Date: August 6, 2021

Cancer Immune Monitoring and Analysis Center
Dana-Farber Cancer Institute
Harvard Medical School
Contact PI: Catherine J. Wu, MD (e-mail: cwu@partners.org)

Performance lab:
Translational Immunogenomics Lab (TIGL)
Catherine J. Wu, MD (Faculty Advisor)
Kenneth J. Livak, PhD (Technology Lead Scientist)
Shuqiang Li, PhD (Technology Senior Scientist)

Single Cell ATAC-Sequencing (scATAC-seq) Analytical Performance Report, version 1.1

This report describes scATAC-seq of six cryopreserved, characterized peripheral blood mononuclear cell (PBMC) samples obtained from a patient with relapsed JAK2^{V617F+} secondary AML after allogeneic stem cell transplantation and with transient response to PD-1 blockade (Penter et al., *Blood Advances* 2021). We demonstrate that scATAC-seq of six consecutive samples from the same patient yielded consistent, high quality sequencing results and reproducible chromatin peaks. Additionally, we show that scATAC-seq revealed similar cell type kinetics before PD-1 blockade, at time of response to PD-1 blockade and at relapse as mass cytometry and single cell RNA sequencing that were performed on the same samples.

1. Purpose of the assay

Initiation of transcription occurs at open accessible chromatin of nuclear DNA. Evaluating this fundamental cellular property can provide robust information on cell differentiation and state. Unlike techniques like ChIP-seq that focus on the characteristics of individual transcription factors or histone marks, ATAC-seq provides a global assessment of epigenetic states. With scATAC-seq it is now possible to determine chromatin accessibility in thousands of single cells and thus define physiologic cell identities at unprecedented resolution. Like scRNA-seq data, scATAC-seq data for individual cells are sparse. Thus, scATAC-seq is not well suited to monitoring specific biomarkers. Rather, scATAC-seq profiles are snapshots of the overall epigenetic state of single cells and thus provide a window into cell identity and state at the regulatory level. In this report we show the utility of scATAC-seq to track immune cells in peripheral blood in response to PD-1 blockade.

Table 1. Summary of analytical performance findings for scATAC-seq

Accuracy	Accurate assay performance determined by fragment size/peak distribution of libraries (**Fig. 1**) and QC metrics of sequenced data (**Table 2**) as compared to reference ranges provided by 10x Genomics. Similar distribution of cell types when compared to two independent analysis methods (**Fig. 4**).
Precision	As indicated by the Standard Deviation column in **Table 2**, the six samples analyzed here had comparable values across all the QC metrics.
Analytical sensitivity	Targeting 500 – 10,000 isolated nuclei, depending on collection and preparation method and recovery rate. Here, we start with 1 million cells for nuclei isolation and target 7000 nuclei for library preparation.
Analytical specificity including interfering substances	Preservation and storage condition (fresh vs frozen) of samples as well as storage and dilution of Nuclei Buffer (PN-2000153, 10x Genomics) could affect transposase binding and subsequent activity (10x Genomics technical notes). Here we processed cryopreserved samples for nuclei isolation according to an optimized protocol with 10x Genomics recommended 1x diluted Nuclei Buffer.

Reportable range	Relative chromatin accessibility over the whole genome as read out by level of transposition. Sample dependent.
Reference interval (normal range)	Generated data represent the reference ranges of QC metrics of publicly available PBMC data reported in technical note CG000202 by 10x Genomics, see **Table 2**.
Standardization, harmonization, reproducibility, and ruggedness	Standardized by pooling unique indexed samples and sequencing on single lane of NovaSeq6000 Sequencer. All analysis for a given sample set performed against the same reference database.
Quality control and improvement procedures	Processing of scATAC sequencing data was done with the cell ranger atac (10x Genomics) pipeline. The QC metrics are summarized in **Table 2**.
Any other performance data	Not applicable.

2. Materials and methods

Thawing of cryopreserved PBMC samples
Slowly melt the contents of cryovials by placing the vials above the water of a water bath for about 10 minutes. Fill up each cryovial with 1 mL thawing medium (PBS + 10% grade II DNase I [cat. no. 10104159001, Sigma Aldrich] + 10% FBS) preheated to 37°C. Transfer contents of each cryovial to 15 mL Falcon tube. Slowly top up each Falcon tube with thawing medium over a period of about 3-5 minutes. Centrifuge Falcon tubes at 300×*g* for 5 minutes and discard supernatant. Resuspend cell pellet in preheated RPMI + 10% FCS + 10% DNase (cat. no. 07900, Stemcell Technologies). Cells were pelleted and resuspended in PBS + 0.04% bovine serum albumin before proceeding to nuclei isolation.

