Cell Systems

Joint analysis of chromatin accessibility and gene expression in the same single cells reveals cancer-specific regulatory programs

Highlights

- Parallel-seq enables scalable, cost-effective joint scATAC-and-scRNA profiling
- Generates >200,000 high-quality joint scATAC-and-scRNA profiles across 40 lung tumor samples
- Maps copy-number variations and extrachromosomal DNA heterogeneity in tumor cells
- Reveals regulatory events and enhancer mutations involved in tumor progression

Authors

Lei Tang, Jinsong Zhang, Yanqiu Shao, ..., Xiang Yan, Changjiang Feng, Qiangfeng Cliff Zhang

Correspondence

fengcj301@163.com (C.F.), qczhang@tsinghua.edu.cn (Q.C.Z.)

In brief

Tang et al. developed Parallel-seq, a high-throughput method for simultaneous single-cell chromatin accessibility and gene expression profiling. Applied to 40 lung tumor samples, it generated >200,000 high-quality joint scATAC-and-scRNA profiles, uncovering cancer-specific regulatory programs, oncogenic enhancer mutations, and therapeutic vulnerabilities, demonstrating its utility in advancing multiomics cancer research.

Tang et al., 2025, Cell Systems *16*, 101266
May 21, 2025 © 2025 The Authors. Published by Elsevier Inc.
https://doi.org/10.1016/j.cels.2025.101266

CellPress

Please cite this article in press as: Tang et al., Joint analysis of chromatin accessibility and gene expression in the same single cells reveals cancer-specific regulatory programs, Cell Systems (2025), https://doi.org/10.1016/j.cels.2025.101266

Cell Systems

CellPress
OPEN ACCESS

Methods

Joint analysis of chromatin accessibility and gene expression in the same single cells reveals cancer-specific regulatory programs

Lei Tang,[1,2,3,6] Jinsong Zhang,[1,2,3,6] Yanqiu Shao,[1,2,3,6] Yifan Wei,[1,2,3] Yuzhe Li,[1,2,3] Kang Tian,[1,2,3] Xiang Yan,[4] Changjiang Feng,[5,*] and Qiangfeng Cliff Zhang[1,2,3,7,*]

[1]State Key Laboratory of Membrane Biology, Beijing Advanced Innovation Center for Structural Biology & Frontier Research Center for Biological Structure, School of Life Sciences, Tsinghua University, Beijing 100084, China
[2]MOE Key Laboratory of Bioinformatics, Center for Synthetic and Systems Biology, School of Life Sciences, Tsinghua University, Beijing 100084, China
[3]Tsinghua-Peking Center for Life Sciences, Beijing 100084, China
[4]Department of Medical Oncology, the Fifth Medical Center, Beijing 301 Hospital, Beijing 100039, China
[5]Department of Thoracic Surgery, the First Medical Center, Beijing 301 Hospital, Beijing 100039, China
[6]These authors contributed equally
[7]Lead contact
*Correspondence: fengcj301@163.com (C.F.), qczhang@tsinghua.edu.cn (Q.C.Z.)
https://doi.org/10.1016/j.cels.2025.101266

SUMMARY

Biological analyses conducted at the single-cell scale have revealed profound impacts of heterogeneity and plasticity of chromatin states and gene expression on physiology and cancer. Here, we developed Parallel-seq, a technology for simultaneously measuring chromatin accessibility and gene expression in the same single cells. By combining combinatorial cell indexing and droplet overloading, Parallel-seq generates high-quality data in an ultra-high-throughput fashion and at a cost two orders of magnitude lower than alternative technologies (10× Multiome and ISSAAC-seq). We applied Parallel-seq to 40 lung tumor and tumor-adjacent clinical samples and obtained over 200,000 high-quality joint scATAC-and-scRNA profiles. Leveraging this large dataset, we characterized copy-number variations (CNVs) and extrachromosomal circular DNA (eccDNA) heterogeneity in tumor cells, predicted hundreds of thousands of cell-type-specific regulatory events, and identified enhancer mutations affecting tumor progression. Our analyses highlight Parallel-seq's power in investigating epigenetic and genetic factors driving cancer development at the cell-type-specific level and its utility for revealing vulnerable therapeutic targets.

INTRODUCTION

The cellular heterogeneity and plasticity of the tumor microenvironment (TME) strongly affect both cancer progression and the responsivity of tumors to therapies.[1] The advent of single-cell sequencing technologies has revealed intratumor heterogeneity at the genomic,[2] epigenomic,[3] transcriptomic,[4] and proteomic levels.[5] These technologies can profile the multimodal landscape (i.e., "multiomics") of a sample in a collective way; however, each technology was originally "mono-omic," only analyzing a data modality in each single cell.[6] Joint analysis of multiple modalities in the same single cells from a tumor sample can reveal a more complete picture of the regulatory processes occurring at distinct layers and facilitate the discovery of vulnerable cell populations and molecular targets for therapeutic development.[7]

Multiple methods for joint analysis of the same single cells have been reported. Notable examples include G&T-seq,[8] which integrates genome and transcriptome analyses, ScTrio-seq,[9] which analyzes the genome, DNA methylome, and transcriptome, and CITE-seq,[10] which measures the transcriptome and protein levels of cell surface markers. Methods for joint analysis of chromatin accessibility and gene expression in single cells have also been reported.[11–15] These methods have been instrumental in dissecting the cell composition of organs and to uncover epigenetic regulatory mechanisms and developmental transitions. Recently, such methods have been applied for cancer study and identified candidate cancer-associated epigenetic drivers.[16]

Here, we present paired analysis of RNA and accessible chromatin in the same single cell by sequencing (Parallel-seq), a single-cell analysis technique that combines combinatorial indexing and droplet-based single-cell sequencing (dsc-seq) to simultaneously measure chromatin accessibility and gene expression in the same cell (i.e., joint scATAC-and-scRNA profiling). By "overloading" chromium droplets with multiple cells (~10 cells, making highly efficient use of generated droplets) and performing both pre- and post-indexing to distinguish the cells within each droplet, Parallel-seq achieves technological

This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Please cite this article in press as: Tang et al., Joint analysis of chromatin accessibility and gene expression in the same single cells reveals cancer-specific regulatory programs, Cell Systems (2025), https://doi.org/10.1016/j.cels.2025.101266

reliability, high data quality, and a vast barcoding space that supports ultra-high-throughput and cost-effective analysis, outperforming state-of-the-art competing methods for joint scATAC-and-scRNA profiling. We performed Parallel-seq on 40 lung tumor and tumor-adjacent samples and obtained an extensive dataset encompassing over 200,000 high-quality joint scATAC-and-scRNA profiles, demonstrating its application scope and power for the identification of cell-type-specific regulatory programs in both tumor cells and TME immune cells.

RESULTS

Parallel-seq analyzes RNA and accessible chromatin in the same single cells

Parallel-seq combines combinatorial indexing with dsc-seq overloading[17–19] (Figure 1A; STAR Methods). Briefly, (1) Parallel-seq starts by fixing and permeabilizing cells distributed in a 96-well plate, followed by Tn5 transposition to label open chromatin regions with a well-specific barcode (BC1); (2) mRNAs are reverse transcribed using poly(T)$_{15}$ and random hexamer primers with the same sets of barcodes (BC1); (3) cells from different wells are combined and randomly distributed into a 96-well plate to ligate splint adapters containing the second sets of well-specific barcodes (BC2) to transposed chromatin and first strand cDNA; (4) after removal of RNA using RNases, a randomly primed second-strand synthesis is performed, which also appends the second PCR anchor to cDNA; (5) all the cells are pooled together and overloaded into a chromium scATAC-seq chip, followed by within-droplet cell lysis and adding of a droplet-specific indexed p5 adapter (BC3) with linear amplification; (6) droplets are aliquoted to tubes and DNA products from each tube are purified and divided into two parts for separate transcriptome and accessible chromatin library construction using corresponding primers (BC4); (7) finally, joint profiles of chromatin accessibility and gene expression with the same barcode combination are identified computationally from the same single cell.

In comparison with combinatorial indexing-based high-throughput methods (e.g., Paired-seq and SHARE-seq), Parallel-seq is distinct in four notable ways. First, Parallel-seq minimizes the number of splint adapter ligation rounds, which attenuates both the loss of ligated molecules and the damage to cells that occurs with each successive ligation. Second, by incorporating second-strand synthesis instead of template switching, Parallel-seq increases cDNA library construction efficiency.[20] Third, the use of in-droplet linear amplification obviates the need for pre-amplification purification, which substantially reduces both material loss and also amplification bias as compared with PCR used by other methods. Finally, the use of a general permeabilization protocol (adapted from Omni-ATAC[21]) enables Parallel-seq to analyze samples containing a broad range of cell types, making it especially suitable for investigating complex tissue samples.

Collectively, these improvements allow the reliable and sensitive profiling of chromatin accessibility and gene expression in the same single cells. Moreover, Parallel-seq exploits overloading on the chromium system to enable ultra-high-throughput and cost-effective profiling with a low expected collision rate.[19] In principle, the barcoding space after four rounds of barcoding is $\sim 1.47 \times 10^{10}$ ($96 \times 96 \times 100{,}000 \times 16$), which endows Parallel-seq the potential to assess more than 1 million cells in one experiment.

Parallel-seq obtains high-quality joint scATAC-and-scRNA profiles in cell lines

We benchmarked Parallel-seq's performance using a mixture of K562 (human), HEK293T (human), GM12878 (human), and NIH/3T3 (murine) cells. We obtained joint scATAC-and-scRNA profiles for 3,010 cells, with a median RNA unique molecular identifier (UMI) ranged from 3,636 (GM12878) to 9,125 (HEK293T) and a median number of unique ATAC reads in peaks ranged from 3,807 (GM12878) to 13,992 (K562) after stringent quality filtering (Figures S1A and S1B; Table S1; STAR Methods). All reads were assigned to 802 mouse cells and 2,208 human cells, with very few doublets, indicating a low collision rate of 0.1%$\sim$0.2% (Figure 1B).

The aggregated data of Parallel-seq's scATAC libraries exhibited a typical size distribution of nucleosome binding pattern for ATAC-seq insert fragments[22] (Figure S1C) and was highly enriched around the transcription start sites (TSSs), with TSS enrichment scores ranged from 9.93 (K562) to 12.03 (GM12878) (Figure S1D). Further, the aggregated Parallel-seq's scATAC-and-scRNA data showed high concordance with published bulk RNA sequencing (RNA-seq) ($R = 0.81$), bulk ATAC-seq ($R = 0.83$),[23] scRNA-seq ($R = 0.77$),[24] and scATAC-seq ($R = 0.74$)[25] datasets for K562 cells, respectively (Figures 1C and S1E–S1H).

We compared Parallel-seq with competing methods, evaluating data quality (Figures 1D and S1I–S1N; Table S1), throughput (Table S2), and costs (Table S3) on multiple cell lines. Taking K562 cells as an example, for scRNA data, we found that the median numbers of RNA UMIs and RNA-detected genes obtained by Parallel-seq were 7,014 and 3,181, which were similar to that obtained by 10× Multiome (7,523 and 2,865) and were substantially higher than that of ISSAAC-seq (4,830 and 2,307), SHARE-seq (248 and 212), and SNARE-seq (412 and 329). For scATAC data, we found that the median number of unique ATAC reads in peaks obtained by Parallel-seq is 13,992, which was lower than 10× Multiome (20,191) but higher than ISSAAC-seq (11,361), SHARE-seq (292), and SNARE-seq (507).

By combining droplet overloading with combinatorial indexing, Parallel-seq allows analysis for dozens of samples in one experiment and enables an ultra-high analysis throughput, much higher than 10× Multiome and ISSAAC-seq (200,000 cells vs. 10,000 cells) (Table S2). It bears emphasis that, Parallel-seq achieved high data quality with very low cost, with an expense of $\sim$1/30 ($0.013/cell) of 10× Multiome ($0.38/cell) and $\sim$1/15 of ISSAAC-seq (>$0.2/cell) (Table S3). Overall, these results demonstrate Parallel-seq's capacity to achieve high data quality and ultra-high-throughput at a remarkably low cost.

Parallel-seq obtains high-quality joint scATAC-and-scRNA profiles for lung tumor samples

Much effort has been dedicated to characterizing the composition and states of cell types within the heterogeneous microenvironment of human lung tumors using scRNA-seq.[26–30] We applied Parallel-seq to 40 clinical samples (22 lung tumor samples and 18 tumor-adjacent samples) from 22 lung tumor

Please cite this article in press as: Tang et al., Joint analysis of chromatin accessibility and gene expression in the same single cells reveals cancer-specific regulatory programs, Cell Systems (2025), https://doi.org/10.1016/j.cels.2025.101266

CellPress
OPEN ACCESS

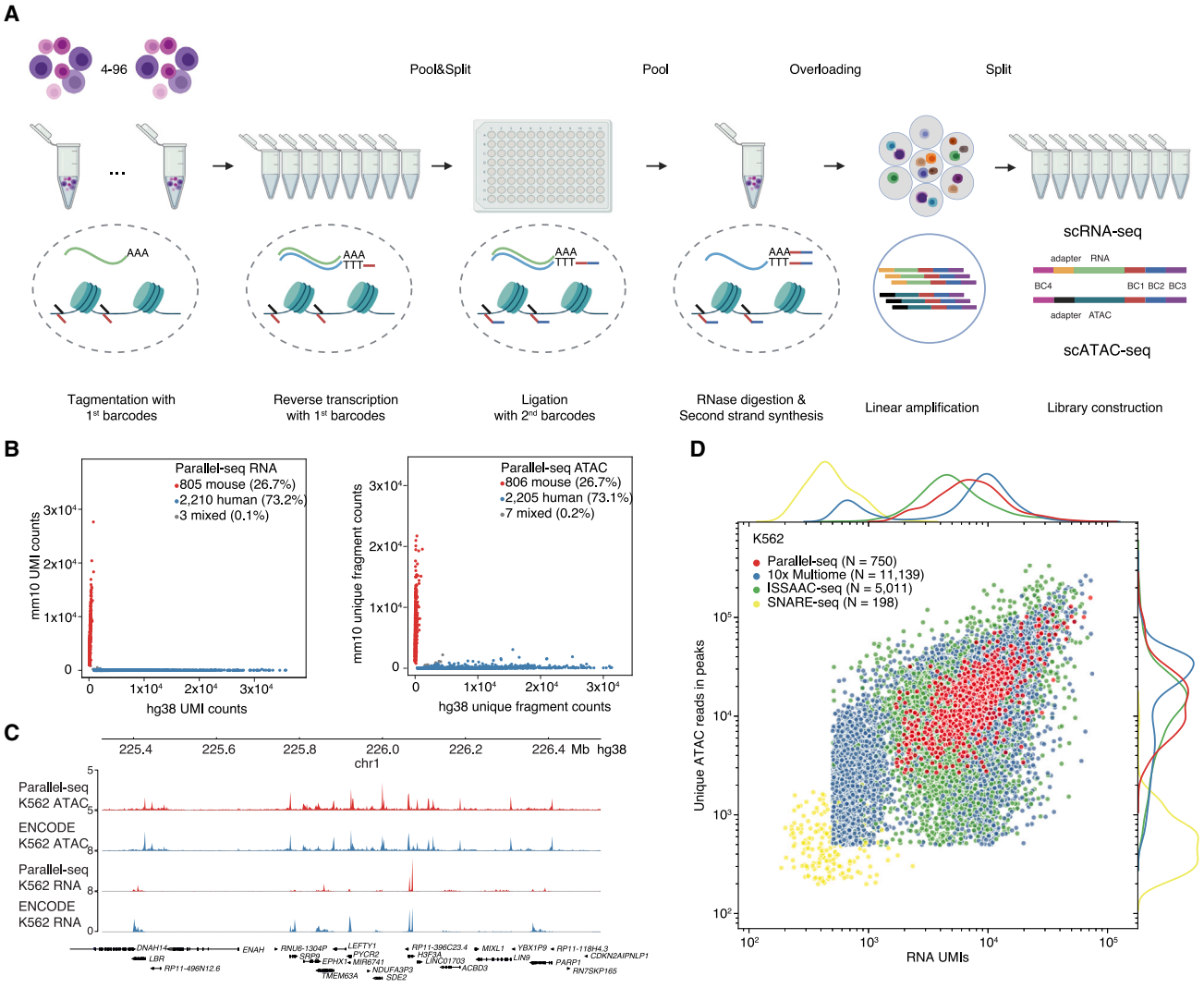


Figure 1. Parallel-seq enables ultra-high throughput simultaneous profiling of accessible chromatin and gene expression in single cells

(A) Parallel-seq combines combinatorial indexing and droplet overloading to measure chromatin accessibility and gene expression in the same single cells.

(B) The number of UMI counts of scRNA (left) and unique fragment counts of scATAC (right) data, mapped to the human and mouse genomes. This experiment examined a mixture of K562 (human), HEK293T (human), GM12878 (human), and NIH/3T3 (murine) cells.

(C) The aggregated single-cell chromatin accessibility and gene expression profiles from a Parallel-seq analysis accurately reproduced the bulk ATAC-seq (ENCFF048XLS) and nucleus RNA-seq (ENCFF631TDY) profiles in K562 cells from ENCODE, respectively.

(D) Scatterplot showing the number of unique ATAC reads in peaks and RNA UMIs per cell for Parallel-seq, 10× Multiome, ISSAAC-seq, and SNARE-seq in K562. Every dot represents a single cell, with aggregated distribution shown along the x and y axes. We only compared the droplet version of ISSAAC-seq in our study. Owing to the inability to match the barcodes between scATAC and scRNA in the SHARE-seq dataset, we were unable to include the SHARE-seq dataset in the figure.

patients (Figures 2A and S2A). After data quality filter (STAR Methods), we obtained 206,263 high-quality joint scATAC-and-scRNA profiles with 1,050 median number of unique ATAC reads in peaks and 1,233 median RNA UMIs (Figures 2B and S2A).

