Enhancing sensitivity and versatility of Tn5-based single cell omics

Isabelle Seufert[1,2], Pooja Sant[3], Katharina Bauer[3], Afzal P. Syed[3],

Karsten Rippe[1,*], Jan-Philipp Mallm[3,*]

[1] Division of Chromatin Networks, German Cancer Research Center (DKFZ) and BioQuant, Heidelberg, Germany

[2] Faculty of Biosciences, Heidelberg University, Germany

[3] Single Cell Open Lab, German Cancer Research Center (DKFZ), Heidelberg, Germany

* Correspondence: Jan-Philipp Mallm, e-mail: j.mallm@dkfz.de and Karsten Rippe, e-mail: karsten.rippe@dkfz.de

Running title: Enhanced Tn5-based single cell omics

Key words: Gene regulation, single cell ATAC-seq, single cell RNA-seq, single cell multiomics

Abstract

The analysis of chromatin features in single cells centers around Tn5 transposase and exploits its activity to simultaneously fragment target DNA and integrate adapter sequences of choice. This reaction provides a direct readout in the assay for transposase-accessible chromatin in single cells (scATAC-seq) to map open chromatin regions. However, a current limitation is the sparse coverage of open chromatin loci that are detected in a given single cell by droplet-based methods. Thus, enhancing Tn5 activity to improve genomic coverage of scATAC-seq or facilitating multi-omics readouts of chromatin features via Tn5 together with the transcriptome is of great interest. Here, we address these issues by optimizing scATAC-seq for an increased number of integrations per cell. In addition, we provide a protocol that combines mapping of histone modification with scRNA-seq from the same cell by targeting Tn5 to antibody-bound chromatin epitopes. Our experimental workflows improve the results obtained from the downstream data analysis and serve to better resolve epigenetic heterogeneity and transcription regulation in single cells.

Introduction

For the analysis of chromatin features in single cells, Tn5 transposase is a crucial enzyme as it can simultaneously fragment target DNA across the genome and integrate adapter sequences of choice (Adey, 2021). This "cut-and-paste" reaction is referred to as tagmentation and has been initially used for preparation of DNA sequencing library. Usage of hyperactive Tn5 enzymes (Picelli et al., 2014) makes tagmentation an efficient process that can be conducted *in situ*, which is a prerequisite for its application in single cell sequencing (sc-seq). The direct application of tagmentation to analyze chromatin is the assay for transposase-accessible chromatin in single cells (scATAC-seq) (Buenrostro et al., 2015;Klemm et al., 2019). It maps enhancers and promoters that are in an open and *bona fide* active state and thus provides valuable information on transcription regulation mechanism and cellular heterogeneity with respect to cis-regulatory elements (Jiang et al., 2023). Furthermore, Tn5 is used in sc-seq multiomics readouts that yield, for example, scATAC-seq and scRNA-seq data from the same cell, albeit frequently with a lower sensitivity as compared to applying the two readouts in separate experiments (Dimitriu et al., 2022). Other types of single cell multiome analysis (e.g., RNA with CUT&Tag-seq of histone modifications) require specific modifications of the Tn5 mediated tagmentation reaction like using specific adapters loaded on a Protein A-tagged Tn5 in a so-called cut-and-tag reaction (Lahnemann et al., 2020;Lee et al., 2020). For all these applications, improvements to the Tn5 tagmentation reaction would advance the

single cell analysis, which is frequently limited by data sparsity. For example, a typical coverage in scATAC-seq experiments with the Chromium drop-seq platform from 10x Genomics is 7,000 accessible sites per cell out of >100,000 sites that are detected by (pseudo-)bulk ATAC-seq (Li et al., 2021). Thus, the number of open loci detected in a given cell is rather sparse, which limits the resolution of cellular heterogeneity based on the ATAC signal. Aggregation of similar cells is one approach to address this issue (Pliner et al., 2018) but this leads to the loss of information on stochastic differences of chromatin accessibility between individual cells. Likewise, there is also a need to improve the sensitivity for the chromatin readout in multiomics experiments that map the transcriptome in the same cell together with ATAC or another chromatin feature via CUT&Tag (C&T) approaches.

Here, we optimized Tn5 activity to increase the sensitivity of scATAC-seq and described how its adapter loading can be made compatible with existing scATAC and multiome workflows for which the adapter configuration is unknown. We demonstrate that our protocols improve data quality and reduce sparsity without affecting scRNA-seq data quality for a more versatile multiome analysis.

Results

We developed optimized experimental workflows that use differently loaded Tn5 preparations to improve Tn5-based analyses of single cells (**Fig. 1A**). Our work comprised three different reaction types: (i) In a protocol termed scTurboATAC we optimized the detection of open chromatin in scATAC-seq experiments over the standard protocol used with the Chromium platform from 10x Genomics. (ii) For the multiome protocol that combines scRNA-seq and scATAC-seq from the same cell (scMultiome-ATAC) we provide a protocol that uses phosphorylated adapters. (iii) The application for scC&T-seq together with scRNA-seq (scMultiome-C&T) involves Tn5 tagged with Protein A to target the enzyme in situ to the chromatin bound antibody. For this type of sc-seq analysis, we described a multiome analysis where histone H3 lysine 27 trimethylation is mapped alongside gene expression. Subsequent to the Tn5 reaction, the downstream processing followed standard protocols that resulted in efficient library preparation for the three applications described here (**Fig. 1B**).

scTurboATAC increases the number of open sites detected

By varying experimental parameters, we found that the activity of Tn5 is a limiting factor for scATAC-seq. We tested the activity of different concentrations of an in-house Tn5 preparation (Tn5-H) loaded with the adapters listed in **Table 1**. Its activity was compared in a qPCR assay

to Tn5 enzymes from 10x Genomics (Tn5-TXG) as well as the Illumina TDE1 enzyme (Tn5-ILMN) by qPCR (**Fig. S1**).
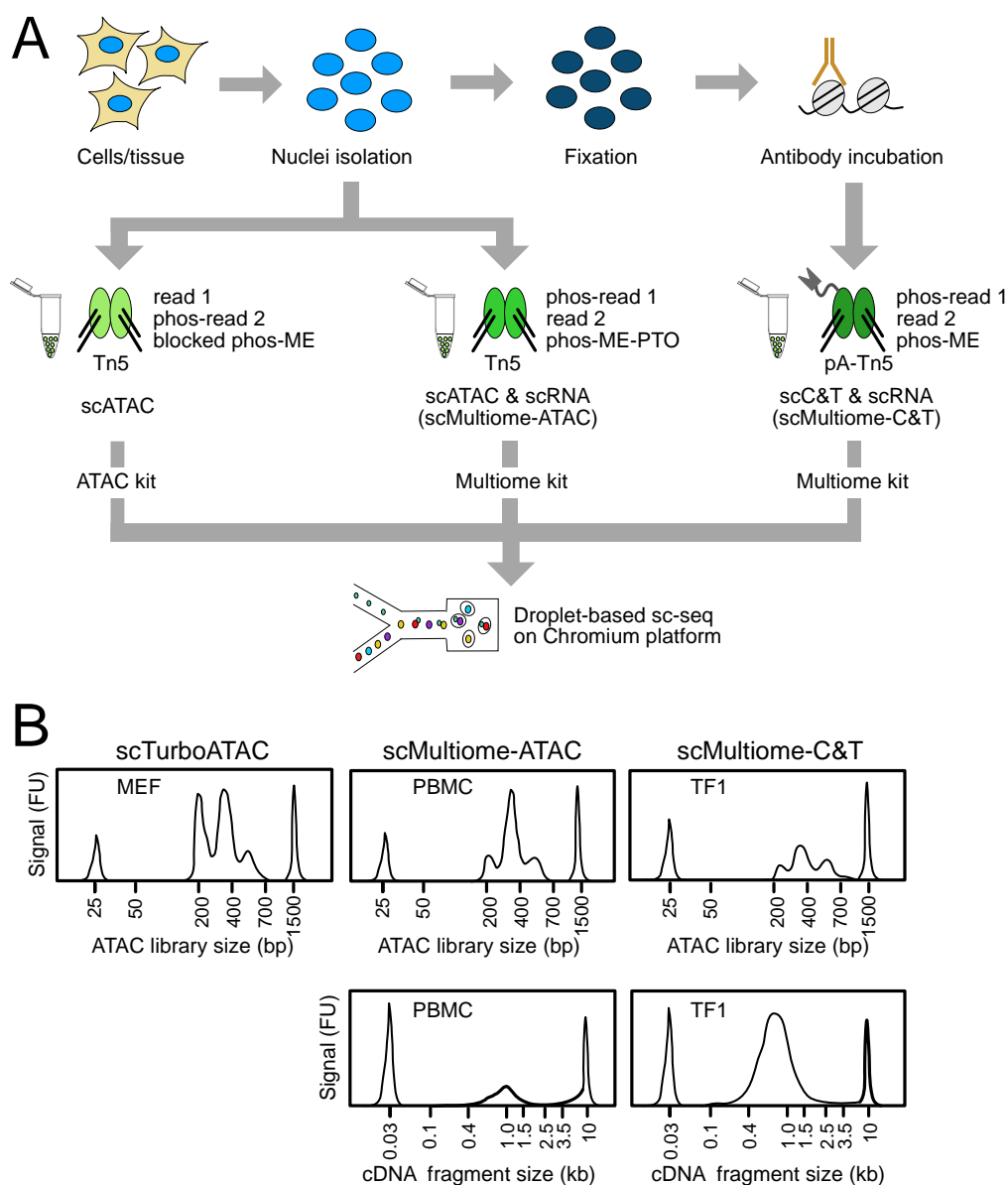

Fig. 1 Experimental workflow. (**A**) Experimental procedure and use of the differently loaded Tn5 preparations. For scATAC and the multiome workflow nuclei were prepared and incubated with Tn5. Importantly, adapters varied between the protocols. For scMultiome-ATAC we used phosphorylated adapters, while the Tn5 for scMultiome-C&T was tagged with Protein A to facilitate binding to the antibody against a given chromatin feature. Following the Tn5 reaction, the downstream processing follows standard protocols. (**B**) Exemplary tapestation profiles are displayed for either the scATAC-seq or scC&T-seq libraries and corresponding cDNA fragments for the multiome protocols. MEFs and TF1 cells were directly used from cell culture while PBMCs were obtained from viably frozen aliquots. All profiles shown were generated with the Tn5-H100 enzyme preparation.

The highest concentration of in-house Tn5 (Tn5-H100) that we used showed a four-fold higher activity per volume in terms of product yield as compared to the Tn5 enzyme provided in the

10x Genomics scATAC-seq version 2 (v2) kits (Tn5-TXGv2) and an about 1.3-fold higher activity per volume than Tn5-TXGv1.1 (**Fig. S1A**). We also tested different buffer compositions and found that the buffer provided in the 10x Genomics kit performed best. The Tn5-ILMN enzyme activity was comparable to Tn5-TXGv1.1 (**Fig. S1A, B**).

We then compared our Tn5-H100 and Tn5-H30 preparations to the Tn5-TXGv1.1 enzyme using mouse embryonic fibroblasts (MEFs) treated with interferon β (IFNβ) for 6 hours. We found that Tn5-H100 yielded the best results with respect to different QC parameters (**Fig. 2A**, **Fig. S2A**): Recovered cell barcodes showed (i) higher TSS enrichment scores, (ii) higher number of unique fragments, and (iii) smaller fragments. In the following, we refer to the use of Tn5-H100 with the buffers provided by 10x Genomics as scTurboATAC vs scATAC for the use with Tn5-TXG.

