SCIENCE CHINA PRESS

SCIENCE CHINA Life Sciences
life.scichina.com link.springer.com
https://doi.org/10.1007/s11427-023-2561-0

SPECIAL TOPIC: Single-cell omics analysis: methods and applications

REVIEW

Single-cell omics: experimental workflow, data analyses and applications

Fengying Sun[1†], Haoyan Li[2†], Dongqing Sun[3,4†], Shaliu Fu[3,5,6,7†], Lei Gu[8†], Xin Shao[2,9†], Qinqin Wang[8†], Xin Dong[3,4†], Bin Duan[3,5,6,7†], Feiyang Xing[3,4†], Jun Wu[10†], Minmin Xiao[1*], Fangqing Zhao[11*], Jing-Dong J. Han[12*], Qi Liu[3,5,6,7*], Xiaohui Fan[2,9,13*], Chen Li[8*], Chenfei Wang[3,4*] & Tieliu Shi[1,10,14*]

[1]Department of Clinical Laboratory, the Affiliated Wuhu Hospital of East China Normal University (The Second People's Hospital of Wuhu City), Wuhu 241000, China
[2]Pharmaceutical Informatics Institute, College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058, China
[3]Key Laboratory of Spine and Spinal Cord Injury Repair and Regeneration (Tongji University), Ministry of Education, Orthopaedic Department, Tongji Hospital, Bioinformatics Department, School of Life Sciences and Technology, Tongji University, Shanghai 200082, China
[4]Frontier Science Center for Stem Cells, School of Life Sciences and Technology, Tongji University, Shanghai 200092, China
[5]Translational Medical Center for Stem Cell Therapy and Institute for Regenerative Medicine, Shanghai East Hospital, Bioinformatics Department, School of Life Sciences and Technology, Tongji University, Shanghai 200082, China
[6]Research Institute of Intelligent Computing, Zhejiang Lab, Hangzhou 311121, China
[7]Shanghai Research Institute for Intelligent Autonomous Systems, Shanghai 201210, China
[8]Center for Single-cell Omics, School of Public Health, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China
[9]National Key Laboratory of Chinese Medicine Modernization, Innovation Center of Yangtze River Delta, Zhejiang University, Jiaxing 314103, China
[10]Center for Bioinformatics and Computational Biology, Shanghai Key Laboratory of Regulatory Biology, the Institute of Biomedical Sciences and School of Life Sciences, East China Normal University, Shanghai 200241, China
[11]Beijing Institutes of Life Science, Chinese Academy of Sciences, Beijing 100101, China
[12]Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Center for Quantitative Biology (CQB), Peking University, Beijing 100871, China
[13]Zhejiang Key Laboratory of Precision Diagnosis and Therapy for Major Gynecological Diseases, Women's Hospital, Zhejiang University School of Medicine, Hangzhou 310006, China
[14]Key Laboratory of Advanced Theory and Application in Statistics and Data Science-MOE, School of Statistics, East China Normal University, Shanghai 200062, China

†Contributed equally to this work
*Corresponding authors (Minmin Xiao, email: 1084173819@qq.com; Fangqing Zhao, email: zhfq@biols.ac.cn; Jing-Dong J. Han, email: jackie.han@pku.edu.cn; Qi Liu, email: qiliu@tongji.edu.cn; Xiaohui Fan, email: fanxh@zju.edu.cn; Chen Li, email: cli@shsmu.edu.cn; Chenfei Wang, email: 08chenfeiwang@tongji.edu.cn; Tieliu Shi, email: tlshi@bio.ecnu.edu.cn)

Received 7 December 2023; Accepted 18 April 2024; Published online 23 July 2024

Cells are the fundamental units of biological systems and exhibit unique development trajectories and molecular features. Our exploration of how the genomes orchestrate the formation and maintenance of each cell, and control the cellular phenotypes of various organismsis, is both captivating and intricate. Since the inception of the first single-cell RNA technology, technologies related to single-cell sequencing have experienced rapid advancements in recent years. These technologies have expanded horizontally to include single-cell genome, epigenome, proteome, and metabolome, while vertically, they have progressed to integrate multiple omics data and incorporate additional information such as spatial scRNA-seq and CRISPR screening. Single-cell omics represent a groundbreaking advancement in the biomedical field, offering profound insights into the understanding of complex diseases, including cancers. Here, we comprehensively summarize recent advances in single-cell omics technologies, with a specific focus on the methodology section. This overview aims to guide researchers in selecting appropriate methods for single-cell sequencing and related data analysis.

single-cell sequencing | genome | epigenome | proteomics | metabolomics | multimodal | spatial transcriptomics | CRISPR screening

CONTENTS

Citation: Sun, F., Li, H., Sun, D., Fu, S., Gu, L., Shao, X., Wang, Q., Dong, X., Duan, B., Xing, F., et al. (2025). Single-cell omics: experimental workflow, data analyses and applications. Sci China Life Sci 68, 5–102. https://doi.org/10.1007/s11427-023-2561-0

Introduction

The exploration of individual cells enhances our understanding of cellular diversity, disease processes, and the organization of multicellular organisms. Technologies for measuring biological systems at the single-cell level have made exciting advances and are now at the forefront of research. Single-cell RNA sequencing (scRNA-seq) technique allows the dissection of gene expression at single-cell resolution, revolutionizing transcriptomic studies. Since its initial discovery in 2009, more than 60 scRNA-seq protocols have been developed so far (Table S1 in Supporting Information). The maturation of scRNA-seq provides researchers with unique opportunities to catalog human cell types, understand their development, variation between individuals, and unravel their involvement in disease. With the rapid development of single-cell sequencing technology and reduction of cost, this has been widely used to solve critical biomedical problems.

The rapid development of scRNA-seq technology has facilitated the exploration of other omics, including genomics, epigenome, proteomics, and metabolomics. Novel technologies, such as multi-omics technology, spatial scRNA-seq, and CRISPR screening, have also emerged to gain a comprehensive understanding of complex cellular behavior through multi-omics data integration and the incorporation of additional information. Figure 1 illustrates the expanded landscape of single-cell sequencing technologies.

This paper will review the latest developments of single-cell omics technologies from the following eight aspects: (i) Single-cell transcriptome sequencing; (ii) Single-cell whole-genome sequencing; (iii) Single-cell epigenome sequencing; (iv) Single-cell proteomics technology; (v) Single-cell metabolomics technology; (vi) Single-cell multimodal sequencing technology; (vii) Single-cell spatial transcriptomics technology; (viii) Single-cell CRISPR screening technology. The aims are to systematically summarize and discuss in detail currently available single-cell omics technologies, the computational approaches to decipher the single-cell dataset, and their advantages, disadvantages, and applications.

Chapter 1 Single-cell transcriptome sequencing

The advent of single-cell RNA sequencing (scRNA-seq) technology is a state-of-the-art technique for analyzing cellular complexity and heterogeneity, providing a wealth of information across diverse scientific domains. The high resolution of this technology makes it possible to discuss novel biological issues by offering a unique opportunity to explore the transcriptional landscape of single cells. This cutting-edge method, since the first scRNA-seq protocol was introduced in 2009, has seen substantial advancements in method development (Figure 2). These approaches incorporate essential improvements and modifications in single-cell isolation, capture, reverse transcription, cDNA amplification, library preparation, sequencing, and data analysis to enhance throughput and automation while decreasing time and costs. This chapter presents a comprehensive overview of single-cell transcriptome sequencing technologies, bioinformatics analysis methods, and their uses in medical and biological sciences research to help researchers make informed choices.

Overview of scRNA-seq

ScRNA-seq, the pioneered single-cell sequencing technology, has witnessed wide popularity and encompasses a variety of approaches. Despite the diversity in methods, they all follow a similar general process involving four primary steps: (i) isolation of single cells, (ii) reverse transcription (RT), (iii) cDNA amplification, and (iv) sequencing library construction and sequencing (Hedlund and Deng, 2018). Major steps of the scRNA-seq workflow are shown in Figure 3. This section outlines some techniques and solutions related to single-cell isolation and sequencing library construction.

(1) Single-cell isolation

In scRNA-seq, the first and critical step is single-cell isolation, in which tissue dissociation and single-cell separation are considered significant contributors to contamination, batch effects, and procedural disparities (Tung et al., 2017). Thus, to perform high-throughput and unbiased single-cell sequencing, a reliable and accurate capturing of single cells with high efficiency is the key determinant. Early methods for single-cell isolation including limited serial dilution (Gross et al., 2015), manual micromanipulation (Hu et al., 2016a), and laser capture microdissection (LCM) (Emmert-Buck et al., 1996) were low-throughput, time-consuming, inefficient, and technically challenging but are still used to analyze low number of cells (e.g., rare cells) (Dal Molin and Di Camillo, 2019).

Fluorescence-activated cell sorting (FACS), a commonly used high-throughput technique, offers specific and automated isolation of thousands of individual cells but requires a large input volume (numbers of cells for isolation >10,000) (Hu et al., 2016a). Moreover, this technique is inadequate for certain cells exhibiting low marker expression due to the faint or weak

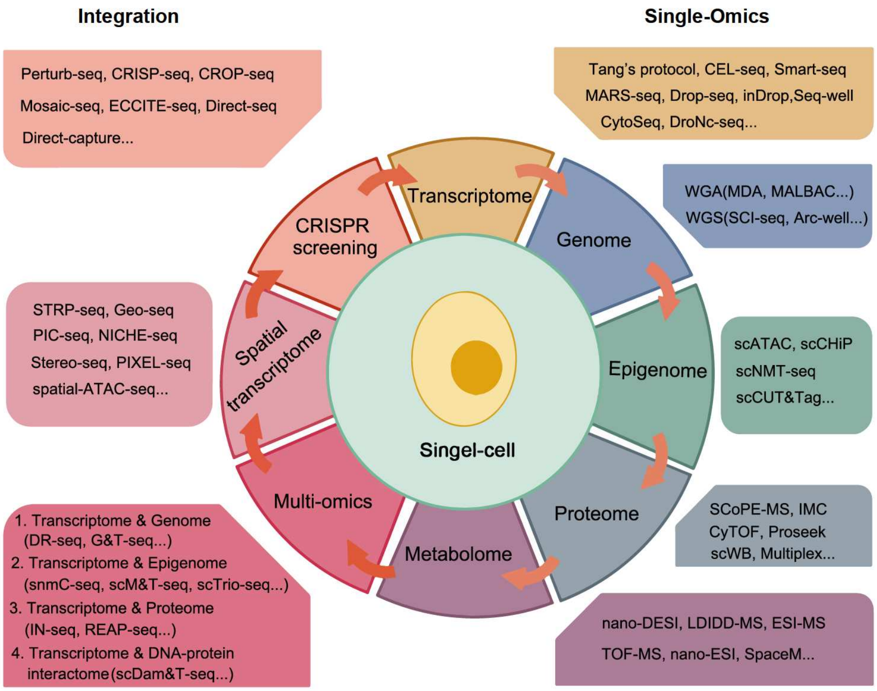

Integration
Single-Omics
Perturb-seq, CRISP-seq, CROP-seq
Mosaic-seq, ECCITE-seq, Direct-seq
Direct-capture...
Tang's protocol, CEL-seq, Smart-seq
MARS-seq, Drop-seq, inDrop,Seq-well
CytoSeq, DroNc-seq...
Transcriptome
CRISPR screening
Genome
WGA(MDA, MALBAC...)
WGS(SCI-seq, Arc-well...)
STRP-seq, Geo-seq
PIC-seq, NICHE-seq
Stereo-seq, PIXEL-seq
spatial-ATAC-seq...
Spatial transcriptome
Epigenome
scATAC, scCHiP
scNMT-seq
scCUT&Tag...
Singel-cell
Multi-omics
Proteome
SCoPE-MS, IMC
CyTOF, Proseek
scWB, Multiplex...
Metabolome
1. Transcriptome & Genome
(DR-seq, G&T-seq...)
2. Transcriptome & Epigenome
(snmC-seq, scM&T-seq, scTrio-seq...)
3. Transcriptome & Proteome
(IN-seq, REAP-seq...)
4. Transcriptome & DNA-protein
interactome(scDam&T-seq...)
nano-DESI, LDIDD-MS, ESI-MS
TOF-MS, nano-ESI, SpaceM...

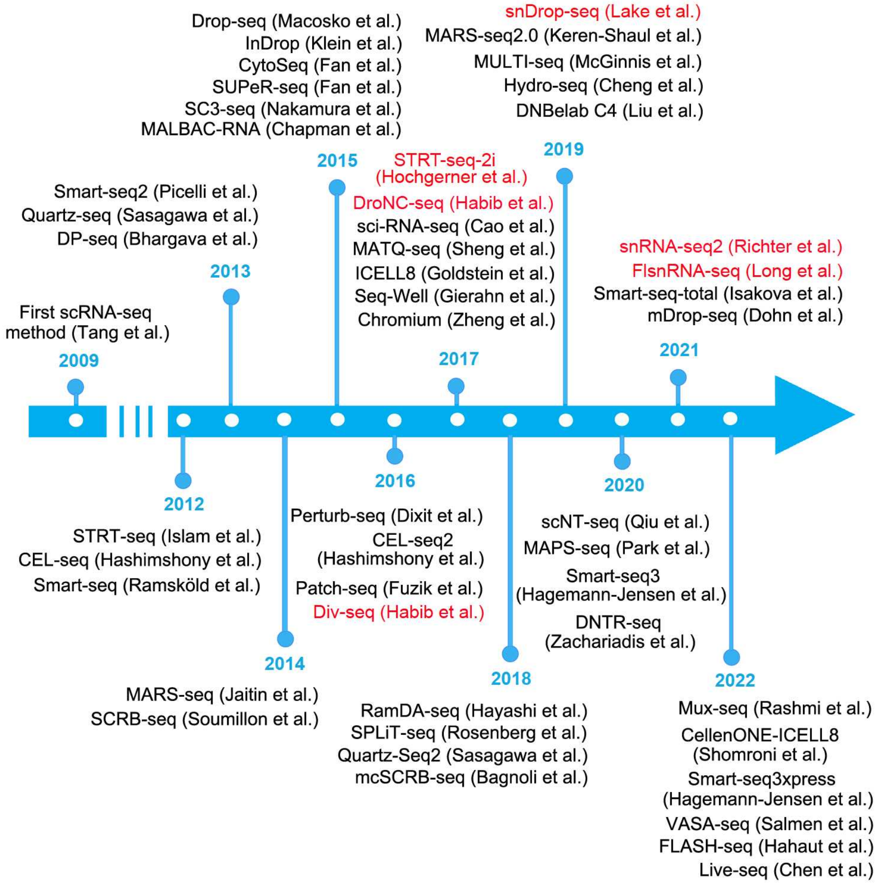
Figure 1. Schematic diagram of single-cell sequencing technologies. Since the inception of the first scRNA-seq in 2009, single-cell sequencing technology has been rapidly expanded to other omics levels and diverse integration approaches. Single omics-level sequencing technologies now include transcriptome, genome, epigenome, proteome and metabolome. Integrated sequencing technologies involve multiple omics data integration, such as transcriptome & genome, transcriptome & epigenome, transcriptome & proteome, transcriptome & DNA-protein interactome. Additionally, these integrated approaches incorporate sequencing data with other layers of information, including spatial data and the CRISPR screening technique. Each type of expansion is represented by the technology listed in the corresponding color box.

fluorescence signal, making it challenging to differentiate subpopulations with similar marker expression (Yasen et al., 2020). Magnetic-activated cell sorting (MACS) is another high-throughput isolation technique designed to separate various cell types based on enzymes, lectins, antibodies, or streptavidin conjugated to magnetic beads, facilitating the binding of specific proteins on the target cells (Hu et al., 2016a). The MACS system boasts a notable advantage, achieving >90% purity for specific cell populations (Miltenyi et al., 1990). However, MACS has inherent limitations compared with FACS due to immunomagnetic techniques that can only isolate cells into negative and positive populations. Moreover, it cannot isolate cells based on low or high expression of a molecule, a capability present in FACS (Hu et al., 2016a). In the current landscape of high-throughput sequencing platforms, methods involving microfluidic-based single-cell manipulation have emerged as the leading technique for single-cell separation in transcriptome studies and have significantly enhanced the scale, efficiency, and accuracy of the isolation process. Microfluidics devices, in which reaction chambers or droplets are used to capture the cells followed by individual steps nanoliter reactions, offer a cost-effective and sample-efficient analysis. These devices are primarily categorized into microwell-based methods, droplet-based methods, and integrated fluidic circuits (IFCs). The integration of microfluidics systems in scRNA-seq has significantly enhanced sequencing throughput, enabling the simultaneous processing and analysis of tens of thousands of single cells. A comprehensive overview of current single-cell isolation technologies, including their advantages and limitations, is presented in Table S2 in Supporting Information.

(2) Reverse transcription and cDNA amplification

A single mammalian cell contains approximately 10 picograms aggregate quantity of RNA, with a predominant portion composed of ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs), whereas messenger RNAs (mRNAs) constitute only 1%–5% of the total (Liu et al., 2014; Wang et al., 2023c). Since mRNA is present in extremely low amounts in a single cell, it is essential to amplify cDNA after the RT process to obtain significant quantities for sequencing library preparation. cDNA amplification can be achieved through either exponential amplification using polymerase chain reaction (PCR) or linear amplification based on *in vitro* transcription (IVT).

Presently, the predominant method for library construction involves PCR-based cDNA amplification, including poly(A) tailing and template switching (TS) methods (Kolodziejczyk et al., 2015). The poly(A) tailing method employs an oligo-dT primer that binds the mRNA 3′-poly(A) tail to reverse transcribe mRNA into cDNA. The poly(A) tailing method is speedy but cannot capture nonpolyadenylated (Poly(A)−) RNA (Hebenstreit, 2012), and the capture efficiency is low, which is reported to be around 10%–15% for current protocols (Islam et al., 2014). Additionally, the termination of the reverse transcriptase reaction may lead to diminished coverage rate of the 5′ end of mRNA in transcription. The TS technique involves the utilization of Moloney murine leukemia virus (MMLV) reverse transcriptase for the TS process. MMLV reverse transcriptase can append a poly (C) tail to the trailing end of the newly synthesized single strand. The poly(C) tail can bind to the 5′-end poly(G) tail of template switching oligonucleotide (TSO) adapter sequences. Following this interaction, a "switch" takes place: the reverse transcriptase utilizes the TSO as a template to synthesize cDNA, to complete the adaptor conversion (Picelli, 2017). The TS method exhibits a reduced susceptibility to nucleic acid loss, but it comes with a lower sensitivity compared with the poly(A) tailing method. PCR-



2009
First scRNA-seq method (Tang et al.)
2012
STRT-seq (Islam et al.)
CEL-seq (Hashimshony et al.)
Smart-seq (Ramsköld et al.)
2013
Smart-seq2 (Picelli et al.)
Quartz-seq (Sasagawa et al.)
DP-seq (Bhargava et al.)
2014
MARS-seq (Jaitin et al.)
SCRB-seq (Soumillon et al.)
2015
Drop-seq (Macosko et al.)
InDrop (Klein et al.)
CytoSeq (Fan et al.)
SUPeR-seq (Fan et al.)
SC3-seq (Nakamura et al.)
MALBAC-RNA (Chapman et al.)
2016
Perturb-seq (Dixit et al.)
CEL-seq2 (Hashimshony et al.)
Patch-seq (Fuzik et al.)
Div-seq (Habib et al.)
2017
STRT-seq-2i (Hochgerner et al.)
DroNC-seq (Habib et al.)
sci-RNA-seq (Cao et al.)
MATQ-seq (Sheng et al.)
ICELL8 (Goldstein et al.)
Seq-Well (Gierahn et al.)
Chromium (Zheng et al.)
2018
RamDA-seq (Hayashi et al.)
SPLiT-seq (Rosenberg et al.)
Quartz-Seq2 (Sasagawa et al.)
mcSCRB-seq (Bagnoli et al.)
2019
snDrop-seq (Lake et al.)
MARS-seq2.0 (Keren-Shaul et al.)
MULTI-seq (McGinnis et al.)
Hydro-seq (Cheng et al.)
DNBelab C4 (Liu et al.)
2020
scNT-seq (Qiu et al.)
MAPS-seq (Park et al.)
Smart-seq3 (Hagemann-Jensen et al.)
DNTR-seq (Zachariadis et al.)
2021
snRNA-seq2 (Richter et al.)
FlsnRNA-seq (Long et al.)
Smart-seq-total (Isakova et al.)
mDrop-seq (Dohn et al.)
2022
Mux-seq (Rashmi et al.)
CellenONE-ICELL8 (Shomroni et al.)
Smart-seq3xpress (Hagemann-Jensen et al.)
VASA-seq (Salmen et al.)
FLASH-seq (Hahaut et al.)
Live-seq (Chen et al.)

Figure 2. Significant works in the field of scRNA-seq over the past 10 years. Black represents scRNA-seq technologies; red represents snRNA-seq technologies.

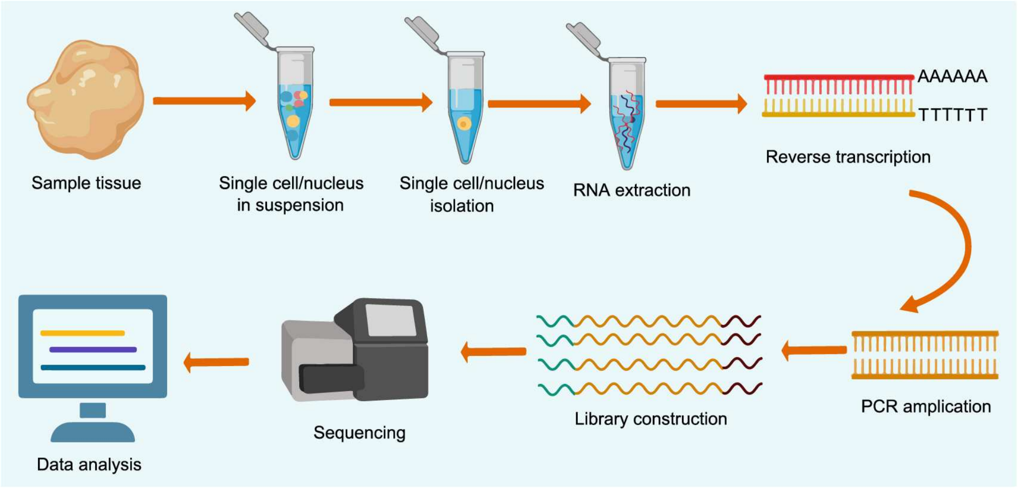


Figure 3. Major steps of scRNA-seq workflow.

based amplification, while capable of rapidly and efficiently amplifying substantial quantities of cDNAs in a short timeframe, is constrained by the inherent characteristics of the exponential amplification process. PCR tends to favor the amplification of shorter and less G-C-rich amplicons, resulting in quantification bias and the accumulation of non-specific transcripts, leading to the loss of original transcript information (Aird et al., 2011). Additionally, PCR can cause the over-presence of highly expressed transcripts in the final library (Aird et al., 2011).

IVT represents another method for cDNA synthesis and amplification. The key strength of IVT is its linear amplification, a feature that mitigates what is commonly known as amplification bias, rendering it considered more precise and reproducible in comparison to PCR (Chen et al., 2017a; Grün and van Oudenaarden, 2015). In this strategy, an oligo-dT primer, encompassing (i) a unique molecular identifier (UMI, featuring random nucleotide sequences that label individual mRNA molecules and are employed for quantifying unique transcripts and correcting amplification biases); (ii) a unique cell barcode; (iii) an Illumina adapter; and (iv) a T7 promoter, initially binds to the 3′-poly(A) tail of mRNA to synthesize both strands of cDNA. Subsequently, the cDNA fragments, each uniquely barcoded, are combined, and the T7 polymerase identifies the T7 promoter sequence, initiating IVT to generate additional RNA molecules. Another round of RT is required to reconvert the amplified RNA molecules into cDNAs for sequencing library construction. Consequently, the resultant cDNAs exhibit a pronounced bias in coverage toward the 3′ end. Moreover, this approach is characterized by its time-intensive and labor-intensive nature, contributing to its less widespread use compared with PCR (Hashimshony et al., 2012).

As a unique step of scRNA-seq, all transcripts originating from a single cell obtain a unique barcode-a brief oligonucleotide sequence introduced into the cDNA during RT to distinguish the transcriptome of individual cells. Leveraging the unique barcode information, these transcripts can be readily attributed to their respective cells. Consequently, a multitude of single-cell transcriptomes, each uniquely labeled with cell barcodes, are amalgamated for the library construction and subsequent sequencing in a single run. This approach significantly diminishes costs and augments sequencing throughput. The cell barcode serves as a highly effective strategy for enabling parallel processing and was initially used in the STRT-seq method in 2011 (Islam et al., 2011). Subsequently, in 2014, Islam et al. (2014) further innovated by introducing a UMI to identify each cDNA molecule within a cell. Like cell barcodes, a UMI is likewise a 6–8 base pair random oligonucleotide sequence that can be integrated into each transcript within a cell during RT. With a substantial pool of UMIs, each transcript is tagged with a distinct barcode, thereby ensuring that all duplicated molecules produced through PCR amplification retain the original transcript's UMIs. By counting the unique barcodes, the number of original copies of the transcript can be accurately quantified, thereby eliminating the bias of PCR amplification.

Regarding the three pivotal issues in single-cell RNA sequencing: (i) isolating individual cells; (ii) minimizing RNA loss during RT; and (iii) producing enough DNA for sequencing, researchers have undertaken explorations into technological advancements in recent years. This has led to the proposal of various scRNA-seq methodologies tailored for the comprehensive study of single-cell transcriptomics.

The different scRNA-seq technologies are rooted in shared fundamental principles, yet they vary in at least one of the following aspects: (i) single-cell isolation; (ii) RT; (iii) amplification of cDNA; (iv) transcript coverage; (v) UMI; (vi) strand-specificity. Each approach has its unique strengths and limitations. The key features of these diverse scRNA-seq technologies are consolidated in Table S1 in Supporting Information. Consequently, researchers have the flexibility to choose a suitable scRNA-seq method based on technical characteristics, advantages, cost factors, and throughput demands.

Currently available scRNA-seq technologies

Low-throughput scRNA-seq technologies

Over 60 distinct approaches have been developed for scRNA-seq, as detailed in Table S1 in Supporting Information. In general, these approaches can be categorized into two main types: low-throughput and high-throughput methods. The fundamental chemistry developed in low-throughput approaches is geared towards enhancing sequencing sensitivity and accuracy, while reducing costs and technical noise. Particularly, sensitivity stands out as the foremost critical feature, serving as a fundamental indicator of method performance. High-throughput methods have evolved from the essential chemistries of several classic low-throughput approaches. We will discuss these technologies in detail below.

(1) Tang's protocol

Tang's protocol, developed by the Surani group in 2009, stands as the pioneering scRNA-seq method. In this technique, a microscope is used to manually select a single mouse blastomere. Following that, the cells are lysed, and the mRNAs are converted into cDNAs through RT using a poly(T) primer featuring an anchor sequence (UP1). A poly(A) tail is then added to the trailing end of the first strand end with the help of terminal transferase. Following this, the second strand cDNAs are synthesized using the second poly(T) primer featuring another anchor sequence (UP2). The cDNAs are then efficiently amplified through PCR using UP1 and UP2 primers, and libraries are constructed for sequencing on the SOLiD system. This method can generate nearly full-length cDNA of transcripts and detect ~13,000 genes (Tang et al., 2009). This technology mainly helps in the unveiling of new transcripts and alternative splicing isoforms.

Despite representing a significant advancement for the emerging field of scRNA-seq at the time, this approach has considerable limitations. First, the approach can only identify mRNA that has a poly(A) tail; it cannot capture mRNAs without poly(A) tails, such as histone mRNAs, miRNA, circular RNA (circRNA), and nascent RNA. This is because the method depends on poly(T) primers for the capturing of mRNA via the poly(A) tail for initiating RT reactions. Second, inefficiencies in the enzymatic reactions contribute to a reduction in sequencing sensitivity, resulting in the loss of low-expression transcripts. Third, this method is not strand-specific and cannot distinguish between sense and antisense transcripts. Thus, it is not commonly used.

(2) STRT-seq

In 2011, Islam et al. (2011) developed a single-cell tagged reverse transcription sequencing (STRT-seq) method, a highly multiplexed approach for scRNA-seq on the Illumina platform. This method introduced a barcode and an upstream primer-binding sequence through the template-switching mechanism during RT, facilitating strand-specific amplification of 3′ ends and high throughput 96-cell multiplexing. Using an array-based strategy, STRT-seq supports the processing of up to 800 individual cells. A key advantage of this method is that it can add a unique barcode sequence to each cell during the RT process, thereby enabling large-scale detection of various mixed cell samples, such as highly heterogeneous tumor cell samples. Compared with Tang's method, STRT-seq significantly reduces costs and processing time through its early barcoding strategy. However, it employs multiple cycles of PCR, potentially introducing PCR bias. This method has many applications in the biomedical field, including the characterization of tumor heterogeneity and the identification of potential novel biomarkers or drug targets for disease diagnosis and treatment (Cui et al., 2021; Song et al., 2022a; Tian et al., 2022).

(3) Smart-seq

In 2012, Ramsköld et al. (2012) developed Smart-seq, a reliable and consistent approach for scRNA-seq, a switching mechanism at the leading end of the transcript, which demonstrated significant improvement, surpassing 40% efficiency in full-length cDNA synthesis for transcripts. The publication of the technology is a landmark in the field of scRNA-seq studies. The core principle of Smart-seq involves utilizing poly(T) primers and SMART-TS technology to convert polyadenylated (poly(A)+) RNA into full-length cDNA. The resultant cDNA molecules, after the amplification using PCR, are then used to create Illumina sequencing libraries by the Nextera Tn5 transposome technique. This method considerably improves the ability to detect alternatively spliced exons and low-abundance expressed transcripts. Smart-seq method has been widely used in medicine, such as analyzing gene expression profiles of $CD177^+$ cells in the liver of a mouse model with biliary atresia (Zhang et al., 2022d); to analyze the genome-wide expression of skeletal muscle stem, niche cells, and single myofibers (Blackburn et al., 2019; Blackburn et al., 2021); and investigating differences in dermal $CD4^+$ Trm cells between patients with acute cutaneous lupus erythematosus and normal controls (Zhao et al., 2022c).

Smart-seq2 was developed to overcome the limitations of coverage, less productivity, and sensitivity issues observed in Smart-seq (Picelli et al., 2013). In order to increase cDNA library yield and length, several improvements were implemented in Smart-seq2 including enhancement of reverse RT, TSOs, and the reamplification of PCR. A notable enhancement in cDNA yield, approximately twofold, was achieved by incorporating a locked nucleic acid (LNA) guanylate at the 3′ end of TSO as opposed to Smart-seq. This improvement can be attributed to the heightened thermal stability of LNA-DNA base pairs. Moreover, the addition of the methyl group donor betaine, along with a higher concentration of $MgCl_2$, also led to a substantial increase in cDNA yield. Commencing the addition of deoxyribonucleoside triphosphates (dNTPs) before RNA denaturation, as opposed to incorporating them in the RT master mix, enhanced the average length of the preamplified cDNA. This improvement is likely attributed to increased stability in the hybridization of RNA to the oligo-dT primer. The use of KAPA HiFi Hot Start DNA polymerase improved cDNA generation and achieved greater cDNA length. Smart-seq2 transcriptome libraries outperform Smart-seq in terms of detection strength, coverage rate, bias, and accuracy. Smart-seq2 transcriptome libraries can be generated with off-the-shelf reagents even at a lower cost, allowing the in-depth analysis of entire exons of each transcript and also detecting different splice variants. This method also facilitates thorough analysis of single nucleotide polymorphisms (SNPs) and mutations. However, it has limitations, such as the lack of strand specificity and the incapability to detect poly(A)− RNAs (Picelli et al., 2014). Additionally, the cell isolation process using micropipettes is time-consuming and low-throughput.

Smart-seq3 represents an enhancement in sensitivity achieved through the optimization of RT and TS conditions (Hagemann-Jensen et al., 2020). The optimal parameters include the utilization of Maxima H-minus reverse transcriptase, transitioning the RT salt from KCl to NaCl or CsCl, performing RT in the presence of 5% polyethylene glycol (PEG), and incorporating GTPs or dCTPs to enhance and stabilize the TS reaction. A distinctive feature of Smart-seq3 is its integration of full-length transcriptome coverage with a 5′ UMI RNA counting strategy, which elevates the precision of transcript counting without sacrificing overall coverage. In this approach, a TSO primer is constructed, comprising a partial Tn5 motif, an 11 base pair tag sequence, an 8 base pair UMI sequence, and three riboguanosines. Following sequencing, the 11 base pair tag is utilized to unequivocally differentiate 5′ UMI-labeled reads from internal reads. Within a single sequencing reaction of Smart-seq3, both 5′ UMI-labeled reads and internal reads that span the entire transcript without UMIs are collected. This approach allows for the accurate quantification of the original transcripts using UMI reads, correcting for the nonlinear PCR amplification bias. The reconstruction of full-length transcripts is achieved through the use of internal reads. Many current sequencing library construction methods incorporate cell barcodes and UMI tag strategies. However, as these can only be introduced at the ends of the cDNA, the resulting sequenced cDNA sequences are limited to one end of the transcript, leading to the loss of significant sequence information in the middle of the transcript. Consequently, tag-based methods are primarily employed for gene expression quantification and are unsuitable for isoform identification or splicing. Despite their ability to capture full-length transcripts, Smart-seq and Smart-seq2 are unable to utilize barcodes or UMIs to tag transcripts, making them incompatible with high-throughput, parallel single-cell sequencing. Additionally, without a UMI tag, these methods cannot address the amplification bias introduced by PCR. However, Smart-seq3 overcomes the limitation of incompatibility between full-length transcript coverage and UMI by utilizing a special TSO primer.

Smart-seq-total was designed to address the limitation of capturing only poly(A)+ RNA molecules in the previous Smart-seq technologies (Isakova et al., 2021). The primary advancement in this approach lies in the utilization of Escherichia coli poly(A) polymerase to add adenine tails to the 3′ end of RNA molecules. Consequently, all poly(A)+ RNAs are reverse-transcribed using a poly(T) primer that incorporates a UMI, along with the TSO. This modification enables Smart-seq-total to capture diverse RNA forms concurrently, encompassing protein-coding, long-noncoding, microRNA, and other noncoding RNA transcripts within a single cell. Such an approach facilitates the exploration of regulatory connections between coding and noncoding transcripts in a cell, offering insights into the intricate regulatory landscape. However, it is important to note that Smart-seq-total does have some limitations. Firstly, it cannot assess circRNA. Secondly, it results in the loss of the endogenous polyadenylation status of transcripts. Despite these drawbacks, Smart-seq-total exhibits significant potential for uncovering noncoding regulatory patterns governing cellular functions and contributing to the definition of cellular identity.

(4) CEL-seq

Cell expression by linear amplification and sequencing (CEL-seq) was the pioneering method to utilize linear strand-specific in IVT for RNA amplification from single cells (Hashimshony et al., 2012). Initiation of the procedure includes the synthesis of the first-strand cDNA using a primer featuring an anchored poly(T), a specific barcode, the 5′ Illumina sequencing adaptor, and a T7 promoter. Subsequently, the second strand is generated to produce double-stranded cDNA containing a T7 promoter. Combined cDNA samples from several cells undergo IVT, initiated by the T7 promoter, enabling linear amplification of cDNA. The resulting amplified RNAs are then converted into cDNAs for sequencing. By employing linear amplification, CEL-seq minimizes amplification bias, delivering a more sensitive and reproducible outcome compared with full-length cDNA coverage techniques like Smart-seq. However, CEL-seq has limitations, such as the inability to detect miRNAs and other poly(A)− transcripts and difficulty in distinguishing alternative splice forms due to its strong 3′ bias.

CEL-seq2, an improved version of CEL-seq increased sensitivity, reduced costs, and decreased hands-on time (Hashimshony et al., 2016). To mitigate mRNA molecule counting biases, CEL-seq2 introduces 5 base pair UMIs upstream of the barcode. Utilizing the SuperScript II Double-Stranded cDNA Synthesis Kit, coupled with a modification of the CEL-seq primer length, notably enhances the efficiency of the RT reaction. CEL-seq2 outperforms the original CEL-seq method by detecting twice as many transcripts and 30% more genes per cell. However, it is important to note that despite its notable 3′ bias, CEL-seq2 does not offer information on the majority of splicing events. Nevertheless, its increased sensitivity and capability for individual transcript counting provide a clear advantage for various applications in transcriptomics.

(5) SUPeR-seq

In single-cell universal poly(A)-independent RNA sequencing (SUPeR-seq), random primers with fixed anchor sequences are utilized instead of the commonly used oligo-dT primers in cDNA synthesis. This enables the detection of both poly(A)+ and poly(A)− RNAs within a single cell (Fan et al., 2015b). The process involves the utilization of the random primers with a fixed anchor sequence (AnchorX-T15N6) to reverse-transcribe total RNAs into first-strand cDNAs. Following the synthesis of the initial cDNA strand, ExoSAP-IT is employed to digest excess primers, preventing the formation of primer-dimer complexes. Using terminal deoxynucleotidyl transferase and dATP with 1% ddATP, a poly(A) tail is appended to the 3′ end of the initial cDNA strand. Subsequently, poly(T) primers with an alternative anchor sequence (AnchorY-T24) are applied to generate the second-strand cDNA, which undergoes amplification through PCR using AnchorY-T24 and AnchorX-T15 primers for subsequent sequencing. SUPeR-seq has been utilized to investigate the regulatory mechanisms of circRNAs during early embryonic development in mammalians. However, it presents challenges for high-throughput sequencing as well as molecule counting due to the absence of UMIs and cell barcodes.

(6) MATQ-seq

In 2017, Sheng et al. (2017) introduced the multiple annealing and dc-tailing-based quantitative single-cell RNA-seq (MATQ-seq) method. This technique incorporates barcodes and UMIs to sequence both polyA+ and polyA− RNAs, distinguishing it from SUPeR-seq. The procedure encompasses converting total RNAs into first-strand cDNA using primers designed for multiple annealing and looping-based amplification cycles (MALBAC). These primers contain mainly G, A, and T bases, along with MALBAC-dT primers. After RT, the first-strand cDNA undergoes dC tailing, followed by the synthesis of second-strand cDNA using G-enriched MALBAC primers. UMIs are presented during second-strand synthesis. Unlike Smart-seq2 and SUPeR-seq, MATQ-seq utilizes UMIs to significantly reduce leading or trailing-end bias in HEK293T transcripts. Additionally, MATQ-seq exhibits higher sensitivity than Smart-seq2 and SUPeR-seq in capturing polyA− RNAs, with a capture efficiency of 89.2%±13.2%. This improvement enhances the efficiency of detecting the low-abundance genes. MATQ-seq's high accuracy and sensitivity allow for the detection of subtle differences in gene expression among individual cells within the same population. However, similar to SUPeR-seq, the time-consuming cell isolation method involving a mouth pipette limits MATQ-seq's throughput.

(7) FLASH-seq

FLASH-seq, a swift and highly profound full-length scRNA-seq method, was developed by Hahaut et al. (2022). Several key modifications were introduced to enhance the efficiency of the Smart-seq2 protocol: (i) it combined the reverse transcription and cDNA preamplification, streamlining the process; (ii) Superscript IV, a more processive reverse transcriptase, replaced Superscript II, and the RT reaction time was shortened; (iii) the amount of dCTP was increased to favor the C-tailing activity of Superscript IV and enhance the template-switching reaction; (iv) the 3′-terminal locked nucleic acid guanine in the template-switching oligonucleotide was replaced with riboguanosine; (v) the reaction volume was reduced to 5 μL (Hahaut et al., 2022). These modifications collectively led to a substantial decrease in both time and cost. FLASH-seq can be completed in approximately 4.5 h, making it 2–3.5 h faster than other methods like Smart-seq2. The cost per cell is lower than other commercial and non-commercial methods, comparable to Smart-seq3, and amounts to less than $1 per cell. Furthermore, FLASH-seq can detect a substantial number of SNPs. It is considered suitable for researchers seeking affordable, automation-friendly, robust, and efficient methods for single-cell transcriptional profiling.

High-throughput scRNA-seq technologies

(1) Strategy for developing high-throughput scRNA-seq

In the early stages of scRNA-seq, micromanipulators or LCM were employed to isolate individual cells for separate transcriptome amplification and library construction. However, these methods had limitations as they could only analyze a few cells in a single experiment. The introduction of cell-specific barcodes revolutionized the field, allowing thousands of single-cell transcriptomes to be mixed together for library construction and sequencing in a single run, enabling high-throughput parallel sequencing. A notable advancement in this direction was the development of the MARS-seq method, which combined FACS with automatic liquid handling to successfully sequence thousands of cells in a single experiment (Jaitin et al., 2014). Three levels of barcodes were used to label cells, plates, and mRNAs, facilitating the mixing of all materials for subsequent automated processing. Similar to MARS-seq, STRT-Seq-2i aimed to increase sequencing throughput by implementing a specialized FACS and barcoding protocol (Hochgerner et al., 2017). This method utilized a custom aluminum plate with 9,600 wells arranged in 96 subarrays of 100 wells each, enabling the simultaneous sequencing of 9,600 cells in one run. However, despite the improvements in throughput achieved by these plate-based methods, the number of cells that could be analyzed was still limited.

The introduction of microfluidic techniques has fundamentally addressed the challenges associated with high-throughput single-cell operations. In 2012, the Fluidigm C1 system became the first commercially available automated microfluidic platform designed for the automatic isolation of cells, cell lysis, cDNA synthesis, amplification, and library preparation for 96 single cells simultaneously. However, the processing capacity of this system was limited and fell short of meeting the demands for high-throughput parallel sequencing. A revolutionary breakthrough occurred in 2015 with the advent of droplet-based microfluidic scRNA-seq technologies. These approaches, exemplified by methods such as those developed by Klein et al. (2015) and Macosko et al. (2015), enabled the concurrent processing of thousands of cells (Klein et al., 2015; Macosko et al., 2015). This marked a significant advancement, allowing for truly massive parallel sequencing in the field of single-cell genomics.

Subsequently, additional high-throughput parallel sequencing strategies were developed, including sci-RNA-seq (Cao et al., 2017) and SPLiT-seq (Rosenberg et al., 2018). These approaches employ a combinatorial indexing method to label cells without the physical compartmentalization of single cells. Notably, these technologies exhibit high cell labeling efficiencies, are straightforward to operate, and can substantially reduce the cost of sequencing. In the following, we will delve into a detailed discussion of these innovative technologies.

(2) Plate-based high-throughput scRNA-seq methods

High-throughput scRNA-seq methods based on microplates involve sorting cells onto microplates through FACS and utilizing barcodes to label cell transcripts for subsequent high-throughput sequencing. These methods offer the advantage of processing any number of individual cells without significantly impacting the cost per cell. On the other hand, alternative high-throughput approaches, such as droplet-based technologies like Drop-seq (Macosko et al., 2015), InDrop (Klein et al., 2015), and 10x Chromium (Zheng et al., 2017), tend to be cost-effective primarily when analyzing a very large number of cells simultaneously. The sequencing throughput of plate-based methods is constrained by the number of microplates utilized. The following section will provide a detailed discussion of these technologies.

1) Quartz-seq. To decipher the biological functions and fundamental causes of non-genetic cellular heterogeneity, Sasagawa et al. (2013) developed the simple yet highly quantitative Quartz-seq technique. Quartz-seq enhances the simplicity and quantitative performance of whole-transcript amplification by addressing three critical aspects. Firstly, in order to overcome the problem of overabundance of byproducts in previous poly(A) tail reaction-based whole-transcript amplification techniques, Quartz-seq combines exonuclease I treatment, a regulated poly(A) tail, and an optimized suppression PCR. This strategic approach completely eliminates the synthesis of byproducts, simplifying subsequent scRNA-seq analysis. Secondly, the technology adopts a robust DNA polymerase (MightyAmp DNA Polymerase) optimized for a single-tube reaction. This choice of PCR enzyme enhances cDNA yield, improves the reproducibility of whole-transcript amplification replication, and reduces the number of required PCR cycles. Finally, Quartz-seq optimizes the efficiencies of RT and second-strand synthesis by adjusting the annealing temperature. Notably, all steps of this method are consolidated into a single PCR tube, eliminating the need for purification and involving only six reaction steps per single cell. This streamlining significantly simplifies the Quartz-seq approach, facilitating its high-throughput implementation. Beyond its simplicity, Quartz-seq exhibits high quantitative accuracy, reproducibility, and sensitivity. Consequently, it can discern various types of non-genetic cellular heterogeneity and differentiate between distinct cell types and cell-cycle phases within the same cell type.

Quartz-seq2 (Sasagawa et al., 2018), an innovative high-throughput scRNA-seq method, was developed as an extension of Quartz-seq. This approach involves sorting single cells using FACS into a 384-well plate, followed by RT using a primer that combines an oligo-dT sequence, a cell barcode sequence, and a UMI sequence. By applying advancements in multiple molecular biological stages, including a major upgrade of poly(A) tagging, Quartz-seq2 achieves excellent UMI conversion efficiency. Notably, Quartz-seq2 utilizes a poly(A) tagging strategy based on the combination of T55 buffer and the increment temperature condition, resulting in an approximately 3.6-fold increase in the amount of cDNA.

The UMI conversion efficiency of Quartz-seq2 is notably high, ranging from 32% to 35%, surpassing that of other methods such as CEL-seq2, SCRB-seq, and MARS-seq (7%–22%). Quartz-seq2 can identify more transcripts from fewer sequence reads at a lower cost because of its increased efficiency. Similar to MARS-seq, Quartz-seq2 employs FACS for cell sorting, a process that requires skilled workers. Despite this requirement, the technology's enhanced UMI conversion efficiency and cost-effectiveness make it a valuable method for scRNA-seq applications.

2) MARS-seq. MARS-seq, an automated and highly parallel RNA single-cell sequencing technology developed in 2014 (Jaitin et al., 2014), revolutionized the field by enabling the counting of unique RNA molecules through the introduction of UMIs in the oligo-dT primer. Single cells are sorted using FACS into 384-well plates as part of the MARS-seq procedure. The RT and library construction processes follow the CEL-seq protocol, ensuring a systematic and reproducible approach. One of the key innovations of MARS-seq lies in its automation, with every step of the method being executed by a liquid-handling robot. This automation improves the technique's repeatability and leads to a significant boost in throughput. The high-throughput and automated nature of MARS-seq makes it applicable to diverse tissues and organs in both normal and disease states. By delineating the cell-type and cell-state compositions of tissues, MARS-seq contributes to a comprehensive understanding of these biological entities, linking this information to detailed genome-wide transcriptional profiles.

In 2019, the research team developed an integrated pipeline for index sorting and massively parallel single-cell RNA sequencing (MARS-seq2.0), building on the foundation of the MARS-seq method (Keren-Shaul et al., 2019). MARS-seq2.0 offers the capability to efficiently sequence 8,000–10,000 cells in a single run, representing a significant enhancement in throughput. Notably, the method achieves an eight-fold reduction in the volume of the RT reaction, decreasing it from 4 μL to 500 nL. The cost of preparing single-cell libraries will drop six-fold as a result of this volume reduction. MARS-seq2.0 is a 3′-based scRNA-seq technique, which limits its use for determining alternative splicing isoforms or particular sequences at the 5′ end of the gene. This is an essential point to remember. Despite this limitation, the integration of indexed FACS sorting with scRNA-seq in MARS-seq2.0 proves beneficial for identifying rare subpopulations and processing rare cells in human clinical samples. Moreover, MARS-seq2.0 provides a flexible platform that enables simultaneously obtain multiple layers of information on the same single cell, encompassing genetics, signaling, epigenetics and spatial location by combining unbiased transcriptional mapping with large numbers of fluorescent markers. This multifaceted approach contributes to a deeper molecular understanding of physiological processes and diseases.

3) SCRB-seq. To economically characterize the major patterns of gene expression variation across heterogeneous ppulations, Soumillon et al. (2014) developed single-cell RNA barcoding and sequencing (SCRB-seq) based on the Smart-seq protocol to profile mRNAs from large numbers of cells using minimal reagents and sequencing reads per cell. In SCRB-seq, FACS is used to sort individual cells into a 384-well plate. Poly(A)+ mRNA is converted to cDNA by RT, which uses a template-switching reverse transcriptase and RT primers made of barcodes, UMIs, and a poly(T) primer. After that, strand information is preserved and the decorated cDNA from several cells is combined and amplified for sequencing using a modified transposon-based fragmentation technique that enriches for the 3′ end. SCRB-seq technology is capable of sequencing about 12,000 single cells, providing deep and full-length transcriptome coverage sequencing. Moreover, it requires roughly two times fewer enzymatic reactions, purifications, and liquid transfer steps than the MARS-seq approach (Jaitin et al., 2014). SCRB-seq, in contrast to Smart-seq, adds unique cell barcodes during RT, making it easier to identify reads that come from the same cell and boosting sequencing throughput. Nevertheless, sequencing a larger number of single cells still faces challenges.

4) STRT-seq-2i. STRT-seq-2i is a dual-index 5′ single-cell and nucleus RNA-seq method designed to significantly increase throughput (Hochgerner et al., 2017). It utilizes a specially designed 9,600-microwell plate, contributing to enhanced efficiency. The microwell array allows for imaging verification of single-cell wells, reducing the occurrence of doublets in a single well. In addition to maintaining several advantages, such as 5′-end reads that reveal transcription start sites, the addition of UMIs for absolute quantification, and the use of single-read sequencing rather than paired-end sequencing to maximize cost efficiency, STRT-seq-2i is still compatible with the previously described STRT-seq method. This technique has demonstrated its adaptability to various experimental contexts by being used to examine the transcriptional profile of both fresh single mouse cortical cells and frozen post-mortem human cortical nuclei.

(3) Microfluidics-based high-throughput scRNA-seq methods

Droplet- and microwell-based platforms stand out as the predominant technologies for high-throughput scRNA-seq, capable of profiling transcriptomes from approximately 10,000 individual cells in a single experiment. Both methods basically involve separating individual cells into many nanoliter-sized containers (such as microwells or water-in-oil droplets) that contain the chemicals required for RT. The integration of cell barcoding strategies into these microfluidic platforms significantly enhances throughput while concurrently reducing costs when compared with both nonmicrofluidic methods and microfluidic methods featuring valves. This advancement is particularly advantageous for biomedical research applications demanding the comprehensive transcriptomic profiling of a vast number of cells.

1) Droplet-based scRNA-seq technologies. (i) InDrop. The fundamental technology of inDrop (Klein et al., 2015) consists of encasing single cells into droplets that contain lysis buffer, RT reagents, and barcoded hydrogel microspheres (BHMs). Each BHM contains ~10^9 photocleavable barcoded primers (147,456 distinct barcodes). The microfluidic device employed in this technology incorporates inlets for carrier oil, cells, lysis/RT reagents, and BHMs, along with an outlet for droplet collection. Each BHM is covalently linked to barcoded primers, featuring a T7 RNA polymerase promoter, an Illumina sequencing adaptor, a unique cell barcode, UMI, and a poly(T) tail, connected via a photo-releasable bond. All BHMs within a sample share the same cell barcode for sample distinction but possess distinct UMIs for precise transcript counting. After encapsulation, ultraviolet (UV) exposure facilitates the photo-release of barcoded primers, allowing mRNA from lysed cells to be barcoded during cDNA synthesis while remaining confined to the same droplet. Subsequently, cDNAs from all cells are pooled post-droplet breakage for library construction and sequencing, following the CEL-seq protocol (Hashimshony et al., 2012). InDrop's intrinsic scalability, unlike conventional methods, is not constrained by the number of reaction chambers. Additionally, the operation processes are simplified by conducting lysis and RT within the droplets. However, a notable drawback of inDrop is its relatively low mRNA capture efficiency (~7%), rendering genes with transcript abundances below 20–50 transcripts challenging to reliably detect in a single cell.

