Enhancing sensitivity and versatility of Tn5-based single cell omics

Isabelle Seufert[1,2,3], Pooja Sant[4], Katharina Bauer[4],
Afzal P. Syed[1,2,4], Karsten Rippe[1,2]* and Jan-Philipp Mallm[2,4]*

[1]German Cancer Research Center (DKFZ) Heidelberg, Division of Chromatin Networks, Heidelberg, Germany, [2]Center for Quantitative Analysis of Molecular and Cellular Biosystems (BioQuant), Heidelberg University, Heidelberg, Germany, [3]Faculty of Biosciences, Heidelberg University, Heidelberg, Germany, [4]German Cancer Research Center (DKFZ) Heidelberg, Single Cell Open Lab, Heidelberg, Germany

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 loci. However, a current limitation is the sparse coverage of these open sites 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

Tn5 transposase is a crucial enzyme for the analysis of chromatin features in single cells 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 (Dimitriu et al., 2022). These methods 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. 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; Zhu et al., 2021). For all these applications, improvements to the Tn5 tagmentation reaction would advance the single cell analysis, which is frequently limited by data sparsity. 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). This rather sparse coverage of open loci detected in a given cell limits the resolution of cellular heterogeneity based on the ATAC signal. Aggregation of similar cells improves the detection of open sites (Pliner et al., 2018) but, at the same time, averages out the stochastic differences in 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 describe the use of custom Tn5 preparations in single-cell sequencing analysis to increase the sensitivity of scATAC-seq, provide protocols to measure Tn5 activity, describe Tn5 adapter loading and its use in multiome workflows. 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 (Figure 1A). Our work comprised three different reaction types: 1) 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. 2) 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. 3) 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 conducted a multiome profiling 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 different applications (Figure 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 then devised a qPCR assay with lambda DNA as substrate to measure relative Tn5 activities against a standard curve (Supplementary Tables S1, S2). It was used to compare the activity of different concentrations of an in-house Tn5 preparation (Tn5-H) loaded with the adapters listed in Table 1 to Tn5 enzymes from 10× Genomics (Tn5-TXG) as well as the Illumina TDE1 enzyme (Tn5-ILMN) (Figure 1C).

The highest concentration of in-house Tn5 (Tn5-H100, 83 µg/ml or 1.6 µM) 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 (Figure 1C, Supplementary Table S2). 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 similar to that of Tn5-TXGv1.1 (Figure 1C).

We then compared our Tn5-H100 and Tn5-H30 (25 µg/ml or 0.48 µM) 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 (Figure 2A, Supplementary Figure S2A; Supplementary Figures S1A, B, Supplementary Table S3): Recovered cell barcodes showed 1) higher TSS enrichment scores, 2) higher number of unique fragments, and 3) 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 (Figure 2A, red rectangle, Supplementary Table S4). The scTurboATAC protocol yielded an about 2-fold higher number of unique fragments normalized to sequencing depth than cells from the regular scATAC workflow (Figure 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 (Figure 2C, Supplementary Figure S1C). Embedding and clustering of both data sets revealed an increased resolution of the scTurboATAC data for both MEF (Figure 2D) and PBMC data (Figure 2E).

Two distinct MEF subtypes (epithelial- and mesenchymal-like MEFs) were annotated with marker genes from our previous work (Muckenhuber et al., 2023) after removing minor clusters 1 and 2 (scATAC) and cluster 1 (scTurboATAC) that showed increased apoptosis module scores (Supplementary Figure S1D). Cluster 4 of scATAC and cluster 2 of scTurboATAC showed increased mesenchymal-like MEF marker module scores (Figure 2D, Supplementary Figure S2A). The epithelial-like MEF markers were enriched in cluster 3 of scATAC and clusters 3 and 4 of scTurboATAC (Figure 2D, Supplementary Figure S2B). A differential analysis of gene activity scores yielded 552 differentially active genes between epithelial clusters C3 and C4 of scTurboATAC (FDR ≤ 0.05 and log2FC ≥ 0.5). The most enriched GO term of these differentially active genes was GO: 0061844 "antimicrobial humoral immune response" (Fisher p-value of 3.6e-06), likely arising from the IFNβ treatment. When computing module scores of previously identified interferon stimulated genes (ISGs) in MEFs for clusters C2, C3 and C4 of scTurboATAC, cluster C4 showed the lowest activity which points to a weaker response (Supplementary Figure S2C).

