Prep

PBMCs were prepared from a single healthy donor for single-cell sequencing using the Illumina Bio-Rad SureCell WTA 3′ Library Prep Kit for the ddSEQ System (Illumina, Catalog No. 20014280).[8] PBMCs were processed from a total of eight sample chambers from two cartridges onto the Bio-Rad ddSEQ Single-Cell Isolator (Bio-Rad, Catalog No. 12004336). After completion of cDNA synthesis, cDNA from two chambers were combined into a single reaction before tagmentation, resulting in generation of four indexed libraries (Table 1).

To read a validated protocol for combining two chambers of isolated PBMCs from a single cartridge for the Bio-Rad ddSEQ Single-Cell Isolator, visit www.illumina.com/surecell

Sequencing and Data Analysis

Pooled libraries were sequenced on the NextSeq™ 500 System at ~150,000 reads/cell and downsampled to different read depths where indicated. Sequencing data were analyzed using the SureCell RNA Single-Cell App in BaseSpace Sequence Hub, downsampled to ~70,000 reads (Table 1). Knee plots were used to identify individual cells statistically and filter out empty beads based on genic unique molecular identifier (UMI) counts per cell barcode (Figure 1).

Clustering Analysis with Seurat

Unbiased clustering analysis was performed using the Seurat R toolkit for single-cell genomics developed by the Satija Lab at New York Genome Center. Following the guided tutorial for PBMCs, a total of nine cell clusters were identified with 3354 cells (Figure 2 and Table 2).

Table 1. Library Pooling Strategy

Library	Total No. of Observed Bead Barcodes	Total No. of Cells Passing Knee Filter	Median No. of Genic UMIs per Cell Passing Filter	Median No. of Genes Detected in Cells Passing Filter	% Valid Barcodes	% Aligned Reads	No. of PF Reads per Cell	Average No. of Cells per Chamber
Library 1	138,157	1187	882	417	78.87	72.51	70,767	594
Library 2	138,865	840	900	436	79.38	74.86	70,238	420
Library 3	142,711	754	1242	511	78.43	74.45	70,557	377
Library 4	128,788	573	981	425	80.13	79.14	63,493	287

Abbreviations: UMI, unique molecular identifier; PF, passing filter.

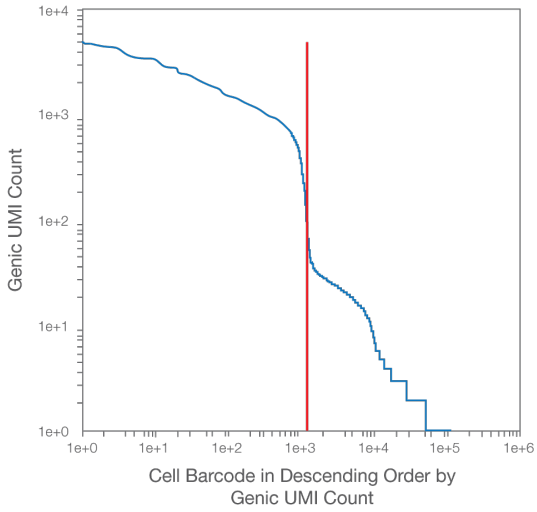


Fig. 1. Filtering Out Noncellular Barcodes from Downstream Analysis—Plotting genic UMI counts against cell barcodes in descending order by genic UMI count enables statistical identification of "true" cells. Cell barcodes to the left of the threshold (vertical red line) have genic UMI counts in the thousands, representing true cells. Cell barcodes to the right of the threshold have genic UMI counts of 1–100, typically below what is expected for live, intact cells. These likely represent empty beads.

Advanced Data Visualization with SeqGeq Software

SeqGeq Software is a desktop application for advanced data analysis, exploration, and visualization of single-cell gene expression data developed by FlowJo, LLC for the Illumina Bio-Rad Single-Cell Sequencing Solution. SeqGeq Software features powerful data reduction and population identification tools. Direct integration with BaseSpace Sequence Hub enables visualization and analysis of expression data with statistic color-mapping of individual cells, summary heatmaps, and drag-and-drop report editors (Figure 3).