(ii) Drop-seq. Similar to inDrop, Drop-seq is designed for the high-throughput analysis of mRNA expression in thousands of single cells. It achieves this by co-encapsulating each cell with a uniquely barcoded bead within nanoliter-scale water-in-oil droplets for simultaneous processing (Macosko et al., 2015). In contrast to inDrop, which utilizes barcoded hydrogel microspheres, Drop-seq employs beads made of an unchanging hard resin. These resin beads are directly synthesized with barcoded primers, which include a poly(T) sequence for mRNA capture, a cell barcode, a UMI, and a universal PCR handle for amplification. Each bead contains over 10^8 different primers. Following cell lysis in the droplet, released mRNAs hybridize with the primers' poly(T) tails on companion beads to generate single-cell transcriptomes that are affixed to microparticles (STAMPs). Following droplet breakup, thousands of STAMPs are pooled together, reverse-transcribed, PCR-amplified, and sequenced in a single reaction. The Drop-seq method boasts a significantly larger number of unique barcodes (16,777,216) compared with inDrop (147,456), enabling cost-effective and rapid high-throughput analysis.

(iii) 10x Genomics. The 10x Genomics system (Zheng et al., 2017) is a fully commercial platform that shares some similarities with inDrop and Drop-seq. The gel bead in emulsion (GEM) is the fundamental component of this method. To create GEM, an 8-channel microfluidic device is used. In about 6 min, this chip can produce 100,000 GEMs, each of which can contain thousands of cells. Every gel bead in the GEM contains barcoded oligonucleotides comprising Illumina adapters, 10x barcodes, UMIs, and a poly(T) tail, which primes poly(A)+ RNA transcripts. After co-encapsulation of cells and gel beads into droplets, cell lysis occurs immediately, releasing mRNAs. Following this, gel beads dissolve and release the barcoded oligonucleotides, enabling RT of poly(A) + RNAs. The RT reaction takes place within each individual droplet, resulting in cDNA molecules that possess a shared barcode per GEM, a unique UMI, and end with a TSO at the 3′ end. The barcoded cDNA molecules are then combined for PCR amplification, adhering to the Smart-seq methodology, after the emulsion has been broken.

InDrop, Drop-seq, and Chromium are three similar platforms that employ droplet-microfluidic approaches to isolate single cells for high-throughput sequencing. All three methods can process tens of thousands of cells rapidly each day. The three technologies differ in the following four aspects:

Firstly, inDrop and 10x Genomics use hydrogel microspheres, while the beads used in Drop-seq are hard resin. Encapsulation of the soft and pliable hydrogel beads used in inDrop and 10x Genomics, which are densely packed in the microfluidic channel, can be synchronized to produce a super-Poisson distribution. A double Poisson distribution governs the encapsulation of tiny hard beads for a single cell and bead in the same droplet. Thus, inDrop and 10x Genomics can achieve significantly higher cell capture efficiency compared with Drop-seq. It has been reported that the cell capture rate of 10x Genomics is about 50% (Zheng et al., 2017).

Secondly, inDrop and 10x Genomics use hydrogel beads, allowing the primers to be fixed inside the beads, whereas the primers of Drop-seq can only be fixed on the surface of the smaller hard beads. After encapsulation, inDrop uses UV-irradiation-induced cleavage to release primers. The 10x Genomics releases all of the primers into the solution by directly dissolving the beads to improve the capture efficiency of mRNA. In contrast, the primers of Drop-seq cannot be released from the beads, and mRNA molecules hybridize to the poly(T) tail on the beads to form STAMPs for RT, which could be a drawback of Drop-seq compared with inDrop and 10x Genomics.

Thirdly, in Drop-seq, droplets are split immediately as soon as the mRNA and primer hybridize, and all STAMPs are mixed together to perform RT. Instead, in the inDrop and 10x Genomics methods, RT reagents are co-encapsulated into droplets, and RT reactions are conducted independently within each droplet, which is beneficial for improving the specificity of cDNA conversion, enhancing relative yield, and reducing reagent consumption (Streets et al., 2014).

Finally, the three platforms adopt different library construction strategies. While Drop-seq and 10x Genomics use a template-switching procedure akin to the well-known Smart-seq chemistry, inDrop uses the CEL-Seq approach. They so inherit the benefits and drawbacks of Smart-seq and CEL-seq, respectively.

(iv) MULTI-seq. To deliver cell-specific barcodes during RT, all of the droplet-based techniques previously discussed generally use the co-encapsulation strategy, which entails simultaneously encapsulating cells and barcoded beads. The MULTI-seq method (sample multiplexing for single-cell and single-nucleus RNA sequencing using lipid-tagged indices) was recently introduced by McGinnis et al. (2019b) as an alternative strategy that makes unique use of cell barcodes. In this technology, DNA barcodes are labeled onto the plasma membranes of single cells by hybridization to an "anchor" lipid-modified oligonucleotide (LMO). The hydrophobic 5′ lignoceric acid amide of the "anchor" LMO binds to membranes; hybridization to a "co-anchor" LMO with a 3′ palmitic acid amide amplifies the hydrophobicity of the complex, extending membrane retention. The 3′ poly(A) capture sequence, the 8 bp sample barcode, and the 5′ PCR handle make up the LMO. Each single cell or nuclei carried by the LMOs is co-encapsulated with an mRNA capture bead into an emulsion droplet generated by the 10x Genomics system. Sample demultiplexing is made possible by the release of endogenous mRNAs and LMOs upon in-drop cell lysis, which both hybridize to the mRNA capture bead and attach to a common cell barcode during RT. Endogenous mRNAs and LMOs are separated by size selection following amplification, enabling pooled sequencing at user-defined ratios. Any cell type or nucleus from any species with an accessible plasma membrane can be barcoded using this method. Moreover, this approach involves minimal sample handling, preserving cell viability and endogenous gene expression patterns.

2) Microwell-based scRNA-seq methods. (i) CytoSeq. CytoSeq, a highly scalable scRNA-seq method, can simultaneously analyze a few thousand cells and can be easily scaled to 10,000 or 100,000 cells, with the detection of approximately 100 genes per cell (Fan et al., 2015a). This method employs a recursive Poisson strategy to adjust the number of cells in suspension, facilitating high-throughput cell settling in 1/10 of 100,000 microwells by gravity. Due to the nanoliter volumes used in the reactions, the cost of library preparation is exceptionally low. Additionally, CytoSeq is advantageous as it is not restricted to specific cell sizes and shapes, enabling the study of expression profiles in large and heterogeneous cell populations. This flexibility allows for the detection of rare cell types within a large background population.

(ii) Seq-Well. Seq-Well is a portable, cost-effective, user-friendly, and efficient scRNA-seq method designed for low-input samples (Gierahn et al., 2017). This approach utilizes a picowell array where barcoded mRNA capture beads and cells are loaded, with each well accommodating one cell and nearly one bead. After the cells settle into the wells by gravity, a semipermeable membrane seals the array, creating a unique environment for each well that allows buffer exchange but prevents the migration of macromolecules. Subsequently, cells are lysed, and the process of amplification and sequencing is carried out. With approximately 86,000 subnanoliter wells, Seq-Well plates enable the simultaneous analysis of transcriptomes in thousands of cells from diverse sources. This method is particularly well-suited for low-sample inputs, such as tissue pinches, fine-needle aspirates, and challenging-to-study cells like hepatocytes and granulocytes (Kumar, 2021). Notably, Seq-Well's implementation requires only a picowell array, a pipette, a polycarbonate membrane, an oven or heat source, a clamp, and a tube rotator to generate a stable cDNA product. This simplicity makes it adaptable to resource-limited environments such as clinic and remote locations (Aicher et al., 2019).

(iii) ICELL8. ICELL8 is a microwell-based microfluidic system that enhances throughput by incorporating a microchip containing 5,184 nanowells, allowing for the capture and processing of approximately 1,300 single cells (Goldstein et al., 2017). Each nanowell contains preprinted oligonucleotides, which include an oligo-(dT_{30}) primer, a well-specific barcode sequence, and a UMI. The process involves dispensing single-cell suspensions into the microchip nanowells using the multi-sample nanodispenser (MSND) and subsequently lysing the cells through freeze-thaw cycles. Following cell lysis, RT is carried out to synthesize cDNAs, employing the SCRB-seq method. Ultimately, cDNAs from hundreds of cells are pooled into a single tube for library construction. The method offers several advantages: i) the use of the MSND instrument for accurate dispensing of cells into nanowells, eliminating errors associated with manual pipetting; ii) incorporation of imaging software to identify nanowells containing viable single cells, ensuring that only wells with single cells are processed into sequencing libraries, resulting in a low cell multiplet rate (<3%); iii) the capability to load up to eight experimental conditions across one array using a multi-sample nanodispenser, enabling the simultaneous processing of 800–1,400 cells on one chip with 5,184 nanowells in a single experiment.

To enhance the capture rate of ICELL8 and enable the simultaneous processing of more than 5,000 cells for sequencing libraries, Shomroni et al. (2022) introduced a novel scRNA-seq method called CellenONE-ICELL8. This method combines the

ICELL8 processing instrument with the CellenONE isolation and sorting system. CellenONE relies on image-based single-cell isolation, enabling the selection of highly purified individual cells based on parameters such as cell morphology, size, or fluorescence markers before subsequent sample processing and sequencing. The integration of both instruments in CellenONE-ICELL8 significantly improves cell capture efficiency compared with the ICELL8 system alone, raising the number of captured cells from the typical 1,200 to 1,400 cells to over 3,300 cells. Furthermore, the utilization of full-length chemistry (SMART-seq technology) in this method can detect non-coding RNAs, especially lengthy intergenic non-coding RNAs, as well as intronic and intergenic sequences.

(4) Combinatorial indexing-based high-throughput scRNA-seq technologies

1) Sci-RNA-seq. Cao et al. (2017) developed the first combinatorial indexing method for high-throughput scRNA-seq, termed sci-RNA-seq. This innovative method facilitates the analysis of the transcriptomes of large numbers of single cells or nuclei, providing 3′ coverage and high-depth sequencing through the use of double UMI barcoding. The scalability of sci-RNA-seq allows for the generation of approximately 4×10^4 individual cell transcriptomes in a single experiment when employing indexing up to 576×960. This scalability enables the processing of more cells with sublinear cost scaling by incorporating additional rounds of indexing or/and using more barcoded RT and PCR primers. But it's crucial to remember that sci-RNA-seq is not without its drawbacks. These include laborious experimental procedures, the high expense of high-throughput transposition reactions, and a notable cell loss brought on by FACS sorting.

2) SPLiT-seq. SPLiT-seq represents another innovative combinatorial indexing method designed for scRNA-seq analysis, allowing the examination of fixed tissues preserved in 1.33% formaldehyde (Rosenberg et al., 2018). Unlike sci-RNA-seq, SPLiT-seq inserts second and third-round barcodes into cDNA by ligation. This approach offers a simpler, gentler, and more cost-effective workflow. The first-round barcodes of SPLiT-seq can act as sample identifiers, thus enabling highly multiplexed parallel sample processing. SPLiT-Seq is especially well suited for the analysis of fixed, difficult-to-completely disaggregate cells or nuclei produced from clinically relevant tissue samples. However, some limitations should be considered, including potential chemical modifications to mRNA caused by aldehyde-based cell fixation, suboptimal reaction efficiency of RT and ligation within fixed cells due to the intricate cross-linked intracellular environment, and potential degradation of RNA quality during the fixation process, leading to reduced detected gene levels.

Single-nuclei RNA-sequencing

ScRNA-seq is a potent tool for exploring cell types, functional processes, and dynamic states in intricate tissues. However, its application is limited when dealing with archived samples or tissues that cannot be easily dissociated, preventing the exploration of new cell types or crucial information related to immunity and disease. To address this difficulty, scientists have resorted to single-nuclei RNA sequencing (snRNA-seq) technologies, in which RNA sequencing experiments are performed using nuclei as proxies rather than entire cells. Several snRNA-seq methods have been developed to analyze RNA in single nuclei obtained from frozen, lightly fixed, or fresh tissues, including DroNc-Seq (Habib et al., 2017), Div-seq (Habib et al., 2016), snDrop-seq (Lake et al., 2019). These approaches extend the applicability of transcriptomic analyses to a broader range of sample types and conditions.

The use of snRNA-seq techniques has proven valuable in various research areas due to the high correlation observed between genes detected by snRNA-seq and traditional scRNA-seq methods (Fischer and Ayers, 2021). These techniques find application in diverse sample types, such as fresh tissues like the brain (Affinati et al., 2021), heart (Nicin et al., 2021), kidney (Barwinska et al., 2021; Muto et al., 2021), pancreas (Basile et al., 2021), muscle (Petrany et al., 2020) or adipose tissue (Sun et al., 2020b), archived tissues (Basile et al., 2021), plant cells (Conde et al., 2021). Despite these advantages, isolated nuclei are often more adhesive compared with isolated cell types. Therefore, precautions should be taken to prevent clumping, which can lead to inflated doublet rates. Furthermore, it's important to note that certain gene transcripts may exhibit enrichment differences between snRNA-seq and scRNA-seq datasets. For instance, long non-coding RNAs (lncRNAs) are enriched in snRNA-seq datasets, while mitochondrial transcripts, residing in the cytosol, are only present in scRNA-seq datasets (Fischer and Ayers, 2021). Researchers need to carefully evaluate whether such sublocalized genes are crucial for their specific research project before opting for individual nuclei for sequencing.

Computational methods for scRNA-seq data

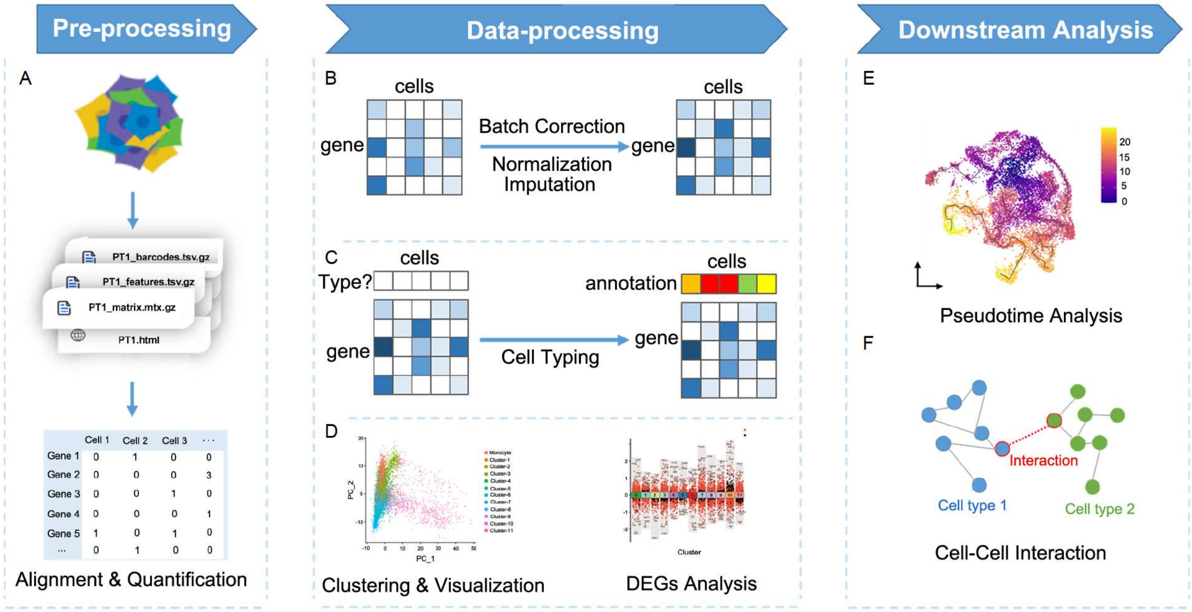
The original data format of scRNA-seq and most of the current scRNA-seq analysis processes are based on FASTQ files (or compressed format fq.gz). Illumina platform sequencing data generates BCL format files by default, which can be converted through CellRanger mkfastq. The analysis processes of scRNA-seq include data preprocessing, processing and extended downstream analysis (Figure 4), among which data preprocessing includes quality control, read alignment and expression quantification; data processing includes normalization, batch effect correction, imputation, feature selection (HVG selection), dimension reduction and clustering, cell typing annotation, differential expression analysis (DEGs), visualization; extended downstream analysis includes pseudotime, cell-cell interaction (CCI), pathway enrichment analysis, gene regulatory network (GRN) and other downstream analysis. On the whole, scRNA-seq analysis methods have mushroomed emerge in endlessly, there is no absolutely perfect method that applies to all scenes, it is important to obtain the biological information from analysis tools and the difficulty is selecting the most appropriate method. Here, we proposed to summarize common single-cell transcriptome analysis methods with their advantages and disadvantages as well as the scope of application to make suggestions.

Data pre-processing

The original sequencing data by filtering out low-quality reads and environmental interference were aligned and quantified with reference genomes. Consequently, the feature count matrix for each cell and auxiliary files recording other information were obtained, which were used for downstream data analysis (Figure 4A).

(1) Quality control

Low-quality sequencing data was inevitably produced due to



	Cell 1	Cell 2	Cell 3	···
Gene 1	0	1	0	0
Gene 2	0	0	0	3
Gene 3	0	0	1	0
Gene 4	0	0	0	1
Gene 5	1	0	1	0
...	0	1	0	0

Figure 4. Overview of the single cell analysis workflow. A, The cell-gene matrix count is formed based on sequencing data through single cell read alignment and quantification methods in pre-processing stage. B, High quality cell matrix used for analysis is obtained by processing the original gene expression matrix, make batch correction to remove batch effect, normalize to reduce biological differences and fill in genes that were missed in sequencing. C, Make cell types assignment based on prior references or not. D, Cells with similar transcriptome characteristics are grouped into one cell group called cluster and visualization of cells can be realized by the method of dimension reduction. Differential expression analysis (DEGs) checks the significance of the classification between groups. E, Pseudotime analysis can restore the dynamic process of cellular transcript change. F, Transcriptome regulatory relationships between cells can be inferred through cell-cell interaction analysis.

the sequencing instrument problem, artificial operation, cell spontaneous cases, or the existence of empty droplets, doublets, dead cells, etc. (Chen et al., 2019a; Hao et al., 2021b). Empty droplets usually appear when the droplet captures extracellular background transcripts instead of cells (Ilicic et al., 2016; Kolodziejczyk et al., 2015). A highly subjective method is to determine a UMI threshold according to the knee point and filter out cells with low UMI count. DropEst (Petukhov et al., 2018), EmptyDrops (Lun et al., 2019), and DIEM (Alvarez et al., 2020) were then used to enhance the filtering effect. DropletQC (Muskovic and Powell, 2021) quantifies the nuclear fraction score of unspliced pre-mRNA content. The choice of the MT gene threshold requires a comprehensive consideration of cell physiology factors, though it is a dead cell measurement (Subramanian et al., 2022). In recent years, deep-learning-based methods, such as neural-network-based EmptyNN (Yan et al., 2021), and deep-generation-models-based CellBender (Fleming et al., 2019), have also emerged and enabled the effective identification of the background transcripts in empty droplets.

Doublet is the case that two cells are contained in one single drop, which can be divided into homo-doublet and hetero-doublet based on the transcriptional distribution, both obeying Poisson statistics (Bloom, 2018). The vast majority of methods are based on gene expression calculations, using prior knowledge or deep learning to obtain the differences between unimodal and bimodal cells, and then train the classifier for screening, e.g., nearest-neighbor-based DoubletFinder (McGinnis et al., 2019a), Scrublet (Wolock et al., 2019); deconvolution-based DoubletDecon (DePasquale et al., 2019), variational-autoencoder-based Solo (Bernstein et al., 2020), and ensemble- algorithm-based Chord (Xiong et al., 2021a). Besides, Scds is another screening method relying on the co-expression-based doublet scoring and binary-classification-based doublet scoring strategy to achieve doublet separation over the scRNA-seq expression data (Bais and Kostka, 2020). A few methods use other features, such as the demuxlet which uses natural genetic variation information guidance experiments and filters computationally (Kang et al., 2018).

Reasonable quality control needs to comprehensively consider both technical and biological factors, which is also the main direction of the current research. A biological data-driven self-learning unsupervised quality control method called ddqc was recently proposed to determine specific thresholds of various GC metrics (Macnair and Robinson, 2023).

(2) Read alignment and expression quantification

The remaining high-quality cells after quality control require mapping these short reads to a specific reference genome for alignment to make the quantification of gene expression levels. RNA read alignment is usually divided into two steps: alignment of reads for indexing and mapping RNA splicing sequence, the former step was shared with DNA read alignment, solving the mismatch problem and setting up index references; the latter step is unique for RNA read alignment and provides connectivity information.

The early second-generation sequencing results were dozens of pair length base reads. Seed-to-extend methods (Buhler, 2001) (including MAQ (Li et al., 2008a), SOAP (Li et al., 2008b), CloudBurst (Schatz, 2009), ZOOM (Lin et al., 2008)), Burrows-Wheeler Transforming methods (Burrows and Wheeler, 1994) (including SOAP2 (Li et al., 2009), Bowtie (Langmead et al., 2009), BWA (Li and Durbin, 2009)), Needleman-Wunsch method (including Novocraft (Hercus, 2009)), and suffix-tree algorithm method (including MUMmer 2 (Delcher et al., 2002)) are effective tools for reads alignment of million level short-chain DNA sequencing. For Bowtie, an FM-index method dependent on Burrows-Wheeler Transforming is used, the result only reports one if the reads have multiple accurate matches, it greatly optimizes the running memory and alignment speed compared with MAQ (Ferragina and Manzini, 2001). BWA is another BWT-based alignment method, using the new SAM (Sequence Alignment/Map) format to output the alignment results. Based on MAQ and Bowtie two short-chain DNA alignment algorithms, Cole Trapnell proposed TopHat, the first RNA-seq alignment method for NGS data in 2009 using the 2-bit-per-base encoding to achieve efficient alignment of reads to splice sites in the mammalian genome without any prior knowledge of the splice sites (Trapnell et al., 2009).

The above methods decrease the alignment accuracy precipitously when the base pair length exceeds 50 bp (Gupta et al., 2018; Lebrigand et al., 2020). Two main categories are used in NGS single-cell sequencing analysis: Bowtie2-based methods and seed-strategy-based methods (Langmead and Salzberg, 2012). Bowtie2 is an upgrade to the Bowtie, retaining the FM-index dependent BWT algorithm core, which permits gapped alignment and uses single-instruction multiple-data (SIMD) to extend to the long sequencing alignment while increasing the running speed. Based on Bowtie2, Daehwan Kim propose TopHat2 (Kim et al., 2013) and HISAT (Kim et al., 2015) successively. The main methods for seed strategy are STAR (Dobin et al., 2013) and Subread (Liao et al., 2013). Based on the Maximal Mappable Prefix (MMP) ideas, STAR adopted the strategy of sequential retrieval to set the longest partial reads matching with the reference as seed 1, the rest read will continue to match, in turn is called from seed 2 to seed *n*. It is worth noting that Rsubread implements the first read alignment and gene quantification process based entirely on the R language platform (Liao et al., 2019).

The gene expression quantification can be divided into pseudo-alignment quantification and read-alignment-based quantification. Pseudo-alignment refers to the alignment of all reads mapping to the reference genome without the rigorous two-step method described above, including the selected k-mers alignment method (Sailfish (Patro et al., 2014), Kallisto (Bray et al., 2016), Salmon (Patro et al., 2017), RapMap (Srivastava et al., 2016)) and Barcode-UMI-Set (BUS) alignment method BUStools (Melsted et al., 2019). Kallisto-BUStools is the latest workflow that uses the BUS file format for initial data pre-processing, like the BUStools, the pseudo-alignment result and quantification counts are saved in the BUS files (Melsted et al., 2021). On the other hand, read-alignment-based methods rely on the result of the RNA read alignment method to quantify the gene. CellRanger is the official open-source data pre-processing software designated by 10x Genomic company to replace Longranger (Zheng et al., 2017). STARsolo is a tool to replace the mapping/quantification function for Cellranger, it can realize the analysis of multi-platform sequencing data and the quantification of transcriptome features beyond gene expression (Kaminow et al., 2021). Other read-alignment-based gene expression quantification like UMI-tools (Smith et al., 2017), zUMIs (Parekh et al., 2018), Alevin-fry (He et al., 2022), DropEst (Petukhov et al., 2018), RainDrop (Niebler et al., 2020), baredSC (Lopez-Delisle and Delisle, 2022), BCseq (Chen and Zheng, 2018) use various quality filter and barcode/UMI treatment strategy to improve the performance of CellRanger to some extent.

Both CellRanger and STARsolo have a good running speed while processing all kinds of single-cell transcriptome datasets, including 10x Chromium, with extremely high accuracy. However, under the premise of obtaining almost identical results, the latter increased the running speed by at least five times compared with the former, which also verifies the purpose of using STARsolo to replace CellRanger by Alexander Dobin et al. (Brüning et al., 2022; Chen et al., 2021a; You et al., 2021).

Data processing

After making the necessary adjustments to the expression matrix (Normalization, Batch Effect Correction, Imputation), biological information can be fully mined from single-cell transcriptomic data for analysis. Seurat and Scanpy perform the modular and scalable processing of the above processes based on R and Python respectively and are currently the mainstream analysis process of single-cell transcriptomic data (Satija et al., 2015; Wolf et al., 2018). The conventional analysis process and expected processing results can be found in total analytical framework (Figure 4B–D).

(1) Normalization

In sequencing processing, due to technical reasons or biological differences between cells themselves, may cause library size differences in the same samples (between cells) or between different samples (Marinov et al., 2014). The infinite number methods process according to the library size, according to the specific principle, they can be roughly divided into global-scaling-based normalization, spike-in normalization, and other data transformation model normalization.

The global scaling method was originally developed for the bulk RNA analysis by scaling the global data with a specific scaling factor (Finak et al., 2015). Counts per ten-thousand (CPT) transformation and count per million (CPM) transformation are common linear scaling methods, without considering the spike-in count, they scale all per UMI/total UMI count equidistantly. Other normalization methods include reads per million (RPM) (Mortazavi et al., 2008), trimmed mean of M values (TMM), DESeq (Robinson and Oshlack, 2010), upper-quartile scaling (Bullard et al., 2010), FPKM (Trapnell et al., 2010), RPKM (Tu et al., 2012) have better stability for extreme values compared with linear scaling, therefore have a wider range of applications like the RPKM/FPKM. However, when using such methods alone for the normalization of the single-cell transcriptome, because of the sparsity and inflated false positives of the data, the effect is not acceptable (Evans et al., 2018). Improvements are often needed when combined with specific methods. SCnorm uses a quantile regression method to evaluate the scale factors among different sequencing depth dependence cell groups (Bacher et al., 2017). Based on the assumption of negative binomial (NB) distribution for gene original count with the true count, bayNorm uses an integrated Bayesian model for scRNA-seq data normalization (Tang et al., 2020).

Spike-in normalization method can be regarded as another expansion of the global scaling method, as the scaling factor is calculated from spike-in genes. It is noted that adding information about RNA spike-ins to other methods can also improve the effect of standardization like SCnorm. GRM is a method based on the gamma distribution of the spike-in ERCC molecule concentration in which ERCC is a calibration material commonly used in sequencing (Ding et al., 2015). BASiCS is an automated Bayesian normalization method applying the Poisson hierarchical model to spike-in (technical) genes for cell specific normalization constants inference (Vallejos et al., 2015).

In the above methods, genes were scaled under the assumption of constant intracellular RNA number which can be deceitful, so other transformation models adopted different strategies. Due to the problem of zero-inflated in the single-cell transcriptomic data, some of the models were designed for this purpose, for example, the relative log expression (RLE) method ascend (Senabouth et al., 2019) and the NB-based models like Dino (Brown et al., 2021), scTransform (Hafemeister and Satija, 2019). Other transformation model normalization method like MUREN uses the least trimmed squares (LTS) regression algorithm (Feng and Li, 2021); Sanity uses the log transcription quotients (LTQs) inferred from UMI count as the input of a Bayesian framework to avoid the Poisson fluctuations, as the LTQs vector changes estimate the gene expression values (Breda et al., 2021); PsiNorm is an unsupervised Pareto distribution scale parameter based method to make improvements to the normalization efficiency and accuracy (Borella et al., 2021). Charles Wang made a comparison of total of eight normalization methods, including the sctransform, TMM, and DESeq, wherein the sctransform and logCPM (the built-in processing method of Seurat) are least affected by data and are most stable over variable datasets (Chen et al., 2021a).

(2) Batch effect correction

Due to the experimental design, sequencing platform, sequencing time, and personnel operation process, different single-cell transcriptome sequencing data will differ significantly in mRNA capture efficiency, and sequencing depth to generate the batch effect among samples (Chen et al., 2019a; Hwang et al., 2018; Tung et al., 2017). Theoretically, the technical variation can be eliminated through experimental strategies, but due to the objective limitations of the experimental process and sequencing instrument errors, the batch effect will inevitably be introduced more or less. Correction with computational methods is necessary to solve imperfect experimental design, usually used methods can be divided into mutual nearest neighbor (MNN) method, latent-space-based method, graph-based method, DL method, and other methods.

MNN first identifies the most similar cells of the same cell type between different batches, and then uses these cells for batch effect correction, including batchelor (Haghverdi et al., 2018), Scanorama (Hie et al., 2019), Canek (Loza et al., 2022). Another class of methods using MNN is based on latent space after dimension reduction, like Seurat (Satija et al., 2015), BEER (Zhang et al., 2019b), SMNN (Yang et al., 2021a), iSMNN (Yang et al., 2021b). For example, Seurat uses the MNN pairs (called "anchors") in the canonical correlation analysis (CCA) latent space to match similar cells while BEER uses the principal component analysis (PCA) sub-spaces for screening poor similar subgroups. SMNN and iSMNN adopt supervised machine learning and iterative supervised machine learning separately to refine MN-pairs trained from information on pre-correction cell clustering or iterative cell clustering.

Latent space-based methods refer to the method of performing batch effect correction in the hidden space or embedding after dimensionality reduction, besides the MNN cluster-based strategy, they also contain PCA-related space method harmony (Korsunsky et al., 2019), FIRM (Ming et al., 2022), Monet (Wagner, 2020); t-distributed stochastic neighbor embedding (t-SNE) space method sc_tSNE (Aliverti et al., 2020) and ZINB-WaVE (Gao et al., 2019). Harmony is widely used to remove batch effects between samples, sequenced cells are fed into a single common embedding using the PCA method, then circulated iteratively between maximum diversity clustering and linear batch correction until a specific correction factor is assigned to each cell which can be used for subsequent batch effect removal. Sc_tSNE method introduces gradient descent's algorithm for the traditional t-SNE algorithm optimization, and a linear correction is used subsequently (Aliverti et al., 2021). ZINB-WaVE was originally designed to perform gene extraction in single-cell data, Risso et al. (2018) extended this method to mini-batch optimization.

Graph-based methods use cell-gene expression matrix to transform the digital information into the spatial constructed graph, where nodes represent different types of batches and weights of edges are based on different calculation methods. BBKNN uses the k-nearest neighbor cells to construct a graph (KNN graph), and the batch effect correction is implemented by merging the graph of individual cells across different data sets using the uniform manifold approximation and projection (UMAP) method, which is also the default method in Scanpy workflow (Polański et al., 2020; Wolf et al., 2018). Bo Wang proposed "ghost cell" (k-means algorithm cluster center by default) in OCAT to make a bipartite graph for cell connection (Wang et al., 2022a).

In recent years, the rapid development of deep learning methods has also provided new ideas for batch-effect correction, realizing efficient and large-throughput data processing, like INSCT (Simon et al., 2021) (triplet neural networks), CLEAR (Han et al., 2022) (self-supervised contrastive learning), BERMUDA (Wang et al., 2019e) (transfer learning), iMAP (Wang et al., 2021a) (VAE-GAN), ResPAN (Wang et al., 2022e) (Wasserstein GAN). Some new methods are shown to have better results in batch effect correction; for example, based on biological prior knowledge from the annotated datasets learned by SciBet, SSBER can remove batch effect in a large RNA sequencing data set (Zhang and Wang, 2021). It is suggested that before the integration of single-cell transcriptomic data, multiple methods should be tested first based on the actual situation of the data, and then the most appropriate batch effect removal method should be selected. For example, Jinmiao Chen group and Charles Wang group conducted a benchmark in 2020 and 2021 for most of the first three methods mentioned in this review 2.2, respectively, they proved that Harmony and Seurat V3 achieve good batch effect correction results in most cases, which is in line with the fact that these two methods are still widely used today, but there is still a lack of good indicators for deep learning methods (Chen et al., 2021a; Tran et al., 2020).

(3) Imputation

Large numbers of 0 values will be introduced during sequencing (probably over >90% zero values in the high-throughput large-scale 10x Genomic sequencing data) (Stegle et al., 2015; Talwar et al., 2018). It will interfere with the analysis of downstream biological differences, and therefore, the missing data values in the original gene expression matrix must be conducted imputation, while effectively distinguishing between the technical noise null value and the biological null value.

The gene/cell separated method is mostly applied in the early imputation which considers separately the cell similarity (MAGIC (van Dijk et al., 2018), Sclmpute (Li and Li, 2018), VIPER (Chen and Zhou, 2018), RESCUE (Tracy et al., 2019), scRMD (Chen et al., 2020a), scRoc (Ran et al., 2020)) or gene-to-gene relationship (SAVER (Huang et al., 2018a), SAVER-X (Wang et al., 2019a), G253 (Wu et al., 2021e), DCA (Eraslan et al., 2019), DeepImpute (Arisdakessian et al., 2019)). Overall, these methods lack consideration for the data set as a whole and can easily lead to excessive imputation or introducing errors (Zhang et al., 2019d). The comprehensive method comprehensively considers the connection of cells and genes with each other: CMF-Impute and netNMF-sc are the earliest methods to effectively utilize the association between cells and genes for imputation (Elyanow et al., 2020; Xu et al., 2020a). scIGANs processes the gene expression matrix by a specific GAN model, using generated cells training GANs model to impute the dropout (Xu et al., 2020b). In recent years, new methods are still being proposed to better solve the impact of technical noise on the data outside of dropout and to achieve a better differentiation of biological zero values. AutoClass (Li et al., 2022c) achieves processing without supervision, while the ALRA method mainly aims at the biological zero values (Linderman et al., 2022). scMOO makes a fundamental change to use the latent structure of the data to learn deep associations in cell similarity vertical structure and total low-rank structure, thus achieving a better imputation effect than a single gene expression matrix as an input, but it also puts more memory requirements (Jin et al., 2022a). sc-PHENIX utilizes the PCA-UMAP initialization method to achieve a nonlinear interpolation of the gene expression (Padron-Manrique et al., 2022). At present, there is no definite conclusion on which imputation can achieve the best effect. Due to the data set itself, the purpose of downstream analysis will have different choices, but there is no doubt that the best imputation method will be able to effectively distinguish between technical noise zero value and biological zero value with lower calculation requirements (Jiang et al., 2022a; Wen et al., 2022).

(4) Feature selection

In order to reduce data dimension to enhance computational analysis efficiency, reduce technical noise interference and the risk of model over-fitting, we often choose feature selection strategy to select highly variable genes in different cells, instead whole data set genes as subsequent analysis, such as cluster (Brennecke et al., 2013; Jackson and Vogel, 2022; Svensson et al., 2017).

In bulk RNA-seq analysis, methods for finding differential genes generally include fold change (FC) based method, statistical-tests-based method, and FC-statistical tests method, the last one has the best screening results and credibility obviously (Chung and Storey, 2015).

Early single-cell feature selection approaches lack a correction between the mean expression and variances resulting in an excessive proportion of the highly expressed gene in the results (Brennecke et al., 2013). EDGE uses an ensemble learning method of massive weak learners to learn inter-cellular similarity probabilities, significant contributions based on information entropy are extracted as the highly variable genes (Sun et al., 2020c). Similarly, SAIC achieves an optimal cell cluster separation based on Iterative Clustering final output (Yang et al., 2017). Recently, some new feature extraction strategies have been proposed and proved for their stability and effectiveness, but the authoritative verification of the performance between them is still lacking: including gene expression distribution matrix-based method SCMER (Liang et al., 2021b), RgCop (Lall et al., 2021), scPNMF (Song et al., 2021a), SIEVE (Zhang et al., 2021e); entropy-based method IEntropy (Li et al., 2022g), infohet (Casey et al., 2023); comprehensively considered cluster-based method Triku (Ascensión et al., 2022), FEAST (Su et al., 2021), etc. Since the vast majority of the above methods ignore the integrity of gene dependency, comprehensive methods are proposed, such as Triku using a k-nearest neighbor graph method to comprehensively explore and classify gene expression patterns, achieve screening for more biologically meaningful feature genes without bias; FEAST ranks the feature by f-test on consensus cluster and extracts HVG based on feature evaluation algorithm (Wang et al., 2022c).

A few other methods use features other than highly variable genes to represent the data set, for example, the scVEGs and scSensitiveGeneDefine methods, using high coefficients of variation (CV) as a feature extraction; the BASiCS method utilizes the information of spike-in genes (Chen et al., 2016b; Chen et al., 2021b). Overall, based on the perspective of accuracy, and biological interpretability, the main goal of the current feature selections is to effectively extract the HVG for an effective downstream analysis of high-dimensional transcriptome data.

(5) Dimension reduction

As the single-cell transcriptome typically includes tens of thousands or more genes, it is not conducive to extracting effective information directly. In the actual analysis process, we usually need to reduce the dimensionality of the original sequencing data. Besides processing the high-dimensional single-cell transcriptome sequencing data using the feature selection method mentioned above, dimension reduction is also an effective method, which can be classified as linear dimension reduction (latent Dirichlet allocation (LDA)-based method, PCA-based method) and nonlinear dimension reduction (t-SNE-based method, UMAP-based method) according to the dimension reduction strategy (Andrews and Hemberg, 2018; Becht et al., 2019; Laurens and Hinton, 2008; Peres-Neto et al., 2005).

In linear dimension reduction, LDA and PCA are two widely used algorithms, LDA distinguishes features from the aspect of the largest classification, while PCA orthogonally extracts the main components from the angle of the largest variance. Despite the improved algorithms of JPCDA, and LDA-PLS, the dimension reduction effect of the LDA model in single-cell transcriptome data is still not optimal (Tang et al., 2014; Zhao et al., 2020). PCA is another linear transformation, Seurat usually determines the amount of the PCs numbers according to the inflection point of the standard deviation-PC diagram or the proportion test result *P*-value (the ScoreJackStraw function) of the PCs. Other variants PCA based dimension reduction methods include the pcaReduce (Žurauskienė and Yau, 2016), GLM-PCA (Townes et al., 2019), RPCA (Gogolewski et al., 2019), tRPCA (Candès et al., 2011), scPCA (Boileau et al., 2020), PCAone (Li et al., 2022l). GLM-PCA extends the traditional PCA analysis to non-normal distributions, directly handles the original matrix by introducing an exponential family likelihoods strategy to make the PCA free from normalization restriction, and then ranks and extracts the gene implementation using bias (Collins et al., 2002). ScPCA uses contrastive PCA and sparse PCA to remove the technical noise and the data, respectively, which further increases the stability of the PCA (Abid et al., 2018; Zou et al., 2006). As most scRNA-seq datasets are difficult to effectively represent by simple linear dimension reduction, one first strategy for solving this is based on a rapid PCA analysis approach. PCAone proposes a new fast randomized singular value decomposition (RSVD) strategy, which completes the analysis of 1.3 million mice brain cells single-cell data within 35 min (Li et al., 2022l).

Nonlinear dimensionality reduction is another solution, like non-parametric dimension reduction methods t-SNE and UMAP, both need to set the hyperparameters of clustering in advance; and in classification effect, the former tends to discrete the formation of cells in the data. In the case of the reasonable use of parameters for specification, there is no significant difference between UMAP and t-SNE, which means after using the same method of information initialization, they can produce approximate analytical efficiency while preserving the global structure of the data set (Do and Canzar, 2021; Kobak and Linderman, 2021). Modified methods for t-SNE include net-SNE, qSNE, FIt-SNE, and Joint t-SNE (Cho et al., 2018a; Linderman et al., 2019; Wang et al., 2022b), while the improvement of the UMAP mainly comes from the self-improvement of the method by the Leland McInnes' group (McInnes et al., 2018). To better visualize the dimension reduction results of t-SNE or UMAP, Hyunghoon Cho proposes the den-SNE/densMAP approaches for the transcriptome variability information based on local radius dependent optimization to iteratively optimize the function of conventional t-SNE/UMAP; Stefan Canzar proposes the j-SNE/j-UMAP to improve the multimodal omics data joint visualization results to reduce misleading of visualization (Do and Canzar, 2021; Narayan et al., 2021).

(6) Clustering

In the analysis of single-cell transcriptome data, clustering is performed to divide the cells into subgroups and we are therefore able to characterize the different cell types in multicellular organisms which helps us to accurately analyze different tissues or developmental processes from the perspective of cell heterogeneity. The actual effect of clustering can be affected by the pre-data processing steps like bath effect normalization, imputation, dimension reduction, etc.

After feature gene selection and dimension reduction, the vast majority of single cells are clustered based on distance. The concept of the K-means clustering algorithm was used for applications like SCUBA, SC3, and RaceID (Grün et al., 2015; Kiselev et al., 2017; Macqueen, 1967; Marco et al., 2014). On parameter selection improvement, SAIC iteratively optimizes multiple initial centers K and P-value by the Davies-Bouldin index to obtain the optimal solution; LAK applies a parameter selection algorithm to datasets for automatic selection of parameters (Davies and Bouldin, 1979; Hua et al., 2020; Yang et al., 2017). In the operation of ultra-high-dimensional data, LAK adds the Lasso penalty to make standardization and mbkmeans achieves rapid clustering at the million-cell level using mini-batch k-means (Hicks et al., 2021). SMSC applies a spectral clustering method to improve the clustering performance but loses some accuracy for ultra-high-dimensional data (Qi et al., 2021). Another broad class of widely used distance clustering methods depends on sharing the nearest neighbor graphs structure and graph cluster, among the most widely used are Louvain or Leiden (Blondel et al., 2008; Xu and Su, 2015). The identification of rare cells needs to be improved in combination with specific methods, such as dropClust using the locality sensitive hashing workflow for screening the nearest neighbor followed by Louvain cluster, it uses the exponential decay function to retain more transcriptomic features of the rare cells (Sinha et al., 2018). Other distance-based clustering methods use different algorithm cores: SIMLR uses a Gaussian kernels learning model to construct kernel matrix for the potential C cell populations in the datasets while Conos proposes a joint mutual nearest-neighbor (mNN) graph cluster to achieve integrative analysis of multiple different single-cell transcriptome samples (Barkas et al., 2019; Wang et al., 2017a). Density-based clustering uses the closeness of the sample distribution for the cluster, DBSCAN is the most classical algorithm (Ester et al., 1996; Fukunaga and Hostetler, 1975). For single-cell sequencing, densityCut and FlowGrid are designed based on this principle (Ding et al., 2016; Fang and Ho, 2021). Hierarchical clustering is a bottom-up clustering method that repeatedly calculates cell-to-cell similarity for classification until the preset number of clusters is completed without advance through unsupervised learning (Guo et al., 2015). Subsequently, the RCA cluster uses a conventional hierarchical clustering method to cluster the cells mapped to global reference panel; HGC constructs a hierarchical tree on the SNN graph (Li et al., 2017; Zou et al., 2021). To solve the defects that the conventional hierarchical clustering method hardly clusters a certain group of cells and only allows the same set of signature genes for clustering, K2Taxonomer uses the constrained k-means algorithm to expand to sample groups, integration calculations are performed recursively based on multiple genes sets to capture subgroups ("taxonomy-like cells") under various resolutions (Reed and Monti, 2021). Mrtree applies hierarchical clustering's strategy to multiple partitions of flat cluster and constructed a multi-resolution reconciled tree to use as cell clustering (Peng et al., 2021a). Recently, Zelig and Kaplan (2020) propose a KMD clustering method, which eliminates the hyperparameter K while clustering through an average linkage hierarchical clustering model, greatly reducing judgment errors caused by subjectivity.

The deep learning cluster method is a combination of the machine learning method and the above single-cell transcriptome clustering strategy, which can achieve more efficient clustering results in the form of unsupervised, supervised, or semi-supervised. These methods tend to learn a nonlinear transformation, obtaining the best low-dimensional representation by mapping the original high-dimensional data into a smaller latent space. Overall, this approach avoids the impact of traditional clustering methods on the choice of pre-cluster data processing methods. Unsupervised clustering methods include ADClust, DESC, SAUCIE, and VAE-SNE, they usually do not require the parameters such as a preset number of clusters to complete the analytical processing of the data set in the way of autonomous learning (Amodio et al., 2019; Graving and Couzin, 2020; Li et al., 2020c; Zeng et al., 2022c). Although the unsupervised clustering method avoids parameters such as manual input cluster number and extends to ultra-high-dimensional cell clustering, sometimes using high-quality annotated data sets or other prior knowledge auxiliary constraints for supervised or semi-supervised clustering can achieve more accurate cell type classification and improve clustering performance (Bai et al., 2021). Transfer learning based ItClust, mutual supervised ZINB auto-encoder and graph neural network (GNN)-based scDSC, soft K-means convolutional auto-encoder based ScCAEs, Cramer-World distance max-mean penalty Gaussian mixture auto-encoder based SeGMA, time series clustering network based STCN are all widely used supervised clustering (Gan et al., 2022; Hu et al., 2022a; Hu et al., 2020a; Ma et al., 2021b; Smieja et al., 2021). Furthermore, Zhang group (Yang et al., 2023b) have utilized hierarchical GAN to design another widespread DL method IMDGC for single-cell transcriptome data analysis to construct cellular embedding cluster in a generated manner.

For the special cases in the clustering, targeted purposes clustering methods are designed as follows: GiniClust (Jiang et al., 2016) (updated to GiniClust 3 (Dong and Yuan, 2020)), MicroCellClust (Gerniers et al., 2021) for rare cell subpopulations clustering; EDClust (Wei et al., 2022), ENCORE (Song et al., 2021b) and MLG (Lu et al., 2021) for noise reduction and batch effect removal; ClonoCluster (clonal origin information) (Richman et al., 2023), IsoCell (alternative splicing information) (Liu et al., 2023) clustering with additional information. Wu and Yang evaluated the cluster methods from the perspective of the effect of feature selection on cluster, they proved that more representative feature selection enhances the level of cell clustering, methods based on "cluster similarity" (most distance-based clustering methods mentioned in our review) generally have a wide range of high clustering type performance; however, high accuracy and high running speed need targeted selection according to the actual data set (Su et al., 2021; Yu et al., 2022). Double dipping presents a significant issue wherein the same expression data are used both in cell clustering and differential expression genes, resulting in an excessively high false discovery rate (FDR) of DE genes when the cell cluster is incorrect. For example, if only one specific cell cluster is present, no gene should be considered as differential genes. To address this problem systematically, ClusterDE adopts a cluster contrast learning strategy for post-clustering DE testing. It demonstrates better FDR control across different threshold ranges compared with the truncated normal (TN) test and the Countsplit method (Song et al., 2023a).

(7) Cell typing annotation

Cell typing annotation refers to the usage of specific information to annotate cells or cell subsets in single-cell sequencing dataset, which is the basis for subsequent biological analysis. The most commonly used strategy is unsupervised clustering of cells followed by annotation based on the marker genes such as scCATCH and SCSA (Cao et al., 2020b; Shao et al., 2020a). However, it is difficult to treat complex high-dimensional datasets (Franzén et al., 2019; Luecken and Theis, 2019; Zhang et al., 2019c). Currently, multiple methods to automatically cell typing have been developed and can be roughly divided into two categories, i.e., reference-dependent and reference-free annotation methods.

Reference-dependent annotation methods require users to provide pre-annotated high-quality single-cell transcriptome datasets or prior knowledge from the PanglaoDB database, ScType database, etc. for alignment (Ianevski et al., 2022). According to the different principles of the method, it can be divided into hierarchy-tree-based methods (CHETAH (de Kanter et al., 2019), Garnett (Pliner et al., 2019), HieRFIT (Kaymaz et al., 2021), scHPL (Michielsen et al., 2021), scMRMA (Li et al., 2022e)), similarity-based methods (SingleR (Aran et al., 2019), scmap (Kiselev et al., 2018), deCS (Pei et al., 2023), scID (Boufea et al., 2020), scMatch (Hou et al., 2019), Symphony (Kang et al., 2021)), signature-gene-based methods (Cellassign (Zhang et al., 2019a), Cell-ID (Cortal et al., 2021), scMAGIC (Zhang et al., 2022g), SciBet (Li et al., 2020b)) and other DL methods. As an early method, ACTINN is a deep learning approach using a 3 hidden layers neural network for annotation classification (Ma and Pellegrini, 2020). SCPred then proposes a method using machine-learning probability-based prediction based on the unbiased feature selection from embeddings (Alquicira-Hernandez et al., 2019). Other methods such as Seurat project query cells in PCA space and train cell typing annotation through weighted vote classifier; scSorter adopts a Gaussian mixture model and GraphCS uses a virtual adversarial training (VAT) loss modified GNN to expand to multi-species, large-scale datasets of cellular annotation (Guo and Li, 2021; Zeng et al., 2022a).

Non-reference annotation approach uses a pre-trained deep learning model and can directly perform cell classification using the query dataset as input alone. scDeepSort uses single-cell atlas from human cell landscape (HCL) and mouse cell atlas (MCA) database as the input for the pre-trained weighted GNN models, which is suitable for human and mouse cell annotation with good results (Han et al., 2018b; Han et al., 2020; Shao et al., 2021b). Similarly, Pollock is a pretrained human cancer reference VAE model to classify the multimodal cells in the cancer environment (Storrs et al., 2022). Although it is more convenient to use, it is difficult to achieve a better cellular annotation effect for significantly different query datasets, and it is also difficult to expand the application due to the accuracy and the number of pre-training reference datasets. There are also some other cell annotation tools for targeted field research, for example, DevKidCC (Wilson et al., 2022) for human kidney cell annotation, ikarus (Dohmen et al., 2021) for the identification of cancer and normal cells. Overall, the performance of the non-reference annotation approach is restricted by the coverage and accuracy of pretrained reference datasets.