We performed clustering analysis based on the joint scATAC-and-scRNA profiles and revealed 8 major clusters of cells (Figure 2C; STAR Methods). Cells from different tissues and patients exhibit a well-mixed distribution of normal cell types (Figures S2B and S2C), demonstrating the advantage of Parallel-seq in profiling multiple samples in a single experiment to mitigate batch correction. We annotated these clusters using marker gene expression and identified T cells, B cells, plasma cells, macrophages, dendritic cells, epithelial cells, endothelial cells, and fibroblasts (Figure S2D). Almost all (99%) of the fluorescence-activated cell sorting (FACS)-sorted CD45+ and CD3+ cells were accurately annotated as immune cells and T cells, respectively (Figures S2E and S2F; STAR Methods), demonstrating that our workflow for cell clustering and annotation based on marker genes is consistent with that based on sorted cell surface markers. We also conducted clustering exclusively based on either Parallel-seq's scATAC profiles or scRNA profiles (Figures S2G and S2H). We then compared the results obtained

Please cite this article in press as: Tang et al., Joint analysis of chromatin accessibility and gene expression in the same single cells reveals cancer-specific regulatory programs, Cell Systems (2025), https://doi.org/10.1016/j.cels.2025.101266

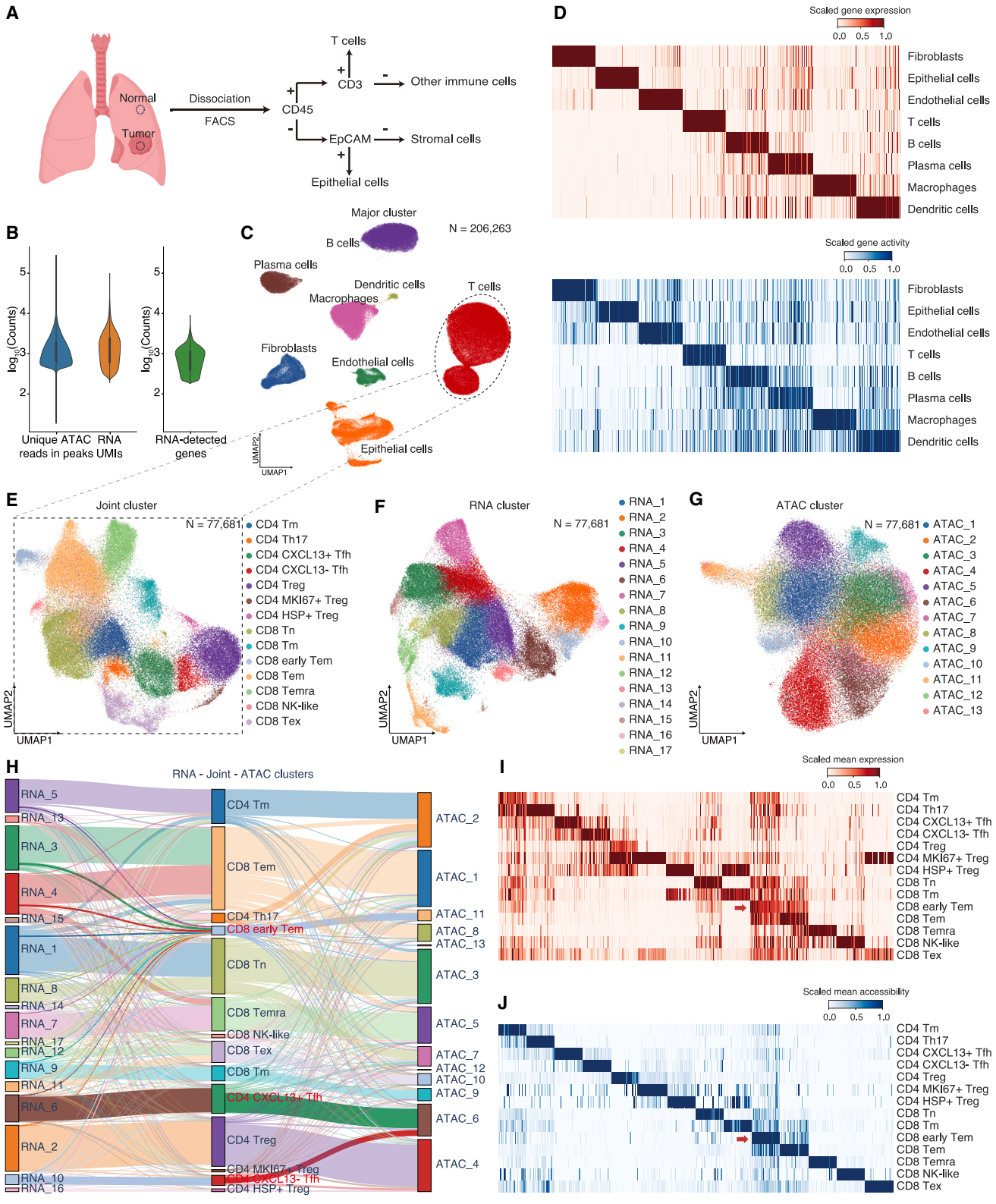


Figure 2. Parallel-seq obtained high-quality joint profiles and revealed distinct T cell subtypes in lung tumor microenvironment

(A) The overall experimental design for the application of Parallel-seq to characterize the microenvironment of human lung tumor. The tumor and paired tumor-adjacent samples were collected for each patient. After dissociation, the single cells were sorted into four groups based on their cell surface markers, and their abundance was balanced for the input of the experiment.

(legend continued on next page)

Please cite this article in press as: Tang et al., Joint analysis of chromatin accessibility and gene expression in the same single cells reveals cancer-specific regulatory programs, Cell Systems (2025), https://doi.org/10.1016/j.cels.2025.101266

from the three different approaches and found that the major cell-type annotations derived from the joint analysis exhibited good agreement[31] with those obtained exclusively from Parallel-seq's scATAC profiles or scRNA profiles (adjusted Rand index [ARI] score: 0.680 or 0.793. Figures S2I–S2L).

We performed an analysis to identify cell-type "signature genes" that exhibited significantly increased expression within each cell type comparing with all other cell types (STAR Methods). We calculated gene activity scores[32] from Parallel-seq's scATAC data for the top 100 signature genes of each major cell type ranked by fold changes of expression level (STAR Methods). We observed that these scores showed high concordance with the gene expression profile for each cell type (Figures 2D and S3A). Specifically, the gene activity scores of the marker gene showed very specific patterns in their matched cell types compared with the others (Figure S3B). These findings supported the above observation of the overall congruence between Parallel-seq's scATAC and scRNA profiles for each major cell type. We also calculated transcription factor (TF) motif activity from Parallel-seq's scATAC data for each cell type with chromVAR.[33] We identified many well-known cell-type-specific TFs, such as TBX21 in T cells,[34,35] POU2F2 in B and plasma cells,[36] SPIB in macrophage cells,[37] and SOX9 in endothelial cells[38,39] (Figure S3C). Interestingly, we observed that the specificity of TF gene expression was markedly lower than that of TF motif activity (Figure S3D). This divergence suggests a temporal disconnect, potentially arising from the fact that TF gene expression often precedes the establishment of motif accessibility.[40]

Joint clustering analysis reveals T cell subtypes of distinct gene expression and chromatin accessibility marks

Tumor-infiltrating T cells are a hallmark of the cancer immune surveillance system, and characterization of T cells in the TME has been used for both prognostic and therapeutic applications.[30] We focused on the T cell populations in our samples and identified 14 T cell subtypes, based on clustering of joint scATAC-and-scRNA profiles (Figures 2E and S3E–S3J; STAR Methods).

In contrast to the strong agreement observed for major cell-type annotations, we observed substantial differences for T cell subtype annotations based on Parallel-seq's joint profiles and based on Parallel-seq's scATAC profiles alone (ARI score: 0.400) or scRNA profiles alone (ARI score: 0.494) (Figures 2E–2H, S3K, and S3L). For instance, we noted two subtypes of T follicular helper (Tfh) cells (i.e., the CD4 CXCL13+ Tfh cells and the CD4 CXCL13− Tfh cells) in the clustering results from Parallel-seq's joint profiles, which both aligned with the ATAC_6 cluster of Parallel-seq's scATAC profiles (Figures 2H, S3K, and S3L).

Notably, we observed a subtype of previously reported "CD8 early T effector memory (Tem) cells"[41] in the clustering results from Parallel-seq's joint profiles. These CD8 early Tem cells came from four distinct clusters of Parallel-seq's scRNA profiles (e.g., the RNA_1, RNA_3, RNA_4, and RNA_5 clusters), but most of these cells (82%) formed one cluster of Parallel-seq's scATAC profiles (Figures 2H, S3K, and S3L). We identified signature genes, as well as signature chromatin accessibility peaks, for each T cell subtype using the above-mentioned pipeline (Figures 2I and 2J; STAR Methods). Intriguingly, we found that the top 30 signature genes of CD8 early Tem were shared by the other CD8 Tem cells (Figure 2I), while the top 30 signature accessible peaks were specifically open in the CD8 early Tem cells (Figure 2J). This suggests that although these CD8 early Tem cells exhibit similar gene expression to the other CD8 Tem cells, they possess a distinct chromatin accessibility state.

In sum, these results illustrate that joint scATAC-and-scRNA profiles can reveal unique gene expression and/or chromatin accessibility marks in some cell populations (e.g., CD4 CXCL13+ Tfh cells, CD4 CXCL13− Tfh cells, and CD8 early Tem cells) that would likely elude detection in analyses relying exclusively on RNA data or ATAC data.

Parallel-seq joint analysis reveals aggressive tumor cell subclones through improved CNV detection

Copy-number variations (CNVs) describe duplications or losses of large genomic segments >1,000 bp in size and are a major driving force underlying cancer development.[42] Traditionally, scATAC-seq and scRNA-seq data have been employed independently to identify CNVs in cancer research.[3,43] We here compared CNVs calling based on Parallel-seq's joint scATAC-and-scRNA profiles with those based on Parallel-seq's scATAC

(B) Violin plot of the numbers of unique ATAC reads in peaks and RNA UMIs (left), and RNA-detected genes (right) per single cell of lung tumor samples (N = 206,263). White dot: median; black bar: interquartile range (IQR).

(C) UMAP visualization of embedding of joint scATAC-and-scRNA profiles of the Parallel-seq lung tumor dataset for major clusters (N = 206,263). Joint scATAC-and-scRNA profiles were processed with Seurat v4 for weighted nearest neighbor (WNN) analysis. Clusters of joint scATAC-and-scRNA profiles were annotated based on marker gene expression.

(D) Heatmap showing the standardized mean gene expression (top) and gene activity (bottom) of 100 top-ranked cell-type-specific genes in the indicated major cell types.

(E–G) UMAP visualization of embedding of joint scATAC-and-scRNA profiles (E), scRNA (F), and scATAC (G) of the Parallel-seq lung tumor dataset for T cells (N = 77,681). RNA profiles were processed with Seurat v4 for principal-component analysis (PCA); ATAC profiles were processed by single-cell ATAC-seq analysis via latent feature extraction (SCALE) for dimension reduction, and joint scATAC-and-scRNA profiles were processed with Seurat v4 for WNN analysis. Clusters of joint scATAC-and-scRNA profiles were annotated based on marker gene expression. Identified subtypes include CD4 T memory (Tm) cells, CD4 T helper 17 (Th17) cells, CD4 CXCL13+ T follicular helper (Tfh) cells, CD4 CXCL13− Tfh cells, CD4 T regulatory (Treg) cells, CD4 MKI67+ Treg cells, CD4 HSP+ Treg cells, CD8 T naive (Tn) cells, CD8 T memory (Tm) cells, CD8 early T effectory memory (Tem) cells, CD8 Tem cells, CD8 T effector memory re-expressing CD45RA (Temra) cells, CD8 natural killer (NK)-like cells, and CD8 T exhausted (Tex) cells.

(H) Sankey plot shows the correspondence between T cell subtypes identified using joint scATAC-and-scRNA profiles, as well as the clusters exclusively identified within scRNA or scATAC profiles.

(I and J) Matrixplot shows the top 30 signature genes (by gene expression fold-of-change) (I) and accessible peaks (by chromatin accessibility fold-of-change) (J) of T cell subtypes. The red arrow highlights the signature genes and signature accessible peaks of the CD8 early Tem cells. CD8 early Tem signature genes are widely expressed in other T cell subtypes, especially CD8 Tn and CD8 Tem cells, while signature accessible peaks are specifically open in CD8 early Tem cells, representing a unique chromatin accessibility state.

Please cite this article in press as: Tang et al., Joint analysis of chromatin accessibility and gene expression in the same single cells reveals cancer-specific regulatory programs, Cell Systems (2025), https://doi.org/10.1016/j.cels.2025.101266

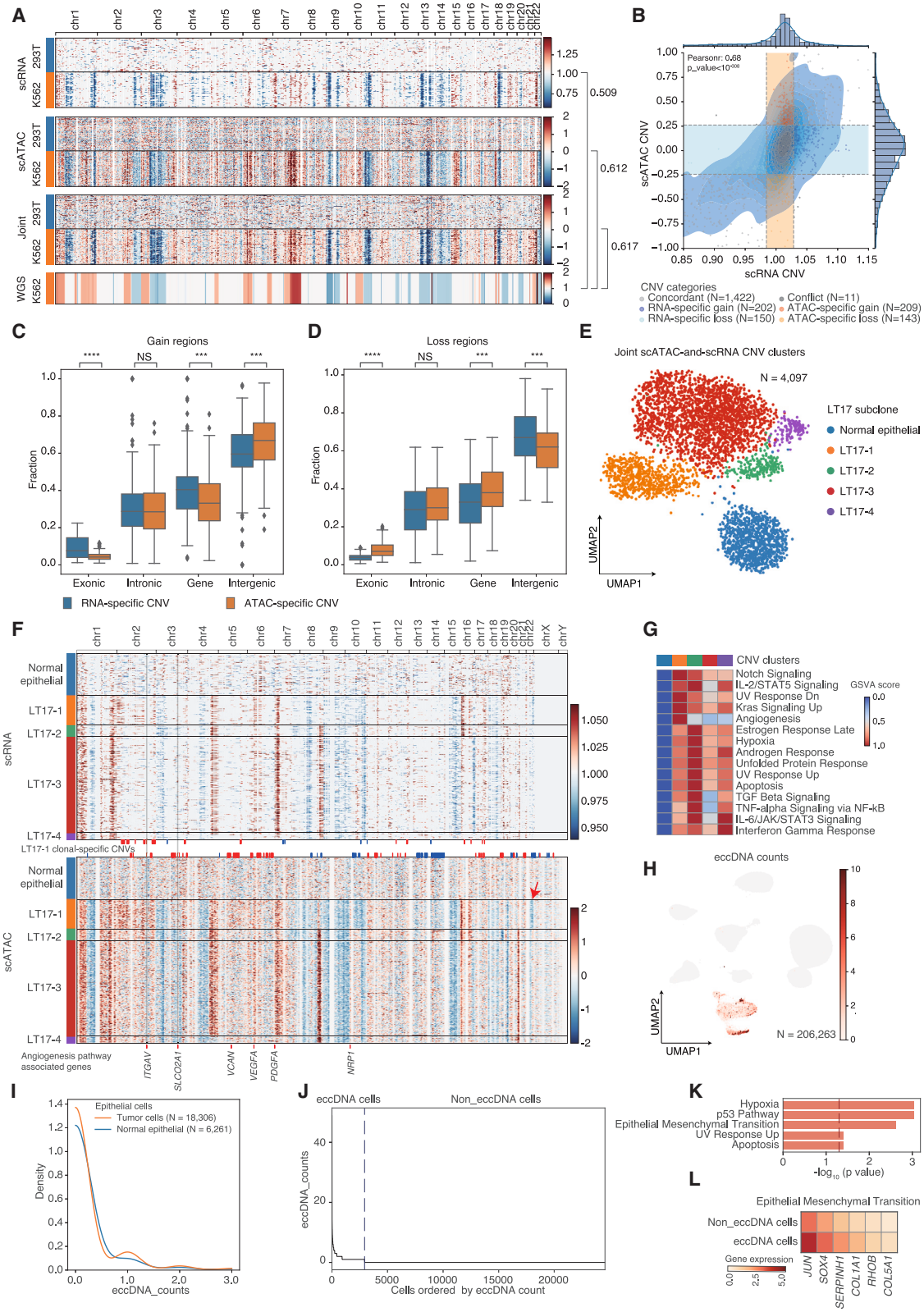


Figure 3. Parallel-seq improves CNV and eccDNA studies by joint analysis of ATAC-seq and RNA-seq data in the same single cell

(A) Comparison of CNVs detected with the examined Parallel-seq derived scRNA, scATAC, and joint scATAC-and-scRNA data (using the CNVs detected from a WGS dataset in the same K562 cell line as a reference). Color bars indicate CNV score ranges. Pearson correlation coefficients are indicated on the right.

