High-throughput chromatin accessibility profiling at single-cell resolution

Anja Mezger1,2, Sandy Klemm1, Ishminder Mann3, Kara Brower4, Alain Mir3, Magnolia Bostick3, Andrew Farmer3, Polly Fordyce1,4,5,6, Sten Linnarsson2 & William Greenleaf1,6,7

Here we develop a high-throughput single-cell ATAC-seq (assay for transposition of accessible chromatin) method to measure physical access to DNA in whole cells. Our approach integrates fluorescence imaging and addressable reagent deposition across a massively parallel (5184) nano-well array, yielding a nearly 20-fold improvement in throughput (up to ~1800 cells/chip, 4–5 h on-chip processing time) and library preparation cost (~81¢ per cell) compared to prior microfluidic implementations. We apply this method to measure regulatory variation in peripheral blood mononuclear cells (PBMCs) and show robust, de novo clustering of single cells by hematopoietic cell type.

DOI: 10.1038/s41467-018-05887-x
A central challenge of systems biology is to determine the epigenome of phenotypically distinct cellular states within complex primary tissue. Toward this goal, single-cell chromatin accessibility measurements provide an important epigenetic view of the regulatory landscape within individual cells by capturing the physical accessibility of putative functional elements across the genome. Methods for measuring chromatin accessibility at single-cell resolution, however, are low throughput, depth limited, or require complex molecular processing to generate cellular indexing reagents. For ultra-high throughput accessibility profiling applications, combinatorial indexing approaches offer significant promise, yet these methods capture fewer accessible fragments per cell than single-cell isolation technologies and are not amenable to integration with single-cell microscopy or other multi-omic assays that require whole, live cells. In this report, we describe a high-throughput implementation of single-cell ATAC-seq (scATAC-seq) that directly integrates fluorescence imaging and provides an extensible foundation for multi-omic epigenetic profiling in single cells.

Results

Implementation of scATAC-seq on nanoliter-scale wells. We have implemented scATAC-seq in small volumes (µATAC-seq) using a recently developed nanoliter-scale liquid deposition system (ICELL8 Single Cell System, Takara Bio USA). This approach reduces reagent costs and achieves equal or higher per-cell fragment counts than prior state-of-the-art implementations. The workflow—illustrated in Figure 1a—is comprising of the following steps: (1) isolated single cells are stained with Hoechst and lysed & transposed; (2) transposase is released & indexed; (3) release buffer & indexed; (4) PCR amplification; (5) dispense & image.

Fig. 1 µATAC-seq: a nano-well scATAC-seq implementation on the ICELL8 platform. a µATAC-seq workflow. b Distribution of cell counts per well measured by fluorescence microscopy (Hoechst). c µATAC-seq library complexity for null, mouse, and human targeted wells using two separate polymerases (e2Tak and Q5) for well barcoding and amplification (n = 5000 wells). For each sample, the box denotes the interquartile range centered at the median (red line), while the whiskers span the 5th and 95th percentile range. d Correlation between nano-well chips processed with either a e2Tak (replicate 1) or Q5 polymerase (replicate 2) across all accessible loci. e Inter-well mixing of mouse and human µATAC-seq fragments. f Representative population and single-cell ATAC-seq genome tracks for the Gapdh locus. g Signal-to-background (percent reads in peaks) as a function of read depth (n = 792). Only cells lying in the upper right quadrant (marked by dashed lines) are retained for downstream analysis.
propidium iodide and stochastically loaded under Poisson statistics (~1 cell per well on average) across 5184 wells under active humidity and temperature control; all wells are then imaged via multi-color microscopy to identify those containing a single live cell; (2) transposition reagents are added to a selected set of wells (e.g., those containing a single live cell) and incubated at 37 °C for 30 min; (3) the transposition reaction is quenched by incubation with EDTA; (4) MgCl₂ is added in equimolar concentration to quench the chelating capacity of EDTA in preparation for subsequent PCR amplification; (5) PCR reagents are added and µATAC-seq fragment libraries are amplified using barcoded primers provided in the prior two steps (see Supplementary Table 1 for reagent loading chart). Following on-chip library construction, indexed µATAC-seq libraries are extracted from all nano-wells by centrifugation, purified, and then further amplified as necessary for sequencing (Methods section).