Currently, to improve cell annotation tools to uniformly assign cell types across large platforms and multicell patterns is the mainstream of cellular annotation research directions, the latest Cellar and ELeFHAnt methods have made some attempts in this regard and achieved initial results (Hasanaj et al., 2022; Thorner et al., 2021). Overall, similarity-based annotation methods are computationally intensive, when applied to very large query and reference data sets, they often make a trade-off between accuracy and speed, it is therefore generally only suitable for cell classification in smaller datasets; for larger-scale datasets, it is recommended to use F-test feature selection or MLP classifier (Hu et al., 2020a; Huang and Zhang, 2021; Ma et al., 2021c). Moreover, the method of semi-supervised transfer learning, such as Itclust, has good results in discovering new cell subtypes. In recent years, new methods based on the above reference annotation method classification have been continuously improved, and deep learning models such as VAE have also been applied in this field.

(8) Differential expression analysis (DEGs)

Statistical tests are commonly used in the differential gene analysis of Bulk RNA-seq, similar to the section 2.4 HVG Selection algorithm: *P*-values and log-fold changes are usually used as important parameters. Statistical tests include *t*-test (two sample based), Wilcoxon test, Kolmogorov Smirnov test (KS-test), and Kruskal-Wallis test (KW-test), some of which are also widely used in the test of single-cell transcriptome DEGs. Based on this, the corresponding detection tools are developed: limma (Ritchie et al., 2015), edgeR (Robinson et al., 2010), and DESeq2 (Love et al., 2014). Both the limma and edgeR algorithms are proposed by Smyth GK, the former is based on a normal or approximate normal distribution model while the latter is based on an overdispersed Poisson distribution model. DESeq2 is based on the NB distribution model for hypothesis testing and uses the empirical Bayes procedure for DEGs. Currently, limma has large errors in RNA count analysis due to specific distribution model assumptions, although both edgeR and DESeq2 utilize the Bayes model to normalize over-dispersion, the latter has better analysis results as promoting the screening of CPM threshold through the average value of data set reads and outlier detection.

Single-cell transcriptome DEGs can be roughly divided into early parametric tests on zero-value, non-parametric tests, and other methods according to time and analytical methods. Since there are vast zero numbers in the scRNA-seq data, most of the early methods are based on this observation to make parameter tests, such as Monocle (Trapnell et al., 2014), SCDE (Kharchenko et al., 2014), MAST (Finak et al., 2015), scDD (Korthauer et al., 2016), D3E (Delmans and Hemberg, 2016), TASC (Jia et al., 2017), DEsingle (Miao et al., 2018), and HIPPO (Kim et al., 2020b). The evaluation of some methods above shows that although they generally achieve good results in the analysis of single-cell datasets, there is no significant performance improvement over the DEA method for bulk data (Soneson and Robinson, 2018). It is possible that the best distribution model is not used for different datasets, and thus one alternative solution is to consider non-parametric DEA methods.

Non-parametric test or distribution-free test does not need to make prior assumptions about the data distribution form and it is therefore applicable to the analysis of multiple datasets, common methods are Swish (Zhu et al., 2019a), IDEAS (Zhang et al., 2022d), ccdf (Gauthier et al., 2021), distinct (Tiberi et al., 2022). Swish evaluates transcript level by Salmon Gibbs and then the FC value is calculated by the Mann-Whitney Wilcoxon test. IDEAS is a pseudo F-statistic test using Jensen-Shannon divergence (JSD) or Wasserstein distance (Was) for gene different expression measurement, the *P* value is generated by PERMANOVA based distance tester kernel based regression. Ccdf is a conditional independence test relying on the conditional cumulative distribution function, the DEGs is predicted by a multiple regressions model. Distinct proposes a hierarchical non-parametric permutation method, the total distance of empirical cumulative distribution function (ECDF) is used for DEGs identification. Alternative methods include deep learning strategies MRFscRNAseq (Li et al., 2021a), pseudotime inference based PseudotimeDE (Song and Li, 2021), non pre-cluster based singleCellHaystack (Vandenbon and Diez, 2020), multiple scores based MarcoPolo (Kim et al., 2022). It is suggested that different single-cell transcriptomic datasets should employ data-specific DE genes detection strategies for optimal DEGs analysis, based on the scCODE workflow, the most optimized DEGs method can be found using indicators involving CDO (DE genes order) and AUCC (area under concordance curve) (Zou et al., 2022). In addition, the research method will have a specific research orientation under different research backgrounds, for example, in dose-response studies after administration DEGs analysis, LRT linear test, and Bayesian multiple group test have better results than other methods (Nault et al., 2022).

(9) Visualization

Single-cell transcriptome data analysis visualization refers to the visual presentation of the above analysis results in the form of graphs, ggplot2 is the most extensive R visualization tool and is commonly used in R to greatly enhance drawing power (Wickham, 2009). ARL is another R package that specifically displays marker gene Association Plots and can display its features in each cluster (Gralinska et al., 2022). Also, there are other specific packages for marker gene visualization like Complex Heatmap that are not described in detail here. HVG visualization is usually presented in the form of volcano plot, by default, the left and right part of the graph are the underrepresented and overrepresented genes, respectively, while the middle is the constant gene. Enhanced Volcano is a specialized R package used for drawing a volcano graph, and ggplot2 can also be used to achieve better results by default. Cluster visualizations are often presented in PCA plot, t-SNE plot and UMAP plot, but it is noteworthy that the results of visualization are very deceptive, since some small cell subpopulations may represent a large number of cells shown in the UMAP figure. Improved methods like den-SNE/densMAP and j-SNE/j-UMAP have been proposed to solve these problems (Macqueen, 1967; Marco et al., 2014). Furthermore, FastProject can output a 2D display of the annotated cluster and PieParty can draw color maps for each gene in the cluster 2D graph (DeTomaso and Yosef, 2016; Kurtenbach et al., 2021).

Meanwhile, the interactive visualization of single-cell transcriptome data is currently a hot field; software such as Single Cell Explorer can achieve interactive visualization to certain extent, but it is still necessary to increase the interaction freedom to provide a more comprehensive 3D presentation of single-cell transcriptome data (Cakir et al., 2020; Feng et al., 2019). To this end, CellexalVR uses VR theory for interaction visualization; CellView is a Web-based tool, including the Explore tab, Co-expression tab, Subcluster-analysis tab modules for different uses; Cellxgene VIP is a cellxgene framework-based plugin and extends to the interactive visualization of ST data based on combination of multiple modules (Bolisetty et al., 2017; Legetth et al., 2021; Li et al., 2022f).

(10) Single-cell simulators

As single-cell transcriptome methods continue to expand, the pressing challenge lies in the benchmarking, with the key issue being the requirement for stable and reliable data, as direct sequencing of single-cell transcriptome may lack ground truth. The realistic single-cell simulator data provided a known truth for benchmarking, allowing training with real data while matching the characteristics of actual data. Additionally, simulated data provide greater flexibility than real data, enabling analysts to adjust parameters like dropout rate based on specific testing methodologies.

Splatter is a two-step simulator framework that initially simulates estimated parameters from real data and then incorporates additional parameters from users (Zappia et al., 2017). Its six pre-designed pipeline module interfaces ensure the repeatability of data generation. Recent updates have focused on specialization and generalization. In the specialization domain, splaPop generates population-scale data with genetic effects (quantitative trait loci), while dyngen simulates dynamic cellular processes like developmental trajectories (Azodi et al., 2021; Cannoodt et al., 2021). In the generalization field, Li's team introduced the six concepts of an ideal simulator including authenticity, preservation of genes, capture of gene correlations, robustness, parameter tunability, and efficiency (Song et al., 2023b; Sun et al., 2021). Subsequently, scDesign2 is proposed to meet all six properties (Sun et al., 2021), followed by scDesign3, addressing the gap in single-cell omics statistical simulation (Song et al., 2023b). The increased accuracy and transparency of the simulator enhance benchmarking between different single-cell data processing methods, guiding the selection of the most appropriate approach for specific data and licensing needs.

Extended downstream analysis

(1) Pseudotime

In order to more truly restore the real process in the organism, integration of multiple transcriptome data using pseudo-timing analysis is needed to reconstruct cellular developmental trajectories by inferring cell information at different time points, including state, distribution, number and gene expression (Bar-Joseph et al., 2012; Bendall et al., 2014; Ding et al., 2022). This dynamic analysis of transcriptome features is known as Pseudotime analysis (Figure 4E). Based on whether it depends on gene expression, pseudotime analysis methods can be divided into gene (exons) expression-based method and RNA-velocity-based method.

Pseudotime analysis based on gene expression level was first proposed, it usually clustering methods such as dimensional reduction is used to construct multi-branching graphs model in a low-dimensional space to mimic the developmental trajectory of cells: minimal spanning tree (MST) based method monocle (Trapnell et al., 2014), monocle 2 (Qiu et al., 2017), TSCAN (Ji and Ji, 2016); PAGA based method PAGA (Wolf et al., 2019), monocle 3 (Cao et al., 2019); other graph architectures method Wishbone (Setty et al., 2016), p-Creode (Herring et al., 2018) are all for this purpose. MST is a model that connects all points in a 2-dimensional plane and has the lowest total connection weight, it was used first to solve the traveling salesman problem, Qiu et al. (2011) applied MST model constructed with Boruvka's algorithm to analyze cellular hierarchy in 2011. Monocle maps cells into a high-dimensional Euclidean space and reduces dimension using ICA, Monocle 2 updates the monocle and uses the reversed graph embedding (RGE) strategy to construct cell path; cells are subsequently distributed to the spanning tree constructed using centroids. PAGA (partition-based graph abstraction) preserves the global topology structure of the dataset, by statistical connectivity measures of the neighborhood graph weights (KNN graph by default), PAGA graphs at multiple resolutions are produced to conduct pseudotime analysis based on an expanded diffusion pseudotime (DPT) method. Monocle 3 combines the advantages of monocle 2 and PAGA to form multiple PAGA graphs on the UMAP space, then uses the SimplePPT algorithm to learn the principal graph and then constrained by other PAGA graphs, the final derived cell developmental trajectories can be adapted to large datasets with compositional complexity. Overall, PAGA and monocle 3 comprehensively consider the computational speed, accuracy, and robustness, and are currently the best methods for the pseudotiming analysis of the single-cell transcriptome. In addition to the graph method, other gene expression based methods include CSHMMs which use HMM model to calculate the distance between each cell to root cell and then complete cell trajectory assignments iteratively; SCUBA which uses a bifurcation analysis model; SLICE which proposes a scEntropy directed model as highly differentiated cells have minimized scEntropy (Guo et al., 2017; Lin and Bar-Joseph, 2019; Marco et al., 2014).

The RNA-velocity-based method relies on the content of RNA velocity, which is proposed first by Peter V. Kharchenko group (La Manno et al., 2018) in 2018, they think that the ratio of unspliced/spliced mRNA can be used in infer transcriptional dynamics as cells with a higher proportion of uncleaved mRNA are younger (as a later cell differentiation state). At the same time, they also propose a dedicated analysis software, velocyto (available through the R package of the velocyto.R) as a steady-state model to quantify RNA velocity for developmental trajectory analysis. scVelo is another analysis tool designed specifically for RNA velocity, it uses the likelihood-based dynamical model to solve the cell trajectory inference with steady-state mRNA levels and situation violates the central assumption of the common splicing rate (Bergen et al., 2020). But there is still room for methodological improvement in velocity projection methods: constant degradation and nuclear export assumptions still need to be proved. This also provides a direction for the subsequent RNA velocity based method (Bergen et al., 2021). Methods concerning deep learning have been widely used in the modeling prediction of RNA velocity to further enhance processing power for large-complex datasets, like the Bayesian hierarchical model BRIE2 (Huang and Sanguinetti, 2021); velocity auto-encoder model based VeloAE (Qiao and Huang, 2021); variational auto-encoder model DeepCycle (Riba et al., 2022).

(2) Cell-cell interaction

Cell-cell interaction (CCI) is an important feature for maintaining the normal physiological function in multicellular organisms, which determines the fate of cells exploring the mechanism of disease occurrence, exploring the genetic variation process and other regulatory processes (Shao et al., 2020b; Singer, 1992). Cell interaction network intuitively embodies the interaction relationship between cells (Figure 4F).

Direct CCI based on the neighborhood structure refers to the extraction and analysis of the CCI with a possible direct contact, using the physical distance between the cells. ProximID method completes the physical cellular network construction on eligible cells with a predetermined interaction distance (Euclidean distance) (Boisset et al., 2018). Neighbor-seq identifies the cell types using a random forest classifier, CCI network is constructed by the igraph method using enrichment scores calculation score (Csardi and Nepusz, 2006; Ghaddar and De, 2022). Due to the great limitations of this analysis method, it is not currently used alone. This KNN connected graph is commonly used as one of the inputs to the CCI for deep learning, and its physical distance has also become an important hypothesis in single-cell CCI studies (two adjacent cells physically in direct contact are more likely to have some form of interaction than two random cells) for global CCI analysis.

The complete process of CCI relationships for indirect contact should include ligands, receptors, signaling proteins, transcriptional factors (TFs), and target genes. Common indirect CCI methods mainly use a priori ligand-receptor pairs databases (like the CellTalkDB database which integrates information from validated 3,398 human LR pairs and 2,033 mouse LR pairs (Shao et al., 2021a)), a cell-cell connection matrix is made in which each value represents the co-expression level of LR pairs. Then a cell connection graph is constructed for CCI analysis, the main analytical method packages include single-cell CCI inference method SoptSC (Wang et al., 2019d), Scriabi (Wilk et al., 2024); LR pairs based cluster CCI method NATMI (Hou et al., 2020), SingleCellSignalR (Cabello-Aguilar et al., 2020), scCrossTalk (Shao et al., 2024), CellPhoneDB (Efremova et al., 2020), Nichenet (Browaeys et al., 2020), CellChat (Jin et al., 2021), CellCall (Zhang et al., 2021d), ICELLNET (Noël et al., 2021), scMLnet (Cheng et al., 2021), CytoTalk (Hu et al., 2021b), Tensor-cell2cell (Armingol et al., 2022), LRLoop (Xin et al., 2022); other information based clusters CCI method InterCellDB (Jin et al., 2022b), EBOCOST (Zheng et al., 2022b). The LR pair-based approach constructs the database using the literature database curated or previous self-validated LR pairs: NATMI uses the connectomeDB2020 database by default (1,751 of 2,293 LR pairs were from the validated draft map by author in 2015) to construct weighted directed multi-edge networks (Ramilowski et al., 2015). CellPhoneDB proposes a certain SQLite database to retain specific subunit architecture of LR pairs, mean expression threshold is used to determine the interacting cells, and a geometric sketching subsample framework is used for enhanced power to large datasets and excluded noise. Similarly, ICELLNET takes use of the multi-subunit structure of ligands with receptors for heteromeric complexes. NicheNet uses model-based parameter optimization on an LR prior model to optimize CCI intensity by adding intracellular signaling information (target gene). It overcomes the problem that the above methods directly use the receptor gene expression level to represent the amount of receptor protein in the cells and combine the downstream signaling pathway with GRN to improve CCI analysis. Therefore, in the analysis of the single-cell transcriptome CCI, the CellPhoneDB and NicheNet are usually used together to achieve the best analysis results (Dimitrov et al., 2022).

The latest methods of single-cell CCI adopt the strategy of DL and improve the application performance to some extent. DeepLinc uses a VGAE model to reconstruct full range intercellular CCI network (Li and Yang, 2022). TraSig is a continuous-state Hidden Markov Model that uses pseudotime ordering to calculate dynamic interaction scores for CCI inference (Li et al., 2022a). In addition, as now spatially resolved transcriptomics (ST) provides the gene information with crucial spatial information, the inference of spatial cell-cell communications remains a great challenge. SpaOTsc can reconstruct the spatial properties of scRNA-seq data and build the CCI network relying on a structured optimal transport method (Cang and Nie, 2020). Giotto uses a cell-cell proximity graph to infer the signaling pathways (Dries et al., 2021b). However, both SpaOTsc and Giotto hardly resolve the spot-based ST data. Recently, Fan's lab (Shao et al., 2022a) present SpaTalk which uses a knowledge graph and graph network to construct a ligand-receptor-target network between spatially adjacent cells for both single-cell and spot-based ST data.

(3) Pathway enrichment analysis

Gene pathway enrichment analysis refers to using the gene of interest as a foreground gene and known specific database associations to establish gene-biological process links which are used to explain the physiological functions of differentially expressed genes, upstream and downstream pathways, etc (Creixell et al., 2015). Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) are the first batch proposed databases used as an enrichment analysis (Ashburner et al., 2000; Ogata et al., 1999). Gene set enrichment analysis (GSEA) is another widely used method calculating the enrichment score to determine whether gene set S will occur in both sides of ranker DEGs List L as well as the significance test value (Subramanian et al., 2005). All pathways in Ingenuity Pathway Analysis (IPA) software have been experimentally verified and it can predict the changing trend of the entire pathway when activated compared with other analysis. Other common database pathway enrichment methods also include over-representation analysis (ORA) (Khatri et al., 2012), network topology-based analysis (NTA) (Wang et al., 2013), Reactome gene sets (Fabregat et al., 2018), CORUM complexes (Ruepp et al., 2010). The difference in the source pathways and enrichment means of the reference datasets will directly affect the pathway enrichment results. Web-based online analysis tools with different integrated databases have been proposed to easier analysis (Wang et al., 2017b; Zhang et al., 2005). Metascape involves the transcriptome databases (KEGG, GO, CORUM, TRRUST, etc.) and protein-protein interactions databases (STRING, BioGrid, OmniPath, etc.); there are a total of 25 databases that can be used for genetic and proteomic enrichment analysis in 8 species including human and mouse (Zhou et al., 2019b).

In conclusion, although we have enumerated the most common parts of single-cell transcriptome downstream analysis (Tables S3 and S4 in Supporting Information), there are still many methods not covered including gene regulatory network analysis, immune analysis, cell cycle assignment, gene variants exploration, alternative splicing analysis. Overall, single-cell transcriptome analysis methods are diverse and still evolving, the starting point and the ultimate goal of all the analysis methods is to use the accurate mining of the biological information from the single-cell transcriptome sequencing data for biological explanation.

Applications of scRNA-seq

ScRNA-seq has become a potent instrument, empowering scientists to delve into the intricate realm of individual cells and unveil their distinct molecular characteristics. Leveraging scRNA-seq, investigators now have the capability to delve into cellular diversity, study developmental biology, scrutinize disease progression, and advance drug development with unparalleled precision. This methodology has unlocked fresh possibilities for revealing novel biomarkers, pinpointing therapeutic targets, and forging pathways toward personalized medicine. In this segment, we illuminate and deliberate on several noteworthy applications of scRNA-seq in the realms of biomedical and clinical research.

Application of scRNA-seq in embryonic, tissue, and organ development research

(1) Embryonic development research

ScRNA-seq proves to be pivotal in embryonic development, particularly in the identification and categorization of distinct cell types and lineages.

In a research by Cao et al. (2019), 38 primary cell types and 655 subtypes were identified by scRNA-seq analysis of 2,072,011 single cells from mouse embryos. This comprehensive information sheds light on the developmental trajectories of different cell types during organogenesis in mammalians, culminating in the creation of a developmental-specific trajectory map for skeletal muscle cells. This research significantly contributes to advancing our knowledge in the realm of mammalian developmental biology. Moreover, researchers leverage scRNA-seq to reconstruct developmental trajectories, unraveling the regulatory networks that govern cell fate decisions through the analysis of gene expression at different developmental stages (Wu et al., 2024). Scialdone et al.'s (Scialdone et al., 2016) study, which analysed 1,205 mouse cells from the early gastrula formation stage, used single-cell transcriptome analysis, created gene expression profiles for healthy mammals during the early developmental stage, and investigated the function of the important hematopoietic transcription factor Tal1, serves as an instructive example.

Nestorowa et al. (2016) harnessed the potential of scRNA-seq to profile over 1,600 single hematopoietic stem and progenitor cells, unveiling the trajectories of hematopoietic stem cells and shedding light on the molecular events orchestrating blood cell differentiation. Additionally, scRNA-seq also plays an essential role in identifying pivotal regulatory factors and signaling pathways dictating cell fate decisions during embryogenesis. In a study by Lescroart et al. (2018), Mesp1 emerged as a crucial transcription factor involved in the specification of cardiac progenitor cells during mouse heart development, revealed through scRNA-seq. By integrating single-cell transcriptomic analyses of human mesoderm cells derived from embryonic stem cells and embryos, Wen et al. (2024) identified and defined the molecular characteristics of human hematopoietic mesoderm cells biased towards hematopoietic lineages. Cui et al. (2019) used scRNA-seq to explore the gene expression landscapes of almost 4,000 cardiac cells from human embryos. They identified four main cell types and uncovered important signalling pathways that may be essential for the maturation and differentiation of several cell types. This wealth of information lays the foundation for a deeper understanding of the mechanisms governing *in vivo* human cardiac development.

(2) Tissue and organ development research

Apart from studying embryonic cells, scRNA-seq is a useful method to describe the evolution of certain cell populations in organs or tissues, providing information about important transitions in the development of animal tissues and organs (Mu et al., 2019; Paik et al., 2020). Subsequently, we will delve into the applications of scRNA-seq in tissue and organ research, emphasizing its significance in unveiling cellular diversity, identifying rare cell populations, and comprehending disease mechanisms.

In 2016, the Human Cell Atlas (HCA) project was initiated as a large-scale international collaboration with the goal of mapping and characterizing all distinct cell types in the human body. The primary objectives include understanding the spatial organization and functional relationships of these cell types, advancing our knowledge of human biology, and ultimately enhancing the diagnosis and treatment of diseases. Since its inception, numerous scRNA-seq analyses have contributed to depicting the landscapes of human cells. Using microwell-seq, Han et al. (2020) profiled 702,968 single cells from seven different types of cell culture and 60 human tissue types, revealing cell heterogeneity in a variety of human tissues that had not been known before. For example, their examination of human kidney tissues in the fetal and adult stages identified a new intercalated cell-tran-principal cell type in the adult kidney and previously unidentified kinds of S-shaped body cells in the fetal kidney. This construction of a human cell landscape at the single-cell level serves as a valuable resource for advancing our understanding of human biology. Similarly, studies by Jones et al. (2022b), Eraslan et al. (2022), Domínguez Conde et al. (2022), as well as Suo et al. (2022), reported pan-tissue single-cell transcriptome atlases encompassing more than 500 different cell types and over a million cells from 68 different donors in more than 30 human tissues. These researches identified unusual cell types, tissue-agnostic traits, tissue-specific cell states, and even disease-associated cell types through cross-tissue comparisons of cell types and their transcriptional properties. These pan-tissue investigations mark a significant milestone in constructing a comprehensive human single-cell atlas.

In addition to constructing organ atlases, scRNA-seq has proven instrumental in unveiling cell type-specific gene expression changes associated with various diseases, offering crucial insights into disease mechanisms. Wilson et al. (2019) performed unbiased snRNA-seq on cryopreserved human diabetic kidney samples. They found that key inflammatory markers TNFRSF21 and ILR1 were significantly increased in infiltrating immune cells, potentially serving as biomarkers for disease progression or targets for early intervention in diabetic kidney disease. In another study, Koenig et al. (2022) used a combination of snRNA-seq and scRNA-seq to examine cardiac tissue from 27 healthy donors and 18 people with dilated cardiomyopathy. They deciphered transcriptional programs distinct to each major cardiac cell type, found gene expression profiles linked with the condition, and illuminated the molecular mechanisms behind dilated cardiomyopathy.

In summary, scRNA-seq has played a pivotal role in constructing comprehensive atlases of human tissues and organs. It has unveiled intricate gene expression networks and developmental trajectories, offering insights into potential factors and targets implicated in the pathogenesis of individual organs or tissues. This wealth of information has the potential to unravel disease mechanisms and provides a solid foundation for advancing disease treatment strategies.

Application of scRNA-seq in tumor biology research

ScRNA-seq, by enabling the comprehensive analysis of the entire transcriptome at a single-cell resolution, has transformed our comprehension of tumor biology. It has been instrumental in unveiling the heterogeneity within tumors, identifying distinct subclones, characterizing interactions between tumor and immune cells, revealing signaling pathways associated with tumors, and predicting responses and resistance to drugs. This technology allows for the creation of detailed cell maps, the discovery of novel biomarkers, and the identification of therapeutic targets. As scRNA-seq technology continues to advance, it holds significant promise for enhancing patient outcomes and expediting the development of personalized treatments.

(1) Tumor heterogeneity

Tumors exhibit a diverse array of cell types, each characterized by unique gene expression profiles and functional attributes. Although traditional bulk RNA sequencing is very informative, it provides an averaged gene expression profile of the entire tumor, masking the internal cellular heterogeneity. ScRNA-seq technology is capable of identifying and characterizing distinct cell populations within tumors, offering a nuanced perspective of the tumor ecosystem.

For instance, Hu et al. (2020b) employed scRNA-seq to identify six subtypes of fallopian tube epithelium cells in healthy human fallopian tube tissues, as well as serous ovarian cancer (SOC) subgroups linked to patient prognosis, and reveal intra-tumoral heterogeneity in SOC. Another study by Liang et al. (2021a) analyzed scRNA-seq data from eight high-grade SOC cases and identified 20 tissue-specific cell clusters. By taking use of the heterogeneity of ovarian cancer immune cells, the study developed a two-gene signature prognostic stratification approach (CXCL13 and IL26) to precisely assess prognostic risk. Tumor heterogeneity poses a significant challenge to the precise diagnosis and treatment of gastric adenocarcinoma (GA). Zhang et al. (2021b) analysed 27,677 cells from nine GA samples and three non-tumor samples using unbiased scRNA-seq technology.

This investigation revealed differentiation and cellular heterogeneity both within and across GA patients, providing insight into the molecular makeup of an uncommon chief cell-predominant GA type (GA-FG-CCP). The authors proposed a biomarker panel for distinguishing between benign and malignant epithelium based on their findings. Zhong et al. (2022) utilized scRNA-seq to explore cellular heterogeneity and regular networks in 9 patients with multiple myeloma. Through analysis, they discovered unique molecules, networks, and crosstalk pairs in different stages of the disease, offering valuable insights into its prognosis and treatment. Therefore, by unraveling tumor heterogeneity, scRNA-seq aids researchers in comprehending the various cell types present, their interactions, and their contributions to tumor development, progression, and responses to treatment.

(2) Tumor microenvironment

The tumor microenvironment (TME) constitutes an intricate ecosystem comprising diverse cell types, including cancer cells, immune cells, stromal cells, and vascular cells. It holds a pivotal role in tumor development, progression, and responses to therapy. ScRNA-seq can unravel the complexity of TME by identifying and classifying distinct cell populations based on their gene expression profiles.

Through the profiling of individual cell transcriptomes, scRNA-seq allows the identification of rare cell populations within the TME, such as tumor-infiltrating lymphocytes (TILs), cancer-associated fibroblasts (CAFs), and myeloid-derived suppressor cells (MDSCs). An exemplary study conducted by a research group from China utilized scRNA-seq to analyze the transcriptomes of 47,304 cells from nine patients with gastric cancer. The study unveiled multiple immune cell subsets, including regulatory T cells, $CD4^+$ T cells, $CD8^+$ T cells, natural killer cells, and innate lymphocyte cells (Li et al., 2022k). Notably, the study found an enrichment of regulatory T cells in gastric tumor tissues, marked by increased expression of immune suppression-related genes like *DUSP4*, *IL2RA*, *TNFRSF4*, *LAYN*, and *LGALS1*, indicating an immunosuppressive microenvironment in gastric tumors.

Understanding TME cellular heterogeneity and gene expression may lead to the creation of novel targeted cancer therapeutics as well as cutting-edge early diagnostics. Significant heterogeneity in the infiltrating T cell population was found in a study by Savas et al., which involved the analysis of 6,311 intratumoral T cells extracted from 123 breast cancer patients using scRNA-seq. According to the study, individuals with breast cancer who had a high TIL count had $CD8^+$ T cells that had characteristics of tissue-resident memory T (TRM) cell development. Moreover, these $CD8^+$ TRM cells exhibited significant quantities of cytotoxic effector proteins (PRF1 and GZMB) and immunological checkpoint molecules (PDCD1 and CTLA4) (Savas et al., 2018). Moreover, in early-stage triple-negative breast cancer, the gene profiles found inside the $CD8^+$ TRM cluster were significantly linked to favorable patient survival. This highlights the ability of scRNA-seq to detect small subpopulations of TILs that are associated with immune surveillance or immunosuppression. These various immune cell types may be used as therapeutic targets or prognostic variables for breast cancer.

(3) Therapeutic selection and monitoring

The development of tailored treatment plans is a significant use of scRNA-seq in cancer research. The population of cells that make up tumors is diverse and includes endothelial, stromal, immunological, and malignant cells. These cell types can all play a role in treatment resistance, metastasis, and tumor formation. By analyzing the gene expression profiles of individual cells within a tumor, scRNA-seq can help identify specific cell populations that play key roles in tumor progression and therapy resistance.

In a study conducted by Tirosh et al. (2016), scRNA-seq was employed to analyze the heterogeneity of melanoma tumors and pinpoint distinct cell states linked to therapy resistance. The investigation revealed a specific subpopulation of tumor cells characterized by elevated AXL gene expression, which was associated with resistance to targeted therapies. This discovery paved the way for the development of combination therapies targeting both the AXL pathway and the targeted therapy pathway, resulting in improved treatment responses. Not every patient responds to immune checkpoint inhibitors, such as anti-PD-1 and anti-CTLA-4 antibodies, which have revolutionised cancer treatment by increasing the immune system's capacity to identify and fight cancer cells. To unravel the underlying resistance mechanisms and enhance treatment outcomes, researchers at the Broad Institute of MIT and Harvard conducted a study using scRNA-seq to analyze the gene expression profiles of individual cells within tumor samples from 33 melanoma patients (Jerby-Arnon et al., 2018). The scRNA-seq analysis unveiled a distinct subset of cancer cells known as the T cell exclusion program (TEX). These TEX cells actively suppressed the recruitment and activation of T cells in the tumor microenvironment, forming an immunosuppressive barrier that shielded cancer cells from immune attack. The TEX program was associated with resistance to immune checkpoint blockade therapies and poor response to anti-PD-1 treatment and may serve as a potential therapeutic target to overcome immune resistance.

Application of scRNA-seq in immune system research

The immune system, which is made up of immune molecules, immune cells, and immunological organs, is a crucial component of the body's internal environment. Its job is to identify and eliminate antigenic foreign substances from the body (Akar-Ghibril, 2022; See et al., 2018). The immune system may produce autoantigenic reactions in the course of combating infections, which can result in immunological disorders and harm tissues or organs (Li et al., 2022m; Suo et al., 2022). The complexity and diversity of immune illness mechanisms make the timely identification of disease triggers essential for the treatment of immunological diseases (Zhao et al., 2021b). As a powerful technology, scRNA-seq can discover new cell subpopulations, reveal the developmental lineage of immune cells, and identify the regulatory programs of immune responses in immune diseases, thereby further elucidating the pathogenesis of immune diseases at the single-cell level and exploring new therapeutic strategies to benefit more patients.

(1) Research on immune cell heterogeneity

ScRNA-seq can characterize individual cells within tissues and organs and identify rare and previously unknown cell populations. In immune system diseases, this technique has unveiled distinct immune cell subsets and their functional states.

He et al. (2023) analyzed 26,456 immune cells from old zebrafish brains by scRNA-seq and revealed the crucial role of microglia and T cells in the neurodegenerative process in aging.

A study from the USA analyzed the transcriptomes of approximately 276,000 single peripheral blood mononuclear cells (PBMCs) from 33 children with systemic lupus erythematosus (cSLE) and 11 matched healthy controls using scRNA-seq (Nehar-Belaid et al., 2020). This investigation identified two novel cell subpopulations (ISG^{hi} T-SC4 and CD8^{+} T cells expressing high levels of cytotoxic proteins) and revealed SLE-restricted activated NK cells and ISG^{hi} NK-SC associated with disease severity. This comprehensive profiling of SLE heterogeneity at the single-cell level contributed to a deeper understanding of the cellular composition and functional diversity within the immune system, shedding the underlying mechanisms driving disease progression. Zheng et al. (2022a) obtained the single-cell landscape associated with lupus pathogenesis by scRNA-seq. The study elucidated the heterogeneous characteristics present in cutaneous lesions between discoid lupus erythematosus (DLE) and SLE, contributing to a better identify potential avenues for therapeutic intervention.

(2) Research on the mechanism of immune disease

ScRNA-seq technology empowers researchers to dissect the intricate cellular composition of immune diseases, pinpoint dysregulated pathways, and uncover novel cell types or subtypes that may contribute to disease pathogenesis. The utilization of scRNA-seq promises substantial progress in the diagnosis, treatment, and management of immune diseases in the future. Using 10x Genomics, Gaydosik et al. (2021) focused on 3,729 CD3^{+} lymphocytes from skin biopsies of 10 healthy donors and 27 patients with active systemic sclerosis (SSc). This study revealed different tissue-resident and circulating T cell subpopulations in both healthy and SSc skin and identified the cytokines that contribute to inflammatory immune disorders. The findings advance our understanding of the immunological mechanisms underlying disease processes and hold potential for the development of novel, tailored therapy approaches in SSc. Xu et al. (2022) analyzed the pathogenic mechanism of vitiligo through single-cell transcriptome technology, revealed the relationship between skin fibroblasts and vitiligo, and further clarified the location preference of vitiligo onset, which has important guiding significance for the development of new therapeutic strategies for the treatment of vitiligo.

Application of scRNA-seq in infectious diseases research

ScRNA-seq has revolutionized the field of infectious disease research by enabling the study of host-pathogen interactions, characterizing the host immune response, and investigating the impact of infectious diseases on host tissues. With the use of this technology, the complexity of infectious diseases has been untangled at the single-cell level, revealing hitherto unexplored insights into the variety and functional states of cellular responses during infection.

A variety of diseases, including interstitial pneumonia with consolidation, granulomatous lesions with non-necrotic or caseous necrotic centres, and cavitary liquefied lesions, are indicative of Mycobacterium TB infection, which results in pulmonary tuberculosis (Hunter et al., 2014). Six tuberculosis patients' lung tissues were subjected to scRNA-seq by Wang et al. (2023b) with the goal of investigating the heterogeneity and intercellular interaction in regions with 18F-FDG avidity and nearby uninvolved tissues. The scRNA-seq analysis identified a total of 29 distinct cell subsets, encompassing both immune and parenchymal cells, each characterized by specific marker genes. The detailed characterization of these cell types and their associated marker genes offers a comprehensive understanding of the distinct immune and non-immune populations present in tuberculosis-infected lungs. This information is crucial for deciphering the complex interactions between these cells during tuberculosis infection and may aid in identifying potential new therapeutic targets. Zhao et al. (2023) employed scRNA-seq to comprehensively analyze the gene expression profiles of immune cells in draining lymph nodes responding to *Y. pestis* infection, which may contribute to understanding of the plague pathogenesis during the early stage of infection. Chua et al. (2020) utilized scRNA-seq on nasopharyngeal and bronchial samples from 19 patients with COVID-19 to identify molecular correlates of disease severity. By better understanding the underlying molecular pathways behind infectious diseases, researchers may be able to design diagnostic tools, treatments, and preventive measures that are more successful.

Application of scRNA-seq in drug discovery and development research

The drug discovery process is often hindered by inefficiencies due to a limited understanding of human biology, including cellular heterogeneity, disease mechanisms, drug responses, and therapeutic targets. Since its inception in 2009, scRNA-seq technology has come a long way and offers a promising solution for drug development. This method can be integrated at different stages of the drug discovery and development pipeline, as it can capture individual cell whole-transcriptome profiles. During the initial phases, scRNA-seq can assist in discovering novel cellular and molecular targets. By deepening our understanding of diseases through subtyping based on altered cell compositions and states, scRNA-seq contributes to a more nuanced comprehension of pathological mechanisms. Furthermore, this technique provides insight into the actions of compounds that are particular to cell types, off-target effects, and heterogeneous responses, all of which help in the process of choosing potential new drugs. During clinical development, scRNA-seq plays a pivotal role in identifying biomarkers for patient stratification. It helps unravel drug mechanisms of action or resistance and allows for the monitoring of drug responses and disease progression. By providing insights into the cellular and molecular landscape, scRNA-seq serves as a valuable tool in optimizing the drug discovery process, ultimately facilitating the development of more effective and targeted therapeutics.

(1) Target identification

ScRNA-seq has brought about a paradigm shift in the initial stages of drug discovery, particularly in the identification of therapeutic targets crucial for disease pathogenesis. This technology facilitates target identification by unveiling dysregulated cell types and states in disease conditions. Profiling the transcriptomes of individual cells allows the identification of specific genes and pathways that exhibit differential expression in disease-associated cell populations compared with healthy counterparts. This information becomes instrumental in guiding the selection of potential therapeutic targets and prioritizing candidate molecules for further exploration.

For instance, Abdelfattah et al. (2022) used scRNA-seq to analyze over 200,000 human glioma, immunological, and stromal cells at the single-cell level in glioblastoma. They discovered S100A4 to be a novel target for immunotherapy in glioblastoma using this method. Interestingly, eliminating this target in non-cancerous cells showed an amazing capacity to rewire the immune system, resulting in a notable increase in survival. In the context of chronic pancreatitis, single-cell sequencing of pancreatic immune cells and T cell receptors has shed light on potential therapeutic targets. The identification of the CCR6-CCL20 signaling pathway in genetic chronic pancreatitis opens avenues for targeted interventions in humans (Lee et al., 2022). In order to examine the relationship between tumors and surrounding immune cells, a study from the University of Texas MD Anderson Cancer Centre, USA, used scRNA-seq on 186,916 cells from 5 early-stage lung adenocarcinomas and 14 multi-region normal lung tissues (Sinjab et al., 2021). The results of this study indicate that CD24 expression in tumor epithelium is dramatically elevated and is connected with pro-tumor immune phenotypes and decreased survival. These findings imply that CD24 could be a promising target for the treatment of early-stage lung adenocarcinome.

(2) Drug screening and optimization

ScRNA-seq has played a pivotal role in enhancing the efficiency and precision of drug screening and optimization. Traditional screening methods often rely on cell populations that may not fully capture the heterogeneity present in the target tissue or organ. Leveraging scRNA-seq, researchers can identify and isolate specific cell types or subpopulations of interest, allowing for a more nuanced assessment of their response to various drug candidates.

Cao et al.'s (Cao et al., 2020a) high-throughput single-cell RNA and VDJ sequencing of antigen-enriched B cells from 60 recovering patients serves as an instructive example. Using this method, they quickly isolated 14 strong neutralising antibodies against SARS-CoV-2 from a large collection of 8,558 IgG1$^+$ clonotypes that bind to antigen. Among the antibodies against SARS-CoV-2, BD-368-2 was found to have the strongest neutralising impact. Additionally, its therapeutic and preventive activity was verified in hACE2-transgenic mice infected with SARS-CoV-2. This work shows how human neutralising antibodies can be effectively discovered by high-throughput single-cell sequencing, especially during pandemics of infectious diseases.

(3) Drug mechanisms of action

ScRNA-seq offers a valuable tool to gain insights into the cellular and molecular changes induced by drugs, enabling a comprehensive characterization of their mechanisms of action. An example of this application is seen in the work of Taukulis et al. (2021), who employed scRNA-seq to investigate acute cisplatin-induced ototoxicity in a mouse model. By comparing the transcriptomes of cisplatin-treated adult stria vascularis with unperturbed adult stria vascularis, the researchers identified cell type-specific regulatory networks. Their findings highlighted that marginal and intermediate cells in the stria vascularis are preferentially affected by cisplatin exposure. Additionally, scRNA-seq data revealed specific gene expression changes associated with chemotherapy-induced ototoxicity. Notably, genes such as *Alcam*, *Atp1b2*, *Spp1*, and *Car12* were downregulated, while *Klf10*, *Cldn3*, and *Tspan1* were upregulated in cisplatin-treated stria vascularis. These differentially expressed genes present potential novel therapeutic targets to mitigate ototoxicity caused by chemotherapy. Zhang et al. (2022f) investigate the immunomodulatory mechanisms of dihydroartemisinin using scRNA-seq in combination with cellular and biochemical methods. Their research revealed that dihydroartemisinin beneficially regulated immune cell heterogeneity and splenic immune cell homeostasis by activating the SOD3-JNK-AP-1 pathway to treat autoimmune diseases. Understanding the mechanisms of drug action is vital for optimizing therapeutic efficacy and minimizing adverse effects.

(4) Patient stratification

ScRNA-seq is a useful ally of personalised medicine, which seeks to customise treatment plans based on unique patient features. This technology contributes to patient stratification by profiling the transcriptomes of individual cells, allowing for the identification of markers relevant to disease prognosis or therapeutic response.

In the context of infant acute lymphoblastic leukemia (iALL), where relapse occurrence is often fatal (Pieters et al., 2019). ScRNA-seq has shown promise in prognostic risk assessment of iALL. Using samples from patients with MLL-rearranged infant acute lymphoblastic leukemia (MLL-r iALL), Candelli et al. (2022) performed scRNA-seq. By measuring the percentage of cells found to be either sensitive or resistant to therapy, the researchers were able to forecast when MLL-r iALL would relapse. This approach outperformed current risk stratification schemes, showcasing the potential of scRNA-seq in refining prognostic markers for better treatment outcomes.

Our knowledge of cellular heterogeneity, disease processes, and treatment responses at the single-cell level has been completely transformed by the use of scRNA-seq in drug discovery and development. Its contributions span the identification of therapeutic targets, improvement of drug screening and optimization, elucidation of mechanisms of action, and facilitation of patient stratification. The integration of scRNA-seq in drug discovery holds significant promise for developing more effective and personalized therapies, ultimately leading to improved patient outcomes.

Summary

The evolution of single-cell transcriptomic atlases through advancements in scRNA-seq technology has provided unprecedented resolution, offering insights into complex cellular events and enhancing our understanding of cell composition and interactions across humans, model animals, and plants. This chapter underscores the progress in different aspects of scRNA-seq technology, emphasizing distinct features and strengths in various areas. It is crucial to recognize that each single-cell sequencing method has its own advantages and limitations. Ongoing developments in this field aim to design improved methods that enhance robustness and coverage, allowing for comprehensive detection of cellular composition at multiple levels and the depiction of cell landscapes within different species. The expectation is that future innovations in scRNA-seq technologies will contribute significantly to the advancement of biological and clinical medicine, offering powerful tools for in-depth exploration and understanding of cellular dynamics.

Chapter 2 Single-cell whole-genome sequencing

The microscopization of life science research proves that cell population-based methods may not be suitable for certain areas of study, such as tumor heterogeneity and early embryonic development. In response, the introduction of single-cell transcriptome sequencing technology in 2009 marked a significant advancement (Tang et al., 2009). Building on this, Navin et al. (2011) introduced single-cell whole genome sequencing technology (scWGS) in 2011 by combining whole genome amplification (WGA) with high-throughput sequencing. This innovative approach addresses challenges related to obtaining information about heterogeneity between different cells in tissue samples and enables the study of individual cells when conventional sequencing might be impractical due to small sample sizes.

By sequencing DNA at the single-cell level, scWGS provides a new dimension for studying the behavior and mechanisms of individual cells. Applications of scWGS have expanded across various research fields, including neuroscience, germline evolution, organogenesis, oncology, clinical diagnosis, immunology, microbiology, embryo development, and prenatal genetic diagnosis. Recognizing its potential, scWGS was highlighted as one of the most anticipated technologies in 2013 by the journal *Nature Methods*.

The development of scWGS has indeed opened up avenues for researchers to delve into inter-cell heterogeneity at the single-cell level and explore various aspects such as single-nucleotide variants, short insertions or deletions, and copy number variants. This technology has proven particularly valuable for studying the genomes of rare cells that hold biological or clinical significance, including circulating tumor cells and cells used in third-generation *in vitro* fertilization preimplantation genetic diagnosis/screening. The scWGS process typically involves three main steps: single-cell isolation, single-cell whole genome amplification (scWGA), and the sequencing and analysis of the amplified products. The critical challenge in this process is to amplify the genome of a single cell effectively, obtaining sufficient material for downstream analyses while minimizing artifacts such as amplification bias, genome loss, mutations, and chimeras. Addressing these challenges is essential to ensure the accuracy and reliability of the genetic information obtained from single cells.

In this section, we first focus on the advancements in scWGA technology. Subsequently, we provide a detailed introduction to several prominent scWGA chemistries, elucidating their crucial biochemical reaction strategies. Our focus then shifts to high-throughput scWGS methods, which enable the parallel sequencing of tumor cell genomes on a massive scale. This approach opens up opportunities to significantly broaden the scope of intratumoral characterization. Key milestones in the development of scWGA and high-throughput scWGS technologies are visually represented in Figure 5. Lastly, we offer a summary of the most recent practical breakthroughs in scWGS within the field of biomedicine. This overview outlines a vision for applying single-cell genomic sequencing in clinical research, highlighting its potential impact on advancing our understanding of biological processes and disease mechanisms.

ScWGA methods

Given the limited DNA content in a single cell (approximately 6 pg/cell), which falls short of the detection requirements of sequencers, it is imperative to amplify the trace amounts of whole-genome DNA in single cells before sequencing. This amplification process aims to generate a complete genome with high coverage, ensuring accurate and comprehensive sequencing results in subsequent high-throughput sequencing. Over time, major changes have occurred in WGA technology to meet these demands. Notable methods include degenerate oligonucleotide-primed polymerase chain reaction (DOP-PCR) (Telenius et al., 1992), multiple displacement amplification (MDA) (Dean et al., 2001), and multiple annealing and looping-based amplification cycles (MALBAC) (Zong et al., 2012). Subsequent innovations, such as linear amplification via transposon insertion (LIANTI) (Chen et al., 2017a), single-stranded sequencing using microfluidic reactors (SISSOR), primary template-directed amplification (PTA) (Gonzalez-Pena et al., 2021), and multiplexed end-tagging amplification of complementary strands (META-CS) (Xing et al., 2021) have further expanded the WGA

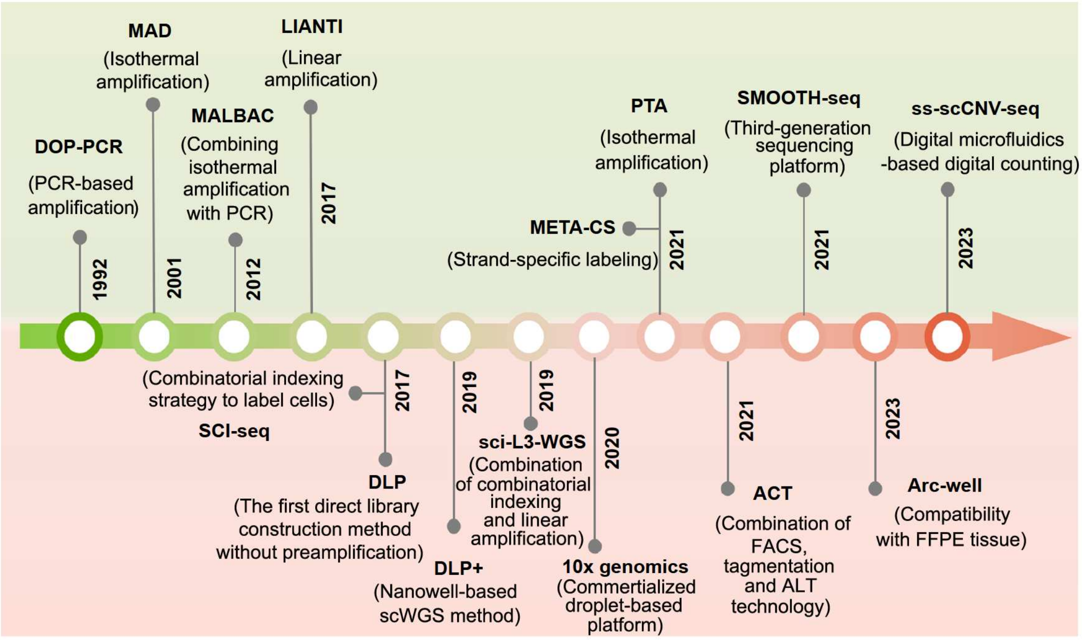

DOP-PCR
(PCR-based amplification)
1992
MAD
(Isothermal amplification)
2001
MALBAC
(Combining isothermal amplification with PCR)
2012
LIANTI
(Linear amplification)
2017
META-CS
(Strand-specific labeling)
2021
PTA
(Isothermal amplification)
SMOOTH-seq
(Third-generation sequencing platform)
2021
ss-scCNV-seq
(Digital microfluidics -based digital counting)
2023
(Combinatorial indexing strategy to label cells)
SCI-seq
2017
DLP
(The first direct library construction method without preamplification)
2019
DLP+
(Nanowell-based scWGS method)
2019
sci-L3-WGS
(Combination of combinatorial indexing and linear amplification)
2020
10x genomics
(Commertialized droplet-based platform)
2021
ACT
(Combination of FACS, tagmentation and ALT technology)
2023
Arc-well
(Compatibility with FFPE tissue)

Figure 5. Timeline of the development of single-cell whole genome amplification and sequencing technology. The upper half of the graph shows the main events marking the development of scWGA and low-through scWGS technologies, and the bottom half the development in throughput.

toolkit. Table 1 provides an overview of the general characteristics of these methods. In the following sections, we conduct a review of several major WGA methods, focusing on their coverage, uniformity, and accuracy. In short, there is no obvious winner in amplification performance, and each strategy has advantages according to the parameter that matters.

PCR-based amplification

PCR-based WGA methods come in various forms, including primer extension preamplification PCR (PEP-PCR) (Zhang et al., 1992), DOP-PCR (Telenius et al., 1992), tagged random primer PCR (T-PCR) (Grothues et al., 1993) and ligation-mediated PCR (LM-PCR) (Klein et al., 1999). These techniques were essential in obtaining the amplification of genomic DNA from single cells, meeting particular research objectives, and acting as models for scWGA technology. One of the first scWGA techniques used was PEP-PCR, which was later superseded by the more popular DOP-PCR. These techniques have led to the development of commercial kits, such as the PicoPLEX WGA Kit (Rubicon Genomics, USA) and the GenomePlex Single Cell Whole Genome Amplification Kit (Sigma-Aldrich, USA).

DOP-PCR operates on the principle of utilizing a partially random primer for a two-step PCR amplification of template genomic DNA (Telenius et al., 1992). The degenerate primers consist of a random six base sequence in the middle flanked by fixed sequences at each end (5′ CGACTCGAGNNNNNNATGTGG 3′). The short ATGTGG sequence at the 3′ end of the primer has an extremely high distribution frequency in genomic DNA, guiding the initial low-temperature annealing step and determining the starting site of amplification with a bias toward specific sequences. The middle six degenerate bases create 46 different sequences, and during annealing, one or more of these degenerate bases, along with the 3′ end specific bases, simultaneously bind to the template DNA, enhancing the primer's binding efficiency. The PCR amplification occurs in two steps: the first few cycles (3–5) involve low-temperature annealing (e.g., 30°C), followed by strand extension at an elevated temperature. In the second step, the products from the first step undergo further amplification using a primer targeting the 5′ fixed sequence at a higher annealing temperature (62°C).