It is noted that differences in sequencing depth can affect the resolution and might contribute to the differences observed. However, when sub-sampling the ATAC data to obtain the same read numbers for each Tn5 preparation, we still obtained higher number of fragments per cell with an additional cluster for scTurboATAC (Supplementary Figures S2D, E, Supplementary
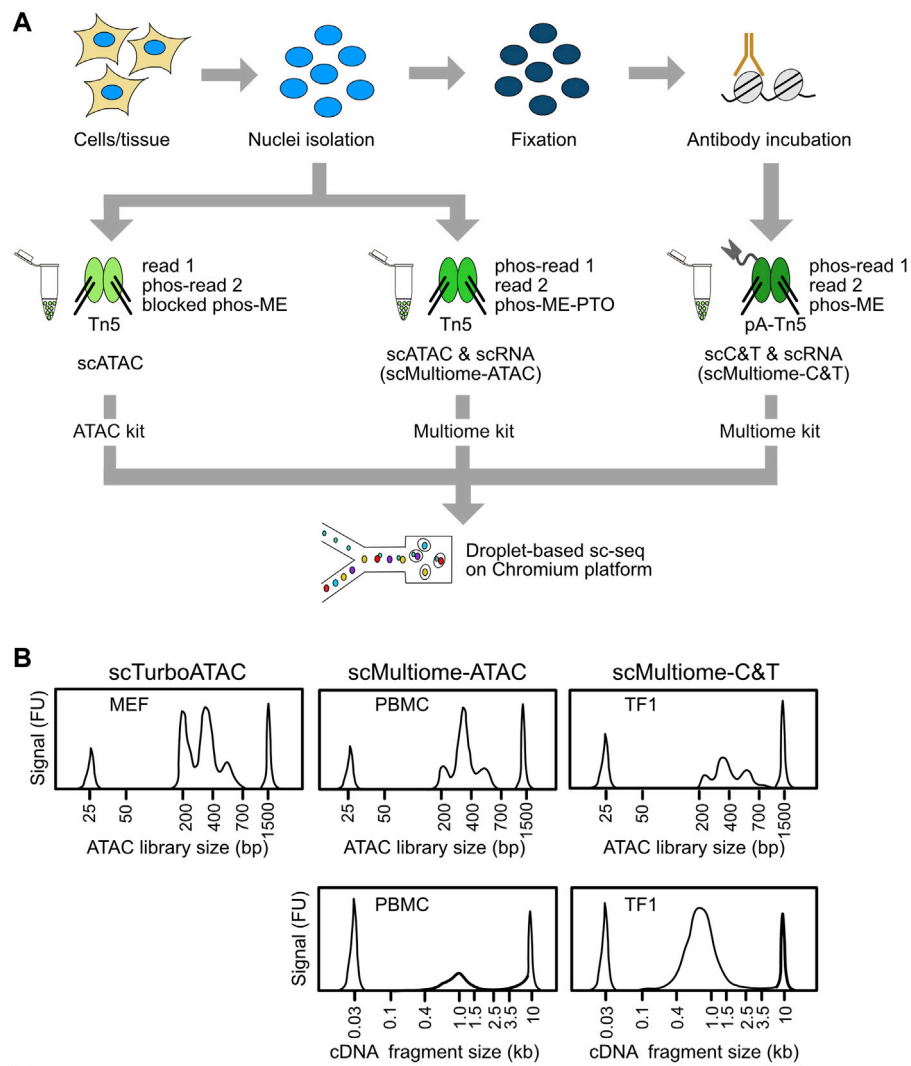
FIGURE 1
Experimental workflow. **(A)** Experimental procedure and use of the differently loaded Tn5 preparations (for details please refer to the method section). For scATAC and the multiome workflow nuclei were prepared and incubated with Tn5. Importantly, adapters varied between the protocols. For 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. **(C)** Tn5 activity measurements by qPCR. Left: 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. Right: 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 resulted in a
(Continued)

FIGURE 1 (Continued)
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. *P*-value from Student's t-test (two-sided, unpaired) is indicated as *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

TABLE 1 Oligonucleotides used for Tn5 loading

Oligonucleotide[a]	Sequence	Assay	Sample
read 1	5′-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3′	TurboATAC	MEFs plus IFNβ, PBMCs
phos-read 2	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-read 1	5′-[phos]TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3′	TurboATAC & RNA	PBMCs
read 2	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-read 1	5′-[phos]TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3′	C&T & RNA	TF1 wt, TF1 IDH2mut
read 2	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. The PTO backbone is used to avoid self-dimerization which can occur due to the higher Tn5 concentration in TurboATAC. This modification is not required in the C&T protocol, which includes additional washing steps. Phosphorylation for read 1 is needed when the multiome-kit is used since the barcode oligos on the beads are added to the fragmented chromatin by ligation. Phosphorylation of the ME oligo is required for tagmentation. Blocking of the 3' end prevents premature elongation by the polymerase in oil droplets. Phosphorylation of read 2 is optional for TurboATAC.

Table S5). Furthermore, the scTurboATAC PBMC data also displayed an increased number of clusters (14 instead of 10) as compared to scATAC at comparable read depth. Thus, the scTurboATAC protocol resolves two epithelial-like clusters, while only one cluster was detected with the standard scATAC protocol. The 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.
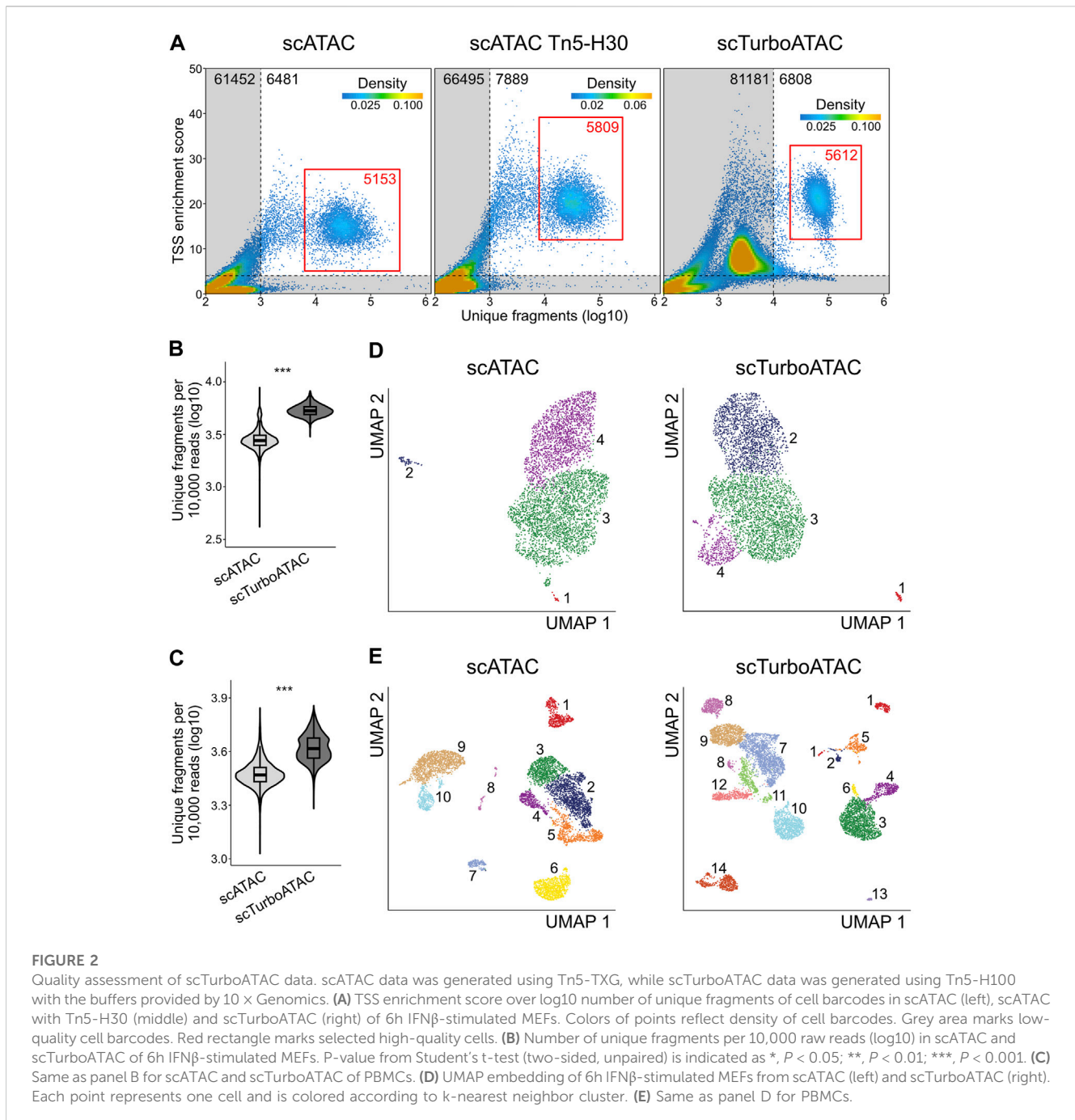
scTurboATAC improves the analysis of transcription regulation