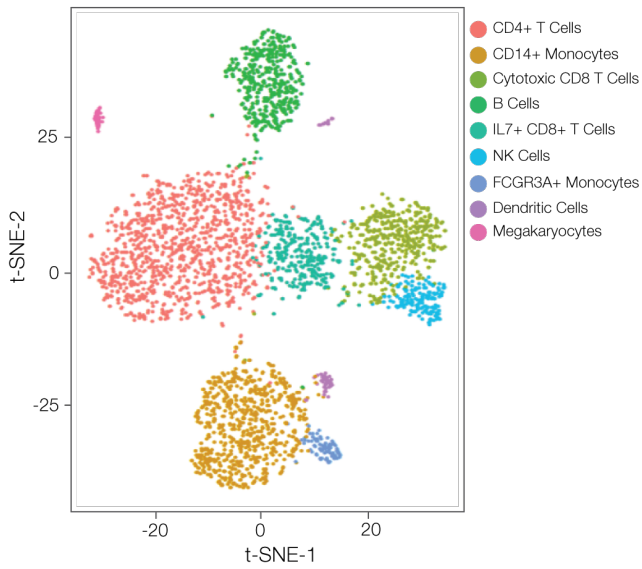


Fig. 2. Unbiased Cluster Analysis of PBMCs in Seurat—Nine cell clusters were identified with 3354 cells down-sampled to approximately 70,000 reads per cell, with a resolution setting of 0.80 and 100 genes as a cut off.

Table 2. Cell Populations Identified with Seurat

Cell Population	Markers	% of Cells
CD4+ T Cells	CD8A- | IL7R+ | CD3D+	34.3%
CD14+ Monocytes	CD14+ | LYZ+	20.9%
CD8+ Cytotoxic T Cells	CD8A+ | GZMB+ | CD3D+	12.9%
B Cells	MS4A1+	12.5%
IL7+/CD8+ T Cells	CD8A+ | IL7R+ | CD3D+	9.0%
NK Cells	NKG7+ |GNLY+	4.7%
FCGR3A+ Monocytes	FCGR3A+ | MS4A7+	2.5%
Dendritic Cells	FCER1A+	1.8%
Megakaryocytes	PPBP+	1.3%

To access a detailed tutorial on clustering analysis of PBMCs with Seurat, visit satijalab.org/seurat/pbmc3k_tutorial.html

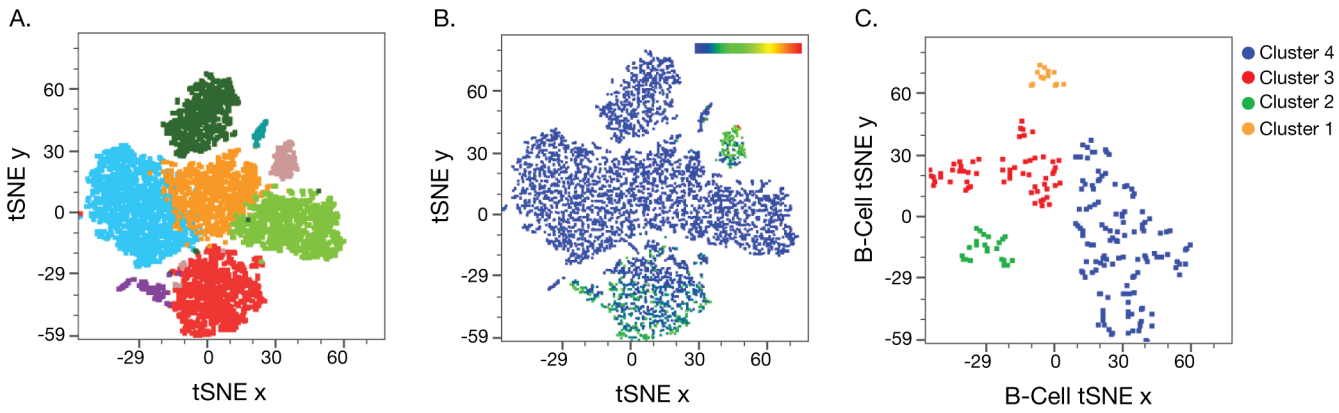


Fig. 3. Simplified Clustering Analysis in SeqGeq—(A) Unbiased clustering analysis of PBMCs based on differential expression of cell-specific genes/markers. (B) Example of using gene set enrichment to identify B cells (green) based on expression of B-cell genes (blue corresponds to low expression and red to high in heat map). (C) B cells identified in (B) undergo further unbiased clustering using PCA guided tSNE to identify subpopulations of B cells. Learn more at www.flowjo.com/seqgeq

Effect of Read Depth on Clustering Analysis

To investigate the effect of read depth on cluster detection, a separate experiment was conducted using PBMCs across four chambers of one ddSeq chip generating two indexed libraries totaling ~2000 cells. Libraries were downsampled from 150,000 to 50,000 reads per cell and unbiased clustering analysis was conducted using Seurat as previously described. With this data set, Seurat detected seven clusters from 70,000 to 150,000 reads per cell (Table 3). At 50,000 reads per cell, a single cluster dropped out.