(legend continued on next page)

Please cite this article in press as: Tang et al., Joint analysis of chromatin accessibility and gene expression in the same single cells reveals cancer-specific regulatory programs, Cell Systems (2025), https://doi.org/10.1016/j.cels.2025.101266

profiles alone or scRNA profiles alone (Figure 3A; STAR Methods). For benchmark, we used the known CNVs detected by CNVpytor[44] from a whole genome sequencing (WGS) dataset for K562 cells (GSE176762)[23] as a reference. We observed that, overall, the CNV gain and loss patterns obtained from the three different approaches were all well correlated across the genome (Pearson r = 0.68, Figure 3B). Yet, the CNVs detected using the joint profiles were (1) the most similar to the reference CNVs, as indicated by the highest Pearson correlation coefficients and the smallest Euclidean distance (Figures 3A and S4A), and (2) achieved the highest sensitivity, specificity, and F1 scores overall, especially for CNV loss regions (Figures S4B–S4J).

Notably, some CNVs were only identified from Parallel-seq's scRNA profiles and some were only identified from the scATAC profiles. We respectively termed these CNVs as "RNA-specific" and "ATAC-specific" CNVs (Figure 3B). We discovered that, for CNV gains, RNA-specific CNVs tended to co-occur with exonic regions and gene regions, while ATAC-specific CNVs tended to co-occur with intergenic regions, but for CNV losses, the results were opposite (Figures 3C and 3D). Overall, these results support that integration of scATAC-and-scRNA data improves CNVs detection.

Next, we leveraged Parallel-seq's joint scATAC-and-scRNA profiles to identify CNVs in 22 lung tumor samples (Figures S5A–S5E). Normal epithelial cells were consistently clustered together, whereas malignant cells formed patient-specific clusters, highlighting the inter-tumor heterogeneity among non-small cell lung cancer (NSCLC) patients. Tumor cells from patient LT17 were divided into two distinct CNV clusters, designated Malignant 3 and Malignant 12, which were clearly separate from the tumor cells of other patients (Figures S5A–S5G). Further analysis of LT17 epithelial cells based on CNV profiles (Figures S5H–S5K) identified four tumor subclones[45] (Figures 3E, 3F and S5L; STAR Methods). To investigate the functional differences among these subclones, we performed gene-set variation analysis (GSVA)[46] to examine oncogenic signatures and identified distinct cancer hallmark pathways enriched for these subclones (Figure 3G).

We were particularly interested in the "LT17-1" subclone, which exhibited a highly enriched oncogenic signature associated with angiogenesis compared with other subclones (Figure 3G). We noted higher expression levels of several genes associated with the angiogenesis pathway in this LT17-1 subclone compared with the other subclones (Figure S5M). Specifically, *SLCO2A1* were linked to an ATAC-specific CNV gain in chromosome 3q, while *ITGAV* were linked to an RNA-specific CNV gain in chromosome 2q, which both are exclusively identified in LT17-1 subclone (Figure 3F). Interestingly, elevated levels of *ITGAV* predicted unfavorable survival outcomes in lung adenocarcinomas (LUAD) (Figure S5N; STAR Methods). Moreover, we noted specific CNV loss on chromosome 22q for this LT17-1 subclone (Figure 3F). The allelic loss of chromosome 22q is commonly observed across various cancer types.[47–51] Particularly, in advanced-stage NSCLC, the incidence of allelic loss in the 22q region is significantly higher compared with early-stage NSCLC.[52] These findings collectively suggest that the LT17-1 subclone is in a more aggressive state during the development of lung cancer.

Parallel-seq joint analysis characterizes the eccDNA heterogeneity of tumor cells

Extrachromosomal circular DNAs (eccDNAs) were recently found to be abundant in multiple human tumor cells and were shown to promote cancer development by amplifying the expression of certain oncogenes.[53–55] As ATAC-seq can cleave DNA from eccDNAs,[56] we reasoned that Parallel-seq's joint profiling data obtained could identify eccDNAs in lung tumor cells and support association analysis of gene expression in the same cell. Pursuing this, we used Circle_finder[56] to identify eccDNAs based on the Parallel-seq's scATAC profiles from the lung tumor samples. We confirmed that the length distribution of the identified eccDNA was similar to previous reports[56]

(B) Scatter-density plot showing the CNV scores calculated from Parallel-seq derived scRNA and scATAC data in K562 cells for sliding windows along the whole genome. The two vertical dashed lines correspond to the 20th and 80th percentiles of CNV scores from scATAC data, while the two horizontal dashed lines correspond to the 20th and 80th percentiles of CNV scores from scRNA data. The orange bins are CNVs uniquely identified from scATAC data (termed as "ATAC-specific CNV"), while the blue bins are CNVs uniquely identified from scRNA data (termed as "RNA-specific CNV"). Pearson correlation coefficient and *p* value are shown.

(C and D) The fraction of exonic, intronic, gene, and intergenic regions of ATAC-specific CNV and RNA-specific CNV gain (C) and loss (D) regions. Each box plot displays the median (center line), interquartile range (IQR, box), and data points beyond 1.5× IQR (whiskers). Statistical significance was calculated by Wilcoxon rank-sum test, NS, not significant; $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, $^{****}p < 0.0001$. $N_{\text{RNA-specific gain}} = 202$, $N_{\text{RNA-specific loss}} = 150$, $N_{\text{ATAC-specific gain}} = 209$, $N_{\text{ATAC-specific loss}} = 143$.

(E) UMAP visualization of embedding of joint CNVs called from joint scATAC-and-scRNA data of Parallel-seq for epithelial cells ($N = 4{,}097$) of LT17 patient with the annotation of subclones.

(F) Heatmap of the scRNA (top) and scATAC (bottom) CNV scores of LT17 epithelial cells, grouped by subclones. Color bars on the right of each heatmap indicate the CNV score range. Color-coded short lines indicate ATAC-specific or RNA-specific gains (red) or losses (blue) exclusively identified within the LT17-1 subclone. The gray lines highlight clonal-specific CNVs in LT17-1 harboring specific angiogenesis pathway genes. The highly expressed angiogenesis pathway genes in LT17-1 are indicated at the bottom. Red arrowheads highlight specific CNV losses on chromosome 22q for LT17-1 subclone.

(G) Heatmap of the standardized GSVA scores of the enriched cancer hallmark pathways for each subclone of LT17.

(H) UMAP visualization of embedding of joint scATAC-and-scRNA profiles of the Parallel-seq lung tumor dataset, colored by the number of eccDNAs in tumor cells and normal epithelial cells. The color bar indicates eccDNA counts.

(I) Density plot shows the distribution of the number of eccDNAs identified in tumor cells ($N = 18{,}306$) and normal epithelial cells ($N = 6{,}261$). *p* value = 8.85×10^{-17}, Mann-Whitney U test.

(J) The distribution of eccDNA number detected in tumor cells. The tumor cells were divided into eccDNAs detected (eccDNA cells) and eccDNAs non-detected (Non_eccDNA cells) groups.

(K) Cancer hallmark pathways enriched within eccDNA signature genes. The red line indicates a cutoff of *p* value < 0.05 calculated by GSEApy.

(L) Heatmap showing the expression based on log transformed transcripts per million (TPM) of the eccDNA signature genes of the epithelial-mesenchymal transition pathway in eccDNA and Non_eccDNA cell groups. The color bar indicates the log transformed TPM values.

Please cite this article in press as: Tang et al., Joint analysis of chromatin accessibility and gene expression in the same single cells reveals cancer-specific regulatory programs, Cell Systems (2025), https://doi.org/10.1016/j.cels.2025.101266

(Figure S5O), and also found, as expected, that tumor cells possessed more eccDNAs compared with normal epithelial cells (p value = 8.85 $\times 10^{-17}$) (Figures 3H and 3I).

We classified the tumor cells into two groups: cells with eccDNAs detected (“eccDNA cells”) and cells without eccDNAs (“Non_eccDNA cells”) (Figure 3J; STAR Methods). We then identified genes that were differentially expressed in the eccDNA cells compared with the Non_eccDNA cells (“eccDNA signature genes”). We noted that the eccDNA signature genes were enriched for known cancer hallmark pathways (Figure 3K), including the apoptosis and “UV response up” pathways, a finding consistent with a report that eccDNAs are apoptotic products and that UV treatment results in strong induction of eccDNA production.[57] Of particular note, we observed an upregulated expression of *SOX4*, a well-established epithelial-mesenchymal transition (EMT) master regulator[58] known to promote metastasis,[59,60] in eccDNA cells (Figure 3L). These results illustrate how Parallel-seq joint analysis can reveal the heterogeneity of eccDNA in tumor cells and support investigations of their potential functions in cancer progression.

Parallel-seq joint profiles predict cell-type-specific gene expression regulatory events

The ability to capture both chromatin accessibility and gene expression in the same single cells provide a unique opportunity to predict gene expression regulatory events mediated by *cis*-regulatory elements.[13,14] We developed an analysis framework based on mutual information (MI)[61] to link accessible chromatin peaks to each gene (STAR Methods). Briefly, we calculated MI values between peak accessibility and gene expression and identified those peak-gene pairs with significantly high MI as “peak-gene associations.” We evaluated the analysis pipeline using K562 datasets generated through Parallel-seq, 10× Multiome, and ISSAAC-seq. With a limited number of cells, 10× Multiome and ISSAAC-seq identified more peak-gene associations than Parallel-seq, likely due to their higher overall data quality. However, we found that increasing the cell count enhances the number of identified peak-gene associations (Figure S6A). Most chromatin peaks in these associations were located close to the TSSs (Figure S6B). Interestingly, we found that most chromatin peaks were associated with only one gene, whereas most genes were associated with more than one peak (Figures S6C and S6D). A majority (93.7%) of the predicted associations involved distal chromatin peaks overlapped with *cis*-regulatory elements annotated by the ENCODE project,[62] suggesting that these associated peaks plausibly represent putative enhancers. To validate the predicted associations, we compared them with the publicly available chromatin loop datasets for the corresponding cell lines obtained with HiChIP using an antibody against H3K27ac.[63] Compared with randomly shuffled associations, we detected enrichment for chromatin loops identified by HiChIP within these predicted peak-gene associations (Figure S6E).

We applied the analysis on the Parallel-seq lung tumor dataset and identified 466,959 peak-gene associations. We then defined cell-type-specific peak-gene associations based on whether an association involves (1) genes specifically expressed in particular cell types or with (2) chromatin peaks specifically detected in particular cell types (consequently, the chromatin peaks were defined as cell-type-specific enhancers). We identified tens of thousands of cell-type-specific peak-gene associations and associated enhancers (Figures S6F–S6H; Table S4). Interestingly, the cell-type-specific enhancers showed higher cell-type specificity than genes or promoters, which is consistent with a previous report[64] (Figure S6I).

The lung tumor samples included 3 patients with lung squamous cell carcinoma (LUSC) and 16 patients with LUAD. To investigate the distinct regulatory programs underlying these two patient groups, we compared the predicted peak-gene associations within the tumor cells of the two lung tumor subtypes. This analysis identified several peak-gene associations, which only occur in LUSC tumor cells, may promote *MIR205HG* expression (Figure 4A; Table S4). For context, it has been reported that *MIR205HG* is highly expressed in LUSC clinical samples with expression level positively correlated with the degree of malignancy.[65] Suppression of *MIR205HG* expression in LUSC cell lines reduces cell proliferation, migration, and EMT progression, while promoting cell apoptosis.[66]

We found a set of tumor-cell-specific enhancers that positively correlate with *MIR205HG* expression in LUSC tumor cells. Both *MIR205HG* gene specifically expressed, and these enhancers specifically opened in LUSC tumor cells (Figure 4A). Among these enhancers, some are supported by previous H3K27ac HiChIP analysis of lung squamous cancer cell lines.[67] We performed a motif analysis and discovered that these enhancers are enriched with the TP63-binding motif (Figure 4B; Table S5). A previous study on miR-205 in prostate cancer indicated TP63 binding to the chr1:209415937-209416437 region.[68] These results, together with the elevated expression of *TP63* in the three LUSC patients (Figure 4A), suggest a potential mechanism through which TP63 may contribute to the elevated *MIR205HG* expression in LUSC patients.

We also analyzed T-cell-subtype-specific regulatory events (Figure S6J; Table S6) and identified peak-gene associations that may regulate *BACH2* expression in CD8 Tn cells (Figure 4C). For context, *BACH2* functions in maintaining T cells in a naive state[69] and regulates the differentiation of naive $CD8^+$ T cells into effector and memory cells, thereby ensuring an effective and balanced immune response.[70] Our analysis found a set of enhancers that positively correlated with *BACH2* expression. We found an ETS-1-binding motif was enriched in these enhancers (Figure 4D; Table S7); this finding, along with the expression of *ETS-1* in CD8 Tn cells, suggests that ETS-1 may promote *BACH2* expression in CD8 Tn cells. It bears note that CD8 Tn marker genes, including *BACH2*, predicted favorable prognostic results for LUAD in The Cancer Genome Atlas (TCGA) cohort[71] (Figure 4E; STAR Methods). Interestingly, *BACH2* overexpression skews exhaustion phenotypes toward stem-like CD8 T cells states.[72] These findings suggest that targeting these regulatory events could be harnessed to enhance the antitumor responses in chimeric antigen receptor (CAR)-T cell immunotherapy for lung tumors.

Parallel-seq joint profiles predict enhancer mutations that affect target gene expression in tumor cells

We further leveraged Parallel-seq data for the identification of candidate somatic enhancer mutations that may differentially regulate target gene expression. We predicted such “enhancer mutations” based on (1) comparative genomic analysis searching

Please cite this article in press as: Tang et al., Joint analysis of chromatin accessibility and gene expression in the same single cells reveals cancer-specific regulatory programs, Cell Systems (2025), https://doi.org/10.1016/j.cels.2025.101266

CellPress
OPEN ACCESS

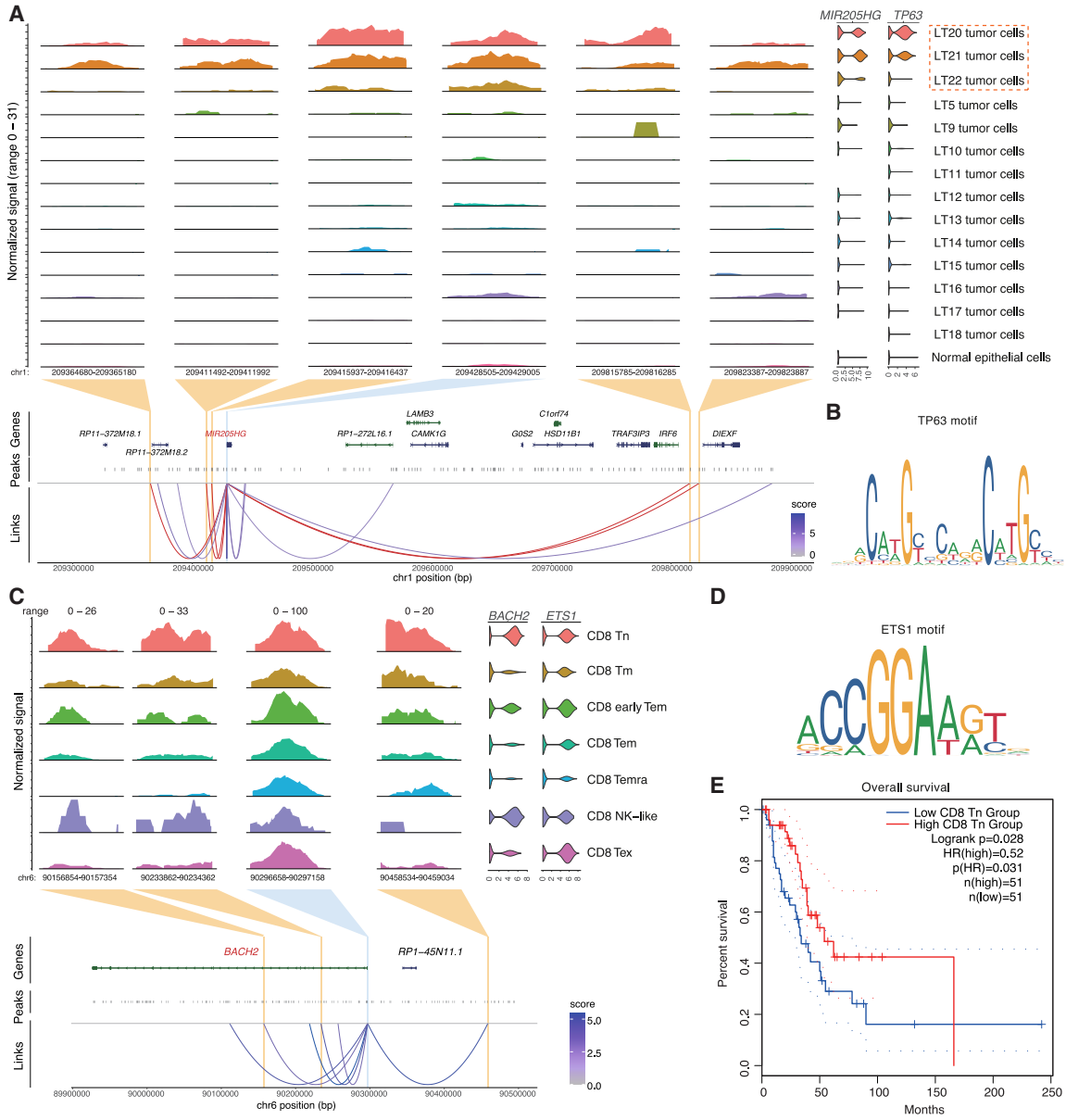


Figure 4. Parallel-seq joint profiles enable detection of cell-type-specific regulatory events in the lung tumor microenvironment

(A) The upper tracks show the aggregated scATAC profiles of the promoter (blue triangle) and associated enhancers (yellow triangle) (chr1: 209364680–209365180, chr1: 209411492–209411992, chr1: 209415937–209416437, chr1: 209815785–209816285, and chr1: 209823387–209823887) of *MIR205HG* in different tumor cells. The violin plots on the right show the gene expression of *MIR205HG* and *TP63*. The dashed box highlights patients with LUSC. The loops below denote the predicted peak-gene associations for *MIR205HG*, colored by their *Z* scores, wherein peak-gene associations certified by H3K27ac HiChIP of lung squamous cancer cell lines[67] highlighted in red.

