Review

Single-Cell Analysis Using Machine Learning Techniques and Its Application to Medical Research

Ken Asada [1,*], Ken Takasawa [1], Hidenori Machino [1], Satoshi Takahashi [1], Norio Shinkai [1,2], Amina Bolatkan [1,3], Kazuma Kobayashi [3], Masaaki Komatsu [1], Syuzo Kaneko [3], Koji Okamoto [4] and Ryuji Hamamoto [2,3,*]

1   Cancer Translational Research Team, RIKEN Center for Advanced Intelligence Project, 1-4-1 Nihonbashi, Chuo-ku, Tokyo 103-0027, Japan; ktakazaw@ncc.go.jp (K.T.); hmachino@ncc.go.jp (H.M.); satoshi.takahashi.fy@riken.jp (S.T.); norio.shinkai@riken.jp (N.S.); amina.bolatkan@riken.jp (A.B.); maskomat@ncc.go.jp (M.K.)
2   Department of NCC Cancer Science, Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, 1-5-45 Yushima, Bunkyo-ku, Tokyo 113-8510, Japan
3   Division of Medical AI Research and Development, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan; kazumkob@ncc.go.jp (K.K.); sykaneko@ncc.go.jp (S.K.)
4   Division of Cancer Differentiation, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan; kojokamo@ncc.go.jp
*   Correspondence: ken.asada@riken.jp (K.A.); rhamamot@ncc.go.jp (R.H.); Tel.: +81-3-3547-5271 (R.H.)

Abstract: In recent years, the diversity of cancer cells in tumor tissues as a result of intratumor heterogeneity has attracted attention. In particular, the development of single-cell analysis technology has made a significant contribution to the field; technologies that are centered on single-cell RNA sequencing (scRNA-seq) have been reported to analyze cancer constituent cells, identify cell groups responsible for therapeutic resistance, and analyze gene signatures of resistant cell groups. However, although single-cell analysis is a powerful tool, various issues have been reported, including batch effects and transcriptional noise due to gene expression variation and mRNA degradation. To overcome these issues, machine learning techniques are currently being introduced for single-cell analysis, and promising results are being reported. In addition, machine learning has also been used in various ways for single-cell analysis, such as single-cell assay of transposase accessible chromatin sequencing (ATAC-seq), chromatin immunoprecipitation sequencing (ChIP-seq) analysis, and multi-omics analysis; thus, it contributes to a deeper understanding of the characteristics of human diseases, especially cancer, and supports clinical applications. In this review, we present a comprehensive introduction to the implementation of machine learning techniques in medical research for single-cell analysis, and discuss their usefulness and future potential.

Keywords: single-cell analysis; next-generation sequencing; machine learning; multi-omics analysis

1. Introduction

Cells are the smallest units of life, and a wide variety of them exist in living organisms. Through the developmental differentiation process, they form tissues and organs with different lineages and functions, making the particular body a living organism. In addition, it operates biological activities through mutual reactions and intercellular communication in response to signals from inside and outside the cell, as well as environmental changes. Therefore, understanding life at the single-cell level is essential for medical research to understand and control diseases, which relate to alterations of the mechanisms of life. For example, cancer tissues are composed of cancer cells and their surrounding non-cancerous cells, such as cancer-associated fibroblasts, macrophages, other blood system cells, and blood endothelial cells, and these intercellular networks play an important role in cancer survival and proliferation [1]. In particular, cancer cells are not a population of cells with uniform characteristics; genetic mutations and epigenetic modifications give rise

to a population of cancer cells with different characteristics, such as cancer stem cells and differentiated cells [2,3]. Such cellular diversity in cancer tissues is the root of the contribution of malignant traits such as tumorigenicity, anticancer drug resistance, and metastatic potential of cancer cells, and it is considered to be the essential cause of refractory cancer. Therefore, analysis at a single-cell level is important because understanding cancer tissue diversity is expected to contribute to the establishment of cancer therapies.

An attempt to isolate cells from tissues and understand their properties of individual cells was first reported in 1992; the construction of an expression profile of a single living cell from the rat hippocampus was demonstrated through quantitative analysis using DNA gel blotting [4]. Subsequently, RT-PCR analysis in single cells was performed [5], and in 1998, fluorescence in situ hybridization (FISH) enabled quantitative analysis of mRNA molecules in single cells [6]. A comprehensive one-cell gene expression analysis was initially performed using microarray analysis [7]. Subsequently, a method that drastically improved the efficiency of cDNA synthesis from a single cell was reported [8]; based on that method, the first single-cell RNA-seq (scRNA-seq) method using next-generation sequencing was reported in 2009 [9]. This was the beginning of a rapid spread of research using comprehensive one-cell gene expression analysis. However, various problems have been identified as challenges in single-cell analysis. Among them, the critical issues are "sensitivity and accuracy" and "linkage with spatial information"; hence, the introduction of new technology is required.

In recent years, the expectations for artificial intelligence (AI) have increased greatly, and machine learning (ML) technology has made remarkable progress, including the rise of deep learning technology [10,11]. It is important to note that ML has been actively used in the fields of medical and biological research [12–23], and it has already been used in actual clinical practice [24–26]. Single-cell analysis is another field in which ML techniques have been used to overcome challenges and make more effective use of its results. Given the ability of ML technology to extract features from large-scale data, we believe that its use is essential for single-cell analysis, and it can be applied in various ways. In this review, we present a comprehensive overview of the use of ML techniques in the field of single-cell analysis and discuss their usefulness and future development.

2. History of Single-Cell Analysis

In 1992, gene expression analysis using a single neuron was first reported in a quantitative analysis using DNA gel blotting [4]. Subsequently, RT-PCR and FISH in single cells were reported [5,6], and in the late 1990s, microarray technology was used to analyze the expression of thousands of genes [27–29]. The central nervous system and neurons consist of multiple cell types; in other words, different regions correspond to different functions. Therefore, single-cell analyses have been actively investigated in this field. For example, the combination of two methodologies, i.e., laser capture microdissection for selective sample collection, and the aforementioned microarray technology for analyzing thousands of genes in a single sample, were used to disclose cellular landscapes [7,30–32].

Next-generation sequencing (NGS) technology is now widely used in laboratories as well as in hospitals to aid decision-making for guided precision medicine [33]. Additionally, it is necessary to use big data and multi-omics data obtained by NGS to achieve precision oncology [34]. However, NGS analysis requires hundreds to thousands of cells as a material; this creates a technical barrier as samples (or cells) are difficult to obtain. Thus, apart from the biological aspect, the first scRNA-seq analysis using the SOLiD system was reported from a realistic experimental point of view, and it was quickly and widely accepted [9]. Importantly, it was reported only a year after the first RNA-seq with bulk cell samples was published [35,36]. The scRNA-analysis introduced by Tang et al. was composed of two parts: (1) isolation of a single blastomere by mouth pipetting, and (2) cDNA synthesis from a single cell. Their method enables single-cell transcriptomic analysis that can detect 75% more gene expression than previous microarray-based methods. The basic approach as proposed by Tang et al., and especially the flow chart of cell isolation followed by NGS,

remains unaltered. Yet other researchers have improved their application. For example, Islam et al. developed a custom-built semi-automated cell picker that collects a single cell into each 96-well capture plate [37], while Kivioja et al. reported the addition of unique molecular identifiers (UMIs) to cDNA during a single-cell reverse-transcription reaction to minimize PCR amplification errors [38]. In this method, a random 10-base-pair DNA tag was added to the 5′ adaptor for sequencing. Theoretically, more than 100 million (4^{10} = 1,048,576) UMIs are available and they bind to each mRNA; therefore, we can accurately estimate the amount of input mRNA by counting UMIs.

The first reported scRNA-seq analysis using a microfluidic system (Fluidigm) could analyze up to 2000 single cells at a time [39], which laid the basis for high-throughput analysis. Pico-well and nano-droplet technology lowered the cost of scRNA-seq analysis and further facilitated high-throughput analysis [40–42]. In 2017, the combinatorial indexing method sci-RNA-seq (GemCode), which uses gel bead-in-emulsion (GEM) from 10x Genomics, enhanced the reproducibility and accelerated the high-throughput performance [43,44]. In Table 1, we summarize the widely accepted scRNA-seq methods.

