Integration of single-cell transcriptome and chromatin accessibility and its application on tumor investigation

Chunyuan Yang¹, Yan Jin¹, and Yuxin Yin¹,²,*

¹Institute of Systems Biomedicine, Department of Pathology, School of Basic Medical Sciences, Peking University, Peking-Tsinghua Center for Life Sciences, Peking University Health Science Center, Beijing 100191, China
²Institute of Precision Medicine, Peking University Shenzhen Hospital, Shenzhen 518036, China
*Correspondence: yinyuxin@hsc.pku.edu.cn
Abstract

The advent of single-cell sequencing techniques has not only revolutionized the investigation of biological processes but also significantly contributed to unraveling cellular heterogeneity at unprecedented levels. Among the various methods, single-cell transcriptome sequencing stands out as the best established, and has been employed in exploring many physiological and pathological activities. The recently developed single-cell epigenetic sequencing techniques, especially chromatin accessibility sequencing, have further deepened our understanding of gene regulatory networks. In this review, we summarize the recent breakthroughs in single-cell transcriptome and chromatin accessibility sequencing methodologies. Additionally, we describe current bioinformatic strategies to integrate data obtained through these single-cell sequencing methods, and highlight the application of this analysis strategy on a deeper understanding of tumorigenesis and tumor progression. Finally, we also discuss the challenges and anticipated developments in this field.

Key words single-cell transcriptome sequencing (scRNA-seq), single-cell chromatin accessibility sequencing, tumor microenvironment, single-cell multi-omics, assay for transposase-accessible chromatin (ATAC)
Introduction

Gene expression serves as the foundation of most cellular activities, which is comprised of transcription, translation, and then, in many cases, post-translational modifications. Transcription is tightly regulated by the activity of transcription factors (TFs). The genomic DNA is tightly packed with histones to form nucleosomes, and in a quiescent state, is inaccessible to TFs [1]. Therefore, genome accessibility emerges as one of the most crucial indicators for gene regulatory network (GRN). The establishment of sequencing technologies detecting chromatin accessibility has transformed our comprehensions of cellular activities, which includes chromatin immunoprecipitation followed by sequencing (CHIP-seq) [2], Deoxyribonuclease I (DNase I)-hypersensitive site sequencing (DNase-seq) [3], assay for transposase-accessible chromatin with sequencing (ATAC-seq) [4], and formaldehyde-assisted isolation of regulatory elements sequencing (FAIRE-seq) [5]. However, these methods primarily rely on bulk tissues, limiting their application on cellular heterogeneity investigations. Consequently, there arises a compelling need for the employment of single-cell sequencing techniques on the analyses of complex tissues.

Cancer ranks as one of the most lethal diseases worldwide, and improvements in tumor treatments are limited primarily due to the great cellular heterogeneity within bulk tumors. In recent decades, the application of single-cell RNA sequencing (scRNA-seq) has broadened the understanding of both tumor cells and tumor microenvironments (TMEs) [6]. This knowledge has revolutionized tumor treatment strategies by providing novel potential treatment targets. Nevertheless, intricate mechanisms underlying these transcriptomic alterations and TME dysfunctions remain incompletely understood.

Here, we review the advancements of scRNA-seq and scATAC-seq, provide an overview of current data integration methods, and elucidate the advantages of employing this single-cell multi-omics strategy on tumor investigations. Finally, we discuss present challenges and future perspectives for the application of single-cell multi-omics on cancer research.
Progress in single-cell transcriptomic sequencing

Since the introduction of scRNA-seq method by Tang et al. in 2009 [7], this sequencing technology has rapidly garnered global attention. Based on different cell division strategies, scRNA-seq methods can be classified into plate-based assays and droplet-based assays [8]. In plate-based assays, cells are divided into micro-wells using flow cytometry, followed by RNA extraction and DNA library construction within the same micro-well. This methodology allows the sequencing of full-length transcripts, which helps to distinguish RNA at allele and isoform resolution. Over the past decade, the application of plate-based scRNA-seq techniques have expanded from coding transcripts to non-coding transcripts, including long non-coding, short non-coding and non-polyadenylated protein-coding transcripts [9]. However, limited by the flow cytometry cell separation strategy, the throughput of plate-based scRNA-seq is relatively restricted. On the other hand, droplet-based assays employ a microfluid system to simultaneously separate and label cells with unique identifiers, facilitating high-throughput sequencing. The recently developed Smart-seq3 balanced the long-read sequencing and high-throughput sequencing requirements, enabling large-scale characterization of cell types and states [10]. Notably, the Chromium platform by 10× Genomics [11] has been dedicated to increasing sequencing throughput, currently allowing the simultaneous sequencing of up to one million cells collected from up to 128 specimens in a single run. This remarkable high-throughput method has significantly accelerated the discovery of rare cell populations. Considering the distinct advantages of plate-based assays and droplet-based assays, both techniques have found wide applications in the past decade (Fig. 1), leading to enormous advancements in tumor biology.