For further analysis, we selected high-quality cell populations of 5,000 to 6,000 cells for the three samples (**Fig. 2A**, red rectangle). The scTurboATAC protocol yielded an about 2-fold higher number of unique fragments normalized to sequencing depth than cells from the regular scATAC workflow (**Fig. 2B**). Similar to the mouse data, a higher number of unique fragments (absolute and normalized to sequencing depth) was detected in cryopreserved peripheral blood mononuclear cells (PBMC) as a test case when using Tn5-H100 as compared to Tn5-TXGv2 (**Fig. 2C, Fig. S2B**). Embedding and clustering of both data sets revealed and increased resolution of the scTurboATAC data for both MEF (**Fig. 2D**) and PBMC data (**Fig. 2E**). Using marker genes from our previous work (Muckenhuber et al., 2023) two distinct MEF subtypes (epithelial- and mesenchymal-like MEFs) were identified after removing minor clusters 1 and 2 (scATAC) and cluster 1 (scTurboATAC) that showed increased apoptosis module scores (**Fig. S2C**). Cluster 4 of scATAC and cluster 2 of scTurboATAC showed increased mesenchymal-like MEF marker module scores (**Fig. 2D, S3D**). The epithelial-like MEF markers were enriched in cluster 3 of scATAC and clusters 3 and 4 of scTurboATAC (**Fig. 2D, S3C**). Thus, the scTurboATAC protocol resolves two epithelial-like clusters, while only one cluster was detected with the standard scATAC protocol. Likewise, the PBMC data set displayed an increased number of clusters (14 instead of 10). These resolution improvements obtained with the scTurboATAC protocol are likely to reflect a higher complexity of the data that arises from the higher number of unique fragments.
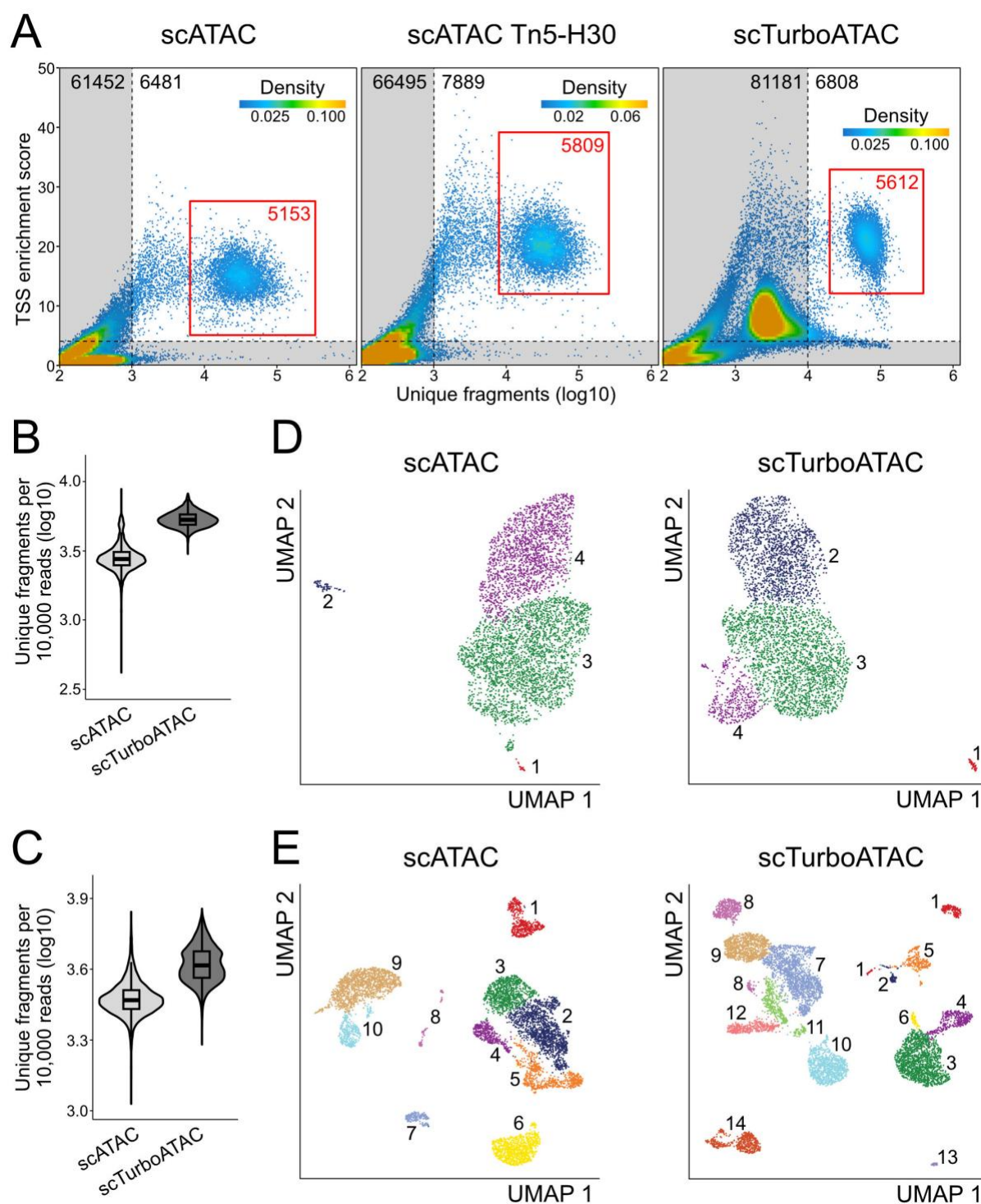
Fig. 2 Quality assessment of scTurboATAC data. (**A**) TSS enrichment score over log10 number of unique fragments of cell barcodes in scATAC (left), scATAC with Tn5-H30 (middle) and scTurboATAC (right) of 6h IFNβ-stimulated MEFs. Colors of points reflect density of cell barcodes. Grey area marks low-quality cell barcodes. Red rectangle marks selected high-quality cells. (**B**) Number of unique fragments per 10,000 raw reads (log10) in scATAC and scTurboATAC of 6h IFNβ-stimulated MEFs. (**C**) Same as panel B for scATAC and scTurboATAC of PBMCs. (**D**) UMAP embedding of 6h IFNβ-stimulated MEFs from scATAC (left) and scTurboATAC (right). Each point represents one cell and is colored according to k-nearest neighbor cluster. (**E**) Same as panel D for PBMCs.

scTurboATAC improves the analysis of the transcription regulation

scTurboATAC was further evaluated for the analysis of transcription factor (TF) binding and gene regulatory mechanisms in MEFs stimulated with IFNβ for 6 hours. Calling peaks from the pseudo-bulk data yielded a higher number of accessible peaks per cell with the scTurboATAC protocol (**Fig. S3A, Fig. S3B**).
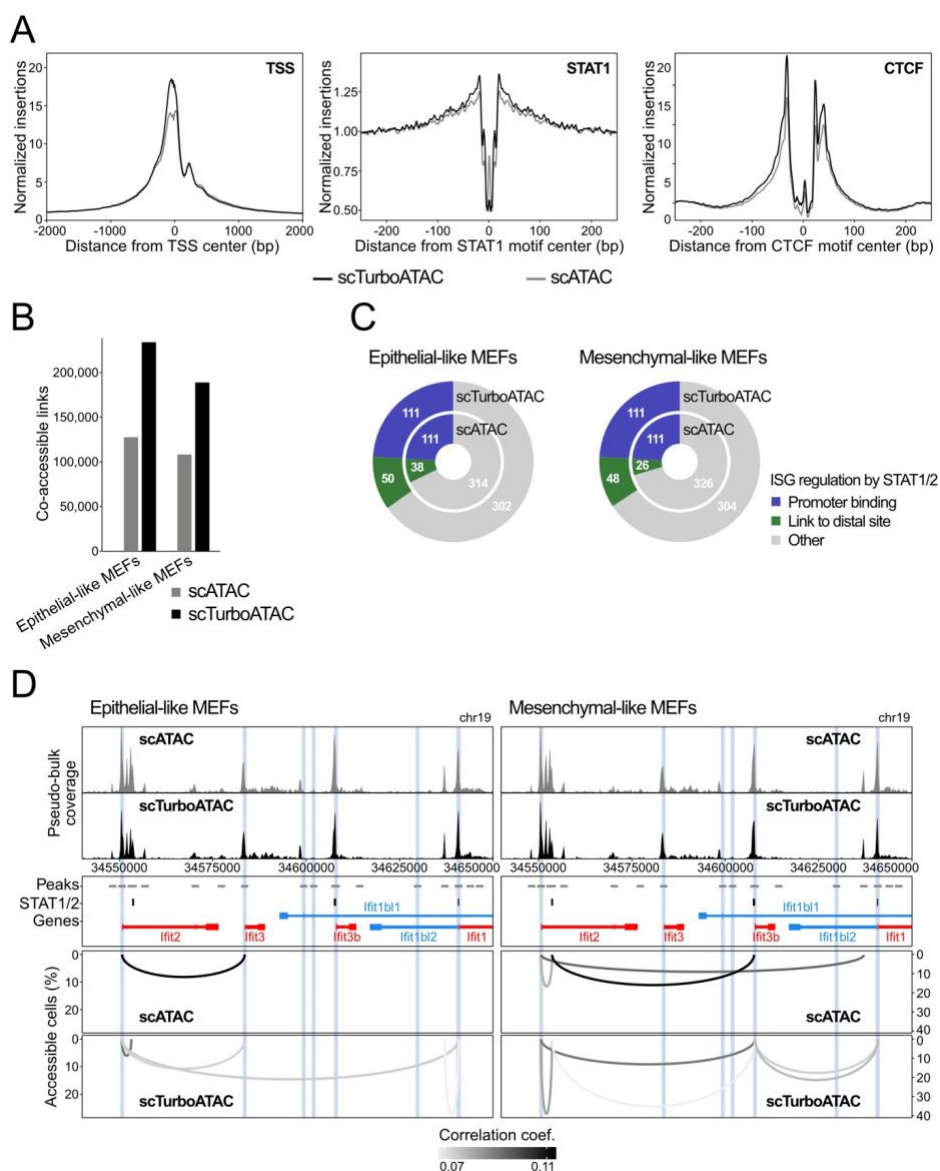

Fig. 3 Comparison of scATAC and scTurboATAC. MEFs stimulated with IFNβ for 6 hours were analyzed. (**A**) Accessibility footprints at TSSs (left), STAT1 motifs (middle) and CTCF motifs (right) from scATAC and scTurboATAC. (**B**) Number of co-accessible links in scATAC and scTurboATAC of epithelial- and mesenchymal-like MEFs. (**C**) ISG regulation mechanisms according to STAT1/2 binding after IFNβ treatment in epithelial- (left) and mesenchymal-like MEFs (right). Promoter bound STAT1/2 (independent of the presence of additional links to distal STAT1/2 sites), blue; co-accessible link to a distal STAT1/2 peak, green; other ISGs, grey. (**D**) Co-accessibility in epithelial- (left) and mesenchymal-like MEFs (right) at the Ifit ISG cluster. Top: browser tracks of aggregated pseudo-bulk chromatin accessibility from single cells of scATAC and scTurboATAC. Middle: 2 kb regions around peaks from pseudo-bulk chromatin accessibility (grey); STAT1/2 peaks from bulk ChIP-seq (black); gene annotation (red, blue). 1 kb regions around ISG TSSs from bulk RNA-seq are marked in blue. Bottom: co-accessible links from scATAC and scTurboATAC.

Furthermore, accessibility footprints at transcription start sites (TSSs) and at STAT1 and CTCF binding motifs showed higher enrichment in pseudo-bulk scTurboATAC than scATAC (**Fig. 3A**).