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 compared to the scATAC protocol (Supplementary Figures 2F, G). Here, scATAC recovered roughly 15,000 and scTurboATAC 26,000 accessible peaks per cell on average. Furthermore, accessibility footprints at transcription start sites (TSSs) and at STAT1 and CTCF binding motifs showed similar or better enrichment in pseudo-bulk scTurboATAC compared to scATAC (Figure 3A).

Next, we applied our previously described approach for co-accessibility analysis with the RWire software to the MEF data set (Muckenhuber et al., 2023). RWire reveals enhancer-promoter links from correlation between open sites of scATAC-seq data. For epithelial-like MEFs of scTurboATAC, we only analyzed cluster 3 as it represents a homogeneous cell population. 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 (Figure 3B). Co-accessible links from scTurboATAC showed higher correlation coefficients and percent accessible cells than co-accessible links from scATAC (Supplementary Figure S2; Supplementary Figure S2H).

Transcription of interferon stimulated genes (ISGs) is partly regulated by simultaneous STAT1 and STAT2 (STAT1/2) binding to ISG promoters as well as to 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- and mesenchymal-like MEFs from scTurboATAC (Figure 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. For both epithelial- and mesenchymal-like MEFs newly linked promoters by TurboATAC belonged to genes associated with "cellular response to interferon-beta" (p-values $7x10^{-11}$ and $1.9x10^{-12}$ respectively) highlighting the relevance of the identified connections. An example for the improved detection of ISG regulation by distal STAT1/2 binding events in the scTurboATAC data is shown for the *Ifit* gene cluster in Figure 3D. The 150 kb locus depicted contains six ISGs (*Ifit1, Ifit1bl1, Ifit1bl2, Ifit2, Ifit3,* and *Ifit3b*) (Muckenhuber et al., 2023). For *Ifit1, Ifit1bl2* and *Ifit3b* promoters, direct STAT1/2 binding was observed in IFNβ-stimulated MEFs. Three co-accessible links were detected between ISG promoters and distal STAT1/2 binding events in epithelial- and mesenchymal-like MEFs of scATAC. These co-accessible links were also detected by scTurboATAC where the percent accessible cells was higher. In addition, six co-accessible links between ISG promoters and STAT1/

FIGURE 2
Quality assessment of scTurboATAC data. scATAC data was generated using Tn5-TXG, while scTurboATAC data was generated using Tn5-H100 with the buffers provided by 10 × Genomics. **(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. P-value from Student's t-test (two-sided, unpaired) is indicated as *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$. **(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.

