Simultaneous Profiling of Gene Expression and Chromatin Accessibility in Single Cells

HHS Public Access

Author manuscript

Adv Biosyst. Author manuscript; available in PMC 2019 December 18.

Published in final edited form as:

Adv Biosyst. 2019 November ; 3(11): . doi:10.1002/adbi.201900065.

Simultaneous profiling of gene expression and chromatin accessibility in single cells

Miguel Reyes[1,2], **Kianna Billman**[1], **Nir Hacohen**[1,3], **Paul C. Blainey**[1,2,*]

[1]Broad Institute of MIT and Harvard, Cambridge, MA, USA.

[2]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA.

[3]Center for Cancer Research, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA

SUMMARY

Profiling multiple omic layers in a single cell enables the discovery and analysis of biological phenomena that are not apparent from analysis of mono-omic data. While methods for multi-omic profiling have been reported, their adoption has been limited due to high cost and complex workflows. Here, we present a simple method for joint profiling of gene expression and chromatin accessibility in tens to hundreds of single cells. We assess the quality of resulting single cell ATAC- and RNA-seq data across three cell types, examine the link between accessibility and expression at the *CD3G* and *FTH1* loci in human primary T cells and monocytes, and compare the accuracy of clustering solutions for mono-omic and combined data. The new method allows biological laboratories to perform simultaneous profiling of gene expression and chromatin accessibility using standard reagents and instrumentation. This technique, in conjunction with other advances in multi-omic profiling, will enable highly-resolved cell state classification and more specific mechanistic hypothesis generation than is possible with mono-omic analysis.

MAIN TEXT

Single cell technologies have developed rapidly in recent years, with expression profiling (scRNA-seq) emerging as a popular approach in many biological research laboratories (1–5). scRNA-seq has been applied to reveal many new cell states, but its utility in unravelling biological mechanisms that are regulated at different molecular layers and at different time-scales is limited. scRNA-seq only profiles a single layer of molecular information, messenger RNAs. While transcription can produce multiple copies of each messenger RNA species, their expression is stochastic, and the rapid turnover of messenger RNA that drives cells' responsivity to natural cues also makes scRNA-seq susceptible to artifacts of sample preparation procedures (6–8). Multi-omic profiling techniques hold the potential to provide

*Correspondence should be addressed to P.C.B. (pblainey@broadinstitute.org).

CONFLICTS OF INTEREST

P.C.B. is an extramural faculty member of MIT's Koch Institute for Integrative Cancer Research and a consultant to and equity holder in two companies in the microfluidics and single-cell industries, 10X Genomics and General Automation Lab Technologies. The Broad Institute and MIT may seek to commercialize aspects of this work.

more biological insight, not just in the state of cells and variability within populations, but also in the molecular mechanisms by which maintain their identity and response to cues (9, 10). Multi-omic methods will allow for more robust and nuanced definitions of cell states and integrated analysis of cellular trajectories across multiple timescales. Importantly, with this increased resolution and the significant number of cells that can be processed, these methods will enable the identification of elements that co-vary between molecular layers, nominating candidate *trans* mechanistic links between these elements.

Multiple layers of molecular information have been profiled at the single cell level, either separately on different cells from the same population, or in tandem from each individual cell -- providing opportunities to learn from co-variation across the seemingly random variation within single cells (11). Among informational layers accessible in cells, chromatin accessibility and gene expression (RNA-seq) are a popular pair. The combined measurement of accessibility and expression allows the analysis of cell states across different time scales, as gene expression responses occur rapidly, within minutes to hours, while epigenomic changes reflected in accessibility occur on longer timescales, from days to weeks (12). In addition, accessibility reveals information about cell state not accessible from expression measurements alone. For example, by identifying which distal regulatory elements may be engaged in driving the expression of a gene.