Table 3. Effect of Read Depth on PBMC Cluster Identification with Seurat

No. of Reads per Cell	150K Reads	100K Reads	70K Reads	50K Reads
Total No. of Cells	2021	2030	2025	1997
No. of Clusters Detected	7	7	7	6
Median No. of Genes Detected per Cell	461	442	428	414

Clustering Analysis of Fresh and Frozen PBMCs

Single-cell RNA-Seq approaches are performed typically on fresh cells processed directly following sample acquisition. Limiting analysis to freshly prepared samples hinders complex study designs involving cells collected at different times or locations. The ability to preserve samples by freezing them, while maintaining the state of the transcriptome, would disconnect time and place of sample collection from subsequent processing steps.[9] This would significantly broaden the scope of applications and accessible specimens for single-cell RNA-Seq.

To demonstrate the capability of the SureCell WTA 3′ Library Prep Kit, PBMCs from the same healthy donor were processed as described using either freshly prepared or previously frozen cells. Resulting libraries were sequenced and analyzed as described. Unbiased clustering of merged data from both fresh and frozen PBMC samples was performed using the Seurat program shows an even distribution of fresh and frozen PBMCs across all eight clusters that were identified (Figure 4). The performance metrics between fresh and frozen samples were similar (Table 4).

Table 4. Comparison of Clustering Analysis of Fresh vs. Frozen PBMCs

Metric	Fresh	Frozen
No. of Cells	2868	2313
Genes per Cell	355	386
UMI per Cell	826	860

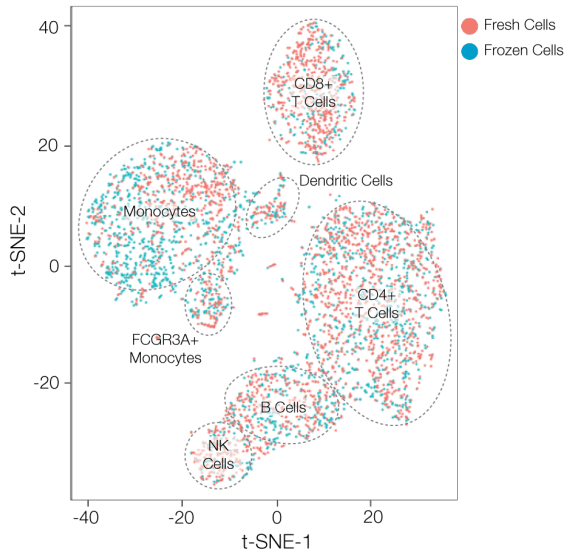


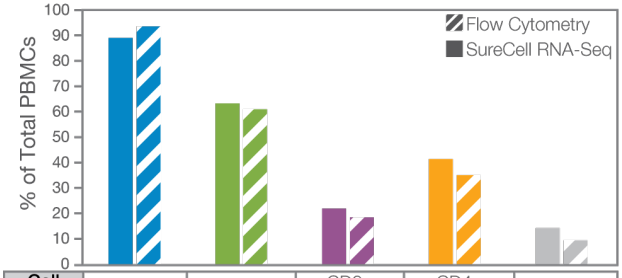
Fig. 4. Clustering Analysis of Fresh vs. Frozen PBMCs—Unbiased clustering analysis of bioinformatically merged data from both fresh and frozen PBMCs identified eight clusters using the Seurat pipeline. Fresh and frozen cells were evenly distributed among all seven clusters. A resolution setting of 0.65 was used for cluster determination.

Comparison of RNA-Seq with the Illumina Bio-Rad Single-Cell Sequencing Solution to Flow Cytometry

Flow cytometry represents the gold standard for classification of cell types by defining cell surface markers.[10] PBMC subtypes identified by flow cytometry were generated to benchmark data produced with the Illumina Bio-Rad Single-Cell Sequencing Solution. The percent of PBMCs from a healthy donor expressing each of five cell-type specific cell surface markers (Figure 5A) was determined by flow cytometry. In parallel, SureCell libraries were generated from the same PBMC sample, sequenced on the NextSeq 500 System and analyzed with the SureCell RNA Single-Cell App.

Unbiased clustering analysis with Seurat was used to identify distinct cell populations from the SureCell dataset. Each cluster was evaluated for expression of the gene markers (Figure 5A) that correspond to the cell types and markers identified by flow cytometry. There is a high correlation ($R^2 = 0.995$) between the percentage of cells assigned to each subpopulation between flow cytometry and single-cell RNA-Seq (Figure 5B). This demonstrates equivalent performance in PBMC subtype identification by the Illumina Bio-Rad Single-Cell Sequencing Solution, as compared to the gold standard method of flow cytometry. Furthermore, flow cytometry is very limited in the number of markers that can be probed in a single experiment, while single-cell RNA-Seq allows for researchers to evaluate expression of thousands of gene markers in any individual cell.

A.



Cell Type	Leukocytes	T Cells	CD8+ T Cells	CD4+ T Cells	B Cells
Cell Surface Marker	CD45+	CD3+	CD8+	CD4+	CD19+
Gene Marker	*PTPRC+*	*CD3D+*	*CD8A+*	*IL7R+ CD8A-*	*MS4A1+*

B.