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 in PBMCs (Figure 4A, Supplementary Table S6). The scTurboATAC protocol yielded a higher number of accessible pseudo-bulk peaks per cell (24,500 on average) compared to the scATAC protocol (19,000 on average, Supplementary Figure S3A). This difference resulted in an improved detection of progenitor cells, classical monocytes/dendritic cells and provided a higher resolution of the B cell cluster (Figure 4A, B). We identified a distinct B cell cluster with scTurboATAC (cluster 1) that was not resolved in scATAC data by integrating B cells from scATAC and scTurboATAC (Supplementary Figure S3B). When investigating the B cell clusters further, we found differential TF activities for
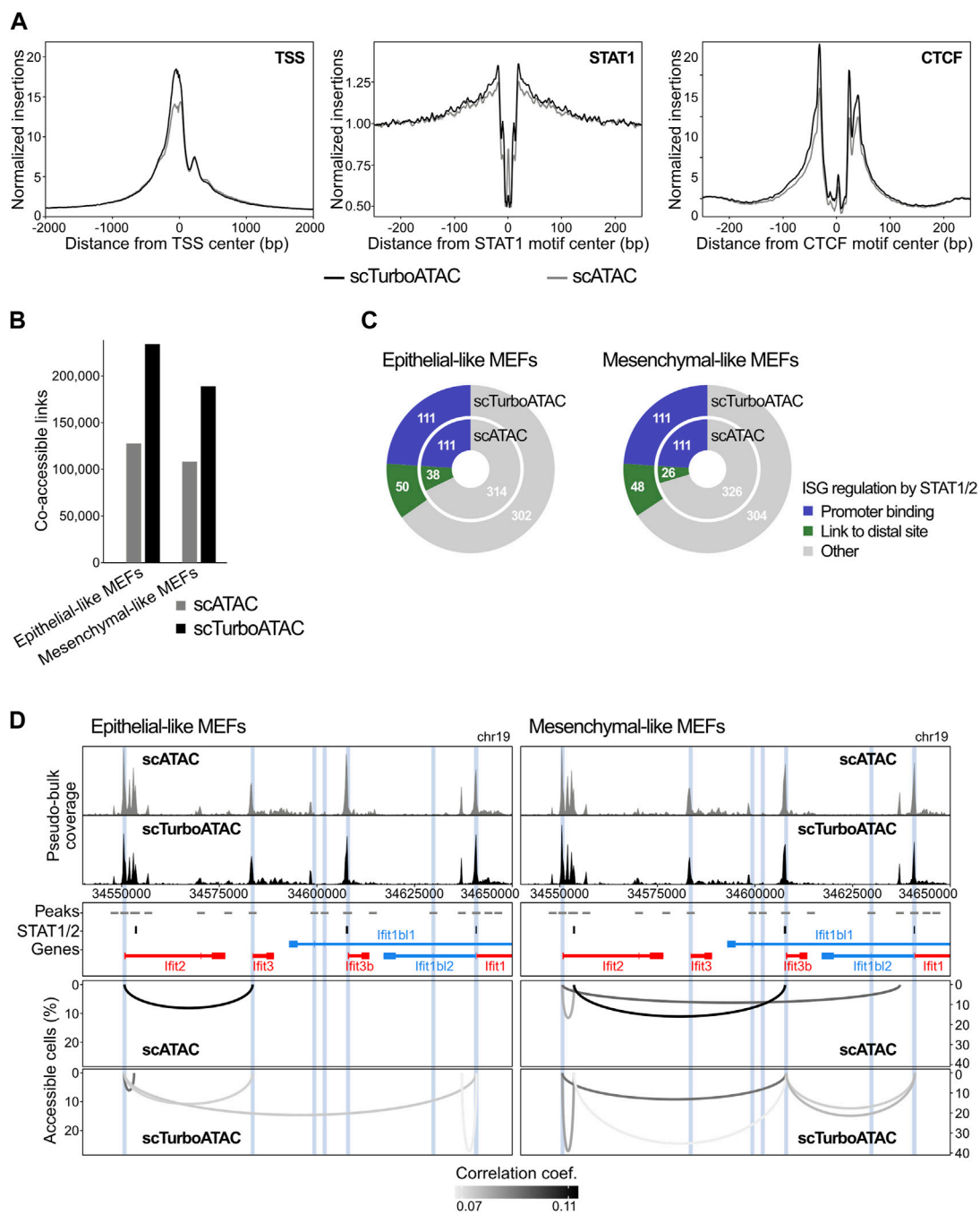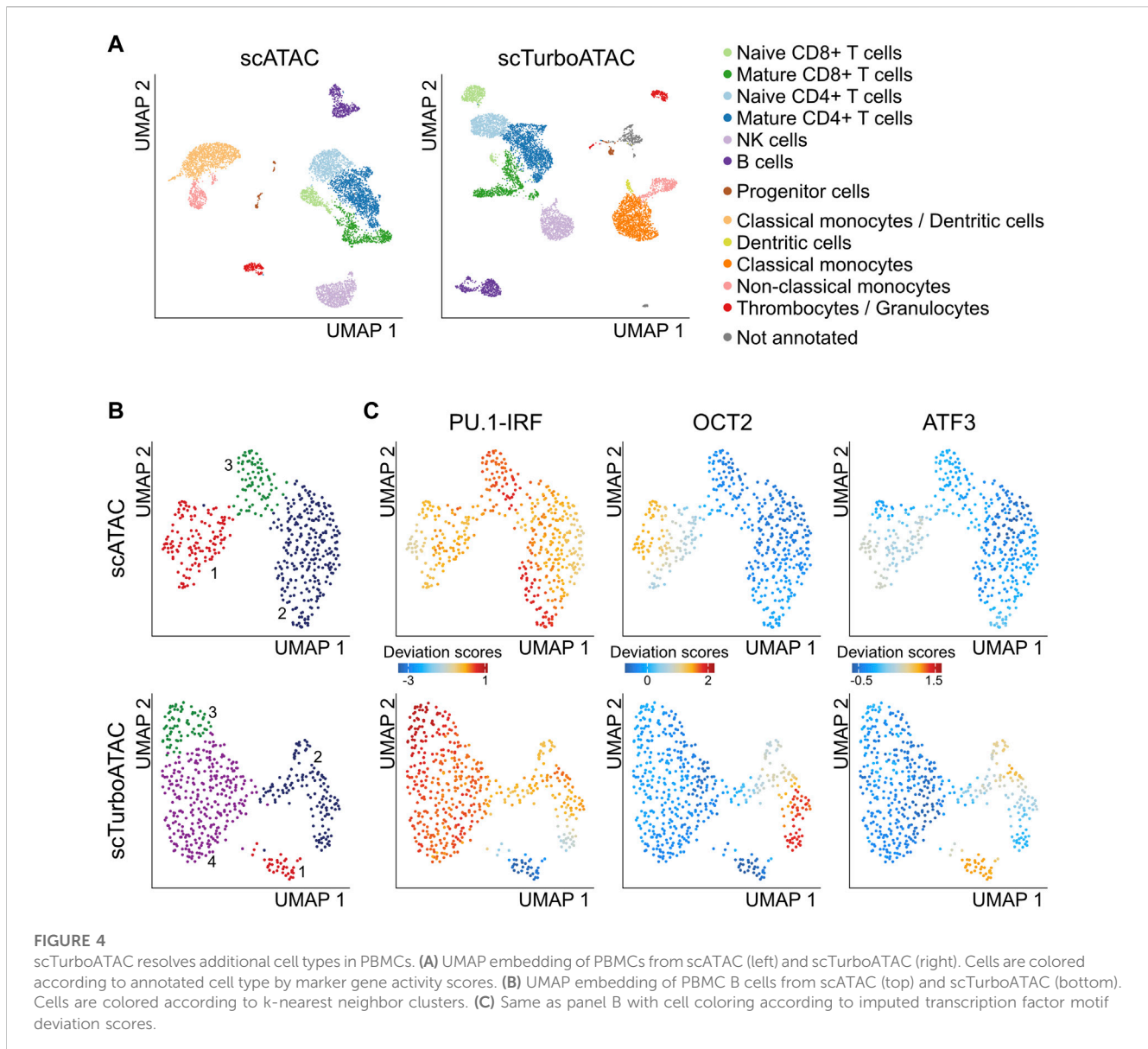
FIGURE 3
Comparison of scATAC and scTurboATAC. MEFs stimulated with IFNβ for 6 hours were analyzed. **(A)** Accessibility footprints at TSSs (left), STAT1 motifs (middle) and CTCF motifs (right) from scATAC and scTurboATAC. **(B)** Number of co-accessible links in scATAC and scTurboATAC of epithelial- and mesenchymal-like MEFs. P-value from Student's t-test (two-sided, paired) is 0.087 for scATAC vs. scTurboATAC. **(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. Widths are scaled according to numbers. **(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.

