# Single Cell ATAC-seq for characterization of complex biological systems

## Abstract

During organismal development, a single genome is interpreted into a multitude of morphologically and functionally distinct cell types and states. This interpretation of a common genetic code into cell-type-specific transcriptional states is controlled by gene regulatory networks (GRNs), interactions between DNA binding proteins and RNAs to regulatory DNA elements, to produce specific gene expression profiles. A fundamental challenge in biology is to precisely link the input regulatory signals with the output gene expression that defines each cell type. While assaying transcriptional output at single cell resolution has become more common, parsing the diverse regulatory signals in single cells at scale remains a challenge. Since both protein binding and regulatory region activity are tightly coupled to nucleosome positioning, chromatin accessibility is an important tool for understanding the flow of information between the regulatory machinery and functional molecules in a cell. With the introduction of the Chromium Single Cell Assay for Transposase Accessible Chromatin (ATAC) Solution, we provide a robust and scalable approach to profile the chromatin landscape of single cells (Figure 1), pushing the frontier of genomics from describing how cells are different to understanding why.

## Highlights

- Profiled open chromatin landscapes for 9,000+ nuclei from PBMCs (Peripheral Blood Mononuclear Cells) using the Chromium Single Cell ATAC Solution
- Clustering performed with Cell Ranger ATAC identified nine distinct cell types that have been previously characterized in PBMCs
- Chromatin accessibility at DNA motifs can be used to associate transcription factor (TF) enrichment to specific cell types
- Assaying chromatin accessibility at single cell resolution enabled us to parse cellular heterogeneity and identify cell type specific regulatory patterns

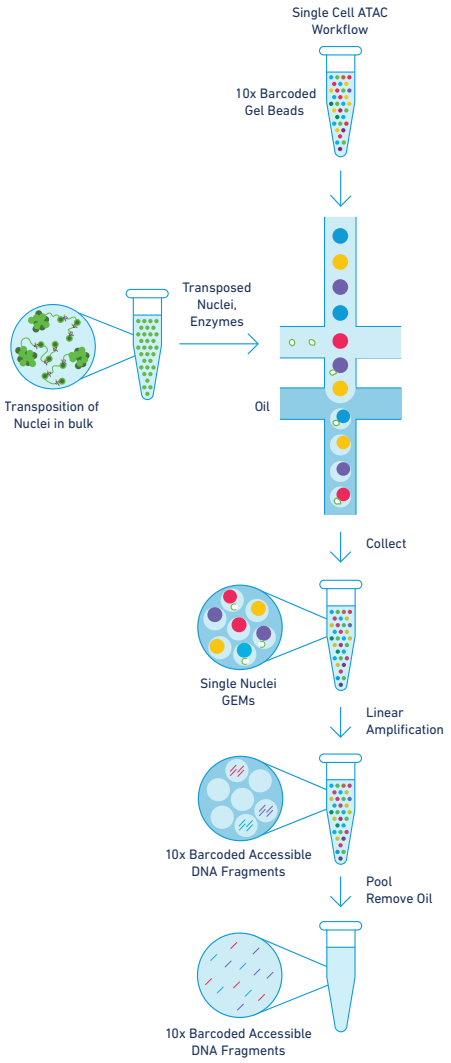


**Figure 1. The Chromium Single Cell ATAC workflow**[1]. Nuclei are transposed in bulk, followed by partitioning on a microfluidic chip into nanoliter-scale GEMs. The transposed DNA of each individual nuclei is indexed with a unique 10x barcode. Libraries are generated and sequenced, and 10x barcodes are used to associate individual reads back to individual partitions, and thereby, to each individual cell.

1 Available for use only with the indicated Chromium Controller (PN-120223;120246) or Chromium Single Cell Controller (PN-120263;120212)

## Introduction

All DNA-templated processes, such as transcription, require access to the genome. However, in a cell, only some regions of the genome are open and accessible to cellular transcriptional machinery, while the majority of the genome is packaged into chromatin and closed off [1]. The open and closed states are highly dynamic features that define the trajectory of a cell during important biological processes, such as differentiation of progenitor cells into specific cell types and cell fate decisions during the mounting of an immune response [2].