A limited number of methods for profiling expression and chromatin accessibility in the same cell have been developed recently, but these techniques have limitations that present significant hurdles to their adoption in many biological laboratories (13–15). scNMT-seq (13), which uses GpC methyltransferase activity and subsequent bisulfite conversion to measure chromatin accessibility, provides single base resolution but incurs a significant cost due to the specialized enzymatic treatment step and the extra library preparation steps involved in bisulfite-sequencing. CAT-seq (14), which calls for mildly lysing cells to separate nuclei from cytoplasms and carrying out single cell ATAC-seq and single cell RNA-seq separately in two plates, produces good quality libraries yet is reagent-intensive due to the addition of Tn5 enzyme, the dominant cost in ATAC-seq, to a large number of single-cell library construction reactions. Finally, a combinatorial indexing-based method (15) allows multi-omic profiling of thousands of single cells in parallel, but requires custom synthesis of an array of indexed enzyme reagents. Due to such technical limitations, combined accessibility and expression profiling has not been widely applied in single cells.

Here, we developed a simple method (Figure 1) for simultaneous chromatin accessibility and gene expression profiling that requires neither bisulfite conversion nor indexed transposomes. In addition, the tagmentation step is performed first in a pooled step, significantly reducing the cost, time, and effort required compared to other techniques. Our method starts with the fixation of cells with dithio-bis(succinimidyl propionate) (DSP), a commercially available tissue stabilizer that has been previously shown to preserve messenger RNA, is compatible with downstream single cell RNA-seq (16), and prevents subsequent cell handling procedures from creating artifactual cellular responses. After fixation, 50,000 – 100,000 cells are permeabilized and tagmented with a commercially available transposase enzyme. Their nuclei are then stained with DAPI (4’,6-diamidino-2-phenylindole) and single cells are sorted into a 96-well plate containing lysis buffer. After

lysis, mRNA are captured using oligo-dT beads and transferred to a second 96-well plate, leaving tagmented genomic DNA (gDNA) behind in the original plate. The two plates are then processed according to standard protocols: Smart-seq2 for the bead-bound mRNA (17) and indexed PCR for the previously tagmented genomic DNA fragments (18). The resulting ATAC-seq and RNA-seq libraries are paired after sequencing and matched by cell-of-origin based on the plate well indices.

We performed the full workflow on three different human cell types: the K562 cell line, CD14+ primary monocytes, and CD3+ primary T cells isolated from human peripheral blood (n = 36, 38, and 42 cells, respectively). The distribution of fragment sizes in the ATAC-seq libraries show the characteristic nucleosome phasing both in single cell and aggregate (combined single-cell) libraries (Figure 2a, Supplementary Figure 1). In addition, the libraries are enriched for reads that map proximal to transcription start sites (TSS) (Figure 2b), both in single cell and aggregate libraries. These patterns are consistent across the cell types we profiled (Figure 2c–d), albeit with slight differences in the overall TSS enrichment and visibility of the nucleosome phasing. We obtained an average of 10^3–10^4 unique fragments per cell, with 40% −65% of reads mapping to peaks (Supplementary Figure 2b), comparable with data obtained from other reported single cell ATAC-seq datasets (19, 20). In addition, we detected 1500 to 3000 genes per cell in our RNA-seq libraries from the same cells (Figure 2f), similar to levels detected in other cellular and nuclear scRNA-seq datasets of the same cell types (21, 22). Interestingly, the RNA-seq reads contained a high fraction of intronic reads (Supplementary Figure 2a), signifying that many of the RNA molecules we capture are nascent polyadenylated RNA from the nucleus. Aggregate single cell ATAC-seq and RNA-seq data from the combined workflow reported here also correlate well with data generated from the corresponding mono-omic bulk methods (5,000 cells) applied to the same cell populations, supporting the accuracy of the combined method in reliably generating both data types (Supplementary Figure 2d–e). To determine whether our fixation step was successful and to check whether cross-talk occurred during the pooled tagmentation step, we performed a human-mouse species mixing experiment. Our results (Figure 2g) show that RNA-seq reads from each well on the plate map exclusively to one species or the other and that the only well with reads mapping to both species is the bulk control which contains a 1:1 mix of cells and was expected to show dual-species mapping. Overall, the libraries from our initial experiments show good quality that are on par with the results of contemporary mono-omic single-cell methods for gene expression and chromatin accessibility profiling.