The efficiency of DOP-PCR amplification relies on primer concentration and polymerase activity. At a low annealing temperature, the primers can bind to multiple genomic loci, resulting in amplification products that cover nearly the entire genome. DOP-PCR stands out as a representative method in PCR-based WGA, finding application in amplifying minute amounts of human genomic DNA in single cells for analyses related to tumor heterogeneity, assessment of copy number variations, and detection of aneuploidies (Knouse et al., 2014; McConnell et al., 2013; Navin et al., 2011). However, DOP-PCR often generates low genome coverage (typically less than 10%) (Navin et al., 2011), a characteristic associated with the exponential nature of PCR amplification. Additionally, the PCR amplification reaction has a high base mismatch rate, rendering this amplification technology less suitable for the detection of single-nucleotide variations due to the elevated false positive rate.

Isothermal amplification

MDA is the most representative method among isothermal scWGA methods, initially developed by Dean et al. (2001). Operating under isothermal conditions, MDA employs a 6-base pair random primer that randomly anneals to the genome, initiating a strand displacement amplification reaction catalyzed by phi29 DNA polymerase with robust strand displacement activity. The single-stranded sequence produced through displacement can be extended randomly by annealing with random primers, resulting in the formation of multibranched amplification structures. Due to the potent DNA synthesis ability of phi29 DNA polymerase, the synthesized DNA fragments are typically 50–100 kb in length. Additionally, the high replication fidelity of phi29 DNA polymerase, characterized by an error rate of about one nucleotide per 10^8, owing to its 3′→5′ exonuclease and proofreading activities, makes MDA suitable for accurate single nucleotide variation (SNV) calling. This feature has led to its application in single-cell genome lineage tracing (Lodato et al., 2015). Furthermore, MDA provides significantly higher genome coverage compared with initial PCR-based methods. However, a drawback of MDA is its exponential amplification process, similar to DOP-PCR, which introduces sequence-dependent bias and hinders coverage uniformity. It is worth noting that the sequence-dependent bias of MDA is not consistently reproducible across the genome from one cell to another, making copy number variation (CNV) measurements noisy and normalization less effective.

To address amplification bias and enhance uniformity and coverage, various improved isothermal amplification technologies, including emulsion MDA (eMDA) (Fu et al., 2015), digital droplet MDA (ddMDA) (Sidore et al., 2016), TruePrime (Picher et al., 2016), SISSOR (Chu et al., 2017) and PTA (Gonzalez-Pena et al., 2021) have been developed based on MDA technology. Both eMDA and ddMDA involve dispersing the amplification process into millions of small droplets, aiming to improve amplification uniformity and correct bias. The TruePrime technique substitutes the N6 primer in the MDA method with a unique DNA primase called TthPrimPol to enhance amplification uniformity. SISSOR enhances sequencing accuracy by randomly distributing megabase-sized single-stranded DNA fragments from homologous chromosome pairs into numerous nanoliter compartments for enzymatic amplification within a microfluidic device. By adding exonuclease-resistant terminators to the reaction, the PTA technique generates smaller double-stranded amplification products that perform limited subsequent amplification. This causes the reaction to change from an exponential to a quasilinear process, increasing the amount of amplification that comes from the primary template and enhancing the coverage and uniformity of genome amplification. Currently, MDA-based products are well-established such the REPLI-g Single Cell Kit from Qiagen.

MALBAC

Reported in 2012 by Zong et al. (2012), MALBAC (multiple annealing and looping-based amplification cycles) is a scWGA method designed to mitigate bias associated with nonlinear amplification. MALBAC primers feature a common 27 nucleotide sequence at the 5′ end and 8 random nucleotides at the 3′ end. The amplification process begins with the hybridization of these 8 random nucleotides to the genomic DNA template at 0°C. Then, as the temperature is increased to 65°C, DNA polymerases with strand-displacement activity are used to create semiamplicons of varying lengths (0.5–1.5 kb). At 94°C, the semiamplicons are then denatured from the template. The semiamplicons are amplified further to produce entire amplicons with complemen-

Table 1. Comparison of mainstream WGA methods[a)]

WGA method	Amplification principle	Enzyme	Product length	Coverage (%)	Uniformity	Accuracy	Application	Reference
PEP-PCR	PCR-based	DNA polymerase	<2 kb	~50	+++	+	CGH, LOH, STR, etc.	(Zhang et al., 1992)
DOP-PCR	PCR-based	DNA polymerase	<2 kb	~45	+++	+	FISH, SNP, SSCP, etc.	(Telenius et al., 1992)
MDA	Isothermal amplification	phi29 DNA polymerase	<100 kb	~87	++	+++	NGS, SNV, SNP, STR, single-cell sequencing, etc.	(Dean et al., 2001)
eMDA	Isothermal amplification	phi29 DNA polymerase	<100 kb	72	+++	+++	CNV, SNV, single-cell sequencing, etc.	(Fu et al., 2015)
SISSOR	Isothermal amplification	phi29 DNA polymeras	<100 kb	~70	+++	++++	single-cell sequencing, haploid analysis, etc.	(Chu et al., 2017)
PTA	Isothermal amplification	phi29 DNA polymeras	~150 bp	>95	+++++	+++++	CNV, SNV, single-cell sequencing, etc.	(Gonzalez-Pena et al., 2021)
MALBAC	Combining isothermal amplification with PCR	Bst enzyme; Taq DNA polymerase	<2 kb	~93	+++	++	single-cell sequencing, NGS, STR, CGH, SNV, CNV, etc.	(Zong et al., 2012)
LIANTI	Linear amplification	Reverse transcriptase; T7 RNA polymeras	~400 bp	~97	++++	+++	CNV, SNV, CGH, single-cell sequencing, chromosome analysis, etc.	(Chen et al., 2017a)
META-CS	End-tagging-based PCR amplification	Q5 DNA polymerase	<2 kb	64	+++	++++	SNV, single-cell sequencing, etc.	(Xing et al., 2021)

a) CGH, comparative genomic hybridization; LOH, loss of heterozygosity; STR, short sequence repeat; FISH, fluorescence in situ hybridization; SNP, single nucleotide polymorphism; SSCP, single-strand conformational polymorphism; NGS, second generation sequencing; CNV, copy number variation; SNV, single nucleotide variation.

tary ends, forming hairpins when the temperature is lowered (to 58°C). Full amplicons form loops that may impede further amplification and cross-hybridization. Following five rounds of preamplification, oligonucleotides containing the common 27 nucleotide sequence are used as primers in PCR to exponentially amp up the entire amplicons. This process yields the micrograms of DNA required for next-generation sequencing. Additionally, commercial kits based on the MALBAC method have been created, such as the MALBAC single-cell WGA test (Qiagen).

MALBAC stands out as more than just a combination of DOP-PCR and MDA; it is fundamentally distinct due to its quasilinear amplification, which circumvents sequence-dependent bias associated with exponential amplification, thereby enhancing amplification uniformity. In the initial phase of MALBAC technology, multiple displacement reactions are employed to achieve a whole genome coverage of the amplification product up to 93%. MALBAC boasts both extensive coverage and uniform amplification, making it suitable for the genome-wide detection of SNPs and CNVs in a single cell (Hou et al., 2013; Zong et al., 2012). The technology's implementation has contributed to the advancement of clinical assisted reproductive technology (Yao et al., 2018). Furthermore, MALBAC demonstrates a low false negative rate in SNV detection. However, when compared with MDA technology, MALBAC does exhibit a higher false positive rate in SNV detection due to the lower fidelity of the current DNA polymerase used in comparison to the phi29 polymerase.

LIANTI

In contrast to previous WGA methods, such as the exponential PCR reaction with degenerate priming in DOP-PCR (Telenius et al., 1992), the strand-displacing DNA polymerase-driven exponential amplification of single-stranded DNA in MDA (Dean et al., 2001), and the quasilinear amplification in MALBAC (Zong et al., 2012) through looping-based amplicon protection followed by PCR all of which involve nonspecific priming and exponential amplification leading to bias and errors, a novel scWGA method was developed by Xie et al.'s team (Chen et al., 2017a) in 2017, known as linear amplification via transposon insertion (LIANTI). LIANTI uses the Tn5 transposase technology to sustain linear amplification during the whole genome amplification process. This method randomly slices genomic DNA, and then fills in the cut locations with a predetermined sequence. By utilizing the Tn5 transposase property, LIANTI inserts the T7 promoter into genomic DNA. The genomic DNA fragments labelled with T7 promoters are then linearly amplified into thousands of copies of RNAs using T7 RNA polymerase for IVT. RT and second-strand synthesis come next, producing double-stranded LIANTI amplicons that are prepared for DNA library construction. In comparison to earlier techniques (DOP-PCR, MDA, and MALBAC), LIANTI exhibits superior sensitivity and accuracy in identifying CNVs and SNVs, respectively. Additionally, LIANTI outperforms other WGA techniques (DOP-PCR, MDA, and MALBAC) with a 97% genome coverage and a 17% allele dropout rate.

META-CS

True SNV must be located at the same position on both DNA strands, while polymerase mistakes and DNA damage often happen at random on one of the two strands. Therefore, sequencing both complementary strands of double-stranded DNA (dsDNA) is essential to reduce false positives on single strands and improve accuracy. Multiplexed end-tagging amplification of complementary strands (META-CS), a revolutionary scWGA technique, was presented by Xie and colleagues in 2021 (Xing et al., 2021). Because of DNA complementarity, META-CS can clearly identify and amplify the two DNA strands in a one-tube reaction, reducing almost all false positives. De novo SNVs can be reliably identified using this method from a single cell.

META-CS is built upon the previously reported multiplexed end-tagging amplification method developed by this research team. Initially, a combination of 16 unique transposon sequences is mixed with Tn5 transposase in an equal molar ratio to create transposon complexes. These complexes randomly cut the genomic DNA from a single cell. Subsequently, genomic DNA fragments tagged by two random transposon sequences undergo denaturation through heating, releasing two single strands. To obtain strand-specific labelling, these strands are subsequently preamplification using two consecutive polymerase extension processes. After each polymerase extension reaction, exonuclease I is employed to eliminate excess primers. The products, separately amplified from the sense and antisense strands of the original DNA, can be differentiated by mapping them to the reference genome. As a result, SNVs are determined with information from both strands, significantly improving accuracy.

High-throughput scWGS methods

Since the introduction of the first scWGS method for profiling copy numbers in human tissues (Navin et al., 2011), the field of single-cell genomics has witnessed rapid progress over the past decade. Early techniques were confined to profiling a small number of cells at a time and were based on WGA chemicals (Chen et al., 2017a; Dean et al., 2001; Telenius et al., 1992). However, the study of tumor cell genome evolution and reconstruction often requires the analysis of mutation signatures in a large number of single-cell tumor genomes simultaneously. Low-throughput scWGS methods pose challenges due to their labor-intensive, costly, time-consuming nature, and limited efficiency. Recent developments in combinatorial indexing, nanowell, and microdroplet techniques have greatly increased cell throughput and decreased costs (Laks et al., 2019; Minussi et al., 2021; Vitak et al., 2017; Yin et al., 2019c). In the following sections, we will delve into a detailed discussion and comparison of high-throughput scWGS methods based on different strategies. The characteristics of several key high-throughput scWGS methods are summarized in Table S5 in Supporting Information.

Microfluidic-based high-throughput scWGS methods

(1) Direct library preparation method (DLP)

A direct DNA transposition single-cell library production (DLP) approach was presented by Zahn et al. (2017). This method creates indexed libraries straight from single cells. The DLP method utilizes a specially designed microfluidic device to capture and lyse single cells. Tn5 transposome complexes randomly fragment the genomic DNA of a single cell, tagging each fragment at the 5′ end with a distinct adaptor sequence. The index barcodes and sequencing adaptors are then added to both ends of the tagmented DNA inserts using eleven PCR cycles. Following indexing, the libraries are combined for multiplexed sequencing.

Prior to building libraries, early single-cell techniques used WGA to capture entire genomes (Gawad et al., 2014; Navin et al., 2011; Wang et al., 2014b). However, preamplification introduces amplification biases and reduces coverage uniformity, hindering the detection of CNVs (Macaulay and Voet, 2014; Wang et al., 2012; Zong et al., 2012). DLP is an example of the first direct production of single-cell libraries without preamplification using tagmentation. This method produces genomes with high uniform coverage and enables multiplexing of many cells, which makes it appropriate for high-throughput and reasonably priced CNV detection. Compared with DOP-PCR, DLP is more cost-effective (approximately $0.50 per cell versus $15 per cell) and time-efficient (2.5 h versus 3 d). However, the use of microfluidic devices limits the throughput of cells and necessitates a certain size of cells because very small cells may slip through traps unless the devices are specifically made for that type of cell. Large cells may also clog channels. Thus, additional optimization is still required.

(2) Microfluidic droplet method

Increasing the throughput of cellular sequencing faces challenges due to limitations in single-cell partitioning methods, difficulties in amplifying genomic DNA from single cells, and the complexity of enzymology steps for library preparation (Lan et al., 2017; Vitak et al., 2017). In response to these challenges, Andor et al. (2020) proposed a solution to enable large-scale scWGS. Their method makes use of a two-stage microfluidic droplet-based technique to automatically generate scWGS libraries with a high cell number. Similar to previous single-cell transcriptome investigations, microfluidic droplets are loaded with a barcoded hydrogel bead that labels DNA. To create cell beads (CBs), individual cells are first encased in a hydrogel matrix. Reagents for cell lysis and protein digestion are then added to these CBs to lyse and unpackage DNA. To create cell bead-gel bead (CBGB) emulsion, a second microfluidic chip is utilized in which a gel bead (GB) is functionalized with millions of copies of a distinct droplet-identifying barcode and co-encapsulated with the hydrogel CB and enzymatic reaction mix. CBGB dissolves to release contents after encapsulation. Genomic DNA fragments labelled with a sequencing adaptor and a barcode sequence are obtained by a two-step isothermal incubation process. After breaking and purifying the emulsion, the library is ready for Illumina sequencing. This microfluidic droplet-based cellular isolation technology can isolate tens of thousands of cells in a single experiment at a throughput that surpasses that of conventional DLP techniques.

Nanowell-based high-throughput scWGS methods

(1) High-throughput direct transposition scWGS method (DLP+)

When compared with preamplification-based techniques, the earlier DLP method, which made use of microfluidic devices, effectively decreased biases (Zahn et al., 2017). Despite the good performance of microfluidic-based DLP analysis, the usage of customised microfluidic devices limits cell size and presents obstacles to their general adoption and scalability. Additionally, some droplet-based methods face similar constraints on cell size (Andor et al., 2020). In response to these problems, Emma Laks and colleagues (Laks et al., 2019) developed the DLP+ platform, a higher-throughput direct transposition scWGS system built on commercially available "off the shelf" picoliter volume piezo-dispensing technology and commodity high-density nanowell arrays. In the DLP+ method, single cells are isolated through limiting dilution and dispensed into nanowell arrays. To achieve an almost flawless single-cell isolation rate, chosen cells are deposited into reaction chambers selectively using spotting software, which locates them inside the dispensing nozzle. A 10x inverted fluorescence microscope scans each nanowell chip to verify single-cell occupancy and gather data on the cell state. Before the library preparation reagents are spotted, imaging takes place, enabling the exclusion of doublets, empty wells, or contaminated cells from the procedure. Adding reagents, spinning, sealing, and heating the chip are the procedures involved in creating DLP+ libraries from unamplified single cells. Using standard Illumina procedures and HiSeq equipment, the resulting libraries are pooled during recovery and sequenced at the required coverage depth.

With its transparent dispensing nozzle and inbuilt camera, DLP+ provides a unique benefit that makes it possible to take high-resolution microscope images of objects prior to dispensation. By actively selecting individual cells, this function helps to prevent the sequencing of detritus or doublet cells. Furthermore, by methodically modifying a number of variables, including cell lysis volume and buffer type, transposase (Tn5) concentration, post-indexing PCR cycles, and cell lysis/DNA solubilization duration, DLP+ has refined the physical reaction determinants for producing high-quality libraries. The DLP approach is the foundation for these optimizations (Zahn et al., 2017). Moreover, DLP+ has shown genome coverage uniformity that is comparable to that of the microfluidic-based DLP approach, but at a throughput that is significantly higher, scaling from hundreds to tens of thousands of cells per experiment across a range of tissue types.

(2) Archival nanowell sequencing method (Arc-well)

The previously developed high-throughput scWGS methods, including DLP (Laks et al., 2019), SCI-seq (Vitak et al., 2017), and ACT (Minussi et al., 2021), have a common limitation-they require fresh or snap-frozen tissue samples, rendering them unsuitable for the analysis of archival formalin-fixed paraffin-embedded (FFPE) tissue samples. Addressing this challenge, a novel method called Arc-well (archival nanowell sequencing) was introduced by Wang et al. (2023a). To perform Arc-well, FFPE blocks are sectioned and deparaffinized to produce single-nucleus suspensions, which are subsequently used for FACS sorting. After the sorted nuclei are distributed into a 5,184-well nanowell chip, the nanowells can be imaged to select single cells and prevent doublets, deteriorated nuclei, and empty wells. Then, a five-step equal volume dispensing step is performed by using the ICELL8 cx system (TaKaRa Bio), which is used to dispense downstream reagents into nanowell chips. First, lysis reagents are dispensed to lysis selected nuclei and release the genomic DNA. Next, reagnets for labeling reaction (Tn5 transposome) and Tn5 inactivation are dispensed. Furthermore, by depositing dual indices (72×72 combinations) and amplifying the PCR result, every nanowell is given a distinct barcode combination. The barcoded libraries are then pooled and sequence on the Illumina platforms.

The acoustic cell tagmentation (ACT) technique was first presented by the researchers in 2021. This technique made use of acoustic liquid transfer (ALT) technology, direct tagmentation of genomic DNA, and FACS of single nuclei to enable high-throughput single-cell DNA sequencing at single-molecule resolution (Minussi et al., 2021). When comparing the ACT method with Arc-well, it was found that Arc-well exhibited higher throughput (1,900–2,600 cells per experiment), lower reagent costs, and reduced technical variability. Importantly, Arc-well demonstrated the capability to amplify degraded DNA fragments commonly found in archival FFPE tissues, making it compatible with such tissue samples.

Combinatorial indexing-based high-throughput scWGS methods

(1) SCI-seq

FACS is used in the SCI-seq method to sort individual cells into 96-well plates (Vitak et al., 2017). Subsequently, genomic DNA from a single cell undergoes random fragmentation by Tn5 transposase, and each resulting fragment is tagged with index 1 and an adaptor. The introduction of index 2 is achieved through a PCR reaction. Ultimately, the distinct libraries are combined for sequencing. Nucleosome depletion is used in a combinatorial indexing procedure by SCI-seq, which makes it possible to produce thousands of single-cell genome sequencing libraries at once. This approach also has the benefit of not requiring specialized microfluidics equipment or droplet emulsification procedures, in addition to its high throughput. However, it is noted that SCI-seq technology introduces a certain bias during the PCR amplification process.

(2) SCI-L3-WGS

To address amplification bias, Yin et al. (2019c) introduced sci-L3, a method that integrates combinatorial indexing with linear amplification. With the help of a 3-level indexing technique, sci-L3-WGS considerably increases LIANTI's throughput, allowing it to sequence at least thousands or even millions of cells per experiment while minimizing amplification biases. The sci-L3-WGS process is delineated into three key steps: (i) Tn5 transposase randomly cleaves genomic DNA from a single cell and attaches barcode 1 to each fragment. (ii) A second set of barcodes is ligated to the ends of DNA fragments, along with a T7 promoter positioned outside both barcodes. (iii) The introduced T7 promoter initiates IVT, followed by RT and second-strand synthesis. A third set of barcodes and UMIs are introduced during second-strand synthesis. Duplex DNA molecules can be prepared in accordance with conventional library preparation procedures. Each molecule contains three barcodes that identify the cell of origin. The sci-L3 strategy has a number of benefits over current methods and any straightforward combination of SCI-seq (Vitak et al., 2017) and LIANTI (Yin et al., 2019c). First off, using IVT, it accomplishes the same linear amplification as LIANTI. Second, because it uses three rounds of barcoding, its theoretical throughput surpasses one million cells per experiment at a cheap cost of library preparation (Cao et al., 2019). Third, sci-L3 is a flexible strategy for linear amplification combined with high-throughput cellular indexing; it also can be used for other single-cell sequencing analysis besides scWGS, such as single-cell RNA/DNA co-assays.

Applications of scWGS in biomedicine

With its ability to reveal differences in single-cell genomic architecture, scWGS technology is a potent tool that is used in many different fields, including tumor biology, somatic mutation and mosaicism, organismal development, germ cell mutation and development, fertility, and microbial research. It has become a major area of study in the life sciences. The applications of this technology in the fields of fertility and tumor biology will be the focus of the discussion that follows.

(1) Tumor biology

Tumor is a multifaceted and diverse disease characterized by genomic instability and the accrual of somatic mutations, and its Intratumoral heterogeneity poses a significant challenge to personalized cancer medicine. Traditional bulk sequencing methods have offered valuable insights into the genomic makeup of cancer; however, they often overlook the inherent heterogeneity within tumors, resulting in an incomplete portrayal of the disease. The advent of scWGS has proven instrumental in overcoming this limitation. By enabling the analysis of individual cancer cells at a single-molecule level, scWGS has exhibited significant potential in various facets of cancer research, such as elucidating intratumoral heterogeneity, interpreting the evolution of clonal processes, comprehending invasion and metastasis, investigating circulating tumor cells (CTCs), and evaluating treatment outcomes.

A study concentrating on breast malignancies revealed the first investigation of intratumoral heterogeneity utilizing scWGS based on DOP-PCR. This study identified subclonal lineages inside breast tumors using copy number changes (Navin et al., 2011). Subsequent scWGS studies, utilizing diverse cancer types such as ovarian (McPherson et al., 2016), bladder (Li et al., 2012), brain (Francis et al., 2014), renal (Xu et al., 2012), colorectal (Leung et al., 2017; Liu et al., 2017), liver (Hou et al., 2016), lung (Ferronika et al., 2017), and hematological (Gawad et al., 2014; Hughes et al., 2014a) cancers, have expanded our understanding of intratumoral heterogeneity at the levels of CNVs and SNVs. These investigations have revealed a correlation between tumor subtype and subclonal diversity in specific cases. For example, Baslan et al. (2020) conducted a comprehensive analysis of 2,086 breast cell genomes from 16 breast cancer samples using a DOP-PCR-based sequencing method. They observed that estrogen receptor-negative breast cancers exhibit higher subclonal diversity compared with estrogen receptor-positive breast cancers.

Phylogenetic analyses based on intratumoral heterogeneity profiles obtained through single-cell DNA sequencing (scDNA-seq) provide valuable insights into identifying driver mutations-genetic alterations that play a significant role in cancer development and progression. Analyzing the genomes of individual cancer cells allows researchers to pinpoint specific mutations driving tumor growth, offering crucial information for the development of targeted therapies that address these driver mutations. In a work by Wang et al. (2014b), hundreds of breast cells were profiled using a combination of targeted duplex single-molecule sequencing and scWGS. In two individuals with breast cancer, the researchers looked into mutational evolution and clonal diversity. Their research showed that SNVs gradually evolved, resulting in a high degree of clonal diversity. On the other hand, aneuploid rearrangements happened early in the genesis of tumors and stayed very stable during clonal growth. The investigation discovered many nonsynonymous mutations in genes linked to cancer, such as PIK3CA, CASP3, FBN2, and PPP2R5E, in a sample of invasive ductal carcinoma that was positive for oestrogen receptors. Interestingly, it is known that the most frequent driver mutation in luminal A breast tumors is PIK3CA (Ellis et al., 2012; Network, 2012).

CTCs, originating from primary tumors and entering the peripheral blood, have the potential to contribute to metastasis. ScWGS of CTCs presents a promising approach for noninvasive sampling of tumors, offering insights into noninvasive prognosis or even diagnosis. In a study by Riebensahm et al. (2019), scWGS was employed to analyze the mutation characteristics of genes in CTCs from breast cancer brain metastasis patients. The study identified mutated genes such as *TP53*, *ARID1A*, *CDH1*, and *TTN*, with *ARID1A*, involved in chromatin remodeling, highlighted as a potential druggable target. The MALBAC approach was employed by Ni et al. (2013) to examine the genomes of individual CTCs obtained from patients with lung cancer. The analysis revealed the presence of insertions/deletions (indels) and SNVs that are linked to cancer in the CTC exomes. This mutation information provided potential clinical guidance for personalized therapy. Additionally, CTCs have been utilized for noninvasive monitoring of treatment response (Dago et al., 2014).

In summary, scWGS stands as a groundbreaking technology in cancer research, providing a comprehensive understanding of the genomic landscape of individual cancer cells. By uncovering clonal evolution, identifying driver gene mutations, tracking chromosomal abnormalities, studying the tumor microenvironment, and detecting minimal residual disease, scWGS offers valuable insights for cancer diagnosis, prognosis, and targeted therapy. As scWGS technology continues to develop, it holds great promise for advancing personalized cancer medicine.

(2) Fertility

Preimplantation genetic diagnosis (PGD) and preimplantation genomic screening (PGS) for embryos created through *in vitro* fertilization (IVF) are two clinical uses for scWGS. This helps prevent the inheritance of harmful mutations and chromosomal abnormalities by enabling a thorough study of chromosomes. For this, a variety of genome analysis systems are used, including multiplex quantitative PCR, comparative genomic hybridization (CGH) arrays, and SNP arrays (Rubio et al., 2013; Tobler et al., 2014; Treff et al., 2012). scWGS technologies have improved conventional methods of analyzing embryo biopsies by enabling simultaneous identification of aneuploidy and mutations throughout the genome (Kumar et al., 2015; Treff et al., 2013; Wells et al., 2014). The rapid advancement of high-throughput sequencing methods has further decreased expenses and improved the accuracy and resolution of PGD/PGS at the chromosomal level. This approach holds promise for enhancing the accuracy of selecting healthy embryos during IVF procedures, improving the success rates of assisted reproduction, and reducing the risk of genetic disorders in newborns. Below, we describe several application examples of scWGS in PGD/PGS.

The use of scWGS in PGS and PGD during IVF has been shown in a number of research: Wells et al. (2002) utilized DOP-PCR-based WGA to perform scWGA on the first polar body, successfully detecting chromosomal abnormalities in embryos using CGH technology. Daina et al. (2013) conducted monogenic analysis on fourteen embryos for a family affected by Lynch syndrome, achieving successful double-factor PGD using the MDA method and leading to the birth of two healthy children. Hou et al. (2013) employed MALBAC-based sequencing technology to analyze the genomes of single human oocytes from eight healthy donors. They demonstrated how to accurately and cost-effectively select normal fertilized eggs for embryo transfer through MALBAC-based PGS during IVF. Huang et al. (2014) collected 23 frozen cleavage embryos from three pregnant women donors and performed single-cell CGH, SNP, and MALBAC sequencing for 24-chromosome aneuploidy analysis. MALBAC sequencing results showed a high concordance rate with CGH and SNP, indicating its application value in PGD/PGS. Shang et al. (2018) extended the application of MALBAC-scWGS to PGD/PGS detection of mitochondrial disorders, demonstrating the versatility of this technology in addressing various genetic conditions.

ScWGS has revolutionized PGD/PGS detection by enabling the analysis of individual cells within embryos. This powerful technique provides detailed information on chromosomal abnormalities, structural variations, and mutational landscapes. The ability to examine the genomic content of individual cells within embryos enhances the precision of genetic analysis, offering valuable insights for selecting embryos with the highest likelihood of success during IVF. As a result, scWGS has contributed to improving the success rates of IVF procedures.

Summary

The progress in single-cell genomics technology has not kept pace with that of transcriptomics, mainly due to challenges in achieving even genomic coverage during DNA capture. Nonetheless, single-cell genome sequencing has brought about significant insights into various previously inaccessible biological questions. This technology has found applications in diverse research fields, including somatic mutagenesis, understanding genome function, studying organismal development, and exploring microbiology. Single-cell genome sequencing shows great potential in clinical and translational research and practical applications, especially for the oncology and assisted reproduction field.

Chapter 3 Single-cell epigenome sequencing

The epigenome of a cell regulates its cell type-specific gene expression. Understanding epigenetic variations is crucial to reveal transcriptional mechanisms that determine tissue and cellular heterogeneities during development, disease formation, and progression. The epigenome involves a variety of precisely regulated epigenetic features, such as nucleic acid methylation, chromatin states, nucleosome positions, histone modifications (HM), TF bindings, and high-order chromatin structures. These features interact with one another to influence nearby genome activity without changing DNA sequences, which further controls cellular activities and results in heritable phenotypes. Single-cell epigenome sequencing techniques, as well as corresponding computational analysis methods, have been developed and widely used in many research areas, especially in cancer immunology, embryonic development, and neurobiology. In this chapter, we survey the recent advances in sequencing techniques and computational tools developed for single-cell epigenome data analysis.

Techniques for sequencing the single-cell epigenome

Methylation

Methylation is a type of epigenetic modification that adds methyl groups (CH3) to nucleic acids. In vertebrates, DNA is mostly methylated at the carbon atom occupying the fifth position of the cytosine ring (5mC). The majority of cytosine methylation generally occurs in the context of CpG dinucleotides, which usually group in CpG-dense regions called CpG islands (CGIs). These regions show high associations with gene promoters, resulting in methylation-regulated gene expression in a *cis* manner. There are also other DNA modifications, such as 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC), which are intermediate products of DNA demethylation and play critical roles in many biological processes. Based on whether they use the bisulfite conversion or not, two mainstream types of sequencing methods have been adopted for profiling DNA methylation at the single-cell level. Furthermore, N^6-methyladenosine (m^6A) is also an abundant RNA methylation that affects RNA regulation and cellular functions. Another technique for profiling m^6A from RNA at the single-cell level has been developed as well.

Over the past decades, bisulfite sequencing has become the gold standard for profiling genome-wide DNA methylation. With the sodium bisulfite treatment, the unmethylated cytosines are deaminated to uracil, while the methylated cytosines remain unchanged. In the following PCR amplification and sequencing, the unmethylated cytosines are then read as thymine, while the methylated cytosine is still read as cytosine. The efficiency of bisulfite treatment reaches about 95%, and the readout of bisulfite-based sequencing methods achieves single base-pair resolution, which enables them to become the dominant methods. Whole-genome bisulfite-sequencing (WGBS) (Cokus et al., 2008) can cover almost all the CpG sites of the whole genome, but the requirement for very deep sequencing makes it costly. Then, reduced representation bisulfite sequencing (RRBS) (Gu et al., 2010) has been developed as a cost-efficient method. It utilizes restriction enzyme (MspI) digestion and size fractionation to enrich CpG-dense regions so that it reduces the complexity and size of the sequence library.

To overcome the massive loss of DNA when detecting methylation in single cells, the single-cell RRBS (scRRBS) (Guo et al., 2013) protocol was introduced. It integrated all of the experimental processes in a single-tube reaction without any purification steps before the bisulfite conversion process, then performed two rounds of PCR amplification and deep sequencing. To remove the PCR amplification bias, UMIs were introduced by quantitative RRBS (Q-RRBS) (Wang et al., 2015). To avoid the bisulfite-induced loss of intact sequencing templates and avoid amplification bias, post-bisulfite adaptor tagging (PBAT) was adopted in scBS-seq (Clark et al., 2017; Smallwood et al., 2014) and scPBAT (Kobayashi et al., 2016). scWGBS (Farlik et al., 2015) implements the PBAT but without the requirement for the preamplification step, which is suitable for high-throughput analysis at low sequencing coverage. Genome-wide CpG coverage is not always needed and expensive; therefore, single-cell locus-specific bisulfite sequencing (SLBS) (Gravina et al., 2015) can be a cheaper choice and is able to directly detect epimutations in DNA methylation patterns. With the prosperous single-cell barcode or separation techniques invention for high-throughput sequencing, such as microfluidics device, combinatorial index, and nucleus sequencing, they were adopted in microfluidic diffusion-based reduced representation bisulfite sequencing (MID-RRBS) (Ma et al., 2018), single-cell combinatorial indexing for methylation analysis (sci-MET) (Mulqueen et al., 2018) and single-nucleus methylcytosine sequencing (snmC-seq) (Luo et al., 2017), respectively. To enrich more regions where CpG methylation is functionally relevant, including promoters, CpG islands, CTCF insulators, and enhancers, a single-cell extended representation bisulfite sequencing (scXRBS) (Shareef et al., 2021) was established by leveraging an early barcoding step for high sensitivity and sample multiplexing.

A limitation of bisulfite treatment is that unmethylated cytosine, 5fCs, and 5caCs, are all converted to uracil and cannot be discriminated from each other, which hampers the investigation of DNA demethylation. Single-cell methylase-assisted bisulfite sequencing (scMAB-seq) (Wu et al., 2017b) was established to solve this problem by pretreatment of the DNA with the CpG methylation enzyme M.SssI, which converts only the cytosines to 5mCs, protects Cs but not 5fCs and 5caCs, and enables direct detection of 5fCs and 5caCs as uracils.

Besides bisulfite treatment, methylation-sensitive restriction enzymes (MSREs) are also widely used in detecting and sequencing DNA methylation. Restriction enzyme-based single-cell methylation assay (RSMA) (Kantlehner et al., 2011) is easy to implement, but its results are reported by gel electrophoresis, and it is not a quantitative method. A similar method, single-cell restriction analysis of methylation (SCRAM) (Lorthongpanich et al., 2013), also detects the methylation by MSREs but uses real-time quantitative PCR (RT-qPCR) as readout. Both of these methods fail to distinguish between heterozygous and hemizygous methylated alleles in diploid cells. Along with SCRAM and single-cell genotyping by next-generation sequencing (NGS), single-cell analysis of genotype, expression, and methylation (sc-GEM) (Cheow et al., 2016) allows for a more reliable assessment of methylation status at specific sites. Genome-wide CGI methylation sequencing for single cells (scCGI-seq) (Han et al., 2017) achieved high single-cell CGI coverage, which extended the use of MSREs from a limited number of loci to CGIs at the genome-scale. To allow genome-wide detection of 5hmC marks in single cells, the restriction endonuclease AbaSI was used in single-cell hydroxymethylation sequencing (scAba-seq) (Mooijman et al., 2016). Without using MSREs, reporter of genomic methylation (RGM) (Stelzer et al., 2015) adopted a fluorescent reporter system, which allows for visualization and tracing of dynamic changes in DNA methylation.

Apart from these two conventional methods, the enzyme conversion-based methods have emerged as less damaging alternatives to bisulfite treatment and thus have been applied to single-cell analysis. EM-seq identifies 5mC and 5hmC by using two sets of enzymatic reactions. The initial reaction involves TET2 and T4-BGT converting 5mC and 5hmC into products resistant to deamination by APOBEC3A. Subsequently, the second reaction, employs APOBEC3A to deaminate unmodified cytosines, transforming them to uracils (Vaisvila et al., 2021). Recently, sciEM combined single-cell combinatorial indexing with enzymatic conversion marks significant advancement as the first non-bisulfite single-cell DNA methylation sequencing method (Chatterton et al., 2023). Similar strategies have also been adopted for RNA methylation detection. Global RNA m^6A profiling reveals its functions in gene expression control, physiological processes, and disease states. Deamination adjacent to RNA modification targets (DART-seq) utilizes a fusion protein consisting of the m^6A-binding YTH domain tethered to the cytidine deaminase APOBEC1 (APOBEC1-YTH) to conduct C-to-U editing at cytidine residues. DART-seq is antibody-free, which allows for mapping m^6A from ultra-low-input amounts of RNA. Therefore, the same group established the single-cell DART-seq (scDART-seq) (Tegowski et al., 2022) to identify RNA m^6A sites in single cells.

Chromatin accessibility and nucleosome positioning

Chromatin accessibility is a widely studied characteristic of the eukaryotic genome. Open chromatin is a necessary condition for DNA to interact with other factors, such as TFs or non-coding RNAs, which play crucial roles in remodeling chromatin or initiating transcriptions. Also, the nucleosome comprises 8-unit histones and is wrapped with naked DNA to form chromatin. The movement of nucleosomes on the genome, or nucleosome positioning, affects chromatin accessibility. At the bulk level, assays for transposase-accessible chromatin (ATAC-seq) (Buenrostro et al., 2013) and Deoxyribonuclease I digestion (DNase-seq) (Song and Crawford, 2010) have been widely used to reveal that chromatin accessibility is a key component of the epigenetic landscape. The dynamics of chromatin accessibility drive cell differentiation and precise gene regulation. Profiling and analyzing chromatin accessibility at the single-cell level can help reveal the nature of cell heterogeneity and gene expression.

ATAC-seq and DNase-seq have been applied to single cells, which can explore the different chromatin states and cell heterogeneity in massive cells. scATAC-seq (Buenrostro et al., 2015) combines microfluidics and Tn5 tagmentation with sequencing barcodes, while scDNase-seq utilizes FACS to sort single cells and digest them with DNase I. The scDNase-seq can detect more DNase I hypersensitive sites (DHSs) with specific properties related to gene expression. However, both methods have relatively low cell throughput due to the microfluidic equipment. To improve the cell throughput in a single experiment, scATAC-seq in small volumes (μATAC-seq) (Mezger et al., 2018) integrates fluorescence imaging and addressable reagent deposition across a parallel nano-well array to improve the cell throughput to ~1,800 cells per chip and yield higher enrichment. Another multiple index barcode method was introduced to them in single-cell profiling of chromatin accessibility by combinatorial cellular indexing (sci-ATAC-seq) (Cusanovich et al., 2015) and indexing single-cell DNase sequencing (iscDNase-seq) (Gao et al., 2021b). These approaches significantly improved the cell throughput to ~15,000 cells. Furthermore, droplet-based single-cell combinatorial indexing for ATAC-seq (dsciATAC-seq) (Lareau et al., 2019) integrated droplet-microfluidics-based method and combinatorial indexing, which makes profiling chromatin accessibility in ~500,000 single cells possible. Also, the single nucleus assay for transposase-accessible chromatin using sequencing (snATAC-seq) (Muto et al., 2021) only uses the cell nucleus for sequencing, which alleviates the mitochondrial contamination to yield higher quality cells and lower noise.

Nucleosome organization and positioning are also involved in forming chromatin compaction and accessibility. Single-cell micrococcal nuclease sequencing (scMNase-seq) (Lai et al., 2018) adopts FACS sorting, lysis, and digestion by MNase to build the library to profile genome-wide nucleosome positions. It reports cell heterogeneity of nucleosome positioning and nucleosome spacing at DHSs.

Histone modification and transcription factor binding

Different HMs indicate different chromatin states and activity of chromatin states, which also affect TF binding and transcription. Antibody-based methods, such as chromatin immunoprecipitation assays with sequencing (ChIP-seq) (Kim and Ren, 2006), have been widely used to profile HMs and TFs landscape on the whole genome. Droplet-based chromatin immunoprecipitation followed by sequencing (Drop-ChIP) (Rotem et al., 2015) and later single-cell ChIP-seq (scChIP-seq) (Grosselin et al., 2019) first separate cells into droplets that contain lysis buffer and MNase, and then barcode them before the immunoprecipitation step. They increase the efficiency of the pull-down step and give low background results. To improve the read number per cell, simultaneous indexing and tagmentation-based ChIP-seq (itChIP-seq) (Ai et al., 2019) adopted the Tn5 transposase-based tagmentation coupled with simultaneous addition of primers for barcoding and PCR amplification. It achieves ~9,000 reads per cell, close to that in scATAC-seq assays. Due to the low affinity and efficiency of antibodies, all these methods are used to profile HMs instead of TFs.

Cleavage under targets & release using nuclease (CUT&RUN) (Skene and Henikoff, 2017) utilizes chromatin immune-cleavage on native chromatin, which is a convenient and efficient low-input method. It has also been adapted to the following similar techniques, including single-cell chromatin integration labeling (scChIL-seq) (Harada et al., 2019), single-cell chromatin immune-cleavage sequencing technique (scChIC-seq) (Ku et al., 2019), combinatorial barcoding and targeted chromatin release (CoBATCH) (Wang et al., 2019b), antibody-guided chromatin tagmentation sequencing (iACT-seq) (Carter et al., 2019), ultra-low-input cleavage under targets and release using nuclease (uliCUT&RUN) (Patty and Hainer, 2021), single-cell cleavage under targets and tagmentation (scCUT&Tag) (Bartosovic et al., 2021), and indexing single-cell immune-cleavage sequencing (iscChIC-seq) (Ku et al., 2021). In particular, scChIC-seq, uliCUT&RUN, and iscChIC-seq use the protein A-micrococcal nuclease (pA-MNase) as the cleavage enzyme, and others use the Tn5 transposase-protein A (pA-Tn5) because of the release of MNase-cleaved fragments into the supernatant, which is not suitable for single-cell platforms. Interestingly, Tn5-based approaches, including CoBATCH, uliCUT&RUN, and scCUT&Tag, profile not only the histone modifications but also several abundant TFs, such as RNA polymerase II (POL II), NANOG, OLIG2, and RAD21.

Single-cell DNA adenine methyltransferase identification (scDamID) (Kind et al., 2015) was applied to the detection of how the chromosomes are spatially organized inside interphase nuclei. DNA adenine methyltransferase (Dam) methylates adenines that are adjacent to positions where the protein of interest interacts with the DNA. These methylated adenines are amplified by PCR and identified by NGS. Combining single-cell DamID with messenger RNA sequencing (scDam&T-seq) (Rooijers et al., 2019) successfully profiled the RING1B binding sites paralleling the transcriptome, providing a powerful tool to identify protein-mediated mechanisms that regulate cell-type-specific transcriptional programs in dynamic processes and heterogeneous tissues.

3D genome structure

Chromatin is spatially and structurally organized and compartmentalized in the cell nucleus, contributing to the effects of *cis*-regulatory elements (CRE) and *trans*-regulatory factors. Chromosome conformation capture (3C) (Hagège et al., 2007) detects genomic regions located in close proximity to each other. With the continuous development of conformation-based techniques, high-throughput sequencing-based Hi-C has enabled genome-wide chromatin interaction detection. Similar to other single-cell sequencing approaches, the isolation or barcoding of individual cells is a primary task for single-cell Hi-C (scHi-C) (Nagano et al., 2013). scHi-C reduces the scale of the traditional Hi-C protocol and sorts cells into multi-well plates for tagmentation. Single-nucleus Hi-C (snHi-C) (Flyamer et al., 2017) amplifies the entire genome and eliminates the biotin fill-in step. Diploid chromatin conformation capture (Dip-C) (Tan et al., 2018) simplifies the experimental protocols with a tagmentation-based strategy. Combinatorial indexing was introduced in Single-cell combinatorial indexed Hi-C (sciHi-C) (Ramani et al., 2017), avoiding the need to isolate cells. To capture long-range and higher-order interactions that are limited by proximity ligation, single-cell split-pool recognition of interactions by tag extension (scSPRITE) (Arrastia et al., 2022) detects both inter- and intra-chromosomal interactions and more DNA contacts per cell.

The key distinctions, limitations, and biological materials used in the original research of the reviewed techniques are summarized in Figure 6 and Table S6 in Supporting Information. Several challenges need to be overcome in the future. First, due to the low rate of DNA capture and lower DNA content than RNA in a single cell, single-cell epigenome data is currently highly sparse. Second, existing methods still have difficulty detecting the precise binding location of TFs, particularly for TFs that are not evenly distributed over the whole genome. Third, the elaboration of the mechanism of gene regulation from DNA to cell states and phenotypes continues to demand the further development of single-cell multi-omic approaches.

Computational methods for single-cell epigenome data

Reads preprocessing, quality control, and quantification

The read adaptor trimmers and mappers, which are designed for bulk tissues, can also be used for single-cell reads. Fastp (Chen et al., 2018), and Trimmomatic (Bolger et al., 2014) are used for removing the adapter sequence to facilitate the read mapping. For DNA methylation data, especially data generated by bisulfite-based methods, Bismark (Krueger and Andrews, 2011), BSMAP (Xi and Li, 2009), and Bsseeker (Chen et al., 2010) were adopted to map the reads to the genome. Bisulfite conversion induces largely depleted cytosines of the genome sequences, which causes multiple mapping sequencing reads, and this situation is more serious when it comes to single-cell data. The scBS-map (Wu et al., 2019a) was developed by remapping chimerical reads, which is the majority of the unmapped reads, with a local alignment approach, and dramatically improving the overall mapping efficiency. For scATAC-seq or other non-converted DNA sequences, BWA (Li and Durbin, 2009), bowtie2 (Langmead et al., 2019), and minimap2 (Li, 2018) were widely used to perform the mapping. Recently, chromap has brought pseudo-alignment to DNA mapping, which significantly improves the mapping efficiency with comparable mapping rate, and has been adopted in several analysis pipelines.

For quality control (QC), FastQC is often used to control the quality at the reading level. Limitations on the number of mapped reads and mitochondrial reads per cell filter out low-quality cells. For single-cell DNA methylation, the count matrices are built from cytosine summary tables or any custom-defined features of interest. The methylation status of cytosines in CG, CH, or both genomic contexts in every feature is counted and summarized in the matrices. MethylStar (Shahryary et al., 2020) and EpiScanpy (Danese et al., 2021) both have a built-in function for quantifying the methylation reads. BPRmeth introduced generalized linear model (GLM) regression to quantify methylation profiles. For scATAC-seq, the count matrices take mapped BAM files or fragment files, like 10x Cell Ranger output, as input. There are two mainstream ways to define the features. The first solution is that the cells that pass read-level QC are merged to call peak with software that is used in bulk, like MACS2 (Zhang et al., 2008) or chromHMM (Ernst and Kellis, 2012). The peak file is regarded as the region of interest and is used to count the reads in each peak. This solution significantly reduces the feature number that accelerates the downstream analysis but may lose the information and heterogeneity of the rare cell population. Dr. seq2 (Zhao et al., 2017), MAESTRO (Wang et al., 2020), scitools (Sinnamon et al., 2019), APEC (Li et al., 2020a), and Signac (Stuart et al., 2021) use the merged cell peak as features. Another solution is to count the reads with a segmented genome or so-called bin-based. SnapATAC (Fang et al., 2021) adopts this strategy to capture the rare population but generates numerous

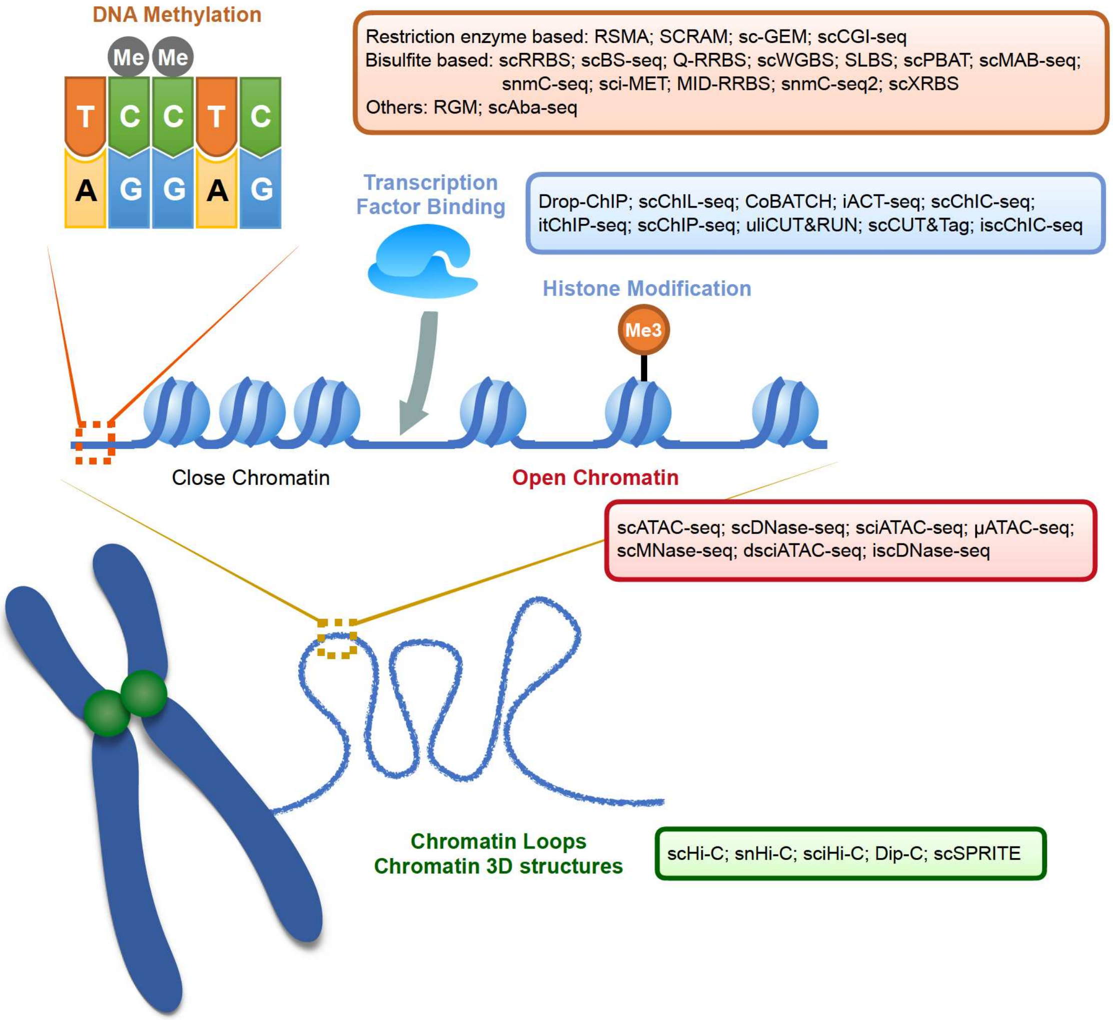

DNA Methylation
Me Me
T C C T C
A G G A G
Restriction enzyme based: RSMA; SCRAM; sc-GEM; scCGI-seq
Bisulfite based: scRRBS; scBS-seq; Q-RRBS; scWGBS; SLBS; scPBAT; scMAB-seq; snmC-seq; sci-MET; MID-RRBS; snmC-seq2; scXRBS
Others: RGM; scAba-seq
Transcription Factor Binding
Drop-ChIP; scChIL-seq; CoBATCH; iACT-seq; scChIC-seq; itChIP-seq; scChIP-seq; uliCUT&RUN; scCUT&Tag; iscChIC-seq
Histone Modification
Me3
Close Chromatin
Open Chromatin
scATAC-seq; scDNase-seq; sciATAC-seq; μATAC-seq; scMNase-seq; dsciATAC-seq; iscDNase-seq
Chromatin Loops
Chromatin 3D structures
scHi-C; snHi-C; sciHi-C; Dip-C; scSPRITE

Figure 6. Overview of singe-cell epigenome sequencing techniques. Single-cell epigenome sequencing is mainly performed at four levels, namely DNA modification, transcription factor binding or histone modification, chromatin accessibility and chromatin three-dimensional structure. DNA modification mainly refers to DNA methylation, and the sequencing methods for DNA modification are mainly based on restriction enzyme digestion and bisulfite conversion, which are shown in orange in the figure. The sequencing methods for transcription factor binding and histone modification are shown in blue. The sequencing methods for chromatin accessibility are shown in red. The sequencing methods for chromatin three-dimensional structure are shown in green.

features that need to be carefully filtered in the downstream analysis. The count matrices for scATAC-seq are often binarized because only double DNA strands are in a cell.

Imputation

As previously mentioned, the epigenome data from single cells is extremely sparse, which impacts the sensitivity and accuracy of downstream analysis for biological findings. Many methods have been developed for predicting and fulfilling the missing values as a result of bias from techniques.