Importantly, scRNA-seq stands as a pivotal tool for the identification of significant yet rare cell populations that are seldomly detected through bulk sequencing methods. Examples of such cells include cancer stem cells (CSC) [27] and treatment-resistant tumor cells [28]. In colorectal carcinoma (CRC), LGR5+ tumor cells have been proved to be the major CSC population, while metastatic sites are predominantly initiated by LGR5− cells [29, 30]. The transformation dynamics of these two populations in diseases remain elusive. Vasquez E.G. et.al. have uncovered the dynamic states between Lgr5+ve crypt base columnar cells with Lgr5−ve regenerative stem cells, and depicted different ratios of these cell sub-populations in malignant
and non-malignant diseases [31]. In gastric cancer, scRNA-seq has identified multiple subtypes of CSCs, including CD44V6 CD133+CD166+ cells, CD44V6 CD133+CD166+ cells, and their upstream cell populations, namely TROP2+CD133+CD166+ cells [32]. Furthermore, a sub-population of quiescent stem-like cells plays an important part in chemoresistance and poor outcomes of acute myeloid leukemia revealed by scRNA-seq. Drug resistant signaling pathways have also been identified in CRC [33]. Cells showing activated oxidative phosphorylation and ATP metabolic states predict chemoresistance potential [34]. Furthermore, scRNA-seq has greatly advanced our understanding of TME components, encompassing cancer-associated fibroblasts (CAFs), endothelial cells [35], and immune cells [36]. During the tumorigenesis process, fibroblasts interact with tumor cells, and were trained into CAFs. Through scRNA-seq, CAF heterogeneity in various types of solid tumors have been identified, including pancreatic adenocarcinoma (PDAC) [37, 38], breast cancer [39, 40], lung cancer [41], liver cancer [42], and so on. CAF sub-populations, including myofibroblastic CAFs, inflammatory CAFs, and antigen-presenting CAFs, regulate extracellular matrix remodeling, immune cell recruitment, immune cell activation regulation, and interactions with tumor cells [43]. As for endothelial cells, scRNA-seq has revealed critical metabolic remodeling characteristics during tumorigenesis [44]. Tumor-specific tip-like endothelial cells participant in angiogenesis and tumor immune escape, thereby promoting tumor progression across multiple types of tumor [45]. Furthermore, the understanding of immune cell sub-populations has also greatly benefitted from scRNA-seq. Sub-clusters of T cells [46], macrophages [47], dendritic cells [48, 49], B cells [50, 51], and neutrophils [52] have been identified. The discovery of tumor-specific immune statuses not only unravels TME heterogeneity, but also provides substantial new targets for tumor therapies.

**Progress in single-cell chromatin accessibility sequencing**

To fully depict GRN and trace back for transcriptional regulatory mechanisms, various epigenetic analysis methods have been developed. Over the past decade, sequencing techniques evaluating DNA methylation, histone modification, and chromatin accessibility are emerging. As an important gene regulatory mechanism, DNA methylation at CpG island detection using bisulfite conversion followed by next-generation sequencing profiles DNA methylome, whereas the low throughput and high cost limited the wide application of this technique [53, 54]. Histone
modification sequencing techniques, including CHIP-seq [55], and subsequent cleavage under targets and tagmentation (CUT&Tag) [56], assess DNA-protein interaction to evaluate gene regulatory mechanisms. However, the intricate protocols and high costs associated with these methods increase the difficulty of broad utilization. In contrast, chromatin accessibility sequencing detects cis-regulatory element activity and TF binding sites, thereby unraveling gene expression regulatory network. Among the various techniques, ATAC-seq has gained widespread use due to its ease of operation, high throughput, and relatively low cost [4]. Hence, we focus on chromatin accessibility detection, especially using ATAC-seq, to unveil gene regulatory mechanisms.