Table 1. scRNA-seq methodology used for analysis.

scRNA Technology	Description	Characteristics	Year	References
1. STRT (Single-cell tagged reverse transcription)-Seq	This method is based on using reverse transcriptase that possesses template-switching activity.	High accuracy of the position of 5′-end mRNA. Low cost and short time owing to the barcode strategy.	2011	[37]
2. Smart (Switching mechanism at 5′ end of RNA template)-Seq		This method enriches 5′-end mRNA and provides robust and reproducible results. SMARTer Ultra Low RNA kit for Illumina sequencing is available.	2012	[45]
3. Smart-Seq2		Improvement of reverse transcription, template switching, and pre-amplification efficacy. Exchanging one single guanylate for a locked nucleic acid (LNA) at the template-switching oligonucleotides 3′ end leads to a two-fold increase in cDNA yield.	2013	[46]
1. CEL-Seq	This method is based on in vitro transcription to reduce PCR-induced amplification bias.	This method provides highly strand specific, reproducible, linear, and sensitive results compared with PCR-based amplification and allows detection of 3′-end mRNAs using in vitro transcription. Small amounts of input RNA can be used.	2012	[47]
2. CEL-Seq2		This method is improved to achieve less amplification bias of genome sequencing, higher sensitivity, lower cost, and less working time.	2016	[48]
1. Quartz-Seq	This is a combined method of suppression PCR with poly(A) tails.	Robust suppression of byproduct synthesis with the combined techniques of poly(A) tailing with PCR amplification.	2013	[49]
2. Quartz-Seq2		The efficiency of converting initial reads into unique molecular identifiers (UMIs) has been improved because of the major improvement of poly(A) tagging, allowing for the detection of more genes.	2018	[50]
Microfluidic platform	This method is based on a microfluidic platform; single cells are captured and lysed in a microfluidic device.	The analysis of individual cells can be automated and parallelized, and cDNA can be synthesized in small-scale reactions using low-input RNA.	2014	[51]
1. Drop-Seq	This method is based on a microfluidic device that creates droplets with a single cell and reagents (such as a bead).	Digital counting of mRNA in thousands of single cells is possible.	2015	[42]
2. inDrop		A theoretical capacity to barcode tens of thousands of cells in a single run, allowing randomly labeling 3,000 cells with 99% unique labeling; many more cells can be processed by splitting a large emulsion into separate tubes.	2015	[41]
Seq-Well	This method enables the detection of a single cell in a PDMS array of more than 80,000 subnanoliter wells.	This method achieves efficient cell lysis with rapid solution exchange, while increasing the capture rate of transcripts and reducing cross-contamination by trapping biological macromolecules.	2017	[52]

scRNA Technology	Description	Characteristics	Year	References
Microwell-seq	This method uses microwells, which are technically simple and cost-effective through using an inexpensive device (agarose plate) for scRNA.	A simple method to profile thousands of single cells utilizing an agarose-constructed microwell array and barcoded beads to establish a convenient, simple, and cost-effective single-cell technology. This method combined existing methodologies.	2018	[53]
RamDA-seq	This method can detect a full-length total RNA expression in a single cell.	This method is highly sensitive to non-poly(A) RNAs such as lncRNAs, covers near-complete full-length transcripts, and profiles recursive splicing in >300-kb introns, detects enhancer RNAs, and their cell type-specific activity in single cells.	2018	[54]
C1 CAGE	This method is a 5′ RNA-sequencing using a C1 microfluidic system and cap analysis gene expression (CAGE) technique.	This method is an automated scRNA-seq platform using the C1 system, which can quantitatively detect the 5′ end of transcripts without bias.	2019	[55]

The fundamental idea for conducting a single-cell analysis is mainly to profile cellular divergence, analyze a cell population, clarify heterogeneity of microenvironment, and map the developmental landscape spatiotemporally [56–66]. Additionally, function-associated methodologies such as Patch-seq, a technique that combines whole-cell patch-clamp recording, immunohistochemistry, and scRNA-seq to improve classification of neural cell types as well as functional characterization [67–69], and in situ hybridization for single-cell analysis to reveal mRNA splicing variants have been reported [70].

3. Single-Cell mRNA Analysis Using Machine Learning

Currently, single-cell analysis contributes significantly to not only transcriptomic analysis but also genomic analysis, such as in identifying single-nucleotide mutations in cancer [71–74] and epigenomics such as DNA methylome [75–77], ChIP-seq analysis and chromatin-accessibility [78–80] (Figure 1).

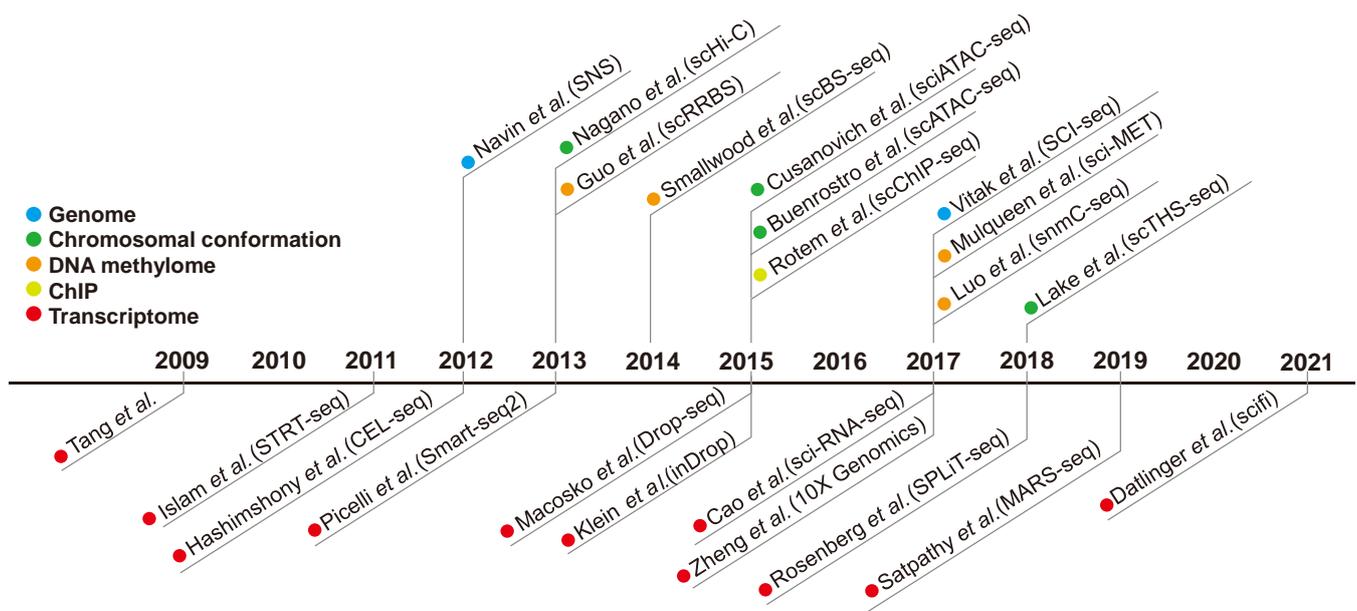
Figure 1. Timeline of technical progress of single-cell analysis. Each method was divided into five categories (genome, chromosomal conformation, DNA methylome, ChIP, and transcriptome) and identified by different colors.
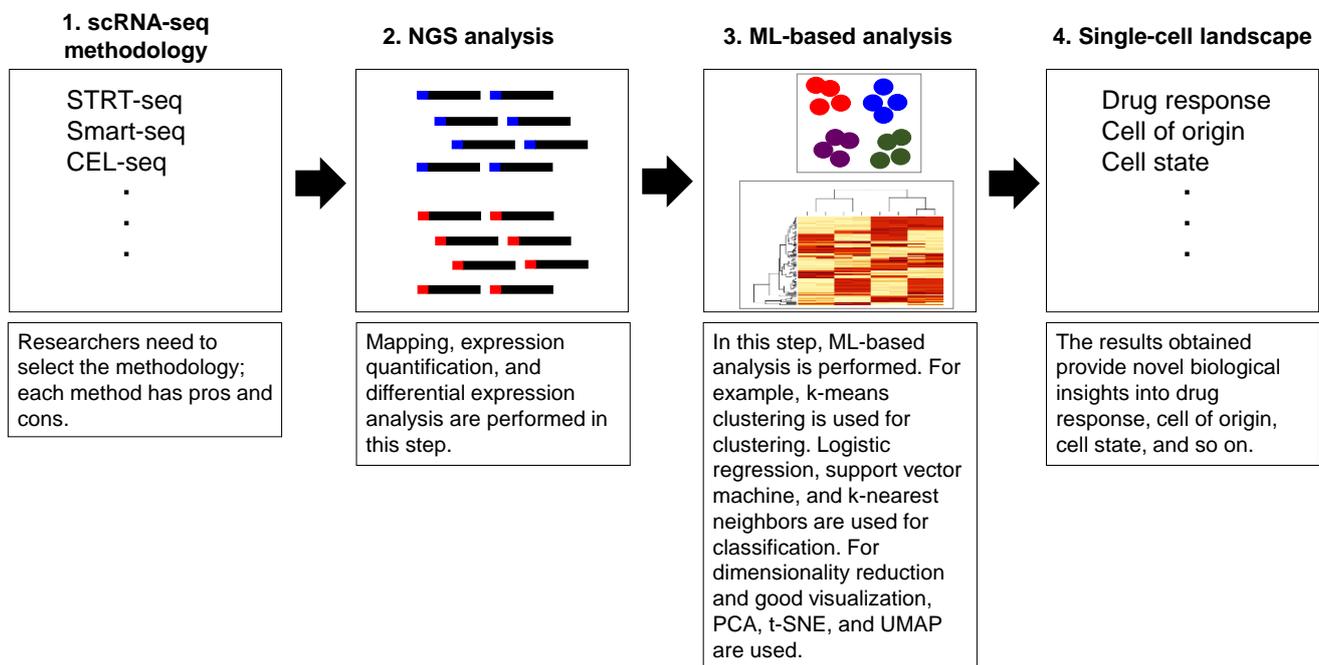
In this section, we discuss the technical aspects of scRNA-seq with ML. scRNA data can be obtained from whole cells or fractionated samples. Similar to the usual RNA-seq

methods, we can analyze scRNA-seq data from whole cell samples either to classify cells of origin or cell types [81]. For fractionated samples in particular, nuclear fractionation, known as single-nucleus RNA-seq (snRNA-seq), has been reported [82,83]; moreover, snRNA-seq is applicable even in frozen human tumor tissue samples [84]. Although we do not discuss the details here, single-cell technology utilizes other omics data such as single-cell DNA sequencing (scDNA-seq), which can detect copy number variations (CNVs) as well as clonal populations [72]; additionally, it can be used for single-cell whole genome analysis [85], DNA methylation [86], or in situ sequencing known as FISSEQ [87], MERFISH [88], and seqFISH [89]. Moreover, scRNA-seq driven classification in metastatic lung adenocarcinoma shows the feasibility of revealing cellular dynamics, tumor intrinsic factors, and novel signatures that are associated with patient outcome [90].

