A rapid and robust method for single cell chromatin accessibility profiling

Xi Chen1, Ricardo J. Miragaia1,4, Kedar Nath Natarajan1,5 & Sarah A. Teichmann1,2,3

The assay for transposase-accessible chromatin using sequencing (ATAC-seq) is widely used to identify regulatory regions throughout the genome. However, very few studies have been performed at the single cell level (scATAC-seq) due to technical challenges. Here we developed a simple and robust plate-based scATAC-seq method, combining upfront bulk Tn5 tagging with single-nuclei sorting. We demonstrate that our method works robustly across various systems, including fresh and cryopreserved cells from primary tissues. By profiling over 3000 splenocytes, we identify distinct immune cell types and reveal cell type-specific regulatory regions and related transcription factors.
Due to its simplicity and sensitivity, ATAC-seq has been widely used to map open chromatin regions across different cell types in bulk. Recent technical developments have allowed chromatin accessibility profiling at the single cell level (scATAC-seq) and revealed distinct regulatory modules across different cell types within heterogeneous samples. In these approaches, single cells are first captured by either a microfluidic device or a liquid deposition system, followed by independent tagmentation of each cell. Alternatively, a combinatorial indexing strategy has been reported to perform the assay without single cell isolation. However, these approaches require either a specially engineered and expensive device, such as a Fluidigm C1 or Takara ICELL8, or a large quantity of custom-modified Tn5 transposase.

Here, we overcome these limitations by performing upfront Tn5 tagging in the bulk cell population, prior to single-nuclei isolation. It has been previously demonstrated that Tn5 transposase-mediated tagmentation contains two stages: a tagging stage where the Tn5 transposase binds to DNA, and a fragmentation stage where the Tn5 transposase is released from DNA by heat or denaturing agents, such as sodium dodecyl sulfate (SDS). As the Tn5 tagging does not fragment DNA, we reasoned that the nuclei would remain intact after incubation with the Tn5 transposase in an ATAC-seq experiment. Based on this idea, we developed a simple, robust and flexible plate-based scATAC-seq protocol, performing a Tn5 tagging reaction on a pool of cells (5000–50,000) followed by sorting individual nuclei into plates containing lysis buffer. With this easy and quick workflow, it only takes a few hours to prepare sequencing-ready libraries, and the method can be implemented by any laboratory using standard equipment.

Results

Benchmark and comparison to Fluidigm C1 scATAC-seq. We first tested the accuracy of our sorting by performing a species mixing experiment, where equal amounts of HEK293T and NIH3T3 cells were mixed, and scATAC-seq was performed with our method. Using a stringent cutoff (Online Methods), we obtained median library sizes ranging from 52,747 (mSF) to 104,608.5 (mCPC_E8.5) unique fragments (Fig. 2a). The amount of mitochondrial DNA also varied across cell types but was low in all samples (<13%). All four tested samples had very high signal-to-noise ratio, with a median FRIP ranging from 0.50 (mSF) to 0.60 (hSF) (Fig. 2a). The insert size distributions of the aggregated single cells from all four samples exhibited clear nucleosomal banding patterns (Fig. 2b), which is a feature of high quality ATAC-seq libraries. Finally, visual inspection of aggregate of single cell profiles showed clear open chromatin peaks around expected genes (Fig. 2c, d), Details of all tested cells/tissues are summarised in Supplementary Data 1.

Validation using different cryopreserved cells. To evaluate the technical robustness of our plate-based method, we further tested it by generating the chromatin accessibility profiles of 3648 splenocytes (after red blood cell removal) from two C57BL/6Jax mice. In total, we performed two 96-well plates and nine 384-well plates. By setting a stringent quality control threshold (>10,000 reads and >90% mapping rate), 3385 cells passed the technical cutoff (>90% successful rate) (Supplementary Figure 3b). The aggregated scATAC-seq profiles exhibited good coverage and signal and resembled the bulk data generated from 10,000 cells by the Immunological Genome Project (ImmGen) (Fig. 3a). The library fragment size distribution before and after sequencing both displayed clear nucleosomal banding patterns (Fig. 3b and Supplementary Figure 4a). In addition, sequencing reads showed strong enrichment around transcriptional start sites (TSS) (Fig. 3c), further demonstrating the quality of the data was high.