To confirm the link between the gene expression and chromatin accessibility in our dataset, we examined a 10 kb region around two genes, *CD3G* and *FTH1*. These genes are highly expressed in T cells and monocytes, respectively, in our data (Figure 3a,c) and in previous scRNA-seq studies (23, 24). Due to the sparsity of ATAC-seq data, we aggregated reads from all single T cells and monocytes for visualization (38 and 42 cells, respectively). Our data show an enrichment in reads around the promoters of *CD3D* and *CD3G* in T cells but not in monocytes, and is consistent with bulk ATAC-seq data (5,000 cells, Figure 3b). Similarly, the loci around the *FTH1* gene show broad enriched peaks in monocytes, but not T cells (Figure 3d), in agreement with previous data. To demonstrate the advantage of multi-omic profiling over mono-omic methods, we compared analyses of the cells using only their

RNA-seq, ATAC-seq, or combined RNA-seq and ATAC-seq profiles (Figure 3e). The agreement between cluster assignments and cell-type labels is higher when the two data-types are combined, as shown by significant increases of the mean adjusted Rand index of the clustering solutions (Figure 3f, Supplementary Figure 3, Methods). These analyses show that our paired approach can capture cell-type specific differences in both the transcriptome and chromatin accessibility profiles and demonstrates the potential of multi-omic methods to improve the resolution of cell state classifications by expanding the basis for classification.

In this work, we demonstrated a simple protocol for simultaneous profiling of gene expression and chromatin accessibility in single cells. We show that the quality of our ATAC libraries is comparable to previous protocols, while the RNA-seq libraries have similar gene counts but are enriched for intronic reads. We note that despite the large number of genes detected in our analysis, the clustering performance using the RNA-seq data is limited in comparison with previous mono-omic RNA-seq studies (23–25). In addition, the data quality resulting from the current protocol limits unbiased discovery of marker genes through differential peak and expression analysis, but known marker genes can be mapped onto the data to confirm their identity (Supplementary Figure 4). The intergenic and intronic mapping rates in our data are high compared to other single-cell RNA-seq protocols (22, 26, 27) (Supplementary Figure 2), which could in principle be explained by genomic DNA contamination in the protocol or relative enrichment of nuclear or nascent transcripts. Although our protocol is designed to exclude genomic DNA from the RNA-seq libraries, we could more stringently exclude genomic DNA by optimizing bead washes during the RNA/DNA separation step or by treating the isolated mRNA with DNAse. These observations highlight the need to holistically analyze diverse criteria when assessing the quality of RNA-seq libraries, as gene count, the most commonly used summary metric, does not completely reflect the quality information content of transcriptome data. Nevertheless, we were able to perform unbiased clustering with the RNA-seq data and highlight the improvement in cell state resolution when multi-omic data are used. We also show that we are able to resolve cell-type specific differences in both the RNA-seq and chromatin accessibility molecular layers.

With further optimization of the protocol and an increase in the number of profiled cells, we envision this method could be used in the multi-omic definition of cell states and the unbiased discovery of regulatory links in the genome across several cell types. This simple protocol can be used produce multi-omic libraries and analyze correlations between the expression of genes and accessibility of neighboring genomic loci including regulatory sequences like promoters and enhancers. Such an approach can be used to validate discoveries based on prior RNA-seq (mono-omic) findings and explore underlying biological mechanisms. The multi-omic protocol presented here, together with other emerging tools for multi-omic profiling techniques in single cells, will enable the discovery of novel molecular regulatory mechanisms across multiple cell types.

EXPERIMENTAL SECTION

Cell Culture and PBMC Isolation

K562 cells were cultured in Iscove's Modified Dulbecco's Medium (Gibco) with 10% FBS (Gibco). L1210 cells were cultured in Dulbecco's Modified Eagle's Medium (Gibco) with 10% FBS (Gibco). Cells were kept at the recommended conditions (37°C, 5% CO_2) and maintained at 10^6 cells/mL.

Human blood samples were obtained from Research Blood Components (MA, USA). Cells were isolated from whole blood using density gradient centrifugation. Whole blood was diluted 1:1 with 1X PBS, layered on top of Ficoll-Paque Plus (GE Healthcare), and centrifuged at 1200g for 20 min. The PBMC layer was retrieved, resuspended in 10 mL RPMI-1640 (Gibco), and centrifuged again at 300 g for 10 min. The cells were counted using a manual hemocytometer, resuspended in FBS (Gibco) with 10% DMSO (Sigma), and aliquoted in 1 mL cryopreservation tubes at a concentration of 5 M cells/mL. The tubes were kept at −80 °C overnight, then transferred to liquid nitrogen for long-term storage. Prior to processing, cells were thawed at 37 °C for 3 min, resuspended in 10 mL RPMI-1640 supplemented with 10% FBS (Gibco), and centrifuged at 300 g for 5 min. The cells were then resuspended in the same buffer prior to the experiment at a concentration of 1 M. The following panel was used to sort T cells ($CD3^+$) and monocytes ($CD14^+$) on an SH800 Cell Sorter (Sony): DAPI, CD45 BV605, CD3 AF700, CD4 FITC, CD8 PE, CD14 APC, CD19 PE-Cy7, CD56 BV650 (all IgG1K, BioLegend).