For DNA methylation, predicting missing methylation states and improving the incomplete CpG coverage is critical to analyzing genome-wide methylation status. DeepCpG (Angermueller et al., 2017) utilizes convolutional neural networks (CNN) to learn the associations between DNA sequence features and methylation states between neighboring CpG sites, both within a cell and across cells. MOFA (Argelaguet et al., 2018) and MOFA+ (Argelaguet et al., 2020) infer an interpretable low-dimensional data representation with PCA to impute the missing values and assays. MELISSA (Kapourani and Sanguinetti, 2019), scMET (Kapourani et al., 2021), and Epiclomal (P. E. de Souza et al., 2020) use Bayesian mixture models to leverage similar methylation patterns in similar cells and impute missing values.

For scATAC-seq, ChromA (Gabitto et al., 2020) also adopts a Bayesian statistical approach with hidden semi-Markov models (HSMM) to overcome the sparsity from scATAC-seq data. ScOpen integrates an unsupervised learning model based on a non-negative matrix factorization (NMF), which does not require making assumptions about the data distribution. AtacWorks (Lal et al., 2021) uses the ResNet (residual neural network) architecture to train a deep learning model from high-quality bulk ATAC-seq datasets and predicts improved signal tracks at the base-pair level and the accessible genomic locations with noisy scATAC-seq tracks. SCATE (Ji et al., 2020b) integrates co-activated peaks, similar cells, and publicly available bulk data to predict the signals of each peak. These imputation methods also enhance cell clustering performance.

For scHi-C, scHiCluster (Zhou et al., 2019a) considers chromosome interactions as a network and uses the random walk algorithm to propagate the smoothed interactions to tackle the sparsity of data. HiCImpute (Xie et al., 2022) considers the spatial dependencies of 2D data structure and borrows information from similar single cells and bulk data. scHiCEmbed (Liu and Wang, 2022) borrows the scHiCluster's result and uses graph auto-encoders to learn node embeddings, which enables the imputation of the chromosome contact matrices and topologically associating domains (TAD) detection. Higashi (Zhang et al., 2022e) transforms the scHi-C data into a hypergraph and imputes the scHi-C contact maps by predicting missing hyperedges within the hypergraph. Another imputation task for scHi-C is to reconstruct the 3D genome structures. Si-C (Meng et al., 2021) applies the Bayesian theory framework to reconstruct genome 3D structures from scHi-C data. SCL (Zhu and Wang, 2019) regards the 3D structure of a chromosome as beads-on-a-string and reconstructs the structure inside a 3D cubic lattice, and uses 2D Gaussian imputation to estimate the propensity for the bead-pairs without scHi-C contacts. Also, a data-driven method, SIMBA3D (Rosenthal et al., 2019) first utilizes bulk Hi-C data to aid in recovering the interactions missed in scHi-C contact maps, then infers 3D chromosome structures with a generalized Bayesian framework.

Clustering

Clustering similar cells together assigns identities to cells to better find the rare cell population, understand the gene regulatory patterns in specific cell states, and alleviate the noise signals. The clustering algorithms, such as tSNE (Laurens and Hinton, 2008), UMAP (McInnes et al., 2018), graph abstraction (Wolf et al., 2019), Louvain clustering (Fortunato, 2009), Leiden clustering (Guo et al., 2019), and diffusion pseudotime (Haghverdi et al., 2016), which are used in single-cell transcriptomes, have been applied to single-cell epigenomes too. ALLCools (Liu et al., 2021), EpiScanpy (Danese et al., 2021), Signac (Stuart et al., 2021), ArchR (Granja et al., 2021), SnapATAC (Fang et al., 2021), and other analysis pipelines integrated these algorithms as built-in functions to facilitate easy clustering of cells.

Although the clustering algorithms used in the single-cell epigenome are similar to those used in single-cell transcriptome data, single-cell epigenome data suffers from more sparse and numerous features. To overcome sparsity, the imputation methods mentioned in the last section can be used to improve the clustering performance by fulfilling the missing features as well as keeping the cell heterogeneity. scABC (Zamanighomi et al., 2018) tries to alleviate the noise from cells with low sequencing depth by implementing a weighted version of the K-medoids clustering algorithm, which gives a low weight to the low sequencing depth cells.

Another difference in clustering algorithms for single-cell epigenome data from single-cell transcriptome data is the reduction of features or dimensions. PCA is the most commonly used method to reduce the dimensions of cluster features. Seurat v3 (Stuart et al., 2019) incorporates latent semantic indexing (LSI) on the scATAC-seq feature count matrix to reduce the dimensionality. CisTopic (Bravo González-Blas et al., 2019) uses LDA with a collapsed Gibbs sampler to identify the cis-regulatory topics. It also facilitates the prediction of TF binding sites and chromatin states. PeakVI (Ashuach et al., 2022) employs a deep generative model to learn a probabilistic low-dimensional representation. ScVAEBGM (Duan et al., 2022) integrates a Variational Autoencoder (VAE) with a Bayesian Gaussian-mixture model (BGM) to process scATAC-seq data. It takes advantage of the BGM to estimate cluster numbers from data.

Besides only using the information from single-cell epigenome data, borrowing the information from sequence features, bulk datasets, and single-cell transcriptome datasets also helps with the task of clustering. Some methods developed for multi-ome experiments, such as MAPLE (Uzun et al., 2021), scAI (Jin et al., 2020a), LIGER (Welch et al., 2019), scMC (Zhang and Nie, 2021), and scGCN (Song et al., 2021c), improved the clustering performance by integration with scRNA-seq. chromVAR (Schep et al., 2017), BROCKMAN (de Boer and Regev, 2018), scFAN (Fu et al., 2020), and scBasset (Yuan and Kelley, 2022) consider the sequence features, including motifs or specific k-mer, to reduce the dimension from peak level to k-mer level or TF level. Furthermore, CellWalkR (Przytycki and Pollard, 2022) integrates scATAC-seq with cell type labels and bulk epigenetic data to better illustrate the CREs active in specific cell types. SCRIP (Dong et al., 2022) incorporates many bulk ChIP-seq data sets, which also use peak set similarity to convert the feature matrix from the peak count to TF count. These methods not only enhance the clustering performance but also provide biological information on which peaks or sequence features are important to specific regulatory factors.

For scHi-C data, SCL and scHiCEmbed increase the clustering performance by alleviating the sparsity of data with imputation. Recently, scHiCStackL (Wu et al., 2022) proposed a computational framework by constructing a two-layer stacking ensemble model for classifying cells and outperformed other methods on the task of clustering cell types.

Cell type annotation and trajectory inference

Even while single-cell approaches allow for the parallel analysis of genomic data among numerous cells, we usually need to know the cell types or differentiation stages of each cluster. Annotating cells using single-cell epigenome data typically requires inferring the gene activity to assist in distinguishing cell types. This is in contrast to scRNA-seq, which can identify cell states by gene markers.

ArchR and MAESTRO both provide statistical models to infer the gene score at the cluster level from scATAC-seq peaks. ArchR incorporates the exponential decay model while accounting for the expanded gene body and gene border. MAESTRO also uses an exponential decay model but considers the exons of each gene and removes the effects of nearby genes. Garnett borrows the methods of calculating gene activity scores from Cicero (Pliner et al., 2018) and applies their predefined markup language and pre-trained classifier to scATAC-seq data. Besides using inferred gene scores as markers to annotate cells, another way is to use well-annotated bulk data as references. SCRAT (Ji et al., 2017) compiles a regulome database consisting of ENCODE (de Souza, 2012) DNase-seq profiles from a wide variety of cell types to infer the likely cell type of each cell. Moreover, MAESTRO not only uses the data from the ENCODE project but also the data from the Cistrome Data Browser (Mei et al., 2017; Zheng et al., 2019; Zheng et al., 2020), which has collected the most comprehensive previous public DNase-seq and ATAC-seq datasets.

The transcription of RNA takes time, therefore single-cell epigenome data is more sensitive in capturing cell differentiation events than scRNA-seq. To infer cellular trajectories, STREAM (Chen et al., 2019b) first uses PCA to extract the most informative features. Modified locally linear embedding (MLLE), a non-linear dimensionality reduction technique, is then used to project cells into a low-dimensional space before the implementation of the Elastic Principal Graph. MIRA (Lynch et al., 2022) uses topic modeling to infer cell states and represent those states in an interpretable latent space, allowing for the inference of cell state trees and the identification of important regulators of branch point fate decisions. Also, many pipeline tools, like EpiScanpy and Signac, incorporate PAGA (Wolf et al., 2019) or Monocle (Trapnell et al., 2014) to infer the cell trajectories. However, understanding the biological system as well as the underlying assumptions is necessary when modeling trajectories using single-cell data. Therefore, to interpret the results of trajectories, well-annotated clustering is often a requirement.

Differential analysis and features selection

With differential analysis, it is crucial to determine which features are related to particular cell states. This approach connects cell states and phenotypes to genomic regions or CREs. A recent report claimed that the Wilcoxon rank-sum test outperforms other differential test methods in large-sample-size data because it does not require any assumptions (Li et al., 2022j). In fact, the Wilcoxon rank-sum test is the most commonly used test method for detecting differential expression genes in the majority of pipeline tools.

Although it is not difficult to perform the differential analysis with current tools, a tricky thing is how to define the useful features of single-cell epigenome data. Bin-based methods and peak-based methods are adopted for scATAC-seq. scMET aggregates the input data within regions, such as promoter regions or enhancers. These genome features rely on the aggregation of individual regions. Recently, a deep generative model PeakVI infers a representation for each cell in high-dimensional, which enables statistically robust inference of single-region-level differential accessibility and cell state annotation.

Gene regulation inference

Inferring TF activity using single-cell epigenomics data is an intriguing potential application that provides clues on how epigenetics influences gene expression and cell phenotypes. ChromVAR, scFAN, scBasset, TRIPOD (Jiang et al., 2022b), and SCRIP all support inferring TF activity at the single-cell level from scATAC-seq data. ChromVAR infers TF activity by estimating the gain or loss of accessibility within peaks sharing the same TF motifs. scFAN pre-trains deep learning-based models on genome-wide bulk ATAC-seq, DNA sequence, and ChIP-seq data and applies the model to single-cell ATAC-seq to predict TF binding in individual cells. scBasset introduces CNNs to leverage the DNA sequence information underlying scATAC-seq peaks to achieve TF activity inference. TRIPOD combines scRNA-seq, scATAC-seq, and DNA sequence features to infer the TF activity related to gene expression associations, accounting for literature-based knowledge. However, the DNA sequence features, such as motifs, lose the cell-type-specific information of TFs and cannot distinguish between TFs with similar motifs, such as the GATA family. Recently, SCRIP incorporated thousands of bulk-level ChIP-seq datasets and scATAC-seq to infer the TF activity based on the peak set similarity, which successfully distinguished the similar motif TF activity at the single-cell level.

Although scATAC-seq identifies the open chromatin regions as CREs, how the CREs link distal regulatory elements with their target genes is also a key question in gene regulation. Cicero samples and aggregates similar cells to quantify correlations between putative CREs and links CREs to target genes based on the correlation using a graphical lasso model. To alleviate uncorrelated technology noise and false positive results in Cicero, JRIM (Dong and Zhang, 2021) uses the group lasso penalty to find similar patterns of sparsity across all the regulatory networks to reconstruct the cis-regulatory interaction networks. To accurately identify the loci of key CREs of different cell types, scEpiLock (Gong et al., 2022) adopts a CNN model to detect the chromatin accessibility regions and refine the peak boundary using gradient-weighted class activation mapping (Grad-CAM). Similarly, DIRECT-NET (Zhang et al., 2022c) adopts eXtreme Gradient Boosting (XGBoost) to identify functional CREs and infer the TF binding sites with known motif patterns. The aforementioned methods successfully link the CREs to target genes, DeepTFni (Li et al., 2022d) implements a GNN with a variational graph auto-encoder (VGAE) to infer TF regulatory networks, which can show the relationship between TFs and TFs. SMGR (Song et al., 2022b) takes both scRNA-seq and scATAC-seq as input and utilizes a generalized linear regression model to identify the latent representation of consistently expressed genes and peaks, as well as identify co-regulatory mechanisms.

ScHi-C allows for exploring gene regulation patterns in a 3D manner at the single-cell level. Topologically associating domains (TAD) segment the genome based on the 3D genome structure. There are more DNA-DNA interactions within TADs than between one TAD and other TADs. deTOKI (Li et al., 2021c) can predict TAD-like domain structures at the single-cell level with NMF from sparse scHi-C data. Chromatin loops are smaller structures that link CREs to target genes physically. SnapHiC (Yu et al., 2021) and SnapHiC2 (Li et al., 2022i) enable identifying chromatin loops at 10 kb resolution with a random walk with restart (RWR) algorithm from scHi-C data.

Multi-function analysis pipelines

The selection and organization of these tools to effectively extract the underlying information from data have become a challenge with the development of numerous computational approaches for single-cell epigenomic data. For example, Chen et al. (2019c) benchmarked 10 computational methods that were developed for scATAC-seq and concluded that different methods have their advantages and limitations. Multi-function pipelines provide one-shot solutions with parameters based on best practices, freeing biologists from menial coding and parameter tuning so that they can focus on the biological results.

Dr.seq2, SCRAT, Scasat, Destin, scitools, scATAC-pro, EpiScanpy, Signac, and SnapATAC are designed especially for single-cell chromatin accessibility or methylome accessibility. scHiC-Tools is a pipeline that is designed for scHi-C data. They include the functions of basic qualification, filtering low-quality cells or features, motif analysis, clustering, differential analysis, and visualization. Since many techniques have been developed for parallelly profiling the transcriptome and epigenome, many computational methods and pipelines have been developed for integration. Seurat v3, APEC, MAESTRO, scAI, ArchR, and ALLCools provide the functions that are mentioned above as well as functions for integration of the epigenome data and transcriptome to better interpret the gene regulation mechanism.

Besides these computational methods, g-chromVAR (Ulirsch et al., 2019) uses fine-mapped variant posterior probabilities and quantitative measurements of regulatory activity to measure the enrichment of regulatory variants in each cell state. Methylscaper (Knight et al., 2021) is specifically developed for visualizations of single-cell DNA methylation and chromatin accessibility patterns. Several integration methods have been developed to analyze scRNA-seq and single-cell epigenome data together. These include MATCHER (Welch et al., 2017), coupled NMF (Duren et al., 2018), coupleCoC (Zeng et al., 2021), coupleCoC+ (Zeng and Lin, 2021), scAMACE (Wangwu et al., 2021), epiConv (Lin and Zhang, 2022), scMVP (Li et al., 2022b), scREG (Duren et al., 2022), and MIRA. These computational methods for integration provide a more thorough and multifaceted perspective in which to understand the gene regulatory process. Table S7 in Supporting Information lists the programming language, key features, limitations, and benchmark dataset that were applied in the original analysis of the reviewed computational approaches (Figure 7).

Applications of single-cell epigenomes

Single-cell technologies provide unprecedented opportunities to investigate a variety of biological processes and gene regulation patterns. Applying these single-cell technologies to various biological systems sheds light on discovering the cell differentiation events and mechanisms of disease occurrence at the single-cell level. These single-cell epigenome sequencing methods have been adopted in many fields. Here, we reviewed their applications in early embryonic development, cancer, and neurobiology.

Early embryonic development

During gamete development and the early stages of embryogenesis, cells undergo significant and drastic alterations and reprogramming in the epigenome, which causes cell differentiation and diverse phenotypes of cells. Therefore, embryonic stem cells are widely used as material in the development of single-cell epigenomic sequencing techniques.

Zhu et al. (2018a) applied scWGBS to human preimplantation embryos. They discovered three waves of global demethylation in mouse preimplantation embryos, indicating that the dynamic balance between global demethylation and drastic remethylation occurs during preimplantation development. Later, the same group, Li et al. (2018a) applied scCOOL-seq to six stages of human preimplantation development and discovered that the pluripotency master TF binding regions and proximal and distal nucleosome-depleted regions were primarily enriched in the genomic regions showing the largest changes in chromatin accessibility. Additionally, they discovered that, compared with mice, human zygotes had reduced access to the maternal genome's chromatin in oocytes and had a delayed balance between parental alleles until the 4-cell stage, which indicated the species-specific features of chromatin accessibility. Argelaguet et al. (2019) performed scNMT-seq on the stages of mouse gastrulation. They found that cells committed to mesoderm and endoderm undergo widespread coordinated epigenetic rearrangements at enhancer marks, driven by ten-eleven translocation (TET)-mediated demethylation and a concomitant increase in chromatin accessibility. In addition, they found that while in the early epiblast, the methylation and accessibility landscape of ectodermal cells had already been established.

These studies shed light on how the epigenome influences cellular differentiation and lineage commitment. In the future, investigations into cell populations using single-cell multi-omics techniques give us the chance to understand the process of orchestrated epigenomic reprogramming, which has the potential to change our understanding of cell fate decisions and benefit the field of stem cell biology.

Tumor immunology

Malignant and non-malignant cells coexist in a tumor, which is a highly heterogeneous structure. Both types of cells play critical roles in the development of cancer. Methods for single-cell epigenome sequencing are being developed to help distinguish the non-genetic factors that contribute to the course of cancer from the complexity of tumors.

Satpathy et al. (2019) applied scATAC-seq on primary tumor biopsies from basal cell carcinoma (BCC) patients receiving PD-1 blockade treatment. They investigated chromatin regulators of therapy-responsive T cell subsets and observed a common regulatory pathway that controls the development of CD4$^+$ T follicular helper cells and intratumoral CD8$^+$ T cell exhaustion. Not only are immune cells investigated by single-cell epigenome sequencing, but also malignant cells show heterogeneities in TME. Meir et al. (2020) employed scRNA-seq and methylome analysis to show that various cancer cell types had clonally stable epigenetic memory. Additionally, they discovered DNA methylation landscapes reflect a separate clock-like methylation loss mechanism while correlating with epithelial-mesenchymal transition (EMT) identities that are identified by transcriptome analysis in clonal colon cancer cell populations. Wu et al. (2021c) employed scCUT&Tag to characterize H3K27me3 before and after therapy in a patient with a brain tumor. They profiled a brain tumor H3K27me3 in the primary sample and after the treatment and discovered various cell types in the TME and heterogeneity in the polycomb group activity.

Epigenetic mechanisms are critical for the interactions between tumor cells and immune cells. Understanding the fundamental processes of epigenetic modifications in immune and tumor cells paves the way for the creation of drugs and immunotherapy techniques.

Neurobiology

Understanding both the normal functions of the brain and the mechanisms of dysfunction and disease requires a better understanding of cellular composition. Lake et al. (2018) detected the transposon hypersensitive sites in the human adult brain at the single-cell level. They identified the cell subpopulations in the human adult cortex and cerebellar hemisphere and used epigenomic data to link genetic risk variants with cell-type-specific cCREs. In a cohort of cognitively healthy people, Corces et al. (2020) examined the single-cell chromatin accessibility landscapes and three-dimensional chromatin interactions of various adult brain regions. They created a machine-learning classifier to include this multi-omic framework and predicted several functional SNPs for Parkinson's and Alzheimer's disease. Yang et al. (2023a) profiled single nucleus-accessible chromatin landscape of the pig hippocampus at different developmental stages and revealed notable enrichment of transposable elements in cell type-specific accessible chromatin regions. This study helps deepen our understanding of human neurodegenerative diseases. Future research on the single-cell level will be fascinating in examining dynamic regulations of the epigenome, specifically alterations to the genome during learning and memory that are reliant on neuronal activity.

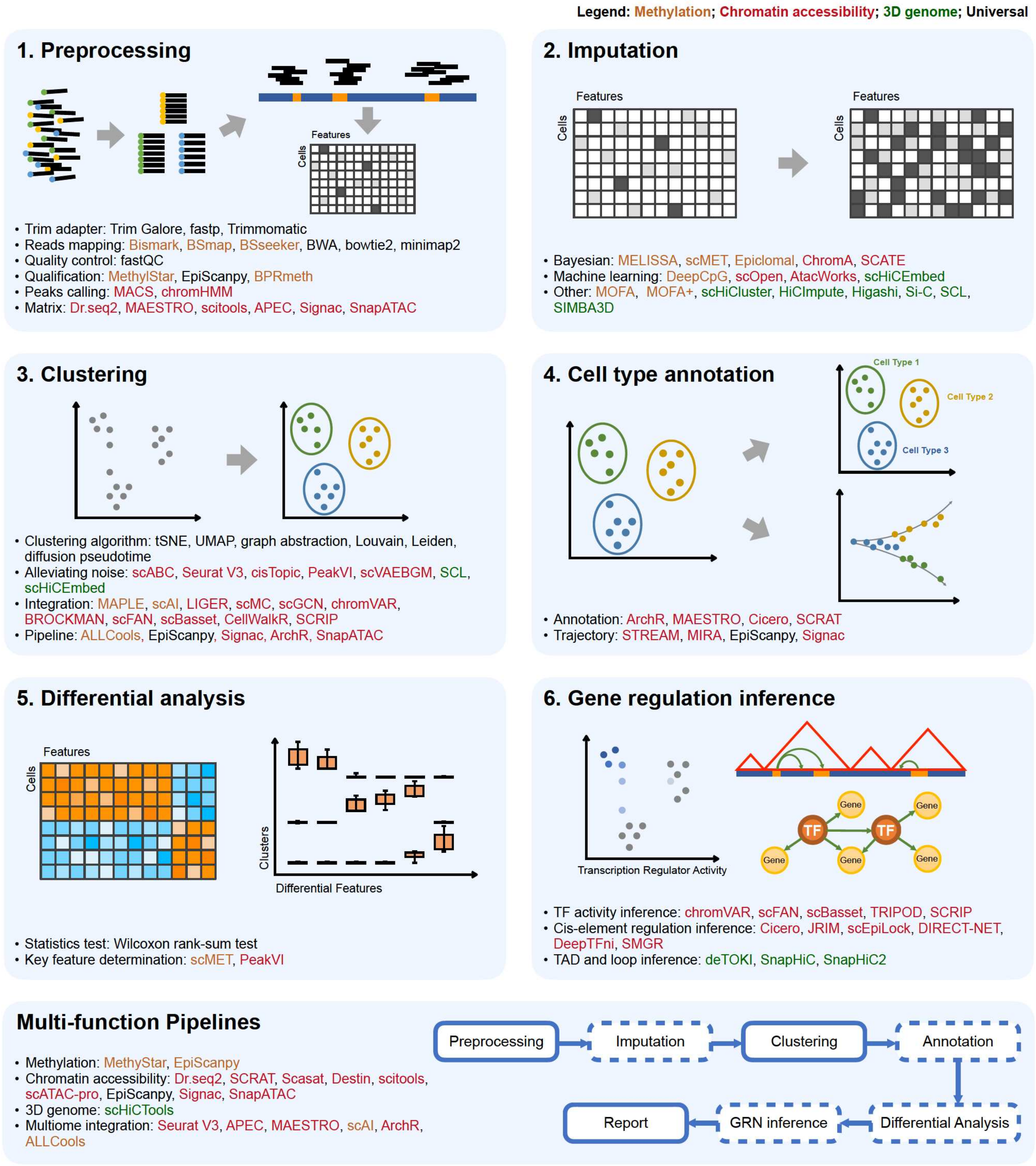

Legend: Methylation; Chromatin accessibility; 3D genome; Universal
1. Preprocessing
Features
Cells
• Trim adapter: Trim Galore, fastp, Trimmomatic
• Reads mapping: Bismark, BSmap, BSseeker, BWA, bowtie2, minimap2
• Quality control: fastQC
• Qualification: MethylStar, EpiScanpy, BPRmeth
• Peaks calling: MACS, chromHMM
• Matrix: Dr.seq2, MAESTRO, scitools, APEC, Signac, SnapATAC
2. Imputation
Features
Cells
Features
Cells
• Bayesian: MELISSA, scMET, Epiclomal, ChromA, SCATE
• Machine learning: DeepCpG, scOpen, AtacWorks, scHiCEmbed
• Other: MOFA, MOFA+, scHiCluster, HiCImpute, Higashi, Si-C, SCL, SIMBA3D
3. Clustering
• Clustering algorithm: tSNE, UMAP, graph abstraction, Louvain, Leiden, diffusion pseudotime
• Alleviating noise: scABC, Seurat V3, cisTopic, PeakVI, scVAEBGM, SCL, scHiCEmbed
• Integration: MAPLE, scAI, LIGER, scMC, scGCN, chromVAR, BROCKMAN, scFAN, scBasset, CellWalkR, SCRIP
• Pipeline: ALLCools, EpiScanpy, Signac, ArchR, SnapATAC
4. Cell type annotation
Cell Type 1
Cell Type 2
Cell Type 3
• Annotation: ArchR, MAESTRO, Cicero, SCRAT
• Trajectory: STREAM, MIRA, EpiScanpy, Signac
5. Differential analysis
Features
Cells
Clusters
Differential Features
• Statistics test: Wilcoxon rank-sum test
• Key feature determination: scMET, PeakVI
6. Gene regulation inference
Transcription Regulator Activity
Gene
TF
TF
Gene
Gene
Gene
Gene
• TF activity inference: chromVAR, scFAN, scBasset, TRIPOD, SCRIP
• Cis-element regulation inference: Cicero, JRIM, scEpiLock, DIRECT-NET, DeepTFni, SMGR
• TAD and loop inference: deTOKI, SnapHiC, SnapHiC2
Multi-function Pipelines
• Methylation: MethyStar, EpiScanpy
• Chromatin accessibility: Dr.seq2, SCRAT, Scasat, Destin, scitools, scATAC-pro, EpiScanpy, Signac, SnapATAC
• 3D genome: scHiCTools
• Multiome integration: Seurat V3, APEC, MAESTRO, scAI, ArchR, ALLCools
Preprocessing
Imputation
Clustering
Annotation
Differential Analysis
GRN inference
Report

Figure 7. Overview of the major steps for single-cell epigenomic analysis and related computation methods. The analysis of single-cell epigenetic data can be roughly divided into 6 parts as shown in the figure. First, the data needs to be preprocessed. The preprocessing part includes removing sequencing adapters, sequence alignment, quality control, data quantification, peak calling and building feature matrix. Due to the extreme sparsity of single-cell epigenetic data, there are currently methods based on Bayesian inference and machine learning to impute missing data. After that, the data is used for dimensionality reduction and clustering. After clustering, different cell types are annotated and pseudotime analysis can be performed. Then, based on different cell types, different genomic features of different cells can be identified. For studying gene transcription regulation, transcription factor binding activity, cis-regulatory elements regulation and TAD regions and loops identification can be inferred. There are also pipelines that integrate multiple functions for batch processing data and output multiple results. Among them, preprocessing, clustering, reporting and other functions are common in all pipelines, while imputation, cell type annotation, differential analysis and gene regulatory network inference are optional functions in different pipelines. In the figure, methods marked in brown are specifically for processing DNA methylation data, methods marked in red are specifically for processing chromatin accessibility data, methods marked in green are specifically for processing three-dimensional genome data and methods marked in black are applicable to two or more types of data.

Summary

In this chapter, we summarized the techniques, computational methods, and applications for single-cell epigenome sequencing. The recent availability of single-cell sequencing technology has expanded the scope of study into biological processes and diseases. These approaches have previously shown their effectiveness in illuminating the parts of complex tissues and revealing novel insights, despite some limitations. Future sequencing technologies with higher coverage and sensitivity, as well as dedicated, advanced, and well-developed computational methodologies, promise to usher in a new era of understanding biology and pave the way for the treatment of diseases.

Chapter 4 Single-cell proteomics technology based on mass spectrometry

Although proteome performed in bulk samples has matured and been widespreadly applied in scientific research and clinical medicine, numerous new challenges have arisen when the initial sample size is slashed to hundreds of cells or even one cell. Since the total protein in one somatic cell is only about 100–200 pg (Wiśniewski et al., 2014), any loss can have an immeasurable impact, and also put more stringent requirements on sensitivity and accuracy of detection methods.

Based on the detection principles, today, single-cell proteomic (SCP) methods can be classified into the following categories: (i) antibody-based assays, such as cytometry by time of flight (CyTOF) (Bandura et al., 2009; Bendall et al., 2011), single-cell western blotting (Hughes et al., 2014b), microengraving (Han et al., 2010; Love et al., 2006; Schubert et al., 2016), (ii) PCR-sequencing based assays, such as proseek multiplex (Assarsson et al., 2014), CITE-seq (Stoeckius et al., 2017), REAP-seq (Peterson et al., 2017), (iii) mass spectrometry (MS) based assays, such as nanoPOTS (nanodroplet processing in one pot for trace samples) (Zhu et al., 2018d; Zhu et al., 2018e), SCoPE2 (Specht et al., 2021). The antibody-based and sequencing-based assays were developed earlier and have made unignorable contributions, which have been well reviewed before (Labib and Kelley, 2020; Levy and Slavov, 2018; Liu et al., 2020a; Xie and Ding, 2022). However, the former was limited by the specificity and the availability of antibodies, while the latter did not directly detect proteins, and both assays were limited in the number of proteins that could be analyzed at a time. Mass spectrometry has been the mainstream analysis tool in bulk-size proteomics due to its advantages of both high accuracy and high throughput. More importantly, unlike the hypothesis-oriented antibody-based methods, MS-based analysis method is discovery-oriented which can play a unique role in biological research. With the progress of mass spectrometry and the innovation of preparation process in recent years, single-cell proteomics based on MS has shown a blowout.

Here we focus on the state-of-the-art MS-based single-cell proteomic tools over the last 5 years, discuss the outstanding innovative points from cell isolation to sample preparation and MS detection, and prospect the future development directions.

MS-based SCP workflow

The workflow of conventional bulk-size proteomics has been well established, but when dealing with the single-cell, each simple step needs to be well optimized to recover as much information as possible from the trace protein. Although the ways of implementation vary, the workflows of MS-based SCP include three main steps: single-cell isolation, sample preparation, and MS analysis.

Single-cell isolation

Isolation of single-cell is a unique requirement in SCP workflows, its accuracy and efficiency lay the foundation for the entire SCP workflow. How to precisely pick the interesting cell types from a complicated cell mixture, and how to maximize cell activity and minimize the impact on the molecular level are the great challenges in this step. Among the existing SCP tools, isolation of single-cell was mostly achieved through manual cell picking, FACS, microfluidic, lab-on-a-chip devices (Gross et al., 2015), or LCM (Mund et al., 2022a).

Sample preparation

Cell lysis and protein digestion are the essential preparation steps for both bulk-size and single-cell proteomic workflows. Since the protein amount contained in one cell is extremely small, any step that may be negligible for bulk-size proteomics can be fatal to the SCP. Even the nonspecific adsorption during sample transfer steps can cause massive protein and peptide loss (Sun and Kumar, 2022; Wu et al., 2019b). Across the existing SCP tools, the improvement strategies can be summarized into (i) simplifying the preparation steps, (ii) minimizing the sample volume, and (iii) automating the preparation.

In bulk-size proteomics, harsh chemical environments such as sodium dodecyl sulfonate (SDS) and urea were required for millions of cell lysis, which need complex buffer exchange steps followed to adapt the enzymatic processes and liquid chromatography-mass spectrometry (LC-MS) analysis. To avoid loss during these steps, most SCP tools chose MS compatible lysis reagents such as DDM (n-Dodecyl-β-D-Maltopyranoside), RapiGest, trifluoroethanol (TFE) (Ctortecka et al., 2022a; Li et al., 2018b; Schoof et al., 2021; Wang et al., 2022d; Williams et al., 2020; Zhu et al., 2018c; Zhu et al., 2018d), and supplemented with heating or sonication to promote cell lysis. Slavov's group (Specht et al., 2021) developed a sample preparation method called mPOP which can lyse mammalian cells in pure water by a freeze-heat cycle (−80°C to 90°C), completely avoiding the introduction of redundant chemical reagents. Moreover, almost all SCP tools took the one-pot preparation to avoid loss during sample transfer, or even completing overall preparation in highly integrated microfluidic chips or capillary, such as iPAD-1 (Shao et al., 2018), iProChip (Gebreyesus et al., 2022).

Decreasing sample volume is another solution to the loss of nonspecific adsorption, which can increase sample concentration meanwhile. Compared with bulk-size proteomics, the protein amount in SCP has a thousand-fold decrease, but the sample volume has not decreased to the same extent, leading to a great reduction of protein concentration. When the concentration is low enough, the tiny amount of protein in the large volume of solution can cause "swimming pool effect" that dramatically reduces the reaction efficiency of enzyme or label reagent with proteins or peptides. Increasing the concentration of enzyme or label reagent can partially alleviate this effect, but an excess of the enzyme or label reagent relative to the protein or peptide level will inevitably cause signal interference in downstream analysis and increase costs exponentially. Therefore, minimization of sample volume can greatly solve the interference caused by insufficient sample concentration and significantly enhance SCP performance. Most recent SCP tools did not exceed the initial volume of 1 μL in the sample preparation step. With the help of capillary-based sampling method or picoliter-level liquid dispensing technology, the initial volume can be controlled to as low as 2 or 8 nL, and the overall reaction system volume is not greater than 50 nL during the entire preparation process (Leduc et al., 2022). When the sample volume is only a few microliters or even tens of nanoliters, how to avoid liquid evaporation during the preparation process becomes a new problem. Precise control of temperature and humidity throughout the reaction environment is necessary. Oil or water-insoluble organic solvent is also used to encapsulate protein droplets to prevent volatilization. For example, OAD (nanoliter-scale oil-air-droplet) used fluorinated oil FC40 to cover the cell sample (Li et al., 2018b), proteoCHIP chose hexadecane which solidified at preparation temperature (Ctortecka et al., 2022a), and WinO designed the entire preparation process in the droplet coated with ethyl acetate (Masuda et al., 2022).

As the sample size decreases, the perturbations caused by human manipulation of the entire SCP workflow are further amplified. To improve throughput and micromanipulation accuracy, some laboratories have developed specific robots to complete the integrated preparation process, such as nanoPOTS dispensing robot (Zhu et al., 2018d) and SODA (the sequential operation droplet array) system used in OAD (Li et al., 2018b; Zhu et al., 2013), as well as other SCP tools employing commercial robotic workstations to cover part or whole preparation steps. The automation of SCP sample preparation has been well reviewed by Alexovič et al. (2021). Automation is an inevitable trend to further improve throughput and reproducibility so that single-cell proteomes can be truly applied to large-scale investigations or clinical studies.

MS analysis

The complexity of proteins has always been one of the biggest barriers in proteomic research. When MS-based proteomics is applied to the single-cell field, the extremely tiny sample size presents new challenges. The types of protein can reach over 10,000 with different expression levels and different physicochemical properties in one cell (Zhang et al., 2013). How to minimize the loss during MS analysis process without reducing the resolution of complex samples, and how to improve the sensitivity and accuracy of detection meanwhile are huge problems that need to be considered from three aspects: analysis strategy, injection method, and chromatography-mass spectrometry performance.

(1) Analysis strategy

Isobaric label-based quantification, such as tandem mass tags (TMT), has been one of the most popular protein quantification methods. TMT label-based bulk-size proteomics has been shown to result in a 15%–20% increase in proteins identified with higher quantitative accuracy (Muntel et al., 2019). Slavov's group (Budnik et al., 2018) developed the SCoPE-MS (single-cell ProtEomics by mass spectrometry) which pioneered the application of TMT to single-cell proteomics. They introduced the idea of "carrier channel" which consisted of about two hundred cells to share most of the loss from single-cell channels caused by non-specific adsorption. Meanwhile, the carrier channel provided the most signal for MS analysis, reducing the required sensitivity 10- to 100-fold. Several SCP tools have been developed subsequently based on isobaric labeling, including SCoPE2, WinO (Masuda et al., 2022; Specht et al., 2021). The improvement in MS analysis throughput is another advantage of isobaric labeling SCP tools that cannot be ignored. Labeled with the TMTpro18-plex, one injection can analyze more than 14 single cells (Leduc et al., 2022). The appropriate number of cells used in carrier channel, however, is still debatable. Cheung et al. (2021) revealed that high levels of carrier channels may adversely affect quantitative accuracy. With the fast development of mass spectrometer, higher detection sensitivity helped reduce the number of cells in the carrier channel. Using 25 cells as carrier channels or eliminating carrier channels, proteoCHIP identified an average of 1,812 or 1,477 proteins from one mammalian somatic cell respectively (Ctortecka et al., 2022a).

The label-free SCP tools that analyze one cell at a time are the equally important development direction in the field of single-cell proteomics. Data-independent acquisition (DIA) is becoming mainstream in bulk-sized proteomics because of its accurate quantification with low missing values and high analytical depth. With the generation of a project-specific library, DIA mode can help label-free proteomics achieve a much higher analytical depth. DIA mode has begun to be applied to label-free SCP tools recently (Brunner et al., 2022; Gebreyesus et al., 2022; Wang et al., 2022d). Compared with the most data-dependent acquisition (DDA)-based label-free SCP work which identified about 1,000 proteins, DIA mode helped identified protein numbers rise up to more than 2,000 in one single cell. Using the same SCP workflow to detect the same type of cells, DIA mode can increase the protein identification number of one single cell by up to 188% compared with is not compatible with the common perception, some groups have already tried to develop the multiplexDIA method such as DIA-TMT and plexDIA to improve data integrity and reliability without reducing throughput (Ctortecka et al., 2022b; Derks et al., 2022).

(2) Injection method

Despite the advantages of minimizing sample volume, the nanoliter-level sample droplet is not compatible with most commercial LC autosamplers. To solve this problem, nanoPOTS group developed the complicated manual loading procedures that aspirated nanodroplet samples to a section of capillary, then eluted the sample onto a solid-phase extraction (SPE) column, and finally inserted the SPE column with an analytical column for gradient separation and MS detection (Zhu et al., 2018d). These procedures are not only complex and time-consuming but also highly dependent on the proficiency of the operators. As an improvement, a nanoliter-scale autosampler integrating nanoPOTS-based sample preparation with automated LC-MS platforms was developed, which enhanced the analysis throughput based on label-free nanoPOTS from 6 cells to 24 cells one day (Specht et al., 2018). Integration of sample preparation with LC-MS analysis has been an important development trend of SCP tools for its robustness and high-throughput. Some integrated tools were based on the self-development autosamplers such as autoPOTS (Specht et al., 2018; Woo et al., 2021) and self-aligning monolithic (SAM) devices (Li et al., 2018b; Wang et al., 2022d), while others were developed based on a high-integrated microfluidic chip or device such as proteoCHIP (Ctortecka et al., 2022a), and iPAD-1 (Shao et al., 2018). Although most of the integrated SCP methods require customized equipment which limits their accessibility, from the perspective of minimizing the loss during sample loading and optimizing the detection effect, the integration of sample preparation and LC-MS analysis is still an inevitable road.

(3) Chromatography-mass spectrometry performance

The overall sensitivity of the chromatography-mass spectrometry system is crucial for the analysis of extremely tiny amounts of peptide samples. Decreasing the chromatographic flow rate and narrowing separation columns' inner diameter are widely used to enhance the separation performance and ionization efficiency. Most bulk-size proteomics conventionally uses the 75 μm i.d. reversed-phase LC columns which operate at 300 nL min^{-1}. Zhu et al. have demonstrated that using 30 μm i.d. columns operating at 50 nL min^{-1} can remarkably improve the proteome coverage and have applied to most nanoPOTS-relative work (Specht et al., 2018; Woo et al., 2021; Zhu et al., 2018f). Although narrower columns and lower flow rate were also tried, the challenges in column package and longer chromatographic gradients limited routine use. A variety of prospective LC technologies also have been explored to improve separation efficiency and have been applied in low-input proteomics, including capillary electrophoresis (Lombard-Banek et al., 2016; Lombard-Banek et al., 2019), porous layer open tubular (PLOT) columns (Li et al., 2015), monolithic capillary columns (Greguš et al., 2020), micropillar array columns (μPAC) (Stadlmann et al., 2019). It is worth looking forward to their applications in the SCP field with advanced mass spectrometry.

Mass spectrometry has undergone substantial development in the past decades, reflected in the improvement of data acquisition speed, detection limit, resolution, and accuracy. Orbitrap series mass spectrometers are the most commonly used in both bulk-size and single-cell proteomics because of their outstanding performance in both resolution and accuracy. With updates to Orbitrap platforms, the protein information available from single cells has increased significantly. For example, when analyzing 2 ng peptide sample, nearly 3-fold unique peptides can be identified by an Orbitrap Fusion Lumos compared with an LTQ Orbitrap XL mass spectrometer. Further, comparing the Orbitrap Fusion Lumos with a newer Orbitrap Eclipse mass spectrometer, the protein coverage from one single cell increased by about 20% (Kelly, 2020). Another notable breakthrough in mass spectrometry is the introduction of ion mobility which added a new separation dimension and resulted in the transition from 3D-Proteomics (retention time, m/z, and ion intensity) into 4D-Proteomics. TimsTOF series mass spectrometers are representative and have been applied in several recent SCP tools such as PiSPA (Wang et al., 2022d), UE-SCP (Gu et al., 2022b), and T-SCP (Brunner et al., 2022). Combined with parallel accumulation-serial fragmentation (PASEF), timsTOF can achieve almost 100% ion utilization and more than 10-fold increase in sensitivity (Meier et al., 2018). FAIMS Pro$^{\mathrm{TM}}$ interface is another popular technique to combine ion mobility with mass spectrometry and can be used in conjunction with Orbitrap series mass spectrometers (Shvartsburg et al., 2006). Applied in the SCP field, field asymmetric ion mobility spectrometry (FAIMS) has been shown to increase protein coverage by 2.3-fold in a single HeLa cell (Cong et al., 2020).

State-of-the-art SCP tools

Recently a variety of SCP tools have sprung up. With the comprehensive advance in cell isolation, sample preparation, and MS analysis mentioned above, the number of proteins identified from one cell has jumped from about 100 to more than 3,000 nowadays. The mainstream bulk-size MS-based proteomics performs complex sample preparation and off-line sample loading steps separately. The single-cell proteomics, however, developed a series of integrated tools to reduce the loss of tracing peptides during pretreatment and sample loading. As a double-edged sword, integrated tools usually require specially customized equipment which limits their promotion and application among other laboratories. Many unintegrated and easy-to-use tools have been developed at the same time (Table 2).

Integrated tools

Chen et al. (2015b) established an integrated proteome analysis device called iPAD-100 in 2015 which completed the whole progress from cell preparation to injection sample into the LC-MS system. iPAD-100 can accomplish cell lysis and protein digestion in a fused-silica capillary simultaneously in only 1 h and robustly identify 635 proteins from 100 living DLD-1 cells. As an updated version, iPAD-1 chose the 22 μm o.d. capillary for single-cell picking and sample preparation, reduced the reactor volume to 2 nL, and compressed the preparation time to 30 min (Shao et al., 2018) (Figure 8A). With further optimized ultrasensitive nano-LC-MS/MS system, a maximum of 328 proteins were identified from one Hela cell.

Li et al. (2018b) designed a droplet-based microfluidics chip called OAD chip which is composed of 4-layer cube structure (Figure 8B). With the isolation by oil in the isolation layer and oil layer, about 100 nL sample droplet can be encapsulated in droplet layer to avoid evaporation and contamination. The entire preparation process took place in the droplet layer which was manipulated through a 3D printing fabricated SAM device with cylinder geometry. The enzymolytic peptide sample was then directly loaded into the nanoliter-level separation column in a pressured manner. With this approach, 51 and 355 proteins were identified in one Hela cell and one mouse oocyte respectively. Recently, Dang et al. (2023) applied OAD to human pre-implantation embryos and achieved a median of 3,736 protein identification from single 2-cell stage human embryos.

NanoPOTS is another microfluidics chip based nanodroplet processing platform that was developed by Zhu et al. (2018d). NanoPOTS chip is composed of a nanowell-patterned glass slide, a glass spacer, and a membrane-coated glass slide (Figure 8C). The surface area of each nanowell was only 0.8 mm^2, which greatly reduced the nonspecific adsorption loss on the reaction vessel surface. Interfacing with FACS made nanoPOTS become an excellent and robust SCP tool that can identify 670 proteins from one HeLa cell when employed the MaxQuant match between run (MBR) algorithm (Zhu et al., 2018c). Although the sample preparation was accomplished in the highly integrated chip, the early nanoPOTS still needed a complicated system to load sample to LC-MS manually. An autosampler was developed to solve this problem soon afterward and improve the throughput from 6 to 24 cells per day in label-free experiments (Specht et al., 2018). TMT label was also introduced to nanoPOTS and an improved boosting to amplify signal with isobaric labeling (iBASIL) strategy was put forward (Tsai et al., 2020). 1,424 proteins could be identified from a single cell by using TMT10plex label and a boost channel containing 10 ng peptides. The throughput was further increased to 77 cells per day. Recently, nested nanoPOTS (N2) chip derived from classical nanoPOTS was developed (Woo et al., 2021). It was designed with tighter nanowell array and increased the cell number that can be analyzed in one chip to 243 (Figure 8D). By further reducing the nanowell volume to about 30 nL, the protein recovery was increased by 230%.

ProteoCHIP is also a highly integrated SCP tool that was designed with two parts: (i) a nanowell layer that included 12 fields each containing 16 nanowells, (ii) a funnel layer that can

Table 2. Recent SCP tools

SCP tools	Customized equipment	Label	Cell type	Isolation method	MS approach	Pretreatment throughput	MS throughput	Average identified protein number per cell	Depth of proteome coverage	Reference
iPAD-1	Need	label free	HeLa	capillary-based isolation	LC-MS/MS	1 cell per run	24 cells per day	271	406 (n=10)	(Shao et al., 2018)
OAD	Need	label free	HeLa	SODA	LC-MS/MS	1 cell per run	4 cells per day	51	–	(Li et al., 2018b)
nanoPOTS	Need	label free	HeLa	FACS	LC-MS/MS	27 cells per run	–	669	–	(Zhu et al., 2018c)
	Need	label free	MCF10A	FACS	LC-MS/MS	–	24 cells per day	764	1,093 (n=10)	(Williams et al., 2020)
		TMT10-plex labelled	MOLM-14, K562, CMK				77 cells per day	1,281	2,558 (n=152)	
autoPOTS	No need	label free	HeLa	FACS	LC-MS/MS	–	10 cells per day	301	–	(Liang et al., 2021c)
nested nano-POTS	Need	TMT16-plex labelled	C10, RAW, SVEC	cellenONE	LC-MS/MS	243 cells per run	108 cells per day	1,716	2,457 (n=108)	(Woo et al., 2021)
proteoCHIP	Need	TMT16-plex labelled	HeLa, HEK-293	cellenONE	LC-FAIMS-MS/MS	592 cells per run	384 cells per day	1,940 (20× carrier) 1,598 (no-carrier)	3,674 (n=276)	(Ctortecka et al., 2022a)
iProChip	Need	label free	MEC-1	chip device	LC-MS/MS	9 cell per run	9 cells per day	455	–	(Gebreyesus et al., 2022)
SciProChip	Need	label free	PC-9	–	–	20 cell per run	16 cells per day	1,500	1,995 (n=10)	
PiSPA	Need	label free	A549	SODA	LC-TIMS-TOF	1 cell per run	–	3,008	5,093 (n=37)	(Wang et al., 2022d)
SCoPE	No need	TMT10-plex labelled	Jurkat, U-937	manual picking	LC-MS/MS	8 cells per run	48 cells per day	–	767 (n=24)	(Budnik et al., 2018)
SCoPE2	No need	TMT16-plex labelled	monocyte and macrophage cells	FACS	LC-MS/MS	–	200 cells per day	1,000	3,042 (n=1,490)	(Specht et al., 2021)
nPOP-SCoPE2	Need	TMT18-plex labelled	U-937, WM989	cellenONE	LC-MS/MS	2,016 cells per run	212 cells per day	997	2,844 (n=1,543)	(Leduc et al., 2022)
A multiplexed scMS workflow	No need	TMT16-plex labelled	OCI-AML8227	FACS	LC-FAIMS-MS/MS	336 cells per run	112 cells per day	987	2,723 (n=2,050)	(Schoof et al., 2021)
UE-SCP	No need	TMT6-plex labelled	HeLa, HEK-293T	cellenONE	LC-TIMS-TOF	308 cells per run	96 cells per day	2,249	4,230 (n=128)	(Gu et al., 2022b)
Mad-CASP	No need	label free	HeLa	FACS	LC-MS/MS	–	16 cells per day	1,240	–	(Li et al., 2022h)
WinO	No need	TMT10-plex labelled	RPMI8226	SH800S Cell sorter	LC-MS/MS	–	144 cells per day	845	–	(Masuda et al., 2022)
T-SCP	No need	label free	HeLa	FACS	LC-TIMS-TOF-SCP	308 cells per run	41 cells per day	2,083	2,501 (n=231)	(Brunner et al., 2022)

pool samples from the same TMT set via centrifugation and online connect to LC autosampler (Figure 8E) (Ctortecka et al., 2022a). proteoCHIP eliminated all manual sample handling steps and resulted in a high-throughput and high sensitivity analysis, which can identify an average of 1,812 or 1,477 proteins from one mammalian somatic cell using 25 cells as carrier channels or eliminating carrier channels respectively. This was the first attempt to eliminate carrier channels in TMT label-based SCP and has achieved remarkable performance.

Unlike most microfluidic chips used in SCP were open, Gebreyesus et al. (2022) designed a confined, highly integrated microfluidic chip called iProChip (Figure 8F). This chip was composed of 9 units and each of them contained a cell capture, imaging and lysis chamber, a protein reduction, alkylation and digestion vessel, and a peptide desalting column. Size-based single-cell capture was achieved by the wedge-shaped twin pillar arrays and the following preparation process can be accomplished online in the chip. With the optimized DIA-MS analysis, 1,160 proteins were identified from one PC-9 cell. SciProChip was derived from iProChip and dedicated to 20-plex processing of single cells. It showed an improvement in cell usage efficiency of ~40% and in protein coverage of 1.53-fold. From one PC-9 cell, SciProChip-DIA can identify about 1,500 proteins.