Benchmarking analysis of µATAC. As an initial test of µATAC-seq, we loaded samples into 5000 wells across two nano-well ICELL8 chips. On each chip, 200 wells were loaded with PBS (designated null wells); 1150 wells were loaded with mouse embryonic stem cells (mESCs, ~1 cell per well); and 1150 wells loaded with human lymphoblastoid GM12878 cells (~1 cell per well). This yielded a total of 4600 wells targeted with either human or mouse cells across both chips. Imaging of Hoechst and propidium iodide fluorescence revealed the anticipated fraction of wells containing live single cells (35%, 1616 single cells), consistent with near optimal loading that maximizes the number of single-cell containing wells (Fig. 1b). Barcoded sequencing of each of the 5000 targeted wells revealed 14.3 × 10⁶ (8.1 × 10⁵) median fragments per single human (mouse) cell containing wells (n = 1616)—reflecting a two orders of magnitude enrichment over null wells (Fig. 1c and Supplementary Figure 1a,b). These library complexes compare favorably with microfluidic cell capture (5.8 × 10⁵ fragments per GM12878 cell) as well as combinatorial indexing (2.5 × 10⁵ fragments per GM12878 cell) approaches. The µATAC-seq libraries capture both sub-nucleosome as well as nucleosome length fragments, yet, the median fragment length is shorter than that observed using the Fluidigm C1 platform. Consistent with prior bulk and single-cell ATAC-seq libraries, we observe a more than tenfold enrichment for fragments proximal to transcription start sites (TSS) relative to distal regions, reflecting a high fraction of fragments captured within open rather than closed chromatin (Supplementary Figure 2a). Furthermore, we find a high degree of concordance (97.9%) between nano-well chips even when µATAC-seq fragments are amplified with different polymerases (Fig. 1d). We further tested the deposition fidelity of the ICELL8 platform, observing both human and mouse cells in fewer than 0.2% of wells (Fig. 1e).

Aggregate single-cell profiles recapitulate population measurements broadly across the accessible genome (Supplementary Figure 2b) as well as specifically at individual genomic loci (Fig. 1f). At single-cell resolution, accessibility profiles are enriched for open chromatin (Fig. 1f, g) in both mESCs (29% reads in peaks, Fig. 1g) and GM12878 cells (22% reads in peaks, Supplementary Figure 2c). Collectively, these data establish the proposed nano-well implementation as a high-throughput framework for scATAC-seq library construction.

Epigenetic signature distinguishes PBMC types. We next asked whether µATAC-seq epigenetic profiles are sufficient to distinguish cell types within complex primary tissue. For this purpose, we performed µATAC-seq on human peripheral blood mononuclear cells (PBMCs) as well as B, T, CD4⁺ T, CD8⁺ T, and monocyte cells isolated directly from whole blood (Fig. 2a), yielding 2333 single cells passing all quality control criteria (Methods section). Using ChromVar, a bioinformatic approach described previously, we calculated the relative accessibility of transcription factor (TF) binding motifs in individual cells and found that isolated B, T, and monocyte cells robustly cluster by cell type (Supplementary Figure 3a). By aggregating fragments within single cells that are proximal to a TF motif, this epigenetic signature captures the variation in putative TF binding site accessibility across a population of cells. A relatively small fraction of cells are incorrectly assigned to clusters; however, the frequency of these events as well as the random distribution of these cells within apposing clusters both suggest that isolation impurity upstream of the µATAC-seq assay is the primary source of these errors (Supplementary Table 2). PBMC subpopulations co-cluster precisely with the isolated cell types (Fig. 2b, c), showing highly concordant cell type-specific accessibility patterns within appropriate ISNE (t-Distributed Stochastic Neighbor Embedding) clusters (Fig. 2c) as well as k-means clustering across highly variable TF binding motif accessibility patterns (Fig. 2b). Consistent with published gene expression data, we find that the PU.1 binding motif is differentially accessible in monocytes and B cells relative to T cells (Fig. 2c, upper right panel)11,12, the C/EBPα motif is exclusively accessible in monocytes (Fig. 2c, lower left panel)13,14, and RUNX1 motif accessibility is appropriately enriched in T cells—reflecting the broad regulatory role of the RUNX protein family in T lymphocytes (Fig. 2c, lower right panel)15. These results are highly robust to biological (three human blood donors) and technical variation (Supplementary Figure 3b). To further establish the robustness of clustering by cell type, we independently purified CD4⁺ and CD8⁺ T cells and found that these subtypes co-cluster with independently isolated T cells (Fig. 2c, upper left panel). Collectively, these data suggest that µATAC-seq signatures are sufficient for de novo clustering of PBMCs by hematopoietic cell type.