Generally, scRNA-seq is performed with the aim of cell-type identification (classification); it is important for metastatic tumors, cell state analysis to categorize immune cells, tumor microenvironment to analyze heterogeneity, or can be performed for expression subtype analysis (clustering). To achieve the aforementioned research purpose, scRNA analysis is often performed using ML techniques. The general implementation of scRNA-seq analysis is shown in Figure 2. Although scRNA-seq is a powerful tool to address many scientific interests, technically, different batches of scRNA-seq data cause batch effects that reflect different cell populations, protocols, or technical variations. To manage these datasets for analysis, the mutual nearest neighbors (MNNs) batch-effect correction method, deep transfer learning (BERMUDA) method, and another MNN-based method named Scanorama that identifies datasets with the same cell types to cluster (not by different batches) were reported to correct batch effects [91–93]. Additionally, an autoencoder is a useful technique for extracting features that are associated with clinical outcomes from a high dimensional dataset [12,19,94]. The unsupervised deep embedding algorithm obtained the initial parameters from an autoencoder and learned non-linear mapping from the original scRNA-seq data space to a low-dimensional space; it showed biological interpretability through improving the clustering accuracy by removing batch effects [95]. In addition to batch effects, scRNA-seq has certain limitations such as transcriptional noises, including variability of gene expression and mRNA degradation. To overcome these problems, single-cell variational inference (scVI) based on a hierarchical Bayesian model using stochastic optimization and deep neural networks for probabilistic representation and analysis of gene expression have been reported [96]. Another autoencoder-based algorithm, i.e., single-cell decomposition using a hierarchical autoencoder (scDHA), showed the visualization of the transcriptome landscape as well as cell classification and pseudo-time inference with the combination of a non-negative kernel autoencoder to remove insignificant features; the second module of a stacked Bayesian autoencoder was used to compress data [97]. The application of a sparsely connected autoencoder was used to query cell subpopulations to investigate functional features rather than to classify cell subpopulations [98]; the single-cell graph neural network (scGNN) comprised multiple autoencoders, such as feature, graph, cluster, and imputation autoencoders; they successfully represented Alzheimer's disease-related neural development and mechanisms [99].

From the aspect of multi-omics analysis, scRNA-seq can be combined with other NGS methods such as DNA methylation or FISH [100,101]. Seurat was initially developed to profile scRNA-seq datasets of different cell conditions (i.e., either resting or stimulated), and beyond species (either humans or mice) [102]. Later, integrated analysis with other omics data—RNA-seq, single-cell protein, and spatial analysis—was reported for a comprehensive understanding of cellular events using the advanced R tool Seurat version 3 [103]. The linked inference of genomic experimental relationships (LIGER) algorithm leveraging integrated non-negative matrix factorization (NMF) described common and uncommon features of cell identity. The use of scRNA-seq and DNA methylation data revealed the mechanisms of cell-type-specific epigenomic regulation [104]. Additionally, single-cell molecular epigenomics with ML has the potential to reveal a cutting-edge field, i.e., typi-

cally highly heterogenous cancer tissues and cancer-associated fibroblasts, which are major constituents and significantly heterogeneous to decode a tumor microenvironment [66,105].

1. scRNA-seq methodology

STRT-seq
Smart-seq
CEL-seq
.
.
.

Researchers need to select the methodology; each method has pros and cons.

2. NGS analysis

Mapping, expression quantification, and differential expression analysis are performed in this step.

3. ML-based analysis

In this step, ML-based analysis is performed. For example, k-means clustering is used for clustering. Logistic regression, support vector machine, and k-nearest neighbors are used for classification. For dimensionality reduction and good visualization, PCA, t-SNE, and UMAP are used.

4. Single-cell landscape

Drug response
Cell of origin
Cell state
.
.
.

The results obtained provide novel biological insights into drug response, cell of origin, cell state, and so on.

Figure 2. Schematic diagram of pipeline of scRNA-seq analysis. NGS data (input data or big data) is converted to actionable data using ML. Although each scRNA-seq technology has its own optimized protocol, in general, tissues or cells are first dissociated to disaggregated into single cells. At this point, high cell viability (>70%) and less than 30 μm of cell size in diameter are required from fresh samples to extract RNA. Then, using isolated RNA, cDNA is synthesized for NGS. NGS data (input data or big data) is converted to actionable data using ML such as cell type identification (classification) or subtypes identification (clustering). The obtained results (output) should be carefully reviewed if they support the research hypothesis. Besides, it is also possible to come up with new ideas with obtained scRNA-seq results.

On top of that, scRNA and chromatin accessibility (single-cell assay for transposase-accessible chromatin sequencing; scATAC-seq), have been reported to identify transcriptional and functional differences between hematopoietic stem cells from liver and bone marrow [106]. Thus, this topic is discussed in the next section.

4. Single-Cell ATAC-seq and ChIP-seq Analysis Using Machine Learning

During the last decade, analyses of chromatin accessibility have become essential methods to investigate the epigenetic profiles of various cell types. DNase-seq [107] and FAIRE-seq [108] are pioneering technologies for analyzing chromatin accessibility; however, the requirement of a large amount of starting material limits its wide clinical application. Subsequently, ATAC-seq [109] was introduced in 2013 and has rapidly expanded its application, taking advantage of simple protocol involving Tn5 insertion followed by PCR and less requirements of sample volume. The notable advantage is its success in single-cell analysis, which was reported by two independent research groups as "scATAC-seq" and "sciATAC-seq" in 2015 [79,80]. Furthermore, the recently developed droplet-based single-cell combinatorial indexing for ATAC-seq (dsciATAC-seq) enabled large-scale and high-throughput profiling of single-cell epigenomes [110].

Simultaneously, various analytical tools have been developed to investigate single-cell epigenomes using single-cell ATAC-seq data. Because of the high dimensionality and sparsity of the data, the computational analysis of single-cell ATAC-seq is more challenging than that of single-cell RNA-seq. To overcome this difficulty, several unsupervised ML algorithms have been proposed. First, the chromVAR utilized transcription factor (TF) motif occurrence in the open chromatin regions to project the vector of bias-corrected deviations

from individual cells onto two dimensions using t-SNE. The advantage of chromVAR is that it can be used to compute the TF binding profiles associated with significant chromatin accessibility. Furthermore, chromVAR can be applied to a collection of k-mers, enabling de novo discovery of previously unannotated motifs [111]. A related approach of chromVAR is seen in BROCKMAN, where k-mer factorization is applied [112]. SCRAT uses a number of predefined features, such as TF motif, ENCODE cluster, and MSigDB gene sets [113]. By contrast, scABC relies solely on the patterns of read counts within genomic regions to cluster cells by using unsupervised k-medoids clustering [114].

Second, natural language processing techniques have been applied to cluster cells based on similarities in chromatin accessibility. A large-scale study for constructing a single-cell atlas of in vivo mouse organs by Cusanovich et al. used latent semantic analysis (LSA) to identify cell clusters [115]. Another example is cisTopic, a probabilistic framework to discover co-accessible enhancers and stable cell states, which utilizes latent Dirichlet allocation (LDA) with a collapsed Gibbs sampler to infer "cis-regulatory topics" [116].

Third, Cicero presents an ML framework to predict cis-regulatory DNA interactions using graphical lasso [117]. Cicero uses sampling and aggregation of groups of similar cells to quantify correlations between putative regulatory elements and links these regulatory elements to target genes using unsupervised ML. Interestingly, these predicted interactions are compatible with other chromatin 3D structure data such as ChIA-PET and Hi-C [117]. Cicero is potentially focused on predicting gene expression and chromatin 3D structure, as well as chromatin accessibility from single-cell ATAC-seq data [117,118].

Finally, comprehensive pipelines of single-cell ATAC-seq such as Scasat [119] and SnapATAC [120] were recently implemented; they could make single-cell ATAC-seq analysis easier to access for every researcher regardless of their specialty.

Next, we will describe some research examples of single-cell ATAC-seq using an unsupervised ML algorithm. Cusanovich et al. constructed a single-cell atlas of adult mouse tissues and characterized 85 distinct chromatin patterns across 13 different tissues [115]. After combining various ML algorithms (e.g., LSA, Cicero, and Basset) [121], they annotated key regulatory sequences and TFs in diverse mammalian cell types. Moreover, it is anticipated that this dataset can be utilized to investigate common human traits and diseases through comparing human genome with mouse genome (Figure 3).

Ranzoni et al. performed an integrative analysis of single-cell RNA-seq and ATAC-seq of human developmental hematopoiesis, revealing that there are multiple subpopulations that differ in their overall chromatin accessibility as well as lineage-specific TF activity within transcriptionally homogeneous cell clusters. They also supported the view that regulatory programming for future differentiation is primed at the chromatin level within the same cell cluster defined by scRNA-seq [106].