Recently, Wang et al. (2022d) developed the pick-up single-cell proteomic analysis (PiSPA) workflow which accomplished single-cell sorting, multi-step preparation and injection of peptides to the LC column integrally by the automated pick-up operation system based on capillary probes. In order to avoid losses of sample transfer, this workflow directly dispensed single cells into a commercial insert tube, using the conical bottom tip of the insert tubes as the nanoliter microreactors for sample pretreatment of single cells (Figure 8G). These insert tubes coupled with

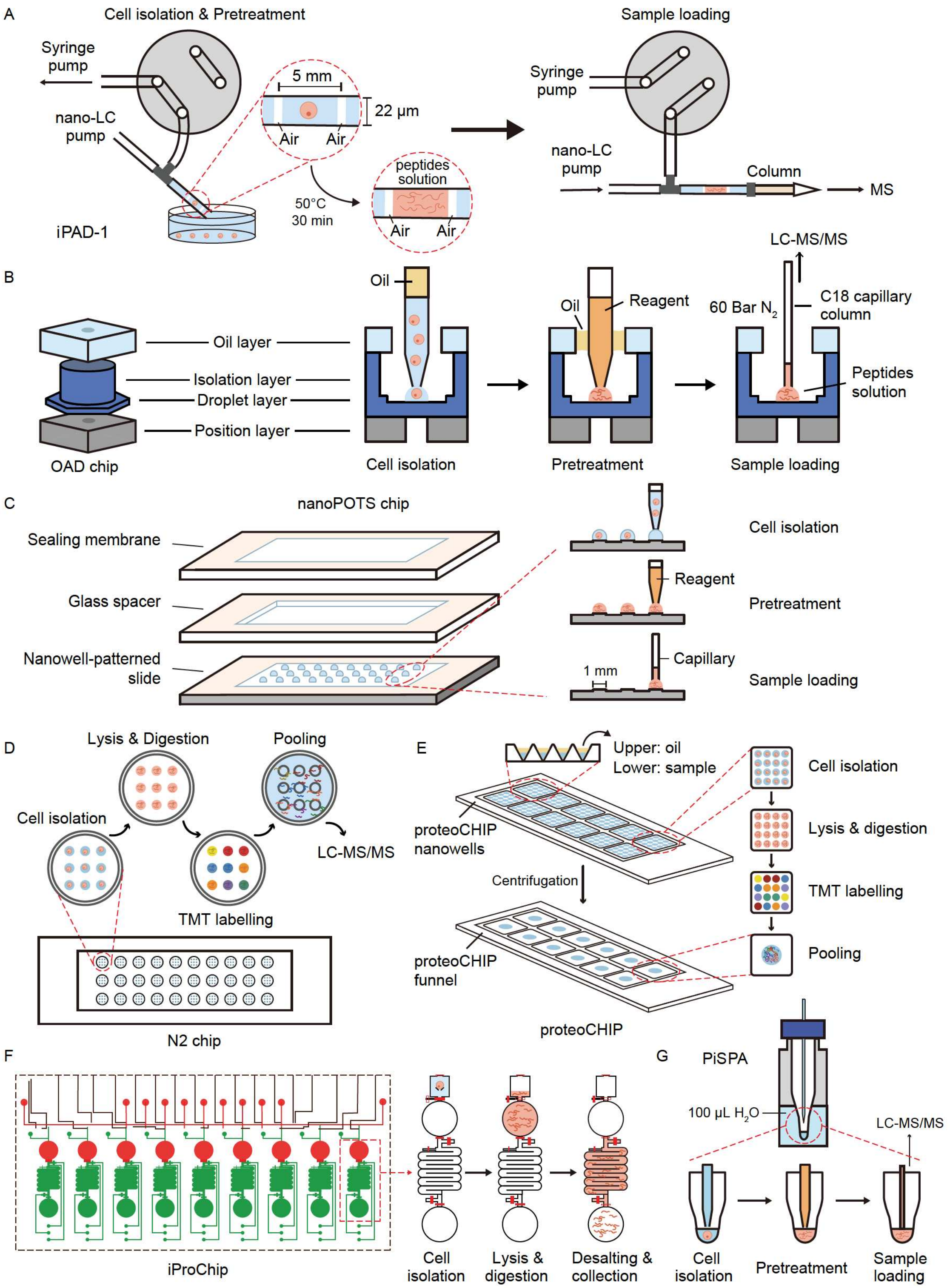


Figure 8. Workflows of recent SCP tools. A, iPAD-1 (Shao et al., 2018). B, OAD (Li et al., 2018b). C, nanoPOTS (Zhu et al., 2018d). D, nest nanoPOTS (Woo et al., 2021). E, proteoCHIP (Ctortecka et al., 2022a). F, iProChip (Gebreyesus et al., 2022). G, PiSPA (Wang et al., 2022d).

sample vials were subsequently used as sample tubes for the autosampler of the LC system to perform the sample injection. With the help of the high sensitivity LC-timsTOF system and DIA mode, 2,467 proteins can be identified from one A549 cell. PiSPA achieved the largest protein number identified from one single mammalian somatic cell in label-free SCP up to now.

Easy-to-use tools

SCoPE pioneered the application of TMT labeling strategies in single-cell proteomics and introduced a "carrier channel" containing hundreds of cells to share the most non-specific adsorption loss as well as to provide most signal for MS analysis (Budnik et al., 2018). Single cells were picked manually into reaction tubes containing H_2O and then mechanically lysed by adaptive focused acoustics (AFA). Any detergent that would interfere with the LC-MS system was not introduced. With TMT10plex labelling, 767 proteins were identified from single U-937 or Jurkat cell using 200 cells as carrier. To overcome the weaknesses of expensive AFA equipment and low-throughput manual cell isolation, a second-generation tool, SCoPE2, was developed (Specht et al., 2021). Single cells were sorted by FACS into the commercial 384-well plates containing 1 μL H_2O and then lysed through freeze-thawing (Figure 9A). Compared with SCoPE, SCoPE2 decreased lysis volumes by 10-fold, reduced the cost of consumables and equipment by over 100-fold, increased the throughput of sample preparation by over 100-fold, and increased the identified proteins from one cell up to 1,000. This was the first SCP tool that could identify more than 1,000 proteins in one cell without any customized equipment or expensive instrument. Recently, the same group developed a new preparation method called nPOP which employed cellenONE for cell sorting and liquid operation (Leduc et al., 2022). nPOP enabled the simultaneous and automated preparation of over 2,000 single cells in droplets on a special fluorocarbon coated glass slide surface. Although automatic precision operations and high throughput attenuated the batch effect, nPOP became less accessible. Schoof et al. (2021) developed a similar SCP tool inspired by ScoPE2. They replaced water with TFE as the lysis reagent, which has been shown to produce more protein and especially peptide identifications. It is worth mentioning that they developed a computational workflow, SCeptre (single cell proteomics readout of expression), for the analysis of SCP MS data. SCeptre was implemented in Python and enabled quality control, normalization of batch effects and biological interrogation of multiplexed SCP MS data. Our group developed UE-SCP (an ultra-sensitive and easy-to-use multiplexed single-cell proteomic workflow) which was also inspired by ScoPE2 recently (Gu et al., 2022b). UE-SCP employed the cellenONE for sorting single cell softly and reduced the cell number in carrier channel to 100 for better quantification. With the help of high sensitivity LC-timsTOF system, the median number of proteins identified from one HeLa cell can exceed 2,300, which achieved the largest identified protein number from one cell without any customized equipment up to now.

To avoid sample loss caused by nonspecific adsorption of tubes and LC columns, Li et al. (2022h) brought a new idea called Mad-CASP (mass-adaptive coating-assisted single-cell proteomics). They designed a hydrophobic peptide, which was mainly composed of hydrophobic amino acids (AAs) F and V with K inserted every 4 AAs (Figure 9B). Using tubes coated with these peptides to prepare single-cell samples, the number of identified proteins can increase by 63%. During trypsin digestion, these hydrophobic peptides could be digested into 5-AA peptide fragments. These low-mass fragments would be excluded in MS data acquisition and simultaneously play the role of carriers to reduce the loss of single-cell peptides due to the adsorption of the LC column. With this novel preparation and data acquisition strategy, they identified an average of 1,240 proteins from a single HeLa cell.

Masuda et al. (2022) developed another novel SCP tool, called WinO (a water droplet-in-oil digestion). It was based on carboxyl-coated beads and phase transfer surfactants (Figure 9C). Single cells were directly sorted into 96-well plates containing 50 μL ethyl acetate and formed as a suspending droplet. Then the entire preparation was accomplished in this water droplet, minimizing the contact area between sample and containers to reduce the loss of proteins and peptides by adsorption. It was the first attempt to use magnetic beads in SCP to enhance the recovery of hydrophobic proteins and peptides. 96.2% of identified peptides showed higher intensity in samples prepared with the beads than in those without beads. This workflow has been successfully performed on 96-well plates and identified 845 proteins from one cell based on TMT10plex labelling.

Recently, Brunner et al. (2022) reported a true single-cell-derived proteomics (T-SCP) that aimed to combine the most advanced technologies that could achieve ultra-high sensitivity and be commercially available at the same time. Cell sorting was achieved by FACS and the entire preparation happened in the commercial 384-well plate in microliter-level volume that was easy to operate. Then peptides were concentrated in an EvoTip device for desalt and on-line sample loading. With the help of timsTOF SCP MS and diaPASEF mode, 2,083 proteins were identified from one HeLa cell using a HeLa DIA spectral library with about 4,000 protein groups.

Applications

Although the SCP technologies are not as developed as single-cell transcriptomics (SCT), there are still several studies that have demonstrated its indispensability in biological and clinical research. Here we mainly introduce its applications in cell differentiation, disease heterogeneity, and cell cycle.

Cancer heterogeneity

Cell heterogeneity is an increasing concern in diseases, especially cancer research. Tsai et al. (2021) developed a novel SCP tool termed surfactant-assisted one-pot sample preparation (SOP)-MS and applied it to single luciferase 2-tdTomato (L2T) tumor cells derived from a patient CTC-derived xenograft (PCDX) mouse model, revealing different protein signatures between primary tumors and early lung metastases. The differentially expressed proteins are involved in tumor immunity, epithelial cell differentiation and EMT, indicating the possibility of selective pressure in immune evasion and cell state plasticity. These results provide a clear path for future research into the mechanisms of cancer metastasis and have the potential to guide targeted cancer therapies

Biomarker discovery

MS-based proteomics is an ideal tool for discovering differences in protein abundance levels in patients and healthy individuals, and therefore, in principle, a powerful technology for biomarker discovery. Karayel et al. (2022) performed large-scale cerebrospinal fluid proteomics analyses on Parkinson's disease patients and quantified more than 1,700 proteins. They discovered lysosomal and immune-related biomarker signatures specific to Parkinson's disease patients with LRRK2 G2019S carriers. Du et al. (2023) performed proteome profiling of 144 urinary and 44 urinary exosomes from type 2 diabetes mellitus patients with albuminuria in varying degrees. By analyzing, they found several potential biomarkers, such as SERPINA1 and transferrin, that could be used for diabetic kidney disease diagnosis or disease

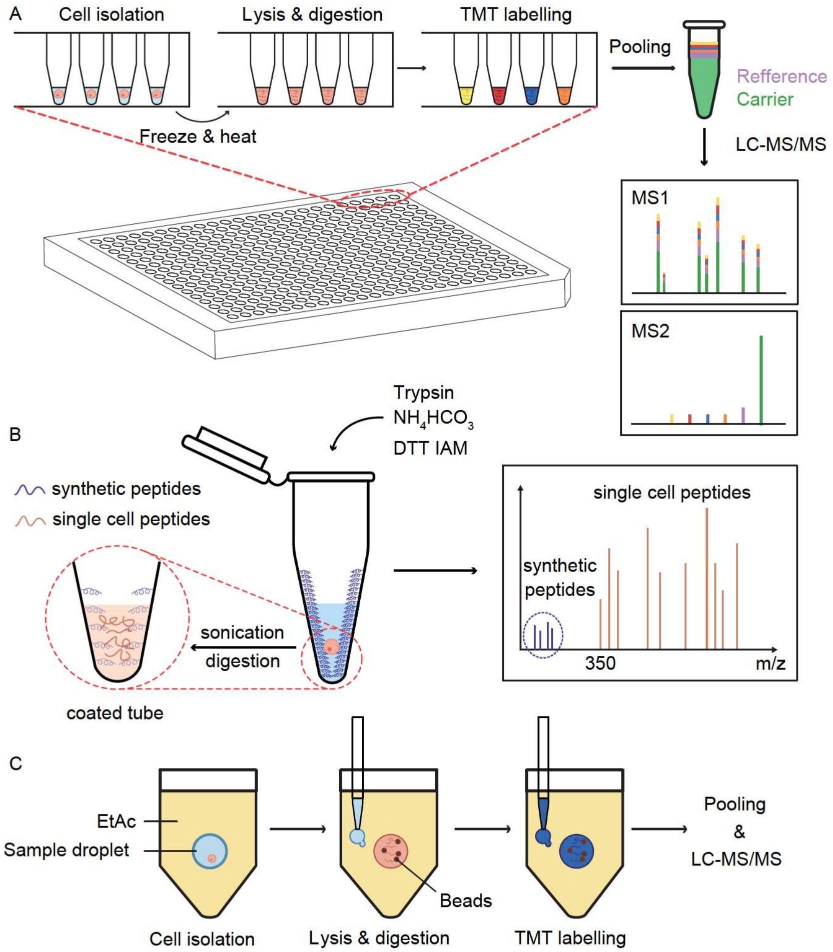


Figure 9. Workflows of recent easy-to-use SCP tools. A, SCoPE2 (Specht et al., 2021). B, WinO (Masuda et al., 2022). C, Mad-CASP (Li et al., 2022h).

monitoring.

Cell differentiation

Cell differentiation processes are subject to various disturbances that lead to different cell fates. Single-cell proteome has been applied to reveal the heterogeneity and dynamics during cell differentiation. Using SCoPE-MS, Budnik et al. (2018) quantified the single-cell proteome of ES cell in days 3, 5, and 8 after differentiation induction. They revealed the corresponding correlation vectors between days 5 and 8 were more similar than between days 3 and 5, indicating the more advanced differentiation changes on those days. Compared their SCP data with SCT data, they further found there is a coordinated mRNA and protein covariation at the single-cell level, proving the quantitative accuracy and necessity of SCP research. Using a multiplexed SCP workflow derived from SCoPE-MS, Schoof et al. (2021) explored the protein profiles of cells in different differentiative stages from a primary acute myeloid leukemia (AML) culture model. They successfully distinguished differentiation stages in this complex cellular hierarchy and found there might be two parallel differentiation trajectories for leukemic stem cells (LSC).

Cell cycle

Analysis of the same cell type in different cell cycle phases is another challenge in single-cell omics which requires higher sensitivity and accuracy of detection. Brunner et al. (2022) applied T-SCP to demonstrate the protein profiles of HeLa cells which were arrested cell cycle by drug. They investigated the differentially expressed proteins between different cell cycle stages, found a large number of known cell cycle regulators were significantly regulated, and also identified some new cell cycle-associated proteins. Using the upgraded preparation method nPOP and LC-MS analysis method pSCoPE, the research team of SCoPE2 recently quantified cell division cycle (CDC)-related protein covariation within a cell type (Leduc et al., 2022). They identified differentially expressed proteins among G1, S, and G2/M phases in monocyte and melanoma cells, and constructed CDC markers.

Discussion and prospects

At the genomic and transcriptomic levels, single-cell sequencing has become a powerful tool for studying cell heterogeneity and identifying different phenotypic cell types. In contrast, MS-based single-cell proteomics is still in its infancy. A recent study by Brunner et al. (2022) sheds light on the necessity for single-cell proteomics. When comparing the SCP measurements with similar single-cell RNA sequencing data, the protein expression completeness reached 49% on average, whereas gene expression completeness was only 27% in SMART-seq2 and even as low as 8% for droplet-based data. Many of the transcripts are expressed at less than one copy per cell on average and result in a mass of shot noise in SCT data. Thus, the amount of transcriptional information that can be captured from one cell is very limited, while single-cell proteomics can provide relatively complete information on the protein level of one cell. This is particularly important for precious cell types such as CTCs and embryos. On the other hand, several published works have demonstrated a low correlation between SCT and SCP. Although bulk transcriptomic and proteomic data showed a moderate correlation, the correlation at the single-cell level decreased to 0.2–0.4 (Brunner et al., 2022; Woo et al., 2021). These results illustrated that the proteome and transcriptome levels of the same gene can vary greatly, but this difference was obscured in the bulk measurements. These views further suggested the necessity for measuring proteome directly at the single-cell level.

The development direction of single-cell proteomics is always focused on more identified proteins and higher throughput. Profit from both advances in instruments and the development of new SCP workflows, the protein number identified from one mammalian somatic cell has jumped to the level of 3,000. More than 5,000 proteins can be cumulatively quantified in ~40 single-cells, which has been able to achieve a similar data level as bulk-size proteomics (Wang et al., 2022d). By using the TMT labeling strategy with shorter liquid chromatography gradients, more than 2,000 single cells have been completely analyzed in less than ten days, which has also been close to matching the SCT throughput (Leduc et al., 2022). Although the deepest protein coverage and the highest throughput have not been achieved in one SCP tool, there are some methods that perform well in both. proteoCHIP achieved analysis of more than 300 cells in about 2 days and identified 3,674 proteins from 276 single cells. Our recent work, UE-SCP, identified 4,320 proteins from 128 single-cells and the entire analysis can be completed within 3 d.

At present, most SCP tools are limited to the laboratory where they were developed and require highly specialized equipment or operator level. Ease of use is an ineluctable issue for SCP to become a viable tool for scientific and clinical researches. There are several SCP tools that can perform excellently without any customized equipment, such as ScoPE2, UE-SCP, and T-SCP. At the same time, some highly integrated microfluidic chips showed promise for commercialization, such as proteoCHIP and N2 chips. Given that SCP is still in its early stages, most of the losses caused by the sample transfer process are still difficult to solve in unintegrated manners. We summarized the latest SCP tools from accessibility, throughput, and analysis depth shown in Figure S1 in Supporting Information. How to better balance these parameters is the next question to consider.

With the help of high-speed cell sorting machines, simplified sample preparation processes, and automated liquid operators, the steps prior to MS analysis have reached the throughput of more than 2,000 single cells per day (Leduc et al., 2022). However, under the premise of considering the analytical performance, the chromatographic separation and mass spectrometry detection time of one SCP sample is still about 1 h. LC-MS analysis time has become the bottleneck of SCP throughput improvement. Multiplex labeling is a feasible method to improve the throughput of LC-MS analysis which has been applied in many SCP tools yet. As TMT labels have been expanded to 18 plex, the development of higher plex reagents may require expensive investments but is still worth looking forward to. Nonisobaric isotopologous mass tags such as mTRAQ have been used in low-input proteomics and combined with DIA-MS to improve throughput and protein coverage at the same time (Derks et al., 2022). As nonisobaric isotopologous mass tags may be easier to expand to multiplex than isobaric mass tags, its application in high-throughput single-cell proteomics is promising. Dephoure and Gygi (2012) described a hyerplexing method that enabled the analysis of samples from multiple conditions simultaneously by combining two different labeling methods, which may have implications for throughput enhancement in single-cell proteomics.

Here we focused on proteomics based on living single-cells in suspensions, but it is important to note that SCP based on trace cells from formalin-fixed paraffin-embedding (FFPE) tissues has come hand in hand. Although it is not yet possible to identify thousands of proteins from a single cell in FFPE tissues because of the impact of formaldehyde-mediated cross-links, there are several approaches that are approaching this goal. Using the optimized nanoPOTS, Nwosu et al. (2022) have identified an average of 1,312 from mouse liver tissues as small as 0.0025 mm^2×10 μm which corresponded to about 10 cells. We developed a spatially resolved proteomic tool called LCM-MTA which can quantify 536 proteins from 0.005 mm^2×8 μm human placenta FFPE tissue (about 15 cells) and 1,477 proteins from 0.1 mm^2×8 μm tissue (Gu et al., 2022a). Applied the LCM-MTA on clinical colorectal cancer (CRC) tissues, the functional differences of different cell types were accurately distinguished. Mund et al. (2022b) introduced the Deep Visual Proteomics (DVP), which combined artificial-intelligence-driven image analysis of cellular phenotypes with automated single-cell or single-nucleus laser microdissection and ultra-high-sensitivity mass spectrometry. By collecting about 100 cells for one sample, they have successfully characterized the expression of the proteome from melanocytes, melanoma in situ to invasive melanoma. Spatially proteomics in single-cell resolution can provide a new dimension to single-cell proteomics. Many hospitals and research institutions have massive amounts of FFPE tissue stored in their repositories. If reliable SCP tools can be applied to them, it will bring a great boost to biomedical research.

Summary

Overall, single-cell proteomics is in the early stage of explosive development. Just in 2019, analysis of the proteome from single cells was described as a “dream”, but today there have been several promising tools developed (Marx, 2019). We believe that with the optimization of accessibility and the further improvement of throughput, the truly large-scale applications of single-cell proteomics in scientific and clinical research, such as organ maps, drug screening, and precise disease classification, are within reach.

Chapter 5 Single-cell metabolomics technology

Most human cells are approximately 5 to 25 μm in diameter with as low as sub-pl intracellular volumes and highly dynamic metabolite concentrations ranging from a few copies to more than 100,000 (Zenobi, 2013). Compared with other omics, the genome is approximately static, the proteome and transcriptome change in minutes or hours, whereas the metabolome changes on a time scale of milliseconds to seconds (Weibel et al., 1974). For the present objects, the metabolome includes small molecules (usually lesser than 1.5 kD in size, but excluding nucleic acids, minerals, and salts), lipids, peptides, drugs, and their xenobiotics (Minakshi et al., 2019; Wishart et al., 2007). All of them are characterized by structural diversity, which makes discrimination difficult. And typically, a single cell can be detected tens to hundreds of analytes but only ~10% can be assigned using high-resolution MS and database search methods (Yin et al., 2018). Thus, extracting small volume content, snapshotting the quick turnover of metabolites, discriminating the molecular species diversity, improving detection sensitivity and selectivity, and boosting detection limits, all are inevitable challenges in single-cell metabolomics (SCM) research.

In the process of SCM research, a large number of research technologies, analytical platforms, and applications have emerged. Here, we review the development of SCM in the last ten years, including the classes of research techniques, mainly analytical workflow, applications, and possible breakthroughs.

Research techniques in single-cell metabolomics

To date, there are various research techniques for analyte measurement of a single cell, which are mainly divided into microscope-based, spectroscopy-based, and mass spectrometry-based platforms (Galler et al., 2014). Microscope-based technologies could observe cellular structures at the nanoscale, such as stimulated emission depletion (STED) microscopy, stochastic optical reconstruction microscopy (STORM), and photoactivated localization microscopy (PALM). The advantages of microscope-based analysis for single cells are as obvious as the disadvantages, with the highest spatial-resolution insight into cellular structures but the least biochemical information. Furthermore, most microscope-based methods have a low throughput limitation, and their long detection time is not suitable for dynamic analysis (Zheng and Li, 2012). Spectroscopy-based methods are widely applied, among which nuclear magnetic resonance (NMR) is the most used because of its characteristics of nondestructive detection and high reproducibility. However, multicellular analysis has dominated so far due to its relatively low sensitivity (Galler et al., 2014). Mass spectrometry-based methods are the indispensable tool for the simultaneous detection of a large number of analytes in a short period of time. They provide accurate mass-charge ratios, retention times, and quantitative results for both known and unknown molecules. In addition, molecules below sub-attomolar concentrations could be detected (Villas-Bôas et al., 2005). By contrast, MS wins out among these technologies for its high detection sensitivity and selectivity, broad detection range, fast analysis speed, and strong power of molecular structure identification. MS is considered the most powerful tool for characterizing the chemical profile of a single cell.

State-of-the-art technologies and methods in single-cell metabolomics field based on mass spectrometry

The analytical workflow in single-cell metabolomics based on MS mainly refers to single-cell sampling, content measurement, and data analysis. Single-cell sampling is the core of SCM. Single cells can be sampled directly or cultured on other platforms until metabolite analysis. For the purpose of the sampled content truly reflecting the metabolic profile (for example, neither loss of analyte volume nor misleading results caused by rapid metabolic turnover), sometimes additional treatments are needed to quench cell metabolism, including the addition of cold organic solvents or rapid freezing (Ibáñez et al., 2013). Care must be taken to avoid interfering with the culture media which may lead to the production of abnormal metabolites (Minakshi et al., 2019). During the content measurement by MS, molecules are ionized and converted to the gas phase, followed by passed into MS. Then moving ions are separated according to their mass to charge ratios in the magnetic field or electric field and detected by a detector. There are two main types of ion sources have been applied in single-cell metabolomics, laser desorption ionization and electrospray ionization (Figure 10). However, data acquisition could be challenging and the mode needs to be selected according to the research purpose. A large but indistinguishable number of metabolite features would be obtained if untargeted metabolomics is performed, whereas limited but definitive results would be acquired if targeted metabolomics is performed. Therefore, it is necessary to consider the tradeoff between throughput and coverage before SCM analysis (Tajik et al., 2022). In order to obtain the structural and functional information of metabolites, data analysis is carried out. The process of information mining partly determines the results of the research, which is important for the research. Therefore, we focus on single-cell sampling techniques and data analysis.

Sampling techniques in single-cell metabolomics based on Mass spectrometry

The single-cell sampling techniques for MS analysis can be broadly divided into three categories: (i) desorption ionization, (ii) content extraction, and (iii) sorting and ionization.

Desorption ionization

The intact single cells can be directly subjected to MS analysis where sampling and ionization processes simultaneously occur using the desorption ionization method (Liu and Yang, 2021). Desorption ionization can be divided into vacuum desorption and ambient desorption based on whether analytes are ionized in a vacuum system.

Secondary ion mass spectrometry (SIMS) (Figure 10A) uses a high-energy accelerated primary ion beam (e.g., Cs^+, O_2^+, Ar^+, and Ga^+) to bombard the target surface, which results in the ejection of plume of molecules or ions from the surface. SIMS is an effective technique for subcellular distribution imaging of various molecules with high spatial resolution (50 nm) (Yin et al., 2019a). While traditional primary ion beams induce extensive molecular fragmentation, modern SIMS often use cluster ions as its primary ion beam (e.g., Bi_3^+, SF_5^+, and C_{60}^+) to minimize fragmentation (Rubakhin et al., 2013). Nanoscale secondary ion MS (nanoSIMS) enables the primary ion beam to scan the sample at a perpendicular angle which shortens the working distance and improves secondary ion transmission. It has been applied for quantitation of subcellular chemical distribution with a lateral resolution of ~50 nm (Jiang et al., 2014).

Matrix-assisted laser desorption ionization (MALDI) (Figure 10B) is regarded as a soft ionization method that does not cause excessive fragmentation. Since the increasing ionization efficiency, it has become one of the most widely used laser desorption methods, which relies on the absorption of laser and the transfer of charges by auxiliary matrix molecules to enhance the analyte ionization. It can achieve resolution at the micron to

https://doi.org/10.1007/s11427-023-2561-0

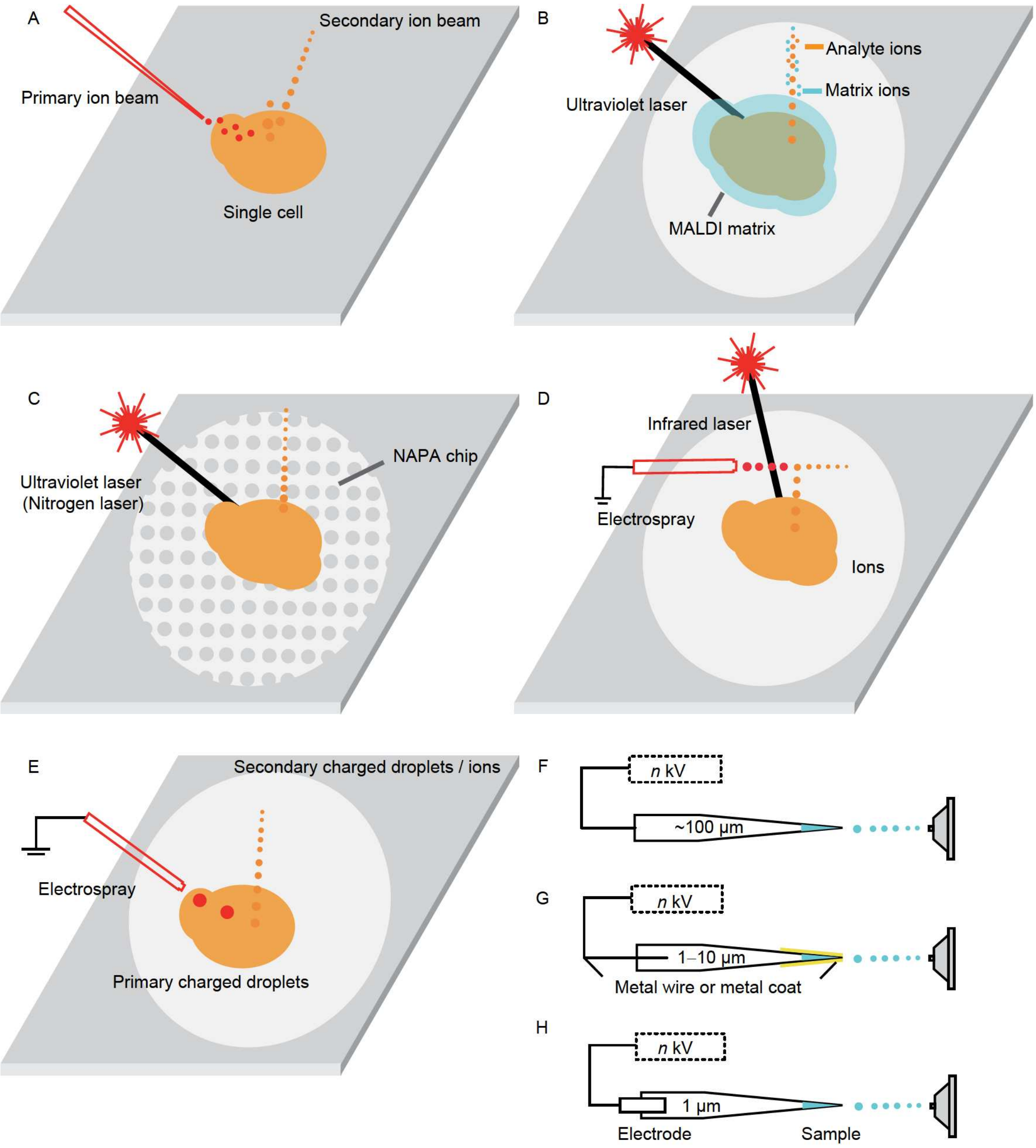


Figure 10. Examples of ionization techniques in single-cell metabolomics based on MS. A, Secondary ion MS (SIMS) (Wu et al., 2017a). B, Matrix-assisted laser desorption/ionization (MALDI) MS. C, Nanopost array (NAPA) MS (Minakshi et al., 2019). D, Laser ablation electrospray ionization (LA-ESI) (Stopka et al., 2018). E, Desorption electrospray ionization (DESI). F, Electrospray ionization (ESI). G, Nano-electrospray ionization (nano-ESI) (Bergman and Lanekoff, 2017). H, Pulsed direct current electrospray ionization MS (Pulsed-DC-ESI) (Wei et al., 2020).

submicron level (Emara et al., 2017) and provide high-fidelity results of native analyte distribution.

However, signals from the matrix molecules may strongly overlap with potential analytes (<500 D) (Ferguson et al., 2014). Multiple matrix-free laser desorption ionization (LDI) methods, such as desorption/ionization on porous silicon (DIOS) MS, Nanostructure-initiator MS (NIMS), and nanopost array (NAPA) MS (Figure 10C) were used. These ionizations rely on the interaction between laser radiation and nanostructures to contribute to the desorption and ionization of the sample, which solve the signal overlap and have comparable spatial resolution to MALDI (Yin et al., 2019a).

Traditional SIMS, MALDI, and matrix-free LDI methods are all vacuum-based. Qualitative and quantitative information on small molecular substances can be provided due to excellent spatial resolution. The optimized technology has high sensitivity and detection limits can reach the fg level (Yin et al., 2019a). A few thousand cells can be analyzed in a single experiment after generating single-cell arrays (Zhang and Vertes, 2018). However, some sample preparation steps, such as frozen dehydration, are introduced to maintain cellular shapes under high vacuum conditions (Zhang and Vertes, 2018). These steps may affect the chemical compositions of cells. The vacuum condition can also potentially interfere with the distribution of metabolites, particularly for volatile and semi-volatile species. Owing to technical innovations, atmospheric-pressure MALDI-MS (AP-MALDI-MS) has been developed with a lateral resolution as low as 1.4 μm (Kompauer et al., 2017).

Ambient ionization refers to the generation of ions under ambient conditions (e.g., native temperature and pressure) requiring little to no sample preparation. Laser ablation electrospray ionization (LA-ESI) (Figure 10D) is a matrix-free technique that utilizes a pulsed mid-infrared region laser beam at the wavelength of 2.94 μm to activate a water-rich target sample. At this wavelength, water strongly absorbs the laser radiation and creates a plume of molecules that are released into the gas phase (Nemes and Vertes, 2007). The desorption plume mixes with an ESI plume can enhance the ionization of analytes. LA-ESI eliminates sample preparation and has a spatial resolution as low as 30 μm (Shrestha and Vertes, 2009). A similar approach is desorption electrospray ionization (DESI) (Figure 10E), in which analyte ions are produced by desorption and ionization using electrospray directed toward the sample surface. However, the limited spatial resolution (>50 μm) usually prevents it from SCM analysis (Taylor et al., 2021). Nanospray desorption electrospray ionization (nano-DESI) utilizes a primary capillary for solvent delivery on cell samples and a secondary capillary for picking up the extracted molecules for MS analysis. Its resolution is determined by the size of the liquid bridge formed between two capillaries, which is controlled by the capillary's size, their position, and the flow rate (Yin et al., 2019b). The addition of shear force probes standardizes capillary-to-sample distance (Nguyen et al., 2017). Then, a pneumatically assisted nano-DESI device was implemented to propel the solvent through the nanospray capillary, which improved sensitivity for metabolite species by 1–3 orders of magnitude and reduced ionization suppression (Duncan et al., 2017). These reduce the dependence on probe-to-surface distance. The resolution that can be achieved with current nano-DESI technology is 8.5 μm (Rao et al., 2015).

The sensitivity, spatial resolution, coverage, and throughput vary with the different desorption ionization methods (Figure S2 and Table S8 in Supporting Information) (Taylor et al., 2021). At present, desorption ionization MS methods are mainly used in mass spectrometry imaging (MSI) which employs an analytical probe (e.g., ion beam, laser, and solvent junction) capable of analytes desorption and ionization *in situ*. MSI can provide additional functional information by mapping the location of small molecules *in situ*, which is promising (Taylor et al., 2021). SpaceM integrated MALDI imaging with light microscopy and digital image processing. It took the first microscope image to capture the relative positions of cells, then collected MALDI imaging of metabolites, followed by a second microscope image to show a visual cue which cell the metabolite came from. SpaceM could detect >100 metabolites from >1,000 individual cells per hour (Rappez et al., 2021). It has the most identifications among the known MSI techniques. If the target sample is sectioned consecutively and each section is used for MSI, the 3D spatial metabolite map will be obtained after the compilation of the 2D MS images. Dueñas et al. (2017) utilized MALDI-MSI to visualize the three-dimensional spatial distribution of phospholipid classes in individual zebrafish embryos.

Content extraction

Electrospray ionization methods (Figure 10F) are also extensively used for biological molecule analysis of single cells, especially for live cells, because of the significantly reduced mechanical and chemical perturbations and greatly simplified sample preparation. ESI-MS generally favors the detection of analytes at relatively high concentrations due to its relatively low ionization efficiency, ion transmission, and relatively high ion suppression. Naturally, the modified techniques have been developed and applied in SCM, including nano-electrospray ionization (nano-ESI) (Figure 10G) (Karas et al., 2000), probe electrospray ionization (PESI) (Gong et al., 2014), induced nano-ESI (InESI) (Huang et al., 2011) and pico-electrospray ionization (pico-ESI) (Wei et al., 2020) (Figure 10H).

Metabolome acquired directly from a living cell *in situ* can result in a more realistic and representative chemical profile of cell metabolism and phenotype. The direct sampling analysis method is content extraction which can be divided into micromanipulation, microextraction, and microjunction probes (Figure 11).

Micromanipulation means manipulating a micropipette to gently pick individual cells and suck out metabolites. Micromanipulation coupled MS mainly uses nano-ESI capillary whose emitter internal diameter is closer to the MS inlet. An application, known as live single-cell MS (live MS) (Figure 11A), was achieved by sucking out the content with a metal-coated microcapillary under video-microscopy, adding an ionization solvent (acetonitrile containing 0.5% formic acid) from the microcapillary rear, and directly feeding the mixture into MS (Mizuno et al., 2008). The extracted content characterized hundreds of molecules at sub-attomolar-level sensitivity within minutes (Fujii et al., 2015). However, the analytes were diluted tens of thousands of times due to the ionization reagent. Subsequently, PESI was used to enrich and extract metabolites by inserting a tungsten probe with a tip diameter of 1 μm into the single cell for ~30 s (Gong et al., 2014). Both have the disadvantage of controlling the imprecise amount of extracted material from cells. Consequently, quantitative extraction techniques of pressure assisted (Zhang and Vertes, 2015) or electroosmotic (Yin et al., 2018) microsampling were developed.

Most small molecules from a single cell can not be directly detected by MS due to the presence of intracellular interfering ions and high concentrations of non-volatile salts. Thus, liquid-liquid extraction serves different analyte classes (Figure 11B). Multiple microextraction devices coupled with MS are proposed to achieve a high coverage metabolic analysis by adding low volumes of an extraction solvent. In short, the capillary tip absorbs organic solvent and aqueous solution respectively in positive and negative ion mode at nano-liter or pico-liter scale, followed by connected to a syringe and a clamp which enable extraction reagent to cover a single cell for a few seconds (e.g., 10 s) under an inverted microscope and to complete the extraction (Wang et al., 2019c). In general, the microextraction partially alleviates the problem of MS incompatibility with intracellular interfering ions and salts, but the high viscosity of the cell contents also needs assistant ionization solvents to obtain ion signals which limit metabolite coverage.

The characteristics of the above content extraction, sampling and then ionization, prohibit the real-time detection. Liquid microjunction probes achieve in real-time, *in situ* metabolite extraction with integrated solvent microextraction and nanoscale micromanipulators. Yang's group (Pan et al., 2014) introduced multiple devices, including single-probe (Figure 11C) and T-probe. Single-probe is fabricated by embedding a fused silica capillary and a nano-ESI emitter into a dual-bore quartz needle. Droplets at the tip of the needle formed by continuously injecting liquid from one side of the needle can extract intracellular material and they are discharged through the other side and sent to MS. T-probe works similarly, with three capillaries embedded into T-shaped grooves, where the solvent-

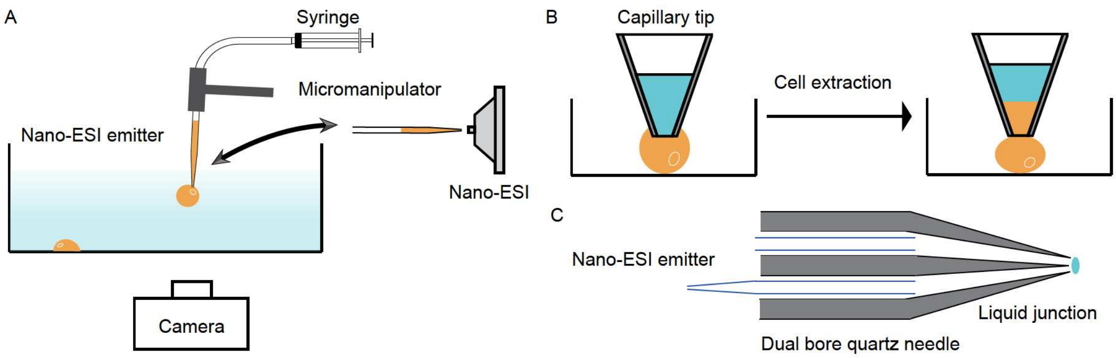


Figure 11. Examples of content extraction techniques in single-cell metabolomics based on MS. A, Live single cell MS (live MS) (Tajik et al., 2022). B, Microextraction strategy (Yin et al., 2018). C, Single-probe diagram (Pan et al., 2020).

providing capillary is in line with the nano-ESI emitter and the sampling capillary is vertically placed (Liu et al., 2018b). Single-probe is used in the single cells residing in microwells and T-probe is modified to analyze live non-adherent cells. In terms of operation difficulty, the content extraction of adhesive cells is easier to achieve than suspended cells (Emara et al., 2017).

In brief, the method of content extraction can not only realize living cell studies but also obtain relatively high metabolite coverage. However, the main limitations of the content extraction method are relatively tedious manipulation, low sample throughput, and time-consuming (3–5 min per cell) (Table S8 in Supporting Information) (Fujii et al., 2015).

Sorting and ionization

Most of the time, single-cell analysis begins with sample preparation in a bid to isolate the target cell without affecting its state. There are a large number of single-cell isolation techniques in a high-throughput manner that have been developed so far. Since single-cell sampling occurs after sorting, we call this method sorting and ionization for short. It can be divided into label sorting-based, microfluidic device-based and LCM-based methods (Figure 12).

Conventional cell sorting methods, including flow cytometry, FACS, and mass cytometry, are label-based. Flow cytometry flows a cell at a time by controlling the cell suspension. Labeled with fluorescent markers, cells can shed light into various properties when a laser beam scatters through them. Similarly, FACS also uses light scattering and fluorescence properties to sort cells into subpopulations. As a combination of flow cytometry and MS, mass cytometry, of which antibodies are labeled with heavy metal ion tags instead and detected by inductively coupled plasma (ICP), has been used for sorting and targeted high-throughput molecular analysis (Bandura et al., 2009). Mass cytometry is currently limited to a couple of dozen available proteins and is not shown in metabolite analysis because the small molecules are difficult to label. Nonetheless, a label-free mass cytometry realizes online sorting and real-time ESI-MS analysis for a single cell. CyESI-MS uses three coaxial capillaries to deliver cell suspension, sheath fluid, and sheath gas, respectively (Figure 12A). Cells are isolated and extracted by the sheath fluid, then the sheath gas assists solvent evaporation and ensures the ions enter MS, which could simultaneously detect hundreds of cellular metabolites in a high-throughput way, approximately 38 cells per minute (Yao et al., 2019).

Microfluidic devices bring a significant enhancement in the throughput and simplification of the workflow. They physically confine individual cells to microfluidic structures, among which micro/nano-wells, droplets, microvalve-controlled channels, and hybrid microfluidic platforms are most extensively used in single-cell analysis (Liu et al., 2019).

Micro/nano-well-based microfluidic devices, also known as chip-based methods, consist of dense arrays of wells that typically are lithographically fabricated onto polydimethylsiloxane (PDMS), glass, or silicon and serve as containers for individual cells (Torres et al., 2014). For example, the early invention of an integrated microfluidic array plate (iMAP) was characterized by the interface of gravity driven flow, open input fluid exchange and cell capture mechanism with approximately 100% capture rate (Dimov et al., 2011). Its design allowed for single-cell capture, reagent addition, and parallel processing operations. Castro et al. (2021) deposited a small volume of buffer containing dense-core vesicles and electron lucent vesicles of Aplysia californica cells onto an indium tin oxide (ITO)-coated glass slide. Ibanez and colleagues developed microarrays for MS (MAMS) (Figure 12B) that allowed thousands of individual cells to be analyzed in a single MS experiment, which featured arrays of hydrophilic wells patterned on an omniphobic surface to enable automated isolation of single cells (Ibáñez et al., 2013). Yang et al. (2016) revised the fabrication process of the microdot array by using the contact printing technique.

Droplet-based microfluidic devices usually use two immiscible fluids to create water-in-oil micro/nanodroplets containing the individual cell as single-cell reaction vessel. A single cell in the droplets could be achieved by limited dilution, but the probability of single-cell events is limited. Combined an inkjet nozzle cell manipulator with PESI-MS, a drop-on-demand inkjet printing device was fabricated and used for lipid profiling, which was capable of producing single-cell events with a probability of about 50% in a fully automatic manner (Chen et al., 2016a). In the other example, Lin's group (Huang et al., 2018b) designed a Dean flow assisted cell ordering system (Figure 12C) to detect multiple cellular lipids, of which a spiral capillary generated a secondary force to separate particles in a single equilibrium position and reduce the agglomeration.

Microvalve-based devices utilize parallel microchannel circuits coupled to pressure-controlled valves or similar control devices (Figure 12D) (Unger et al., 2000). By means of precise control of the microvalve assembly, a series of operations automate and

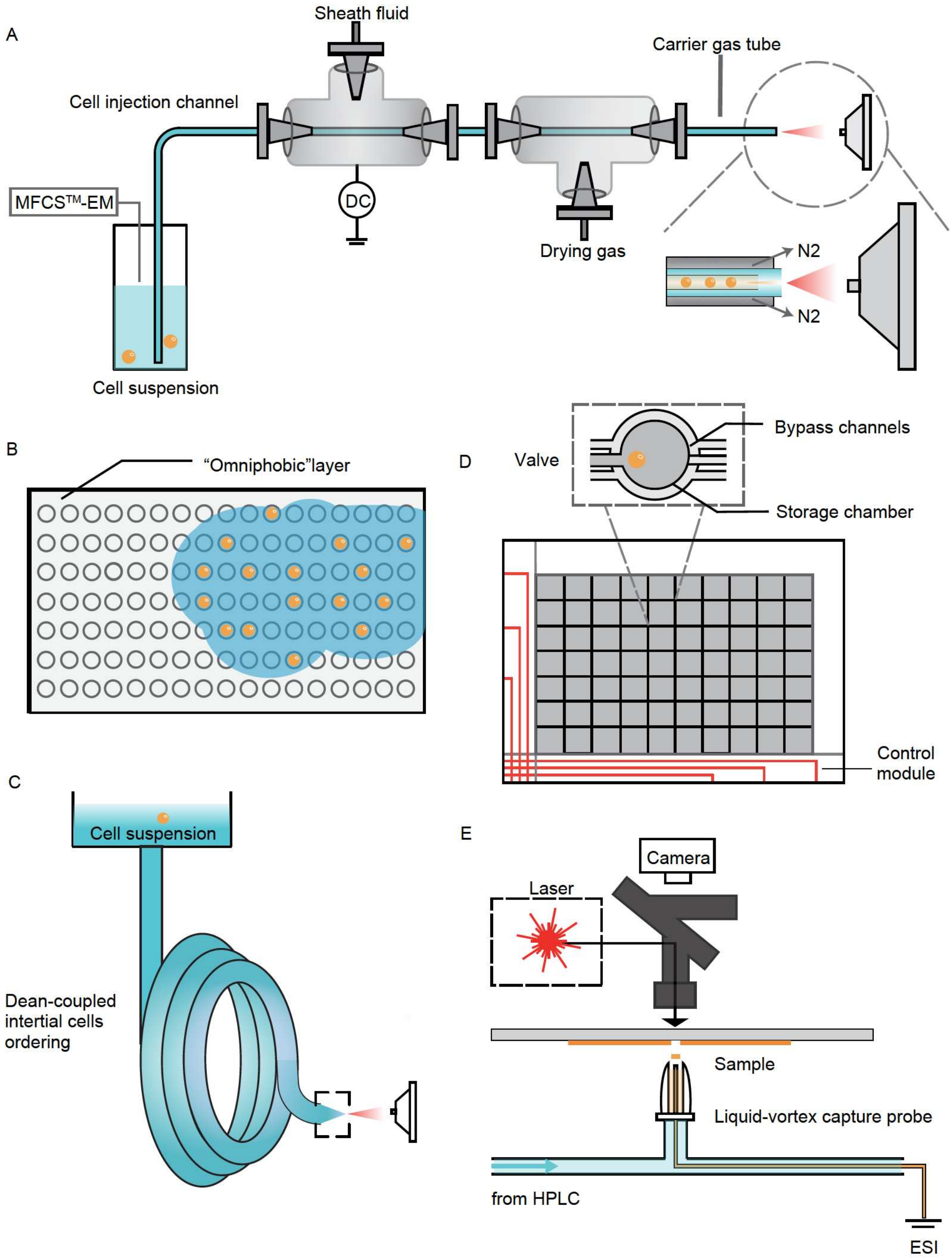


Figure 12. Examples of sorting and ionization techniques in single-cell metabolomics based on MS. A, Label free mass cytometry (CyESI-MS) (Yao et al., 2019). B, Microarrays for MS (MAMS) (Ibáñez and Svatos, 2020). C, Dean flow assisted cell ordering system (Huang et al., 2018b). D, Microvalve-based microfluidic device (Leung et al., 2012). E, Laser capture microdissection/liquid vortex capture MS (LMD-LVC-MS) (Cahill and Kertesz, 2020).

parallelize complex biological analysis.

The microfluidic platforms aforementioned have their own advantages and drawbacks. A promising approach is to combine the core components from different microfluidic platforms and overcome each other's limitations to form a new hybrid platform, whose common name is lab-on-a-chip system. Leung et al. (2012) developed a microfluidic device. The programmable microdroplets-based device combined integrated microvalve technology with the sample compartmentalization and dispersion-free transport to perform single-cell manipulation and analysis. Furthermore, a multi-dimensional organic mass cytometry was established by connecting a simple microfluidic chip to the nanoelectrospray emitter, enabling the identification of about 100 metabolites with a throughput of around 40 cells per minute (Xu et al., 2021a) (Table S8 in Supporting Information). The multi-step integration is not only beneficial to obtain high coverage results but also to save time and labor.

Sorting and selection of single cells or subcellular components can be done by the LCM system which typically consists of a microscopy component for sample visualization, a laser component for selectively dissecting samples, and a collection component for material dissected. It is not considered a high-throughput method, but a high-resolution method suitable for single-cell isolation from tissue section samples. LCM can completely isolate the cell from its natural environment. There has developed a high spatial resolution hybrid laser capture microdissection/liquid vortex capture/mass spectrometry system (LMD-LVC-MS) (Figure 12E) with a liquid vortex capture probe placing directly below the sample substrate that captured the laser-ablated material, dissolved the material into liquid, and transported it to MS for analysis (Cahill and Kertesz, 2018).

It is worth mentioning that although the above sampling techniques can be applied to live cells, developing technologies under near-physiological conditions remains challenges. But there have been some technological breakthroughs. Shao et al. (2022b) came up with an intact living-cell electrolaunching ionization mass spectrometry (ILCEI-MS) method, which used a capillary with an inner diameter slightly smaller than the average cell diameter to achieve cell separation and transport with the help of a small device. It reduced the volume of ionized droplets formed under the combined action of an applied electric field and the surface tension on the port before reaching the MS inlet. A droplet was roughly equal to a cell. Through this method, 51 cells could be analyzed per minute, and 368 metabolites (from 482 cells) could be assigned in a single experiment. Recently, an asymmetric serpentine channel microfluidic chip coupled to pulsed electric field-induced electrospray ionization-high resolution mass spectrometry (chip-PEF-ESI-HRMS) conditions have been developed, which was sheathless and external-force-unused. The single cells were suspended in an aqueous solution (i.e., isotonic salt concentration). Once a single cell reached the tip of nanospary emitter, the high voltage electric field made it disruption and the contents ionization and identification in real time. It allowed for the annotation of approximately 120 metabolites in a single cell and the throughput of up to 80 cells per minute (Feng et al., 2022).

Data analysis in single-cell metabolomics based on Mass spectrometry

The workflow of current single-cell metabolomics data analysis includes data preprocessing, metabolite annotation, statistical analysis, network and pathway analysis, and data visualization. A large number of bioinformatics tools, analytical software, and databases are now available at each step (Liu and Yang, 2021; Misra, 2020) (Table S9 in Supporting Information) Data preprocessing includes two parts. First, convert the raw data only available to the commercial software of specific vendor into a compatible format. Second, extract metabolome related information which involves determining “true” signals, normalizing relative abundances among different cells, and screening out the metabolites present in most cells. Except for the methods mentioned in Table S9 in Supporting Information, more and more custom algorithms have been written to extract information (Liu and Yang, 2021; Shao et al., 2022b). Then some software is needed to recognize metabolites, which is metabolite annotation. Next is statistical analysis. In order to reduce batch effects and technical variations, data processing is carried out, including normalization, transformation, and scaling. After evaluating whether the data has Gaussian distribution or not, parametric or nonparametric univariate analysis (e.g., *t*-tests and analysis of variance) and multivariate analysis (e.g., unsupervised principal components analysis and supervised orthogonal partial least-squares discriminant analysis) are performed to reveal the metabolomic biomarkers, group clustering results, and discrimination between groups according to the experimental design. By integrating conventional statistical methods with machine learning to build complex mathematical models, it can provide high predictive classification results. The open-source statistical analysis tools include, but are not limited to, R (http://www.R-project.org) and Python (https://www.python.org). Mapping of metabolites onto metabolic maps or known biochemical pathways is network and pathway analysis. Finally, the above results are visualized to complete the data analysis.