(B) The binding motif for the transcription factor TP63 is enriched from the *MIR205HG* enhancers.

(C) The upper tracks show the aggregated scATAC profiles of the promoter (blue triangle) and associated enhancers (yellow triangle) (chr6: 90156854–90157354, chr6: 90233862–90234362, and chr6: 90458534–90459034) of *BACH2* in distinct CD8 T cell subtypes. The violin plots on the right show the gene expression of *BACH2* and *ETS1*. The loops below denote the predicted peak-gene associations for *BACH2*, colored by their *Z* scores.

(D) The binding motif for the transcription factor ETS1 is enriched from the *BACH2* enhancers.

(E) Kaplan-Meier plot for a TCGA LUAD patient dataset, indicating that the CD8-Tn-high subgroup exhibited a more favorable prognosis in terms of overall survival compared with the CD8-Tn-low subgroup. The CD8-Tn high and low groups were defined based on the 75% quantile value of the mean expression levels of *BACH2*, *SELL*, *CCR7*, and *LEF1*. *p* value from univariate Cox regression.

Please cite this article in press as: Tang et al., Joint analysis of chromatin accessibility and gene expression in the same single cells reveals cancer-specific regulatory programs, Cell Systems (2025), https://doi.org/10.1016/j.cels.2025.101266

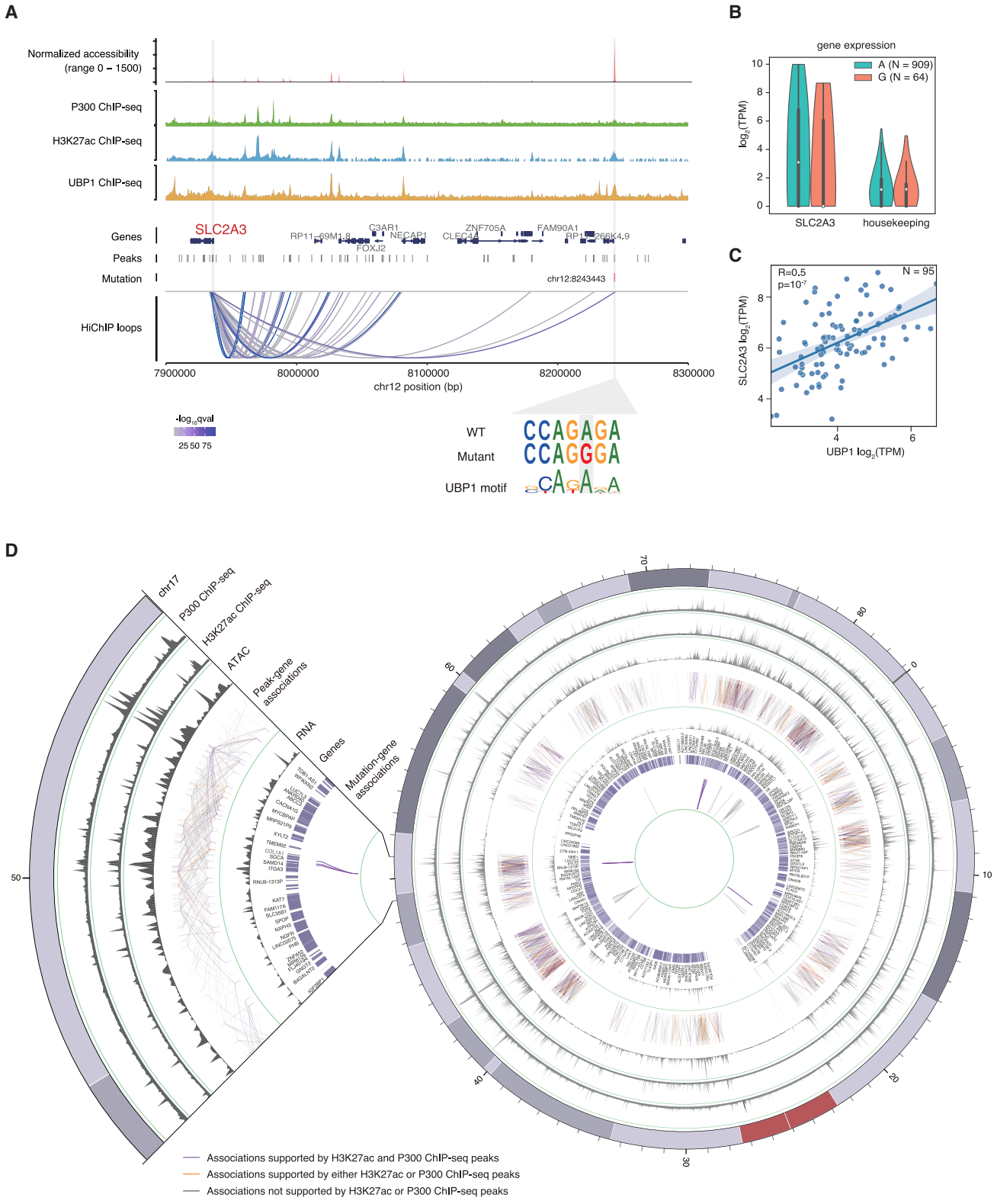


Figure 5. Parallel-seq identifies regulatory enhancer mutations in lung tumors

(A) The tracks show the aggregated data for scATAC profiles of Parallel-seq, P300 ChIP-seq, H3K27ac ChIP-seq and UBP1 ChIP-seq in the regions near SLC2A3. Loops denote H3K27ac HiChIP loops colored by their q values. The sequences of the putative binding site of UBP1 (containing the mutation site chr12: 8243443) are shown, along with the binding motif of UBP1.

(legend continued on next page)

Please cite this article in press as: Tang et al., Joint analysis of chromatin accessibility and gene expression in the same single cells reveals cancer-specific regulatory programs, Cell Systems (2025), https://doi.org/10.1016/j.cels.2025.101266

for differential alleles in tumor vs. normal cells and (2) whether these mutations result in differential expression of the relevant associated target genes (STAR Methods). We refer to the pairing of such mutations and their target genes as "mutation-gene associations" (Figure S7A). In total, we identified 320 and 219 enhancer mutations for LT17 and LT18, corresponding to 884 and 400 mutation-gene associations, respectively (Table S8). Among them, there were 33 mutations and 13 mutation-gene associations shared between the two patients (Figures S7B and S7C). The mutated allele ratio was indeed much higher in tumor cells than in normal cells (Figure S7D). Many mutations did have additional supporting evidence from the H3K27ac HiChIP interaction datasets[63,67] (Figures S7E–S7H, STAR Methods), and many mutation-gene associations involved cancer-related genes, such as chr20:63641960/*DNAJC5*[73] (Figures S7I and S7J).

Among these enhancer mutations identified in LT17 was one mutation chr12:8243443 associated with *SLC2A3* (Figure 5A), which was located in an open chromatin region with a consensus positive signal of H3K27ac chromatin immunoprecipitation sequencing (ChIP-seq) and P300 ChIP-seq, and *SLC2A3* was previously reported as an indicator of poor survival in NSCLC.[74] The mutation from A to G was associated with significantly decreased expression of *SLC2A3* (Figure 5B). We performed a motif search and identified that the UBP1 binding motif was enriched in this loci; this potential UBP1 binding was also supported by a ChIP-seq experiment.[75] Moreover, the expression of *UBP1* and *SLC2A3* are positively correlated (Figure 5C). Together, these imply that UBP1 probably regulates *SLC2A3*, and this mutation may decrease the expression of *SLC2A3* in tumor cells by disrupting the UBP1 binding.

Finally, to visualize the *cis*-regulatory elements and mutations identified in tumor cells of our lung tumor dataset, we prepared a Circos plot comprising tracks for peak-gene associations and for mutation-gene associations, as well as tracks for the scATAC and scRNA of Parallel-seq (Figure 5D). These results highlight the capability of Parallel-seq's joint scATAC-and-scRNA profiles in identifying enhancer mutations that potentially contribute to the dysregulation of target gene expression in tumor cells.

DISCUSSION

Parallel-seq achieves simultaneous measurement of chromatin accessibility and gene expression data in the same single cells. Parallel-seq combines combinatorial indexing with dsc-seq technology (currently the two main approaches for high-throughput single-cell sequencing) using an overloading strategy that has been adopted by the recently developed dsci-ATAC-seq,[17] scifi-RNA-seq,[18] and SCITO-seq.[19] Its design allows Parallel-seq to reduce the number of combinatorial indexing rounds and the consequent loss of target molecules and to reduce damage to cells. This, together with other innovations including the use of second-strand synthesis,[20] in-droplet linear amplification, and a general permeabilization protocol,[21] enables Parallel-seq to reliably obtain high-quality data from complex samples, doing so in an ultra-high-throughput and highly cost-effective fashion.

Methods for joint analysis of chromatin accessibility and gene expression in the same single cells have the potential to reveal the complex regulatory programs of the diverse cells present in tumors. However, it is challenging to generate data of sufficiently high quality for various cell types (immune cells in particular) to enable association analyses between genomic structural variations and mutations and the transcriptome in the same cell. Here, we successfully applied Parallel-seq to 40 lung tumor and tumor-adjacent samples and obtained over 200,000 high-quality joint scATAC-and-scRNA profiles, which allowed us to obtain a more complete picture of for example the CNVs and eccDNAs within single tumor cells, as inferred from both the chromatin accessibility and gene expression profiles.

Perhaps most interestingly, the joint scATAC-and-scRNA profiles obtained by Parallel-seq enabled us to deduce cell-type-specific gene expression regulatory events mediated by *cis*-regulatory elements for both immune cells (e.g., T cells) and tumor cells and also supported the identification of mutations located in open chromatin regions that affect gene expression in tumor cells. These example analyses based on Parallel-seq data from lung tumor samples illustrate the power of Parallel-seq as a tool to investigate genetic and epigenetic factors that affect cancer progression (at the cell-type-specific level) and potentially to reveal vulnerable therapeutic targets.

It should be straightforward to expand the capacity of Parallel-seq to encompass other analytical modalities such as DNA methylation,[76] histone modifications,[77] nascent transcriptomes,[78] proteomics,[79] and genetic perturbations.[80] Of particular note, a very attractive and readily extension could be the combination of Parallel-seq with the spatial barcoding[81,82] to enable joint profiling with spatial information. We anticipate that Parallel-seq is especially well suited for large-scale cancer analyses and should greatly facilitate the discovery of recurrent or context-specific genetic and epigenetic regulatory programs within particular cell types in the TME.

Limitations of the study

While this study has produced encouraging results, we acknowledge that there remain opportunities for further refinement.

First, Parallel-seq's coverage on the transcriptome and open chromatin regions is limited. For the scRNA part, the coverage is currently bias to the 3′ end of transcripts, resulting in insufficient coverage of full-length transcripts. This limitation hampers

(B) Violin plot for the expression of *SLC2A3* in epithelial cells, grouped by the allele types at chr12: 8243443 in the LT17 patient. White dot: median; black bar: interquartile range (IQR). Adjusted p value = 0.0075, two-sided unpaired Student's t test were used and p value were corrected by Benjamini-Hochberg method.
(C) Scatterplot of the expression of *UBP1* and *SLC2A3* in the epithelial cells (N = 95) from the LT17 patient. The R value was calculated using Spearman correlation, and the p value indicates the statistical significance of the correlation.
(D) In the Circos plot, which integrates various data tracks from outer to inner circles, the following information is depicted: the P300 and H3K27ac ChIP-seq data from the ENCODE project, aggregated chromatin accessibility profiles, and gene expression profiles obtained from Parallel-seq. The loops present peak-gene associations and mutation-gene associations of chromosome 17 in LT17, with the region chr17: 49,000,000–51,000,000 enlarged. Associations are color-coded in purple, orange, or gray to indicate whether they are supported by H3K27ac and P300 ChIP-seq peaks, supported by either H3K27ac or P300 ChIP-seq peaks, or not supported by either.

Please cite this article in press as: Tang et al., Joint analysis of chromatin accessibility and gene expression in the same single cells reveals cancer-specific regulatory programs, Cell Systems (2025), https://doi.org/10.1016/j.cels.2025.101266

the investigation of alternative splicing events, which were often dysregulated during tumorigenesis,[83] as well as the characterization of T cell receptor (TCR) and B cell receptor (BCR) repertoires for T cells and B cells, respectively, which provides vital information on adaptive immunity in TME.[84] Integrating technologies such as EasySCI-RNA[85] and SHERRY2[86] may help address this issue. For the scATAC part, ~50% of fragments have identical adapters on both ends, which cannot be successfully sequenced.[87] This leads to a natural loss of about 50% of information, significantly affecting the analysis of chromatin accessibility differences. Adopting techniques such as THS-seq[88] for capturing open chromatin regions may yield more comprehensive information.

Moreover, we observed a decrease in the data quality of Parallel-seq in tumor samples compared with cell lines, a trend also seen with other methods (Table S10). To improve data quality in primary samples, several enhancements could be made. These include performing reverse transcription and RNase digestion before ATAC-seq to reduce RNA degradation and prevent Tn5 from cutting RNA/DNA hybrids,[89] as well as increasing the transposase concentration to enhance transposition efficiency.[17]

Additionally, fully utilizing the barcoding space of Parallel-seq requires dozens of samples and a large number of cells for each sample in a single experiment. This can be challenging for clinical samples, which often require sophisticated preparation to be dissociated into cells and/or a long time frame to collect. Alternatively, using nuclei extracted from properly preserved pre-collected samples could mitigate this challenge.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Qiangfeng Cliff Zhang (qczhang@tsinghua.edu.cn).

Materials availability

This study did not generate new materials.

Data and code availability

- The raw FASTQ files of Parallel-seq data derived from human samples have been deposited in the Genome Sequence Archive (GSA) for Human at the National Genomics Data Center (NGDC), under accession number: HRA006440 and HRA001829. The processed data have been deposited at OMIX database at NGDC: OMIX009208. Both raw and processed data were available upon request for scientific research, in compliance with the national legal guidelines outlined in the Regulations of the People's Republic of China on the Administration of Human Genetic Resources, ensuring the protection of human patient privacy. The publicly available dataset analyzed in this paper with their accession number are listed in the key resources table.
- All original code has been deposited at GitHub: https://github.com/yqshao17/Parallel-seq and Zenodo[90]: https://doi.org/10.5281/zenodo.10728191 with the MIT License and is publicly available as of the date of publication.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

This work was supported by the National Key Research and Development Project of China (grant no. 2022YFF1203100 to Q.C.Z.), the National Natural Science Foundation of China (grants no. 32230018 and 32125007 to Q.C.Z.). Q.C.Z thanks for support from the Beijing Advanced Innovation Center for Structural Biology and the Tsinghua-Peking Joint Center for Life Sciences. We thank members of the Q.C.Z labs for helpful discussions. We would like to acknowledge the State Key Laboratory of Membrane Biology, Facility and high-throughput screening (HTS) Core Facility, Center of Pharmaceutical Technology at Tsinghua University for technical assistance.

AUTHOR CONTRIBUTIONS

Q.C.Z. and C.F. supervised the project. L.T. designed the protocol. L.T. and J.Z. developed the protocol with assistance from Y.W. L.T. generated datasets. C.F. collected patient tissues with assistance from X.Y. Y.S. and J.Z. performed computation analysis with input from Y.L., L.T., and Y.W. Q.C.Z., L.T., Y.S., and J.Z. wrote the manuscript with input from all authors. All authors reviewed the manuscript.

DECLARATION OF INTERESTS

The methods, reagents, and data disclosed in this manuscript may be encompassed by one or more patents and patent applications filed by Tsinghua University.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell Culture
  - Lung tumor sample collection and ethics statement
- METHOD DETAILS
  - Lung tumor sample preparation
  - Parallel-seq protocol
  - Design of the barcodes with phase nucleotides
  - Processing of Parallel-seq data
  - Comparison of data quality to other methods
  - Clustering of the joint profiles of lung tumor samples
  - Calculation of gene activity scores
  - CNV analysis
  - EccDNA analysis
  - Metacells analysis
  - Identification of peak-gene associations
  - Prediction of TF binding motifs at cell type-specific enhancers
  - Identification of mutation-gene associations
  - Overlap of peak-gene associations and mutation-gene associations with H3K27ac HiChIP data
  - Comparison of mutation-gene association analysis between Parallel-seq and 10x Multiome

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cels.2025.101266.

Received: August 18, 2024
Revised: January 19, 2025
Accepted: March 28, 2025
Published: April21, 2025

REFERENCES

1. Dagogo-Jack, I., and Shaw, A.T. (2018). Tumour heterogeneity and resistance to cancer therapies. Nat. Rev. Clin. Oncol. *15*, 81–94. https://doi.org/10.1038/nrclinonc.2017.166.
2. Wang, Y., Waters, J., Leung, M.L., Unruh, A., Roh, W., Shi, X., Chen, K., Scheet, P., Vattathil, S., Liang, H., et al. (2014). Clonal evolution in breast

Please cite this article in press as: Tang et al., Joint analysis of chromatin accessibility and gene expression in the same single cells reveals cancer-specific regulatory programs, Cell Systems (2025), https://doi.org/10.1016/j.cels.2025.101266

cancer revealed by single nucleus genome sequencing. Nature *512*, 155–160. https://doi.org/10.1038/nature13600.