Profiling chromatin accessibility of mouse splenocytes. After this validation of the technical robustness of our plate-based method, we further tested it by generating the chromatin accessibility profiles of 3648 splenocytes (after red blood cell removal) from two C57BL/6Jax mice. In total, we performed two 96-well plates and nine 384-well plates. By setting a stringent quality control threshold (>10,000 reads and >90% mapping rate), 3385 cells passed the technical cutoff (>90% successful rate) (Supplementary Figure 3b). The aggregated scATAC-seq profiles exhibited good coverage and signal and resembled the bulk data generated from 10,000 cells by the Immunological Genome Project (ImmGen) (Fig. 3a). The library fragment size distribution before and after sequencing both displayed clear nucleosomal banding patterns (Fig. 3b and Supplementary Figure 4a). In addition, sequencing reads showed strong enrichment around transcriptional start sites (TSS) (Fig. 3c), further demonstrating the quality of the data was high.

Importantly, for the majority of the cells, less than 10% (median 2.1%) of the reads were mapped to the mitochondrial genome (Supplementary Figure 3a). Overall, we obtained a median of 643,734 reads per cell, whereas negative controls (empty wells) generated only ~100–1000 reads (Supplementary Figure 3b). In most cells, more than 98% of the reads were mapped to the mouse genome (Supplementary Figure 3b), indicating low level of contamination. The median of estimated library sizes is 31,808.5 (Supplementary Figure 3c). At the sequencing depth of this experiment, the duplication rate of each single cell library is ~95% (Supplementary Figure 3d), indicating that the libraries were sequenced to near saturation. Down-sampling the raw reads (from the fastq files) and repeating the analysis suggest that at 20–30% of our current sequencing depth, the majority of the fragments would have already been captured (Supplementary Figure 4a and b). Therefore, in a typical scATAC-seq experiment, ~120,000 reads per cell are sufficient to capture most of the unique fragments, with higher sequencing depth still increasing the number of detected unique fragments (Supplementary Figure 3e).

Next, we examined the data to analyse signatures of different cell types in the mouse spleen. Reads from all cells were merged, and a total of 78,048 open chromatin regions were identified by peak calling with q-values less than 0.01 (Methods). We binarised peaks as “open” or “closed” (Methods) and applied a Latent Semantic Indexing (LSI) analysis to the cell-peak matrix for dimensionality reduction (Methods). Consistent with previous findings, the first dimension is primarily influenced by sequencing depth (Supplementary Figure 3f). Therefore, we only focused on the second dimension and upwards and...
visualised the data by t-distributed stochastic neighbour embedding (t-SNE)\textsuperscript{19}. We did not observe batch effects from the two profiled spleens, and several distinct populations of cells were clearly identified in the t-SNE plot (Fig. 3d). Read counts in peaks near key marker genes (e.g. Bcl11a and Bcl11b) suggested that the major populations are B and T lymphocytes, as expected in this tissue (Fig. 4a). In addition, we found a small population of granulocytes (GN), most clusters as different subtypes of B, T and Natural Killer (NK) cells, as well as a small population of granulocytes (GN), dendritic cells (DC) and macrophages (MF) (Fig. 4b and Supplementary Figure 7). In this way, we identified most clusters as different subtypes of B, T and Natural Killer (NK) cells, as well as a small population of granulocytes (GN), dendritic cells (DC) and macrophages (MF) (Fig. 4b and Supplementary Data 2). An aggregate of all single cells within the same predicted cell type agrees well with the ImmGen bulk ATAC-seq profiles (Supplementary Figure 8). Remarkably, the aggregate of as few as 55 cells (e.g. the predicted MF cell cluster) already exhibited typical bulk ATAC-seq profiles (Supplementary Figure 8). This finding opens the door for a different ATAC-seq experimental design, where Tn5 tagging can be performed upfront on large populations of cells (e.g. $5000–50,000$ cells).