Discussion
In this report, we have described µATAC-seq—a high-throughput, single-cell chromatin accessibility assay that dramatically reduces per-cell costs, requires only commercially available reagents, provides state-of-the-art data quality, and increases throughput nearly 20-fold over existing single-cell capture technologies. Single-cell chromatin measurements present a unique experimental challenge since only two DNA templates are present in a diploid cell. Technical sampling noise as well as biological heterogeneity further confound this problem, resulting in a ~10% observation efficiency of accessible regions in single cells. Consequently, a few hundred cells are typically required to reliably determine the accessibility landscape of each subpopulation within a mixture of cells. Our approach in this work has been to develop an experimental framework for processing more than a thousand of cells in parallel to determine the accessibility of multiple cell types within a complex tissue. In general, nano-well single-cell sequencing approaches such as µATAC-seq are highly extensible, well-suited for multi-omic analysis, and define an important direction for single-cell epigenetic methods development.

Methods
Cell culture. All cell lines were grown at 37 °C with 5% CO₂. GM12878 cells were obtained from the laboratory of Michael Synder (Stanford University) and were cultured in RPMI 1640 media supplemented with t-glutamine (Thermo Fisher Scientific, MA, USA, Cat. #11875–085) and 10% FBS (Thermo Fisher Scientific, Cat. #100990261). mESCs cells (129SIX CASTANEOUS, gift from Howard Chang) were cultured in 15% FBS (HyClone GE Healthcare Life Sciences, SH30070.03E) supplemented with non-essential amino acids, t-glutamine and Leukemia Inhibitory
Figure 2 De novo identification of hematopoietic cell types by μATAC-seq. a Human PBMC isolation and μATAC-seq workflow. b Hierarchical (TF motifs) and k-means (cells) clustering of accessibility deviation z-scores across 2333 single cells (columns) of the 50 most variable TF motifs (rows). Colors correspond to cell types defined in a. c tSNE visualization of accessibility deviations at TF motifs for cells in b. Cells are either colored by cell type (upper left panel) or by the accessibility deviation z-score for the specified TF motif following imaging, the Tn5 transposition mix (2X TD buffer [20% dimethylformamide, 20 mM Tris-HCl, pH 7.6, 10 mM MgCl₂], 100 µl Tn5 transposase [Nextera DNA Library Prep Kit, Illumina, CA, USA] per ml Tn5 transposition mix, 0.2% Tween 20, 0.2% NP40, and 0.02% Digitonin [Promega, WI, USA]) was dispensed. Chips were then sealed with imaging film, centrifuged at ~3000 g for 5 min at 4 °C and incubated for 30 min at 37 °C. To index the whole chip, the i5 and i7 previously published, custom indices (Supplementary Table 3) were dispensed at 6.25 µM concentration with EDTA and MgCl₂, respectively. To release the bound Tn5 transposase, 60 mM EDTA was dispensed together with the i5 indexes. Chips were then sealed with imaging film, centrifuged, and incubated at room temperature for 5 min. Finally, a PCR mix (5x Q5 [NEB] or e2TAK [Takara Bio USA] reaction buffer, 1 mM dNTPs [Thermo Fisher Scientific], and 100 U/ml Q5 [NEB] or 50 U/ml e2TAK polymerase [Takara Bio USA]) was added. After sealing the chip, it was centrifuged at 3000 g for 3 min and incubated for 30 min at 50 °C. Prior to performing PCR on-chip, the chelating capacity of EDTA was suppressed by dispensing 60 mM MgCl₂ together with the i7 indices. Chips were then sealed with imaging film, centrifuged, and incubated at room temperature for 5 min. For sequencing, the PCR product was split across four MinElute columns, eluted in 10 µl volumes, and following purification and amplification, the collected PCR product was purified using MinElute PCR purification columns (Qiagen, Germany) following the manufacturer’s instructions. Due to the large sample volume, the PCR product was split across four MinElute columns, eluted in 10 µl volumes, and
subsequently pooled. To remove free PCR primers, which would induce index-swapping during additional rounds of off-chip amplification, we performed two rounds of bead clean-up using Ampure XP beads (Beckman Coulter, CA, USA) in a 1:1.2 ratio. The beads were incubated for 8 min with the PCR product, washed twice in 70% ethanol, and eluted in 20 μl ultrapure water (Thermo Fisher Scientific). Further amplification was required only for the mouse and human mixing experiment. PBMCs libraries generated on-chip were directly sequenced following column and bead purifications.