The process of chromatin accessibility detection is composed of two steps: assigning TFs to open chromatin regions, namely peaks in many methods, and subsequent assignment of these open chromatin regions to specific genes. Based on these two processes, many bulk chromatin accessibility methods have been developed [2-5]. However, given the cellular complexity of most tissues, it is not surprising that rare cell types are frequently neglected using these bulk methods. Novel methods that can detect cellular chromatin accessibility at single-cell level are urgently needed. As scRNA-seq techniques began to emerge, the detection of chromatin accessibility began to extended from bulk methods to single-cell methods. Single-cell chromatin accessibility techniques, such as scATAC-seq, facilitate the discovery of multiple cellular states regulated by different TFs, and introduce the concept of “regulome” into the assessment of cellular variation [57]. Cell-type-specific chromatin accessibility features have also been identified [58].

The application of these single-cell chromatin accessibility methods has provided novel insights into various physiological and pathological processes [59-64]. In tumors, single-cell chromatin accessibility sequencing has been applied on the investigation of cis-elements in both tumor cells and TME components. During lung cancer tumor progression, chromatin accessibility modules transform from normal to metastasis [65]. Additionally, chromatin accessibility sequencing unravels epigenetic regulators in both tumor cell and tumor-infiltrating immune cell of basal cell carcinoma. A comparison of T cell status before and after programmed cell death protein 1 blockade identifies critical chromatin accessibility regulators during T cell exhaustion [66]. In terms of tumor treatment, chimeric antigen receptor T cells (CAR-T) therapy
has shown promising therapeutic efficacy in various tumor types but may be hindered by T cell differentiation after injection. However, single-cell chromatin accessibility sequencing helps to improve treatment efficacies. Using scATAC-seq, Jiang et al. identified chromatin accessibility alterations in differentiated T cells, initiated by BATF and IRF4. Knockdown of BATF and IRF4 inhibited T cell exhaustion and prolonged the validity of CAR-T therapy [67]. In conclusion, chromatin accessibility sequencing provides novel insights into tumor characteristics, facilitates patient prognosis evaluation, and aids in assessing treatment effectiveness [65-67], thereby indicating a critical dimension for tumor investigation.

Over the past decade, the throughput of scATAC-seq techniques has increased from hundreds [57] to hundreds of thousands (>10⁵ cells) [68], greatly accelerating the discovery of chromatin accessibility during cell-to-cell transitions and prompting the identification of rare cell states. Furthermore, considering the critical functions of mitochondria in cell biology, the investigation of DNA accessibility detection has expanded from genome DNA to mitochondrial DNA [69], introducing a new dimension for GRN investigations (Fig. 2). Notably, to comprehensively depict a whole GRN, a series of multi-omics methods have emerged, including the combinatorial detection of single-cell chromatin accessibility with single-cell transcriptome, single-cell DNA methylation, single-cell genome sequencing, and/or single-cell proteome (Table 1). The comprehensive use of these sequencing techniques not only provide a full landscape of GRN but also helps to identify critical transcriptomic regulators.

Integration of single-cell transcriptome and chromatin accessibility data

Given the distinct types of data obtained for single-cell transcriptome versus chromatin accessibility sequencing, integration of these data types remains a challenging task. From the perspective of cell sources, a comprehensive analysis of scRNA-seq and single-cell chromatin accessibility data can be categorized into either paired sequencing and unpaired sequencing [93]. Paired sequencing data comes from simultaneous detection of both transcriptome and chromatin accessibility, as exemplified by 10x Genomics Chromium Single-Cell Multiome ATAC + Gene Expression products [94, 95]. Multifunctional scATAC-seq analysis packages, especially ArchR [96] and signac [97], offer robust chromatin accessibility analyses as well as direct projection
between transcriptome and GRN. The use of paired data analysis tools minimizes the risk of neglecting GRNs in rare cell populations.