3. Satpathy, A.T., Granja, J.M., Yost, K.E., Qi, Y., Meschi, F., McDermott, G.P., Olsen, B.N., Mumbach, M.R., Pierce, S.E., Corces, M.R., et al. (2019). Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. Nat. Biotechnol. *37*, 925–936. https://doi.org/10.1038/s41587-019-0206-z.
4. Patel, A.P., Tirosh, I., Trombetta, J.J., Shalek, A.K., Gillespie, S.M., Wakimoto, H., Cahill, D.P., Nahed, B.V., Curry, W.T., Martuza, R.L., et al. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science *344*, 1396–1401. https://doi.org/10.1126/science.1254257.
5. Wagner, J., Rapsomaniki, M.A., Chevrier, S., Anzeneder, T., Langwieder, C., Dykgers, A., Rees, M., Ramaswamy, A., Muenst, S., Soysal, S.D., et al. (2019). A Single-Cell Atlas of the Tumor and Immune Ecosystem of Human Breast Cancer. Cell *177*, 1330–1345.e18. https://doi.org/10.1016/j.cell.2019.03.005.
6. Regner, M.J., Wisniewska, K., Garcia-Recio, S., Thennavan, A., Mendez-Giraldez, R., Malladi, V.S., Hawkins, G., Parker, J.S., Perou, C.M., Bae-Jump, V.L., and Franco, H.L. (2021). A multi-omic single-cell landscape of human gynecologic malignancies. Mol. Cell *81*, 4924–4941.e10. https://doi.org/10.1016/j.molcel.2021.10.013.
7. Fan, X., Lu, P., Wang, H., Bian, S., Wu, X., Zhang, Y., Liu, Y., Fu, D., Wen, L., Hao, J., and Tang, F. (2022). Integrated single-cell multiomics analysis reveals novel candidate markers for prognosis in human pancreatic ductal adenocarcinoma. Cell Discov. *8*, 13. https://doi.org/10.1038/s41421-021-00366-y.
8. Macaulay, I.C., Haerty, W., Kumar, P., Li, Y.I., Hu, T.X., Teng, M.J., Goolam, M., Saurat, N., Coupland, P., Shirley, L.M., et al. (2015). G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. Nat. Methods *12*, 519–522. https://doi.org/10.1038/nmeth.3370.
9. Hou, Y., Guo, H., Cao, C., Li, X., Hu, B., Zhu, P., Wu, X., Wen, L., Tang, F., Huang, Y., and Peng, J. (2016). Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. Cell Res. *26*, 304–319. https://doi.org/10.1038/cr.2016.23.
10. Stoeckius, M., Hafemeister, C., Stephenson, W., Houck-Loomis, B., Chattopadhyay, P.K., Swerdlow, H., Satija, R., and Smibert, P. (2017). Simultaneous epitope and transcriptome measurement in single cells. Nat. Methods *14*, 865–868. https://doi.org/10.1038/nmeth.4380.
11. Cao, J., Cusanovich, D.A., Ramani, V., Aghamirzaie, D., Pliner, H.A., Hill, A.J., Daza, R.M., McFaline-Figueroa, J.L., Packer, J.S., Christiansen, L., et al. (2018). Joint profiling of chromatin accessibility and gene expression in thousands of single cells. Science *361*, 1380–1385. https://doi.org/10.1126/science.aau0730.
12. Chen, S., Lake, B.B., and Zhang, K. (2019). High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. Nat. Biotechnol. *37*, 1452–1457. https://doi.org/10.1038/s41587-019-0290-0.
13. Zhu, C., Yu, M., Huang, H., Juric, I., Abnousi, A., Hu, R., Lucero, J., Behrens, M.M., Hu, M., and Ren, B. (2019). An ultra high-throughput method for single-cell joint analysis of open chromatin and transcriptome. Nat. Struct. Mol. Biol. *26*, 1063–1070. https://doi.org/10.1038/s41594-019-0323-x.
14. Ma, S., Zhang, B., LaFave, L.M., Earl, A.S., Chiang, Z., Hu, Y., Ding, J., Brack, A., Kartha, V.K., Tay, T., et al. (2020). Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and Chromatin. Cell *183*, 1103–1116.e20. https://doi.org/10.1016/j.cell.2020.09.056.
15. Xu, W., Yang, W., Zhang, Y., Chen, Y., Hong, N., Zhang, Q., Wang, X., Hu, Y., Song, K., Jin, W., and Chen, X. (2022). ISSAAC-seq enables sensitive and flexible multimodal profiling of chromatin accessibility and gene expression in single cells. Nat. Methods *19*, 1243–1249. https://doi.org/10.1038/s41592-022-01601-4.
16. Terekhanova, N.V., Karpova, A., Liang, W.W., Strzalkowski, A., Chen, S., Li, Y., Southard-Smith, A.N., Iglesia, M.D., Wendl, M.C., Jayasinghe, R.G., et al. (2023). Epigenetic regulation during cancer transitions across 11 tumour types. Nature *623*, 432–441. https://doi.org/10.1038/s41586-023-06682-5.
17. Lareau, C.A., Duarte, F.M., Chew, J.G., Kartha, V.K., Burkett, Z.D., Kohlway, A.S., Pokholok, D., Aryee, M.J., Steemers, F.J., Lebofsky, R., and Buenrostro, J.D. (2019). Droplet-based combinatorial indexing for massive-scale single-cell chromatin accessibility. Nat. Biotechnol. *37*, 916–924. https://doi.org/10.1038/s41587-019-0147-6.
18. Datlinger, P., Rendeiro, A.F., Boenke, T., Senekowitsch, M., Krausgruber, T., Barreca, D., and Bock, C. (2021). Ultra-high-throughput single-cell RNA sequencing and perturbation screening with combinatorial fluidic indexing. Nat. Methods *18*, 635–642. https://doi.org/10.1038/s41592-021-01153-z.
19. Hwang, B., Lee, D.S., Tamaki, W., Sun, Y., Ogorodnikov, A., Hartoularos, G.C., Winters, A., Yeung, B.Z., Nazor, K.L., Song, Y.S., et al. (2021). SCITO-seq: single-cell combinatorial indexed cytometry sequencing. Nat. Methods *18*, 903–911. https://doi.org/10.1038/s41592-021-01222-3.
20. Hughes, T.K., Wadsworth, M.H., Gierahn, T.M., Do, T., Weiss, D., Andrade, P.R., Ma, F.Y., de Andrade Silva, B.J., Shao, S., Tsoi, L.C., et al. (2020). Second-Strand Synthesis-Based Massively Parallel scRNA-Seq Reveals Cellular States and Molecular Features of Human Inflammatory Skin Pathologies. Immunity *53*, 878–894.e7. https://doi.org/10.1016/j.immuni.2020.09.015.
21. Corces, M.R., Trevino, A.E., Hamilton, E.G., Greenside, P.G., Sinnott-Armstrong, N.A., Vesuna, S., Satpathy, A.T., Rubin, A.J., Montine, K.S., Wu, B., et al. (2017). An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat. Methods *14*, 959–962. https://doi.org/10.1038/nmeth.4396.
22. Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218. https://doi.org/10.1038/nmeth.2688.
23. ENCODE; Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74. https://doi.org/10.1038/nature11247.
24. Zhao, S.Q., Hong, C.K.Y., Myers, C.A., Granas, D.M., White, M.A., Corbo, J.C., and Cohen, B.A. (2023). A single-cell massively parallel reporter assay detects cell-type-specific gene regulation. Nat. Genet. *55*, 346–354. https://doi.org/10.1038/s41588-022-01278-7.
25. 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. Nat. Genet. *53*, 403–411. https://doi.org/10.1038/s41588-021-00790-6.
26. Chen, Z., Fillmore, C.M., Hammerman, P.S., Kim, C.F., and Wong, K.K. (2014). Non-small-cell lung cancers: a heterogeneous set of diseases. Nat. Rev. Cancer *14*, 535–546. https://doi.org/10.1038/nrc3775.
27. Wu, F., Fan, J., He, Y., Xiong, A., Yu, J., Li, Y., Zhang, Y., Zhao, W., Zhou, F., Li, W., et al. (2021). Single-cell profiling of tumor heterogeneity and the microenvironment in advanced non-small cell lung cancer. Nat. Commun. *12*, 2540. https://doi.org/10.1038/s41467-021-22801-0.
28. Kim, N., Kim, H.K., Lee, K., Hong, Y., Cho, J.H., Choi, J.W., Lee, J.-I., Suh, Y.-L., Ku, B.M., Eum, H.H., et al. (2020). Single-cell RNA sequencing demonstrates the molecular and cellular reprogramming of metastatic lung adenocarcinoma. Nat. Commun. *11*, 2285. https://doi.org/10.1038/s41467-020-16164-1.
29. Leader, A.M., Grout, J.A., Maier, B.B., Nabet, B.Y., Park, M.D., Tabachnikova, A., Chang, C., Walker, L., Lansky, A., Le Berichel, J., et al. (2021). Single-cell analysis of human non-small cell lung cancer lesions refines tumor classification and patient stratification. Cancer Cell *39*, 1594–1609.e12. https://doi.org/10.1016/j.ccell.2021.10.009.
30. Maynard, A., McCoach, C.E., Rotow, J.K., Harris, L., Haderk, F., Kerr, D.L., Yu, E.A., Schenk, E.L., Tan, W.L., Zee, A., et al. (2020). Therapy-Induced Evolution of Human Lung Cancer Revealed by Single-Cell

Please cite this article in press as: Tang et al., Joint analysis of chromatin accessibility and gene expression in the same single cells reveals cancer-specific regulatory programs, Cell Systems (2025), https://doi.org/10.1016/j.cels.2025.101266

RNA Sequencing. Cell *182*, 1232–1251.e22. https://doi.org/10.1016/j.cell.2020.07.017.

31. Amezquita, R., Lun, A., Hicks, S., and Gottardo, R. (2021). Advanced single-cell analysis with Bioconductor. Version 1.2. 1. https://github.com/OSCA-source/OSCA.advanced.
32. 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., et al. (2018). Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility Data. Mol. Cell *71*, 858–871.e8. https://doi.org/10.1016/j.molcel.2018.06.044.
33. 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. https://doi.org/10.1038/nmeth.4401.
34. Kallies, A., and Good-Jacobson, K.L. (2017). Transcription Factor T-bet Orchestrates Lineage Development and Function in the Immune System. Trends Immunol. *38*, 287–297. https://doi.org/10.1016/j.it.2017.02.003.
35. Jenner, R.G., Townsend, M.J., Jackson, I., Sun, K.M., Bouwman, R.D., Young, R.A., Glimcher, L.H., and Lord, G.M. (2009). The transcription factors T-bet and GATA-3 control alternative pathways of T-cell differentiation through a shared set of target genes. Proc. Natl. Acad. Sci. USA *106*, 17876–17881. https://doi.org/10.1073/pnas.0909357106.
36. Staudt, L.M., Clerc, R.G., Singh, H., LeBowitz, J.H., Sharp, P.A., and Baltimore, D. (1988). Cloning of a lymphoid-specific cDNA encoding a protein binding the regulatory octamer DNA motif. Science *241*, 577–580. https://doi.org/10.1126/science.3399892.
37. Huang, Q., Liu, J., Wu, S., Zhang, X., Xiao, Z., Liu, Z., and Du, W. (2021). Spi-B promotes the recruitment of tumor-associated macrophages via enhancing CCL4 expression in lung cancer. Front. Oncol. *11*, 659131. https://doi.org/10.3389/fonc.2021.659131.
38. Zhao, J., Patel, J., Kaur, S., Sim, S.L., Wong, H.Y., Styke, C., Hogan, I., Kahler, S., Hamilton, H., Wadlow, R., et al. (2021). Sox9 and Rbpj differentially regulate endothelial to mesenchymal transition and wound scarring in murine endovascular progenitors. Nat. Commun. *12*, 2564. https://doi.org/10.1038/s41467-021-22717-9.
39. Oh, S.Y., Kim, J.Y., and Park, C. (2015). The ETS Factor, ETV2: a Master Regulator for Vascular Endothelial Cell Development. Mol. Cells *38*, 1029–1036. https://doi.org/10.14348/molcells.2015.0331.
40. Li, C., Virgilio, M.C., Collins, K.L., and Welch, J.D. (2023). Multi-omic single-cell velocity models epigenome-transcriptome interactions and improves cell fate prediction. Nat. Biotechnol. *41*, 387–398. https://doi.org/10.1038/s41587-022-01476-y.
41. Zheng, L., Qin, S., Si, W., Wang, A., Xing, B., Gao, R., Ren, X., Wang, L., Wu, X., Zhang, J., et al. (2021). Pan-cancer single-cell landscape of tumor-infiltrating T cells. Science *374*, abe6474. https://doi.org/10.1126/science.abe6474.
42. Shlien, A., and Malkin, D. (2009). Copy number variations and cancer. Genome Med. *1*, 62. https://doi.org/10.1186/gm62.
43. Gao, R., Bai, S., Henderson, Y.C., Lin, Y., Schalck, A., Yan, Y., Kumar, T., Hu, M., Sei, E., Davis, A., et al. (2021). Delineating copy number and clonal substructure in human tumors from single-cell transcriptomes. Nat. Biotechnol. *39*, 599–608. https://doi.org/10.1038/s41587-020-00795-2.
44. Suvakov, M., Panda, A., Diesh, C., Holmes, I., and Abyzov, A. (2021). CNVpytor: a tool for copy number variation detection and analysis from read depth and allele imbalance in whole-genome sequencing. GigaScience *10*, giab074. https://doi.org/10.1093/gigascience/giab074.
45. Minussi, D.C., Nicholson, M.D., Ye, H., Davis, A., Wang, K., Baker, T., Tarabichi, M., Sei, E., Du, H., Rabbani, M., et al. (2021). Breast tumours maintain a reservoir of subclonal diversity during expansion. Nature *592*, 302–308. https://doi.org/10.1038/s41586-021-03357-x.
46. Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics *14*, 7. https://doi.org/10.1186/1471-2105-14-7.
47. Castells, A., Ino, Y., Louis, D.N., Ramesh, V., Gusella, J.F., and Rustgi, A.K. (1999). Mapping of a target region of allelic loss to a 0.5-cM interval on chromosome 22q13 in human colorectal cancer. Gastroenterology *117*, 831–837. https://doi.org/10.1016/s0016-5085(99)70341-0.
48. Bryan, E.J., Watson, R.H., Davis, M., Hitchcock, A., Foulkes, W.D., and Campbell, I.G. (1996). Localization of an ovarian cancer tumor suppressor gene to a 0.5-cM region between D22S284 and CYP2D, on chromosome 22q. Cancer Res. *56*, 719–721.
49. Ino, Y., Silver, J.S., Blazejewski, L., Nishikawa, R., Matsutani, M., von Deimling, A., and Louis, D.N. (1999). Common regions of deletion on chromosome 22q12.3-q13.1 and 22q13.2 in human astrocytomas appear related to malignancy grade. J. Neuropathol. Exp. Neurol. *58*, 881–885. https://doi.org/10.1097/00005072-199908000-00010.
50. Miyakawa, A., Wang, X.L., Nakanishi, H., Imai, F.L., Shiiba, M., Miya, T., Imai, Y., and Tanzawa, H. (1998). Allelic loss on chromosome 22 in oral cancer: possibility of the existence of a tumor suppressor gene on 22q13. Int. J. Oncol. *13*, 705–709. https://doi.org/10.3892/ijo.13.4.705.
51. Chung, D.C., Brown, S.B., Graeme-Cook, F., Tillotson, L.G., Warshaw, A.L., Jensen, R.T., and Arnold, A. (1998). Localization of putative tumor suppressor loci by genome-wide allelotyping in human pancreatic endocrine tumors. Cancer Res. *58*, 3706–3711.
52. Shiseki, M., Kohno, T., Nishikawa, R., Sameshima, Y., Mizoguchi, H., and Yokota, J. (1994). Frequent allelic losses on chromosomes 2q, 18q, and 22q in advanced non-small cell lung carcinoma. Cancer Res. *54*, 5643–5648.
53. Turner, K.M., Deshpande, V., Beyter, D., Koga, T., Rusert, J., Lee, C., Li, B., Arden, K., Ren, B., Nathanson, D.A., et al. (2017). Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. Nature *543*, 122–125. https://doi.org/10.1038/nature21356.
54. Wu, S., Turner, K.M., Nguyen, N., Raviram, R., Erb, M., Santini, J., Luebeck, J., Rajkumar, U., Diao, Y., Li, B., et al. (2019). Circular ecDNA promotes accessible chromatin and high oncogene expression. Nature *575*, 699–703. https://doi.org/10.1038/s41586-019-1763-5.
55. Hung, K.L., Yost, K.E., Xie, L., Shi, Q., Helmsauer, K., Luebeck, J., Schöpflin, R., Lange, J.T., Chamorro González, R., Weiser, N.E., et al. (2021). ecDNA hubs drive cooperative intermolecular oncogene expression. Nature *600*, 731–736. https://doi.org/10.1038/s41586-021-04116-8.
56. Kumar, P., Kiran, S., Saha, S., Su, Z., Paulsen, T., Chatrath, A., Shibata, Y., Shibata, E., and Dutta, A. (2020). ATAC-seq identifies thousands of extrachromosomal circular DNA in cancer and cell lines. Sci. Adv. *6*, eaba2489. https://doi.org/10.1126/sciadv.aba2489.
57. Wang, Y., Wang, M., Djekidel, M.N., Chen, H., Liu, D., Alt, F.W., and Zhang, Y. (2021). eccDNAs are apoptotic products with high innate immunostimulatory activity. Nature *599*, 308–314. https://doi.org/10.1038/s41586-021-04009-w.
58. Tiwari, N., Tiwari, V.K., Waldmeier, L., Balwierz, P.J., Arnold, P., Pachkov, M., Meyer-Schaller, N., Schübeler, D., van Nimwegen, E., and Christofori, G. (2013). Sox4 Is a Master Regulator of Epithelial-Mesenchymal Transition by Controlling Ezh2 Expression and Epigenetic Reprogramming. Cancer Cell *23*, 768–783. https://doi.org/10.1016/j.ccr.2013.04.020.
59. Tavazoie, S.F., Alarcón, C., Oskarsson, T., Padua, D., Wang, Q., Bos, P.D., Gerald, W.L., and Massagué, J. (2008). Endogenous human microRNAs that suppress breast cancer metastasis. Nature *451*, 147–152. https://doi.org/10.1038/nature06487.
60. Liao, Y.L., Sun, Y.M., Chau, G.Y., Chau, Y.P., Lai, T.C., Wang, J.L., Horng, J.T., Hsiao, M., and Tsou, A.P. (2008). Identification of SOX4 target genes using phylogenetic footprinting-based prediction from expression microarrays suggests that overexpression of SOX4 potentiates metastasis in hepatocellular carcinoma. Oncogene *27*, 5578–5589. https://doi.org/10.1038/onc.2008.168.