In conclusion, through combining single-cell ATAC-seq with other sequencing techniques and ML algorithms, it is possible to extensively define cell types and their key regulatory elements of every species, including humans. These findings provide useful resources in clinical research to investigate human disease traits and refine precision medicine.
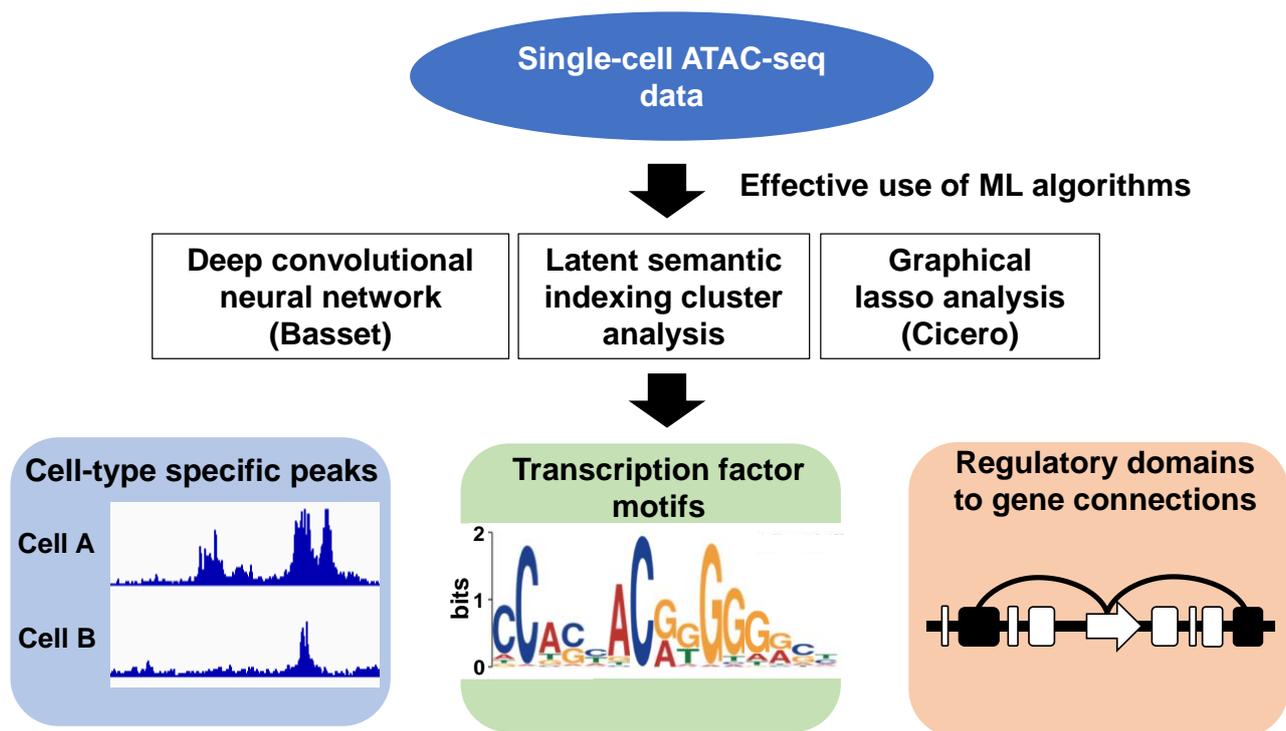
Figure 3. Schematic diagram of single-cell ATAC-seq analysis effectively leveraging ML. Basset is a deep convolutional neural network method to learn the functional activity of DNA sequences from genomics data. In particular, Basset allows us to simultaneously learn that cell's chromatin accessibility code and annotate every mutation in the genome with its influence on present accessibility and latent potential for accessibility. Latent semantic indexing cluster analysis uses dimensional compression to normalize and cluster the data. Cicero uses a graphical lasso model to assign a regularized joint accessibility score, penalized by genomic distance, to each pair of "open" sites in each cluster. Hence, it is possible to find sites that were co-accessible in aggregated groups of cells within each cluster. By effectively using ML techniques in this way, we are able to predict cell clusters, transcription factors binding profiles, and cis-regulatory DNA elements (Reference [115]).

5. Discussion

Recent advances in single-cell epigenomic technologies have enabled the study of genomic regulation and dynamics with unprecedented resolution. In this review, we introduced some of these innovative technologies and how they can identify the trajectory and state while taking advantage of various ML techniques. In particular, scRNA-seq is a rapidly developing technology, with several protocols published in the last few years. As previously mentioned, the most widely used methods are microwell- and droplet-based. These technologies differ in the way they tag cell-derived transcripts and how they generate libraries for sequencing. The choice of which method to use depends on the scientific question, the number of obtained cells, the depth of required information, as well as the cost. For example, droplet-based methods are best suited for characterizing tissue composition because they can handle large numbers of cells. Conversely, if only cells with known surface markers are to be analyzed, it is best to sequence a small number of cells after enrichment using flow cytometry.

Single-cell analysis allows for pseudotime analysis, which makes it possible to trace the developmental trajectory of a single cell. This indicates that single-cell epigenomic technology plays an important role in understanding cellular diversity and discovering gene regulatory mechanisms. Specifically, gene activation and cell identity are regulated by TFs, chromatin regulators, non-coding RNAs, and factors that control the chromatin 3D structure. Furthermore, the importance of cell-to-cell variability in tumor tissues is now being recognized, and single-cell epigenomic methods will be even more powerful in the

future to elucidate the mechanisms of gene regulation in diverse cellular environments regarding development and disease [122].

With the growing number of single-cell epigenomic methods, a key challenge in this field is the computational combination of various single-cell "omics" methods to enable integrated cellular regulatory models. Further advanced ML is required to decode high-dimensional single-cell data [123]. In this context, previously unknown intermediate cellular states may become apparent. Indeed, several reports have evaluated intermediate states that cannot be adequately identified using known cellular markers [124,125]. Although completely reconstructing the identity of a cell may be difficult with the existing pipeline, careful construction using high-precision ML is feasible.

Furthermore, tumor immunity has been attracting attention in the field of oncology due to the recent successful clinical application of immune checkpoint inhibitors. In this context, single-cell analysis has been actively introduced to analyze the heterogeneity of immune system cells, the diversity of T and B cell repertoires, and antigen specificity. For example, to assess the breadth of T-cell activation, Want et al. profiled the TCR repertoire by single-cell sequencing when stimulated with neoantigens and identified sequence motifs that define an oligoclonal and autologous T cell response [126]. It has also been reported that accumulation of cytotoxic T cells and NK cells and increase in effector function are observed when WHSC1 is pharmacologically inhibited by single-cell analysis of immunokinetics in tumors in an experimental system using mice grafted with prostate tumor cells [127]. Importantly, the ability of ML algorithms to learn complex patterns in data is expected to be utilized in the analysis of the complex immune system. In fact, a suit of unsupervised and supervised deep learning methods, called DeepTCR, have also been proposed that can model highly complex TCR sequencing data by learning a joint representation of a TCR by CDR3 sequences and V/D/J gene usage [128]. The use of DeepTCR improved the featurization of TCRs and demonstrated its usefulness in classifying antigen-specific TCRs and extracting antigen-specific TCRs from noisy scRNA-seq and T-cell culture-based assays [128]. These results demonstrate the flexibility and capability of ML technology in analyzing the complex immune system, and further research is expected in the future.

6. Conclusions

As highlighted in this review, ML techniques are currently being applied to single-cell analysis in various ways, and important knowledge has been accumulated. One of the weaknesses of single-cell analysis has been the "problem of spatial information"; however, this problem is now being solved with the release of Visium, a kit for spatial transcriptome analysis from $10\times$ Genomics, after a study using mouse brain and human breast cancer that was published by Stahl et al. in 2016 [129–131]. We believe that spatial transcriptome analysis is an important area in single-cell analysis where ML technology should be actively introduced in the future. In addition, ML technology plays a central role in the current development of AI. Its importance in medical AI is widely recognized, and many AI-equipped medical devices are widely used in actual clinical settings [24,132]. In fact, single-cell analysis is also expected to contribute to the realization of precision medicine by revealing gene expression profiles that lead to the prediction of patient prognosis and stratification of molecular pathologies; clinical applications of single-cell analysis are expected to progress steadily in the future through the effective use of ML technology.

Author Contributions: Conceptualization: K.A., K.T. and R.H.; investigation: K.A., K.T., H.M. and R.H.; writing—original draft preparation: K.A., K.T., H.M., S.K. and R.H.; writing—review and editing: K.A., K.T., H.M., S.T., N.S., A.B., K.K., M.K., S.K., K.O. and R.H. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