However, from a data mining perspective, there is an expectation to match single-cell transcriptome data and single-cell chromatin accessibility data from unpaired cells, such as from those originating from different sequencing experiments of the same organ. Numerous challenges arise in integrating transcriptomic and chromatin accessibility data from unpaired cells. First, the inherent heterogeneities among samples. Individual heterogeneity potentially brings great differences in cell type (or cell state) composition even within the same organ, underlying the fundamental cause of data integration dilemma. Second, sequencing batch effects. The application of different sequencing platforms, distinct sample preparation strategies, variations in sequencing depth all influence the matching of cells in transcriptomics and chromatin accessibility data. To overcome these challenges, various mathematical methods have been introduced to align these two types of data, including manifold alignment (AIscEA [98]), variational autoencoders (GLUE [99]), probabilistic modelling (MultiVI [100]), and more (Table 2). The combination of single-cell transcriptome data and single-cell chromatin accessibility data leverages the accurate cell type annotation of scRNA-seq and GRN investigation of single-cell chromatin accessibility sequencing (Fig. 3). Altogether, this comprehensive sequencing strategy facilitate the discovery of GRNs in biological and pathological processes, including developments [101-103], immune reactions [104, 105], chemotherapy responses [106], and tumorigenesis, which will be detailed in the following section.

**Application of single-cell transcriptome and chromatin accessibility assays on tumor investigations**

The heterogeneity of tumor cells and the complexity of the TME are considered the leading causes of therapeutic failure. Hence, deciphering the development of tumors at single-cell level is the foundation of precision medicine. In the past few years, researches using scRNA-seq methods have made substantial inroads into a better understanding of TME heterogeneity, by enabling the compilation of atlases of the TME of many different tumor types [107]. Nevertheless, mechanisms underlying of gene regulatory networks are frequently neglected. Recently, analysis of a combination of single-cell transcriptome and chromatin accessibility
information has improved our understanding of tumor cells and the TME and may be summarized as follows (Table 3).

The employment of single-cell transcriptome and chromatin accessibility strategy has enabled a more comprehensive exploration of the mechanisms of tumorigenesis. For example, compared with normal cells, malignant acute leukemic cells express higher levels of select TFs including RUNX1, FOS, and NFE2 [108]. In addition, in chronic lymphocytic leukemia, a new subtype of B cell has been identified through comprehensive analysis of scRNA-seq and scATAC-seq, which is featured by high expression level of TOX2 [109]. Moreover, as polyps progress to colorectal cancer, the proportion of stemness cells gradually increases, and pre-CAFs as well as regulatory T cells also show evolution patterns. Chromatin accessibility of the RUNX1 gene in CAF and HNF4A gene in epithelial cells suggests the underlying mechanisms of CRC development [110]. Comprehensive analysis of scRNA-seq and scATAC-seq data has unraveled unique chromatin accessibility features of TME in clear cell renal cell carcinoma, and identified critical prognostic-indicating TFs [111]. Intriguingly, compared with tumor heterogeneity at single-cell transcriptome level, single-cell chromatin accessibility-level heterogeneity has been found to be greater in multiple myeloma, implying a greater sensitivity of epigenomic methods for detecting tumors [112].

Integrative use of single-cell transcriptomic and chromatin accessibility sequencing on describing tumor progression has also identified many important regulators. Liver metastases of colorectal carcinoma have been found to resemble liver cancer instead of CRCs at both transcriptomic and epigenetic levels. The chromatin accessibility of HNF4A is significantly elevated in liver metastasis, suggesting an important role for HNF4A in colorectal carcinoma progression [113]. Epigenetic remodeling of ATRX-deletion glioma has been identified to be absent of CCCTC-binding sites, thereby showing treatment resistance [114]. In another example, treatment effectiveness has been shown to be influenced by CAFs. Using both scRNA-seq and scATAC-seq, three subtypes of cancer associated fibroblasts were identified, and the dynamics among these fibroblast subtypes have been shown to correlate with the prognosis of several cancers, including skin cancer, breast cancer, and pancreatic cancer [115]. Importantly, pan-cancer investigation using both snATAC-seq and scRNA-seq have revealed the cancer-specific differentially accessible chromatin regions (DACRs) during cancer evolution, these DACRs...
primarily locate in cis-regulatory regions, including 53% in enhancers and 37% in promoters. Cancer-type specific TFs have also been revealed, including KLF6 and PPARG in PDAC. Furthermore, compared with primary tumors, metastatic tumors show unique TFs. TWIST1 and PBX3 show higher accessibilities during CRC metastasis. GNA13 shows higher accessibility during melanoma metastasis [116].