ATAC-seq procedure, library preparation, and sequencing
The single-cell ATAC-seq (scATAC-seq) assay (10x Genomics) was performed on six samples after dead cell depletion using the Dead Cell Removal Kit (cat. no. 130-090-101, Miltenyi). Nuclei were prepared by following the Nuclei Isolation for Single Cell ATAC Sequencing protocol (CG000169 Rev D, 10x Genomics). Nuclei were subsequently loaded on a Chromium Chip E (cat. no. 1000155, 10x Genomics) with a targeted recovery of 7,000 nuclei. The library preparation was performed according to manufacturer's instructions using the Chromium Single Cell ATAC Library & Gel Bead Kit v1 (cat. no. 1000110, 10x Genomics). Quality control was performed with a Bioanalyzer High Sensitivity DNA Kit (cat. no. 5067-4626, Agilent) (**Fig. 1**). Pooled libraries were sequenced on a NovaSeq SP platform (Illumina) with 50 nt paired-end reads, 8 nt for index 1 and 16 nt for index 2.

Data preprocessing
Standard data preprocessing included demultiplexing of Illumina raw base call (BCL) files using **cellranger-atac mkfastq** into fastq files. This was followed by filtering, alignment and analysis of single cell reads with **cellranger-atac count** using the hg38 reference genome. Cellranger-atac QC metrics are shown in Table 2.

Data processing
Downstream analyses were performed using the ArchR package in R.[1] High quality cells were filtered based on a transcription start site (TSS) enrichment score >4 and fragments per cell >1000 (**Fig. 2**). The analysis steps followed the ArchR manual (https://www.archrproject.com/) including removal of doublets, dimensionality reduction with iterative Latent Semantic Indexing (LSI), clustering using the Seurat's FindClusters() function and embedding using Uniform Manifold Approximation and Projection (UMAP).[2,3]

Data availability
Single cell ATAC-seq fastq files can be downloaded from GEO (accession number GSE165397).

3. Results

Fragment length distribution of all six samples showed the typical traces expected from ATAC-seq (**Fig. 1**). Sequencing quality metrices provided by cellranger-atac were all within reference values with a median of 4,014 estimated single cell chromatin profiles (range 3,516 – 6,097) and 10,986 fragments per cell (range 8,066 – 12,764) (**Table 2**).

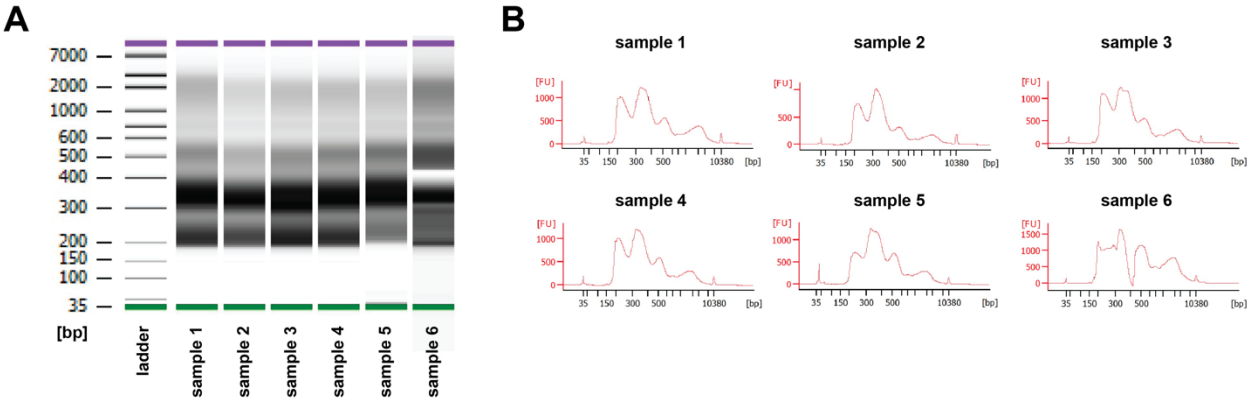


Figure 1. Quality control of scATAC-seq libraries. A Fragment lengths **B** Peak distribution

Basic filtering of high-quality cells was performed using transcription start site (TSS) enrichment and unique fragments per cell (**Fig. 2**). High quality cells were defined as having a TSS > 4 and >1000 unique fragments per cell. All six samples had similar TSS enrichment scores ranging from 12.4 to 14.2 and high-quality cells formed well-defined populations. Doublets were removed using the ArchR functionality for doublet detection.