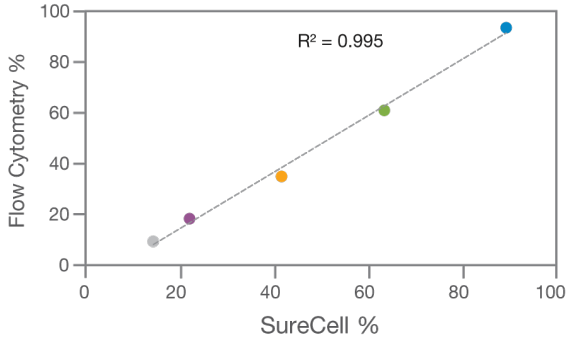


Fig. 5. Correlation of PBMC Identification with Single-Cell RNA-Seq to Flow Cytometry—(A) Comparison of PBMC sub-type identification between flow cytometry (solid bars) and SureCell RNA-Seq (hashed bars). Cell types and corresponding cell surface markers and gene markers are indicated in the table below. (B) Data plot showing the high correlation between flow cytometry and SureCell RNA-Seq.

Summary

The diverse range of cell types that make up PBMCs necessitates the in-depth gene expression analysis provided by single-cell RNA-Seq to gain insight into how individual cells contribute to the function of a complex tissue such as peripheral blood. The Illumina Bio-Rad Single-Cell Sequencing Solution enables transcriptome analysis of hundreds to thousands of single cells in a single experiment. This application note presents results from unbiased clustering analysis of PBMCs with two different software programs. While performance is comparable between the two, SeqGeq is designed to integrate directly with the SureCell RNA Single-Cells App, providing a simplified, user-friendly interface for researchers without command line experience.

Results from a PBMC downsampling experiment demonstrate that the read depth can be reduced without significant impact to cluster identification. Performance of clustering analysis is comparable between freshly prepared and previously frozen PBMCs, which significantly expands flexibility of experimental design and the scope of accessible samples. The high correlation of PBMC sub-type identification between flow cytometry and the Illumina Bio-Rad Single-Cell Sequencing Solution indicate that it is ideally suited for analysis and identification of the diverse cell types that comprise PBMCs.

Learn More

To learn more about the Illumina Bio-Rad Single-Cell Sequencing Solution, visit www.illumina.com/surecell or www.bio-rad.com/ddSEQ

To learn more about SeqGeq, visit: docs.flowjo.com/seqgeq/flowjo-seqgeq-documentation/pbmc-tutorial/

References

1. Abbas, AK, Leichtman, AH. *Basic Immunology: Functions and Disorders of the Immune System*. 3rd ed. Philadelphia, PA: Saunders/Elsevier; 2009.
2. Neller MA, Burrows JM, Rist MJ, Miles JJ, Burrows SR. High frequency of herpesvirus-specific clonotypes in the human T-cell repertoire can remain stable over decades with minimal turnover. *J Virol*. 2013;87(1):697–700.
3. Haen SP, Rammensee HG. The repertoire of human tumor-associated epitopes-- identification and selection of antigens and their application in clinical trials. *Curr Opin Immunol*. 2013;25(2):277–283.
4. Meyer EH, Hsu AR, Liliental J, et al. A distinct evolution of the T-cell repertoire categorizes treatment refractory gastrointestinal acute graft-versus host. *Blood*. 2013;121(24):4955–4962.
5. American Autoimmune Related Diseases Association. The Cost Burden of Autoimmune Disease: The Latest Front in the War on Healthcare Spending. www.diabetesed.net/page/_files/autoimmune-diseases.pdf. Published January 2011. Accessed September 22, 2017.
6. Kolodziejczyk AA, Kim JK, Svensson V, Marioni JC, Teichmann SA. The technology and biology of single-cell RNA sequencing. *Mol Cell*. 2015;58 (4):610–620.
7. Shapiro E, Biezuner T, Linnarsson S. Single-cell sequencing-based technologies will revolutionize whole organism science. *Nat Rev Genet*. 2013;14(9):618–630.
8. Illumina (2016). Illumina Bio-Rad SureCell WTA 3′ Library Prep Kit for the ddSEQ System. www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/surecell-single-cell-rna-seq-data-sheet-1070-2016-014.pdf. Accessed June 27, 2017.
9. Guillaumet-Adkins A, Rodriguez-Esteban G, Mereu E, et al. Single-cell transcriptome conservation in cryopreserved cells and tissues. *Genome Biol*. 2017;18(1):45.
10. Rane AS, Rutkauskaite J, deMello A, Stavrakis S. High-throughput Multi-parametric imaging flow cytometry. *Chem*. 2017;3:588–602.

Pub. No. 1070-2017-004-A QB 5501
Bio-Rad Bulletin 6943 Ver A