Next, we applied our approach for co-accessibility analysis to MEF subtypes in scATAC and scTurboATAC using the RWire software as described previously (Muckenhuber et al., 2023). RWire reveals enhancer-promoter links from correlation between open sites of scATAC-seq data. For epithelial-like MEFs of scTurboATAC, only cluster 3 was considered. We observed twice as many co-accessible links above background co-accessibility and minimal percent accessible cells in epithelial- as well as mesenchymal-like MEFs of scTurboATAC than scATAC (**Fig. 3B**). Additionally, co-accessible links from scTurboATAC showed higher correlation coefficients and percent accessible cells than co-accessible links from scATAC (**Fig. S3E**). In our previous study, we observed that the transcription activation of ISGs is partly regulated by simultaneous STAT1 and STAT2 (STAT1/2) binding to ISG promoters as well as distal regulatory sites (Muckenhuber et al., 2023). We were able to link roughly 10 % of ISGs without STAT1/2 promoter binding to a distal STAT1/2 binding event by co-accessibility analysis of 6h IFNβ-stimulated epithelial- as well as mesenchymal-like MEFs from scTurboATAC (**Fig. 3C**). In contrast, only 5-8 % of ISGs were linked to a distal STAT1/2 binding event by co-accessibility analysis of 6h IFNβ-stimulated epithelial- as well as mesenchymal-like MEFs from scATAC. An example for the improved detection of ISG regulation by distal STAT1/2 binding events with scTurboATAC is shown for the *Ifit* gene cluster in **Fig. 3D**. The 150 kb locus depicted contains six ISGs (*Ifit1, Ifit1bl1, Ifit1bl2, Ifit2, Ifit3,* and *Ifit3b*) that were previously identified in MEFs (Muckenhuber et al., 2023). For *Ifit1, Ifit1bl2* and *Ifit3b* promoters, direct STAT1/2 binding was observed in IFNβ-stimulated MEFs. We detected three co-accessible links between ISG promoters and distal STAT1/2 binding events in epithelial- and mesenchymal-like MEFs of scATAC. All co-accessible links were also detected in epithelial- and mesenchymal-like MEFs of scTurboATAC, where they showed higher confidence by higher percent accessible cells. In addition, six co-accessible links between ISG promoters and STAT1/2 binding events were only detected with scTurboATAC. In mesenchymal-like MEFs of scATAC, we detected one unspecific link between the *Ifit2* promoter and a distal peak without a STAT1/2 binding event or other ISG promoter. This link was not observed in mesenchymal-like MEFs of scTurboATAC.

scTurboATAC enhances the cell type annotation

Next, we compared scTurboATAC and scATAC with the Tn5-TXGv2 enzyme with respect to cell type resolution by clustering and marker gene activity (**Fig. 4A, Tab. S2**). The

scTurboATAC protocol specifically improved the detection of progenitor cells, classical monocytes/dendritic cells and provided a higher resolution of the B cell cluster (**Fig. 4A, B**). We identified a distinct B cell cluster in scTurboATAC (cluster 1) that was not resolved in scATAC data by integrating B cells from scATAC and scTurboATAC (**Fig. S4B**). When investigating the B cell clusters further, we found differential TF activities for PU.1-IRF, OCT2 and ATF3 among others (**Fig. 4B, C, Fig. S4C**). While OCT2 has been observed in late state B cell differentiation (Corcoran et al., 1993), PU.1-IRF plays a role in earlier lymphoid maturation (Scott et al., 1994). This could account for their contrary activity scores. Elevated ATF3 activity was mainly found in cluster 1 of scTurboATAC data, potentially pointing to stress-induced activation of B cells to repress pro-apoptotic genes and cytokine production (Ku and Cheng, 2020).
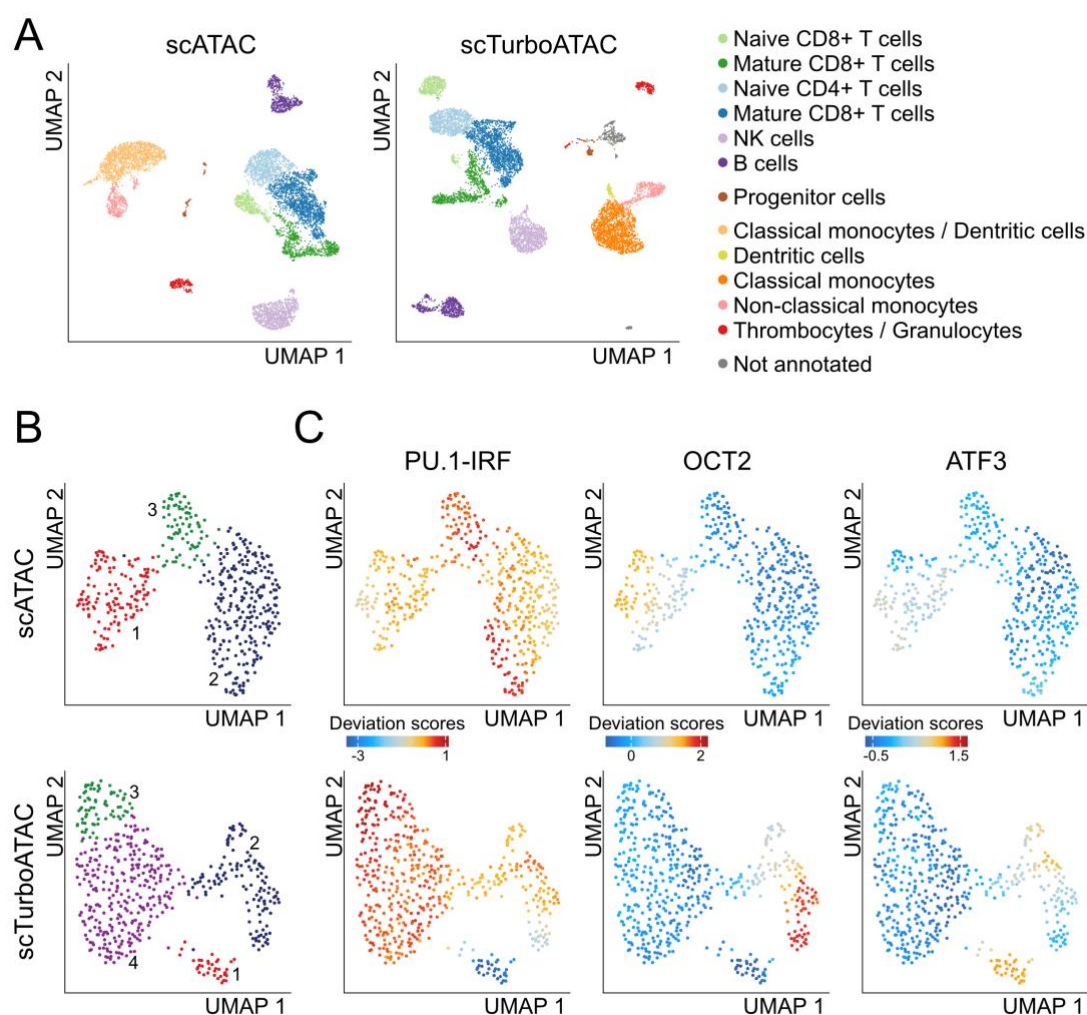

Fig. 4 scTurboATAC resolves additional cell types in PBMCs. (**A**) UMAP embedding of PBMCs from scATAC (left) and scTurboATAC (right). Cells are colored according to annotated cell type by marker gene activity scores. (**B**) UMAP embedding of PBMC B cells from scATAC (top) and scTurboATAC (bottom). Cells are colored according to k-nearest neighbor clusters. (**C**) Same as panel B with cell coloring according to imputed transcription factor motif deviation scores.

scTurboATAC can be integrated into an improved multi-omics analysis

Multi-omics approaches allow to directly link epigenetic heterogeneity with transcriptional profiles. We thus wanted to further exploit scTurboATAC for this type of application and adapted the loading of the transposase to be compatible with the 10x Genomics multiome chemistry. The latter utilizes a spacer oligonucleotide to add cell barcodes to the Tn5 fragments presumably by ligation. To account for this difference, we changed the oligonucleotides for loading (**Table 1**) and additionally used a Protein A-tagged Tn5 for a single-cell C&T-seq assay with RNA information from the same cell (**Fig. 1B**). To avoid self-dimerization for the multiome-ATAC workflow a complementary oligonucleotide with a phosphorothioate (PTO) backbone was used. This modification can be omitted for multiome-C&T due to the Tn5 capture and washing of unbound Tn5. Phosphorylation of Tn5_read1 was found to be essential to facilitate the barcode ligation step in the droplets.

In agreement with the scTurboATAC data, we observed a higher number of unique fragments normalized to sequencing depth when using our adapted Tn5-H100 in comparison to Tn5-TXGv2 with the multiome chemistry in PBMCs (**Fig. 5A left**) while the number of reads in TSSs was similar (**Fig. S5A, B**). UMI counts of transcripts normalized to sequencing depth showed the same yield in comparison to the standard workflow (**Fig. 5A right**). The number of clusters was higher for scMultiome-ATAC with Tn5-H100 (**Fig. 5B**) compared to the standard scMultiome-ATAC with Tn5-TXGv2 (**Fig. 5C**) for both for the ATAC readout alone as well as co-embedded ATAC and RNA data. When using only the gene expression data, the number of clusters remained the same (**Fig. 5B, C**). Thus, the improved ATAC reaction did not compromise the RNA readout in the scMultiome-ATAC protocol.

The scC&T analysis was performed in isogenic TF1 wildtype and mutant cell lines that carry the IDH2 R140Q mutation (Stein, 2016) with histone H3 trimethylation at lysine 27 (H3K27me3) as C&T target. Data from both samples show a similar yield in the number of H3K27me3 fragments (**Fig. 5D left**) with a bias for high-nucleosomal fragments (**Fig. S5C**), a high fraction of reads in peaks (**Fig. S5D**) and comparable UMI transcript counts (**Fig. 5D right**). Very broad regions of H3K27me3 enrichment were observed as depicted for an exemplary region of the HOXC gene cluster in **Fig. 5E left**. In both cell lines all genes within this H3K27me3-rich region were silenced while adjacent genes lacking H3K27me3 signal were highly expressed. UMAP of C&T and RNA data from both cell lines showed separate embedding with some overlap according to IDH2 mutation status (**Fig. 5E right**). Taken together, these data illustrate the versatility and specificity of the scTurboATAC and scMultiome-C&T techniques, which open up new ways for data analysis and experimental readouts.
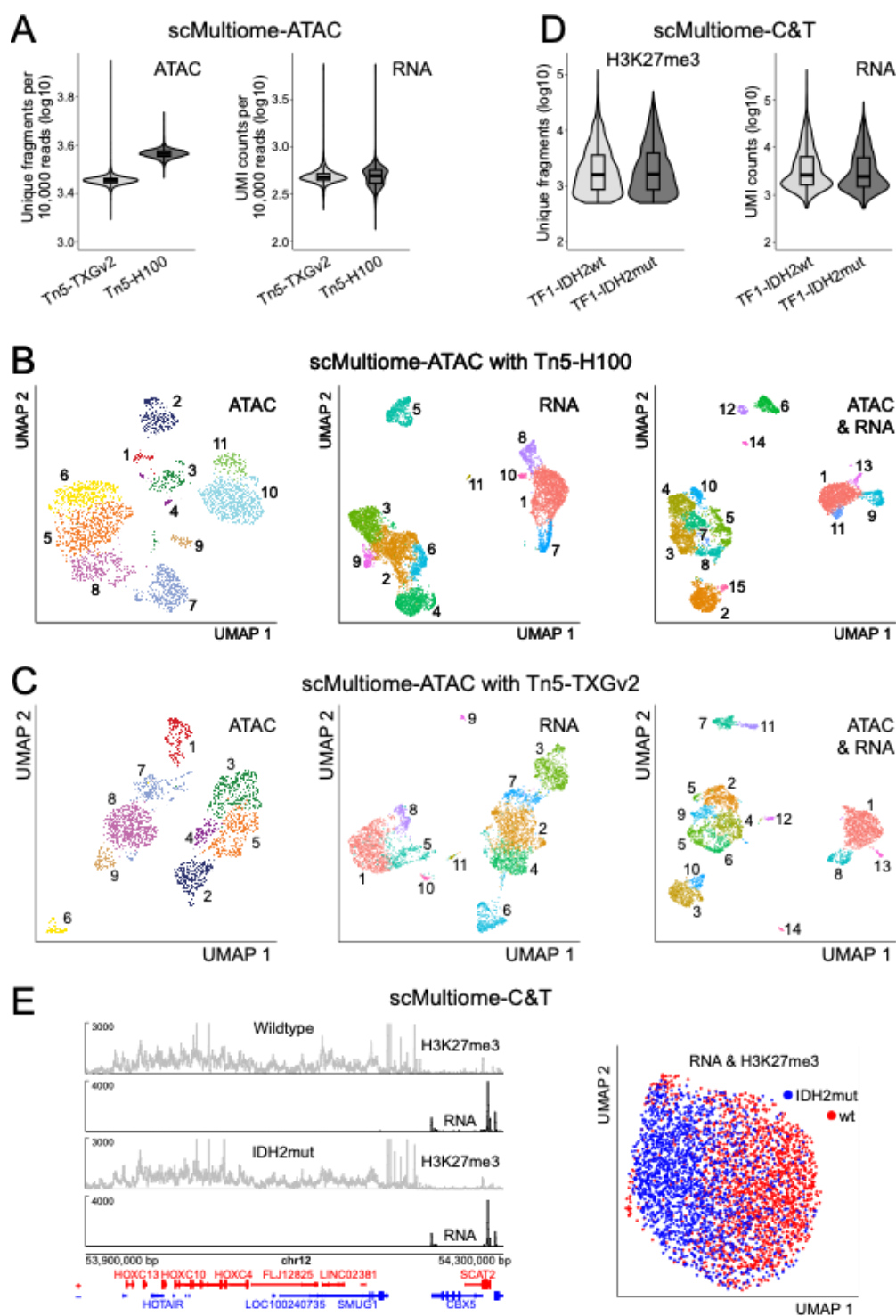
Fig. 5 Multiome analysis. (**A**) scMultiome-ATAC analysis of PBMCs. The log10 number of unique fragments per 10,000 raw reads (ATAC, left) and UMI counts per 10,000 raw reads (RNA, right) is shown. (**B**) UMAP embedding of PBMCs from scMultiome-ATAC with Tn5-H100 based on ATAC (left), RNA (middle) and combined ATAC and RNA (right) readouts. Cells are colored according to k-nearest neighbor clusters. (**C**) Same as panel B but for the scMultiome-ATAC standard protocol with Tn5-TXGv2. (**D**) Analysis of TF1 wildtype and IDH2 mutated cell lines by scMultiome-C&T. The log10 number of unique fragments for H3K27me3 (left) and of UMI counts for RNA (right) are plotted. (**E**) scMultiome-C&T of TF1 wildtype and IDH2 mutated cells. Left, browser tracks of aggregated pseudo-bulk H3K27me3 (grey) and RNA (black) for an exemplary region with the genes on the plus (red) and minus (blue) strand shown at the bottom. Right, UMAP co-embedding of H3K27me3 and RNA. Cells are colored according to sample.