PU.1-IRF, OCT2 and ATF3 among others (Figures 4B, C, Supplementary Figure S3C). 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. Since ATF3 modulates metabolism and immunity, especially immune-responsiveness, this population might represent a

FIGURE 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.

specific cellular state (Ku and Cheng, 2020). Previous studies suggested that ATF3 is a repressor of IRF targets (Garber et al., 2012; Labzin et al., 2015). In line with this view, the population with high ATF3 activity has low PU.1-IRF scores.

scTurboATAC can be integrated into an improved multi-omics analysis

Multi-omics approaches allow to directly link epigenetic heterogeneity with transcriptional profiles. Accordingly, we adapted our scTurboATAC protocol to this type of application and devised a Tn5 loading scheme that is compatible with the 10x Genomics multiome chemistry. The latter utilizes a spacer oligonucleotide to add cell barcodes to the Tn5 fragments possibly 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 (Figure 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.

Our protocol yielded a higher number of unique fragments normalized to sequencing depth when using self-loaded Tn5-H100 in comparison to Tn5-TXGv2 with the multiome chemistry in PBMCs (Figure 5A left, Supplementary Table S7) while the number of reads in TSSs was similar (Supplementary Figures S4A, B). UMI counts of transcripts normalized to sequencing depth showed the same yield in comparison to the standard workflow (Figure 5A right, Supplementary Table S7). The number of clusters was higher for scMultiome-ATAC with Tn5-
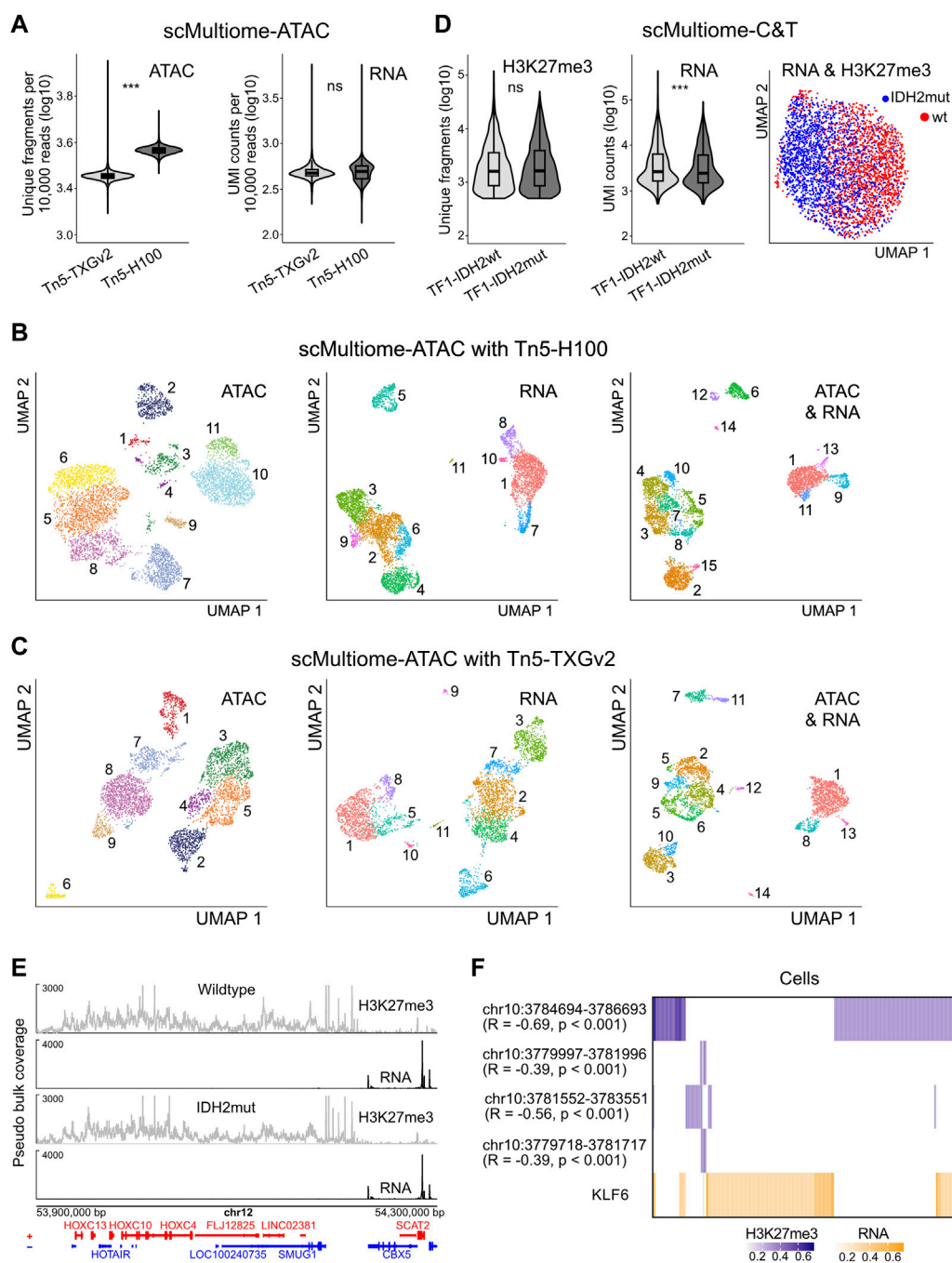
FIGURE 5
Multiome analysis. **(A)** scMultiome-ATAC analysis of PBMCs. The log10 number of unique fragments per 10,000 raw reads (ATAC, left) and UMI counts per 10,000 raw reads (RNA, right) is shown. P-values from Student's t-test (two-sided, unpaired) are indicated as *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$. **(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. P-values from Student's t-test (two-sided, unpaired) are indicated as *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$. Right, UMAP co-embedding of H3K27me3 and RNA. Cells are colored according to sample. **(E)** scMultiome-C&T of TF1 wildtype and IDH2 mutated cells. 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. **(F)** Heatmap of KLF6 gene expression (log10 UMI counts; orange) and KLF6 promoter H3K27me3 (log10 counts; blue) in TF1 wildtype cells. Single cells are clustered hierarchically by complete linkage. Only single cells with KLF6 signal in either RNA or promoter H3K27me3 were used. Spearman correlation between KLF6 expression and promoter H3K27me3 was computed.