Please cite this article in press as: Tang et al., Joint analysis of chromatin accessibility and gene expression in the same single cells reveals cancer-specific regulatory programs, Cell Systems (2025), https://doi.org/10.1016/j.cels.2025.101266

61. Steuer, R., Kurths, J., Daub, C.O., Weise, J., and Selbig, J. (2002). The mutual information: detecting and evaluating dependencies between variables. Bioinformatics *18* (*Suppl 2*), S231–S240. https://doi.org/10.1093/bioinformatics/18.suppl_2.s231.

62. Moore, J.E., Purcaro, M.J., Pratt, H.E., Epstein, C.B., Shoresh, N., Adrian, J., Kawli, T., Davis, C.A., Dobin, A., et al.; ENCODE Project Consortium (2020). Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature *583*, 699–710. https://doi.org/10.1038/s41586-020-2493-4.

63. Mumbach, M.R., Satpathy, A.T., Boyle, E.A., Dai, C., Gowen, B.G., Cho, S.W., Nguyen, M.L., Rubin, A.J., Granja, J.M., Kazane, K.R., et al. (2017). Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. Nat. Genet. *49*, 1602–1612. https://doi.org/10.1038/ng.3963.

64. Kartha, V.K., Duarte, F.M., Hu, Y., Ma, S., Chew, J.G., Lareau, C.A., Earl, A., Burkett, Z.D., Kohlway, A.S., Lebofsky, R., et al. (2022). Functional inference of gene regulation using single-cell multi-omics. Cell Genom. *2*, 100166. https://doi.org/10.1016/j.xgen.2022.100166.

65. Chang, Y., Xue, X., Li, C., Zhao, W., Ma, Y., Xu, F., Wu, Z., Dai, Y., Li, Y., Liu, Y., and Chen, L. (2020). MIR205HG facilitates carcinogenesis of lung squamous cell carcinoma in vitro revealed by long noncoding RNA profiling. Acta Biochim. Biophys. Sin. (Shanghai) *52*, 371–381. https://doi.org/10.1093/abbs/gmaa006.

66. Liu, L., Li, Y., Zhang, R., Li, C., Xiong, J., and Wei, Y. (2020). MIR205HG acts as a ceRNA to expedite cell proliferation and progression in lung squamous cell carcinoma via targeting miR-299-3p/MAP3K2 axis. BMC Pulm. Med. *20*, 163. https://doi.org/10.1186/s12890-020-1174-2.

67. Liu, Y., Wu, Z., Zhou, J., Ramadurai, D.K.A., Mortenson, K.L., Aguilera-Jimenez, E., Yan, Y., Yang, X., Taylor, A.M., Varley, K.E., et al. (2021). A predominant enhancer co-amplified with the SOX2 oncogene is necessary and sufficient for its expression in squamous cancer. Nat. Commun. *12*, 7139. https://doi.org/10.1038/s41467-021-27055-4.

68. Gandellini, P., Profumo, V., Casamichele, A., Fenderico, N., Borrelli, S., Petrovich, G., Santilli, G., Callari, M., Colecchia, M., Pozzi, S., et al. (2012). miR-205 regulates basement membrane deposition in human prostate: implications for cancer development. Cell Death Differ. *19*, 1750–1760. https://doi.org/10.1038/cdd.2012.56.

69. Tsukumo, S., Unno, M., Muto, A., Takeuchi, A., Kometani, K., Kurosaki, T., Igarashi, K., and Saito, T. (2013). Bach2 maintains T cells in a naive state by suppressing effector memory-related genes. Proc. Natl. Acad. Sci. USA *110*, 10735–10740. https://doi.org/10.1073/pnas.1306691110.

70. Roychoudhuri, R., Clever, D., Li, P., Wakabayashi, Y., Quinn, K.M., Klebanoff, C.A., Ji, Y., Sukumar, M., Eil, R.L., Yu, Z., et al. (2016). BACH2 regulates CD8(+) T cell differentiation by controlling access of AP-1 factors to enhancers. Nat. Immunol. *17*, 851–860. https://doi.org/10.1038/ni.3441.

71. Tang, Z., Kang, B., Li, C., Chen, T., and Zhang, Z. (2019). GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. Nucleic Acids Res. *47*, W556–W560. https://doi.org/10.1093/nar/gkz430.

72. Yao, C., Lou, G., Sun, H.W., Zhu, Z., Sun, Y., Chen, Z., Chauss, D., Moseman, E.A., Cheng, J., D'Antonio, M.A., et al. (2021). BACH2 enforces the transcriptional and epigenetic programs of stem-like CD8+ T cells. Nat. Immunol. *22*, 370–380. https://doi.org/10.1038/s41590-021-00868-7.

73. Wang, H., Luo, J., Tian, X., Xu, L., Zhai, Z., Cheng, M., Chen, L., and Luo, S. (2021). DNAJC5 promotes hepatocellular carcinoma cells proliferation though regulating SKP2 mediated p27 degradation. Biochim. Biophys. Acta Mol. Cell Res. *1868*, 118994. https://doi.org/10.1016/j.bbamcr.2021.118994.

74. Younes, M., Brown, R.W., Stephenson, M., Gondo, M., and Cagle, P.T. (1997). Overexpression of Glut1 and Glut3 in stage I nonsmall cell lung carcinoma is associated with poor survival. Cancer *80*, 1046–1051. https://doi.org/10.1002/(Sici)1097-0142(19970915)80:6<1046::Aid-Cncr6>3.0.Co;2-7.

75. Partridge, E.C., Chhetri, S.B., Prokop, J.W., Ramaker, R.C., Jansen, C.S., Goh, S.T., Mackiewicz, M., Newberry, K.M., Brandsmeier, L.A., Meadows, S.K., et al. (2020). Occupancy maps of 208 chromatin-associated proteins in one human cell type. Nature *583*, 720–728. https://doi.org/10.1038/s41586-020-2023-4.

76. Liu, Y., Siejka-Zielińska, P., Velikova, G., Bi, Y., Yuan, F., Tomkova, M., Bai, C., Chen, L., Schuster-Böckler, B., and Song, C.X. (2019). Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution. Nat. Biotechnol. *37*, 424–429. https://doi.org/10.1038/s41587-019-0041-2.

77. Bartosovic, M., Kabbe, M., and Castelo-Branco, G. (2021). Single-cell CUT&Tag profiles histone modifications and transcription factors in complex tissues. Nat. Biotechnol. *39*, 825–835. https://doi.org/10.1038/s41587-021-00869-9.

78. Cao, J., Zhou, W., Steemers, F., Trapnell, C., and Shendure, J. (2020). Sci-fate characterizes the dynamics of gene expression in single cells. Nat. Biotechnol. *38*, 980–988. https://doi.org/10.1038/s41587-020-0480-9.

79. Guilliams, M., Bonnardel, J., Haest, B., Vanderborght, B., Wagner, C., Remmerie, A., Bujko, A., Martens, L., Thoné, T., Browaeys, R., et al. (2022). Spatial proteogenomics reveals distinct and evolutionarily conserved hepatic macrophage niches. Cell *185*, 379–396.e38. https://doi.org/10.1016/j.cell.2021.12.018.

80. Xu, Z., Sziraki, A., Lee, J., Zhou, W., and Cao, J. (2024). Dissecting key regulators of transcriptome kinetics through scalable single-cell RNA profiling of pooled CRISPR screens. Nat. Biotechnol. *42*, 1218–1223. https://doi.org/10.1038/s41587-023-01948-9.

81. Srivatsan, S.R., Regier, M.C., Barkan, E., Franks, J.M., Packer, J.S., Grosjean, P., Duran, M., Saxton, S., Ladd, J.J., Spielmann, M., et al. (2021). Embryo-scale, single-cell spatial transcriptomics. Science *373*, 111–117. https://doi.org/10.1126/science.abb9536.

82. Russell, A.J.C., Weir, J.A., Nadaf, N.M., Shabet, M., Kumar, V., Kambhampati, S., Raichur, R., Marrero, G.J., Liu, S., Balderrama, K.S., et al. (2024). Slide-tags enables single-nucleus barcoding for multimodal spatial genomics. Nature *625*, 101–109. https://doi.org/10.1038/s41586-023-06837-4.

83. Bradley, R.K., and Anczuków, O. (2023). RNA splicing dysregulation and the hallmarks of cancer. Nat. Rev. Cancer *23*, 135–155. https://doi.org/10.1038/s41568-022-00541-7.

84. Rade, M., Grieb, N., Weiss, R., Sia, J., Fischer, L., Born, P., Boldt, A., Fricke, S., Franz, P., Scolnick, J., et al. (2024). Single-cell multiomic dissection of response and resistance to chimeric antigen receptor T cells against BCMA in relapsed multiple myeloma. Nat. Cancer *5*, 1318–1333. https://doi.org/10.1038/s43018-024-00763-8.

85. Sziraki, A., Lu, Z., Lee, J., Banyai, G., Anderson, S., Abdulraouf, A., Metzner, E., Liao, A., Banfelder, J., Epstein, A., et al. (2023). A global view of aging and Alzheimer's pathogenesis-associated cell population dynamics and molecular signatures in human and mouse brains. Nat. Genet. *55*, 2104–2116. https://doi.org/10.1038/s41588-023-01572-y.

86. Di, L., Liu, B., Lyu, Y., Zhao, S., Pang, Y., Zhang, C., Wang, J., Qi, H., Shen, J., and Huang, Y. (2022). Rapid and sensitive single-cell RNA sequencing with SHERRY2. BMC Biol. *20*, 213. https://doi.org/10.1186/s12915-022-01416-x.

87. Mulqueen, R.M., Pokholok, D., O'Connell, B.L., Thornton, C.A., Zhang, F., O'Roak, B.J., Link, J., Yardımcı, G.G., Sears, R.C., Steemers, F.J., et al. (2021). High-content single-cell combinatorial indexing. Nat. Biotechnol. *39*, 1574–1580. https://doi.org/10.1038/s41587-021-00962-z.

88. Sos, B.C., Fung, H.L., Gao, D.R., Osothprarop, T.F., Kia, A., He, M.M., and Zhang, K. (2016). Characterization of chromatin accessibility with a transposome hypersensitive sites sequencing (THS-seq) assay. Genome Biol. *17*, 20. https://doi.org/10.1186/s13059-016-0882-7.

89. Di, L., Fu, Y.S., Sun, Y., Li, J., Liu, L., Yao, J.C., Wang, G.B., Wu, Y.L., Lao, K.Q., Lee, R.W., et al. (2020). RNA sequencing by direct tagmentation of RNA/DNA hybrids. Proc. Natl. Acad. Sci. USA *117*, 2886–2893. https://doi.org/10.1073/pnas.1919800117.

Please cite this article in press as: Tang et al., Joint analysis of chromatin accessibility and gene expression in the same single cells reveals cancer-specific regulatory programs, Cell Systems (2025), https://doi.org/10.1016/j.cels.2025.101266

90. Tang, L., Zhang, J., and Shao, Y. (2024). Code for: A method for joint analysis of chromatin accessibility and gene expression in the same single cells reveals cancer-specific regulatory programs (Zenodo). https://doi.org/10.5281/zenodo.10728191.

91. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. *17*, 10–12. https://doi.org/10.14806/ej.17.1.200.

92. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

93. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

94. 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. https://doi.org/10.1016/j.cels.2018.11.005.

95. Feng, J., Liu, T., Qin, B., Zhang, Y., and Liu, X.S. (2012). Identifying ChIP-seq enrichment using MACS. Nat. Protoc. *7*, 1728–1740. https://doi.org/10.1038/nprot.2012.101.

96. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.

97. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., 3rd, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.

98. Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. *19*, 15. https://doi.org/10.1186/s13059-017-1382-0.

99. Xiong, L., Xu, K., Tian, K., Shao, Y., Tang, L., Gao, G., Zhang, M., Jiang, T., and Zhang, Q.C. (2019). SCALE method for single-cell ATAC-seq analysis via latent feature extraction. Nat. Commun. *10*, 4576. https://doi.org/10.1038/s41467-019-12630-7.

100. 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. https://doi.org/10.1038/s41592-019-0619-0.

101. Fang, Z., Liu, X., and Peltz, G. (2023). GSEApy: a comprehensive package for performing gene set enrichment analysis in Python. Bioinformatics *39*, btac757. https://doi.org/10.1093/bioinformatics/btac757.

102. Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics *27*, 1017–1018. https://doi.org/10.1093/bioinformatics/btr064.

103. Rosenberg, A.B., Roco, C.M., Muscat, R.A., Kuchina, A., Sample, P., Yao, Z., Graybuck, L.T., Peeler, D.J., Mukherjee, S., Chen, W., et al. (2018). Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. Science *360*, 176–182. https://doi.org/10.1126/science.aam8999.

104. Hawkins, J.A., Jones, S.K., Jr., Finkelstein, I.J., and Press, W.H. (2018). Indel-correcting DNA barcodes for high-throughput sequencing. Proc. Natl. Acad. Sci. USA *115*, E6217–E6226. https://doi.org/10.1073/pnas.1802640115.

105. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550. https://doi.org/10.1073/pnas.0506580102.

106. Khan, A., Fornes, O., Stigliani, A., Gheorghe, M., Castro-Mondragon, J.A., van der Lee, R., Bessy, A., Chèneby, J., Kulkarni, S.R., Tan, G., et al. (2018). JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. Nucleic Acids Res. *46*, D1284. https://doi.org/10.1093/nar/gkx1188.

107. Bhattacharyya, S., Chandra, V., Vijayanand, P., and Ay, F. (2019). Identification of significant chromatin contacts from HiChIP data by FitHiChIP. Nat. Commun. *10*, 4221. https://doi.org/10.1038/s41467-019-11950-y.

Please cite this article in press as: Tang et al., Joint analysis of chromatin accessibility and gene expression in the same single cells reveals cancer-specific regulatory programs, Cell Systems (2025), https://doi.org/10.1016/j.cels.2025.101266

CellPress
OPEN ACCESS

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-CD45 PE	BioLegend	Cat#304039; RRID: AB_2562057
Anti-CD3 BV421	BioLegend	Cat#317344; RRID: AB_2565848
Anti-EpCAM PE/Cy7	BioLegend	Cat#324222; RRID: AB_2561506
Biological samples		
Human lung tumor tissues	Beijing 301 Hospital	N/A
Chemicals, peptides, and recombinant proteins		
Dulbecco's Modified Eagle's Medium (DMEM)	ThermoFisher	C11995500BT
fetal bovine serum (FBS)	PAN BIOTECH	P30-3302
PBS	ThermoFisher	C10010500BT
Trypsin-EDTA	ThermoFisher	25200114
RPMI 1640 Medium	ThermoFisher	C11875500BT
MACS Tissue Storage Solution	Miltenyi Biotec	130-100-008
DMEM/F-12	ThermoFisher	11320033
Liberase TL	Sigma Aldrich	05401020001
Elastase	Worthington	NC9301601
DNase I	Sigma Aldrich	11284932001
Dimethyl sulfoxide (DMSO)	Sigma Aldrich	D2650
UltraPure DNase/RNase-Free Distilled Water	ThermoFisher	10977023
Transposase	Robustnique	M0221
Methanol-free formaldehyde	ThermoFisher	28906
Bovine serum albumin (BSA)	VWR	V0332-100G
1M Tris-HCl pH 7.4	Sigma Aldrich	T2663-1L
5M NaCl	ThermoFisher	AM9759
1M $MgCl_2$	ThermoFisher	AM9530G
RiboLock	ThermoFisher	EO0384
SUPERase•In RNase Inhibitor	ThermoFisher	AM2694
10% Nonidet P40 Substitute	Sigma Aldrich	11332473001
10% Tween 20	Sigma Aldrich	11332465001
Digitonin	Sigma Aldrich	D141-100MG
10% Triton X-100	Sigma Aldrich	93443
500 μM EDTA	ThermoFisher	AM9260G
50% PEG 8000	NEB	B1004SVIAL
Klenow fragment	NEB	M0212L
Critical commercial assays		
C-Chip Disposable Hemocytometer	As One	DHC-N01N
1.5 ml Microcentrifuge Tubes	Axygen	MCT-150-C
70 μm Cell Strainers	BIOLOGIX	15-1070
Red Cell Lysis Buffer	TIANGEN	4992957
Trypan Blue Solution	ThermoFisher	15250061
Cell staining buffer	BioLegend	420201
Human TruStain FcX	BioLegend	422302
Calcein-AM	BioLegend	425201
APC Annexin V	BioLegend	640941

(*Continued on next page*)