Discussion

Single-cell sequencing applications are frequently limited by data sparsity, which has been addressed both experimentally and bioinformatically (Lahnemann et al., 2020;Bouland et al., 2023;Sant et al., 2023). One strategy to counteract the sparsity for scATAC-seq data is to aggregate multiple similar cells (Pliner et al., 2018;Chen et al., 2019). However, this leads to the loss of information on stochastic differences between individual cells. The scTurboATAC protocol introduced here improved the experimental data acquisition and resulted in a higher library complexity with a consistently lower number of duplicates and higher number of integration events per cell (**Table S1**). These enhancements led to better results in the downstream analysis as shown for cell clustering and computation of TF motif footprints and activities. Furthermore, scTurboATAC improved the co-accessibility analysis to detect interactions between cis-regulatory elements, which was conducted here without aggregating cells to capture stochastic changes of chromatin accessibility.

Plate-based scATAC methods, similar to RNA analysis, could perform better than the droplet-based method used here (Xu et al., 2021). The droplet-based approaches, however, allow processing of several samples with thousands of cells each, which is needed to resolve cell types in complex tissues. Their throughput can be further improved by multiplexing samples combined with overloading the chip (Han et al., 2022). Thus, it will be interesting to compare the sensitivity of our scTurboATAC protocol to plated-based scATAC-seq methods and to evaluate its integration into multiplexing approaches.

As stated above, the relationship between enhancer activity, transcription factor binding and gene expression is of general interest (Fulco et al., 2019;Riegel et al., 2023;Uyehara and Apostolou, 2023). Techniques exist that allow the integration of separately acquired data types from the same sample. However, it is difficult to separate biological differences from confounding technical biases that need to be removed via the integration. Multiome readouts overcome this problem but frequently provide lower sensitivity (Dimitriu et al., 2022). Thus, improving sensitivity for each individual readout is crucial. To this end, we here also adapted and optimized loading of Tn5 to make it compatible with existing multiome workflows for which the adapter configuration is unknown. Furthermore, we demonstrate that our scMultiome-ATAC protocol improved data quality and reduced sparsity without affecting scRNA-seq data quality.

In comparison to the scMultiome-ATAC analysis, the combination of scRNA with scC&T-seq is more difficult to implement (Lahnemann et al., 2020;Lee et al., 2020), especially for repressive or promoter distal histone marks like H3K9me3 and H3K27me3. To address this point, we here also devised a scMultiome-C&T protocol with specific adapters loaded on a Protein A-tagged Tn5 that can be easily applied by other laboratories. A similar workflow has

been described before but as a single readout without transcriptome information (Kaya-Okur et al., 2019). Here, we used our scMultiome-C&T method to analyze the repressive H3K27me3 histone modification together with the RNA from the same cell. An enrichment of H3K27me3 and repression of transcription was detected at loci that were previously described as silenced by H3K27me3 (Soshnikova and Duboule, 2009;Noordermeer et al., 2011). We expect that our technique will work well for all C&T-grade antibodies and can thus easily be adapted to other histone modifications of interest.

In summary, the scTurboATAC, scMultiome-ATAC and scMultiome-C&T protocols described here enhance the sensitivity and versatility of a variety of approaches for analyzing chromatin features in single cells. The improved data quality translates directly into better results for various types of downstream analysis. Accordingly, we anticipate that the methodological enhancement we have made will help to better resolve heterogeneity and variation of epigenetic gene regulation.

Methods

Tn5 loading for scTurboATAC

Lyophilized adapter oligonucleotides were resuspended in annealing buffer (50 mM NaCl, 40 mM Tris·HCl, pH 8) at 100 µM concentration. Adapters (**Table 1**) were preannealed (read 1 with the mosaic end (ME) sequence and read 2 with ME) separately on a thermocycler for 2 min at 85 °C followed by cooling down to 20 °C with 0.1 °C per minute. Glycerol was then added to a final concentration of 50% (v/v) and the adapters were stored at -20 °C. Purified unloaded Tn5 was obtained from the EMBL Protein Expression and Purification Core Facility, which was produced and stored as described previously (Hennig et al., 2018). For Tn5 assembly, transposase, annealed primer pairs and dilution buffer (DB, 50 mM Tris·HCl, pH 7.5, 100 mM NaCl, 0.1 mM EDTA, 1 mM DTT, 0.1 % NP-40, 50 % glycerol) were mixed with equal volumes and incubated for 30 min at room temperature. The loaded Tn5 was then diluted to 1:3 in DB yielding the Tn5-H100 preparation and stored at -20 °C.

Comparing Tn5-H activity to Tn5-TXG

Activity was assesed by qPCR measuring the efficiency in generating fragmented lambda DNA. Different Tn5 preparations were used at the same volume with 10 ng of lambda DNA, 2.5 µl tagmentation buffer (20 mM Tris-HCl pH 7.5, 20 mM MgCl2, 50 % v/v DMF) or 10x Genomics tagmentation buffer as indicated. For the reaction, 1.25 µl phage DNA (10 ng), and 1.25 µl Tn5 enzyme were mixed and incubated for 10 min at 55 °C. The reaction was stopped by adding 1.25 µl 0.2 % SDS. The mix was diluted 1:100 in nuclease free water. A 10 µl qPCR mix was used that contained 5 µl 2x KAPA SYBR Fast, 0.2 µl 10 µM i7 adapter, 0.2 µl 10 µM i5 adapter, 1 µl diluted template DNA and 3.6 µl nuclease free water. The PCR program was 72 °C for 5 min, 95°C for 30 s, followed by 25 cycles of 95 °C for 10 s, 60 °C for 30 s and 72 °C for 60 s. Relative activities were calculated from the measured Ct values.

scTurboATAC library preparation

For scTurboATAC, the transposase provided with the ATAC kit from 10x Genomics was replaced with Tn5-H100. Purified nuclei were processed using the standard 10x ATAC protocol according to the user guides for Chromium Next GEM Single Cell ATAC v1.1 or v2 kits (CG000496, 10x Genomics) as indicated. MEFs were cultured and treated as previously described (Muckenhuber et al., 2023). PBMCs were isolated as previously described and stored viably frozen in serum containing DMSO (Mallm et al., 2019). Prior to the experiments they were thawed in a water bath and immediately processed. For transposition, 10,000 nuclei

were used and loaded on to the Chromium Next GEM Chip H (PN-1000161). Libraries were generated using standard reagents. Quality control and molarity calculations of the final libraries were performed with the Qubit 3.0 Fluorometer (Q33216, Invitrogen) and the 4200 tapestation system (Agilent Technologies). For multiome analysis the appropriately loaded Tn5 was used (**Table 1**), while all downstream reactions were performed with standard multiome reagents from 10x Genomics.

scC&T & RNA from the same cell

TF1-IDH2 wildtype and mutant cell lines were acquired form ATCC (Manassas, Virginia, USA, wildtype, CRL-2003; IDH2 mutant, CRL-2003). They were cultured following the providers instructions. Cells were harvested, washed once in PBS and $1x10^6$ cells were fixed with 1 ml 0.2 % PFA for five min at room temperature. Fixation was stopped with 125 mM glycine for 5 min followed by one wash with PBS. Cells were then resuspended in 500 µl ice cold nuclei buffer and incubated for 10 min on ice. Cells were pelleted and resuspended in 500 µl antibody buffer (2 mM EDTA, 0.1 % BSA, 0.04 % digitonin, 20 mM HEPES pH 7.5, 150 mM NaCl and 0.5 mM spermidine).

A 10 µl volume of αH3K27me3 (cell signalling, 9733) and 12.5 µl RNase inhibitor (40 U/µl) were added and incubated over night at 4 °C. Cells were washed with wash buffer (0.04 % digitonin, 20 mM HEPES pH 7.5, 150 mM NaCl and 0.5 mM spermidine), resuspended in 500 µl wash buffer containing 10 µl 1 µg/µl anti-rabbit secondary antibody raised in guinea pig (Sigma-Aldrich, SAB3700889) and 1 U/µl RNase inhibitor. After 30 min at room temperature cells were washed three times with wash buffer with five min incubation time per wash and resuspended in Tn5 buffer (0.01 % digitonin, 20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM spermidine and 1X protease inhibitor). Cells were counted and 500,000 cells were used from this point forward. 10 µl of loaded proteinA-Tn5 was added and incubated at room temperature for 30 min. After three washes with wash buffer, cells were resuspended in tagmentation buffer (10 mM Tris pH 7.5, 10 mM MgCl, 5 mM HEPES, 50 mM NaCl, 25 % DMF, 0.01 % digitonin, 1 U/µl RNase inhibitor), transferred to a PCR tube and incubated for one hour at 37 °C. A volume of 15 µl was then used for GEM formation with the multiome kit from 10x Genomics.

Sequencing

Multiplexed library pools were generated at 10 nM concentrations of each library and sequenced on NovaSeq 6000 system (Illumina, v1.5). Sequencing was paired end with Read1N and Read2N sequenced with 50 cycles each, i7 (index) sequenced with 8 cycles and i5 using 10x Genomics barcodes with 16 cycles for ATAC and 24 cycles for multiome-ATAC

and multiome-C&T. Sequencing depth was at least 25,000 read pairs per nucleus. RNA libraries were sequenced paired end with 28 bp and 90 bp for read 1 and read 2, respectively.

Analysis of scATAC-seq data from MEFs

Data were demultiplexed and aligned with Cell Ranger ATAC count v2.0.0 (10x Genomics) using the provided mouse mm10 reference. Further processing of data was conducted with ArchR (v1.0.3) in R (v4.2.2) (Granja et al., 2021). Non-cell barcodes were removed by applying a minimal threshold to log10 number of fragments (3.8 for conventional scATAC, 3.9 for scATAC with Tn5-H30, and 4.3 for TurboATAC) and TSS enrichment score (5 for conventional ATAC, 12 for scATAC with Tn5-H30 and TurboATAC). This yielded 5153 cells for conventional scATAC, 5809 cells for scATAC with Tn5-H30, and 5612 cells for TurboATAC. Cell doublets were removed using Amulet (scDblFinder v1.12.0) (Thibodeau et al., 2021) and single cells were embedded in two-dimensional space using an accessibility matrix of 500 bp tiles, Iterative LSI (LSI components 2-10 for conventional ATAC, 2-12 for TurboATAC) and UMAP. Cells were clustered by SNN modularity optimization with a resolution of 0.1. Apoptotic cell clusters were identified by marker gene modules of single-cell gene activity scores (*Casp3, Parp1, Bax, Bak1, Bid, Bbc3, Pmaip1, Bcl2, Bcl2l1, Mcl1, Trp53, Cycs, Apaf1, Tnfrsf10b*, and *Fas*; (Galluzzi et al., 2018)) and removed for further analyses.