ATAC-seq has been widely used to profile chromatin accessibility across a broad range of tissue and cell types in bulk. While this has been a powerful approach for capturing the broad characteristics of a homogeneous population, tissues or samples are frequently mixtures of cell types. Averaging the signal across a population has the inherent limitation of masking heterogeneity. Subpopulations can be purified by applying filters such as FACS sorting, but this requires prior knowledge of differentiating markers and limits the discovery to known cell types.

Profiling chromatin accessibility at single cell resolution enables the discovery of epigenetic states through identification of cell-to-cell variation in regulatory elements. Single cell ATAC-seq has revealed significant variability within sorted populations using cell surface markers and led to the identification of cell states masked by bulk measurements [3].

Here we generated single cell gene expression and chromatin accessibility data from the same cryopreserved lot of unstim-ulated PBMCs, an accessible sample with a well characterized mixture of cell types. We demonstrate that epigenetic profiling of single cells is able to distinguish different cell subtypes, identify cell-type-specific TF enrichment, and reveal regulatory patterns that are complementary to gene expression.

## Results

Nuclei were isolated from unstimulated PBMCs, transposed in bulk, and ~15,000 nuclei were loaded into the Chromium System for barcode addition within GEMs according to the Chromium Single Cell ATAC Reagents User Guide (CG000168). The recovered library was sequenced to an average depth of 50,651 raw reads per cell, generating chromatin accessibility profiles for 9,542 cells with a median of 8,940 unique fragments per cell. Raw data was processed through the Cell Ranger ATAC pipeline v1.0.0.

Clustering was performed with Cell Ranger ATAC using the first 80 principal components, and we were able to distinguish nine different functional cell types (Figure 2A). Comparison with bulk ATAC-seq data from purified populations confirmed that these clusters were derived from a variety of cell types and states previously characterized in PBMCs, including naïve and memory CD4 and CD8 T cells, CD8 effector memory T cells, monocytes, B cells, natural killer (NK) cells and dendritic cells (DC) (Figure 2B). The heatmap represents the overlap between differentially accessible peaks for each cluster and reads from the respective bulk ATAC-seq datasets ([4], [5] and internally generated).

Additionally, we examined chromatin accessibility at known marker loci by aggregating reads from all cells within a cluster to form 'pseudo-bulk' accessibility profiles. As expected, the openness of chromatin at cell surface marker genes is specifically associated with the cell type of expression by its