Taken together, these examples demonstrate that the combination of single-cell transcriptome and chromatin accessibility sequencing uncovers the tumor-specific chromatin accessibility regions, TFs, transcription regulators, and novel cell types [117], revolutionizing the study of tumor biology (Fig. 4). However, considering the vast heterogeneity of tumor samples, epigenetic characteristics should be observed in a large number of patients from around the world. The etiology of tumors can dramatically differ among patients from different countries. For example, while hepatitis virus infection is a predominant cause of liver cancer in Asia, alcohol and non-alcoholic steatohepatitis prevail as major causes in Western countries. Previous single-cell transcriptomic sequencing results have shown consistent gene expression characteristics in liver cancer worldwide, regardless of the underlying causes [35]. Nevertheless, investigations into the GRNs of tumors from diverse regions remain unexplored. Therefore, future endeavors should encompass a broader spectrum of districts to comprehensively elucidate tumorigenesis mechanisms.

Conclusions

Developments in single-cell transcriptomic sequencing and chromatin accessibility sequencing techniques, coupled with the evolutions of data integration tools, have paved the way of GRN investigation in biological and pathological processes, particularly in tumors. This approach has led to the identification of novel cell sub-populations, cell type-specific TFs, transcription regulators, and chromatin regions associated with pathological conditions, all playing critical roles in tumorigenesis. Altogether, the comprehensive assessment of single-cell transcriptome and chromatin accessibility characteristics enhances our understanding of tumors both in baseline conditions and under treatments.
Despite the numerous achievements in sequencing and data analysis, several challenges persist. From the point of transcriptomic or chromatin accessibility sequencing, the presence of alternative splicing and multiple transcription variants complicates the accurate assessment of cell states. Failure to distinguish between transcription variants may result in inaccurate interpretations of sequencing data. From the point of data integration, despite the availability of various tools, the choice of analysis methods can introduce biases to results. The recently developed data integration tools primarily handle two types of input data: trajectories or groups. The former is more suitable for continuous data analysis, and has been employed to development assessments. In contrast, the latter is more suitable for discrete data analysis, and applies to the investigation of pathological conditions, including tumorigenesis. Given that tumorigenesis and tumor developments are progressive pathologic processes, accurate tool selection at different steps is paramount. More importantly, from the point of application, although there have been various single-cell multi-omics techniques, high cost and intricate experimental protocols limit the widespread of many well-performed techniques. To overcome these challenges, researchers must familiarize themselves with the developments and mechanisms of sequencing techniques as well as data processing methods. Only by deep understandings of both experiments and bioinformatic analyses can the researchers choose adequate approaches to address their scientific inquiries. Furthermore, improving the level of industry commercialization is essential to provide more options for the majority of researchers. Finally, with the advances in single-cell multi-omics sequencing and analysis methods, we anticipate deeper insights into tumorigenesis mechanisms and tumor treatment options.
Acknowledgements

We thank Prof. Susan P.C. Cole for assistance with manuscript editing. This work was supported by grants including the National Natural Science Foundation of China (Nos. 31420103905, 81321003, 81430056, 81372491 to Y.Y.), the Beijing Municipal Science and Technology Commission (No. 7161007 to Y.Y.), and the Lam Chung Nin Foundation for Systems Biomedicine.

Author contributions

Conceptualization, C.Y., and Y.Y.; writing-original draft preparation, C.Y., and Y.Y.; writing-review and editing, C.Y., Y.J. and Y.Y. All authors have read and agreed to the published version of the manuscript.

Conflict of interest

The authors declare no conflict of interests.
References


Granja JM, Corces MR, Pierce SE, Bagdatli ST, Choudhry H, Chang HY, et al. ArchR is a


Table 1. Single-cell multi-omics based on chromatin accessibility sequencing methods.

<table>
<thead>
<tr>
<th>Number of omics</th>
<th>Type of omics</th>
<th>Method</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two omics data integration</td>
<td>Chromatin accessibility + DNA methylation</td>
<td>scNOMe-seq</td>
<td>[72]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>scCOOL-seq</td>
<td>[74]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>iscCOOL-seq</td>
<td>[82]</td>
</tr>
<tr>
<td>Chromatin accessibility + Transcriptome</td>
<td>sci-CAR</td>
<td>Perturb-ATAC</td>
<td>[75]</td>
</tr>
<tr>
<td>Proteome</td>
<td></td>
<td>CRISPR-sciATAC</td>
<td>[88]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ASAP-seq</td>
<td>[81]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Spear-ATAC</td>
<td>[83]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PHAGE-ATAC</td>
<td>[85]</td>
</tr>
<tr>
<td>Chromatin accessibility + Genome</td>
<td>scGET-seq</td>
<td>mtscATAC-seq</td>
<td>[69]</td>
</tr>
<tr>
<td>Three omics data integration</td>
<td>Chromatin accessibility + DNA methylation + Transcriptome</td>
<td>scNMT-seq</td>
<td>[78]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>scChaRM-seq</td>
<td>[86]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>snmCAT-seq</td>
<td>[90]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>scNOMeRe-seq</td>
<td>[92]</td>
</tr>
<tr>
<td>Chromatin accessibility + Transcriptome + Proteome</td>
<td>TEA-seq</td>
<td>DOGMA-seq</td>
<td>[81]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NEAT-seq</td>
<td>[87]</td>
</tr>
</tbody>
</table>
Table 2. Representative methods for integrating scRNA-seq and scATAC-seq data