Please cite this article in press as: Tang et al., Joint analysis of chromatin accessibility and gene expression in the same single cells reveals cancer-specific regulatory programs, Cell Systems (2025), https://doi.org/10.1016/j.cels.2025.101266

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
7-AAD Viability Staining Solution	BioLegend	420404
35 μm BD cell strainers	BD Falcon	352235
10 mM dNTP	NEB	N0447L
Maxima H Minus Reverse Transcriptase	ThermoFisher	EP0753
NEBuffer 3.1	NEB	B7203S
T4 DNA ligase	Robustnique	M0082
RNase Cocktail Enzyme Mix	ThermoFisher	AM2286
RNAse H	Enzymatics	Y9220L
Chromium Next GEM Chip H	10x Genomics	PN-1000162
Chromium Single Cell ATAC Reagent Kits	10x Genomics	PN-1000176
SPRI beads	Beckman Coulter	B23318
Elution Buffer	QIAGEN	19086
Qubit dsDNA HS Assay Kit	ThermoFisher	Q32851
Agilent High Sensitivity D1000 ScreenTape Assay	Agilent	5067-5584 and 5067-5585
Deposited data		
Raw and analyzed data	This paper	GSA: HRA006440 and HRA001829; OMIX database: OMIX009208
Bulk RNA-seq and ATAC-seq data of K562 cell line	Consortium et al.[23]	ENCODE: ENCFF631TDY and ENCFF048XLS
scRNA-seq data of K562 cell line	Zhao et al.[24]	GSE188639
scATAC-seq data of K562 cell line	Granja et al.[25]	GSE162690
Joint scATAC-and-scRNA data of cell lines generated by 10x Multiome, and ISSAAC-seq	Xu et al.[15]	E-MTAB-11264
Joint scATAC-and-scRNA data of cell lines generated by SNARE-seq	Chen et al.[12]	GSE126074
Joint scATAC-and-scRNA data of cell lines generated by SHARE-seq	Ma et al.[14]	GSE140203
Joint scATAC-and-scRNA data of cell lines generated by Paired-seq	Zhu et al.[13]	GSE130399
Joint scATAC-and-scRNA data of cell lines generated by sci-CAR	Cao et al.[11]	GSE117089
scRNA-seq data of T cells generated by 10x Genomics RNA kit	Zheng et al.[41]	GSE156728
scATAC-seq data of T cells generated by 10x Genomics RNA kit	Satpathy et al.[3]	GSE129785
WGS dataset of K562 cell line	Consortium et al.[23]	GSE176762
H3K27ac HiChIP data of K562 cell line	Mumbach et al.[63]	GSE101498
H3K27ac HiChIP data of lung cancer cells	Liu et al.[67]	GSE166234
ChIP-seq data of UBP1	Partridge et al.[75]	GSE104247
Experimental models: Cell lines		
Human: HEK293T	ATCC	CRL-3216
Human: HeLa-S3	ATCC	CCL-2.2
Human: K562	ATCC	CCL-243
Mouse: NIH/3T3	ATCC	CRL-1658
Oligonucleotides		
See Table S9 for a full list	This paper	N/A
Software and algorithms		
Cutadapt v3.4	Martin[91]	https://cutadapt.readthedocs.io/en/stable/
STAR v.2.7.0	Dobin et al.[92]	https://github.com/alexdobin/STAR/releases
samtools	Li et al.[93]	http://samtools.sourceforge.net/

(*Continued on next page*)

Please cite this article in press as: Tang et al., Joint analysis of chromatin accessibility and gene expression in the same single cells reveals cancer-specific regulatory programs, Cell Systems (2025), https://doi.org/10.1016/j.cels.2025.101266

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Parallel-seq pipeline	This paper; Mendeley Data	https://doi.org/10.5252/zenodo.10728191. and https://github.com/yqshao17/Parallel-seq
Scrublet	Wolock et al.[94]	https://github.com/swolock/scrublet
MACS2 v2.2.7.1	Feng et al.[95]	https://github.com/macs3-project/MACS/releases
BEDtools v2.31.0	Quinlan and Hall[96]	https://github.com/arq5x/bedtools2/releases
Seurat v4	Hao et al.[97]	https://satijalab.org/seurat/articles/install_v5.html#install-previous-versions-of-seurat
Scanpy v1.9.6	Wolf et al.[98]	https://scanpy.readthedocs.io/en/stable/installation.html
SCALE v 1.1.2	Xiong et al.[99]	https://github.com/jsxlei/SCALE
Harmony	Korsunsky et al.[100]	https://github.com/immunogenomics/harmony
Cicero	Pliner[32]	https://cole-trapnell-lab.github.io/cicero-release/docs_m3/
inferCNV	Trinity CTAT Project	https://github.com/broadinstitute/inferCNV
GSVA	Hänzelmann et al.[46]	https://github.com/rcastelo/GSVA
Circle_finder	Kumar et al.[56]	https://github.com/pk7zuva/Circle_finder
GSEApy	Fang et al.[101]	https://gseapy.readthedocs.io/en/latest/index.html
FIMO	Grant et al.[102]	https://meme-suite.org/meme/doc/fimo.html

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell Culture

HEK293T, HeLa-S3, and NIH/3T3 cells were cultured in Dulbecco's modified Eagle's medium (DMEM) (C11995500BT, ThermoFisher) supplemented with 10% fetal bovine serum (FBS) (P30-3302, PAN BIOTECH) at 37 °C with 5% CO_2. Cells were rinsed with PBS (C10010500BT, ThermoFisher) and detached by incubating for 3-5 min at 37 °C with 1 ml of 0.25% Trypsin-EDTA (25200114, ThermoFisher). K562 cells were cultured in RPMI 1640 Medium (C11875500BT, ThermoFisher) supplemented with 10% FBS at 37 °C with 5% CO2. Detached HEK293T, HeLa-S3, and NIH/3T3 cells and K562 cell suspensions were harvested via centrifugation, washed with PBS, and counted using a C-Chip Disposable Hemocytometer (DHC-N01N, As One).

Lung tumor sample collection and ethics statement

Tumor samples were collected from resected specimens of 22 patients who were pathologically diagnosed with lung tumor and underwent surgery at the Department of Thoracic Surgery of the Beijing 301 Hospital. The paired normal samples were collected at least 5 cm away from the tumor border in the same surgery operation. This study was approved by the Medical Research Ethics Committee of the Beijing 301 Hospital (S2019-310-01), and written informed consent was obtained from all participants included in this study.

METHOD DETAILS

Lung tumor sample preparation

Fresh, solid tumor samples for lung tumor and paired normal samples were collected and placed in pre-chilled MACS Tissue Storage Solution (130-100-008, Miltenyi Biotec) (2–8 °C). The sample was completely covered with the MACS Tissue Storage Solution and transported from the hospital to the research laboratory. For the dissociation mixture, we used 4677 μl DMEM/F-12 (11320033, ThermoFisher); 250 μl of 2.5 mg/ml Liberase TL (05401020001, Sigma Aldrich) to a final concentration of 250 μg/ml; 23 μl of 2 mg ml^{-1} elastase (NC9301601, Worthington) to a final concentration of 9.2 μg/ml; and 50 μl of 10 mg/ml DNase I (11284932001, Sigma Aldrich) to a final concentration of 100 μg/ml. The tissue was minced using scissors in 1.5 ml Microcentrifuge Tubes into pieces under 0.4 mm. The dissociation mix was incubated at 37 °C and rotated horizontally at 90 rpm for 60 min. Single cell suspensions were filtered through 70 μm Cell Strainers (15-1070, BIOLOGIX) and centrifuged at 500 ×g for 5 min at 4 °C. Cells were resuspended with 1 ml PBS and 3 ml Red Cell Lysis Buffer (4992957, TIANGEN) was added. The mix was incubated at room temperature for 5 min and centrifuged at 500 ×g for 5 min at 4 °C. Cells were resuspended in 500 μl FBS, then 5 μl cells were aliquoted and mixed with 5 μl Trypan Blue Solution (15250061, ThermoFisher) for cell number counting with a C-Chip Disposable Hemocytometer. The single cell suspension was diluted with FBS supplemented with Dimethyl sulfoxide (DMSO) (D2650, Sigma Aldrich) to a final concentration of 10%.

We cryopreserved the cells with each tube containing $1x10^6$ single cells. Before experiments, cells were gently thawed at 37 °C for 5 min and centrifuged at 500 ×g for 5 min at 4 °C. Cells were resuspended with 80 μl cell staining buffer (420201, BioLegend) and 5 μl Human TruStain FcX (422302, BioLegend) and incubated for 5 min at 4 °C. The cell mixture was stained by adding 5 μl anti-CD45 PE (304039, BioLegend), 5 μl anti-CD3 BV421(317344, BioLegend), and 5 μl anti-EpCAM PE/Cy7 (324222, BioLegend) antibody and

Please cite this article in press as: Tang et al., Joint analysis of chromatin accessibility and gene expression in the same single cells reveals cancer-specific regulatory programs, Cell Systems (2025), https://doi.org/10.1016/j.cels.2025.101266

incubated for 15 min at 4 °C with protection from light. The stained cells were washed with 1 ml PBS and centrifuged at 500 ×g for 5 min at 4 °C. The supernatant was discarded and single cells were resuspended with 1 mL PBS supplemented with 0.02 μM Calcein-AM (425201, BioLegend) followed by incubation at room temperature for 15 min while protected from light. Then, cells were centrifuged and resuspended with 90 μl Annexin V Binding Buffer, 5 μl APC Annexin V (640941, BioLegend), and 5 μl 7-AAD Viability Staining Solution (420404, BioLegend), followed by incubation at room temperature for 10 min. Finally, we added 400 μl PBS to the mixture and filtered cells with 35 μm BD cell strainers (352235, BD Falcon). Single cells stained as Calcein-AM positive, 7-AAD negative, and Annexin V negative were sorted with MoFlo Astrios EQ Cell Sorter (Beckman Coulter). We found that tumor cells or T cells may dominate the sequencing data. To balance the sequencing effort on different cell types, we first sorted cells based on their staining with CD45 and EpCAM antibody into four groups of cells as illustrated in Figure 2A, and balanced their abundance in the input of an experiment.

Parallel-seq protocol

Transposon preparation

Tn5Merev, Tn5ME-A, Tn5ME-B, and barcoded R1BxME (x for 1-96) were ordered from Sangon Biotech (Shanghai) Co., Ltd. And sequences are listed in Table S9. Tn5Merev (10 μM) was annealed with 10 μM Tn5ME-B and 10 μM R1BxME of each barcode separately with the protocol: denaturing at 95 °C for 2 min, cool to 20 °C with ramp rate of 0.1 °C/s, and hold at 4 °C. To assemble the barcoded transposons, we combined 2 μl annealed Tn5ME-B (Parallel-seq), 2 μl annealed Tn5ME-R1Bx, 2 μl 10x TPS, 4 μl Transposase (M0221, Robustnique), and 10 μl UltraPure DNase/RNase-Free Distilled Water (10977023, ThermoFisher) and incubated the mixture at room temperature for 30 min. Assembled transposons were split into 4 μl in each tube for storage no more than one month at -20 °C.

Fixation

We counted 50k single cells for cell lines or sorted 50k primary cells for lung tumor samples as the starting materials for each tube. Single cells were centrifuged at 500 ×g for 5 min at 4 °C and resuspended with 250 μl PBS. Fixations were performed by adding 750 μl PBS containing 1.33% methanol-free formaldehyde (28906, ThermoFisher) and incubating on ice for 10 min. Then 50 μl 20% bovine serum albumin (BSA) (V0332-100G, VWR) was added, and the fixed cells were collected by swinging-bucket centrifugation at 1000 ×g for 3 min at 4 °C followed by fixed-angle centrifugation at 1000 ×g for 1 min. After centrifugation, the cell pellet was on one side of the 1.5 ml microcentrifuge tubes (MCT-150-C, Axygen) and the supernatant was carefully removed with two pipetting steps without disturbing the cell pellet. We found that adding BSA and the centrifugation strategy used here helped recover more primary cells, thus the centrifugation step in the following protocol used the same strategy if not specified.

Permeabilization

We prepared the 50 ml 2x RSB buffer as described in Omni-ATAC[21] by combining 1 ml 1M Tris-HCl pH 7.4 (T2663-1L, Sigma Aldrich), 200 μl 5M NaCl (AM9759, ThermoFisher), 300 μl 1M $MgCl_2$ (AM9530G, ThermoFisher), and 48.5 ml UltraPure DNase/RNase-Free Distilled Water. Permeabilization buffer was prepared for fixation by combining 50 μl 2x RSB buffer, 1 μl RiboLock (EO0384, ThermoFisher), 1 μl SUPERase•In RNase Inhibitor (AM2694, ThermoFisher), 1 μl 10% Nonidet P40 Substitute (11332473001, Sigma Aldrich), 1 μl 10% Tween 20 (11332465001, Sigma Aldrich), 1 μl 1% Digitonin (D141-100MG, Sigma Aldrich), 5 μl 20% BSA, and 40 μl UltraPure DNase/RNase-Free Distilled Water for each sample. Wash buffer was prepared by combining 500 μl 2 x RSB buffer, 10 μl 10% Tween 20, 1 μl RiboLock, 50 μl 20% BSA, and 439 μl UltraPure DNase/RNase-Free Distilled Water for each sample. Permeabilization was performed by adding 100 μl permeabilization buffer immediately after removing the fixation reagent and pipetting eight times, followed by incubation on ice for 5 min. After permeabilization, 1 ml wash buffer was added into tubes of each sample. The cells were collected by centrifugation.

Transposition

We prepared the ATAC-seq reaction solution by combining 10 μl 5x LM buffer (M0221, Robustnique), 16.5 μl PBS, 0.5 μl RiboLock, 0.5 μl SUPERase•In RNase Inhibitor, 0.5 μl 10% Tween 20, 0.5 μl 1% Digitonin, and 17.5 μl UltraPure DNase/RNase-Free Distilled Water for each sample. The permeabilized single cells were resuspended with 46 μl ATAC-seq reaction solution supplemented with 4 μl barcoded transposon (BC1) into each tube. ATAC-seq reaction was conducted at 37 °C at 550 rpm with a heated lid on the Thermomixer. After the ATAC-seq reaction, cells were washed with 949 μl PBS, 10 μl 10% Triton X-100 (93443, Sigma Aldrich), 1 μl RiboLock, and 50 μl 20% BSA and collected by centrifugation.

In-cell reverse transcription

We prepared 16 μl suspension solution for each sample by combining 8 μl PBS, 0.5 μl RiboLock, 0.5 μl SUPERase•In RNase Inhibitor, and 7 μl UltraPure DNase/RNase-Free Distilled Water. The reverse transcription mixture was prepared by combining 8 μl 5 x RT buffer, 2 μl 10 mM dNTP (N0447L, NEB), 0.5 μl RiboLock, 0.25 μl SUPERase•In RNase Inhibitor, 4 μl Maxima H Minus Reverse Transcriptase (EP0753, ThermoFisher), and 5.25 μl UltraPure DNase/RNase-Free Distilled Water. Then 20 μl reverse transcription mixture was distributed into each tube and the 2 μl random and 2 μl polyT reverse transcription primers with matched barcode (BC1) were added into corresponding tubes (Table S9). We resuspended the transposed cells with 16 μl resuspension solution and added them into barcode matched PCR tubes. The reaction was mixed thoroughly and the reverse transcription reaction was conducted at 50 °C for 10 minutes, followed by three thermal cycles (8 °C for 12 s, 15 °C for 45 s, 20 °C for 45 s, 30 °C for 30 s, 42 °C for 120 s, and 50 °C for 180 s), then incubating at 50 °C for 5 minutes and finally holding at 4 °C. The reverse transcription reactions from different tubes were combined into one 1.5 ml tube on ice. Cells were centrifuged and washed with 1 ml PBS supplemented with 10 μl 10% Triton X-100 and 50 μl 20% BSA.

Please cite this article in press as: Tang et al., Joint analysis of chromatin accessibility and gene expression in the same single cells reveals cancer-specific regulatory programs, Cell Systems (2025), https://doi.org/10.1016/j.cels.2025.101266

Preparation of ligation plates

Parallel-seq used a ligation reaction to add second round barcodes (Barcode2). The ligation adapters were annealed before in-cell barcode ligation by combining 11 μM dscB' oligos and 12 μM barcoded dscBxx oligos (Table S9), resulting in a final reaction volume of 100 μl per well of a 96-well plate. Then we incubated the plate at 95 °C for 2 minutes, cooled it to 20°C at the ramp rate of 0.1°C per second, and split the plate into 10 ligation plates with each well containing 10 μl barcoded splint ligation adapters.

In-cell ligation

The ligation reaction was conducted using the SPLiT-seq protocol[103] without RNase inhibitor. We prepared 2 ml 1x NEBuffer 3.1 (B7203S, NEB) and 2 ml ligation solution (500 μl 10x T4 DNA ligation buffer, 100 μl T4 DNA ligase (M0082, Robustnique), 50 μl 10% Triton X-100, and 1,350 μl UltraPure DNase/RNase-Free Distilled Water). We resuspended the pooled single cells with 1x buffer 3.1 and thoroughly mixed them with the ligation solution. Then 40 μl cells in the ligation mixture were added into each well of the prepared ligation plate. The plate was incubated at room temperature with rotation at 15 rpm for 1 hour. After ligation, 2 μl 500 μM EDTA (AM9260G, ThermoFisher) was added to each well and cells were pooled. The pooled cells were supplemented with 50 μl 10% Triton X-100 and 50 μl 20% BSA and centrifuged to remove supernatant. Then the cells were washed with 940 μl PBS, 10 μl 10% Triton X-100, and 50 μl 20% BSA one more time.