Fixation, Pooled Tagmentation, and Sorting

Cells were washed twice in cold 1X PBS prior to resuspension in 200 μL 1X DSP (Sigma-Aldrich) (16). The cells were left at room temperature for 30 min, and 4.1 μL Tris-HCl pH 7.5 was added to react the remaining DSP. The cells were then transferred to a 1.5 mL tube and topped up to 1 mL with 1X PBS. The 1.5 mL tubes used were pre-coated with 1% BSA overnight prior to use. After fixation, the cells were tagmented using the Fast-ATAC protocol (28). To stop the tagmentation reaction, 50 μL 10 mM Tris pH 8, 20 mM EDTA was added to the mix and incubated on ice for 10 min. The permeabilized cells were then stained with 1 μg/mL DAPI and single cells were sorted into 5 μL lysis buffer: 20 mM Tris pH 8, 50 mM DTT,0.2% SDS, 50 mM NaCl, 2 μg/mL proteinase K (Qiagen), 1.2 U/μL recombinant RNAse inhibitor (Takara Bio).

mRNA-gDNA Separation and Library Preparation

RNA-seq was performed using Smart-Seq2 (17) with minor modifications. For mRNA capture, a biotinylated oligo (5'-Biotin-AAGCAGTGGTATCAACGCAGAGTAC-30T-VN) (Integrated DNA Technologies) was attached to streptavidin magnetic beads (New England Biolabs) following the manufacturer's protocol. The beads were then used to capture mRNA from the lysates and washed with 4 μL 20 mM Tris pH 8, 10 mM DTT, 3.3 mM $MgCl_2$, 33 mM NaCl, 0.5% Tween-20 thrice. The supernatants were transferred to another plate for amplification of the tagmented gDNA. The beads were then resuspended in the reverse transcription mix, following the same steps as the published protocol. After amplification and clean-up, libraries were quantified using a Qubit fluorometer (Invitrogen) and their size

distributions were determined using the Agilent Bioanalyzer 2100. After normalizing the amplicon concentrations to 0.1–0.2 ng/mL, sequencing libraries were constructed using the Nextera XT DNA Library Prep Kit (Illumina), following the manufacturer's protocol. The plate with tagmented gDNA was amplified with by adding 25 μL 2X KAPA HiFi Mastermix (KAPA Biosystems) with 0.05X ROX (Invitrogen), 0.1X EVA-Green (Biotium). The amplification reaction was monitored via qPCR and stopped once the appropriate number of cycles was reached (18). All ATAC-seq and RNA-seq libraries were sequenced with 38×37 paired-end reads using a MiniSeq or NextSeq (Illumina).

Data Analysis

ATAC-seq and RNA-seq libraries were sequenced to depths of 100,000–200,000 and 1–5 million reads per sample, respectively. RNA-seq reads were aligned to the UCSC hg19 transcriptome using STAR (29) and used as input to generate QC statistics with RNA-SeQC (30). RSEM (31) was used to generate an expression matrix for all samples. ATAC-seq reads were aligned to the UCSC hg19 genome, with removal of duplicate and mitochondrial reads. Peaks were called by taking an aggregate of the reads from all cells of the same cell type and using the resulting BAM file as an input to MACS2 (32). Peak tracks were visualized using IGV (33). To generate a peak matrix for the ATAC-seq data, reads from each cell were mapped to the 11,926 MACS2-derived peaks.

Single cell analyses for ATAC-seq, RNA-seq, and combined ATAC-RNA-seq were all performed using the *scanpy* package (34). The matrices were first log-transformed and total feature counts per cell were regressed out to remove possible confounding effects of cell quality. After which, the data were filtered to select the top 4,000 variable features. Principal components analysis was then performed on the resulting matrix for visualization, and the top 20 components were used as an input to Louvain clustering.