<table>
<thead>
<tr>
<th>Tool</th>
<th>Method</th>
<th>Data Type</th>
<th>Language</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIScEA</td>
<td>Manifold alignment</td>
<td>Unpaired</td>
<td>Python</td>
<td>[98]</td>
</tr>
<tr>
<td>MMD-MA</td>
<td>Manifold alignment</td>
<td>Unpaired</td>
<td>Python</td>
<td>[118]</td>
</tr>
<tr>
<td>Pamona</td>
<td>Manifold alignment</td>
<td>Unpaired</td>
<td>Python</td>
<td>[119]</td>
</tr>
<tr>
<td>SCOT</td>
<td>Gromov-Wasserstein optimal transport</td>
<td>Unpaired</td>
<td>Python</td>
<td>[120]</td>
</tr>
<tr>
<td>UnionCom</td>
<td>Manifold alignment</td>
<td>Unpaired</td>
<td>Python</td>
<td>[121]</td>
</tr>
<tr>
<td>sc-compReg</td>
<td>Correlation-based</td>
<td>Unpaired</td>
<td>R</td>
<td>[109]</td>
</tr>
<tr>
<td>Seurat v3</td>
<td>Correlation-based</td>
<td>Unpaired</td>
<td>R</td>
<td>[122]</td>
</tr>
<tr>
<td>scMTNI</td>
<td>Probabilistic graphical model-based</td>
<td>Unpaired</td>
<td>C++</td>
<td>[123]</td>
</tr>
<tr>
<td>SOMatic</td>
<td>Toroid topology-based</td>
<td>Unpaired</td>
<td>C++/R</td>
<td>[103]</td>
</tr>
<tr>
<td>Dictys</td>
<td>Stochastic process model</td>
<td>Unpaired/paired</td>
<td>Python</td>
<td>[101]</td>
</tr>
<tr>
<td>GLUE</td>
<td>Variational autoencoders</td>
<td>Unpaired/paired</td>
<td>Python</td>
<td>[99]</td>
</tr>
<tr>
<td>Inferelator 3.0</td>
<td>Regularized regression</td>
<td>Unpaired/paired</td>
<td>Python</td>
<td>[124]</td>
</tr>
<tr>
<td>MultiVI</td>
<td>Probabilistic modelling</td>
<td>Unpaired/paired</td>
<td>Python</td>
<td>[100]</td>
</tr>
<tr>
<td>ArchR</td>
<td>Corresponding barcode</td>
<td>Paired</td>
<td>R</td>
<td>[96]</td>
</tr>
<tr>
<td>DeepMAPS</td>
<td>Autoencoder-like neural networks</td>
<td>Paired</td>
<td>Python</td>
<td>[125]</td>
</tr>
<tr>
<td>DIRECT-NET</td>
<td>Constrained optimal cell mapping</td>
<td>Paired</td>
<td>R</td>
<td>[126]</td>
</tr>
<tr>
<td>FigR</td>
<td>Constrained optimal cell mapping</td>
<td>Paired</td>
<td>R</td>
<td>[127]</td>
</tr>
<tr>
<td>GRaNIE</td>
<td>Pearson’s correlation</td>
<td>Paired</td>
<td>R</td>
<td>[104]</td>
</tr>
<tr>
<td>MAGICAL</td>
<td>Hierarchical Bayesian framework</td>
<td>Paired</td>
<td>R</td>
<td>[105]</td>
</tr>
<tr>
<td>Pando</td>
<td>Minimum-cost maximum-flow bipartite matching</td>
<td>Paired</td>
<td>R</td>
<td>[102]</td>
</tr>
<tr>
<td>scAI</td>
<td>Matrix factorization</td>
<td>Paired</td>
<td>R</td>
<td>[106]</td>
</tr>
<tr>
<td>SCENIC+</td>
<td>Unsupervised identification model</td>
<td>Paired</td>
<td>R</td>
<td>[128]</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>signac</td>
<td>Corresponding barcode</td>
<td>Paired</td>
<td>R</td>
<td>[97]</td>
</tr>
</tbody>
</table>
Table 3. Application of single-cell transcriptome and chromatin accessibility methods on tumor investigations