To systematically interrogate various cell populations captured in our experiments, we applied a spectral clustering technique\textsuperscript{21} which revealed 12 different cell clusters (Fig. 4b). Reads from cells within the same cluster were merged together to form ‘pseudo-bulk’ samples and compared to the bulk ATAC-seq data sets generated by ImmGen (Supplementary Figures 6 and 7). Cell clusters were assigned to the most similar ImmGen cell type (Fig. 4b and Supplementary Figure 7). In this way, we identified most clusters as different subtypes of B, T and Natural Killer (NK) cells, as well as a small population of granulocytes (GN), dendritic cells (DC) and macrophages (MF) (Fig. 4b and Supplementary Data 2). An aggregate of all single cells within the same predicted cell type agrees well with the ImmGen bulk ATAC-seq profiles (Supplementary Figure 8). Remarkably, the aggregate of as few as 55 cells (e.g. the predicted MF cell cluster) already exhibited typical bulk ATAC-seq profiles (Supplementary Figure 8). This finding opens the door for a different ATAC-seq experimental design, where Tn5 tagging can be performed upfront on large populations of cells (e.g. $5000–50,000$ cells). Subsequently, cells of interest (for example, marked by surface marker probes) can be isolated by FACS, and libraries generated for subsets of cells only. This will be a simple and fast way of obtaining scATAC-seq profiles for rare cell populations.
To test the feasibility of this idea, we stained mouse splenocytes with an anti-CD4 antibody conjugated with PE and performed tagmentation afterwards. The PE signal remained after tagmentation (Supplementary Figure 9), allowing us to specifically sort out CD4-positive T cells from the rest of the splenocytes for analysis (we named these “TagSort” libraries). As a control, we first purified CD4 T cells using an antibody-based depletion method (Methods), and subsequently performed scATAC-seq on the purified CD4 T cells (we named these “SortTag” libraries). The data of CD4 T cells generated from these two strategies agree very well (Fig. 4c). The library complexity is comparable with median library sizes of 30,953 and 25,830, respectively (Fig. 4c, top left panel). The binding signals around open chromatin peaks are highly correlated (Pearson $r = 0.96$) (Fig. 4c, top right panel). Visual inspection of read pileup profiles around the Cd4 gene locus from single cell aggregates suggested the data are of good quality (Fig. 4c, bottom panel).

This experiment serves as a proof-of-principle test where staining of a surface marker can be done before Tn5 tagging, and a specific population can be sorted by FACS afterwards for scATAC-seq analysis. It should be noted that we have only tested CD4—an abundant marker in a subpopulation of splenocytes. Other surface markers in different tissues would need to be investigated individually. In addition, the ability to investigate rare cell populations using this approach is limited by the frequency of the rare cell types and the amount of cells that can be tagged upfront.

The spectral clustering was able to distinguish different cell subtypes, such as naive and memory CD8 T cells, naive and regulatory CD4 T cells and CD27+ and CD27− NK cells (Fig. 4b). Previous studies have identified many enhancers that are only accessible in certain cell subtypes, and these are robustly identified in our data. Examples are the Ilr2b and Cd44 loci in memory CD8 T cells22 and Ikzf2 and Foxp3 in regulatory T cells23 (Supplementary Figure 10a and b). Interestingly, our clustering approach successfully identified two subtle subtypes of NK cells (CD27− and CD27+ NK cells), as determined by their open chromatin profiles (Fig. 4b, d). It has been shown that, upon activation, NK cells can express CD83, a well-known marker for mature dendritic cells24. In mouse spleen, CD83 expression was barely detectable in the two NK subpopulations profiled by the ImmGen consortium (Supplementary Figure 10c). However, in our data, the Cd83 locus exhibited different open chromatin states in the two NK clusters (Fig. 4d). Multiple ATAC-seq peaks were observed around the Cd83 locus in the CD27+ NK cell cluster but not in the CD27− NK cluster (Fig. 4d). This suggests that Cd83 is in a transcriptionally permissive state in the CD27+ NK cells, and the CD27+ NK cells have a greater potential for rapidly producing CD83 upon activation. This may partly explain the functional differences between CD27+ and CD27− NK cell states26.