H100 (Figure 5B) compared to the standard scMultiome-ATAC with Tn5-TXGv2 (Figure 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 (Figures 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. For both samples the number of H3K27me3 fragments was similar (Figure 5D left, Supplementary Table S8) with a bias for fragments with higher nucleosome numbers (Supplementaty Figure S4C), a high fraction of reads in peaks (Supplementaty Figure S4D) and comparable UMI transcript counts (Figure 5D right, Supplementary Table S8). UMAP of C&T and RNA data from both cell lines showed separate embedding with some overlap according to IDH2 mutation status (Figure 5D far right).

Very broad regions of H3K27me3 enrichment were observed as depicted for an exemplary region of the HOXC gene cluster in Figure 5E. In both cell lines all genes within this H3K27me3-rich region were silenced while adjacent genes lacking H3K27me3 signal were highly expressed.

To evaluate the data quality further we investigated the heterogeneity of gene expression in the TF1 wildtype cell line and its correlation to promoter-associated H3K27me3. The number of RNA UMIs was >2500 on average per cell and provided a good coverage of the single cell transcriptome profile. It revealed heterogeneous expression of the transcription factor KLF6 in TF1 wildtype cells. While the number of unique fragments per cell obtained from the H3K27me3 C&T analysis yields a relatively sparse coverage of the rather broad genome-wide H3K27me3 peaks, the H3K9me3 signal was sufficient to reveal a medium to high negative correlation of KLF6 expression with the H3K27me3 modification at the KLF6 promoter (Figure 5F). Interestingly, the epigenetic deregulation of KLF6 has been linked to AML (Adelman et al., 2019), where recurrent mutations of IDH1/2 are drivers of tumorigenesis and therapeutic target (Tyner et al., 2018). 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.

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 resulted in a higher library complexity with a consistently lower number of duplicates and higher number of integration events per cell (Supplementary Table S3). These enhancements led to better results in the downstream analysis as shown for the resolution of cell types, computation of TF motif footprints and activities as well as co-accessibility analysis between cis-regulatory elements.

We showed for instance that scTurboATAC resolved an additional cell cluster in the B-cell population. The TF activity computed revealed ATF3 as one of the TFs with elevated activity for this specific B-cell cell state. This cell state might be governed by ATF3, since it is a central factor for numerous functions related to immunity. Furthermore, we also observed opposing activities for PU1.-IRF and OCT2. This is in line with ATF3 acting as a repressor of IRF targets as reported previously (Garber et al., 2012; Labzin et al., 2015).

Similarly, we obtained an additional cell cluster in scTurboATAC data for epithelial-like MEFs after treatment with IFNβ. When comparing gene activity scores between the unique epithelial-like MEFs (scTurboATAC cluster C4) and the shared epithelial-like MEFs with scATAC (scTurboATAC cluster C3) we found GO terms linked to immune response as top hits. This suggests that the two epithelial-like MEF clusters mainly differ in their response to IFN stimulation with cluster C4 showing a reduced response.

Co-accessibility of promoter and enhancer is yet another layer to investigate regulatory elements and their impact on transcription. Due to data sparsity this type of analysis is usually done using cell aggregates. Here, the co-accessibility analysis was conducted without aggregating cells to also capture stochastic changes of chromatin accessibility and to identify STAT1/2 mediated co-accessibility upon IFNβ stimulation. We show that the scTurboATAC protocol yielded a higher number of distal STAT1/2 binding sites that could be linked to their ISG promoters. These additional links could be due to generally higher correlation coefficients between promoters and distal sites or due to stochastic events that are not captured with more sparse data.

The droplet-based approaches used here 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). On the other hand, plate-based scATAC methods, similar to scRNA-seq, could perform better than the droplet-based method used here (Xu et al., 2021). 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 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. An alternative approach has been published using two Tn5 reactions—one for chromatin accessibility and a second round after cDNA synthesis (Xu et al., 2022). This study showed similar (but not higher)

sensitivity as compared to the commercial 10x Genomics multiome kit for chromatin accessibility even though a higher Tn5 concentration was used as compared to our study. Since Tn5 activity is key to these improvements, activity measurements, e.g., by the qPCR assay described here, are important to ensure that the integration efficiency is high.

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 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 droplet-based workflow has been described before but as a single readout without transcriptome information (Kaya-Okur et al., 2019). In addition, combinatorial barcoding has been used for detecting histone profiles and RNA (Zhu et al., 2021). 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) and heterogeneity within a cell line could be detected despite the inherent sparsity of the data. 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. Furthermore, no specific equipment to run this assay would be required than the one that might already be in place in labs generating single-cell data.

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

Expression and purification of the Tn5 transposase

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) with some adjustments resulting in the following workflow: The pETM11-Sumo3-Tn5 plasmid encoding His_6-Sumo3-tagged Tn5 E54K L372P was freshly transformed into *E. coli* BL21(DE3) codon + RIL cells (Stratagene). Precultures were grown overnight at 37°C in LB medium supplemented with 30 μg/ml kanamycin and 34 μg/ml chloramphenicol and used to inoculate the large-scale expression cultures. A 10 ml preculture was added to 1 liter of TB-FB medium supplemented with 2 mM $MgSO_4$, 0.05% glucose, 1.5% lactose, 30 μg/ml kanamycin and 34 μg/ml chloramphenicol. The cultures were grown at 37°C until the OD_{600} reached ~0.6,

after which the temperature was reduced to at 18°C. After overnight expression at 18°C, the cultures were harvested by centrifugation (30 min, 5000 x *g*, 4°C) and the pellets were flash-frozen in liquid nitrogen and stored at −80°C until the start of the protein purification.