<table>
<thead>
<tr>
<th>Cancer type</th>
<th>Sample source</th>
<th>Main findings</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute leukemia</td>
<td>Malignant and normal blood cells</td>
<td>identification of key TFs (RUNX1, FOS, and NFE2)</td>
<td>[108]</td>
</tr>
<tr>
<td>Bladder cancer</td>
<td>Primary and recurrent tumor tissue</td>
<td>Identification of transcription regulator (EZH2) and key TF (TCF7)</td>
<td>[129]</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>Primary and metastatic tumor tissue</td>
<td>Identification of chromatin accessible regions in CXCL14-high cancer cells and immune cells</td>
<td>[130]</td>
</tr>
<tr>
<td>Chronic lymphocytic leukemia</td>
<td>Malignant and normal blood cells</td>
<td>Identification of a new B cell subtype featured by TOX2 regulatory function</td>
<td>[109]</td>
</tr>
<tr>
<td>Clear cell renal carcinoma</td>
<td>Tumor and adjacent tissues</td>
<td>Identification of key TFs (HOXC5, VENTX, ISL1, and OTP)</td>
<td>[111]</td>
</tr>
<tr>
<td>Clear cell renal carcinoma</td>
<td>Different tumor subtypes</td>
<td>Identification of distinct chromatin accessibility features in different cancer subtypes</td>
<td>[131]</td>
</tr>
<tr>
<td>Colorectal cancer</td>
<td>Primary and metastatic tumor tissue</td>
<td>Chromatin accessibility shifts upon tumor metastasis, and identification of key TF (HNF4A)</td>
<td>[113]</td>
</tr>
<tr>
<td>Colorectal cancer</td>
<td>Different stages of tumor</td>
<td>Chromatin accessibility shifts during tumor development, and identification of key TF (RUNX1)</td>
<td>[110]</td>
</tr>
<tr>
<td>Colorectal cancer</td>
<td>Pre- and post-chemotherapy tumor tissue</td>
<td>Chromatin accessibility changes upon drug treatment regulating metabolic states</td>
<td>[132]</td>
</tr>
<tr>
<td>Tumor Type</td>
<td>Subtype</td>
<td>Chromatin Accessibility Feature</td>
<td>Reference</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>--------------------------------------</td>
<td>----------------------------------------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Glioma</td>
<td>Different tumor subtypes</td>
<td>Chromatin accessibility difference (CCCTC binding site deletion) according to differences in ATRX expression status</td>
<td>[114]</td>
</tr>
<tr>
<td>Male breast cancer</td>
<td>Different types of tumor</td>
<td>Identification of differences in chromatin accessibility of MYC cis-regulatory elements in male and female breast cancer</td>
<td>[133]</td>
</tr>
<tr>
<td>Multiple myeloma</td>
<td>Subtypes of tumor</td>
<td>Tumor evolution tree mapping according to chromatin accessibility and transcriptome</td>
<td>[134]</td>
</tr>
<tr>
<td>Multiple myeloma</td>
<td>Post-chemotherapy tumor cells</td>
<td>Chromatin accessibility features in venetoclax-resistant multiple myeloma</td>
<td>[135]</td>
</tr>
<tr>
<td>Ovarian and endometrial tumor</td>
<td>Different tumor subtypes</td>
<td>Identification of greater heterogeneity in chromatin accessibility relative to transcriptome</td>
<td>[112]</td>
</tr>
<tr>
<td>Multiple (skin, breast, and pancreatic cancer)</td>
<td>Different tumor types</td>
<td>Identification of a cancer-associated fibroblast subtype according to chromatin accessibility</td>
<td>[115]</td>
</tr>
<tr>
<td>Retinoblastoma</td>
<td>Tumor at different stages</td>
<td>Tumor evolution trajectory, chromatin accessibility identification of RB-positive tumor-initiating cells</td>
<td>[136]</td>
</tr>
</tbody>
</table>
Figure legends

Figure 1. Time line of single-cell RNA sequencing methods.