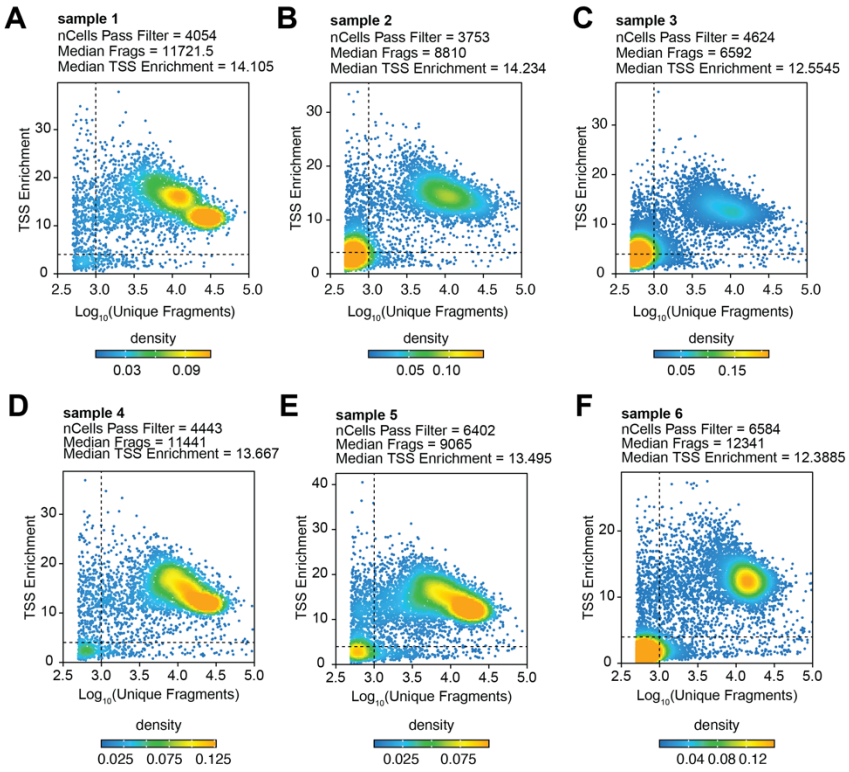


Figure 2. Filtering of high-quality cells using TSS enrichment and number of unique fragments.

We obtained a total of 28,713 single-cell chromatin accessibility profiles that clustered into the expected major cell populations (AML blasts, T cells, B cells, monocytes, megakaryocytic and erythroid cells) (**Fig. 3A**). Chromatin accessibility profiles enabled calculation of gene activity scores which showed canonic expression of genes such as *CD34*, *CD3D*, *PAX5* and *PF4* across these cell populations (**Fig. 3B**). Chromatin peaks of marker genes within the 3 largest cell populations (AML blasts, T cells and B cells) were specific for these cell populations, such as *CD3D* and *CD3G* for T cells, *PAX5* for B cells and *CD33* as well as *SIGLECL1* (encoding CD169) for AML blasts (**Fig. 3C-E**). Further, the chromatin peaks of marker genes were highly concordant within cell subpopulations across all 6 samples (**Fig. 3F-H**).