For the protein purification, all chromatography steps were performed at 4°C and samples were kept on ice at all times. The cell pellet was resuspended in cold running buffer (20 mM Hepes-NaOH pH 7.2, 800 mM NaCl, 5 mM imidazole, 1 mM EDTA, 2 mM DTT and 10% glycerol) supplemented with cOmplete EDTA-free protease inhibitors (Roche). The cells were lysed by 5 passages through a microfluidizer, followed by centrifugation (30 min, 140000 x *g*, 4°C). After centrifugation, a 10% polyethyleneimine (PEI) solution at pH 7.2 was added dropwise to the cleared lysate to a final concentration of 0.5% while constantly stirring the solution. To remove the PEI-precipitated fraction, the lysate was subjected to another centrifugation step at 20000 x *g* for 30 minutes at 4°C. The cleared lysate containing the recombinant His_6-Sumo3-tagged Tn5 protein was then loaded on a 5 ml prepacked cOmplete His-tag purification column (Roche) equilibrated in running buffer. After loading the sample, the cOmplete column was washed with the same buffer until the absorbance signal at 280 nm returned to baseline. The column was eluted with a 1-step elution in running buffer supplemented with 300 mM imidazole. The elution fractions were analysed via SDS-PAGE and the fractions containing the His_6-Sumo3-tagged Tn5 were pooled and dialysed overnight to running buffer in the presence of His_6-tagged SenP2 protease to remove the His_6-Sumo3 fusion tag. The next morning, the dialysed sample was loaded again on a 5 ml cOmplete His-tag purification column equilibrated with running buffer. The untagged Tn5 was collected in the flow through of the column and further purified using size exclusion chromatography. The Tn5 sample was loaded on a HiLoad Superdex 200 16/600 pg column (GE Healthcare) equilibrated with 50 mM Tris pH 7.5, 800 mM NaCl, 0.2 mM EDTA, 2 mM DTT and 10% glycerol. The elution fractions were analysed via SDS-PAGE and the dimer peak fractions were pooled. The final Tn5 sample was aliquoted and stored at −20°C in 25 mM Tris pH 7.5, 800 mM NaCl, 0.1 mM EDTA, 1 mM DTT and 50% glycerol at a concentration of 1 mg/ml.

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. 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 (e.g. 2.5 μl each, 10 μl in total) and incubated for 30 min at room temperature. The loaded Tn5 was then diluted with 20 μl DB yielding the Tn5-H100 (83 μg/ml) 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 10× 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 (CAAGCAGAAGACGGCATACGAGAT), 0.2 μl 10 μM i5 adapter (AATGATACGGCGACCACCGAGATCTACAC), 1 μl diluted template DNA and 3.6 μl nuclease free water. The following program was run on a StepOnePlus (Applied Biosystems) machine (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 (see Supplementary Tables S1, S2). Linear performance was verified using a diluted reference standard.

scTurboATAC library preparation

For scTurboATAC, the transposase provided with the ATAC kit from 10 × Genomics was replaced with Tn5-H100. Purified nuclei were processed using the standard ATAC protocol according to the user guides for Chromium Next GEM Single Cell ATAC v1.1 or v2 kits (CG000496, 10 × 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, United States, wildtype, CRL-2003; IDH2 mutant, CRL-2003). They were cultured following the providers instructions. Cells were harvested, washed once in PBS and 1×10^6 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 1 × 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 10 × 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 (10× 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 (10 × 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 (Supplementary Table S4). 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 (10× 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 (10× 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. Differential gene expression analysis was performed on log10-transformed and scaled data using Wilcoxon test in Seurat (only.pos = FALSE, min.pct = 0.1, logfc.threshold = 0).

Data availability statement

The single cell sequencing data generated in this study are publicly available at Gene Expression Omnibus under accession number GSE235506 via the link https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE235506.

Author contributions

IS: Data analysis and writing the manuscript PS, KB, and AS: Data acquisition KR: Supervision and writing the manuscript J-PM: Conception, data analysis and writing the manuscript. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author J-PM declared that he was an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/freae.2023.1245879/full#supplementary-material

References

Adelman, E. R., Huang, H. T., Roisman, A., Olsson, A., Colaprico, A., Qin, T., et al. (2019). Aging human hematopoietic stem cells manifest profound epigenetic reprogramming of enhancers that may predispose to leukemia. *Cancer Discov.* 9, 1080–1101. doi:10.1158/2159-8290.cd-18-1474

Adey, A. C. (2021). Tagmentation-based single-cell genomics. *Genome Res.* 31, 1693–1705. doi:10.1101/gr.275223.121

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. doi:10.1186/s13059-023-02933-w

Buenrostro, J. D., Wu, B., Litzenburger, U. M., Ruff, D., Gonzales, M. L., Snyder, M. P., et al. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* 523, 486–490. doi:10.1038/nature14590

Chen, H., Lareau, C., Andreani, T., Vinyard, M. E., Garcia, S. P., Clement, K., et al. (2019). Assessment of computational methods for the analysis of single-cell ATAC-seq data. *Genome Biol.* 20, 241. doi:10.1186/s13059-019-1854-5

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. doi:10.1101/gad.7.4.570

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. doi:10.3389/fcell.2022.854317

Fulco, C. P., Nasser, J., Jones, T. R., Munson, G., Bergman, D. T., Subramanian, V., et al. (2019). Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat. Genet.* 51, 1664–1669. doi:10.1038/s41588-019-0538-0

Galluzzi, L., Vitale, I., Aaronson, S. A., Abrams, J. M., Adam, D., Agostinis, P., et al. (2018). Molecular mechanisms of cell death: recommendations of the nomenclature committee on cell death 2018. *Cell Death Differ.* 25, 486–541. doi:10.1038/s41418-017-0012-4

Garber, M., Yosef, N., Goren, A., Raychowdhury, R., Thielke, A., Guttman, M., et al. (2012). A high-throughput chromatin immunoprecipitation approach reveals principles of dynamic gene regulation in mammals. *Mol. Cell* 47, 810–822. doi:10.1016/j.molcel.2012.07.030

Granja, J. M., Corces, M. R., Pierce, S. E., Bagdatli, S. T., Choudhry, H., Chang, H. Y., et al. (2021). ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* 53, 403–411. doi:10.1038/s41588-021-00790-6

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. doi:10.1186/s13059-019-1874-1

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 Genomics Bioinforma.* 4, lqac095. doi:10.1093/nargab/lqac095

Hennig, B. P., Velten, L., Racke, I., Tu, C. S., Thoms, M., Rybin, V., et al. (2018). Large-scale low-cost NGS library preparation using a robust Tn5 purification and tagmentation protocol. *G3 Genes|Genomes|Genetics* 8, 79–89. doi:10.1534/g3.117.300257

Jiang, S., Huang, Z., Li, Y., Yu, C., Yu, H., Ke, Y., et al. (2023). Single-cell chromatin accessibility and transcriptome atlas of mouse embryos. *Cell Rep.* 42, 112210. doi:10.1016/j.celrep.2023.112210

Kaya-Okur, H. S., Wu, S. J., Codomo, C. A., Pledger, E. S., Bryson, T. D., Henikoff, J. G., et al. (2019). CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun.* 10, 1930. doi:10.1038/s41467-019-09982-5