Representative droplet-based (left) and plate-based (right) scRNA-seq techniques are listed from top to bottom according to their invention time, with corresponding references listed behind.

Figure 2. Time line of single-cell chromatin accessibility sequencing methods.

Representative droplet-based (left) and plate-based (right) single-cell chromatin accessibility sequencing techniques are listed from top to bottom according to their invention time, with corresponding references listed behind.

Figure 3. Cell clustering by different single-cell sequencing techniques.

Shown are representative illustrations of (A) the complex cellular composition within tissues, (B) cell clustering using scRNA-seq methods, (C) cell clustering by single-cell chromatin accessibility sequencing methods, and (D) cell clustering by combining single-cell transcriptome and single-cell chromatin accessibility methods.

Figure 4. Graphical abstract.

Schematic graph showing the application of single-cell transcriptome sequencing and single-cell chromatin accessibility sequencing in tumor investigation, which prompts the identification of tumor-specific chromatin accessible regions, transcription factors (TFs), transcription regulators (TRs), and new cell types.
Figure 1

Droplet-based assays

- STRT-seq [12]
- Drop-seq [13]
- inDrop-seq [15]
- SC3-seq [17]
- CytoSeq [19]
- 10X Genomics Chromium [11]
- Seq-well [22]
- Smart-seq3 [10]

Plate-based assays

- CEL-seq [14]
- Smart-seq [16]
- Smart-seq2 [18]
- Quartz-seq [20]
- MARS-seq [21]
- CEL-seq2 [23]
- Microwell-seq [24]
- SCAN-seq [25]
- VASA-seq [9]
- MAS-ISO-seq [26]
Figure 2

Droplet-based assays

<table>
<thead>
<tr>
<th>Year</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015</td>
<td>Single cell ATAC-seq [70]</td>
</tr>
<tr>
<td>2017</td>
<td>scATAC-seq [57]</td>
</tr>
<tr>
<td>2018</td>
<td>SCNOMe-seq [72]</td>
</tr>
<tr>
<td>2019</td>
<td>scCOOL-seq [74]</td>
</tr>
<tr>
<td>2020</td>
<td>sci-CAT-seq [76]</td>
</tr>
<tr>
<td>2021</td>
<td>scNMT-seq [78]</td>
</tr>
<tr>
<td>2022</td>
<td>iscCOOL-seq [82]</td>
</tr>
<tr>
<td>2023</td>
<td>SHARE-seq [60]</td>
</tr>
</tbody>
</table>

Plate-based assays

<table>
<thead>
<tr>
<th>Year</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015</td>
<td>scATAC-seq [57]</td>
</tr>
<tr>
<td>2017</td>
<td>SCNOMe-seq [72]</td>
</tr>
<tr>
<td>2018</td>
<td>scCOOL-seq [74]</td>
</tr>
<tr>
<td>2019</td>
<td>sci-CAT-seq [76]</td>
</tr>
<tr>
<td>2020</td>
<td>scNMT-seq [78]</td>
</tr>
<tr>
<td>2021</td>
<td>iscCOOL-seq [82]</td>
</tr>
<tr>
<td>2022</td>
<td>SHARE-seq [60]</td>
</tr>
<tr>
<td>2023</td>
<td>SCNOMeRe-seq [92]</td>
</tr>
</tbody>
</table>

Other methods:

<table>
<thead>
<tr>
<th>Year</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015</td>
<td>mtscATAC-seq [69]</td>
</tr>
<tr>
<td>2016</td>
<td>Spear-ATAC [83]</td>
</tr>
<tr>
<td>2017</td>
<td>PHAGE-ATAC [85]</td>
</tr>
<tr>
<td>2018</td>
<td>NEAT-seq [87]</td>
</tr>
<tr>
<td>2019</td>
<td>scGET-seq [89]</td>
</tr>
<tr>
<td>2020</td>
<td>ISSAAC-seq [91]</td>
</tr>
<tr>
<td>2021</td>
<td>CRISPR-sciATAC [88]</td>
</tr>
<tr>
<td>2022</td>
<td>smCAT-seq [90]</td>
</tr>
<tr>
<td>2023</td>
<td>sciATAC [68]</td>
</tr>
</tbody>
</table>
Figure 3

A  Tissue cell composition

B  Clustering by transcriptome

C  Clustering by chromatin accessibility

D  Clustering by transcriptome + chromatin accessibility
Figure 4

New cell types

Single cell transcriptome + Single cell chromatin accessibility