The number of required off-chip amplification cycles was determined by running a 20 μl qPCR reaction (2 μl PCR product, 0.5 μM oligo C [Ilumina P5], 0.5 μM oligo D [Ilumina P7], 0.6X SYBR Green I [Thermo Fisher Scientific], and 1X NEBNext High-Fidelity 2X PCR Master Mix (NEB)): 30 s at 98 °C, followed by 20 cycles of 10 s at 98 °C and 30 s at 63 °C and 1 min at 72 °C. The remaining 18 μl PCR product was amplified using the number of PCR cycles corresponding to 1/3 of the maximum fluorescence intensity. The amplified PCR product was then purified and concentrated using a Qiagen MinElute column.

DNA sequencing. All libraries were sequenced on a NextSeq 500 (Illumina) using the high output v2 kit (Illumina) in 76 × 8 × 8 × 76 cycle mode, although 38 bp × 8 × 8 × 38 bp sequencing is sufficient. On average, ~50 K reads were sequenced per cell. Due to the nature of the sequencing libraries 30–40 phix control v3 (Ilumina) was spiked in and 1.5 pM were loaded onto the flow cell.

Per cell cost estimate. The per cell library preparation cost is conservatively estimated (assuming only 1200 single cells captured per chip) at 814/cell: (1) Takara Bio Ace2 (1.8 chip/225 cell), (2) Ilumina TruSeq (244/cell), (3) Qiagen polymerase (44/cell), (4) other reagents contribute <1% additionally. The additional per cell sequencing cost at the depth used for this report (assuming a 75 cycles NextSeq 500/550 High Output v2 Kit) is approximately 17¢/cell.

Data analysis. Illumina sequencing reads in BCL format were demultiplexed by single-cell barcode to fastq files using bcftools (Illumina) according to the manufacturer’s manual. Reads were trimmed using Cutadapt18 (parameters: --a Trans2_rc = C/T/G/C/T/T/T/T/T/T/T/T/A/AC/T/TG/A/G/G/G/G/G/G/G/A/G/G/A/G, Trans1_rc = C/T/G/T/C/T/T/T/T/T/T/T/A/AC/T/TG/G/C/T/G/G/G/G/A/G/G/G/A/G) and aligned to either the human (hg19) or mouse (mm9) genomes using Bowtie219. Mitochondrial reads were removed prior to downstream analysis. PCR duplicates were identified and removed if either the start or end position was shared with another sequencing read. Library complexity estimates were obtained using the Picard Tools MarkDuplicates utility (https://broad-institute.github.io/picard/), except for empty well reads. For empty well reads was present for a robust estimate; in the latter case, the library complexity was estimated as the number of unique reads observed. Accessible complexity was estimated as the number of unique reads observed. Accessible transcription factors (rows) were then hierarchically clustered using the ward.D2 agglomeration method19,20 within the R pheatmap package21, while single cells (columns) were ordered by assigned cluster and cell type (Fig. 2b). Visualizations of clustering and tSNE22 analyses were constructed using R scripts.

Data availability. The sequencing data that support the findings of this study are available in Figgshere under the following DOIs: Metadata: doi: 10.6084/m9.figshare.7006154.v1; Human monocyte cells: doi:10.6084/m9.figshare.7005707.v1; Human lymphoblastoid cell (GM12878): doi:10.6084/m9.figshare.7005713.v1; Human peripheral blood mononuclear cells (PBMCs): doi:10.6084/m9.figshare.7005752.v1; Mouse embryonic stem cells (mESCs): doi:10.6084/m9.figshare.7005710.v1; Human CD8+ T cells: doi:10.6084/m9.figshare.7005698.v1; Human T cells: doi:10.6084/m9.figshare.7005593.v1. All other data are available from the authors upon reasonable request.