Finally, we investigated whether we could identify the regulatory regions that define each cell cluster. To this end, we trained a logistic regression classifier using the spectral
clustering labels and the binarised scATAC-seq count data (Methods). From the classifier, we extracted the top 500 open chromatin peaks (marker peaks) that can distinguish each cell cluster from the others (Fig. 4e and Methods). By looking at genes in the vicinity of the top 50 marker peaks, we recapitulated known markers, such as \textit{Cd4} for the helper T cell cluster (cluster 3), \textit{Cd8a} and \textit{Cd8b1} for the cytotoxic T-cell cluster (cluster 6) and \textit{Cd9} for marginal zone B cell cluster (cluster 4) (Supplementary Figure 11 and Supplementary Data 3). These results are consistent with our correlation-based cell cluster annotation (Fig. 4b).

Whereas the peaks at TSS are useful for cell-type annotation, the majority of the cluster-specific marker peaks are in intronic and distal intergenic regions, in line with the global peak distribution (Supplementary Figure 12). To identify transcription factors that are important for the establishment of these marker peaks, we investigated them in more detail by motif enrichment analysis using HOMER\textsuperscript{27}. The full results of these motif enrichment analyses are included in Supplementary Data 4. As expected, different ETS motifs and ETS-IRF composite motifs were significantly enriched in marker peaks of many clusters (Fig. 4f), consistent with the notion that ETS and IRF transcription factors are important for regulating immune activities\textsuperscript{28}. Furthermore, we found motifs that were specifically enriched in certain cell clusters (Fig. 4f). Our motif discovery is consistent with previous findings, such as the importance of T-box (e.g. Tbx21) motifs in NK\textsuperscript{29} and CD8T memory cells\textsuperscript{30} and POU domain (e.g. Pou2f2) motifs in marginal zone B cell\textsuperscript{31}. This suggests that our scATAC-seq data are able to identify known gene regulation principles in different cell types within a tissue.

**Discussion**

In recent years, other methods, such as DNase-seq\textsuperscript{32}, MNase-seq\textsuperscript{33} and NOME-seq\textsuperscript{34,35} have investigated chromatin status at the single cell level. However, due to its simplicity and reliability, ATAC-seq currently remains the most popular technique for chromatin profiling. Several recent studies have demonstrated the power of using scATAC-seq for investigating regulatory principles, e.g. brain development\textsuperscript{19}, Mouse sci-ATAC-seq Atlas\textsuperscript{36} and pseudotime inference\textsuperscript{37}. The combined multi-omics approaches also began to emerge, such as sci-CAR-seq\textsuperscript{38}, scCAT-seq\textsuperscript{39} and piATAC-seq\textsuperscript{8}. Our study added on top of those methods to provide a simple and easy-to-implement scATAC-seq approach that can successfully detect different cell populations, including subtle and rare cell subtypes, from a complex tissue. More
importantly, it is able to reveal key gene regulatory features, such as cell-type-specific open chromatin regions and transcription factor motifs, in an unbiased manner. Future studies can utilise this method to unveil the regulatory characteristics of novel and rare cell populations and the mechanisms behind their transcriptional regulation.