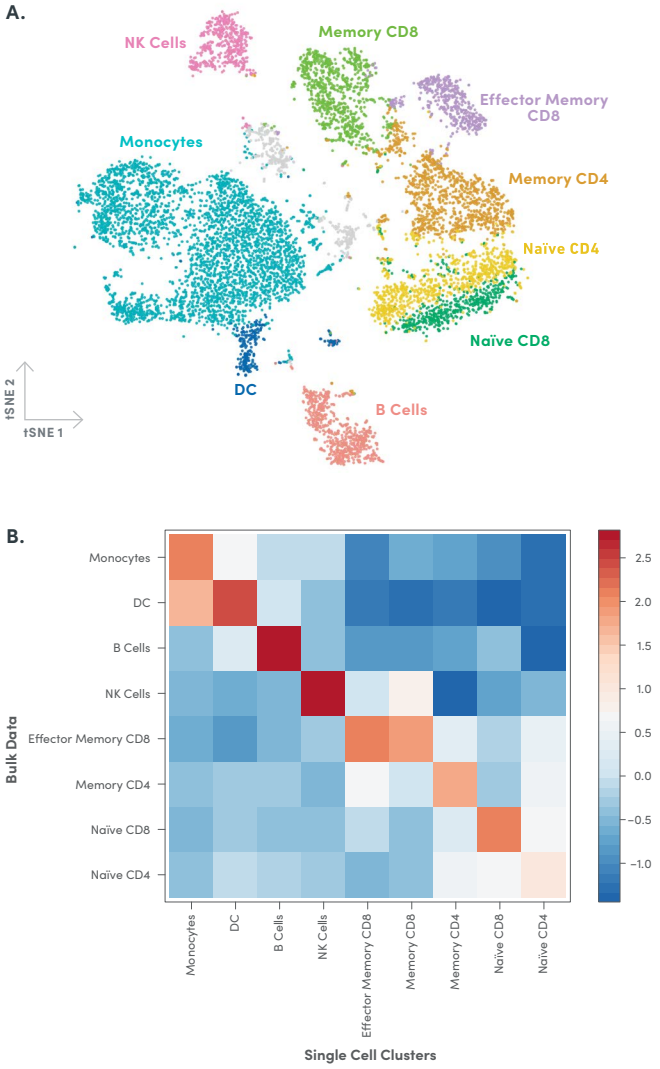


**Figure 2. Heterogeneity in chromatin accessibility delineates cell types.** A. Clustering of single nuclei accessibility profiles reveals nine cell groups in PBMCs. B. Matrix of z-scores representing the intersection between the top 200 cluster-specific peaks and the number of reads from the res pective bulk ATAC-seq datasets from sorted populations. Bulk data was obtained from the following sources: [4], [5] and internally generated.

gene expression, for example CD33 in monocytes and dendritic cells, CD8 and CD4 in their respective T cell populations (Figure 3). In contrast, all lymphoid lineage clusters shared a common pattern of accessibility around CD79A, a known cell surface marker for B cells—except for a single cell-type-defining peak (Figure 3 asterisk). Importantly, memory-associated loci, such as LEF1 (a lineage-determining TF), and effector function loci, such as Gzmb (a serine protease), could be used to distinguish cell states within cell types.

We also assayed gene expression at single cell resolution from the same lot of PBMCs (Figure 4). We used the Chromium Single Cell 3' and 5' reagents and workflow to recover ~2,000 - 4,000 cells sequenced to an average depth of 45,160 reads per cell for single cell 3' RNA-seq (scRNA-seq) and 21,184 for single cell 5' RNA-seq. 1,290 and 1,438 median genes per cell were detected, respectively. Raw data was processed using the Cell Ranger pipeline followed by clustering with Seurat v2.3.4. Nine clusters were identified in the 3' scRNA-seq dataset and eight clusters in the 5' scRNA-seq dataset. The proportion of cell types identified by scATAC-seq correlates strongly with the paired gene expression datasets (Table 1). Thus, clustering based on chromatin accessibility profiles is both consistent with orthogonal data types and biologically meaningful.

The accessibility of DNA motifs in individual cells can also be used to extrapolate the activity of TFs that bind to them. The Cell Ranger ATAC pipeline outputs a matrix tabulating the number of fragments that overlap a list of TF motifs, as defined by the JASPAR database 2018 release, for each single cell. We used this matrix to extract TF accessibility z-scores and performed systematic pairwise comparison across all cell types, selecting TF motifs with the highest variability for

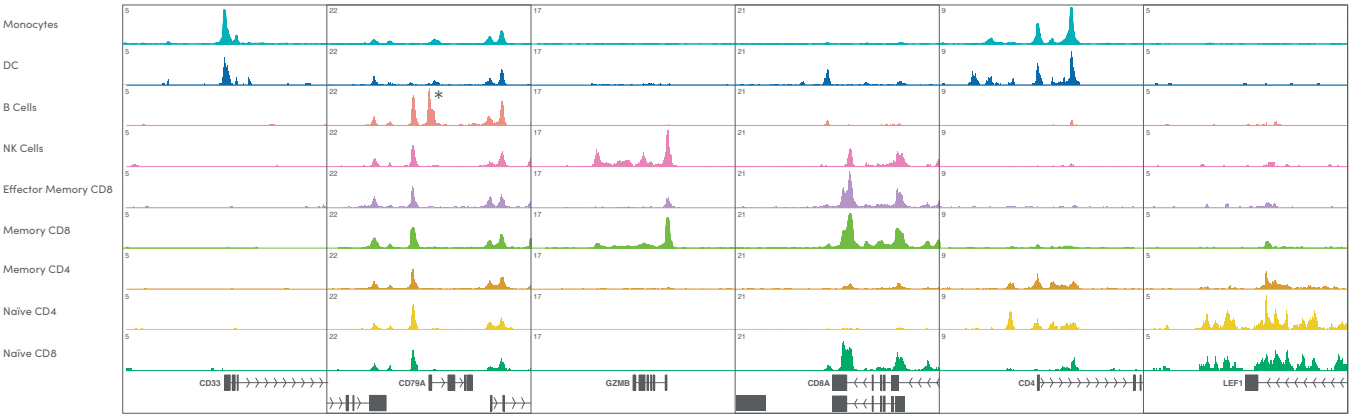


**Figure 3. Open chromatin signals around marker genes are specifically associated with the cell type of expression.** Plots show aggregate chromatin accessibility profiles for each cluster at several marker gene loci.