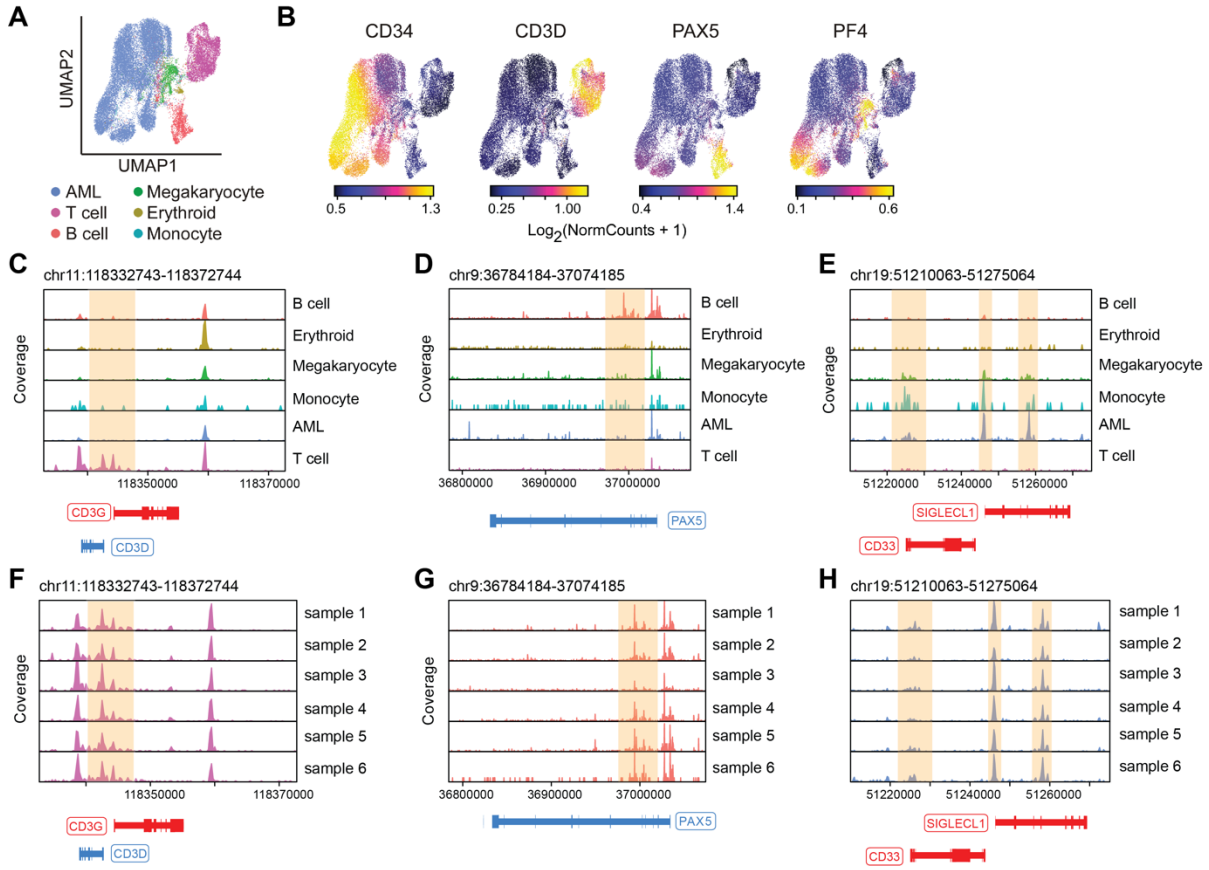


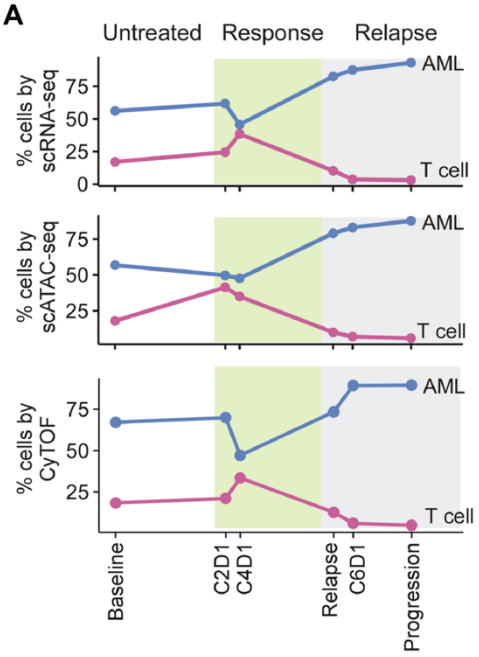
Figure 3. Identification of cell types. A UMAP representation of 28,713 single-cell chromatin profiles. **B** Imputed gene activity scores of *CD34*, *CD3D*, *PAX5* and *PF4*. **C-E** Browser tracks of *CD3D* and *CD3G* (C), *PAX5* (D), *CD33* and *SIGLECL1* (E) across different cell types. **F-H** Reproducible browser tracks of chromatin peaks in T cells (F), B cells (G) and AML blasts (H) across 6 samples .

Finally, we compared the frequency of AML blasts and T cells across all six samples with measurements made using single-cell RNA sequencing (scRNA-seq) and mass cytometry (CyTOF) and observed high concordance (**Fig. 4A**). The percentage of AML blasts and T cells quantified with CyTOF and scATAC-seq correlated well (r = 0.95). The mean ratio comparing scATAC-seq and CyTOF results for both cell populations was 1.06 (range 0.71 – 1.97), indicating consistent quantification of cell populations (**Fig. 4B**). Possible explanations for the differences are variabilities in cell surface marker staining and differences in the total number of cell populations identified in each sample.

In summary, we were able to perform scATAC-seq analysis to identify cell types with reproducible intra-assay quality across six samples and obtained biological insights that agreed with two orthogonal methods. This demonstrates the basic capability of using scATAC-seq to obtain epigenetic profiles of sufficient depth

to distinguish cell types. As scATAC-seq evolves and data density improves, additional criteria may be needed to assess assay quality.