**Methods**

**Ethics statement.** The mice were maintained under specific pathogen-free conditions at the Wellcome Trust Genome Campus Research Support Facility (Cambridge, UK). These animal facilities are approved by and registered with the UK Home Office. All procedures were in accordance with the Animals (Scientific Procedures) Act 1986. The protocols were approved by the Animal Welfare and Ethical Review Body of the Wellcome Trust Genome Campus.
Fig. 4 Identification of different cell types and cell-type-specific open chromatin regions and transcription factor motifs. a The same t-SNE plot as in Fig. 3d, coloured by the number of counts in the peaks near indicated gene locus. b The same t-SNE plot as in Fig. 3d coloured by spectral clustering and cell-type annotation. c Comparisons of spleen CD4+ T cells scATAC-seq obtained by two strategies. TagSort: cells were stained with anti-CD4-PE, tagged with Tn5 and CD4-PE-positive cells were sorted for scATAC-seq; SortTag: CD4 T cells were purified first and scATAC-seq was performed on the purified cells. Top: comparison of library size and binding signal correlation (pearson r = 0.96) around called peaks; bottom: UCSC genome browser tracks of the indicated single cell aggregates around the Cd4 gene locus. d UCSC genome browser tracks around Cd27 and Cd83 gene loci, displaying the aggregate (top panel) and single cell (bottom panel) signals of the two NK clusters. ATAC-seq peaks specific to the CD27+ NK cells are highlighted. For visual comparison reason, we randomly choose 65 out of 75 CD27- NK cells. e z-score of normalised read counts in the top 500 peaks that distinguish each cell cluster based on the logistic regression classifier, across each peak (row) in each cell (column). Top 500 marker peaks were picked per cell cluster, so there are 500 × 12 = 6000 peaks in the heatmap. Cells are ordered by cluster labels. f Heatmap representation of transcription factor motif (rows) enrichments (binomial test p-values) in the top 500 marker peaks in different cell clusters (columns). Some key motifs are enclosed by black rectangles and motif logos are shown to the right. Motif names are taken from the HOMER software suite

Cell isolation. For splenocytes, the spleen from a C57BL/6Jax mouse was mashed by a 2 ml syringe plunger through a 70 μm cell strainer (Fisher Scientific 10788201) into 30 ml 1X DPBS (Thermo Fisher 14190169) supplied with 2 mM EDTA and 0.5% (v/v) BSA (Sigma A9418). Cells were centrifuged down, supernatant was removed, and the cell pellet was briefly vortexed. 5 ml 1X RBC lysis buffer (Thermo Fisher 04-500-003) was used to resuspend the cell pellet, and the cell suspension was vortexed again for 10 s. Cells were incubated on bench for 5 min. Then 45 ml 1X DPBS was added, and cells were centrifuged down. Volume of 30 ml 1X DPBS was used to resuspend the cell pellet. The cell suspension was passed through a Miltenyi 30 μM Pre-Separation Filter (Miltenyi 130-041-047), and the cell number was determined using C-chip counting chamber (VWR DHC-01). All centrifugations were done at 500× g, 4 °C, 5 min. For human and mouse skin fibroblasts, cells were extracted as previously described. For mouse cardiac progenitor cells, cells were extracted as previously described. Cells were cryopreserved in 90% FBS and 10% DMSO and stored in liquid nitrogen until experiments.

Plate-based single-cell ATAC-seq (scATAC-seq). A detailed step-by-step protocol can be found in Supplementary Methods. Briefly, 50,000 cells were centrifuged down at 500× g, 4 °C, 5 min. Cell pellets were resuspended in 50 μl of 1X DPBS with 0.5% BSA was added and the cell suspension was passed through a Miltenyi 30 μM Pre-Separation Filter (Miltenyi 130-041-047), and the cell number was determined using C-chip counting chamber (VWR DHC-01). All centrifugations were done at 500× g, 4 °C, 5 min. For human and mouse skin fibroblasts, cells were extracted as previously described. Cells were cryopreserved in 90% FBS and 10% DMSO and stored in liquid nitrogen until experiments.

Species mixing experiments. A total of 25,000 HEK293T (Human, ATCC® CRL-3216®) and 25,000 NIH3T3 (Mouse, ATCC® CRL-1658®) cells were mixed together, and scATAC-seq was performed as described in Supplementary Methods. The obtained sequencing reads were mapped to a concatenated genome of mouse and human by hisat22. One 384-well plate was performed. We first set a technical cutoff where a successful well must contain more than 10,000 total reads and more than 90% of reads are mapped to the concatenated genome. In all, 307 wells were marked as successful. Among the successful wells, we calculated the ratio of reads that mapped to the human genome and the mouse genome. If the ratio is larger than 0.9, the well is categorised as containing human single cells; if the ratio is less than 0.1, the well is categorised as containing mouse single cells; otherwise, the well is categorised as containing human-mouse doubles.