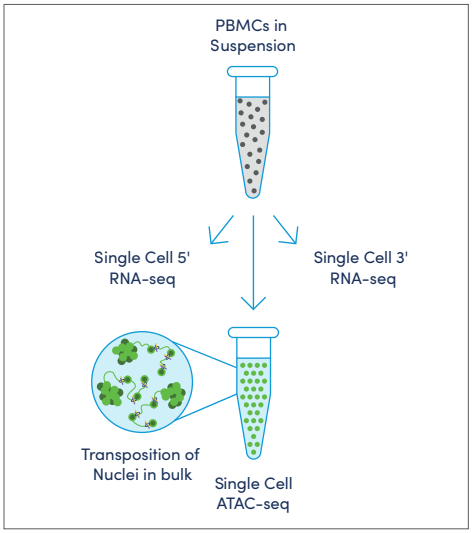


**Figure 4. Paired single cell gene expression and chromatin accessibility assay.** Schematic of the experimental set-up.

| Cell Types | scATAC-seq (%) | 3' scRNA-seq (%) | 5' scRNA-seq (%) |
|---|---|---|---|
| Monocytes | 3,576 (37.5) | 925 (38.3) | 1,324 (28.4) |
| DC | 298 (3.1) | 63 (2.6) | 73 (1.6) |
| B Cells | 681 (7.1) | 158 (6.5) | 247 (5.3) |
| NK Cells | 513 (5.4) | 138 (5.7) | 344 (7.4) |
| Effector Memory CD8 | 533 (5.6) | 143 (5.9) | 350 (7.5) |
| Memory CD8 | 910 (9.5) | 241 (10) | 541 (11.6) |
| Memory CD4 | 1,121 (11.7) | 245 (10.2) | 637 (13.7) |
| Naïve CD4 | 907 (9.5) | 479 (19.9) | 794 (17) |
| Naïve CD8 | 542 (5.7) | --- | 320 (6.9) |
| Plasma B | --- | 21 (0.9) | 36 (0.8) |
| Unknown | 461 (4.8) | --- | --- |
| Total | 9,542 (100) | 2,413 (100) | 4,666 (100) |

**Table 1.** Comparison of PBMC cell types identified in the same lot of single cell ATAC and gene expression experiments.

hierarchical clustering (Figure 5). We find that cell-type-specific accessibility in DNA binding motifs identifies known master regulators of hematopoiesis, such as BACH2 and CEBPA in monocytes; EOMES, TBX21 and IRF2 in NK and T cell subpopulations; and POU5F1 in B cells [6]. The agreement between TF enrichment and known PBMC transcription factor activity indicates that chromatin accessibility, as defined by the Chromium Single Cell ATAC Solution, can accurately identify cluster-specific functions.

Transcription factors tend to bind in open chromatin regions. Thus, the accessibility of a given TF motif is expected to correlate with the expression of the associated TF. We found this to be true in many cases. For example SPI1 is expressed nearly exclusively in monocytes, the same cell type in which we see enrichment of open chromatin over the SPI1 binding motif (Figure 6, top).