Sample-specific and merged peak sets were obtained from pseudo-bulk accessibility data using MACS2 (Zhang et al., 2008) in ArchR (extendSummits = 1000; reproducibility = 1). Transcription factor binding motifs were annotated in sample-specific peaks with Homer mm10 motifs from chromVARmotifs (v0.2.099) (Schep et al., 2017). Transcription factor motif footprints were computed using ArchR (no Tn5 bias normalization, 5 bp smooth window). MEF subtype clusters were identified as epithelial- and mesenchymal-like MEFs by marker gene modules of single-cell gene activity scores of epithelial- and mesenchymal-like MEF markers from our previous study (Muckenhuber et al., 2023).

Simultaneously accessible regions in the same cell were determined by co-accessibility analysis with RWire (https://github.com/RippeLab/RWire-IFN) based on our previously described RWire approach (Mallm et al., 2019;Muckenhuber et al., 2023). Cell numbers between the samples and subtypes were matched to 1,617 cells. The consensus peak set was used for all samples and subtypes. Pearson correlation coefficients were calculated for peaks within 1 Mb for each sample and subtype without aggregation of single cells. Only co-accessible links with positive correlation above background co-accessibility threshold r (0.088 for epithelial-like MEFs from conventional ATAC, 0.079 for mesenchymal-like MEFs from conventional ATAC, 0.066 for epithelial-like MEFs from TurboATAC, 0.066 for mesenchymal-like MEFs from TurboATAC) were considered. The background co-accessibility threshold was

determined by the 99th co-accessibility percentile from randomly shuffled accessibility values over cells and peaks. Co-accessible links were further filtered by a lower threshold of 5 % accessible cells in the linked peaks. STAT1/2 bound sites and TSSs of interferon-stimulated genes from our previous study (Muckenhuber et al., 2023) at co-accessibly linked peaks were annotated.

Analysis of scATAC-seq data from PBMCs

Data were demultiplexed and aligned with Cell Ranger ATAC count v2.1.0 (10x Genomics) using the provided human GRCh38 reference. Further processing of data was conducted with ArchR (v1.0.3) in R (v4.2.2) (Granja et al., 2021). Non-cell barcodes were removed by applying a minimal threshold to log10 number of fragments (4.1 for conventional ATAC, and 4.4 for TurboATAC) and TSS enrichment score (10 for conventional ATAC, 8 for TurboATAC). This yielded 7658 cells for conventional ATAC, and 8309 cells for TurboATAC. Cell doublets were removed using Amulet (scDblFinder v1.12.0) (Thibodeau et al., 2021) and single cells were embedded in two-dimensional space using an accessibility matrix of 500 bp tiles, Iterative LSI (LSI components 2-21) and UMAP. Cells were clustered by SNN modularity optimization with a resolution of 0.2.

Clusters were assigned to cell types by marker gene modules of single-cell gene activity scores of established markers (**Table S2**). Sample-specific and merged peak sets were obtained from pseudo-bulk accessibility data using MACS2 (Zhang et al., 2008) in ArchR (extendSummits = 1000; reproducibility = 1). Transcription factor binding motifs were annotated in sample-specific peaks with Homer hg38 motifs from chromVARmotifs (v0.2.0) (Schep et al., 2017). A subset of only B cells was embedded in two-dimensional space using an accessibility matrix of 500 bp tiles, Iterative LSI (LSI components 2-8) and UMAP. Cells were clustered by SNN modularity optimization with a resolution of 0.25. B cells from conventional ATAC and TurboATAC were integrated using Harmony (v0.1.1) (Korsunsky et al., 2019). ChromVAR deviations were calculated for 200 bp windows around top 1000 transcription factor binding motifs in sample-specific peak sets by decreasing scores (v1.20.0) (Schep et al., 2017). ChromVAR deviations were imputed using MAGIC (van Dijk et al., 2018) in ArchR.

Analysis of scMultiome-ATAC data from PBMCs

Data were demultiplexed and aligned with Cell Ranger ARC count v2.0.2 (10x Genomics) using the provided human GRCh38 reference. Further processing of data was conducted with ArchR (v1.0.3) (Granja et al., 2021) and Seurat (v4.3.0) (Stuart et al., 2019) in R (v4.2.2). Non-cell barcodes in ATAC were removed by applying a minimal threshold of 3.8 and 1000 to log10

number of fragments and reads in TSS, respectively. This yielded 4563 cells for conventional ATAC, 7405 cells for ATAC Tn5-H50, and 6690 cells for TurboATAC. ATAC cell doublets were removed using Amulet (scDblFinder v1.12.0) (Thibodeau et al., 2021), cells with a blacklist ratio above 0.01 and/or nucleosome ratio above 3 were excluded, and single cells were embedded in two-dimensional space using an accessibility matrix of 500 bp tiles, Iterative LSI (LSI components 2-20) and UMAP. Cells were clustered by SNN modularity optimization with a resolution of 0.3.

Non-cell barcodes in RNA were removed by applying a minimal threshold of 100 and 500 to number of detected genes and number of UMI counts, respectively, as well as a maximal threshold of 40 % to percentage of mitochondrial UMI counts. This yielded 4587 cells for RNA of conventional ATAC, 7916 cells for RNA of ATAC Tn5-H50, and 7101 cells for RNA of TurboATAC. RNA cell doublets were removed using Scrublet with a cutoff of 0.15 (v0.2.3) (Wolock et al., 2019) in Python v3.10.4 (pandas v1.4.3, scipy v1.8.1) and single cells were embedded in two-dimensional space using SCTransform (v0.3.5) (Hafemeister and Satija, 2019), PCA (PCs 1-20) and UMAP. Cells were clustered by SNN modularity optimization with a resolution of 0.5. Co-embedding of single cells using ATAC and RNA information was conducted using Signac (v1.9.0) (Stuart et al., 2021). Combining filtered RNA and ATAC cell information yielded 3620 cells for RNA/ATAC of conventional ATAC, 5999 cells for RNA/ATAC of ATAC Tn5-H50, and 5077 cells for RNA/ATAC of TurboATAC. Single cells were embedded using LSI (LSI components 2-25) for ATAC, SCTransform and PCA (PCs 1-30) for RNA, weighted nearest neighbor graph and UMAP. Cells were clustered by SNN modularity optimization with a resolution of 0.8.

Analysis of scMultiome-C&T (H3K27me3) data from TF1 cell lines

Data were demultiplexed and aligned with Cell Ranger ARC count v2.1.0 (10x Genomics) using the provided human GRCh38 reference. Further processing of data was conducted with ArchR (v1.0.3) (Granja et al., 2021) and Seurat (v4.3.0) (Stuart et al., 2019) in R (v4.2.2). Browser tracks of pseudo-bulk RNA and H3K27me3 signal were generated with plotgardener (v1.4.1) (Kramer et al., 2022). For H3K27me3 peak calling, MACS2 was used with the broad setting and keeping one read duplicate resulting in 57,533 peaks that were used in downstream analysis with ArchR. Non-cell barcodes in H3K27me3 were removed by applying a minimal threshold of 500 and 0.1 to number of fragments and fraction of reads in peaks, respectively. This yielded 2786 cells for H3K27me3 of TF1-IDH2wt, and 2653 cells for H3K27me3 of TF1-IDH2mut. Non-cell barcodes in RNA were removed by applying a minimal threshold of 100 and 500 to number of detected genes and number of UMI counts, respectively, as well as a maximal threshold of 25 % to percentage of mitochondrial UMI

counts. This yielded 2928 cells for RNA of TF1-IDH2wt, and 2956 cells for RNA of TF1-IDH2mut. Co-embedding of single cells using H3K27me3 and RNA information was conducted using Signac (v1.9.0) (Stuart et al., 2021). Combining filtered RNA and H3K27me3 cell information yielded 1803 cells for RNA/H3K27me3 of TF1-IDH2wt, and 1800 cells for RNA/H3K27me3 of TF1-IDH2mut. Single cells were embedded using LSI (LSI components 3-20) for H3K27me3, SCTransform (v0.3.5) (Hafemeister and Satija, 2019) and PCA (PCs 1-20) for RNA, weighted nearest neighbor graph and UMAP.

Data availability

Single cell sequencing data are available via GEO accession number GSE235506.

Acknowledgments

References

Adey, A.C. (2021). Tagmentation-based single-cell genomics. *Genome Res* 31**,** 1693-1705.

Bouland, G.A., Mahfouz, A., and Reinders, M.J.T. (2023). Consequences and opportunities arising due to sparser single-cell RNA-seq datasets. *Genome Biol* 24**,** 86.

Buenrostro, J.D., Wu, B., Litzenburger, U.M., Ruff, D., Gonzales, M.L., Snyder, M.P., Chang, H.Y., and Greenleaf, W.J. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* 523**,** 486-490.

Chen, H., Lareau, C., Andreani, T., Vinyard, M.E., Garcia, S.P., Clement, K., Andrade-Navarro, M.A., Buenrostro, J.D., and Pinello, L. (2019). Assessment of computational methods for the analysis of single-cell ATAC-seq data. *Genome Biol* 20**,** 241.

Corcoran, L.M., Karvelas, M., Nossal, G.J., Ye, Z.S., Jacks, T., and Baltimore, D. (1993). Oct-2, although not required for early B-cell development, is critical for later B-cell maturation and for postnatal survival. *Genes Dev* 7**,** 570-582.

Dimitriu, M.A., Lazar-Contes, I., Roszkowski, M., and Mansuy, I.M. (2022). Single-Cell Multiomics Techniques: From Conception to Applications. *Front Cell Dev Biol* 10**,** 854317.

Fulco, C.P., Nasser, J., Jones, T.R., Munson, G., Bergman, D.T., Subramanian, V., Grossman, S.R., Anyoha, R., Doughty, B.R., Patwardhan, T.A., Nguyen, T.H., Kane, M., Perez, E.M., Durand, N.C., Lareau, C.A., Stamenova, E.K., Aiden, E.L., Lander, E.S., and Engreitz, J.M. (2019). Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat Genet* 51**,** 1664-1669.

Galluzzi, L., Vitale, I., Aaronson, S.A., Abrams, J.M., Adam, D., Agostinis, P., Alnemri, E.S., Altucci, L., Amelio, I., Andrews, D.W., Annicchiarico-Petruzzelli, M., Antonov, A.V., Arama, E., Baehrecke, E.H., Barlev, N.A., Bazan, N.G., Bernassola, F., Bertrand, M.J.M., Bianchi, K., Blagosklonny, M.V., Blomgren, K., Borner, C., Boya, P., Brenner, C., Campanella, M., Candi, E., Carmona-Gutierrez, D., Cecconi, F., Chan, F.K., Chandel, N.S., Cheng, E.H., Chipuk, J.E., Cidlowski, J.A., Ciechanover, A., Cohen, G.M., Conrad, M., Cubillos-Ruiz, J.R., Czabotar, P.E., D'angiolella, V., Dawson, T.M., Dawson, V.L., De Laurenzi, V., De Maria, R., Debatin, K.M., Deberardinis, R.J., Deshmukh, M., Di Daniele, N., Di Virgilio, F., Dixit, V.M., Dixon, S.J., Duckett, C.S., Dynlacht, B.D., El-Deiry, W.S., Elrod, J.W., Fimia, G.M., Fulda, S., Garcia-Saez, A.J., Garg, A.D., Garrido, C., Gavathiotis, E., Golstein, P., Gottlieb, E., Green, D.R., Greene, L.A., Gronemeyer, H., Gross, A., Hajnoczky, G., Hardwick, J.M., Harris, I.S., Hengartner, M.O., Hetz, C., Ichijo, H., Jaattela, M., Joseph, B., Jost, P.J., Juin, P.P., Kaiser, W.J., Karin, M., Kaufmann, T., Kepp, O., Kimchi, A., Kitsis, R.N., Klionsky, D.J., Knight, R.A., Kumar, S., Lee, S.W., Lemasters, J.J., Levine, B., Linkermann, A., Lipton, S.A., Lockshin, R.A., Lopez-Otin, C., Lowe, S.W., Luedde, T., Lugli, E., Macfarlane, M., Madeo, F., Malewicz, M., Malorni, W., Manic, G., et al. (2018). Molecular mechanisms of cell death: recommendations of the Nomenclature Committee on Cell Death 2018. *Cell Death Differ* 25**,** 486-541.