Received: 18 April 2018 Accepted: 24 July 2018

NATURE COMMUNICATIONS | DOI: 10.1038/s41467-018-05887-x

References

1. Corces, M. R. et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. Nat. Genet. 48, 1193–1203 (2016).
2. Cusanovich, D. A. et al. The cis-regulatory dynamics of embryonic development at single-cell resolution. Nature 555, 538–542 (2018).
3. Buenrostro, J. D. et al. Single-cell chromatin accessibility reveals principles of regulatory variation. Science 351, 241–246 (2016).
4. Pott, S. Simultaneous measurement of chromatin accessibility, DNA methylation, and nucleosome phasing in single cells. eLife 6, e23203 (2017).
5. Jin, W. et al. Genome-wide detection of DNase I hypersensitive sites in single cells and FFPE tissue samples. Nature 528, 142–146 (2015).
6. Thurman, R. E. et al. The accessible chromatin landscape of the human genome. Nature 489, 75–82 (2012).
7. Cusanovich, D. A. et al. Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. Science 348, 910–914 (2015).
8. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods 10, 1213–1218 (2013).
9. Chen, H. et al. Neutrophils and monocytes express high levels of PU.1–Sphi but not Sph-B. Blood 85, 2918–2928 (1995).
10. Lloberas, J., Soler, C. & Celada, A. The key role of PU.1/SP1-1 in B cells, myeloid cells and macrophages. Immunol. Today 20, 184–189 (1999).
11. Di Tullio, A. et al. CCAAT/enhancer binding protein alpha (C/EBP(alpha))-induced transdifferentiation of pre-B cells into macrophages involves no overt retrodifferentiation. Proc. Natl Acad. Sci. USA 108, 17016–17021 (2011).
12. Haisma, C. V., Stadtfeld, M., Xie, H., de Andres-Aguayo, L. & Graf, T. Reprogramming of committed T cell progenitors to macrophages and dendritic cells by C/EBP alpha and PU.1 transcription factors. Immunity 25, 731–744 (2006).
13. Kohu, K. et al. Pleiotropic roles of runx transcription factors in the differentiation and function of T lymphocytes. Curr. Immunol. Rev. 4, 101–115 (2008).
14. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 17, 10 (2011).
15. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359 (2012).
16. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9, R137 (2008).
17. Ward, J. H. Hierarchical grouping to optimize an objective function. J. Am. Stat. Assoc. 58, 236–244 (1963).
18. Murtagh, F. & Legendre, P. Ward’s hierarchical agglomerative clustering method: which algorithms implement ward’s criterion? J. Classif. 31, 274–295 (2014).
19. Kolde, R. Heatmap: Pretty Heatmaps. (R package, version 1.8.0, 2015). Available at: https://CRAN.R-project.org/package=heatmap.
20. Xu, J. et al. Landscape of monoallelic DNA accessibility in mouse embryonic stem cells and neural progenitor cells. Nat. Genet. 49, 377–386 (2017).

Acknowledgements A. Mezger is supported by the Swedish Research Council (grant 2015–06403). S.K. is supported by a T32 Ruth L. Kirschstein National Research Service Award (Institutional Training Grant in Genome Science NIH 5 T32 HG000044). This work was supported by NIH (P50HG007735 and U11HG09442 and U19AI057266 to W.G.), the Rita Allen Foundation (to W.G.), the Baxter Foundation Faculty Scholar Grant, and the Human Frontiers Science Program grant R1G0065 (to W.G.), and the Joint Institute for Metrology in Biology. W.G. is a Chan Zuckerberg Bioph Case investigator and acknowledges grants 2017–174468 and 2018–182817 from the Chan Zuckerberg Initiative.

Author contributions A. Mezger, S.K., S.L., and W.G. conceived the study and designed all experiments; A. Mezger and S.K. analyzed the data and performed all experiments with the help of J.M.; K.B. assisted with imaging analysis; A. Mir, M.B., A.F., F.P., and W.G. supervised the study; A. Mezger, S.K., and W.G wrote the manuscript with input from all authors.

Additional information Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-05887-x.
Competing interests: I.M., A. Mir, M.B., and A.F. are employees at Takara Bio USA, Inc. S.K. presented the described work on behalf of Takara Bio USA, Inc. at the Advances in Genome Biology and Technology General Meeting (2018), but has no financial interest in this work. W.G. is a scientific co-founder of Epinomics and a consultant for 10X genomics. Stanford University has filed a provisional patent application (US2016006691A1; Status: Components are pending) on the ATAC-seq methods described and W.G. is named as an inventor. A. Mezger, K.B., P.F., and S.L. declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.