References

1. Quail, D.F.; Joyce, J.A. Microenvironmental regulation of tumor progression and metastasis. *Nat. Med.* **2013**, *19*, 1423–1437. [CrossRef] [PubMed]
2. Prasetyanti, P.R.; Medema, J.P. Intra-tumor heterogeneity from a cancer stem cell perspective. *Mol. Cancer* **2017**, *16*, 41. [CrossRef] [PubMed]
3. Kreso, A.; Dick, J.E. Evolution of the cancer stem cell model. *Cell Stem Cell* **2014**, *14*, 275–291. [CrossRef] [PubMed]
4. Eberwine, J.; Yeh, H.; Miyashiro, K.; Cao, Y.; Nair, S.; Finnell, R.; Zettel, M.; Coleman, P. Analysis of gene expression in single live neurons. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 3010–3014. [CrossRef]
5. Sucher, N.J.; Deitcher, D.L. PCR and patch-clamp analysis of single neurons. *Neuron* **1995**, *14*, 1095–1100. [CrossRef]
6. Femino, A.M.; Fay, F.S.; Fogarty, K.; Singer, R.H. Visualization of single RNA transcripts in situ. *Science* **1998**, *280*, 585–590. [CrossRef]
7. Kamme, F.; Salunga, R.; Yu, J.; Tran, D.T.; Zhu, J.; Luo, L.; Bittner, A.; Guo, H.Q.; Miller, N.; Wan, J.; et al. Single-cell microarray analysis in hippocampus CA1: Demonstration and validation of cellular heterogeneity. *J. Neurosci.* **2003**, *23*, 3607–3615. [CrossRef]
8. Kurimoto, K.; Yabuta, Y.; Ohinata, Y.; Ono, Y.; Uno, K.D.; Yamada, R.G.; Ueda, H.R.; Saitou, M. An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis. *Nucleic Acids Res.* **2006**, *34*, e42. [CrossRef]
9. Tang, F.; Barbacioru, C.; Wang, Y.; Nordman, E.; Lee, C.; Xu, N.; Wang, X.; Bodeau, J.; Tuch, B.B.; Siddiqui, A.; et al. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* **2009**, *6*, 377–382. [CrossRef]
10. Hamamoto, R.; Komatsu, M.; Takasawa, K.; Asada, K.; Kaneko, S. Epigenetics Analysis and Integrated Analysis of Multiomics Data, Including Epigenetic Data, Using Artificial Intelligence in the Era of Precision Medicine. *Biomolecules* **2020**, *10*, 62. [CrossRef]
11. Hamamoto, R. Application of Artificial Intelligence for Medical Research. *Biomolecules* **2021**, *11*, 90. [CrossRef]
12. Asada, K.; Kobayashi, K.; Joutard, S.; Tubaki, M.; Takahashi, S.; Takasawa, K.; Komatsu, M.; Kaneko, S.; Sese, J.; Hamamoto, R. Uncovering Prognosis-Related Genes and Pathways by Multi-Omics Analysis in Lung Cancer. *Biomolecules* **2020**, *10*, 524. [CrossRef]
13. Dozen, A.; Komatsu, M.; Sakai, A.; Komatsu, R.; Shozu, K.; Machino, H.; Yasutomi, S.; Arakaki, T.; Asada, K.; Kaneko, S.; et al. Image Segmentation of the Ventricular Septum in Fetal Cardiac Ultrasound Videos Based on Deep Learning Using Time-Series Information. *Biomolecules* **2020**, *10*, 1526. [CrossRef]
14. Kobayashi, K.; Bolatkan, A.; Shiina, S.; Hamamoto, R. Fully-Connected Neural Networks with Reduced Parameterization for Predicting Histological Types of Lung Cancer from Somatic Mutations. *Biomolecules* **2020**, *10*, 1249. [CrossRef]
15. Komatsu, M.; Sakai, A.; Komatsu, R.; Matsuoka, R.; Yasutomi, S.; Shozu, K.; Dozen, A.; Machino, H.; Hidaka, H.; Arakaki, T.; et al. Detection of Cardiac Structural Abnormalities in Fetal Ultrasound Videos Using Deep Learning. *Appl. Sci.* **2021**, *11*, 371. [CrossRef]
16. Jinnai, S.; Yamazaki, N.; Hirano, Y.; Sugawara, Y.; Ohe, Y.; Hamamoto, R. The Development of a Skin Cancer Classification System for Pigmented Skin Lesions Using Deep Learning. *Biomolecules* **2020**, *10*, 1123. [CrossRef] [PubMed]
17. Shozu, K.; Komatsu, M.; Sakai, A.; Komatsu, R.; Dozen, A.; Machino, H.; Yasutomi, S.; Arakaki, T.; Asada, K.; Kaneko, S.; et al. Model-Agnostic Method for Thoracic Wall Segmentation in Fetal Ultrasound Videos. *Biomolecules* **2020**, *10*, 1691. [CrossRef] [PubMed]
18. Yasutomi, S.; Arakaki, T.; Matsuoka, R.; Sakai, A.; Komatsu, R.; Shozu, K.; Dozen, A.; Machino, H.; Asada, K.; Kaneko, S.; et al. Shadow Estimation for Ultrasound Images Using Auto-Encoding Structures and Synthetic Shadows. *Appl. Sci.* **2021**, *11*, 1127. [CrossRef]
19. Takahashi, S.; Asada, K.; Takasawa, K.; Shimoyama, R.; Sakai, A.; Bolatkan, A.; Shinkai, N.; Kobayashi, K.; Komatsu, M.; Kaneko, S.; et al. Predicting Deep Learning Based Multi-Omics Parallel Integration Survival Subtypes in Lung Cancer Using Reverse Phase Protein Array Data. *Biomolecules* **2020**, *10*, 1460. [CrossRef]
20. Kobayashi, K.; Miyake, M.; Takahashi, M.; Hamamoto, R. Observing deep radiomics for the classification of glioma grades. *Sci. Rep.* **2021**, *11*, 10942. [CrossRef]
21. Takahashi, S.; Takahashi, M.; Kinoshita, M.; Miyake, M.; Kawaguchi, R.; Shinojima, N.; Mukasa, A.; Saito, K.; Nagane, M.; Otani, R.; et al. Fine-Tuning Approach for Segmentation of Gliomas in Brain Magnetic Resonance Images with a Machine Learning Method to Normalize Image Differences among Facilities. *Cancers* **2021**, *13*, 1415. [CrossRef]
22. Kawaguchi, R.K.; Takahashi, M.; Miyake, M.; Kinoshita, M.; Takahashi, S.; Ichimura, K.; Hamamoto, R.; Narita, Y.; Sese, J. Assessing Versatile Machine Learning Models for Glioma Radiogenomic Studies across Hospitals. *Cancers* **2021**, *13*, 3611. [CrossRef]
23. Kobayashi, K.; Hataya, R.; Kurose, Y.; Miyake, M.; Takahashi, M.; Nakagawa, A.; Harada, T.; Hamamoto, R. Decomposing Normal and Abnormal Features of Medical Images for Content-Based Image Retrieval of Glioma Imaging. *Med. Image Anal.* **2021**, *74*, 102227. [CrossRef]

24. Hamamoto, R.; Suvarna, K.; Yamada, M.; Kobayashi, K.; Shinkai, N.; Miyake, M.; Takahashi, M.; Jinnai, S.; Shimoyama, R.; Sakai, A.; et al. Application of Artificial Intelligence Technology in Oncology: Towards the Establishment of Precision Medicine. *Cancers* **2020**, *12*, 3532. [CrossRef]

25. Komatsu, M.; Sakai, A.; Dozen, A.; Shozu, K.; Yasutomi, S.; Machino, H.; Asada, K.; Kaneko, S.; Hamamoto, R. Towards Clinical Application of Artificial Intelligence in Ultrasound Imaging. *Biomedicines* **2021**, *9*, 720. [CrossRef]

26. Yamada, M.; Saito, Y.; Imaoka, H.; Saiko, M.; Yamada, S.; Kondo, H.; Takamaru, H.; Sakamoto, T.; Sese, J.; Kuchiba, A.; et al. Development of a real-time endoscopic image diagnosis support system using deep learning technology in colonoscopy. *Sci. Rep.* **2019**, *9*, 14465. [CrossRef] [PubMed]

27. Schena, M.; Shalon, D.; Heller, R.; Chai, A.; Brown, P.O.; Davis, R.W. Parallel human genome analysis: Microarray-based expression monitoring of 1000 genes. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 10614–10619. [CrossRef]

28. Heller, R.A.; Schena, M.; Chai, A.; Shalon, D.; Bedilion, T.; Gilmore, J.; Woolley, D.E.; Davis, R.W. Discovery and analysis of inflammatory disease-related genes using cDNA microarrays. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 2150–2155. [CrossRef] [PubMed]

29. Welford, S.M.; Gregg, J.; Chen, E.; Garrison, D.; Sorensen, P.H.; Denny, C.T.; Nelson, S.F. Detection of differentially expressed genes in primary tumor tissues using representational differences analysis coupled to microarray hybridization. *Nucleic Acids Res.* **1998**, *26*, 3059–3065. [CrossRef] [PubMed]

30. Luo, L.; Salunga, R.C.; Guo, H.; Bittner, A.; Joy, K.C.; Galindo, J.E.; Xiao, H.; Rogers, K.E.; Wan, J.S.; Jackson, M.R.; et al. Gene expression profiles of laser-captured adjacent neuronal subtypes. *Nat. Med.* **1999**, *5*, 117–122. [CrossRef] [PubMed]

31. Fassunke, J.; Majores, M.; Ullmann, C.; Elger, C.E.; Schramm, J.; Wiestler, O.D.; Becker, A.J. In situ-RT and immunolaser microdissection for mRNA analysis of individual cells isolated from epilepsy-associated glioneuronal tumors. *Lab. Investig.* **2004**, *84*, 1520–1525. [CrossRef]

32. Hashida, H.; Goto, J.; Suzuki, T.; Jeong, S.; Masuda, N.; Ooie, T.; Tachiiri, Y.; Tsuchiya, H.; Kanazawa, I. Single cell analysis of CAG repeat in brains of dentatorubral-pallidoluysian atrophy (DRPLA). *J. Neurol. Sci.* **2001**, *190*, 87–93. [CrossRef]

33. Akkari, Y.; Smith, T.; Westfall, J.; Lupo, S. Implementation of cancer next-generation sequencing testing in a community hospital. *Cold Spring Harb. Mol. Case Stud.* **2019**, *5*, a003707. [CrossRef]

34. Asada, K.; Kaneko, S.; Takasawa, K.; Machino, H.; Takahashi, S.; Shinkai, N.; Shimoyama, R.; Komatsu, M.; Hamamoto, R. Integrated Analysis of Whole Genome and Epigenome Data Using Machine Learning Technology: Toward the Establishment of Precision Oncology. *Front. Oncol.* **2021**, *11*, 666937. [CrossRef]

35. Nagalakshmi, U.; Wang, Z.; Waern, K.; Shou, C.; Raha, D.; Gerstein, M.; Snyder, M. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **2008**, *320*, 1344–1349. [CrossRef]

36. Wilhelm, B.T.; Marguerat, S.; Watt, S.; Schubert, F.; Wood, V.; Goodhead, I.; Penkett, C.J.; Rogers, J.; Bahler, J. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* **2008**, *453*, 1239–1243. [CrossRef] [PubMed]