Granja, J.M., Corces, M.R., Pierce, S.E., Bagdatli, S.T., Choudhry, H., Chang, H.Y., and Greenleaf, W.J. (2021). ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nature Genetics* 53**,** 403-411.

Hafemeister, C., and Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol* 20**,** 296.

Han, S., Kim, K., Park, S., Lee, A.J., Chun, H., and Jung, I. (2022). scAVENGERS: a genotype-based deconvolution of individuals in multiplexed single-cell ATAC-seq data without reference genotypes. *NAR Genom Bioinform* 4**,** lqac095.

Hennig, B.P., Velten, L., Racke, I., Tu, C.S., Thoms, M., Rybin, V., Besir, H., Remans, K., and Steinmetz, L.M. (2018). Large-Scale Low-Cost NGS Library Preparation Using a Robust Tn5 Purification and Tagmentation Protocol. *G3 (Bethesda)* 8**,** 79-89.

Jiang, S., Huang, Z., Li, Y., Yu, C., Yu, H., Ke, Y., Jiang, L., and Liu, J. (2023). Single-cell chromatin accessibility and transcriptome atlas of mouse embryos. *Cell Rep* 42**,** 112210.

Kaya-Okur, H.S., Wu, S.J., Codomo, C.A., Pledger, E.S., Bryson, T.D., Henikoff, J.G., Ahmad, K., and Henikoff, S. (2019). CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun* 10**,** 1930.

Klemm, S.L., Shipony, Z., and Greenleaf, W.J. (2019). Chromatin accessibility and the regulatory epigenome. *Nat Rev Genet* 20**,** 207-220.

Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.R., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* 16**,** 1289-1296.

Kramer, N.E., Davis, E.S., Wenger, C.D., Deoudes, E.M., Parker, S.M., Love, M.I., and Phanstiel, D.H. (2022). Plotgardener: cultivating precise multi-panel figures in R. *Bioinformatics* 38**,** 2042-2045.

Ku, H.C., and Cheng, C.F. (2020). Master Regulator Activating Transcription Factor 3 (ATF3) in Metabolic Homeostasis and Cancer. *Front Endocrinol (Lausanne)* 11**,** 556.

Lahnemann, D., Koster, J., Szczurek, E., Mccarthy, D.J., Hicks, S.C., Robinson, M.D., Vallejos, C.A., Campbell, K.R., Beerenwinkel, N., Mahfouz, A., Pinello, L., Skums, P., Stamatakis, A., Attolini, C.S., Aparicio, S., Baaijens, J., Balvert, M., Barbanson, B., Cappuccio, A., Corleone, G., Dutilh, B.E., Florescu, M., Guryev, V., Holmer, R., Jahn, K., Lobo, T.J., Keizer, E.M., Khatri, I., Kielbasa, S.M., Korbel, J.O., Kozlov, A.M., Kuo, T.H., Lelieveldt, B.P.F., Mandoiu, Ii, Marioni, J.C., Marschall, T., Molder, F., Niknejad, A., Raczkowski, L., Reinders, M., Ridder, J., Saliba, A.E., Somarakis, A., Stegle, O., Theis, F.J., Yang, H., Zelikovsky, A., Mchardy, A.C., Raphael, B.J., Shah, S.P., and Schonhuth, A. (2020). Eleven grand challenges in single-cell data science. *Genome Biol* 21**,** 31.

Lee, J., Hyeon, D.Y., and Hwang, D. (2020). Single-cell multiomics: technologies and data analysis methods. *Exp Mol Med* 52**,** 1428-1442.

Li, Z., Kuppe, C., Ziegler, S., Cheng, M., Kabgani, N., Menzel, S., Zenke, M., Kramann, R., and Costa, I.G. (2021). Chromatin-accessibility estimation from single-cell ATAC-seq data with scOpen. *Nat Commun* 12**,** 6386.

Mallm, J.P., Iskar, M., Ishaque, N., Klett, L.C., Kugler, S.J., Muino, J.M., Teif, V.B., Poos, A.M., Grossmann, S., Erdel, F., Tavernari, D., Koser, S.D., Schumacher, S., Brors, B., Konig, R., Remondini, D., Vingron, M., Stilgenbauer, S., Lichter, P., Zapatka, M., Mertens, D.,

and Rippe, K. (2019). Linking aberrant chromatin features in chronic lymphocytic leukemia to transcription factor networks. *Mol Syst Biol* 15, e8339.

Muckenhuber, M., Seufert, I., Muller-Ott, K., Mallm, J.P., Klett, L.C., Knotz, C., Hechler, J., Kepper, N., Erdel, F., and Rippe, K. (2023). Epigenetic signals that direct cell type-specific interferon beta response in mouse cells. *Life Sci Alliance* 6.

Noordermeer, D., Leleu, M., Splinter, E., Rougemont, J., De Laat, W., and Duboule, D. (2011). The dynamic architecture of Hox gene clusters. *Science* 334, 222-225.

Picelli, S., Bjorklund, A.K., Reinius, B., Sagasser, S., Winberg, G., and Sandberg, R. (2014). Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res* 24, 2033-2040.

Pliner, H.A., Packer, J.S., Mcfaline-Figueroa, J.L., Cusanovich, D.A., Daza, R.M., Aghamirzaie, D., Srivatsan, S., Qiu, X., Jackson, D., Minkina, A., Adey, A.C., Steemers, F.J., Shendure, J., and Trapnell, C. (2018). Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility Data. *Mol Cell* 71, 858-871 e858.

Riegel, D., Romero-Fernandez, E., Simon, M., Adenugba, A.R., Singer, K., Mayr, R., Weber, F., Kleemann, M., Imbusch, C.D., Kreutz, M., Brors, B., Ugele, I., Werner, J.M., Siska, P.J., and Schmidl, C. (2023). Integrated single-cell profiling dissects cell-state-specific enhancer landscapes of human tumor-infiltrating CD8(+) T cells. *Mol Cell* 83, 622-636 e610.

Sant, P., Rippe, K., and Mallm, J.P. (2023). Approaches for single-cell RNA sequencing across tissues and cell types. *Transcription*, 1-19.

Schep, A.N., Wu, B., Buenrostro, J.D., and Greenleaf, W.J. (2017). chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat Methods* 14, 975-978.

Scott, E.W., Simon, M.C., Anastasi, J., and Singh, H. (1994). Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages. *Science* 265, 1573-1577.

Soshnikova, N., and Duboule, D. (2009). Epigenetic temporal control of mouse Hox genes in vivo. *Science* 324, 1320-1323.

Stein, E.M. (2016). Molecular Pathways: IDH2 Mutations-Co-opting Cellular Metabolism for Malignant Transformation. *Clin Cancer Res* 22, 16-19.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. *Cell* 177, 1888-1902 e1821.

Stuart, T., Srivastava, A., Madad, S., Lareau, C.A., and Satija, R. (2021). Single-cell chromatin state analysis with Signac. *Nat Methods* 18, 1333-1341.

Thibodeau, A., Eroglu, A., Mcginnis, C.S., Lawlor, N., Nehar-Belaid, D., Kursawe, R., Marches, R., Conrad, D.N., Kuchel, G.A., Gartner, Z.J., Banchereau, J., Stitzel, M.L., Cicek, A.E., and Ucar, D. (2021). AMULET: a novel read count-based method for effective multiplet detection from single nucleus ATAC-seq data. *Genome Biol* 22**,** 252.

Uyehara, C.M., and Apostolou, E. (2023). 3D enhancer-promoter interactions and multi-connected hubs: Organizational principles and functional roles. *Cell Rep***,** 112068.

Van Dijk, D., Sharma, R., Nainys, J., Yim, K., Kathail, P., Carr, A.J., Burdziak, C., Moon, K.R., Chaffer, C.L., Pattabiraman, D., Bierie, B., Mazutis, L., Wolf, G., Krishnaswamy, S., and Pe'er, D. (2018). Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. *Cell* 174**,** 716-729 e727.

Wolock, S.L., Lopez, R., and Klein, A.M. (2019). Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst* 8**,** 281-291 e289.

Xu, W., Wen, Y., Liang, Y., Xu, Q., Wang, X., Jin, W., and Chen, X. (2021). A plate-based single-cell ATAC-seq workflow for fast and robust profiling of chromatin accessibility. *Nat Protoc* 16**,** 4084-4107.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9**,** R137.

Table 1. Oligonucleotides used for Tn5 loading

Oligo-nucleotide [a]	Sequence	Assay	Sample
read1	5′-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3'	TurboATAC	MEFs plus IFNβ, PBMCs
phos-read2	5′-[phos]GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3'		
blocked phos-ME	5'-[phos]C*T*G*T*C*T*C*T*T*A*T*A*C*A*[23ddC]-3'		
phos-read1	5'-[phos]TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3'	TurboATAC & RNA	PBMCs
read2	5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3'		
phos-ME_PTO	5'-[phos]C*T*G*T*C*T*C*T*T*A*T*A*C*A*C*A*T*C*T-3'		
phos-read1	5'-[phos]TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3'	C&T & RNA	TF1 wt, TF1 IDHmut2
read2	5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3'		
phos-ME	5'-[phos]CTGTCTCTTATACACATCT-3'		

[a] Oligonucleotides where synthesized with the indicated modifications: phos, 5'-phosphate; PTO, phosphorothioate backbone indicated by the *. ME stands for mosaic end sequence.

Supplementary Information

Enhancing sensitivity and versatility of Tn5-based single cell omics

Isabelle Seufert, Pooja Sant, Katharina Bauer, Afzal P. Syed, Karsten Rippe & Jan-Philipp Mallm

Inventory of Supplementary Information

Supplementary Figures

Supplementary Tables

Supplementary References
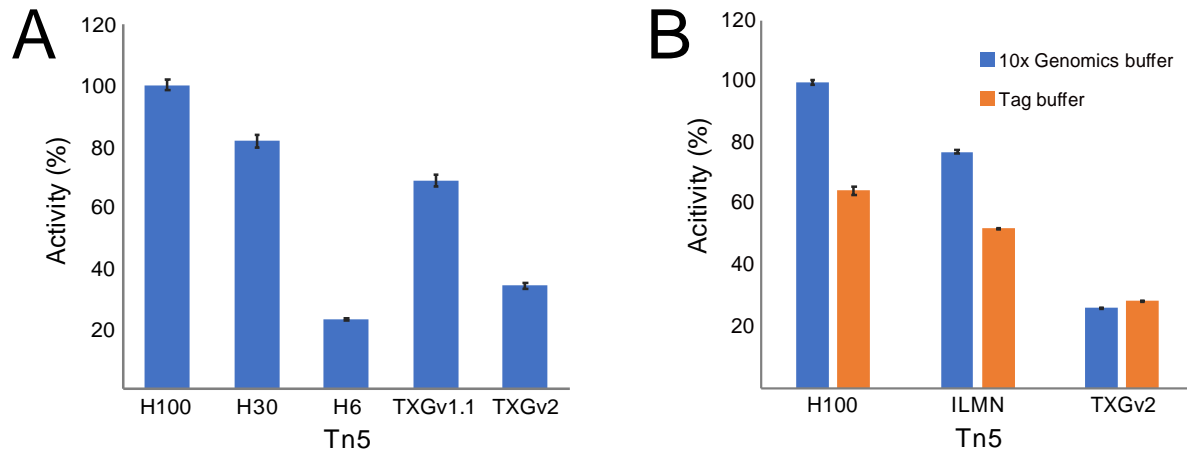
Supplementary Figures