To assess the quality of clustering solutions, adjusted rand indices was calculated between the solution from Louvain clustering and the true cell type labels. Distributions of rand indices for each data type were calculated by subsampling the total number of cells (90% without replacement, 20 times).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

This work was supported by NIH NIAID U24 AI118668 (N.H. and P.C.B.). P.C.B. is supported by a Career Award at the Scientific Interface from the Burroughs Wellcome Fund.

REFERENCES

1. Kolodziejczyk AA, Kim JK, Svensson V, Marioni JC, Teichmann SA, The technology and biology of single-cell RNA sequencing. Mol. Cell 58, 610–620 (2015). [PubMed: 26000846]
2. Haque A, Engel J, Teichmann SA, Lönnberg T, A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. Genome Med. 9, 75 (2017). [PubMed: 28821273]

3. Bacher R, Kendziorski C, Design and computational analysis of single-cell RNA-sequencing experiments. Genome Biol. 17, 63 (2016). [PubMed: 27052890]
4. Hwang B, Lee JH, Bang D, Single-cell RNA sequencing technologies and bioinformatics pipelines. Exp. Mol. Med 50, 96 (2018). [PubMed: 30089861]
5. Wu AR, Wang J, Streets AM, Huang Y, Single-Cell Transcriptional Analysis. Annu. Rev. Anal. Chem. 10, 439–462 (2017).
6. Macneil LT, Walhout AJM, Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. Genome Res. 21, 645–657 (2011). [PubMed: 21324878]
7. Balázsi G, van Oudenaarden A, Collins JJ, Cellular decision making and biological noise: from microbes to mammals. Cell. 144, 910–925 (2011). [PubMed: 21414483]
8. Raj A, van Oudenaarden A, Nature, nurture, or chance: stochastic gene expression and its consequences. Cell. 135, 216–226 (2008). [PubMed: 18957198]
9. Packer J, Trapnell C, Single-Cell Multi-omics: An Engine for New Quantitative Models of Gene Regulation. Trends Genet. (2018), doi:10.1016/j.tig.2018.06.001.
10. Chappell L, Russell AJC, Voet T, Single-Cell (Multi)omics Technologies. Annu. Rev. Genomics Hum. Genet 19, 15–41 (2018). [PubMed: 29727584]
11. Macaulay IC, Ponting CP, Voet T, Single-Cell Multiomics: Multiple Measurements from Single Cells. Trends Genet. 33, 155–168 (2017). [PubMed: 28089370]
12. Kelsey G, Stegle O, Reik W, Single-cell epigenomics: Recording the past and predicting the future. Science. 358, 69–75 (2017). [PubMed: 28983045]
13. Clark SJ et al., scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. bioRxiv (2018), p. 138685.
14. Liu L et al., Deconvolution of single-cell multi-omics layers reveals regulatory heterogeneity. Nat. Commun 10, 470 (2019). [PubMed: 30692544]
15. Cao J et al., Joint profiling of chromatin accessibility and gene expression in thousands of single cells. Science, eaau0730 (2018).
16. Attar M et al., A practical solution for preserving single cells for RNA sequencing. Sci. Rep 8, 2151 (2018). [PubMed: 29391536]
17. Picelli S et al., Full-length RNA-seq from single cells using Smart-seq2. Nat. Protoc 9, 171–181 (2014). [PubMed: 24385147]
18. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ, ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr. Protoc. Mol. Biol 109, 21.29.1–9 (2015).
19. Chen X, Miragaia RJ, Natarajan KN, Teichmann SA, A rapid and robust method for single cell chromatin accessibility profiling. Nat. Commun 9, 5345 (2018). [PubMed: 30559361]
20. Buenrostro JD et al., Single-cell chromatin accessibility reveals principles of regulatory variation. Nature. 523, 486–490 (2015). [PubMed: 26083756]
21. Ziegenhain C et al., Comparative Analysis of Single-Cell RNA Sequencing Methods. Mol. Cell 65, 631–643.e4 (2017). [PubMed: 28212749]
22. Habib N et al., Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons. Science. 353, 925–928 (2016). [PubMed: 27471252]
23. Zheng GXY et al., Massively parallel digital transcriptional profiling of single cells. Nat. Commun 8, 14049 (2017). [PubMed: 28091601]
24. Gierahn TM et al., Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. Nat. Methods 14, 395–398 (2017). [PubMed: 28192419]
25. Macosko EZ et al., Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell. 161, 1202–1214 (2015). [PubMed: 26000488]
26. Habib N et al., Massively parallel single-nucleus RNA-seq with DroNc-seq. Nat. Methods 14, 955–958 (2017). [PubMed: 28846088]
27. Picelli S et al., Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat. Methods 10, 1096–1098 (2013). [PubMed: 24056875]
28. Corces MR et al., Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. Nat. Genet 48, 1193–1203 (2016). [PubMed: 27526324]