RNase digestion

Barcode ligated cells were ready for RNase digestion. We resuspended the cells with RNase digestion reaction (40 μl 5 x RT buffer, 8 μl RNase Cocktail Enzyme Mix (AM2286, ThermoFisher), 8 μl RNAse H (Y9220L, Enzymatics), and 144 μl UltraPure DNase/RNase-Free Distilled Water) and incubated the mixture at 37 °C for 30 min with interval mixing (15s at 300 rpm and 45s hold) on the thermomixer. After digestion, cells were washed by adding 790 μl PBS and 10 μl 10% Triton X-100 and centrifugated. BSA should not be added in this step because the remaining BSA would aggregate with PEG 8000 in the reaction of the next step.

Second strand synthesis

The cells collected from the RNase digestion step were resuspended with second strand synthesis reaction mixture (40 μl 5 x RT buffer, 48 μl 50% PEG 8000 (B1004SVIAL, NEB), 20 μl 10mM dNTP, 2 μl 1mM dN-P3short primer (for Parallel-seq) (Table S9), 5 μl Klenow fragment (M0212L, NEB), and 85 μl UltraPure DNase/RNase-Free Distilled Water) and incubated at 37°C for 1 hour with interval mixing (15s at 300 rpm and 45s hold) on the thermomixer. After second strand synthesis, the cells were washed twice with PBS containing 0.1 Triton X-100 and 1% BSA. Then the cells were resuspended with 40 μl 0.5x PBS and counted with a C-Chip Disposable Hemocytometer in trypan blue.

Overloading with Chromium ATAC-seq kit

Theoretically, 1.53 M cells can be loaded into one channel of a Chromium Next GEM Chip H (PN-1000162, 10x Genomics).[18] In this study, 100k cells were centrifuged and resuspended with 7 μl ATAC-seq buffer B and replenished to 15 μl with 1x nuclei buffer. Then, 56.5 μl Barcoding Reagent B, 1.5 μl Reducing Agent B, and 2 μl Barcoding Enzyme (PN-1000176, Chromium Genomics) were combined with the cells and loaded into one channel of a Chromium Next GEM Chip H. After GEM Generation, the droplets were split into 16 tubes with each tube containing 6.25 μl droplets. Linear amplification was performed as follows: 72°C for 5 mins, 98°C for 30s, then 12 cycles at 98°C for 10 s, 59°C for 30 seconds, 72°C for 1 min, and hold at 15 °C. Then the cells were combined with droplet barcodes (BC3).

Parallel-seq libraries construction

Post GEM incubation cleanup was conducted and scaled down as Chromium Next GEM Single Cell ATAC Reagent Kits V1.1. Recovery Agent (7.8 μl) was added to each tube and the tubes were gently inverted 10 times to mix. Samples were centrifuged briefly and 12.5 μl Dynabeads Cleanup Mix was added. The mixture was pipetted five times and incubated for 10 min at room temperature. The product was eluted with 81 μl Elution Solution I and split into two parts, 40 μl for scATAC and 40 μl for scRNA. The scATAC portion was cleaned with 1.2x SPRI beads (B23318, Beckman Coulter), and the scRNA portion with 0.8x SPRI beads. scATAC libraries were amplified with SI-PCR primer B (PN-2000128, 10 x Genomics) and N7xx primers (BC4) (Table S9). RNA-seq libraries were amplified with SI-PCR primer B (PN-2000128) and P3xx primers (BC4) (Table S9). After amplification, the scATAC portion was cleaned with 1.2x SPRI beads and the scRNA portion with 0.8x SPRI beads. The libraries were eluted with 20 μl Elution buffer (19086, QIAGEN) and the concentration was determined using a Qubit dsDNA HS Assay Kit (Q32851, ThermoFisher). Each sub-library was expected to recover more than 20 ng product. Quality control of the libraries was performed with Agilent High Sensitivity D1000 ScreenTape Assay (5067-5584 and 5067-5585, Agilent).

Sequencing

Parallel-seq libraries were sequenced using an Illumina NovaSeq 6000 Sequencing System with 16-nt i5 index, 8-nt i7 index, and PE150 sequencing.

Design of the barcodes with phase nucleotides

The cellular barcodes and ligation linkers of Parallel-seq were recorded in Read1. To balance sequencing nucleotide composition, we added phase nucleotides between BC2 and the ligation linker. In detail, the 1st-24th barcodes did not have any phase nucleotides added; the 25th-48th barcodes had one T phase nucleotide added; the 49th-72nd barcodes had two phase nucleotides, CA, added; and the 73rd-96th had three phase nucleotides, ACA, added. For cells whose barcode2 was 1-24, the BC1 and BC2 should be located within 36-41 and 11-18th of Read1, respectively. For cells whose BC2 was 25-48, 49-72, or 73-96, the barcode location should be moved one nucleotide step down. The BC3 and BC4 correspond to i5 index and i7 index, respectively. The unique molecular identifier (UMI) is located within 1-10 of Read1.

Please cite this article in press as: Tang et al., Joint analysis of chromatin accessibility and gene expression in the same single cells reveals cancer-specific regulatory programs, Cell Systems (2025), https://doi.org/10.1016/j.cels.2025.101266

Processing of Parallel-seq data

A customized pipeline was used for processing the Parallel-seq data including preprocessing the reads, parsing barcodes, mapping, and quality filtering. Briefly, raw reads were trimmed with cutadapt (v3.4)[91] to remove general sequencing adaptors. The cell barcodes were parsed based on Read1 and a single edit was allowed for each round of barcodes as measured by FREE Divergence.[104] We further filtered out reads with mosaic end sequence for scRNA libraries and reads without the sequence for scATAC libraries. Reads were aligned to the human genome for the human lung tumor sample and a combined genome of human (hg38) and mouse (mm10) for cell line mixture using STAR[92] with paired-end reads served as input for scATAC and Read2 served as input for scRNA.

After mapping, a modified Python script from the SPLiT-seq pipeline[103] was used to collapse UMI and quantify the number of UMIs in the scRNA data or the number of unique fragments in the scATAC data for each barcode combinations. For scATAC data, reads mapped to mitochondria were removed. Then, a ''knee'' plot was applied to automatically choose putative cells. Specifically, the barcode combinations were ranked by their number of unique fragments for scATAC data or the number of UMIs for scRNA data, and those above the transition point, where there is a dramatical decrease in unique fragments or UMIs, were selected for further filtering (Figures S1A and S1B).

We further quantified the number of detected genes in the scRNA data, and calculated the the TSS enrichment score15 in the scATAC data, for each unique barcode combination. Following the approaches of the Kun Zhang and Bing Ren groups,[12,13] we set cutoffs >300 unique fragments and TSS enrichment > 6 for scATAC data, and >200 detected genes for scRNA data, to filter out barcode combinations.

Finally, a digital gene x cell matrix was generated for the scRNA data and a bin x cell matrix was generated by counting the Tn5 insertions on 2-kb bins across the genome for the scATAC data.

We used Scrublet[94] to predict doublet probability and remove doublets at the default threshold. For mixed cell line data, cells with <90% UMIs mapped to one species were regarded as mixed cells.

Comparison of data quality to other methods

We compared Parallel-seq with competing methods, including droplet-based methods 10x Multiome, ISSAAC-seq, and SNARE-seq, and combinatorial indexing based methods SHARE-seq, Paired-seq, and sci-CAR, using K562, HEK293T, GM12878, and NIH/3T3 cell lines, and evaluated both data quality (using a pipeline reported in the ISSAAC-seq study[15]), throughput, and costs. We only compared the droplet version of ISSAAC-seq in our study. We applied a uniform cutoff of 100 unique fragments for scATAC data and 100 UMIs for scRNA data across Parallel-seq and other competing methods to ensure a fair comparison of data quality.

Clustering of the joint profiles of lung tumor samples

RNA profiles were processed with Seurat v4[97] for principal component analysis (PCA) and ATAC profile was processed by SCALE[99] for dimension reduction. Briefly, the raw cell-gene matrix was normalized with 'SCTransform' and reduced dimensions with 'RunPCA' by Seurat v4 with default parameters. Raw ATAC count matrix was input into SCALE (n_top_features=50000, batch_size=64) and the output adata.obsm['latent'] was extracted for further analysis. We found that the latent features of RNA showed batch effects, while ATAC did not. So, we first removed batch effects of the PCA features of RNA with Harmony[100] and then used the corrected PCA features and SCALE latent features for joint clustering by weighted nearest neighbor (WNN) analysis using Seurat v4. Uniform Manifold Approximation and Projection (UMAP) was used for visualization. The confusion matrix illustrates the relationship between cell types identified using joint scATAC-and-scRNA profiles, as well as the clusters identified within separate scRNA and scATAC profiles. The ARI score was used to assess the agreement between the clusters.

Calculation of gene activity scores

We calculated gene activity scores by Cicero as previously described.[32] Briefly, Cicero identifies enhancer-promoter associations based on their co-accessibility. Then gene activity scores are calculated based on the overall accessibility of its promoter and associated distal peaks. Gene activity scores (GA) are further transformed by '$\log_{10}$(GA*1000000 + 1)'.

CNV analysis

ATAC CNV was estimated as previously described.[3] Briefly, insertions were calculated across the genome by 5M bins with a step size of 1M. Then 100 background regions with similar GC content were compared to the target region to calculate amplification. RNA CNV was calculated by inferCNV (https://github.com/broadinstitute/inferCNV).

For comparing ATAC and RNA CNV, we first converted features of the inferCNV matrix from genes to bins which are the same as the bins used in ATAC CNV, by calculating the average data value of genes within one bin as the new score for each bin. Then we calculated the Spearman correlation of average CNV scores along all bins between ATAC and RNA. Those bins with RNA CNV scores between the 20th and 80th percentile and ATAC CNV scores larger than the 80th percentile or smaller than the 20th percentile were termed as ATAC-specific CNV regions, while the counterparts were termed as RNA-specific CNV regions. To calculate the ratio of exonic and intronic regions, we annotated all bins by nucleotide by GENCODE GTF file.

For subclone analysis, we reduced dimensions of ATAC and RNA CNV scores by PCA. Then we used WNN for joint embedding of ATAC and RNA CNVs and the Leiden method for clustering to identify subclones. For pathway analysis of subclones, we used the R package GSVA[46] to calculate the variation of gene set enrichment of cancer hallmark pathways through cells.

Please cite this article in press as: Tang et al., Joint analysis of chromatin accessibility and gene expression in the same single cells reveals cancer-specific regulatory programs, Cell Systems (2025), https://doi.org/10.1016/j.cels.2025.101266

EccDNA analysis

eccDNAs were identified by Circle_finder[56] from scATAC data. Circle_finder identifies eccDNAs from paired-end reads with one end mapping on the body of an eccDNA and the other end mapping on the ligation junction as a split read. We assigned all eccDNAs to single cells and classified all epithelial cells into two groups based on whether eccDNA was detected. Cells with at least one eccDNA detected are called "eccDNA cells" and cells without eccDNA are called "Non_eccDNA cells". Genes that were specifically expressed in eccDNA cells were identified by Wilcoxon rank-sum test (adjusted p-value < 0.05, $\log_2$ foldchange > 1), and were termed as eccDNA signature genes. We used GSEApy[101] to identify the enriched cancer hallmark pathways from MsigDB[105] with eccDNA signature genes.

Metacells analysis

To reduce the sparsity of single-cell profiles, we first performed high-resolution clustering for each major cell type using Seurat's FindClusters function with parameters algorithm = 3, resolution = 50, based on the joint embedding of cells. Then we generated the metacells by averaging the information of the cells from the same clusters identified by the high-resolution clustering.

Identification of peak-gene associations

For each cluster, peaks were called by MACS2[95] (–nomodel –shift -75 –extsize 150) on Tn5 insertions. Then the cluster-specific peaks were merged by BEDtools[96] merge and extended to 500 bp. We counted Tn5 insertions on peaks to generate the cell-peak matrix.

To identify peak-gene associations, we calculated mutual information to measure the mutual dependency between gene expression and peak accessibility. Only peaks within 500 kb from the TSS of the gene were considered.

Specially, we binned the expressions of certain gene g and the chromatin accessibilities of certain peak p across the cells of interest into a fix number of 8 bins, separately, to construct a contingency matrix, and calculated the MI between gene g and peak p using the following formula with the sklearn.metrics.mutual_info_score function:

$$MI(U,V) = \sum_{i=1}^{|U|}\sum_{j=1}^{|V|}\frac{|U_i \cap V_j|}{N} log\frac{N|U_i \cap V_j|}{|U_i||V_j|}$$

where $|U_i|$ is the number of cells in bin i for the expression of gene g, and $|V_i|$ is the number of cells in bin i for the chromatin accessibility of peak p.

To remove the probable bias toward higher accessibility regions, we calculated Z scores against the background distribution of the open chromatin regions. Briefly, we generated 100 random peaks as background for each peak with similar accessibility and GC content. Mutual information between the corresponding gene and background peaks served as the background distribution. The Z score of the peak-gene association was calculated with

$$z = (m - \overline{m})/\sigma$$

in which m is the mutual information between the gene expression and peak accessibility, and $\overline{m}$ and σ are the average and standard deviation, respectively, of mutual information between the gene expression and the background peaks. P-values were determined by a one-sided z-test and adjusted by the Benjamini-Hochberg procedure with FDR=0.05.

We conducted regulatory analysis on the K562 dataset from Parallel-seq, 10x Multiome and ISSAAC-seq.[15] We randomly sampled the same numbers of cells (180, 350, and 700 cells) from both datasets for the comparison.

For the identification of cell type-specific regulatory events in lung tumors, we first used Scanpy[98] to calculate cell type-specific genes and peaks by Wilcoxon rank-sum test with $\log_2$ foldchange > 1 and adjusted p-value < 0.05. Then we identified peak-gene associations within each cell type and required them to involve either cell type-specific genes or peaks.

Prediction of TF binding motifs at cell type-specific enhancers

To predict the TFs occupied at enhancers, we extracted enhancer sequences with SAMtools[93] faidx from the hg38 reference genome. Then we used Find Individual Motif Occurrences (FIMO)[102] to scan motifs in the JASPAR20 database,[102,106] with a cutoff of q value < 0.05.

Identification of mutation-gene associations

We generated allele frequency for each site by pileup output with BAM files using SAMtools on ATAC peaks called from tumor cells. Those sites with at least two kinds of alleles in all epithelial cells and where each allele occurred in at least five tumor cells were kept for downstream analysis. The wild-type allele was identified as the allele that occurred in more than 80% of normal epithelial cells with reads mapped to the particular site. For each site, the epithelial cells were grouped by whether they contained a mutated or wild-type allele for differential gene expression analysis using Scanpy. To increase the confidence of mutation-gene associations, we only kept those genes whose flanking region (500 kb) of TSS contains the mutations as candidates. We then performed a Z-test for each gene, taking into account the distribution of fold changes observed for all detected genes. The candidate genes with p-values < 0.05 were deemed to be associated with the mutation site. We further used the average expression of housekeeping genes (ACTB, GAPDH, PPIA, PGK1) as the control in the illustration.

Please cite this article in press as: Tang et al., Joint analysis of chromatin accessibility and gene expression in the same single cells reveals cancer-specific regulatory programs, Cell Systems (2025), https://doi.org/10.1016/j.cels.2025.101266

Overlap of peak-gene associations and mutation-gene associations with H3K27ac HiChIP data

To evaluate the confidence of peak-gene associations, we mapped the gene promoter and associated peaks to the loops identified by H3K27ac HiChIP of the corresponding cell line.[107] We then compared the overlap with that identified from random peaks by Fisher's exact test. The peak-gene associations were significantly enriched in HiChIP loops in the K562 cell line (odds ratio = 7.13, p-value = 10^{-83}) and the GM12878 cell line (odds ratio = 6.82, p-value = 10^{-37}).

For mutation-gene associations identified from lung tumors, mutation-gene associations mapped to HiChIP interactions from any of the three cell lines (K562, GM12878, and CD4 naïve T cells) were regarded as supported by HiChIP.

Comparison of mutation-gene association analysis between Parallel-seq and 10x Multiome

To evaluate the performance of mutation-gene association analysis across different technologies, we analyzed publicly available 10x Multiome data from B-cell lymphoma samples (obtained from the 10x Genomics website) and compared it with Parallel-seq data from lung tumor samples. Although these datasets originate from different tumor types, this comparison provides insights into the relative capabilities of the two approaches.

For the 10x Multiome dataset, we annotated the major cell types, including B cells, T/NK cells, and macrophages/monocytes (Figure S7E). Based on CNV analysis, we further classified the B-cell population into 792 normal B cells and 2,436 tumor cells (Figures S7F and S7G). Using our mutation-gene association identification pipeline, we detected 13,614 mutations, each supported by at least five cells carrying the wild-type allele and five carrying the mutated allele (consistent with the threshold applied to the Parallel-seq data). Among these, 772 mutations were associated with differential expression of their corresponding target genes, leading to 1,670 mutation-gene associations.

For comparison, in lung tumor samples profiled by Parallel-seq, we identified 2,661 and 1,669 mutations in tumor cells from LT17 and LT18 patients, respectively. Among these, 320 and 219 mutations were associated with differential expression of their target genes, resulting in 884 and 400 mutation-gene associations, respectively.

Approximately 30% of the enhancer mutation-gene associations identified in the Parallel-seq data were supported by H3K27ac HiChIP, whereas only 10% of those identified in the 10x Multiome dataset had such support (Figure S7H).[63,67] Although this comparison involves different tumor types, the results suggest that Parallel-seq may provide higher sensitivity in detecting functionally relevant mutation-gene associations.