37. Islam, S.; Kjallquist, U.; Moliner, A.; Zajac, P.; Fan, J.B.; Lonnerberg, P.; Linnarsson, S. Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res.* **2011**, *21*, 1160–1167. [CrossRef]

38. Kivioja, T.; Vaharautio, A.; Karlsson, K.; Bonke, M.; Enge, M.; Linnarsson, S.; Taipale, J. Counting absolute numbers of molecules using unique molecular identifiers. *Nat. Methods* **2011**, *9*, 72–74. [CrossRef]

39. Wu, A.R.; Neff, N.F.; Kalisky, T.; Dalerba, P.; Treutlein, B.; Rothenberg, M.E.; Mburu, F.M.; Mantalas, G.L.; Sim, S.; Clarke, M.F.; et al. Quantitative assessment of single-cell RNA-sequencing methods. *Nat. Methods* **2014**, *11*, 41–46. [CrossRef]

40. Bose, S.; Wan, Z.; Carr, A.; Rizvi, A.H.; Vieira, G.; Pe'er, D.; Sims, P.A. Scalable microfluidics for single-cell RNA printing and sequencing. *Genome Biol.* **2015**, *16*, 120. [CrossRef] [PubMed]

41. Klein, A.M.; Mazutis, L.; Akartuna, I.; Tallapragada, N.; Veres, A.; Li, V.; Peshkin, L.; Weitz, D.A.; Kirschner, M.W. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **2015**, *161*, 1187–1201. [CrossRef] [PubMed]

42. Macosko, E.Z.; Basu, A.; Satija, R.; Nemesh, J.; Shekhar, K.; Goldman, M.; Tirosh, I.; Bialas, A.R.; Kamitaki, N.; Martersteck, E.M.; et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **2015**, *161*, 1202–1214. [CrossRef] [PubMed]

43. Cao, J.; Packer, J.S.; Ramani, V.; Cusanovich, D.A.; Huynh, C.; Daza, R.; Qiu, X.; Lee, C.; Furlan, S.N.; Steemers, F.J.; et al. Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* **2017**, *357*, 661–667. [CrossRef]

44. Zheng, G.X.; Terry, J.M.; Belgrader, P.; Ryvkin, P.; Bent, Z.W.; Wilson, R.; Ziraldo, S.B.; Wheeler, T.D.; McDermott, G.P.; Zhu, J.; et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **2017**, *8*, 14049. [CrossRef] [PubMed]

45. Ramskold, D.; Luo, S.; Wang, Y.C.; Li, R.; Deng, Q.; Faridani, O.R.; Daniels, G.A.; Khrebtukova, I.; Loring, J.F.; Laurent, L.C.; et al. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.* **2012**, *30*, 777–782. [CrossRef]

46. Picelli, S.; Bjorklund, A.K.; Faridani, O.R.; Sagasser, S.; Winberg, G.; Sandberg, R. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **2013**, *10*, 1096–1098. [CrossRef]

47. Hashimshony, T.; Wagner, F.; Sher, N.; Yanai, I. CEL-Seq: Single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep.* **2012**, *2*, 666–673. [CrossRef]

48. Hashimshony, T.; Senderovich, N.; Avital, G.; Klochendler, A.; de Leeuw, Y.; Anavy, L.; Gennert, D.; Li, S.; Livak, K.J.; Rozenblatt-Rosen, O.; et al. CEL-Seq2: Sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol.* **2016**, *17*, 77. [CrossRef]

49. Sasagawa, Y.; Nikaido, I.; Hayashi, T.; Danno, H.; Uno, K.D.; Imai, T.; Ueda, H.R. Quartz-Seq: A highly reproducible and sensitive single-cell RNA sequencing method, reveals non-genetic gene-expression heterogeneity. *Genome Biol.* **2013**, *14*, R31. [CrossRef] [PubMed]

50. Sasagawa, Y.; Danno, H.; Takada, H.; Ebisawa, M.; Tanaka, K.; Hayashi, T.; Kurisaki, A.; Nikaido, I. Quartz-Seq2: A high-throughput single-cell RNA-sequencing method that effectively uses limited sequence reads. *Genome Biol.* **2018**, *19*, 29. [CrossRef] [PubMed]

51. Streets, A.M.; Zhang, X.; Cao, C.; Pang, Y.; Wu, X.; Xiong, L.; Yang, L.; Fu, Y.; Zhao, L.; Tang, F.; et al. Microfluidic single-cell whole-transcriptome sequencing. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 7048–7053. [CrossRef] [PubMed]

52. Gierahn, T.M.; Wadsworth, M.H., II; Hughes, T.K.; Bryson, B.D.; Butler, A.; Satija, R.; Fortune, S.; Love, J.C.; Shalek, A.K. Seq-Well: Portable, low-cost RNA sequencing of single cells at high throughput. *Nat. Methods* **2017**, *14*, 395–398. [CrossRef] [PubMed]

53. Han, X.; Wang, R.; Zhou, Y.; Fei, L.; Sun, H.; Lai, S.; Saadatpour, A.; Zhou, Z.; Chen, H.; Ye, F.; et al. Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell* **2018**, *172*, 1091–1107.e1017. [CrossRef]

54. Hayashi, T.; Ozaki, H.; Sasagawa, Y.; Umeda, M.; Danno, H.; Nikaido, I. Single-cell full-length total RNA sequencing uncovers dynamics of recursive splicing and enhancer RNAs. *Nat. Commun.* **2018**, *9*, 619. [CrossRef]

55. Kouno, T.; Moody, J.; Kwon, A.T.; Shibayama, Y.; Kato, S.; Huang, Y.; Bottcher, M.; Motakis, E.; Mendez, M.; Severin, J.; et al. C1 CAGE detects transcription start sites and enhancer activity at single-cell resolution. *Nat. Commun.* **2019**, *10*, 360. [CrossRef]

56. Zhang, S.; Cui, Y.; Ma, X.; Yong, J.; Yan, L.; Yang, M.; Ren, J.; Tang, F.; Wen, L.; Qiao, J. Single-cell transcriptomics identifies divergent developmental lineage trajectories during human pituitary development. *Nat. Commun.* **2020**, *11*, 5275. [CrossRef] [PubMed]

57. Geirsdottir, L.; David, E.; Keren-Shaul, H.; Weiner, A.; Bohlen, S.C.; Neuber, J.; Balic, A.; Giladi, A.; Sheban, F.; Dutertre, C.A.; et al. Cross-Species Single-Cell Analysis Reveals Divergence of the Primate Microglia Program. *Cell* **2020**, *181*, 746. [CrossRef]

58. Luo, W.; Lin, G.N.; Song, W.; Zhang, Y.; Lai, H.; Zhang, M.; Miao, J.; Cheng, X.; Wang, Y.; Li, W.; et al. Single-cell spatial transcriptomic analysis reveals common and divergent features of developing postnatal granule cerebellar cells and medulloblastoma. *BMC Biol.* **2021**, *19*, 135. [CrossRef]

59. Azizi, E.; Carr, A.J.; Plitas, G.; Cornish, A.E.; Konopacki, C.; Prabhakaran, S.; Nainys, J.; Wu, K.; Kiseliovas, V.; Setty, M.; et al. Single-Cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment. *Cell* **2018**, *174*, 1293–1308.e1236. [CrossRef]

60. Suraci, N.; Mihos, C.; Baquero, S.; Santana, O. Superior Vena Cava Stent Migration Into the Right Atrium. *J. Invasive. Cardiol.* **2020**, *32*, E75.

61. Jew, B.; Alvarez, M.; Rahmani, E.; Miao, Z.; Ko, A.; Garske, K.M.; Sul, J.H.; Pietilainen, K.H.; Pajukanta, P.; Halperin, E. Accurate estimation of cell composition in bulk expression through robust integration of single-cell information. *Nat. Commun.* **2020**, *11*, 1971. [CrossRef]

62. Muhl, L.; Genove, G.; Leptidis, S.; Liu, J.; He, L.; Mocci, G.; Sun, Y.; Gustafsson, S.; Buyandelger, B.; Chivukula, I.V.; et al. Single-cell analysis uncovers fibroblast heterogeneity and criteria for fibroblast and mural cell identification and discrimination. *Nat. Commun.* **2020**, *11*, 3953. [CrossRef] [PubMed]

63. Kumar, M.P.; Du, J.; Lagoudas, G.; Jiao, Y.; Sawyer, A.; Drummond, D.C.; Lauffenburger, D.A.; Raue, A. Analysis of Single-Cell RNA-Seq Identifies Cell-Cell Communication Associated with Tumor Characteristics. *Cell Rep.* **2018**, *25*, 1458–1468.e1454. [CrossRef] [PubMed]

64. Cuomo, A.S.E.; Seaton, D.D.; McCarthy, D.J.; Martinez, I.; Bonder, M.J.; Garcia-Bernardo, J.; Amatya, S.; Madrigal, P.; Isaacson, A.; Buettner, F.; et al. Single-cell RNA-sequencing of differentiating iPS cells reveals dynamic genetic effects on gene expression. *Nat. Commun.* **2020**, *11*, 810. [CrossRef] [PubMed]

65. Olalekan, S.; Xie, B.; Back, R.; Eckart, H.; Basu, A. Characterizing the tumor microenvironment of metastatic ovarian cancer by single-cell transcriptomics. *Cell Rep.* **2021**, *35*, 109165. [CrossRef]

66. Wu, F.; Fan, J.; He, Y.; Xiong, A.; Yu, J.; Li, Y.; Zhang, Y.; Zhao, W.; Zhou, F.; Li, W.; et al. Single-cell profiling of tumor heterogeneity and the microenvironment in advanced non-small cell lung cancer. *Nat. Commun.* **2021**, *12*, 2540. [CrossRef]