29. Dobin A et al., STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 29, 15–21 (2013). [PubMed: 23104886]

30. DeLuca DS et al., RNA-SeQC: RNA-seq metrics for quality control and process optimization. Bioinformatics. 28, 1530–1532 (2012). [PubMed: 22539670]

31. Li B, Dewey CN, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 12, 323 (2011). [PubMed: 21816040]

32. Zhang Y et al., Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9, R137 (2008). [PubMed: 18798982]

33. Thorvaldsdóttir H, Robinson JT, Mesirov JP, Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform 14, 178–192 (2013). [PubMed: 22517427]

34. Wolf FA, Angerer P, Theis FJ, SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 19, 15 (2018). [PubMed: 29409532]

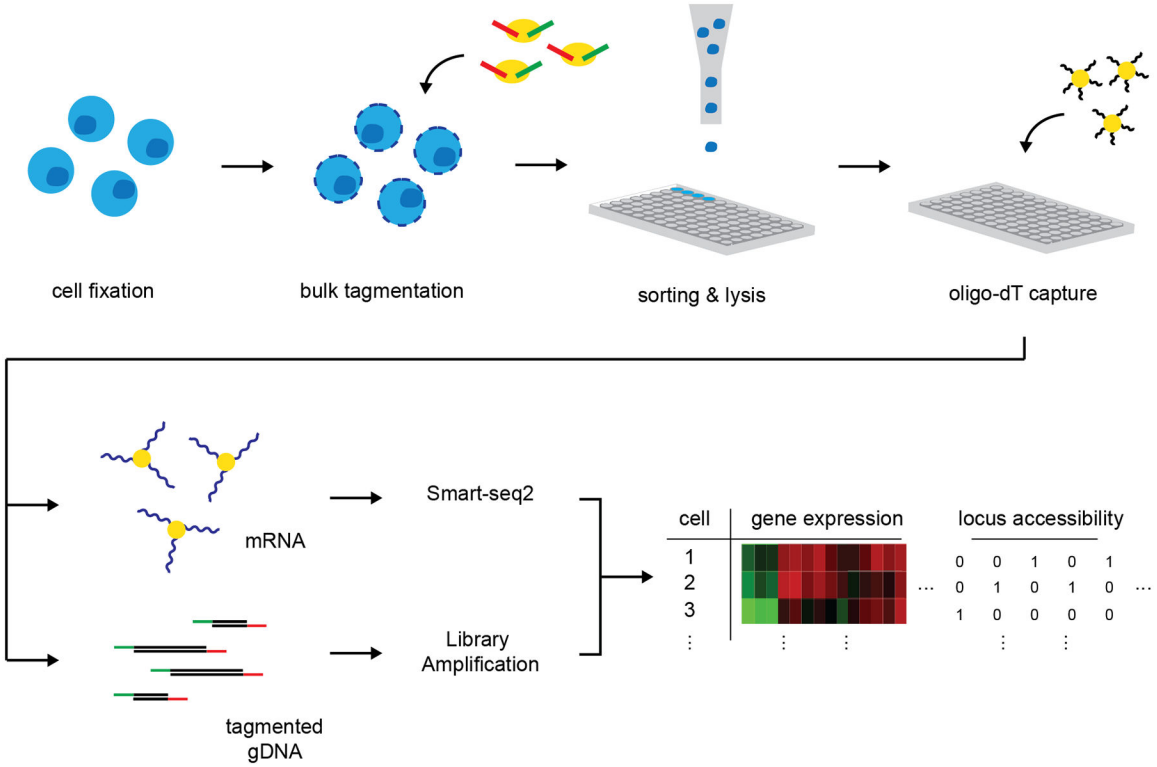

Figure 1. A method for simultaneous profiling of expression and chromatin accessibility in single cells.