Plate scATAC-seq on CD4+ T cells (TagSort vs. SortTag). For the “TagSort” strategy, 50,000 splenocytes were stained with anti-Mouse CD4-PE (Bioscience cat. no. 12-0043-82) at room temperature for 30 min according to the manufacturer instructions. The stained cells were washed with ice-cold 1X PBS twice and pelleted down at 4 °C, 5 min. Experiments were carried out following the procedures described in Supplementary Methods. DAPI and PE double-positive cells were sorted into a 384-well plate for library construction. For the “SortTag” strategy, CD4+ T cells were purified first from mouse splenocytes using the Naive CD4 T-Cell Isolation Kit, Mouse (Miltenyi, cat. no. 130-104-453) following the manufacturer’s instruction without the anti-CD44 depletion step. The purified CD4+ T cells were processed according to the procedures described in Supplementary Methods.

scATAC-seq using Fluidigm C1. Experiments were performed as previously described using the medium-sized (1862x) Open App chip. We followed the manufacturer’s instructions described in the “ATAC Seq No Stain (Rev C)” from the Fluidigm ScriptHub (https://www.fluidigm.com/clopapp/scripthub), except that we replace the detergent NP-40 in the original protocol with digitonin so that the final concentration of digitonin in the reaction chamber is 0.05%. After collecting the pre-amplified material from the Fluidigm chip, the libraries were indexed by library PCR for 14 cycles as previously described.

Costs involved in plate-based and Fluidigm C1 scATAC-seq. For our plate-based scATAC-seq method, most reagents and buffers are available in a standard molecular biology lab. Exceptions are the Tn5 transposase, which can be purchased from Illumina (Cat. No. FC-121-1030), and the PCR master mix, which can be purchased from various vendors (we used the 2X NEBNex® High-Fidelity 2X PCR Master Mix from NEB). As the Tn5 tagging reaction was performed upfront at the bulk level, the Tn5 cost per cell depends on how many cells are sorted during the sorting. Based on our experience, when 50,000 cells are used at the beginning, two to eight 384-well plates can be sorted. Therefore, the cost of Tn5 is negligible. The major cost per unit for the plate-based scATAC-seq is the PCR master mix used during library amplification. Currently, 10 μl of PCR master mix are needed per cell in a 20 μl library amplification reaction, but we have been successfully and consistently generated libraries from half of the volume described in the protocol. For scATAC-seq using the Fluidigm C1, all the aforementioned reagents are needed, and a microfluidic chip is required per 96 wells.

qPCR for library amplification. After assembly of the 20 μl PCR reaction (see Supplementary Methods), a pre-amplification step was performed on a PCR machine (Alpha Cycler 4, PCRmax) with 72 °C 5 min, 98 °C 5 min, 8 cycles of [98 °C 10 s, 63 °C 30 s, 72 °C 20 s]. Of the product, 19 μl of pre-amplified library was transferred to a 96-well-qPCR plate, 1 μl 20X EvaGreen (Biotium #31000) was added, and qPCR was performed on an ABI StepOnePlus system with the following cycle conditions: 98 °C 1 min, 20 cycles of [98 °C 10 s, 63 °C 30 s, 72 °C 20 s]. Data were acquired at 72 °C. We qualitatively chose the cycle number to which the fluorescence signals just about to start going up (Supplementary Figure 1b). In this study, a total of 18 cycles were used to amplify the libraries.

Sequencing data processing. All sequencing data were processed using a pipeline written in snakemake. The software/packages and the exact instructions described in this study can be found in the ‘Snakefile’ provided in the GitHub repository https://github.com/dbrg77/plate_scATAC-seq. Briefly, reads were trimmed with cutadapt to remove the Nextera sequence at the 3’ end of short inserts. The trimmed reads were mapped to the reference mouse genome (UCSC mm10) using hisat2. Reads with mapping quality less than 30 were removed by samtools (−q 30 flag) and deduplicated using the MarkDuplicates function of the Picard tool (http://broadinstitute.github.io/picard). All reads from single cells were merged together using samtools, and the merged BAM file was deduplicated again. Peak calling was performed on the merged and deduplicated bam file by MACS2. For bulk ATAC-seq and single cell aggregate coverage visualisation, bedGraph files generated from MACS2 callpeak were converted to bigWig files and visualised using UCSC genome browser. For individual single cell ATAC-seq visualisation, aligned reads from individual cells were converted to bigBed files. A count matrix over the union of peaks was generated by counting the number of reads from individual cells that overlap the union peaks using coverageBed from the bedTools suite.