67. Cadwell, C.R.; Palasantza, A.; Jiang, X.; Berens, P.; Deng, Q.; Yilmaz, M.; Reimer, J.; Shen, S.; Bethge, M.; Tolias, K.F.; et al. Electrophysiological, transcriptomic and morphologic profiling of single neurons using Patch-seq. *Nat. Biotechnol.* **2016**, *34*, 199–203. [CrossRef]

68. Gokce, O.; Stanley, G.M.; Treutlein, B.; Neff, N.F.; Camp, J.G.; Malenka, R.C.; Rothwell, P.E.; Fuccillo, M.V.; Sudhof, T.C.; Quake, S.R. Cellular Taxonomy of the Mouse Striatum as Revealed by Single-Cell RNA-Seq. *Cell Rep.* **2016**, *16*, 1126–1137. [CrossRef]

69. Munoz-Manchado, A.B.; Bengtsson Gonzales, C.; Zeisel, A.; Munguba, H.; Bekkouche, B.; Skene, N.G.; Lonnerberg, P.; Ryge, J.; Harris, K.D.; Linnarsson, S.; et al. Diversity of Interneurons in the Dorsal Striatum Revealed by Single-Cell RNA Sequencing and PatchSeq. *Cell Rep.* **2018**, *24*, 2179–2190.e2177. [CrossRef]

70. Ren, X.; Deng, R.; Zhang, K.; Sun, Y.; Teng, X.; Li, J. SpliceRCA: In Situ Single-Cell Analysis of mRNA Splicing Variants. *ACS Cent. Sci.* **2018**, *4*, 680–687. [CrossRef]

71. Navin, N.; Kendall, J.; Troge, J.; Andrews, P.; Rodgers, L.; McIndoo, J.; Cook, K.; Stepansky, A.; Levy, D.; Esposito, D.; et al. Tumour evolution inferred by single-cell sequencing. *Nature* **2011**, *472*, 90–94. [CrossRef] [PubMed]

72. Vitak, S.A.; Torkenczy, K.A.; Rosenkrantz, J.L.; Fields, A.J.; Christiansen, L.; Wong, M.H.; Carbone, L.; Steemers, F.J.; Adey, A. Sequencing thousands of single-cell genomes with combinatorial indexing. *Nat. Methods* **2017**, *14*, 302–308. [CrossRef] [PubMed]

73. Xu, X.; Hou, Y.; Yin, X.; Bao, L.; Tang, A.; Song, L.; Li, F.; Tsang, S.; Wu, K.; Wu, H.; et al. Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. *Cell* **2012**, *148*, 886–895. [CrossRef]

74. Francis, J.M.; Zhang, C.Z.; Maire, C.L.; Jung, J.; Manzo, V.E.; Adalsteinsson, V.A.; Homer, H.; Haidar, S.; Blumenstiel, B.; Pedamallu, C.S.; et al. EGFR variant heterogeneity in glioblastoma resolved through single-nucleus sequencing. *Cancer Discov.* **2014**, *4*, 956–971. [CrossRef]

75. Guo, H.; Zhu, P.; Wu, X.; Li, X.; Wen, L.; Tang, F. Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing. *Genome Res.* **2013**, *23*, 2126–2135. [CrossRef]

76. Smallwood, S.A.; Lee, H.J.; Angermueller, C.; Krueger, F.; Saadeh, H.; Peat, J.; Andrews, S.R.; Stegle, O.; Reik, W.; Kelsey, G. Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat. Methods* **2014**, *11*, 817–820. [CrossRef]

77. Luo, C.; Keown, C.L.; Kurihara, L.; Zhou, J.; He, Y.; Li, J.; Castanon, R.; Lucero, J.; Nery, J.R.; Sandoval, J.P.; et al. Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. *Science* **2017**, *357*, 600–604. [CrossRef]

78. Nagano, T.; Lubling, Y.; Stevens, T.J.; Schoenfelder, S.; Yaffe, E.; Dean, W.; Laue, E.D.; Tanay, A.; Fraser, P. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **2013**, *502*, 59–64. [CrossRef] [PubMed]

79. Cusanovich, D.A.; Daza, R.; Adey, A.; Pliner, H.A.; Christiansen, L.; Gunderson, K.L.; Steemers, F.J.; Trapnell, C.; Shendure, J. Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **2015**, *348*, 910–914. [CrossRef]

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

81. Miao, Z.; Moreno, P.; Huang, N.; Papatheodorou, I.; Brazma, A.; Teichmann, S.A. Putative cell type discovery from single-cell gene expression data. *Nat. Methods* **2020**, *17*, 621–628. [CrossRef]

82. Lake, B.B.; Ai, R.; Kaeser, G.E.; Salathia, N.S.; Yung, Y.C.; Liu, R.; Wildberg, A.; Gao, D.; Fung, H.L.; Chen, S.; et al. Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* **2016**, *352*, 1586–1590. [CrossRef]

83. Habib, N.; Li, Y.; Heidenreich, M.; Swiech, L.; Avraham-Davidi, I.; Trombetta, J.J.; Hession, C.; Zhang, F.; Regev, A. Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons. *Science* **2016**, *353*, 925–928. [CrossRef]

84. Slyper, M.; Porter, C.B.M.; Ashenberg, O.; Waldman, J.; Drokhlyansky, E.; Wakiro, I.; Smillie, C.; Smith-Rosario, G.; Wu, J.; Dionne, D.; et al. A single-cell and single-nucleus RNA-Seq toolbox for fresh and frozen human tumors. *Nat. Med.* **2020**, *26*, 792–802. [CrossRef]

85. Zahn, H.; Steif, A.; Laks, E.; Eirew, P.; VanInsberghe, M.; Shah, S.P.; Aparicio, S.; Hansen, C.L. Scalable whole-genome single-cell library preparation without preamplification. *Nat. Methods* **2017**, *14*, 167–173. [CrossRef] [PubMed]

86. Wei, Y.; Lang, J.; Zhang, Q.; Yang, C.R.; Zhao, Z.A.; Zhang, Y.; Du, Y.; Sun, Y. DNA methylation analysis and editing in single mammalian oocytes. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 9883–9892. [CrossRef] [PubMed]

87. Lee, J.H.; Daugharthy, E.R.; Scheiman, J.; Kalhor, R.; Yang, J.L.; Ferrante, T.C.; Terry, R.; Jeanty, S.S.; Li, C.; Amamoto, R.; et al. Highly multiplexed subcellular RNA sequencing in situ. *Science* **2014**, *343*, 1360–1363. [CrossRef] [PubMed]

88. Chen, K.H.; Boettiger, A.N.; Moffitt, J.R.; Wang, S.; Zhuang, X. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **2015**, *348*, aaa6090. [CrossRef] [PubMed]

89. Shah, S.; Lubeck, E.; Zhou, W.; Cai, L. seqFISH Accurately Detects Transcripts in Single Cells and Reveals Robust Spatial Organization in the Hippocampus. *Neuron* **2017**, *94*, 752–758.e751. [CrossRef]

90. 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. Single-cell RNA sequencing demonstrates the molecular and cellular reprogramming of metastatic lung adenocarcinoma. *Nat. Commun.* **2020**, *11*, 2285. [CrossRef]

91. Haghverdi, L.; Lun, A.T.L.; Morgan, M.D.; Marioni, J.C. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.* **2018**, *36*, 421–427. [CrossRef]

92. Wang, T.; Johnson, T.S.; Shao, W.; Lu, Z.; Helm, B.R.; Zhang, J.; Huang, K. BERMUDA: A novel deep transfer learning method for single-cell RNA sequencing batch correction reveals hidden high-resolution cellular subtypes. *Genome Biol.* **2019**, *20*, 165. [CrossRef] [PubMed]

93. Hie, B.; Bryson, B.; Berger, B. Efficient integration of heterogeneous single-cell transcriptomes using Scanorama. *Nat. Biotechnol.* **2019**, *37*, 685–691. [CrossRef] [PubMed]

94. Hinton, G.E.; Salakhutdinov, R.R. Reducing the dimensionality of data with neural networks. *Science* **2006**, *313*, 504–507. [CrossRef] [PubMed]

95. Li, X.; Wang, K.; Lyu, Y.; Pan, H.; Zhang, J.; Stambolian, D.; Susztak, K.; Reilly, M.P.; Hu, G.; Li, M. Deep learning enables accurate clustering with batch effect removal in single-cell RNA-seq analysis. *Nat. Commun.* **2020**, *11*, 2338. [CrossRef] [PubMed]

96. Lopez, R.; Regier, J.; Cole, M.B.; Jordan, M.I.; Yosef, N. Deep generative modeling for single-cell transcriptomics. *Nat. Methods* **2018**, *15*, 1053–1058. [CrossRef]

97. Tran, D.; Nguyen, H.; Tran, B.; La Vecchia, C.; Luu, H.N.; Nguyen, T. Fast and precise single-cell data analysis using a hierarchical autoencoder. *Nat. Commun.* **2021**, *12*, 1029. [CrossRef]

98. Alessandri, L.; Cordero, F.; Beccuti, M.; Licheri, N.; Arigoni, M.; Olivero, M.; Di Renzo, M.F.; Sapino, A.; Calogero, R. Sparsely-connected autoencoder (SCA) for single cell RNAseq data mining. *NPJ Syst. Biol. Appl.* **2021**, *7*, 1. [CrossRef]

99. Wang, J.; Ma, A.; Chang, Y.; Gong, J.; Jiang, Y.; Qi, R.; Wang, C.; Fu, H.; Ma, Q.; Xu, D. scGNN is a novel graph neural network framework for single-cell RNA-Seq analyses. *Nat. Commun.* **2021**, *12*, 1882. [CrossRef]