Cells are fixed, permeabilized, and tagmented in bulk before sorting single cells into a 96-well plate with lysis buffer. mRNA and DNA molecules are separated via oligo-dT bead capture. The libraries are prepared independently and linked via their well-position in the original plate.

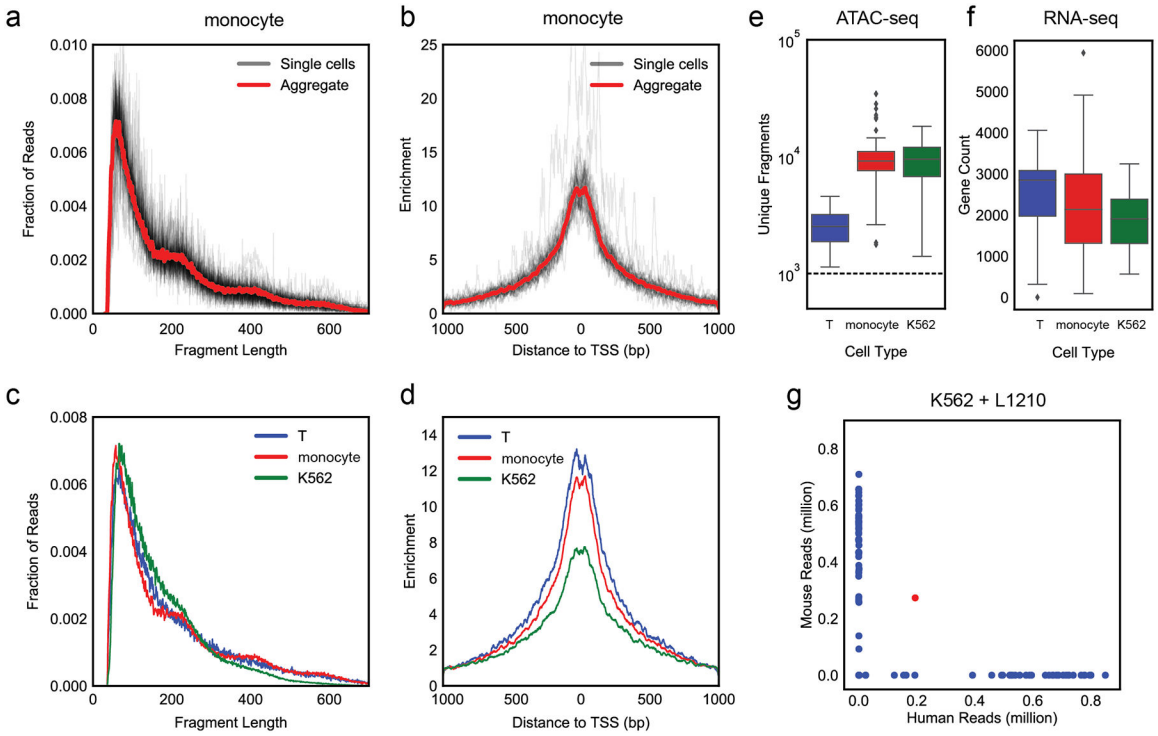

Figure 2. ATAC-seq and RNA-seq library quality assessment.

(a) Fragment size distribution and (b) TSS enrichment for single cell and aggregate ATAC-seq libraries of monocytes. (c) Fragment size distribution and (d) TSS enrichment for aggregate ATAC-seq libraries of three different cell types (K562, T cells, monocytes). (e) Number of unique fragments detected in single cell ATAC-seq libraries across three cell types. (f) Number of genes (count > 1) detected in RNA-seq libraries across three cell types. (g) Barnyard plot showing mapping of single cell libraries in a species-mixing experiment. Blue points represent single cell wells, red point indicates bulk well with 500 cells.