ARTICLE
and we labelled the peak scATAC-seq data, peak calling was performed on reads merged from all cells.

References

1. Buenrostro, J. D. et al. Epigenetics. Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. Nature 438, 910–914 (2015).

2. Buenrostro, J. D. et al. Single-cell chromatin accessibility reveals principles of regulatory variation. Nature 523, 486–490 (2015).

3. Lake, B. B. et al. Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. Nat. Biotechnol. 36, 70–80 (2018).

4. Buenrostro, J. D. et al. The cis-regulatory dynamics of embryonic development at single-cell resolution. Nature 555, 538–542 (2018).

5. Ryan Corces, M. et al. Lineage-specific and single-cell chromatin accessibility chart human hematopoiesis and leukemia evolution. Nat. Genet. 48, 1193–1203 (2016).

6. Merger, A. et al. High-throughput chromatin accessibility profiling at single-cell resolution. Nat. Commun. 9, 3647 (2018).

7. Chen, X. et al. Joint single-cell DNA accessibility and protein epitope profiling reveals environmental regulation of epigenomic heterogeneity. Preprint at https://doi.org/10.1101/310359 (2018).

8. Preisil, S. et al. Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. Nat. Neurosci. 21, 432–439 (2018).

9. Amini, S. et al. Haplo-type resolved whole-genome sequencing by contigity-preserving transposition and combinatorial indexing. Nat. Genet. 46, 1343–1349 (2014).

10. Picelli, S. et al. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. Genome Res. 24, 2033–2040 (2014).

11. Goryshin, I. Y., Jendrisak, J., Hoffman, L. M., Meis, R. & Reznikow, S. W. Insertional transposon mutagenesis by electroperoration of released Tn5 transposition complexes. Nat. Biotechnol. 18, 97–100 (2000).

12. Corces, M. R. et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat. Methods 14, 939–962 (2017).

13. Goldemberger, D., Harschil, L., Ritzel, M. & Altwegg, M. A single universal DNA extraction procedure using SDS and proteinase K is compatible with direct PCR amplification. PCR Methods Appl. 4, 368–370 (1995).

14. Hagai, T. et al. Gene expression variability across cells and species shapes innate immunity. Preprint at https://doi.org/10.1101/137992 (2017).

15. Jia, G. et al. Single cell RNA-seq and ATAC-seq indicate critical roles of Isl1 and Nkx2-5 for cardiac progenitor cell transition states and lineage settlement. Preprint at https://doi.org/10.1101/210930 (2017).

16. Heng, T. S. P. & Painter, M. W., Immunological Genome Project Consortium. The Immunological Genome Project: networks of gene expression in immune cells. Nat. Immunol. 9, 1091–1094 (2008).

17. Zhong, Y. et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9, R137 (2008).

18. van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. J. Mach. Learn. Res. 9, 2579–2605 (2008).

19. Cesta, M. F. Normal structure, function, and histology of the spleen. Toxicol. Pathol. 34, 455–465 (2006).

20. von Luxburg, U. A tutorial on spectral clustering. Stat. Comput. 17, 395–417 (2007).

21. Bevington, S. L. et al. Inducible chromatin priming is associated with the establishment of immunological memory in T cells. EMBO J. 35, 515–535 (2016).

22. Sapsttein, R. M. et al. Foxp3 exploits a pre-existing enhancer landscape for regulatory T cell lineage specification. Cell 151, 153–166 (2012).

23. Maillard, R. B. et al. IL-18-induced CD83+CCR7+ NK helper cells. J. Exp. Med. 202, 941–953 (2005).

24. Zhou, L. J. & Tedder, T. F. Human blood dendritic cells selectively express CD83, a member of the immunoglobulin superfamily. J. Immunol. 154, 3821–3835 (1995).