Figure S1. Tn5 activity measurements by qPCR. (**A**) The in-house Tn5-H preparations at three different concentrations (Tn5-H100, Tn5-H30 and Tn5-H6) were tested against Tn5-TXGv1.1 and Tn5-TXGv2 that are provided in the kits from 10x Genomics. Activity was calculated from the yield of fragmented lambda phage DNA measured by qPCR. Tn5-H100 displayed the highest activity while Tn5-TXGv1.1 was more efficient than the later introduced Tn5-TXGv2 enzyme. Error bars display standard deviation from triplicates. (**B**) Comparison of Tn5 activity in the buffer from 10x Genomics against a standard Tn5 reaction buffer (Tag buffer) for the different Tn5 enzyme preparations. The buffer provided in the 10x Genomics kits results in a higher activity for both the Tn5-H100 as well as the Tn5-ILMN (TDE1 enzyme, #20034197, Illumina) preparations. No difference between the two buffers was observed for Tn5-TXGv2. Error bars display standard deviation from triplicates.
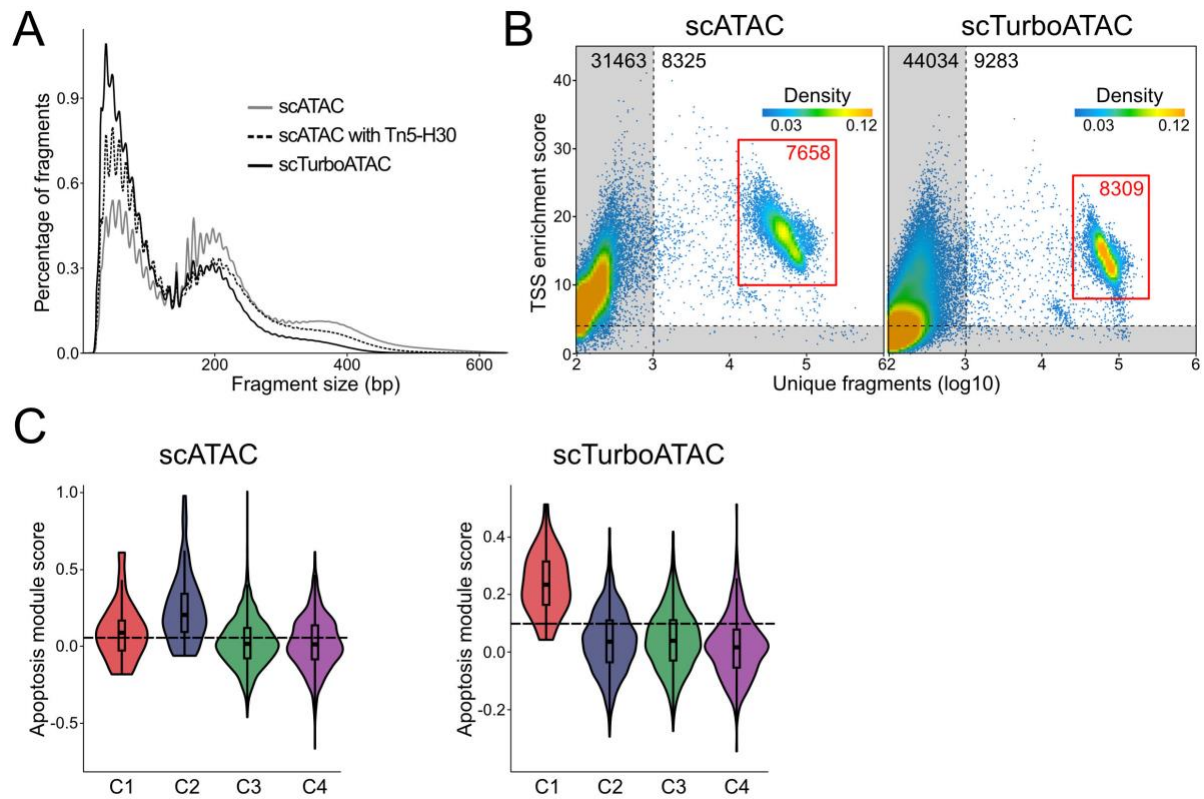
Figure S1. Quality assessment of scTurboATAC data. (**A**) Fragment size distribution in scATAC with Tn5-TXGv1.1 and Tn5-H30 versus scTurboATAC protocols for MEFs stimulated for 6h with IFNβ. (**B**) TSS enrichment score against log10 number of unique fragments of cell barcodes in scATAC with Tn5-TXGv2 (left) and scTurboATAC (right) for PBMCs. Colors of points reflect density of cell barcodes. Grey area marks low-quality cell barcodes. Red rectangle marks selected high-quality cells. (**C**) Apoptosis module score per cell in scATAC with Tn5-TXGv1.1 (left) and scTurboATAC (right) of 6h IFNβ-stimulated MEFs.
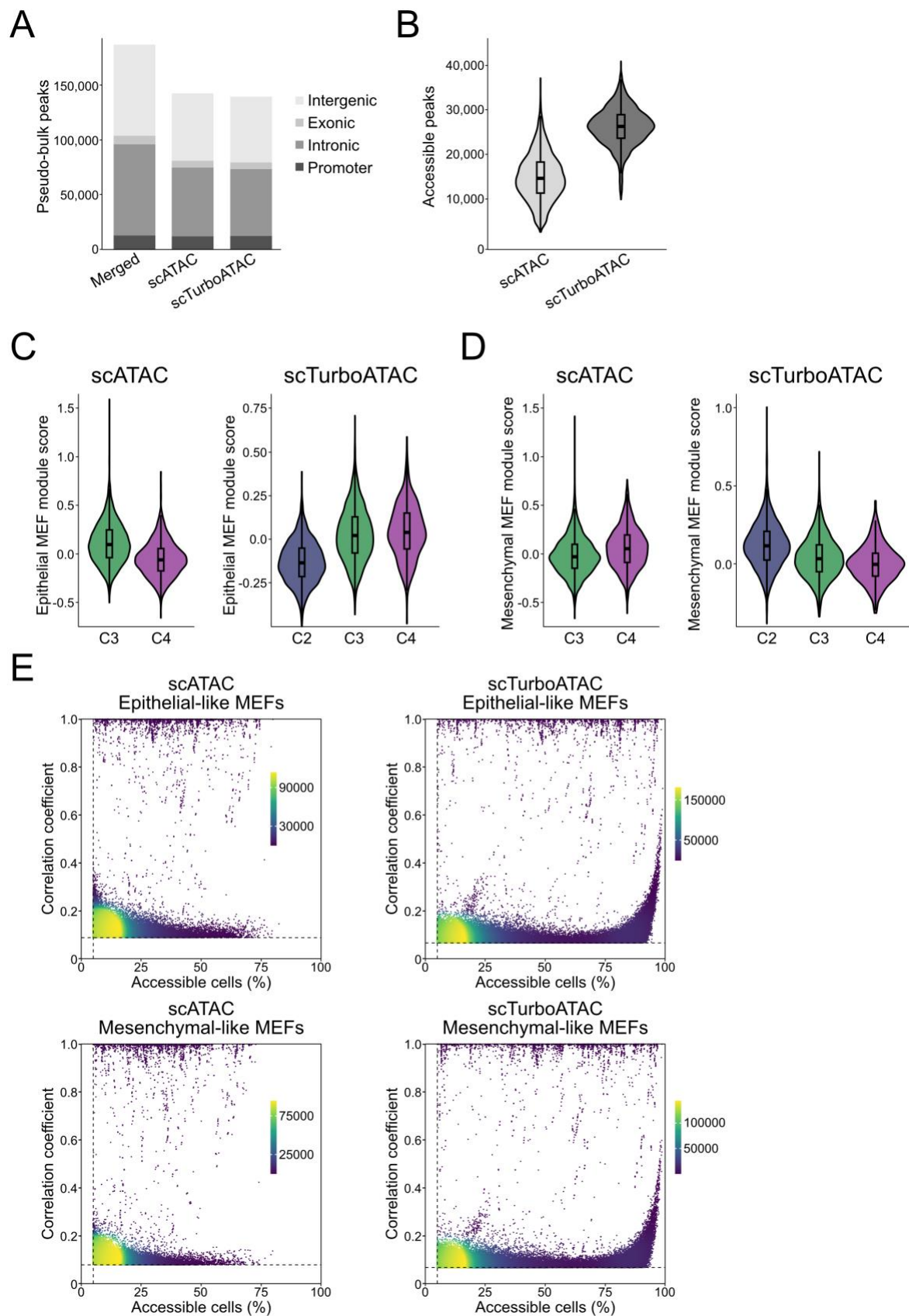
Figure S2. Investigation of transcription regulation with scTurboATAC. MEFs stimulated for 6h with IFNβ were studied. The scATAC data refer to the Tn5-TXGv1.1 enzyme. (**A**)