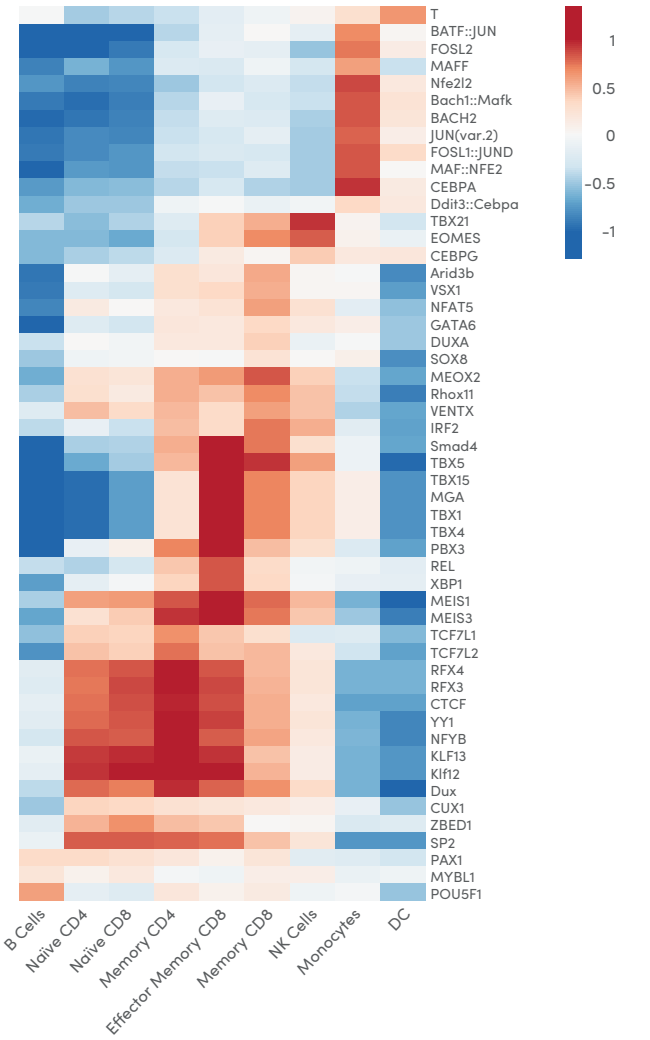


**Figure 5. Transcription factor motif accessibility identifies cell-type-specific regulators.** Displayed are TF motif accessibility z-scores for the top enriched TF motifs identified through pairwise comparison across all cell types.

We also observed patterns where chromatin accessibility and gene expression reveal a more complex model of TF activity. For example, ELF1 appears to be expressed in all cell types but accessibility of ELF1 motifs is restricted to T cells (Figure 6, middle). In contrast, the expression of TBX21 is restricted to NK cells and a subset of CD8 memory T cells while accessibility to its binding site is much more permissive (Figure 6, bottom). Thus, in the cases of ELF1 and TBX21, neither gene expression nor chromatin accessibility alone could paint a complete picture. A more thorough investigation integrating the two data types is needed to parse the roles of these key regulators.

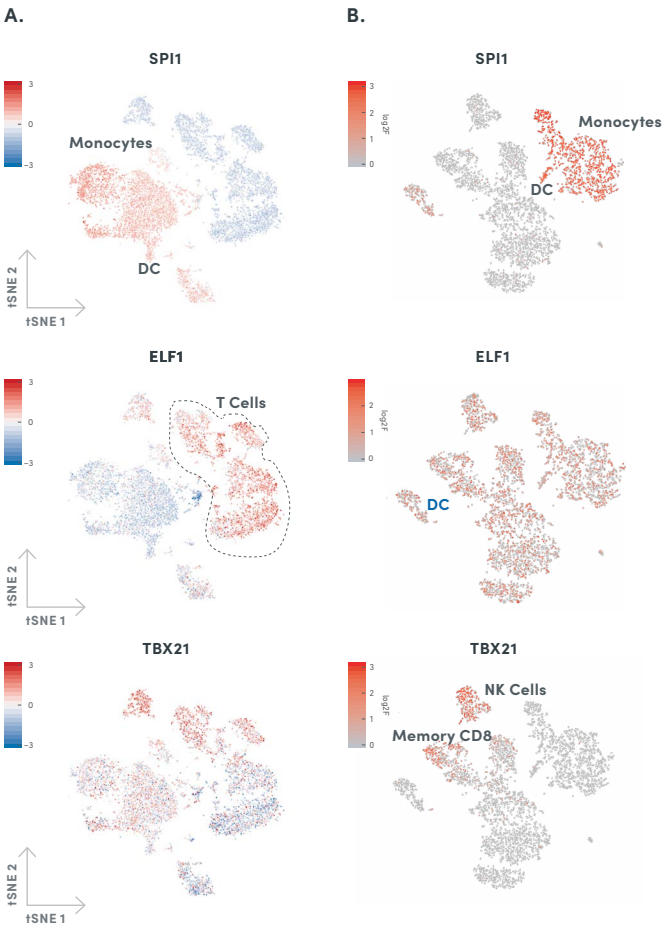


**Figure 6. Complementary patterns of transcription factor expression and binding site accessibility.** A. tSNE projection of scATAC-seq of 9,542 PBMC cells, with each cell colored by the TF motif accessibility z-score for SPI1, ELF1 and TBX21. Enrichment of open chromatin for that motif (red) or lack thereof (blue) compared to the average accessibility (white) across all cells. B. tSNE projection of 4,666 cells from a matched 5′ scRNA-seq PBMC sample, with each cell colored on the basis of log2 fold expression change of SPI1, ELF1 and TBX21 genes.