Applications

Subtle differences from cell to cell may lead to great changes in important biological processes. The bulk analysis of cell population shows the average features of multiple cell types and ignores the rare cells. Therefore, it is necessary to study individual cells. At present, SCM has been applied to various studies and has made more or less progress.

It is commonly used to identify single-cell metabolites, explore the metabolic profiles and compare the up-regulation and down-regulation of metabolites in normal and other states, especially for cancer cells. It is also applied to quantify compounds (Pan et al., 2019), visualize cell heterogeneity (Huang et al., 2018b), differentiate cell subsets (Zhang et al., 2018), and screen drugs (Anchang et al., 2018). For example, abnormal lipid synthesis (e.g., C=C bond position or sn-position isomers formation) could result in different diseases and reflect the prospect of lipidomics in precision medicine (Li et al., 2021d). It served to investigate the role of cells in key biological processes, including drug resistance (Prieto-Vila et al., 2019), immune response (Labib and Kelley, 2020), tumor metastasis (Wu et al., 2020), and cell fate determination (Stirparo et al., 2018). Also, an important application is to study the special properties of rare cells, such as CTCs, cancer stem cells, and antigen-specific T cells. Take CTCs for example, they are released into the bloodstream from primary tumor lesions and cause metastases in distant tissues and organs. Abouleila et al. (2019) revealed the metabolic profile differences between CTCs and lymphocytes and found that the synthesis of GPLs was a key factor in cancer proliferation. Moreover, SCM is also used in plants. For example, Yuan et al. (2023) used metabolomics-assisted breeding in watermelons.

Summary

Current single-cell metabolomics methods can analyze a maximum of about 3,000 cells in a single experiment, mostly 500–1,000 (Feng et al., 2022). Developing automation technology is beneficial. For example, armed with an automated system and a recognition algorithm, a dispenser robot coupled to a motorized *x-y* stage enables to pick up the target cell or organelle quickly (Emara et al., 2017).

The destructive nature of MS limits the repeatability and temporal analysis. Therefore, a combination of different analytical tools is recommended to provide comprehensive and multidimensional information. For example, a modified patch clamp setup was combined with InESI-MS to simultaneously capture the electrophysiological and metabolic state of a neuron (Zhu et al., 2017). At the same time, it is necessary to integrate with single-cell multiomics to plot more detailed characteristic profiles.

How to get highly confident results is a much more important thing than crafts. Since the standardization procedure of single-

https://doi.org/10.1007/s11427-023-2561-0

cell metabolomics, from sample preparation to data analysis, has not been established, the following work should follow it. We summarize the technological progress over the last decade in Table S8 in Supporting Information. Balancing sensitivity, coverage, and throughput, none of these methods are perfect. So, it is necessary to develop new methodologies to improve all aspects. In summary, SCM is still in the early stage, it is going to continue to flourish.

Chapter 6 Single-cell multimodal sequencing technology

Multiomics analysis with bulk sequencing has been widely applied to provide a comprehensive understanding of biological processes like disease development (Hoadley et al., 2018; Hutter and Zenklusen, 2018; Liu et al., 2018a; Malta et al., 2018) and tissue development (Consortium et al., 2020a; Consortium et al., 2020b; He et al., 2020b; Sethi et al., 2020) through the atlas integration of multi-omics datasets like genomic, transcriptomic, epigenomic and proteomic data in multiple species. At the single-cell scale, the applications of unpaired single-cell multi-omic sequencing technologies, which use different cells from the same or similar source for different single-cell experiments, have been applied to discover new cell subpopulations and new biological mechanisms by connecting the single-cell transcriptome for cell type identification to different modalities of other similar cells (Argelaguet et al., 2019; Hao et al., 2021b). Recent advances in single-cell sequencing technologies have further enabled the measurement of multiple omics like DNA, mRNA, epigenomic, and protein in the same cell at single-cell resolution, providing the paired and high-resolution discovery of single-cell status. Also, to integrate the multi-omic single-cell datasets, several bioinformatic algorithms and methods have been developed to help pre-process, integrate, and interpret the emerging multi-omic single-cell datasets.

In this chapter, we first review the (i) recent development of single-cell joint profiling technique capturing multiple views of cell molecules from the same cell, mainly focused on the transcriptome-focused multimodal technologies (Table 3). Next, we review the (ii) recent advances in multi-modal integration analysis methods and tools, including the different categories of applications and algorithms for both unpaired multimodal datasets and paired multimodal datasets. Also, we summarize the performance of popular single multi-omic data integration methods from recent single-cell multi-modal integration benchmark studies.

Single-cell multimodal sequencing technology

Multiple types of molecules can be isolated from the same captured cell by single-cell multi-omics technologies. Several approaches of single-cell multi-omics sequencing designed for capturing genomic DNA (gDNA), transcriptome, proteome, and epigenome have been developed in recent years. Major steps of single-cell multimodal sequencing technology workflow as depicted in Figure 13.

Transcriptome+gDNA

Several multi-omic technologies can simultaneously measure mRNA and gDNA in a single cell. Genome and transcriptome sequencing (G&T-seq) (Macaulay et al., 2015) applied flow cytometry cell isolation along with beads-based mRNA and gDNA separation. gDNA-mRNA sequencing (DR-seq) (Dey et al., 2015) isolated cells by pipette and then amplified and split tagged gDNA and mRNA. Simultaneous isolation of genomic DNA and total RNA (SIDR) (Han et al., 2018a) selected cells with microplates and separated nucleus and cytoplasm by hypotonic cytosis. And TARGET-seq (Rodriguez-Meira et al., 2019) optimized the steps of FACS for cell isolation and reverse transcription polymerase chain reaction (RT-PCR) for amplification, which provided higher cell throughput than previous methods.

Transcriptome+epigenome

Bisulfite (BS) treatment can convert methylated and unmethylated DNA CG sites (Frommer et al., 1992) and analyze the DNA methylation at single nucleotide resolution by PCR and next generation sequencing (Grunau et al., 2001; Harris et al., 2010). Several single-cell bisulfite sequencing methods that measure the methylation level at single-cell scale have been developed, including single-cell reduced representative bisulfite sequencing (scRRBS) (Guo et al., 2013), single-cell whole genome bisulfite sequencing (scWGBS) (Smallwood et al., 2014), single-nucleus methylcytosine sequencing (snmC-seq) (Luo et al., 2017), and single-cell combinatorial indexing for methylation (sci-MET) (Mulqueen et al., 2018) (see "Epigenome sequencing" section for more detail). Recently, several single-cell multi-omic techniques have been developed to capture mRNA and gDNA methylation in the same cell. Firstly, single-cell methylome and transcriptome sequencing (scM&T-seq) (Angermueller et al., 2016) used a similar protocol as G&T-seq (Macaulay et al., 2015), which used flow cytometry cell isolation along with beads-based mRNA and gDNA separation followed by bisulfite treatment. Secondly, simultaneous single-cell methylome and transcriptome sequencing (scMT-seq) (Hu et al., 2016b) used micro pipetting to isolate the nucleus of the single cells, then performed scRRBS and a modified Smart-seq2 procedure to generate DNA methylome and transcriptome data, respectively. As an extension of scMT-seq (Hu et al., 2016b), scTrio-Seq (Hou et al., 2016) can analyze genomic CNVs, the DNA methylome, and the transcriptome for individual cell simultaneously, as the genomic CNVs can be computationally inferred from scRRBS by bulk RRBS data.

Several next-generation-sequencing-based techniques, such as chromatin immunoprecipitation followed by sequencing (ChIP-seq) (Blecher-Gonen et al., 2013; Johnson et al., 2007), Dnase I hypersensitive site sequencing (Dnase-seq) (Boyle et al., 2008), and assay for transposase-accessible chromatin using sequencing (ATAC-seq) (Buenrostro et al., 2013) have been developed to investigate the epigenome profiles such as chromatin structure and histone modifications in many species (Consortium et al., 2020a). Another similar method, nucleosome occupancy and methylome sequencing (NOMe-seq) (Kelly et al., 2012) can label accessible genomic regions using an exogenous M. CviPI GpC methyltransferase and simultaneously measure nucleosome occupancy and cytosine methylation level. Based on these methods, many new protocols have been developed to measure the chromatin accessibility as well as DNA methylation or histone modification in chromatin accessible sites at single-cell resolution, including single-cell Dnase sequencing (scDNase-seq) (Jin et al., 2015), single-cell combinatorial indexing assay for transposase-accessible chromatin with sequencing (sci-ATAC-

Table 3. Recent single-cell multimodal technologies

Multimodal techniques	Modalities	References
Genome and transcriptome sequencing (G&T-seq)	gDNA+mRNA	(Macaulay et al., 2015)
gDNA-mRNA sequencing (DR-seq)	gDNA+mRNA	(Dey et al., 2015)
Simultaneous isolation of genomic DNA and total RNA (SIDR)	gDNA+mRNA	(Han et al., 2018a)
TARGET-seq	gDNA+mRNA	(Rodriguez-Meira et al., 2019)
Single-cell methylome and transcriptome sequencing (scM&T-seq)	mRNA+Methylation	(Angermueller et al., 2016)
Simultaneous single-cell methylome and transcriptome sequencing (scMT-seq)	mRNA+Methylation	(Hu et al., 2016b)
scTrio-Seq	mRNA+Methylation	(Hou et al., 2016)
sci-CAR	mRNA+ATAC	(Cao et al., 2018)
SNARE-seq	mRNA+ATAC	(Chen et al., 2019d)
Paired-seq	mRNA+ATAC	(Zhu et al., 2019b)
SHARE-seq	mRNA+ATAC	(Ma et al., 2020)
10x Multiome	mRNA+ATAC	https://www.10xgenomics.com/cn/blog/introducing-chromium-single-cell-multiome-atac-gene-expression
PEA/STA	mRNA+proteome	(Genshaft et al., 2016)
PLAYR	mRNA+proteome	(Frei et al., 2016)
CITE-seq	mRNA+proteome	(Stoeckius et al., 2017)
REAP-seq	mRNA+proteome	(Peterson et al., 2017)
RAID	mRNA+proteome	(Gerlach et al., 2019)
scNMT-seq	mRNA+methylation+ATAC	(Clark et al., 2018)
scNOMeRe-seq	mRNA+methylation+ATAC	(Wang et al., 2021b)
ECCITE-seq	mRNA+sgRNA+target protein	(Mimitou et al., 2019)
Paired-Tag	mRNA+ATAC+5 histone modifications	(Zhu et al., 2021a)
scCUT&Tag-pro	5 histone modifications+proteome	(Zhang et al., 2022a)

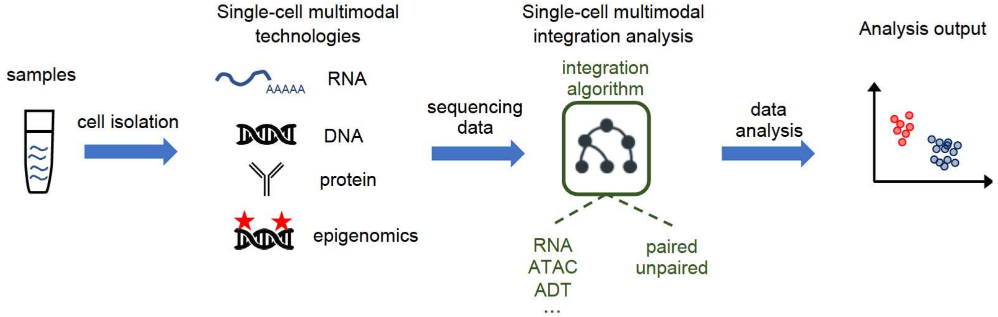


Figure 13. Major steps of single cell multimodal sequencing technology workflow.

seq) (Cusanovich et al., 2015), single-cell assay for transposase-accessible chromatin using sequencing (scATAC-seq) (Buenrostro et al., 2015), single-cell micrococcal nuclease sequencing (scMNase-seq) (Lai et al., 2018) and single-cell chromatin immunoprecipitation sequencing (scChIP-seq) (Rotem et al., 2015) which can measure the H3 lysine 4 tri-methylation (H3K4me3) and di-methylation (H3K4me2) modifications. For more details, please see the Epigenome sequencing section.

Based on these single-cell epigenomic technologies, several single-cell multi-modal high throughput methods targeting both chromatin accessibility and transcriptome have been developed. Cao et al. (2018) developed sci-CAR, the first protocol that can jointly profile the mRNA and ATAC in the same cell. SciCAR applied combinatorial indexing for each cell, and then redistributed cells by FACs and lysate splitting, and then amplified for sequencing. Cao et al. (2018) applied sci-CAR to human and mouse cell line mixture and mouse kidney tissue and identified *cis*-regulatory network from the joint profiling datasets. However, due to the high sparsity in the scATAC modality and limited coverage in scRNA modality of sci-CAR, only a minority of differentially accessible sites and differentially expressed genes in bulk scRNA and scATAC sequencing can be discovered by single-cell datasets in sci-CAR. Chen et al. (2019d) developed the droplet-based single-nucleus chromatin accessibility and mRNA expression sequencing (SNARE-seq), which enhanced the sequencing coverage of both scRNA and scATAC, and improved the coverage limitation in sci-CAR. SNARE-seq used a splint oligonucleotide with sequence complementary to the adaptor sequence inserted by ATAC transposition (5′ end) and the mRNA poly(A) bases (3′ end), which allowed to capture both omics data.

https://doi.org/10.1007/s11427-023-2561-0

Compared with sci-CAR, SNARE-seq detected 4–5 times more chromatin accessible sites in the mouse postnatal brain dataset and adult brain dataset than sciCAR tissue dataset, and improved the throughput by a cellular combinatorial indexing strategy (Preissl et al., 2018). Zhu et al. (2019b) further improved the protocol and developed the parallel analysis of individual cells for RNA expression and DNA accessibility by sequencing (Paired-seq), which adopted a ligation-based combinatorial indexing strategy to simultaneously tag both the open chromatin fragments by the Tn5 transposases and the cDNA molecules by RT of RNA. Zhu et al. applied Paired-seq to mouse embryonic cerebral cortex tissue and applied integrated analysis with ENCODE mouse embryonic cerebral cortex tissue datasets, reconstructed the cellular trajectory, and recovered the cis-regulatory network from the dual-omic dataset. More recently, Ma et al. (2020) developed simultaneous high-throughput ATAC and RNA expression with sequencing (SHARE-seq), which used multiple rounds of hybridization blocking to joint labeling mRNA and chromatin fragments in the same single cell. Compared with sciCAR (Cao et al., 2018), SNARE-seq (Preissl et al., 2018), and Paired-seq (Zhu et al., 2019b), SHARE-seq showed higher scalability on much larger library size for more than 30,000 cells and higher sensitivity with more genes and ATAC peaks detected in each cell than previous multi-modal methods. Based on higher data quality, Ma et al. applied a new definition of domains of regulatory chromatin (DORCs) rather than individual peaks to analyze the regulatory map between chromatin accessibility and gene expression, and identified prior functions of DORCs to gene expression in cell lineage choice and cell fate decision (Preissl et al., 2018). Recently, 10x Genomics developed 10x Multiome, a commercial service platform for joint profiling of scRNA and scATAC in the single cell, which would accelerate the applications of scRNA and scATAC multi-modal techniques in more biological and clinical research.

Transcriptome+proteome

Besides single-cell multi-modal technologies targeting DNA and RNA molecules, several single-cell multi-modal methods that can measure RNA and protein simultaneously in the same cell were developed. Genshaft et al. (2016) developed proximity extension assay/specific RNA target amplification (PEA/STA) method. PEA/STA applied reverse transcriptase as the DNA polymerase for both RT of RNA and extension of proximity extension assay (PEA) DNA oligos for 38 proteins to enable cDNA synthesis and PEA to proceed in the same reaction in the Fluidigm C1^{TM} system (DeLaughter, 2018). Frei et al. (2016) developed proximity ligation assay for RNA (PLAYR), a method for highly multiplexed transcript quantification using flow and mass cytometry, which is also compatible with standard antibody staining. Using the mass cytometry, PLAYR allowed simultaneous measurement of more than 40 mRNAs and proteins, and enabled the characterization of the interplay between transcription and translation at single-cell level. Besides protein-DNA ligation strategy, two methods targeting surface protein and mRNA, cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq) (Stoeckius et al., 2017), and RNA expression and protein sequencing assay (REAP-seq) (Peterson et al., 2017) were developed to detect both mRNAs and cell surface proteins using oligonucleotide-labeled antibodies, enabled the multimodal analysis at single-cell scale by droplet-based single-cell sequencing approaches. These two methods greatly improved the throughput of the transcriptome. For example, REAP-seq can quantify proteins with 82 barcoded antibodies and measure more than 20,000 genes in a single workflow (Peterson et al., 2017). Another method, single-cell RNA and immunodetection (RAID) (Gerlach et al., 2019), can detect intracellular proteins and phosphorylated proteins together with mRNAs. RAID immunostained the intracellular target proteins with antibodies conjugated with RNA barcodes, and then converted proteins into RNAs.

Techniques capturing more than two modalities

Based on these bi-modal single-cell methods we discussed above, several methods were developed to capture more than two omics in the same cell. Single-cell nucleosome, methylation and transcription sequencing (scNMT-seq) (Clark et al., 2018) were developed by combining scM&T-seq (Angermueller et al., 2016) and NOMe-seq to measure nucleosome, transcriptome, and DNA methylome in the same cell. Recently, Wang et al. (2021b) developed scNOMeRe-seq, which enabled the profiling of genome-wide chromatin accessibility, DNA methylation, and transcriptome in the same individual cell and applied this method for a single-cell multi-omics map of mouse preimplantation development. Based on CITE-seq, Mimitou et al. (2019) developed expanded CRISPR-compatible cellular indexing of transcriptomes and epitopes by sequencing (ECCITE-seq) to capture mRNA, sgRNA and designed target proteins for at least five modalities in the same single cell. By adapting Paired-seq (Zhu et al., 2019b) with cleavage under targets and tagmentation (CUT&Tag) strategy (Kaya-Okur et al., 2019), Zhu et al. (2021a) further developed parallel analysis of individual cells for RNA expression and DNA from targeted tagmentation by sequencing (Paired-Tag), a novel protocol which can simultaneously profile scRNA, scATAC and five histone modifications in the same cell. Also, Zhang et al. (2022a) developed scCUT&Tag-pro, which combined CUT&Tag with CITE-seq and captured five histone modifications by CUT&Tag library and proteins by antibody-derived protein tags library.

Multi-omics integration analysis

Recent advances in single-cell multi-modal technologies provided substantial data resources to uncover the molecular mechanism by multi-view high-dimensional and high-resolution datasets. However, it is hard to properly integrate the multi-modal single-cell datasets, arising from high dataset dimensionality, high data sparsity as well as complex variables among multimodal datasets and techniques. An increasing number of algorithms were developed for different applications and tasks in multi-modal data integration analysis, indicating the rapid progression as well as growing attention from researchers to the field of single-cell multi-modal data integration and data analysis. Also, several review papers (Adossa et al., 2021; Argelaguet et al., 2021; Lance et al., 2022; Xu and McCord, 2022) and benchmark studies (Brombacher et al., 2022; Luecken et al., 2022) have performed specialized classification as well as general evaluations for emerging multi-modal integration tools, providing great benefits for other researchers to select appropriate methods from different types of integration tools. Here, we introduce a comprehensive set of multimodal integration tools and related studies from the perspectives of the following two sections, including (i) categories of multimodal integration tools, which introduce recently published single-cell multi-modal integration tools by different classification standards, and (ii) recent benchmark studies for these recently published tools.

Categories of multimodal integration tools

Based on previous review papers (Adossa et al., 2021; Argelaguet et al., 2021; Stanojevic et al., 2022), several criteria can be applied to classify multimodal integration tools. Firstly, multimodal integration tools can be classified with four major integration strategies based on the choice of the shared features (known as anchors) for data integration, including horizontal integration with all genomic features like genes as anchor, vertical integration with all common cells as anchor, diagonal integration with no shared features and mosaic integration with partially shared cells and partially shared genomic features as anchor (Argelaguet et al., 2021). Several tools with similar methodologies can also be developed as different types of integration tools. For example, for algorithms based on nonnegative matrix factorization (NMF), iNMF (Gao et al., 2021a) was designed as a vertical integration tool, coupledNMF (Gao et al., 2021a) was designed as a diagonal integration tool, and UINMF (Kriebel and Welch, 2022) was designed as mosaic integration tool.

Secondly, the multimodal integration tools can also be grouped by the types of multi-modal techniques, including "paired" and "unpaired" integration tools (Brombacher et al., 2022; Stanojevic et al., 2022). The paired multimodal integration tools were specifically designed for multimodal datasets simultaneously captured and sequenced from the same cell. The vertical integration and mosaic integration strategies are usually applied in paired integration tools, as the paired datasets share all or partial common cells between different modalities. The unpaired integration tools were designed for integrating independent single-cell experiments from different modalities, as the cells of one modality cannot find matched cells from the other modalities. Due to the difference in both cells and features, diagonal integration was commonly applied for unpaired data integration. Several tools were developed for paired or unpaired multimodal dataset integration specifically. For example, the Seurat v3 (Stuart et al., 2019) was designed for unpaired multimodal datasets, and the updated version—Seurat v4 (Hao et al., 2021b) was specifically designed for paired multi-modal datasets. Also, some integration tools (Hu et al., 2022b; Lin et al., 2022; Zhang et al., 2021c) can be applied to both paired and unpaired multi-modal datasets.

Thirdly, based on the methodologies, the multimodal integration tools can be categorized into several sub-types (Stanojevic et al., 2022), including mathematical matrix factorization methods, manifold alignment methods, network-based methods, and deep learning methods. The deep learning integration tools can be further categorized by the infrastructure of the deep model, including autoencoder (AE), generative adversarial network (GAN), GNN, and their extended structure such as variational autoencoder (VAE). The selection of the methodology is largely determined by the type of multimodal dataset and the integration tasks. For unpaired multimodal datasets, manifold alignment methods can first reduce the different features of multimodal datasets into the same dimension of latent embeddings/manifolds, and then integrate the heterogeneous modalities by the same manifolds (Argelaguet et al., 2021; Stanojevic et al., 2022). Similarly, matrix factorization methods can be applied for different integration tasks by matrix factorizing unmatched features or cells to the matrix of the same dimension with less information loss than simple dimension reduction methods along with manifold alignment (Stanojevic et al., 2022). The deep learning tools using GAN (Khan et al., 2022; Xu et al., 2021b; Zhao et al., 2022a), VAE (Ashuach et al., 2021; Lotfollahi et al., 2022; Minoura et al., 2021) and transformer (Li et al., 2022b) can learn the common latent embedding from different modalities of same cells or shared cells and then impute the missing cells and features, as GNN (Cao and Gao, 2022; Ma et al., 2021a) model is applied to learn the relationship between different types of features (gene in scRNA and peak in scATAC for etc.) and infer biological network in multimodal data integration (Cao and Gao, 2022; Ma et al., 2021a).

Fourthly, the multimodal integration tools can be classified based on certain omics for integration. Several multimodal integration algorithms were designed for specific multi-modal datasets integration; for example, CiteFuse (Kim et al., 2020a) was designed for CITE-seq (Peterson et al., 2017) analysis, SCIM (Stark et al., 2020) was designed for scRNA and CyTOF integration, while scMVP (Li et al., 2022b) was designed for paired scRNA and scATAC datasets integration. Also, besides these tools restricted to specific omics datatypes, several algorithms like LIGER (Welch et al., 2019) were designed for general integration tasks without restrictions on integration omics types.

The multimodal integration tools can also be categorized by major coding languages like Python and R, and special integration applications like cross modality translation. All multimodal integration tools along with their categories are summarized in Table S10 in Supporting Information.

Benchmarks for single-cell multimodal integration tools

Although plenty of multimodal integration studies have been published for different tasks in multi-modal single-cell analysis, it is still difficult to find state-of-the-art methods from published integration methods. To solve this issue, recently, several benchmark studies have been performed for the evaluation of single-cell multimodal integration tools for different tasks of multi-modal dataset analysis (Brombacher et al., 2022; Lance et al., 2022; Luecken et al., 2022). These benchmark studies would provide comprehensive and objective evaluations of the performance of these candidate integration tools from the perspective of data users. Next, we introduce recent benchmark studies and summarize the performance of multimodal integration tools from these third-party evaluation studies.

Luecken et al. (2022) performed a benchmark analysis of single-cell integration tools for tasks of atlas level data integration and developed a benchmark pipeline for objective, comprehensive, and reproducible evaluation of single-cell integration tools. This study included several unpaired multimodal single-cell integration tools; however, it only focused on the integration tasks for different datasets from the same modalities, like integration of different scRNA datasets and integration of different scATAC datasets, but did not provide evaluation for cross modality integration of paired or unpaired scRNA and scATAC data. Nevertheless, this study provided a stable, comprehensive, and highly scalable benchmark framework for single-cell atlas integration evaluation.

Recently, to better accomplish the analysis challenges arising from data sparsity, technical and biological variability, and high dimensionality from single-cell multimodal integration analysis, NeuraIPS2021 launched an online competition for three major tasks in single-cell multimodal data integration analysis (Lance et al., 2022), including (i) predicting one modality from another, (ii) matching cells between modalities, and (iii) jointly learning representations of cellular identity. Among the three tasks, the second task is specifically designed for unpaired multimodal integration tools, and the third task is designed for paired multimodal integration tools. Also, the competition launcher generated the first single-cell multimodal benchmarking datasets, including a multi-center CITE-seq dataset with 90,000 cells for scRNA and protein integration tasks and a multi-center 10x Multiome dataset with 70,000 cells for scRNA and scATAC integration tasks. Among all three tasks, the CLUE (cross-linked universal embedding) algorithm, a semi-supervised modality matching function in GLUE (Cao and Gao, 2022) package, won the first prize and got all categories winner in the second modality aligning task, showing the top performance of cross modality matching among unpaired integration tools (Lance et al., 2022). However, as the competition only evaluated the algorithms from online submitters, most published single-cell multimodal integration tools were not included in the benchmark evaluation.

For further comparison of published single-cell multimodal integration tools with deep learning framework, Brombacher et al. (2022) first reviewed 18 recently published multimodal integration tools using deep learning model, and then performed the first comprehensive benchmark study for selected popular tools using CITE-seq dataset and 10x Multiome dataset from NeuraIPS2021. For biology preservation tasks, Cobolt (Gong et al., 2021) showed the best performance among benchmark algorithms on both CITE-seq and 10x Multiome datasets, only for larger cell numbers, scMVP (Li et al., 2022b) has better performance than Cobolt (Gong et al., 2021) on 10x Multiome dataset. For technique effect removal tasks, SCALEX (Xiong et al., 2021b) showed consistent top performance on CITE-seq dataset, and scMVP (Li et al., 2022b) showed the highest performance on 10x Multiome dataset.

Summary

In this chapter, we have provided a summary of recent advances in multiple types of multi-omic single-cell sequencing techniques, and their bioinformatics integration methods. With improvements in experimental data quality and the performance of bioinformatics algorithms, single-cell multi-modal technologies have provided comprehensive multi-modal insights into different eras at the single-cell level. Moving forward, expanding the biological applications of single-cell multimodal techniques, as well as increasing the performance and robustness of single-cell multi-modal algorithms, would undoubtedly accelerate new discoveries in biological and medical research. These improvements hold significant potential to revolutionize our understanding of cellular processes and the development of personalized medicine.

Chapter 7 Single-cell spatial transcriptomics technology

In the preceding sections, we systematically reviewed the recent advances in single-cell omics. Although these single-cell sequencing technologies allow the investigation of cellular heterogeneity at an unprecedented resolution, they are far from adequate to get a full understanding of the intricate workings of multicellular organisms. Many studies emphasize that the state of one cell is not only regulated by the intracellular regulatory network but also interfered by the extracellular signals from the environment (Dries et al., 2021a; Junttila and de Sauvage, 2013; Lin and Hankenson, 2011). Both the dissociation of tissues and the isolation of single cells during experimental procedures cause the loss of critical spatial information, including cell positions and their mutual proximities. Spatial transcriptomics (ST) has addressed this limitation, enabling the measurement of gene expression with spatial information preserved. In this section, we will introduce spatial transcriptomics technologies, discuss computational methods for spatial data analysis, and provide a review of their applications in various biological systems. Additionally, we will also delve into the current progress in techniques of spatial multi-omics.

Techniques for spatially resolved transcriptomics

All current spatial transcriptomics techniques can be broadly summarized into three categories majorly based on (i) microdissection, (ii) barcoding, and (iii) imaging, respectively (Figure 14A–C). These ST technologies differ in their approaches to location labeling and transcript profiling, which may determine the spatial resolution, detection efficiency, demanding sample types, and so on. Next, we will discuss the principles of a selection of representative techniques from each category and summarize their characteristics. The curated list of techniques and their corresponding features is shown in Table 4.

Microdissection-based ST techniques

Techniques falling within this category aim to computationally reconstruct the 3D structure of tissues from multiple spatially proximal tissue subregions isolated by various microdissection approaches (Figure 14A). For instance, RNA tomography (tomo-seq) obtains RNA from a series of sequential cryosections along three orthogonal axes in multiple putatively identical biological samples (Junker et al., 2014). The requirement of identical biological samples limits the application of tomo-seq on human samples. By comparison, STRP-seq slices tissues into primary sections and then secondary stripes using a two-level dissection strategy, which assumes that spatial expression patterns are constant between consecutive primary sections spaced 14 μm apart (Schede et al., 2021). Based on cryosectioning, Geo-seq utilizes LCM to section tissues into regions as small as around 10 cells (Chen et al., 2017b). Other methods within this type includes ProximID (Boisset et al., 2018) and PIC-seq (Giladi et al., 2020), which focus on physical cell interaction within two (doublets) or three cells (triplets), rather than the positions or surrounding context in the tissue.

In addition to physical sectioning, microdissection could be accomplished by combining optical marking with fluorescence-based cell selection, or photo-cleavage of gene index oligos. For example, transcriptome *in vivo* analysis (TIVA) loads TIVA tags (i.e., photoactivatable mRNA capture molecules) into live cells and selects cells by laser photoactivation, which subsequently triggers tags' hybridization to mRNA (Lovatt et al., 2014). As an alternative technology, NICHE-seq injects labeled landmark cells into transgenic mice expressing photoactivatable green fluor-

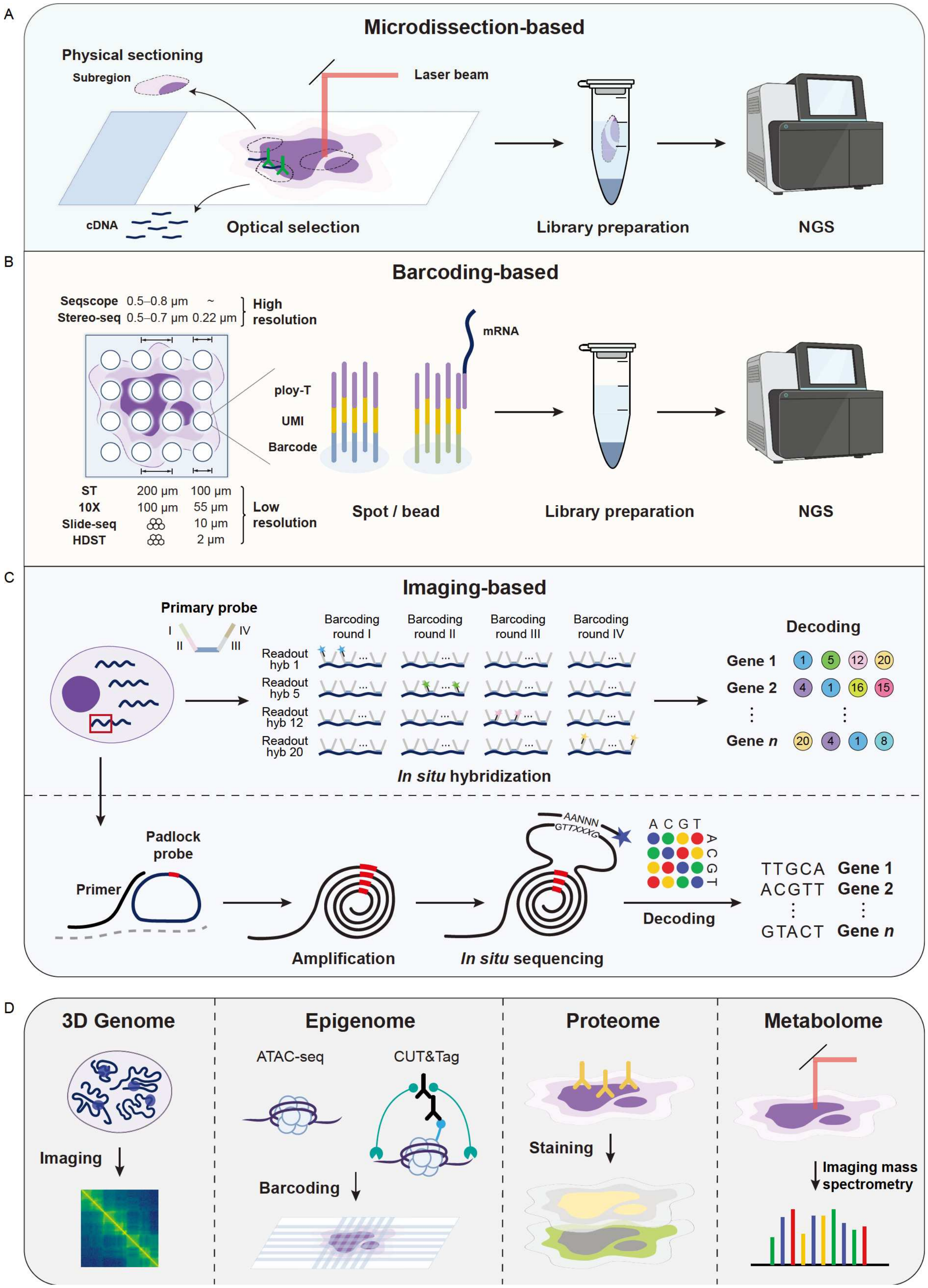


Figure 14. Schematics of techniques for spatial transcriptomics and other omics. A, For microdissection-based techniques, the sub-regions of interest can be isolated from the sample by physical sectioning or optical selection, and collected for library preparation and next-generation sequencing (NGS). B, For barcoding-based techniques, barcodes are attached to spots or beads for position labeling and *in situ* mRNA capturing. Barcoded cDNA are collected for subsequent library preparation and sequencing. C, For imaging-based techniques, with probes designed for genes of interest, *in situ* hybridization (ISH) or *in situ* sequencing (ISS) can be performed for *in situ* profiling. D, Schematics of spatial techniques for other omics, including 3D genome, epigenome, proteome and metabolome.

Table 4. The curated list of spatial transcriptomics techniques and their corresponding features

Techniques	Features	Type	Spatial resolution	Gene coverage	References
tomo-seq	Require identical biological samples; enable the 3D reconstruction of tissues	Microdissection-based	18 μm	Transcriptome-wide	(Junker et al., 2014)
STRP-seq	Require consecutive thin slices	Microdissection-based	10 cells	Transcriptome-wide	(Schede et al., 2021)
Geo-seq	Enable the 3D reconstruction of tissues	Microdissection-based	10 cells	Transcriptome-wide	(Chen et al., 2017b)
PIC-seq	Focus on physical cell interaction rather than spatial positions	Microdissection-based	Cellular	Transcriptome-wide	(Giladi et al., 2020)
ProximID	Focus on physical cell interaction rather than spatial positions	Microdissection-based	Cellular	Transcriptome-wide	(Boisset et al., 2018)
TIVA	Require the loading of capture tag into cells; rely on photoactivation	Microdissection-based	Cellular	Transcriptome-wide	(Lovatt et al., 2014)
NICHE-seq	Work on genetically engineered mice; select region of interest by photoactivation	Microdissection-based	Cellula	Transcriptome-wide	(Medaglia et al., 2017)
GeoMX DSP	Rely on photocleavage; appliable in protein detection	Microdissection-based	20–40 cells	Targeted (~1,500 genes)	(Merritt et al., 2020)
ST	With H&E image	Spatial barcoding	100 μm	Transcriptome-wide	(Ståhl et al., 2016)
10x Visium	With H&E or immunohistochemistry image	Spatial barcoding	55 μm	Transcriptome-wide	(Ståhl et al., 2016)
Slide-seq(V2)	Without histology on the same tissue section	Spatial barcoding	10 μm	Transcriptome-wide	(Stickels et al., 2021)
HDST	Low sensitivity	Spatial barcoding	2 μmr	Transcriptome-wide	(Vickovic et al., 2019)
Stereo-seq	Enable subcellular analysis; allow large field of view	Spatial barcoding	0.5–0.7 μm	Transcriptome-wide	(Chen et al., 2022)
Seq-scope	Enable subcellular analysis	Spatial barcoding	0.5–0.8 μm	Transcriptome-wide	(Cho et al., 2021)
PIXEL-seq	Enable subcellular analysis	Spatial barcoding	1 μm	Transcriptome-wide	(Fu et al., 2021b)
DBiT-seq	Enable simultaneous measurement of RNA and proteins	Spatial barcoding	10 μm	Transcriptome-wide	(Liu et al., 2020b)
seqFISH+	Use a barcode palette of 60 pseudo colours	Fluorescence imaging (ISH-based)	Subcellular	Targeted (~10,000 genes)	(Eng et al., 2019)
MERFISH	Based on multi-bit binary encoding strategy; combine expansion microscopy with *in situ* hybridization	Fluorescence imaging (ISH-based)	Subcellular	Targeted (~10,000 genes)	(Xia et al., 2019)
STARmap	Combine hydrogel-tissue chemistry with *in situ* sequencing	Fluorescence imaging (ISS-based)	Subcellular	Targeted (~1,000 genes)	(Wang et al., 2018)
FISSEQ	Use partition sequencing	Fluorescence imaging (ISS-based)	Subcellular	Transcriptome-wide	(Lee et al., 2015)
ExSeq	Combine expansion microscopy with *in situ* sequencing	Fluorescence imaging (ISS-based)	Subcellular	Transcriptome-wide	(Alon et al., 2021; Shah et al., 2016)
Slide-DNA-seq	Use barcoded bead arrays to capture spatially resolved DNA	Spatial barcoding	10 μm	Genome-wide	(Zhao et al., 2022b)
spatial-CUT&Tag	Combine *in situ* CUT&Tag chemistry with microfluidic deterministic barcoding	Spatial barcoding	20 μm	Genome-wide	(Deng et al., 2022a)
spatial-ATAC-seq	Combine *in situ* Tn5 transposition chemistry and microfluidic deterministic barcoding	Spatial barcoding	20 μm	Genome-wide	(Deng et al., 2022b)
DNA-MERFISH	Enable simultaneous imaging of genomic loci and nascent transcripts	Fluorescence imaging	Subcellular	Genome-wide	(Su et al., 2020)
CosMX SMI	Enable quantification of RNA and proteins	Fluorescence imaging	Subcellular	64 proteins, 1,000 genes	(He et al., 2021a)
SpaceM	Combine light microscopy and MALDI-imaging MS	Fluorescence imaging	Cellular	>100 metabolites	(Rappez et al., 2021)
Perturb-map	Combine a protein barcode system and multiplex imaging	Fluorescence imaging	Cellular	35 genes (120 Pro-Codes)	(Dhainaut et al., 2022)

escent (PA-GFP), allowing *in situ* labeling of niches of interest. After tissue dissociation, activated PA-GFP$^+$ cells are sorted by FACS for single-cell transcriptome profiling (Medaglia et al., 2017). The commercial GeoMX Digital Spatial Profiler (DSP) developed by NanoString employs probes with UV cleavable linkers and automates the optical selection (Merritt et al., 2020).

Overall, microdissection coupled with single-cell or bulk RNA sequencing makes it possible to study the transcriptome within the spatial context. Microdissection can be performed in a physical manner, or an optics-dependent way. The physical sectioning is often implemented manually, making the dissection protocol labor-intensive and time-consuming. In contrast, optics-dependent sectioning generally depends on the loading of specialized tags into live cells or genetic engineering in model organisms, which restricts its application to fresh-frozen or FFPE human samples. No matter how microdissection and sequencing are performed, the exact positions of profiled cells within the selected subregion are unknown, resulting in a generally low spatial resolution.

Barcoding-based ST techniques

Microdissection-based techniques trace the spatial information by manually labeling each subregion. The spatial barcoding techniques enable automatic recording of spatial coordinates (Figure 14B). In 2016, Ståhl et al. (2016) pioneered the application of barcoding techniques in the ST technology. In this approach, barcodes, together with UMIs and poly(T) oligonucleotides are immobilized on glass slides to allow *in situ* capture of mRNA and cDNA synthesis. Each barcoded spot in the array is 100 μm in diameter and is positioned 200 μm center-to-center apart from the adjacent ones, which brings about a resolution of 10–40 cells. 10x Genomics has further enhanced the spatial resolution to 5–10 cells using spots with diameters of 55 μm and center-to-center distances of 100 μm. Instead of directly attaching barcodes to slides, some techniques link barcodes to beads for position labeling and mRNA capture. For example, Slide-seq deposits 10-μm DNA-barcoded beads onto a surface (Rodriques et al., 2019; Stickels et al., 2021). Similarly, HDST places barcoded beads into an array with 2-μm wells (Vickovic et al., 2019). Both of these technologies improve spatial resolution to 1–2 cells. However, as the barcoded beads are randomly distributed on the slide, *in situ* sequencing (ISS) or *in situ* hybridization (ISH) is required to decode each fixed bead's barcode sequence.

Although bead-based techniques can reach a cellular resolution, they are still too coarse to detect subcellular differences. Recently, Seq-scope was developed to achieve a 0.5–0.8 μm center-to-center resolution by repurposing the Illumina sequencing platform (Cho et al., 2021). Another technique enabling sub-micrometer-resolution profiling is Stereo-seq (Chen et al., 2022), where 220 nm DNA nanoballs (DNBs) containing barcodes are deposited on a patterned array with a center-to-center distance of 500 or 715 nm. Both Seq-scope and Stereo-seq require two-round sequencing, in which the first associated barcodes with spatial locations and the second provides information of captured cDNA, as performed in Slide-seq.

To summarize, the barcoding-based approaches combine spatial barcoding techniques with NGS to allow transcriptome-wide profiling of RNA in the spatial context. The technologies involve a trade-off between spatial resolution and detection efficiency. Compared with the original ST technology or commercialized 10x Visium, the improvement in spatial resolution by Seq-scope, Stereo-seq tends to come at the expense of low detection sensitivity and low gene coverage.

Imaging-based ST techniques

Both microdissection-based and barcoding-based techniques extract nucleic acid molecules for NGS sequencing after position labeling. To preserve RNA *in situ*, various *in situ* transcriptomic techniques were developed for the spatial mapping of gene expression, including ISH and ISS (Figure 14C). Because these methods necessitate fluorescence imaging, they are collectively known as imaging-based techniques.

Most ISH-based ST techniques mainly rely on single-molecule RNA fluorescence *in situ* hybridization (smFISH) (Femino et al., 1998) to enable quantitative measurements of targeted transcripts *in situ*. SeqFISH belongs to this type, which allows the simultaneous detection of multiple mRNA molecules by sequential rounds of fluorescent hybridization, imaging, and stripping of readout probes (Lubeck et al., 2014). Using the seqFISH strategy, all the targeted genes are encoded by the combination of rounds of readout probes. SeqFISH+ expands the readout probe palette from four or five colors in seqFISH to 60 “pseudo colors” (Eng et al., 2019), enabling the multiplexing of up to 10,000 genes in a single cell. MERFISH is another smFISH-based technique, which also requires multiple rounds of hybridization, but employs a distinct multi-bit binary encoding strategy (Chen et al., 2015a). To address the issue of optical crowding, expansion microscopy (ExM) was integrated into MERFISH (Xia et al., 2019). The encoding strategy, in conjunction with ExM, allows MERFISH to reduce the number of hybridization rounds. For example, to ensure the detection of ~10,000 genes, using 3-color imaging, seqFISH+ needs 80 (4×20) rounds of hybridization, while MERFISH only needs 23 rounds to construct a 69-bit HD4 code with a Hamming weight of 4 (Zhuang, 2021).

In addition to the techniques based on multiplexed FISH, *in situ* profiling of RNA can also be achieved by ISS, which sequences RNA in the fixed tissue or cell sample with *in situ* signal amplification. Due to the limited cellular space, some of the ISS-based techniques select a part of genes by designing probes to target specific RNA or cDNA. The initial ISS approach published in 2013 uses padlock probes to bind to targets (Ke et al., 2013), followed by rolling-circle amplification (RCA) to generate RCA products for subsequent sequencing by ligation. STARmap uses two-component padlock probes to directly bind to RNA rather than cDNA, avoiding the inefficient step of RNA-to-cDNA and reducing potential noise (Wang et al., 2018). To diminish strong background fluorescence brought by conventional supported oligo ligation detection (SOLiD) sequencing, sequencing with error-reduction by dynamic annealing and ligation (SEDAL) was devised for STARmap, which allows error rejection during sequencing.

Besides targeted ISS methods, ISS could be conducted in an untargeted manner, in which transcripts are reversely transcribed to cDNA, followed by DNA amplification and sequencing without pre-selection of genes. While the untargeted manner could improve the coverage to transcriptome-wide, it can also lead to molecule crowding. To mitigate it, FISSEQ leverages a partition sequencing strategy (Lee et al., 2015), where only a small fraction of amplicons is randomly selected and sequenced using extended sequencing primers, therefore resulting in low detection efficiency. Combined with ExM, FISSEQ was adapted to another approach called ExSeq to discriminate between crowded molecules and increase spatial resolution (Alon et al., 2021).

In general, imaging-based techniques offer high spatial resolution, reaching cellular or even subcellular levels. Among these techniques, ISH-based ones, which rely on prior knowledge of target genes, exhibit high detection efficiency. By comparison, due to the limitations of ISS, ISS-based techniques have comparatively low efficiency, especially in untargeted ones.

Moreover, most of these techniques necessitate specialized equipment for high-resolution imaging, which may limit their broader applicability.

Techniques for spatial multi-omics

To achieve a more comprehensive characterization of cells, considerable efforts have been directed towards the measurement of other modalities in the spatial context, including genome, epigenome, proteome, metabolome, and so on (Figure 14D). The positioning strategies used in ST technologies have been adapted to realize spatial profiling of other omics. For instance, Slide-DNA-seq captures spatially resolved genomic sequences using a barcoded bead array which was initially developed for spatial RNA profiling (Zhao et al., 2022b). Similarly, spatial-CUT&Tag (Deng et al., 2022a) and spatial-ATAC-seq (Deng et al., 2022b) were developed to profile histone modification and chromatin accessibility by combining DbiT-seq's microfluidic deterministic barcoding strategy (Liu et al., 2020b) with *in situ* CUT&Tag chemistry and Tn5 transposition chemistry, respectively. To gain an understanding of 3D chromatin conformation within its native context, a MERFISH-based method was designed to visualize over 1,000 genomic loci for high-resolution chromatin tracing (Su et al., 2020).

In the realm of proteome, protein expression could be readily visualized by multiplexed immunohistochemistry (IHC). IHC can be further coupled with imaging mass cytometry (Giesen et al., 2014) or multiplexed ion beam imaging (MIBI) (Angelo et al., 2014) to allow the simultaneous imaging of ~100 proteins. Moreover, proteins of interest can be targeted by DNA-barcoded antibodies and thus quantified by NGS, as in GeoMx DSP (Merritt et al., 2020). Cell-surface proteins can be bound by antibodies without cell lysis, preventing damage to RNA. Therefore, proteomics could be combined with transcriptomics in both single-cell and spatial omics. For example, the enhanced version of 10x Visium conducts IHC prior to mRNA capturing to enable the co-detection of protein and RNA, albeit only allowing for the detection of 1–2 proteins. By adding antibody-derived tags to fixed tissue slides before flow barcoding, DbiT-seq enables the co-measurement of mRNA and dozens of proteins (Liu et al., 2020b). Additionally, NanoString offers the CosMx SMI platform, enabling the quantification of 1,000 RNA and 64 protein analytes through high-plex imaging (He et al., 2021a).

Metabolites collected from samples are often quantified using MS. For the study of spatially resolved metabolome, various techniques have been developed based on imaging mass spectrometry (IMS). These techniques differ in the manners by which ions are produced from molecules of samples, including MALDI (Rappez et al., 2021), DESI (Yin et al., 2019b), and SIMS (Passarelli et al., 2017). For example, SpaceM is a MALDI-based method for *in situ* single-cell metabolomics (Rappez et al., 2021). It addresses the challenge of assigning metabolite intensities to individual cells by integrating MALDI-imaging with light microscopy followed by computational methods for image segmentation and registration.

In addition to intrinsic genetics, many gene functions are influenced by the spatial context (Haigis et al., 2019). To study spatial functional genomics, Dhainaut et al. (2022) established an approach called Perturb-map, which enables pooled CRISPR screens at the single-cell resolution in the tissue context. This is achieved by employing a protein barcode system and multiplex imaging.

Computational methods for spatial transcriptomics

A standard workflow for single-cell analysis encompasses critical tasks such as cell clustering, cell-type annotation, differential expression analysis, lineage tracing, cell-cell communication, and integration analysis. These tasks also form the backbone of ST data analysis. Spatial transcriptomics, with its unique capacity to provide information about spatial proximity and context, not only broadens the analytical scope but also poses great integration challenges. To address these, a large number of computational methods have been developed to integrate gene expression with spatial information and provide new insights (Figure 15). We will review the methods designed for different purposes in the forthcoming sections. A list of published computational methods is presented in Table S11 in Supporting Information.

Denoising to enhance the signal in spatial transcriptomics

As discussed above, many ST techniques face challenges related to low detection efficiency and significant noise. These issues arise from shallow sequencing for each spatial unit (i.e., spot or bead) or complex experimental steps needed to preserve the tissue structure, or a combination of both (Wang et al., 2022f). Wang et al. (2022f) have shown in 10x Visium and Slide-seq data that the signal noise was reflected in both the drop-outs and the inflation of gene expression. While denoising methods have been developed for scRNA-seq data to address the drop-out problem, they often struggle to correct the "false" high expression. Furthermore, these single-cell methods rely solely on transcriptomics data, and thus could not be directly applied to integrate additional spatial information.