Number of peaks from pseudo-bulk chromatin accessibility of scATAC, scTurboATAC and the merged peak set, which additionally included data from scATAC with Tn5-H30. (**B**) Number of accessible peaks detected per cell in scATAC and scTurboATAC using the merged peak set. (**C**) Epithelial-like MEF module score per cell in scATAC and scTurboATAC. (**D**) Mesenchymal-like MEF module score per cell in scATAC and scTurboATAC. (**E**) Correlation coefficient and percent accessible cells of co-accessible links from scATAC (left) and scTurboATAC (right). Colors of points reflect density of co-accessible links. Dashed lines represent background co-accessibility cutoff on correlation coefficient and minimal 5 % cutoff on accessible cells.
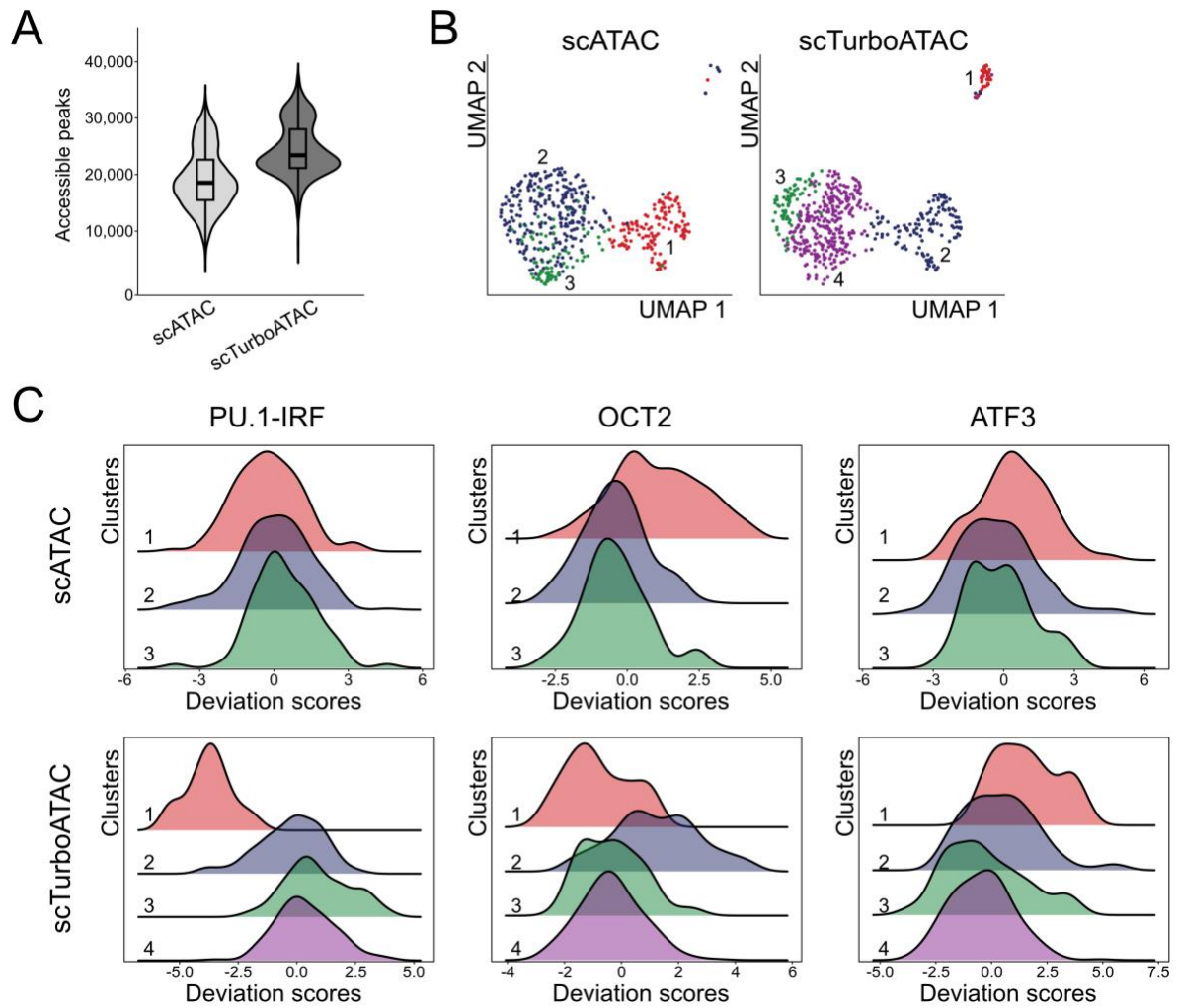
Figure S3. Resolving cell types in PBMCs with scTurboATAC. The scATAC data refer to the Tn5-TXGv2 enzyme. (**A**) Number of accessible peaks per cell in scATAC and scTurboATAC using the merged peak set. (**B**) Integrated UMAP embedding of B cells filtered from PBMC data set for scATAC and scTurboATAC. Integrated UMAP embeddings split by samples scATAC (left) and scTurboATAC (right) are shown. Cells are colored according to sample-specific k-nearest neighbor clusters. (**C**) Transcription factor motif deviation scores (non-imputed) in k-nearest neighbor clusters of scATAC (top) and scTurboATAC (bottom).
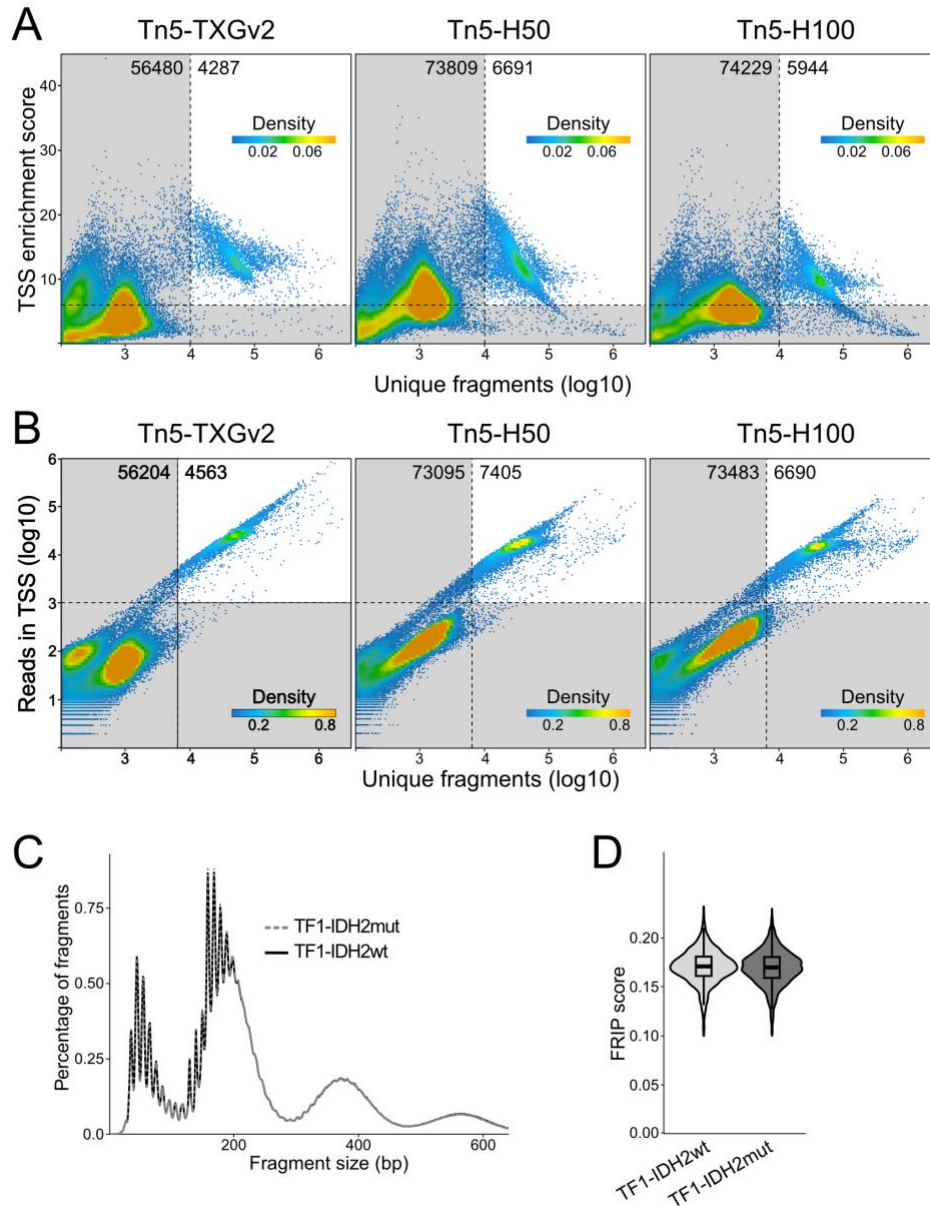
Figure S5. Quality assessment of single cell multiome data. (**A**) TSS enrichment score over log10 number of unique fragments of cell barcodes in scMultiome-ATAC of PBMCs with the indicated Tn5 preparations. Colors of points reflect density of cell barcodes. Grey area marks low-quality cell barcodes. (**B**) Same as panel A for log10 number of reads in TSS over log10 number of unique fragments. (**C**) Fragment size distribution for scMultiome-C&T against H3K27me3 in TF1-IDH2wt and TF1-IDH2mut. (**D**) Fragments of reads in peaks (FRIP) scores for H3K27me3 per cell in TF1-IDH2wt and TF1-IDH2mut from the scMultiome-C&T analysis.

Supplementary Tables

Table S1. Sequencing metrics of scATAC and scTurboATAC analysis

	MEF 6h IFNβ			PBMC	
	scATAC with Tn5-TXGv1.1	scATAC with Tn5-H30	scTurboATAC	scATAC with Tn5-TXGv2	scTurboATAC
Sequenced read pairs	827,356,762	1,221,969,310	1,304,430,940	1,466,934,091	1,468,516,560
Duplicates	70.1 %	75.8 %	45.9 %	66.6 %	52.2 %
Mapped read pairs	92.2 %	88.2 %	87.7 %	94.0 %	90.5 %
Fragments nucleosome-free regions	40.6 %	54.5 %	64.8 %	50.4 %	47.5 %

Table S2. Marker genes of hematopoietic cell types

Cell type	Marker gene	Reference
Naïve CD8+ T cells	CD8A, GZMB, CD3D, CD34, LEF1	(Terstappen et al., 1992; Uhlen et al., 2019)
Mature CD8+ T cells	CD8A, GZMB, CD3D, CD2	(Uhlen et al., 2019)
Naïve CD4+ T cells	CD4, IL7R, CCR7, CD3D, CD34, LEF1	(Terstappen et al., 1992; Uhlen et al., 2019)
Mature CD4+ T cells	CD4,IL7R, S100A4, CD3D, CD2, IL2, TNF, IL21	(Uhlen et al., 2019)
NK cells	GNLY, NKG7, GZMB, NCAM1, FCGR3A	(Uhlen et al., 2019; Karlsson et al., 2021)
B cells	MS4A1, CD79A, CD74, CD19	(Uhlen et al., 2019; Karlsson et al., 2021)
Progenitor cells	CD34, KIT, PROM1, PTPRC (negative marker), CD38 (negative marker)	(Terstappen et al., 1992; Monaco et al., 2019)
Classical monocytes	CD14, LYZ, CST3	(Uhlen et al., 2019)
Non-classical monocytes	FCGR3A, MS4A7	(Uhlen et al., 2019)
Dentritic cells	FCER1A, CST3, CD74	(Uhlen et al., 2019)
Basophils	ENPP3, CD69, CCR3, PTGDR2	(Monaco et al., 2019; Uhlen et al., 2019)
Neutrophils	CEACAM8, ITGAM, FCGR3B, PTPRC	(Uhlen et al., 2019)(Monaco et al., 2019)
Eosinophils	EPX, RNASE2, CLC, SIGLEC8	(Uhlen et al., 2019)
Erythrocytes	HBA2, HBB, HBA1, SLC4A1, EPB41	(Karlsson et al., 2021)
Platelets	PPBP, ITGA2B, PF4	(Shattil et al., 1985; Poncz et al., 1987; Majumdar et al., 1991)

Table S3. Sequencing metrics of scMultiome-ATAC analysis in PBMCs

	Tn5-TXGv2	Tn5-H50	Tn5-H100
Sequenced ATAC read pairs	1,304,897,617	1,431,221,793	1,444,079,941
Sequenced RNA read pairs	714,778,267	644,951,970	716,455,503
Duplicates ATAC	68.5 %	67.2 %	57.3 %
Duplicates RNA	92.3 %	91.8 %	92.2 %
Mapped ATAC read pairs	92.5 %	91.0 %	91.2 %
Mapped RNA read pairs	61.4 %	53.1 %	51.8 %

Table S4. Sequencing metrics of scMultiome-C&T in TF1 cell lines

	TF1-IDH2wt	**TF1-IDH2mut**
Sequenced H3K27me3 read pairs	486,259,010	468,044,936
Sequenced RNA read pairs	42,673,800	36,916,878
Duplicates H3K27me3	96.6 %	96.6 %
Duplicates RNA	27.0 %	29.5 %
Mapped H3K27me3 read pairs	84.1 %	83.5 %
Mapped RNA read pairs	68.5 %	68.5 %

Supplementary References

Karlsson, M., Zhang, C., Mear, L., Zhong, W., Digre, A., Katona, B., et al. (2021). A single-cell type transcriptomics map of human tissues. *Sci Adv* 7(31). doi: 10.1126/sciadv.abh2169.

Majumdar, S., Gonder, D., Koutsis, B., and Poncz, M. (1991). Characterization of the human beta-thromboglobulin gene. Comparison with the gene for platelet factor 4. *Journal of Biological Chemistry* 266(9), 5785-5789. doi: 10.1016/s0021-9258(19)67665-9.

Monaco, G., Lee, B., Xu, W., Mustafah, S., Hwang, Y.Y., Carre, C., et al. (2019). RNA-Seq Signatures Normalized by mRNA Abundance Allow Absolute Deconvolution of Human Immune Cell Types. *Cell Rep* 26(6), 1627-1640 e1627. doi: 10.1016/j.celrep.2019.01.041.

Poncz, M., Surrey, S., LaRocco, P., Weiss, M.J., Rappaport, E.F., Conway, T.M., et al. (1987). Cloning and characterization of platelet factor 4 cDNA derived from a human erythroleukemic cell line. *Blood* 69(1), 219-223. doi: 10.1182/blood.V69.1.219.219.

Shattil, S.J., Hoxie, J.A., Cunningham, M., and Brass, L.F. (1985). Changes in the platelet membrane glycoprotein IIb.IIIa complex during platelet activation. *J Biol Chem* 260(20), 11107-11114.

Terstappen, L.W., Huang, S., and Picker, L.J. (1992). Flow cytometric assessment of human T-cell differentiation in thymus and bone marrow. *Blood* 79(3), 666-677. doi: 10.1182/blood.V79.3.666.666.

Uhlen, M., Karlsson, M.J., Zhong, W., Tebani, A., Pou, C., Mikes, J., et al. (2019). A genome-wide transcriptomic analysis of protein-coding genes in human blood cells. *Science* 366(6472). doi: 10.1126/science.aax9198.