10x GENOMICS

## Conclusion

The Chromium Single Cell Assay for Transposase Accessible Chromatin (ATAC) Solution is a robust and scalable approach to profile the chromatin landscape of single cells. In this application note, we demonstrated the solution on a population of unstimulated PBMCs and generated chromatin accessibility profiles for 9,542 cells. Clustering based on single cell chromatin accessibility profiles identified nine distinct functional cell types, which were consistent with cell types identified with bulk ATAC-seq as well as 3' and 5' single cell gene expression analysis. Additionally, we demonstrate that chromatin accessibility can be used to associate TF enrichment to specific cell types. In conclusion, assaying chromatin accessibility at single cell resolution has the potential to parse cellular heterogeneity and identify cell-type-specific regulatory patterns that drive cell identity and function.

## Resources

| | |
|---|---|
| Datasets | go.10xgenomics.com/scATAC/datasets |
| Seminars | go.10xgenomics.com/scATAC/seminars |
| Application Notes | go.10xgenomics.com/scATAC/app-notes |
| Technical Support | go.10xgenomics.com/scATAC/support |
| Publications | go.10xgenomics.com/scATAC/pubs |

## Support

support@10xgenomics.com
10x Genomics
6230 Stoneridge Mall Road
Pleasanton, CA 94588-3260

## Legal Notices

For 10x Genomics legal notices visit:
10xgenomics.com/legal-notices

## References

1. M. Tsompana and M. J. Buck, Chromatin accessibility: a window into the genome, *Epigenetics Chromatin*, vol. 7, no. 1, p. 33 (2014).
2. G. Kelsey, O. Stegle, and W. Reik, Single-cell epigenomics: Recording the past and predicting the future, *Science*, vol. 358. (2017).
3. J. D. Buenrostro et al., Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation, *Cell*, vol. 173, no. 6, pp. 1535-1548.e16, (2018).
4. D. Calderon et al., Landscape of stimulation-responsive chromatin across diverse human immune cells, *bioRxiv* 409722, (2018).
5. M. R. Corces et al., Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution, *Nat. Genet.*, vol. 48, no. 10, pp. 1193–1203, (2016).
6. A. Mezger et al., High-throughput chromatin accessibility profiling at single-cell resolution, *Nat. Commun.*, vol. 9, no. 1, (2018).

## Resources from 10x Genomics

We are dedicated to helping you get the most out of your 10x Genomics system by offering multiple helpful resources:

### 10x University
Immerse yourself in 10x University, a comprehensive step-by-step learning and training environment containing video tutorials and trainings.
10xgenomics.com/10x-university

### Support
Visit the support site for documentation, software, and datasets that will help you get the most out of your 10x Genomics products.
support.10xgenomics.com

### Solutions and products
Along with our suite of complete solutions, we offer an ever-growing catalogue of services to help you find the answers to your research questions.
10xgenomics.com/solutions

### 10x Genomics library
Easy access to our complete library of product literature, customer publications, application notes, protocols, and many other useful resources.
10xgenomics.com/resources/library

### 10x Genomics compatible products
Access our list of key partner products that have been certified compatible to work with our various solutions.
10xgenomics.com/compatible-products

### 10x Blog
Keep up to date with the 10x Genomics Blog, where you'll find everything from tips and tricks to the latest 10x news.
10xgenomics.com/blog

## Contact us

10x Genomics
6230 Stoneridge Mall Road
Pleasanton, CA 94588-3260

+1 925 401 7300
+1 800 709 1208

info@10xgenomics.com
10xgenomics.com

For more locations in US, EU and Asia visit:
10xgenomics.com/company/#locations


LIT000036 Rev B Single Cell ATAC-Seq for Characterization of Complex Biological Systems