A



B

Sample	Cells	scATAC-seq	CyTOF	Ratio
sample1	T cells	17.94	18.30	0.98
sample1	AML	56.74	67.00	0.85
sample2	T cells	41.30	21.00	1.97
sample2	AML	49.65	69.80	0.71
sample3	T cells	34.94	33.40	1.05
sample3	AML	47.54	47.00	1.01
sample4	T cells	9.99	12.60	0.79
sample4	AML	78.90	73.30	1.08
sample5	T cells	6.95	6.11	1.14
sample5	AML	82.99	89.20	0.93
sample6	T cells	5.89	4.95	1.19
sample6	AML	87.58	89.40	0.98

Figure 4. Trajectories of AML blasts and T cells measured with single cell RNA (scRNA-seq), single cell ATAC (scATAC-seq) and mass cytometry (CyTOF). A. Baseline – before PD-1 blockade; C2D1, C4D1 – nivolumab treatment at time of response; C6D1 – nivolumab treatment at relapse **B.** Percentage of T cells and AML blasts quantified through CyTOF and scATAC-seq, and ratio of percentages.

4. Appendix

Table 2. Quality metrics of samples 1-6 (cellranger-atac).

Cell parameters	Reference	Sample1	Sample2	Sample3	Sample4	Sample5	Sample6	Median	Standard Deviation
Estimated number of cells	500-10,000	3,829	3,516	3,762	4,198	6,002	6,097	4,014	1,169
Median fragments per cell	>500	12,565	9,429	8,066	12,076	9,896	12,764	10,986	1,938
Fraction of fragments overlapping any targeted region	Dependent on cell type and sequencing depth	81.40%	71.00%	67.80%	77.70%	78.00%	76.00%	76.85%	5.0%
Fraction of transposition events in peaks in cell barcodes	>25%	72.70%	52.30%	47.70%	65.60%	65.20%	60.30%	62.75%	9.3%
Sequencing parameters									
Total number of read pairs	User defined	133,633,370	164,608,805	132,313,800	143,088,842	161,930,648	192,915,372	152,509,745	23,197,548
Fraction of read pairs with a valid barcode	>75%	98.00%	98.10%	97.90%	98.00%	97.90%	98.00%	98.00%	0.1%
Q30 bases in Read 1	>65%	93.60%	93.80%	93.60%	93.80%	93.60%	93.50%	93.60%	0.1%
Q30 bases in Read 2	>65%	93.10%	93.30%	93.20%	93.30%	93.10%	93.00%	93.15%	0.1%
Q30 bases in Barcode	>65%	87.00%	88.20%	87.60%	87.30%	87.70%	87.30%	87.45%	0.4%
Q30 bases in Sample Index	>90%	90.70%	88.60%	91.80%	89.40%	87.60%	90.40%	89.90%	1.5%
Targeting									
Enrichment score of TSS	>5	7.19	7.07	6.55	7.42	7.16	6.61	7	0.3
Fraction of fragments overlapping TSS	NA	45.30%	42.30%	36.20%	41.60%	43.50%	41.30%	41.95%	3.1%
Fraction of fragments overlapping called peaks	NA	74.40%	53.60%	49.10%	67.30%	66.80%	62.00%	64.40%	9.4%
Fraction of fragments overlapping any targeted region	>55%	81.40%	71.00%	67.80%	77.70%	78.00%	76.00%	76.85%	5.0%
Fraction of total read pairs mapped confidently to genome (>30 mapq)	>80%	83.90%	72.60%	74.30%	81.80%	81.50%	85.50%	81.65%	5.3%
Fraction of total read pairs that are unmapped and in cell barcodes	NA	1.00%	0.80%	0.70%	0.90%	0.90%	0.80%	0.85%	0.1%
Fraction of total read pairs in mitochondria and in cell barcodes	<40%	4.50%	14.60%	12.10%	6.50%	6.50%	2.00%	6.50%	4.7%
Insert Sizes									
Fragments in nucleosome-free regions	>40%	52.60%	48.80%	48.40%	51.10%	50.60%	44.70%	49.70%	2.8%
Fragments flanking a single nucleosome	Dependent on sample type	39.70%	44.20%	45.20%	41.30%	41.80%	46.30%	43.00%	2.5%

5. References

1. Granja JM, Corces MR, Pierce SE, et al. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nature Genetics*. 2021;53(3):403–411.
2. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv:1802.03426 [cs, stat]*. 2018;
3. Stuart T, Butler A, Hoffman P, et al. Comprehensive Integration of Single-Cell Data. *Cell*. 2019;177(7):1888-1902.e21.