25. Hayakawa, Y. & Smyth, M. J. CD27 dissects mature NK cells into two subsets with distinct responsiveness and migratory capacity. J. Immunol. 176, 1517–1524 (2002).

26. Heinz, S. et al. Simple combinations of lineages-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell 38, 576–589 (2010).

27. Smale, S. T. Transcriptional regulation in the immune system: a status report. Trends Immunol. 35, 190–194 (2014).

28. Simonetta, F., Pradier, A. & Roosnek, E. T-bet and eomesoderm in NK cell development, maturation, and function. Front. Immunol. 7, 241 (2016).

29. Inthefskoer, A. M. et al. Effector and memory CD8+ T cell fate coupled by T-bet and eomesoderm. Nat. Immunol. 6, 1236–1244 (2005).

30. Martin, F. & Kearney, J. F. Marginal-zone B cells. Nat. Rev. Immunol. 2, 323–330 (2002).

31. Jin, W. et al. Genome-wide detection of DNase I hypersensitive sites in single cells and FFPE tissue samples. Nature 528, 142–146 (2015).

32. Liu, R. et al. Principles of genome organization revealed by single-cell micrococcal nuclease sequencing. Nature 562, 281–285 (2018).

33. Pott, S. Simultaneous measurement of chromatin accessibility, DNA methylation, and nucleosome phasing in single cells. Proc. Natl. Acad. Sci. U.S.A. 109, 13421–13425 (2012).

34. Pliner, H. A. et al. Cicero predicts cis-regulatory DNA interactions from methylation, and nucleosome phasing in single cells. Proc. Natl. Acad. Sci. U.S.A. 109, 13421–13425 (2012).

35. Clark, S. J. et al. scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. Nat. Commun. 9, 781 (2018).

36. Cusanovich, D. A. et al. A single-cell atlas of in vivo mammalian chromatin accessibility. Cell 174, 1309–1324 (2018).

37. Pliner, H. A. et al. Cicero predicts cis-regulatory DNA interactions from single-cell chromatin accessibility data. Mol. Cell 71, 858–871 (2018).

38. Cao, J. et al. Joint profiling of chromatin accessibility and gene expression in thousands of single cells. Science 361, 1380–1385 (2018).

39. Liu, L. et al. Deconvolution of single-cell multi-omics layers reveals regulatory heterogeneity. Preprint at https://doi.org/10.1101/316208 (2018).

40. Kim, D., Langbead, N. & Salberg, S. L. HISAT: a fast spliced aligner with low memory requirements. Nat. Methods 12, 357–360 (2015).

41. Köster, J. & Rahmann, S. Snakemake—a scalable bioinformatics workflow engine. Bioinformatics 28, 2520–2522 (2012).

42. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBet. J. 17, 10–12 (2011).

43. Li, H. et al. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).

44. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842 (2010).

45. Pedregosa, F. et al. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830 (2011).
inspiring discussion of the method, the critical reading of the manuscript and the computational help. We also thank Natalia Kunowska, Qianxin Wu and Andrew Knights for the helpful discussion related to the Tn5 transposase. We thank Bee Ling Ng, Chris Hall, Jennie Graham and Sam Thompson for the excellent support of FACS. We thank the DNA pipeline from the Wellcome Sanger Institute for the Illumina sequencing support. We thank Guangshuai Jia and Jens Preussner from Thomas Braun’s lab for sharing the mouse cardiac progenitor cells. We thank Aik Ooi for the initial help with the experimental setup. X.C. is funded by the FET-OPEN grant MRG-GRAMMAR 664918, K.N.N. by the Wellcome Trust Strategic Award “Single cell genomics of mouse gastrulation” and S.A.T. by the European Research Council grant ThDEFINE. Wellcome trust core facilities are supported by grant WT206194.

Author contributions
X.C., K.N.N. and S.A.T. conceived the project. X.C. designed the protocol. X.C., R.J.M. and K.N.N. performed the experiments. X.C. carried out the computational analysis. S.A.T. supervised the entire project. All authors contributed to the writing.

Additional information
Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-07771-0.

Competing interests: The authors 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.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2018