Klemm, S. L., Shipony, Z., and Greenleaf, W. J. (2019). Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* 20, 207–220. doi:10.1038/s41576-018-0089-8

Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., et al. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* 16, 1289–1296. doi:10.1038/s41592-019-0619-0

Kramer, N. E., Davis, E. S., Wenger, C. D., Deoudes, E. M., Parker, S. M., Love, M. I., et al. (2022). Plotgardener: cultivating precise multi-panel figures in R. *Bioinformatics* 38, 2042–2045. doi:10.1093/bioinformatics/btac057

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. doi:10.3389/fendo.2020.00556

Labzin, L. I., Schmidt, S. V., Masters, S. L., Beyer, M., Krebs, W., Klee, K., et al. (2015). ATF3 is a key regulator of macrophage IFN responses. *J. Immunol.* 195, 4446–4455. doi:10.4049/jimmunol.1500204

Lahnemann, D., Koster, J., Szczurek, E., Mccarthy, D. J., Hicks, S. C., Robinson, M. D., et al. (2020). Eleven grand challenges in single-cell data science. *Genome Biol.* 21, 31. doi:10.1186/s13059-020-1926-6

Lee, J., Hyeon, D. Y., and Hwang, D. (2020). Single-cell multiomics: technologies and data analysis methods. *Exp. Mol. Med.* 52, 1428–1442. doi:10.1038/s12276-020-0420-2

Li, Z., Kuppe, C., Ziegler, S., Cheng, M., Kabgani, N., Menzel, S., et al. (2021). Chromatin-accessibility estimation from single-cell ATAC-seq data with scOpen. *Nat. Commun.* 12, 6386. doi:10.1038/s41467-021-26530-2

Mallm, J. P., Iskar, M., Ishaque, N., Klett, L. C., Kugler, S. J., Muino, J. M., et al. (2019). Linking aberrant chromatin features in chronic lymphocytic leukemia to transcription factor networks. *Mol. Syst. Biol.* 15, e8339. doi:10.15252/msb.20188339

Muckenhuber, M., Seufert, I., Muller-Ott, K., Mallm, J. P., Klett, L. C., Knotz, C., et al. (2023). Epigenetic signals that direct cell type-specific interferon beta response in mouse cells. *Life Sci. Alliance* 6, e202201823. doi:10.26508/lsa.202201823

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. doi:10.1126/science.1207194

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. doi:10.1101/gr.177881.114

Pliner, H. A., Packer, J. S., Mcfaline-Figueroa, J. L., Cusanovich, D. A., Daza, R. M., Aghamirzaie, D., et al. (2018). Cicero predicts cis-regulatory DNA interactions from single-cell chromatin accessibility data. *Mol. Cell* 71, 858–871.e8. doi:10.1016/j.molcel.2018.06.044

Riegel, D., Romero-Fernandez, E., Simon, M., Adenugba, A. R., Singer, K., Mayr, R., et al. (2023). Integrated single-cell profiling dissects cell-state-specific enhancer landscapes of human tumor-infiltrating CD8(+) T cells. *Mol. Cell* 83, 622–636.e10. doi:10.1016/j.molcel.2022.12.029

Sant, P., Rippe, K., and Mallm, J. P. (2023). Approaches for single-cell RNA sequencing across tissues and cell types. *Transcription*, 1–19. published online 16 April 2023. doi:10.1080/21541264.2023.2200721

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. doi:10.1038/nmeth.4401

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. doi:10.1126/science.8079170

Soshnikova, N., and Duboule, D. (2009). Epigenetic temporal control of mouse Hox genes *in vivo*. *Science* 324, 1320–1323. doi:10.1126/science.1171468

Stein, E. M. (2016). Molecular pathways: IDH2 mutations-Co-opting cellular metabolism for malignant transformation. *Clin. Cancer Res.* 22, 16–19. doi:10.1158/1078-0432.ccr-15-0362

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., 3rd, et al. (2019). Comprehensive integration of single-cell data. *Cell* 177, 1888–1902.e21. doi:10.1016/j.cell.2019.05.031

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. doi:10.1038/s41592-021-01282-5

Thibodeau, A., Eroglu, A., Mcginnis, C. S., Lawlor, N., Nehar-Belaid, D., Kursawe, R., et al. (2021). AMULET: a novel read count-based method for effective multiplet detection from single nucleus ATAC-seq data. *Genome Biol.* 22, 252. doi:10.1186/s13059-021-02469-x

Tyner, J. W., Tognon, C. E., Bottomly, D., Wilmot, B., Kurtz, S. E., Savage, S. L., et al. (2018). Functional genomic landscape of acute myeloid leukaemia. *Nature* 562, 526–531. doi:10.1038/s41586-018-0623-z

Uyehara, C. M., and Apostolou, E. (2023). 3D enhancer-promoter interactions and multi-connected hubs: organizational principles and functional roles. *Cell Rep.* 42, 112068. doi:10.1016/j.celrep.2023.112068

Van Dijk, D., Sharma, R., Nainys, J., Yim, K., Kathail, P., Carr, A. J., et al. (2018). Recovering gene interactions from single-cell data using data diffusion. *Cell* 174, 716–729.e27. doi:10.1016/j.cell.2018.05.061

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.e9. doi:10.1016/j.cels.2018.11.005

Xu, W., Wen, Y., Liang, Y., Xu, Q., Wang, X., Jin, W., et al. (2021). A plate-based single-cell ATAC-seq workflow for fast and robust profiling of chromatin accessibility. *Nat. Protoc.* 16, 4084–4107. doi:10.1038/s41596-021-00583-5

Xu, W., Yang, W., Zhang, Y., Chen, Y., Hong, N., Zhang, Q., et al. (2022). ISSAAC-seq enables sensitive and flexible multimodal profiling of chromatin accessibility and gene expression in single cells. *Nat. Methods* 19, 1243–1249. doi:10.1038/s41592-022-01601-4

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., et al. (2008). Model-based analysis of ChIP-seq (MACS). *Genome Biol.* 9, R137. doi:10.1186/gb-2008-9-9-r137

Zhu, C., Zhang, Y., Li, Y. E., Lucero, J., Behrens, M. M., and Ren, B. (2021). Joint profiling of histone modifications and transcriptome in single cells from mouse brain. *Nat. Methods* 18, 283–292. doi:10.1038/s41592-021-01060-3