100. Bian, S.; Hou, Y.; Zhou, X.; Li, X.; Yong, J.; Wang, Y.; Wang, W.; Yan, J.; Hu, B.; Guo, H.; et al. Single-cell multiomics sequencing and analyses of human colorectal cancer. *Science* **2018**, *362*, 1060–1063. [CrossRef]

101. Dos Santos, M.; Backer, S.; Saintpierre, B.; Izac, B.; Andrieu, M.; Letourneur, F.; Relaix, F.; Sotiropoulos, A.; Maire, P. Single-nucleus RNA-seq and FISH identify coordinated transcriptional activity in mammalian myofibers. *Nat. Commun.* **2020**, *11*, 5102. [CrossRef]

102. Butler, A.; Hoffman, P.; Smibert, P.; Papalexi, E.; Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **2018**, *36*, 411–420. [CrossRef]

103. Stuart, T.; Butler, A.; Hoffman, P.; Hafemeister, C.; Papalexi, E.; Mauck, W.M., III; Hao, Y.; Stoeckius, M.; Smibert, P.; Satija, R. Comprehensive Integration of Single-Cell Data. *Cell* **2019**, *177*, 1888–1902.e1821. [CrossRef]

104. Welch, J.D.; Kozareva, V.; Ferreira, A.; Vanderburg, C.; Martin, C.; Macosko, E.Z. Single-Cell Multi-omic Integration Compares and Contrasts Features of Brain Cell Identity. *Cell* **2019**, *177*, 1873–1887.e1817. [CrossRef] [PubMed]

105. Bartoschek, M.; Oskolkov, N.; Bocci, M.; Lovrot, J.; Larsson, C.; Sommarin, M.; Madsen, C.D.; Lindgren, D.; Pekar, G.; Karlsson, G.; et al. Spatially and functionally distinct subclasses of breast cancer-associated fibroblasts revealed by single cell RNA sequencing. *Nat. Commun.* **2018**, *9*, 5150. [CrossRef] [PubMed]

106. Ranzoni, A.M.; Tangherloni, A.; Berest, I.; Riva, S.G.; Myers, B.; Strzelecka, P.M.; Xu, J.; Panada, E.; Mohorianu, I.; Zaugg, J.B.; et al. Integrative Single-Cell RNA-Seq and ATAC-Seq Analysis of Human Developmental Hematopoiesis. *Cell Stem Cell* **2021**, *28*, 472–487.e477. [CrossRef] [PubMed]

107. Boyle, A.P.; Davis, S.; Shulha, H.P.; Meltzer, P.; Margulies, E.H.; Weng, Z.; Furey, T.S.; Crawford, G.E. High-resolution mapping and characterization of open chromatin across the genome. *Cell* **2008**, *132*, 311–322. [CrossRef] [PubMed]

108. Giresi, P.G.; Kim, J.; McDaniell, R.M.; Iyer, V.R.; Lieb, J.D. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res.* **2007**, *17*, 877–885. [CrossRef]

109. Buenrostro, J.D.; Giresi, P.G.; Zaba, L.C.; Chang, H.Y.; Greenleaf, W.J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **2013**, *10*, 1213–1218. [CrossRef] [PubMed]

110. 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.; et al. Droplet-based combinatorial indexing for massive-scale single-cell chromatin accessibility. *Nat. Biotechnol.* **2019**, *37*, 916–924. [CrossRef]

111. Schep, A.N.; Wu, B.; Buenrostro, J.D.; Greenleaf, W.J. chromVAR: Inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **2017**, *14*, 975–978. [CrossRef]

112. de Boer, C.G.; Regev, A. BROCKMAN: Deciphering variance in epigenomic regulators by k-mer factorization. *BMC Bioinform.* **2018**, *19*, 253. [CrossRef]

113. Ji, Z.; Zhou, W.; Ji, H. Single-cell regulome data analysis by SCRAT. *Bioinformatics* **2017**, *33*, 2930–2932. [CrossRef]

114. Zamanighomi, M.; Lin, Z.; Daley, T.; Chen, X.; Duren, Z.; Schep, A.; Greenleaf, W.J.; Wong, W.H. Unsupervised clustering and epigenetic classification of single cells. *Nat. Commun.* **2018**, *9*, 2410. [CrossRef]

115. Cusanovich, D.A.; Hill, A.J.; Aghamirzaie, D.; Daza, R.M.; Pliner, H.A.; Berletch, J.B.; Filippova, G.N.; Huang, X.; Christiansen, L.; DeWitt, W.S.; et al. A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. *Cell* **2018**, *174*, 1309–1324.e1318. [CrossRef]

116. Bravo Gonzalez-Blas, C.; Minnoye, L.; Papasokrati, D.; Aibar, S.; Hulselmans, G.; Christiaens, V.; Davie, K.; Wouters, J.; Aerts, S. cisTopic: Cis-regulatory topic modeling on single-cell ATAC-seq data. *Nat. Methods* **2019**, *16*, 397–400. [CrossRef] [PubMed]

117. 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. Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility Data. *Mol. Cell* **2018**, *71*, 858–871.e858. [CrossRef] [PubMed]

118. Sinnamon, J.R.; Torkenczy, K.A.; Linhoff, M.W.; Vitak, S.A.; Mulqueen, R.M.; Pliner, H.A.; Trapnell, C.; Steemers, F.J.; Mandel, G.; Adey, A.C. The accessible chromatin landscape of the murine hippocampus at single-cell resolution. *Genome Res.* **2019**, *29*, 857–869. [CrossRef] [PubMed]

119. Baker, S.M.; Rogerson, C.; Hayes, A.; Sharrocks, A.D.; Rattray, M. Classifying cells with Scasat, a single-cell ATAC-seq analysis tool. *Nucleic Acids Res.* **2019**, *47*, e10. [CrossRef] [PubMed]

120. Fang, R.; Preissl, S.; Li, Y.; Hou, X.; Lucero, J.; Wang, X.; Motamedi, A.; Shiau, A.K.; Zhou, X.; Xie, F.; et al. Comprehensive analysis of single cell ATAC-seq data with SnapATAC. *Nat. Commun.* **2021**, *12*, 1337. [CrossRef]

121. Kelley, D.R.; Snoek, J.; Rinn, J.L. Basset: Learning the regulatory code of the accessible genome with deep convolutional neural networks. *Genome Res.* **2016**, *26*, 990–999. [CrossRef]

122. Wang, W.; Wang, L.; She, J.; Zhu, J. Examining heterogeneity of stromal cells in tumor microenvironment based on pan-cancer single-cell RNA sequencing data. *Cancer Biol. Med.* **2021**, 1–13. [CrossRef]

123. Wagner, A.; Regev, A.; Yosef, N. Revealing the vectors of cellular identity with single-cell genomics. *Nat. Biotechnol.* **2016**, *34*, 1145–1160. [CrossRef]

124. Marr, C.; Strasser, M.; Schwarzfischer, M.; Schroeder, T.; Theis, F.J. Multi-scale modeling of GMP differentiation based on single-cell genealogies. *FEBS J.* **2012**, *279*, 3488–3500. [CrossRef] [PubMed]

125. Chickarmane, V.; Enver, T.; Peterson, C. Computational modeling of the hematopoietic erythroid-myeloid switch reveals insights into cooperativity, priming, and irreversibility. *PLoS Comput. Biol.* **2009**, *5*, e1000268. [CrossRef] [PubMed]

126. Want, M.Y.; Konstorum, A.; Huang, R.Y.; Jain, V.; Matsueda, S.; Tsuji, T.; Lugade, A.; Odunsi, K.; Koya, R.; Battaglia, S. Neoantigens retention in patient derived xenograft models mediates autologous T cells activation in ovarian cancer. *Oncoimmunology* **2019**, *8*, e1586042. [CrossRef] [PubMed]

127. Want, M.Y.; Karasik, E.; Gillard, B.; McGray, A.J.R.; Battaglia, S. Inhibition of WHSC1 Allows for Reprogramming of the Immune Compartment in Prostate Cancer. *Int. J. Mol. Sci.* **2021**, *22*, 8742. [CrossRef] [PubMed]

128. Sidhom, J.W.; Larman, H.B.; Pardoll, D.M.; Baras, A.S. DeepTCR is a deep learning framework for revealing sequence concepts within T-cell repertoires. *Nat. Commun.* **2021**, *12*, 1605. [CrossRef]

129. Stahl, P.L.; Salmen, F.; Vickovic, S.; Lundmark, A.; Navarro, J.F.; Magnusson, J.; Giacomello, S.; Asp, M.; Westholm, J.O.; Huss, M.; et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **2016**, *353*, 78–82. [CrossRef]

130. Bergenstrahle, J.; Larsson, L.; Lundeberg, J. Seamless integration of image and molecular analysis for spatial transcriptomics workflows. *BMC Genom.* **2020**, *21*, 482. [CrossRef]

131. Dries, R.; Zhu, Q.; Dong, R.; Eng, C.L.; Li, H.; Liu, K.; Fu, Y.; Zhao, T.; Sarkar, A.; Bao, F.; et al. Giotto: A toolbox for integrative analysis and visualization of spatial expression data. *Genome Biol.* **2021**, *22*, 78. [CrossRef]

132. Asada, K.; Komatsu, M.; Shimoyama, R.; Takasawa, K.; Shinkai, N.; Sakai, A.; Bolatkan, A.; Yamada, M.; Takahashi, S.; Machino, H.; et al. Application of Artificial Intelligence in COVID-19 Diagnosis and Therapeutics. *J. Pers. Med.* **2021**, *11*, 886. [CrossRef] [PubMed]