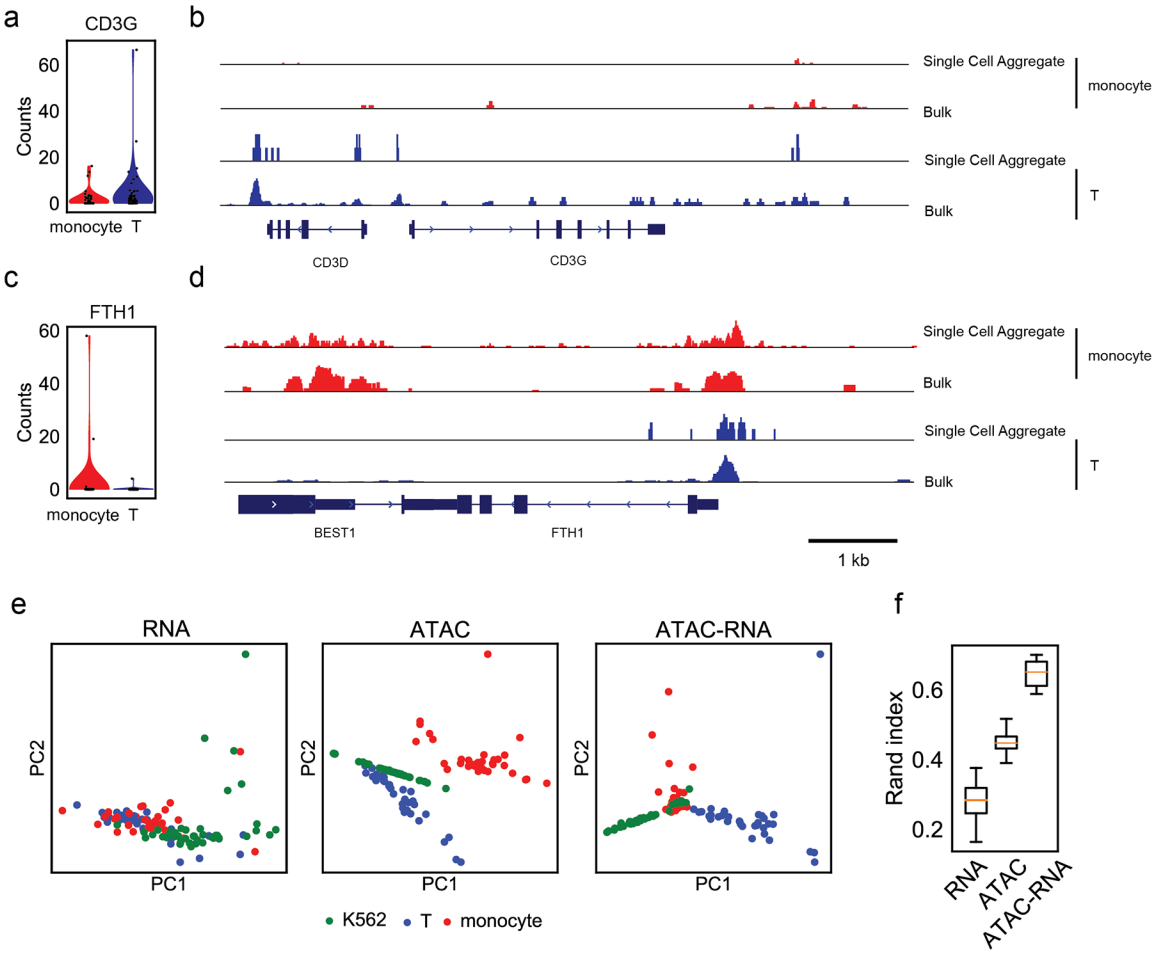

Figure 3. Linking chromatin-accessible regulatory elements and gene expression

Violin plots showing single cell gene expression and genome tracks showing aggregated locus accessibility for *CD3G* (a-b) and *FTH1* (c-d) from monocytes (38 cells) and T cells (42 cells). Bulk ATAC-seq (5,000 cells) tracks for monocytes and CD4 T cells are also shown. (e) Principal components analysis of all cells ($n = 116$) using different data types (RNA only, ATAC only, and combined ATAC + RNA). (f) Adjusted rand score distributions of Louvain clustering solutions for each data type when compared with true cell type labels. Cells are subsampled (90% of total, 20 times) and the rand index was computed for each to calculate the distribution. Boxes show the mean and interquartile range (IQR), with whiskers extending to 1.5 IQR in either direction from the top or bottom quartile.