Several computational methods have been specially developed to tackle the denoising ST data. For example, Sprod could impute gene expression in noisy ST data from barcoding-based techniques based on latent graph learning (Wang et al., 2022f). The denoising process in Sprod involves two steps. First, Sprod builds a graph by leveraging spatial proximity and expression similarity. Importantly, if available, features extracted from the corresponding pathological images could be incorporated for graph construction. Next, Sprod corrects gene expression for each spot/bead by borrowing expression information from the neighborhood revealed in the graph. Another method, spARC, adopts a similar graph-based framework but demonstrates its serviceability on imaging-based ST techniques (Kuchroo et al., 2022). SiGra is also a graph-based method but employs a different approach to build the graph (Tang et al., 2022). It utilizes three graph transformer autoencoders for imaging, transcriptomics, and hybrid, respectively, as well as the attention mechanism, enabling SiGra to enhance the sparse and noisy transcriptomics data with multi-modal spatial information. The SME method in stLearn also allows the integration of image features to normalize spatial gene expression (Pham et al., 2020). It employs a simple strategy of weighted average, where the weights are calculated based on the morphological similarity between close spots.

Rather than random drop-outs or inflation, Ni et al. (2022) believed that the loss and inflation are caused by the bleed of mRNA between and among nearby spots, which is referred to as spot swapping. To adjust for the effects of spot swapping, they proposed a method called SpotClean (Ni et al., 2022). SpotClean employs a probabilistic framework to model gene-specific

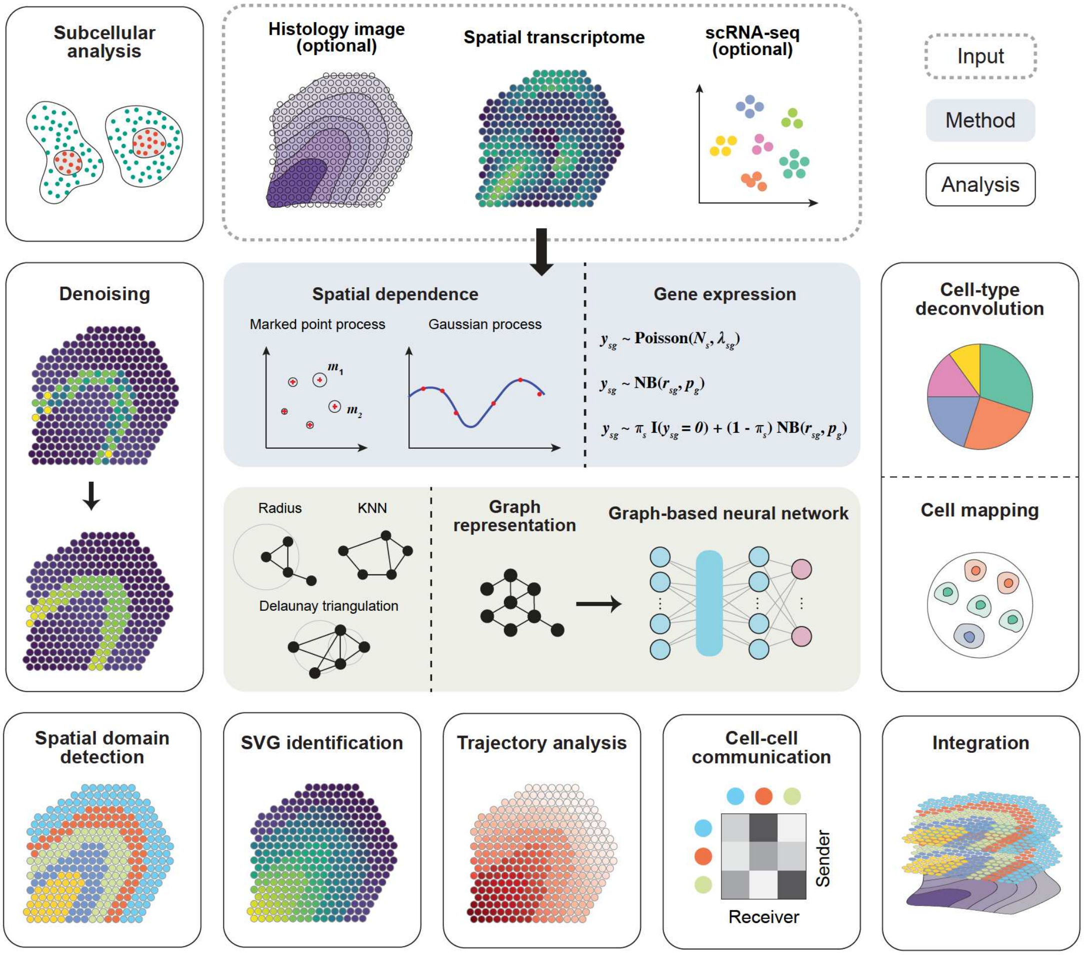


Figure 15. Overview of spatial transcriptomics analysis. With spatial gene expression taken as input, as well as optional histology images and matched scRNA-seq, a variety of analyses could be performed, including data denoising, cell type annotation, spatial domain detection, identification of spatially variable genes (SVG), pseudo-time trajectory analysis, and cell-cell communication analysis. Besides, subcellular analysis is available for high-resolution data such as imaging-based and high-resolution barcoding-based ST data. If multiple samples or multiple modalities are provided, integrative analysis can also be performed. To enable these analyses, most of the computational methods rely on probabilistic modeling or graph building to represent spatial gene expression. For probabilistic modeling-based methods, the spatial dependence between spatial locations could be modeled by marked point process or Gaussian process, and the gene expression could be modeled by Poisson, negative binomial (NB) or zero-inflated negative binomial (ZINB) distributions. For graph-based methods, neighborhood graphs could be constructed by specifying a fixed distance, or by k-nearest neighbors (KNN) or Delaunay triangulation alternatively, and then used as the input of graph-based neural networks for different analysis tasks.

expression at a given spot, which considers reads present in tissue at that spot and reads bleeding out into other spots, and also removes reads bleeding in from other spots. The authors demonstrated that SpotClean could accurately estimate gene-specific UMI counts in technologies such as ST and 10x Visium, where background positions could be identified by the alignment between the ST slide and the matched H&E image.

Subcellular analysis for imaging-based ST data

Imaging-based ST techniques provide good opportunities for cellular or even subcellular analyses but also pose great challenges. For these technologies, every measured pixel represents only one transcript, which is insufficient to infer the cell type it belongs to. How to merge these single pixels to form a cell or a sub-cellular structure will be of great significance. In the current studies, two primary strategies are utilized for the analysis of high-resolution ST data: segmentation-based or segmentation-free approaches.

(1) Cell segmentation-based analysis

Cell segmentation, initially proposed in the processing of microscopic IHC images, provides more information about cell number and cell morphology. Cell segmentation here is to determine the cell boundaries based on the sparse measurement of transcripts, namely to assign transcripts to cells. Conventional cell segmentation relies on features extracted from the staining images, including intensity and textures, some of which could represent the cell boundaries. But for the fluorescent images of RNA, revealing cell boundaries requires specific staining for cell membranes, which hampers the segmentation of cells. Most groups choose to perform additional nucleus staining (e.g., DAPI) to identify putative nuclei, which are then used to guide the transcript assignment (Eng et al., 2019; Wang et al., 2018). Considering that the gene expression in the nucleus region may not equate to the expression within the whole cell, some of the groups combine, auxiliary poly(A) staining to inform the soma of cells (Moffitt et al., 2018; Wang et al., 2018). Several computational methods have been developed to provide alternative solutions.

For instance, Qian et al. (2020) developed pciSeq, which utilized a probabilistic framework to assign RNA spots to their

https://doi.org/10.1007/s11427-023-2561-0

original cells. To be specific, pciSeq treats the nuclei segmentation from the DAPI image as the initial approximation of cells, and models the cellular RNA counts and the gene-cell distance by a negative binomial distribution and a Poisson process, respectively. With paired scRNA-seq as a reference, the method estimates the probability of a transcript belonging to a cell and a cell type simultaneously using variational Bayes inference. JSTA is another method relying on initial nucleus segmentation from DAPI staining and matched scRNA-seq reference (Littman et al., 2021). JSTA can also achieve joint cell segmentation and cell type annotation through iterative pixel assignment with a deep neural network (DNN) as a classifier.

Cell segmentation can be implemented in a scRNA-independent way. For example, Baysor can perform cell segmentation based on the expression of transcripts alone (Petukhov et al., 2022), while also supporting the integration with prior information from cell-type-specific expression profiles obtained by scRNA-seq, as well as segmentation from co-stained images to improve segmentation. Notably, Baysor uses a Markov random field (MRF) to restrict the relationship between the spatially proximal molecules. With each cell modeled with a Gaussian distribution, the entire dataset can be regarded as a mixture of cell-specific distributions, which could be separable by Bayesian mixture models (BMMs). Similarly, Sparcle utilizes the Dirichlet process mixture model for initial cell-type identification and iteratively assigns each transcript to cells by borrowing information from neighboring pixels (Prabhakaran, 2022). Another method, ClusterMap, also leverages expression from the neighborhood to compute a neighborhood gene composition (He et al., 2021b), and then formulates the cell segmentation as a point pattern analysis problem, solvable by the density peak clustering (DPC) algorithm.

Following cell segmentation, cell-level analyses, such as differential expression analysis and cell-cell interaction could be performed as in scRNA-seq. What's more, further exploration of subcellular structures within cells becomes possible. For instance, based on cell segmentation, ClusterMap can further segment cells into subcellular structures including the nucleus and cytoplasm, using K-means clustering (Petukhov et al., 2022). Bento, a toolkit for subcellular analysis of ST data, further enables the identification of 5-class subcellular localization of RNA transcripts (Mah et al., 2022), including nuclear, cytoplasmic, nuclear edge, cell edge, and none of the above.

(2) Segmentation-free

The cell segmentation methods discussed above facilitate single-cell analysis on imaging-based ST data. Nevertheless, challenges arise from technical noise such as uneven intensity signals, and biological variation, including various cell sizes and shapes and different cell densities. These factors can pose difficulties in achieving accurate cell segmentation, potentially resulting in deviation in downstream analyses. Therefore, several segmentation-free methods have been developed to enable robust analysis without performing explicit segmentation. Most of the methods aim to assign each molecule pixel to specific cell types rather than individual cells.

To enable cell-type assignment for pixels, the authors of Baysor also provide a segmentation-free approach (Petukhov et al., 2022). It is based on the assumption that the neighboring RNA molecules are likely to stem from the same cell, collectively reflecting the transcriptomics profile of the corresponding cell type. They compute a neighborhood composition vector (NCV) for each transcript, effectively enhancing one pixel's signal by leveraging information from its neighbors. The NCVs are subsequently treated as "pseudo-cells" for the downstream clustering and annotation analyses. SSAM provides a similar solution (Park et al., 2021), which estimates the mRNA signal for each pixel by borrowing its neighborhood's information. Differently, they apply a kernel density estimation (KDE) with a Gaussian kernel, differing from Baysor, which gives equal weights to considered nearest neighbors.

Deciphering spatial distribution of cell types by integrating scRNA-seq

No matter whether tissue samples are profiled by single-cell or spatial transcriptomics, cell-type annotation is always of great necessity to decipher cell compositions. The annotation strategy designed for scRNA-seq, involving unsupervised clustering and cell-type inference based on expressed marker genes, seems applicable to the analysis of ST data. Unfortunately, the attempt does not usually work owing to the limitations of current ST technologies. First of all, for imaging-based targeted ST techniques, the restricted selection of genes and the presence of readout noise can hinder the identification of unknown cell types. Second, for the low-resolution barcoding-based ST data, the measurement of the mixture of multiple cells or cell types in each spot may be averaged, potentially obscuring cell heterogeneity. Finally, for the high-resolution barcoding-based ST data, the low detection efficiency also challenges both the manifest clustering and proper cell-type annotation. As a result, in most cases, integrating ST data with matched scRNA-seq becomes necessary to understand the cell-type distribution. Generally, the integration can be accomplished by two approaches: mapping or deconvolution.

(1) Cell mapping

Cell mapping includes two aspects: mapping pre-defined cell types to spatial locations and mapping cells from scRNA-seq data to the tissue. The former transfers cell type labels from scRNA-seq to spatial transcriptomics, while the latter predicts the spatial locations for cells from scRNA-seq, which is also taken as the spatial reconstruction of scRNA-seq in some cases.

For cell-type mapping, one could calculate the enrichment score using cell-type-specific gene signatures derived from scRNA-seq. This method has proven effective in the analysis of microarray-based ST data of pancreatic ductal adenocarcinomas. As for the imaging-based ST methods with limited genes, cell segmentation methods mentioned above, such as pciSeq (Qian et al., 2020), JSTA (Littman et al., 2021), and Baysor (Petukhov et al., 2022) could also allow cell-type assignment when scRNA-seq is available. Alternatively, since these imaging-based ST techniques could provide single-cell level expression after cell segmentation, existing methods designed for single-cell data integration can be directly applied to integrate single-cell resolution spatial data and scRNA-seq (Korsunsky et al., 2019; Peng et al., 2021b; Stuart et al., 2019; Welch et al., 2019). For example, Seurat projects cells from ST and scRNA-seq to the shared latent space by canonical correlation analysis (CCA) (Stuart et al., 2019). With the cell pairs identified by mutual nearest neighbors (MNN) as anchors, the cell-type labels from scRNA-seq could be transferred to spatial cells. Similar integration can also be achieved by LIGER and Harmony. By leveraging the common latent space and neighborhood information, these single-cell integration methods can also predict the spatial expression of genes missed by ST, and strengthen the original weak signals of ST-profiled genes.

Spatial reconstruction of scRNA-seq, which predicts the spatial locations of cells from expression with a few spatial landmark genes, was initially proposed before the ST technologies boom. Earlier methods, such as Seurat (v1.0), models ISH reference data with dozens of genes as a binarized expression map, and then probabilistically infers a single cell's location by relating bimodal mixture models derived from scRNA-seq to the binarized expression reference (Satija et al., 2015). Achim et al. (2015) and DistMap (Karaiskos et al., 2017) also utilize binarized ISH references but adopt different methods to calculate cell-location correspondence. Achim et al. (2015) designed a scoring scheme to assess the correspondence between a cell and each spatial location based on the gene specificity ratio in the given cell. DistMap calculates Matthew correlation coefficient (MCC) score using binarized single-cell gene expression and spatial reference, and then softly assigns cells to spatial locations. A recently developed method, Tangram is capable of aligning scRNA-seq to spatial transcriptomics measured by various technologies besides ISH-based data (Biancalani et al., 2021). By maximizing the correlation of gene expression shared by scRNA-seq and ST, Tangram could achieve a probability mapping matrix, which denotes the probability of finding every single cell in each spatial location.

Instead of scoring the cell-location correspondence, recent methods convert the problem of spatial reconstruction of scRNA-seq to a supervised learning problem or an optimization problem. For example, DEEPsc formulates the problem of mapping cells to spatial locations as a supervised classification problem by training a neural network-based classifier with spatial reference treated as scRNA-seq (Maseda et al., 2021). The sufficiently trained DEEPsc network takes the feature vector from a cell as input and predicts the cell's spatial origin according to the likelihood from different spatial locations. Another method, glmSMA frames cell mapping as a convex optimization problem (Gu and Liu, 2021). First, it employs Laplacian matrices to represent the location-to-location physical distance and cell-to-cell expression distance. By minimizing the differences between each cell's and corresponding locations' expression, glmSMA could finally find a mapping from cells in scRNA-seq to spatial locations in ST. SpaOTsc formulates cell mapping as an optimal transport problem, which aims to minimize the transport cost from cells to locations (Cang and Nie, 2020). The transport cost in SpaOTsc is measured majorly based on the gene expression dissimilarity across scRNA-seq and spatial reference and combines two penalty terms to handle the unbalanced sample sizes of two datasets and to preserve the structure within each dataset, respectively. Similarly, novoSpaRc adopts the framework of optimal transport, the core of which is the hypothesis that physically proximal cells share similar expression profiles (Nitzan et al., 2019). novoSpaRc measures the transport cost by the combination of location-to-location physical distance and cell-to-cell expression distance, both computed as the shortest path in their respective kNN graphs. By minimizing the transport costs, novoSpaRc finally obtains a certain mapping by which cells are mapped to locations with the original cell-cell correspondence preserved as much as possible, accounting for the above hypothesis. Notably, novoSpaRc also allows *de novo* reconstruction of scRNA-seq when reference ST data is not available. Most of the reconstruction methods are based on the assumption that the physical proximity could be reflected in expression similarity. However, the assumption cannot represent all the spatial distribution patterns of cells, which makes the inferred cell locations questionable.

(2) Cell type deconvolution

Deconvolution, which aims to estimate the exact cell-type proportions for each spatial location (i.e., spot or bead) is usually used in the integration of scRNA-seq and low-resolution barcoding-based ST data, such as 10x Visium. For the high-resolution barcoding-based ST techniques, such as Stereo-seq, the original pixel-level expression is aggregated in a bin-based manner, and then each bin is treated as a new spatial unit for deconvolution analysis. Current ST deconvolution methods can be basically divided into four categories: regression, factorization, probabilistic modeling, and graph-based.

Regression is one of the most popular methods developed for bulk RNA-seq deconvolution. Due to the limited number of cells covered in each spot, the direct application of bulk RNA-seq deconvolution methods on ST data will lead to noise from unrelated cell types. To overcome this problem, spatialDWLS, an ST deconvolution method based on the dampened weight least square (DWLS) regression, adopts two measures (Dong and Yuan, 2021). Firstly, the cell-type enrichment analysis is performed before the accurate estimation of cell-type proportions to identify possible cell types for each spot. Secondly, after the first round of deconvolution on the enriched cell types, cell types predicted to have low proportions are removed to perform another round of deconvolution.

Methods based on regression highly rely on the selection of marker genes for each cell type. Instead of performing regression on a cell-type-specific expression profile, some methods propose to perform regression on a latent topic profile, which can be decomposed from single-cell expression data by matrix factorization. For example, NMFreg, which was initially developed for cell-type annotation of Slide-seq, combines non-negative matrix factorization (NMF) and non-negative least square (NNLS) (Rodriques et al., 2019). It derives a basis gene-by-factor profile from pre-labeled scRNA-seq using NMF, and then computes the factor loadings for each bead using NNLS regression. With each factor linked to a cell type, the factor loadings serve as cell type proportions. SPOTlight adopts a similar strategy, but uses a seeded NMF, in which the combination of cell-type-specific marker genes and highly variable genes (HVG) is used, and the factor-by-cell profile is initialized with the cell-cell-type belongingness derived from scRNA-seq (Elosua-Bayes et al., 2021).

Deconvolution can also be achieved with factorization alone. For instance, STRIDE employs LDA, a topic modeling method, to derive cell-type-associated topic profiles from scRNA-seq (Sun et al., 2022). Then, the cell-type compositions of each spot can be inferred using the pre-trained topic model. Stdeconvolve is also based on LDA but provides a reference-free solution (Miller et al., 2022). CARD builds upon NMF, but takes spatial correlation between spots into consideration by a conditional autoregressive (CAR) model, which makes CARD a "spatial" deconvolution method (Ma and Zhou, 2022).

In addition to the intuitive regression or factorization-based methods, probabilistic modeling approaches have been developed, assuming that the gene expression in a cell or a spot follows a specific probabilistic model. For example, RCTD models the gene expression in each location by a Poisson distribution and fits each spot as a linear combination of individual cell types (Cable et al., 2022). Notably, RCTD also takes platform-specific effects into account. Cell2location follows a similar concept, but uses NB distribution to model gene expression instead (Kleshchevnikov et al., 2022). Likewise, Stereoscope utilizes the NB model, but it works on the complete set of genes rather than a set of selected marker genes (Andersson et al., 2020). DestVI also uses the NB distribution to model each gene's expression in a cell or a spot, with parameters encoded and decoded by neural networks (Lopez et al., 2022). Most importantly, DestVI not only estimates cell-type proportions, but also recovers cell-type-specific expression in each spot, which captures the continuous expression variation within cells of the same type.

Apart from DestVI, there are several other methods based on neural networks. DSTG first generates pseudo-ST data by randomly mixing cells from scRNA-seq and then constructs a link graph across spots from pseudo-ST and real-ST (Song and Su, 2021). With the link graph, which captures the intrinsic topological similarity between spots, semi-supervised graph convolutional network (GCN) is used to estimate cell-type proportions within each spot in real-ST. CellDART also generates pseudo-ST data—a virtual mixture of cells—but adopts the idea of adversarial domain adaptation (Bae et al., 2022). CellDART integrates two neural-network-based classifiers, in which the source classifier is trained to predict cell-type compositions, and the domain classifier is trained to discriminate real spots and pseudo spots. By iteratively updating two classifiers during training, the well-trained CellDART model could accurately estimate the cell-type proportions of each spot from real ST data. GraphST, another neural-network-based method, adopts a different strategy (Long et al., 2023). GraphST utilizes a graph contrastive self-supervised framework to reconstruct the gene expression for ST data by integrating spatial location information and local context. Using an autoencoder, GraphST can learn the latent representation of scRNA-seq separately. Based on the learned features, a cell-to-spot mapping probability matrix is trained through a contrastive learning mechanism, which can be combined with cell-type annotation of scRNA-seq to provide estimates of cell-type compositions for spots.

Spatial domain identification

Cell type annotation for ST data could depict the spatial distribution of cell types in the tissue. More than discrete distribution, we are also interested in how the cell types are spatially organized to form the tissue architecture and execute functions. Intuitively, physically proximal cells, no matter from the same or different cell types, could constitute a spatial structure, which is usually termed spatial domain. Identification of spatial domains will help us understand the communication between cells within the domain and their biological functions (Jiang et al., 2024). In a sense, a spatial domain can be regarded as a cluster of cells with specific spatial patterns. The standard Louvain clustering method for scRNA-seq is based on the graph built upon gene expression similarity, which does not consider spatial information and is not directly applicable here. Some spatial clustering methods modify graph-based clustering algorithms to incorporate the spatial information. For example, stLearn utilizes Louvain or K-means for global clustering and performs local clustering to find spatially separated sub-clusters or merge spatially proximal singleton spots by considering physical distances (Pham et al., 2020). Another method, MULTILAYER applies Louvain clustering on the gene-pattern co-expression graph (Moehlin et al., 2021). At first, MULTILAYER detects expression patterns for overexpressed genes by an iterative agglomerative strategy. A gene expression pattern here is defined as a region with the gene overexpressed in multiple contiguous locations. Then MULTILAYER constructs a graph where nodes represent previously detected gene patterns, and edges represent the similarity between gene patterns (i.e., gene co-expression degree). Finally, the Louvain algorithm is implemented to partition the gene co-expression patterns into multiple tissue communities.

Instead of incorporating spatial information in indirect ways, many spatial clustering methods encode the information of spatial proximity in an MRF, in which the spatial dependence is formulated by the Potts model. Zhu et al. (2018b) developed smfishHmrf, which applied hidden Markov random fields (HMRF) to the identification of spatial domains from seqFISH data. They first construct a neighborhood graph to represent the spatial relationship between cells, in which the Markov property keeps only relationships between immediately neighboring nodes. Then they model each cell's domain state by a joint probability distribution, which considers both the cell's gene expression and the domain states of neighboring cells. By solving parameters for the equilibrium of the field using expectation-maximization (EM), smfishHmrf enables the detection of spatial domains with spatially coherent gene expression. BayesSpace adopts a fully Bayesian statistical model with an MRF to ensure spots from the same cluster are closer to each other physically (Zhao et al., 2021a). By using Markov chain Monte Carlo (MCMC) and a fixed precision matrix across different clusters, BayesSpace is able to stably estimate model parameters, identify spatial clusters, and even enhance the resolution of spatial transcriptomics. Given that MCMC is computationally intensive and the fixed smoothness parameters may limit the performance in different ST datasets, Yang et al. (2022) proposed SC-MEB to enable both efficient computation and adjustable smoothness parameters. In particular, they applied an efficient iterative-conditional-mode-based expectation-maximization (ICM-EM) scheme to estimate parameters, and selected the cluster number by the modified Bayesian information criterion (MBIC). The above MRF-based methods all assume the hidden cell states to be discrete, limiting our understanding of spatial dependency among cells. In contrast, SPICEMIX integrates NMF into HMRF, in which observed gene expression is modeled as linear mixtures of latent factors, and the mixing weights of latent factors are regarded as the hidden cell states (Chidester et al., 2021). SPICEMIX, to understand it in another way, provides a method of dimension reduction for ST data by considering spatial information, which could be the foundation of downstream clustering. Based on the inferred cell states, hierarchical clustering is further applied by SPICEMIX to define categorical cell types.

Cell type clustering and spatial domain identification can be treated as two separate tasks in ST data analysis. Most of the methods we discussed above focus on identifying spatial domains, except SPICEMIX, where spatial clustering is intended to infer cell types without integration with scRNA-seq. Another method, FICT aims to infer cell types in FISH-based spatial transcriptomics by spatial clustering (Teng et al., 2022). Specifically, FICT models the expression of a cell by a cell-type-specific Gaussian distribution and models the relationship between the cell and its neighboring cells by a multinomial distribution. FICT is capable of assigning cell clusters by maximizing the joint probabilistic likelihood. Similarly, BASS also models the gene expression in a cell by a cell-type-specific normal distribution, but meanwhile, it models the cell type belongingness by a domain-specific categorical distribution (Li and Zhou, 2022). With such a hierarchical probabilistic framework, BASS enables simultaneous cell type clustering and spatial domain detection.

Spatial transcriptomics can be naturally regarded as a spot-spot graph, which is suitable to be fed to graph-based neural networks. Many GNN-based methods have been developed to learn the low-dimensional latent representations from spatial transcriptomics by integrating gene expression and spatial information (Cang et al., 2021; Hu et al., 2021a), which can facilitate the downstream analyses such as spatial domain identification and detection of spatially variable genes. For example, SpaGCN applies a GCN to integrate multiple sources of information, including gene expression, spatial locations, and histology (Hu et al., 2021a). Firstly, a graph is built to represent the relationship among spots, where nodes represent spots, and the distances of the edges are calculated by converting histology image features to the third "z" coordinate and combining it with the spots' original spatial coordinates (x, y). Then a convolutional layer is applied to aggregate gene expression from neighboring spots in the graph. Based on the aggregated gene expression, an unsupervised iterative clustering algorithm is then implemented to identify clusters (i.e., spatial domains).

Other methods introduce additional mechanisms into the basal GCN. As we discuss in the section on cell-type deconvolution, GraphST applies a graph contrastive self-supervised framework to learn the spatial latent representations for ST data by combining gene expression with spatial location information and local context information (Long et al., 2023). Another method, SpaceFlow integrates a deep graph infomax (DGI) framework into the GCN encoder (Ren et al., 2022). In addition to a spatial expression graph (SEG) built from spatial transcriptomics, SpaceFlow also constructs an expression permuted graph (EPG) by randomly permuting expression. The two graphs are both fed to a graph convolutional encoder to get the low-dimensional embeddings, and the DGI enables the encoder to distinguish embeddings of SEG from those of EPG through a discriminator loss. Some methods take autoencoders for spatial embedding. For instance, SEDR employs a deep autoencoder network to learn a low-dimensional latent representation for gene expression, which is later integrated with spatial information using a variational graph autoencoder (VGAE) (Fu et al., 2021a). STAGATE introduces an attention mechanism to the autoencoders, enabling adaptive learning of the edge weights (i.e., spot similarity) (Dong and Zhang, 2022). stMVC constructs a more comprehensive learning framework (Zuo et al., 2022). To be specific, stMVC first learns visual features from histology images through data augmentations and contrastive learning. Then semi-supervised graph attention autoencoders (SGATE) are used to learn view-specific representations based on the extracted visual features and spatial gene expression independently and integrate two graphs via an attention mechanism. The attention-based multi-view graph collaborative learning model proposed by stMVC finally learns a more robust representation of ST data.

Due to ST data's essence of spatial signals, some methods translate the problem of spatial domain identification to the classic image segmentation problem. RESEPT uses GNN to learn a three-dimensional embedding from a spot-spot graph, with gene expression treated as the nodes' attributes and physical adjacency revealed by edge connectivity (Chang et al., 2022). The three-dimensional embedding of each spot is transformed to an RGB scale so that a previous CNN designed for semantic segmentation can be directly applied to segment spatial domains. Another method, Vesalius adopts a similar RGB embedding strategy, but through dimension reduction by UMAP rather than neural networks (Martin et al., 2022).

Detection of spatially variable genes and gene expression patterns

HVG play a critical role in dimension reduction and subsequent cell clustering in the analysis of scRNA-seq. In spatial transcriptomics, the identification of spatially variable genes (SVG) is also important to characterize the functional organization in complex tissues. To identify SVG is to find genes showing great variation in space. HVG detection in scRNA-seq only considering the high variance but ignoring spatial information, cannot be directly applicable in the SVG identification. Various computational methods have been proposed to detect SVG from spatial transcriptomics. Some of the methods identify SVG based on segmented spatial domains. For example, SpaGCN first identifies spatial domains by integrating multiple sources of information as we discussed above and then defines the neighboring domain for each identified domain (Hu et al., 2021a). The spatially variable genes are determined by identifying differentially expressed genes between each target domain and the corresponding neighbor domain using the Wilcoxon rank-sum test. Instead of relying on spatial domain identification, most methods directly incorporate spatial information into the models to study the spatial variance of gene expression. According to the core models, methods could be generally divided into three categories: methods based on statistical modeling, graphs, and other principles.

(1) Based on statistical modeling

Trendsceek models the spatial expression as marked point processes, where the spatial locations are considered as a two-dimensional point process, and the locations' expressions are treated as marks (Edsgärd et al., 2018). For a given gene and a specified distance, the dependency between the spatial distribution of points and their marks is evaluated for all point pairs at the distance. The dependency assessment could be achieved by four summary statistics. Stoyan's mark-correlation, mean-mark, variance-mark, and mark-variogram. The summary statistics will remain constant when the marks and the distribution of marks are independent, but if they are dependent, the statistics will vary across different distances. Significance is estimated by permuting the expression values, and the smallest p-value among different distances is regarded as the significance of the gene. scGCO also utilizes the marked point process to model spatial gene expression but integrates HMRF into the model (Zhang et al., 2022b). For each gene, scGCO segments the graph representation by a graph cuts algorithm. The segments can be used as the candidate regions to test the expression's dependence on the spatial locations under the complete spatial randomness framework, where the distribution of points in 2D space is modeled as a homogeneous Poisson process.

In addition to the marked point process, many methods utilize the Gaussian process (GP) to model spatial gene expression. GP is a collection of random variables indexed by time or space, in which any finite collection of these random variables has a multivariate normal distribution. GP is widely used in geostatistics and has been applied in modeling spatial transcriptomics. For example, SpatialDE, based on Gaussian process regression, models each gene's variability with two components: spatial and non-spatial variance terms. The ratio of these terms can be calculated to quantify the spatial variability (Svensson et al., 2018). Statistical significance could be estimated with a log-likelihood test by comparing the likelihood of the full model with the null model without spatial covariance. SpatialDE could further identify genes with different types of spatial variation, including linear or periodic patterns, by comparing the full model fitted with a linear or periodic (i.e., cosine) covariance function with that of the Gaussian kernel. To meet the assumption of Gaussian distribution, SpatialDE employs a two-step normalization. Specifically, SpatialDE uses a variance-stabilizing transformation method, known as Anscombe's transformation, to transform the NB-distributed raw counts followed by regression of log total counts. Gpcounts also builds on Gaussian process regression, but adapts it by fitting the spatial counts by NB or zero-inflated negative binomial (ZINB) distribution rather than Gaussian distribution (BinTayyash et al., 2021). Similarly, BOOST-GP models gene read counts through a ZINB distribution but adopts a Bayesian framework to infer the parameters (Li et al., 2021b). Another method, SPARK, employs the generalized linear spatial model (GLSM) with GP modeling the spatial relationships between spatial locations and Poisson distribution modeling the expression count data (Sun et al., 2020a). Moreover, SPARK provides a more powerful statistical method to control type I errors, which computes p-values for each parameterized kernel separately and combines them with the Cauchy combination rule.

With the development of ST techniques, previous methods need to be modified to adapt to large-scale spatial transcriptomics data of high sparsity. Based on SPARK, SpatialDE2 improves the computational efficiency by replacing the Cauchy combination with the omnibus test and introducing GPU acceleration of Tensorflow (Kats et al., 2021). In order to reduce the computational complexity and physical RAM requirement, the authors of SPARK proposed a scalable non-parametric test method, SPARK-X (Zhu et al., 2021b). To be specific, SPARK-X builds on a non-parametric covariance test framework, in which two covariance matrices are calculated to measure the expression similarity and spatial proximity, respectively. Then identifying genes with specific spatial trends is converted to testing the dependence between gene expression and spatial locations. Another method, SOMDE incorporates the self-organizing map (SOM) neural network into the Gaussian process regression framework of SpatialDE (Hao et al., 2021a). SOMDE condenses the original spatial locations into SOM nodes with the spatial expression pattern and the topological structure preserved. The original spatial expression is then aggregated to form the node-level gene meta-expression, which significantly reduces the size of the covariance matrix, and thus increases the computational efficiency.

(2) Based on graph representation

As discussed in the section on spatial domain identification, the spatial expression can be represented by a graph. Some graph-based methods have been demonstrated to be successful in SVG identification. The graph Laplacian score, commonly used for graph-based feature selection, can be applied to identify spatially variable genes from graphs. GLISS, for instance, first builds a mutual nearest neighbor graph and computes a Laplacian score for each gene to measure its locality-preserving power (i.e., its association with local structures) (Zhu and Sabatti, 2020). A low Laplacian score, within a fixed graph, indicates that similarity of gene expression occurs in close locations, whereas large variation occurs in more distant locations (He et al., 2005). The statistical significance of each gene is estimated by permuting expression with the graph fixed. RayleighSelection proposed combinatorial Laplacian scores with the graph-based representation extended to the simplicial complex representation of spatial expression data (Govek et al., 2019). Apart from vertices and edges included in graphs, simplicial complexes also contain higher-dimensional elements such as triangles and tetrahedrons, which could capture more complex relations of the data. Accordingly, the combinatorial Laplacian score facilitates the identification of genes with more complex spatial structures.

Some methods introduce spatial gridding into the ordinary graph representation to simplify or optimize the spatial structure. singleCellHaystack, a spatial-gridding-based approach, was initially developed to predict differentially expressed genes from low-dimensional spaces learned from scRNA-seq, independent of cell clustering (Vandenbon and Diez, 2020). It can also be applied to the SVG identification of spatial transcriptomics data using the natural 2D or three-dimensional space. singleCellHaystack first divides the multi-dimensional space into grids and defines grid points, which are used to estimate the reference distribution of cells in the space. Then for each gene, singleCellHaystack classifies all cells into detected and undetected groups according to the binarized expression and estimates the cell distribution separately. Kullback-Leibler divergence is subsequently calculated to measure the gene's divergence by comparing it with the reference cell distribution, and the significance is evaluated by permutation test. MERINGUE is another method based on spatial gridding (Miller et al., 2021). It starts by constructing the neighborhood adjacency relationships using Voronoi tessellation, which is also used for the construction of graph representation in scGCO. Compared with the k-nearest neighbor or k-mutual-nearest neighbor, Voronoi tessellation adapts to varying neighborhood sizes and distances, offering better stability in tissues with diverse cell types and non-uniform densities. Then MERINGUE computes Moran's I for each gene to measure the spatial auto-correlation, which indicates the expression correlation among spatially adjacent locations. Giotto also provides a spatial gridding-based method, BinSpect (Dries et al., 2021b). Similarly, BinSpect relies on Voronoi tessellation to determine the neighborhood relationships. Instead of Moran's I, BinSpect adopts the statistical enrichment analysis. For each gene, BinSpect binarizes the expression using k-means clustering with k=2 or simple thresholding on rank. Next, a contingency table is calculated to reflect the expression dependency between neighboring locations. A Fisher exact test is then employed to obtain an odds ratio and the corresponding P-value. If a gene is found to be significant, it tends to be highly expressed in the neighboring locations.

(3) Based on other principles

In addition to methods rooted in statistical models or graph representation, there are approaches using entirely different principles. Sepal proposed a unique strategy founded on the diffusion theory, which regards the observed gene expression profile as the outcome of transcript diffusion. Within the framework of simulation, sepal assumes that it will take more time for transcripts to form a structured pattern than to reach a homogeneous random state. Hence, inferring the structured degree of gene expression patterns is converted to measuring the diffusion time in the simulation system. Another method, SPADE focuses on identifying important genes associated with morphological features (Bae et al., 2021). SPADE extracts latent image features from histological images by utilizing a CNN. PCA is then performed on the high-dimensional features to summarize the spatial distribution patterns of image features. SPADE uses a linear model to discover genes correlated with the image patterns (i.e., PCs), which have been demonstrated to exhibit specific spatial trends.

To model the spatial variation of gene expression, the methods discussed above only consider the relative distance between locations, ignoring the variation along specific directions. SPATA offers an option for users to manually define a trajectory axis according to prior knowledge (Kueckelhaus et al., 2020). For each gene, multiple functions are fitted to model the spatial variation patterns along the predefined spatial axis, including linear, logarithmic, or gradient ascending/descending, one-, or multiple-peak functions. Among all the functions, the best-fitting one is selected to represent the gene's dynamics by comparing the summed residuals.

After spatially variable genes are identified, some methods further determine archetypal gene patterns through clustering. By an extended Gaussian mixture model with a spatial prior on cluster centroids, SpatialDE conducts clustering to group SVGs with similar spatial expression patterns (Svensson et al., 2018). Similarly, SPARK implements a hierarchical clustering algorithm to classify detected variable genes into different categories (Sun et al., 2020a). Instead of constructing similarity matrices based on expression, MERINGUE derives a cross-correlation matrix by computing a spatial cross-correlation index, which is a modification of Moran's I auto-correlation for each pair of genes. This forms the basis of hierarchical clustering. GLISS fits a spline model on the latent structure, where each gene can be represented by the fitted spline coefficients and genes with similar gene patterns will share similar coefficients. Compared with expression-based similarity, computing gene-gene similarity based on spline coefficients could reduce correlation unrelated to spatial variation. Then GLISS performs spectral clustering on the coefficients to cluster genes into groups.

Pseudo-time trajectory analysis

From scRNA-seq or ST data, we capture only a snapshot of the cellular gene expression. The above spatial domain detection or SVG identification enables us to study the transcriptional dynamics by space in a discrete or continuous way, respectively. Previous efforts in pseudotime analysis of scRNA-seq have provided us with opportunities to reconstruct cell state trajectories from expression data alone. The additional spatial information brought by ST expands the original pseudotime analysis by introducing the dimension of space.

Direct application of single-cell pseudotime methods on ST data may cause cell trajectory to be continuous with time but discontinuous in space. To address the problem, stLearn adapts the original pseudotime algorithm by incorporating spatial information (Pham et al., 2020). stLearn first utilizes the diffusion pseudotime (DPT) algorithm to predict pseudotime from gene expression. Then it computes a pseudo-space-time distance (PSTD) matrix by combining differences in expression-based pseudotime and spatial distance, with a weight to balance between them. Based on the PSTD matrix, stLearn constructs a directed graph and applies a minimum spanning tree algorithm to determine branches (i.e., to infer cell trajectories).

Instead of relying on the initial pseudotime trajectories inferred only from gene expression, several methods emerged to predict the cell trajectories from combined expression and spatial information. SpaceFlow, which has been discussed in the section on spatial domain identification, provides a deep learning framework to learn low-dimensional embeddings from ST data (Ren et al., 2022). The embeddings produced by SpaceFlow could be used to calculate the pseudo-Spatiotemporal Map (pSM) using the DPT algorithm, facilitating the integrative reconstruction of spatiotemporal trajectories from ST data. Consequently, the spatiotemporal order generated by SpaceFlow maintains consistency in both space and pseudotime.

Cell-cell communication and gene-gene interaction

Through the aforementioned analyses, we could get a basic understanding of the spatial distribution of cell types and the expression variations in space. However, the organization of cells or cell types, as well as the regulation of genes to generate such spatial patterns, remain elusive. Many studies have reported that cellular behavior can be shaped by cell signaling pathways from the environment. Spatial transcriptomics offers a unique opportunity to investigate cell-cell communications within the preserved microenvironment. Several methods have been proposed to explore spatial dependence between cells from ST data, among which the most intuitive is to study the proximity or the co-localization of different cell types. Giotto, for instance, adopts a random permutation strategy to identify the enriched cell-type pairs (Dries et al., 2021b). With the structure of the neighborhood network fixed, cell-type labels are shuffled among the nodes to form random neighboring relationships. In this way, the ratio of observed-over-expected frequencies between two cell types is determined, and the corresponding enrichment significance can be estimated. spicyR, originally devised for spatial analysis of *in situ* cytometry, defines a score to measure the degree of cell-type co-localization (Canete et al., 2022). With the spatial distribution of cells modeled by the marked point process, spicyR applies a K-function or variance stabilized K-function (i.e., L-function) to quantify the co-localization between two cell types within a specific distance. Recently, Cang et al. (2023) developed COMMOT, based on a collective optimal transport method, to handle complex molecular interactions and spatial constraints for inferring paracrine-dependent cell-cell communication in spatially resolved transcriptomics.

Beyond observed co-localization of cell types, the spatial dependence among cells can be more complicated, which needs to be modeled by more complex methods. NCEM reconciles variance attribution and cell-cell communications in a node-centric expression model (Fischer et al., 2021). NCEM first uses the graph structure to enforce a neighborhood constraint on the cell communications. With the provided cell-type labels, NCEM applies a function to fit a cell's observed gene expression by its cell type and its spatial context. To accommodate the complexity of the spatial dependencies in different scenarios, NCEM provides three models, including the linear, nonlinear, and generative latent variable models, which are implemented by linear regression, nonlinear encoder-decoder GNN, and conditional variational autoencoder, respectively. By modeling the dependencies of the molecular states of the target cell (i.e., receiver) on the neighborhood (i.e., sender), NCEM can also determine the directionality of the sender-receiver signaling.

Instead of modeling the entire expression profile's dependence on cell-cell communications, several methods quantify the effect of cell-cell interactions on expression for each gene individually. For instance, SVCA models the expression of a target gene across cells with the Gaussian process model and decomposes the gene's variability into three components, including intrinsic effects, environmental effects from unmeasured spatial variables, and cell-cell interaction effects from neighboring cells (Arnol et al., 2019). In this manner, the fraction of variance explained by each term can be estimated for each gene, and the biologically relevant genes participating in cell-cell interactions can be identified. MISTy designs a multi-view framework to account for the expression of individual genes, where cell-cell interactions from different spatial contexts are modeled in different views (Tanevski et al., 2022). Similar to SVCA, MISTy includes intraview, juxtaview, and paraview, which correspond to intrinsic effects from gene expression of other genes in the same location, effects from immediate neighbors, and effects from the tissue architecture (i.e., cells within a radius of the specified cell), respectively. By analyzing each predictor gene's importance to the target gene in each view, the effects from different spatial contexts can be explainable for the gene pair of interest.

SVCA and MISTy can model the gene-gene relations, and discover genes associated with cell-cell interactions, but neither of them can identify explicit gene-gene interaction pairs. Yuan and Bar-Joseph (2020) developed GCNG, a GCN-based supervised computational framework, to predict gene-gene interactions. GCNG takes the graph representation of spatial neighborhood as input, as well as the normalized expression of candidate gene pairs. The output will be the classification of the interacting or non-interacting gene pairs. To enable supervised learning, known ligand-receptor interactions from a curated list are labeled as positive pairs, and randomly selected ligand-receptor pairs are labeled as negative data. With a five-layer GCN structure, GCNG could predict new gene-gene interactions in the studied ST dataset. However, GCNG cannot inform the cell types where interactions occur, and cannot focus on interaction inference within specific local regions of interest either. To address these limitations, some methods rely on the co-expression of ligands and receptors by taking cell-type locations into consideration (Dries et al., 2021b; Garcia-Alonso et al., 2021; Pham et al., 2020). For example, MERINGUE further constrains the spatial cross-correlation calculation between pairs of genes to the curated ligand-receptor pairs and two cell types of interest (Miller et al., 2021). Garcia-Alonso et al. (2021) upgraded their CellphoneDB to v3.0, which identifies ligand-receptor pairs within specific microenvironments where cell types of interest are co-localized. Similarly, based on the cell-type proximity analysis in the previous step, Giotto defines a ligand-receptor interaction score by calculating the weighted average expression of ligands and receptors in the cell subsets of interacting cell types.

Integrative analysis of spatial data

With increased throughput and decreased costs, some studies generate ST slides from multiple individuals to perform large-scale analysis. Some other studies produce a series of ST slides from multiple adjacent layers of the tissue, enabling a global view of the whole tissue. Conducting separate analyses on individual ST slides may diminish the power of multiple samples. Thus, integration methods are required to perform a joint analysis of multiple samples. Moreover, with additional information such as morphologies provided, spatial transcriptomics should be integrated with other modalities to comprehensively characterize the tissue. In this section, we will review computational methods for the integration of multiple samples and the integration of spatial data from various modalities.

(1) Multi-sample integration

The core of multi-sample integration involves placing multiple samples in the same space, referred to as common coordinate framework (CCF). The coordinate system encompasses two facets. On one hand, CCF can represent the natural three-dimensional space, in which multiple plane slides are aligned and stacked to provide a stereoscopic view of tissues. On the other hand, high-dimensional spatial measurements of location from multiple samples could be projected into a shared low-dimensional space for integrative analyses such as joint spatial domain identification.

Some methods have been developed to align multiple sequential slides from the same tissue. PASTE formulates the multi-slide alignment as an optimal transport problem, which computes the probabilistic alignment based on both gene expression and spatial information (Zeira et al., 2022). By minimizing the transport cost, PASTE could achieve a mapping that maximizes gene expression similarity between aligned locations across slides while preserving spatial structure within a slide. PASTE can align multiple sequential slides from the same tissue, but cannot be applied to the integration of slides from different time points. Andersson et al. (2021a) proposed a method, eggplant, which is a landmark-based method to project multiple slides into the common reference. First, eggplant projects the measured spatial locations to the reference by making the distance between landmarks conserved before and after transformation. Next, eggplant applies the Gaussian Process Regression to learn the relationship between the gene expression and the distance to the landmarks for all landmark-excluded locations, allowing prediction of gene expression for each location in the reference. With the strategy of location transformation combined with expression prediction, multiple slides at different time points or from different individuals could be transferred to the same reference for integrative analysis. However, eggplant necessitates not only the selection of landmark locations but also the definition of reference, which is usually a canonical structure representing the tissue domain. Both requirements limit eggplant's application on more complicated tissues such as tumors. To address this issue, Jones et al. (2022a) developed GPSA, which is also based on the Gaussian process model. GPSA constructs a two-layer Gaussian process framework, where the first layer maps the measured spatial locations to a common coordinate system, and the second layer describes the spatial gene expression within this system. Compared with eggplant, GPSA could iteratively estimate the common coordinate system *de novo*, but it also offers an option for template-based alignment with a pre-defined common coordinate system by fixing one slide.

Instead of mapping spatial locations from multiple slides to the CCF in the natural 3D space, several methods focus on projecting multiple samples to a shared low-dimensional space. In this case, integration methods should be capable of removing unwanted variations from different batches and preserving the meaningful biological variations as in scRNA-seq. But different from single-cell integration methods, ST integration methods should take into account spatial information. Liu et al. (2022) proposed PRECAST, a unified and principled probabilistic model, to jointly estimate low-dimensional embeddings and perform spatial clustering across multiple tissue slides. PRECAST performs dimension reduction on the normalized gene expression using the intrinsic conditional autoregressive (CAR) model, which could preserve the original spatial dependence among neighbors in the low-dimensional space. The resulting latent low-dimensional embedding could be further employed to perform spatial clustering using an HMRF model. As we mentioned above, BASS enables multi-scale analysis for simultaneous cell type clustering and spatial domain detection. It also allows the multi-sample integration analysis by jointly modeling the Harmony-corrected spatial transcriptomics with a hierarchical Bayesian framework. Another method, MAPLE proposed a hybrid framework for joint spatial clustering of multiple sections, following the spatially aware low-dimensional embedding learning via a GCN-based model (Allen et al., 2022).

(2) Multi-modal integration

As we discussed above, single-cell and spatial transcriptomics are usually integrated to decipher the spatial distribution of cell types through cell mapping or cell-type deconvolution. Among the integration methods we reviewed, Tangram stands out by enabling the mapping of data from other modalities onto the spatial transcriptomics through integration with multi-modal single-cell data. For example, once the single cells from SHARE-seq are mapped to spatial locations by gene expression similarity, the spatial patterns of chromatin accessibility can be unveiled.

Considering that many ST technologies provide corresponding histological images, many computational methods leverage the additional image information to enhance the analytic performance at each step. For example, stLearn leverages morphological similarity to normalize the expression data, thereby reducing the impact of technical noise of dropouts (Pham et al., 2020). spaGCN takes the histology image features into consideration when calculating spot-spot distances to construct a graph for spatial transcriptomics (Hu et al., 2021a). stMVC employs graph networks with the attention mechanism to integrate multi-source information including histological features, and ultimately learns the low-dimensional embedding of ST data (Zuo et al., 2022). Likewise, methods such as conST (Zong et al., 2022) and MUSE (Bao et al., 2022) also use deep learning architectures to integrate cell morphologies and transcriptional states for joint representation. Instead of the complex deep learning-based mechanism, SPADE directly associates the spatial variance of gene expression with the spatial distribution patterns of image features using a linear regression model (Bae et al., 2021).

In addition to facilitating the analysis of spatial transcriptomics, the histological images could also be used to predict spatial gene expression. Many methods have been developed to address such a problem. To overcome the limitation of low resolution in some barcoding-based ST technologies, Bergenstråhle et al. (2022) proposed a deep generative model to infer the super-resolved expression maps from high-resolution histology images, both within and between the originally measured locations. Rather than focusing on improving the resolution of spatial gene expression, some methods generalize the spatial transcriptome prediction to the histopathology images without matched expression data. For example, He et al. (2020a) introduced a deep learning algorithm, ST-Net, to capture gene expression heterogeneity by combining spatial transcriptomics and histology images. With the model trained with a BRCA spatial transcriptomics dataset comprising 68 ST slides of breast tissue sections, it can predict the spatially resolved transcriptome of other breast cancer datasets directly from histology images. However, ST-Net does not account for spatial dependencies between spots. HisToGene employs a modified Vision Transformer model to enable the prediction of spatial gene expression with the spot dependency considered (Pang et al., 2021). Building upon HisToGene, Hist2ST additionally includes a Convmixer module to capture the internal relations of 2D vision features within image patches (Zeng et al., 2022b).

Applications

The recent and rapid progress in spatial transcriptomics has promoted its widespread application across various biological systems. ST techniques have been instrumental in spatially characterizing the cell states of healthy tissues, and some of them aim to decipher the spatial architecture of tissues at specific developmental stages. Notably, among the tissues, the nervous system has been a focal point of investigation. Numerous studies have made substantial contributions to constructing detailed spatial atlases for the brain. Moreover, ST techniques have proven invaluable in exploring the microenvironments of injured or diseased tissue, including mouse lungs infected with virus (Boyd et al., 2020), human hearts with myocardial infarction (Kuppe et al., 2022), as well as a range of different tumor types (Ji et al., 2020a; Qi et al., 2022; Wu et al., 2021a; Wu et al., 2021b; Wu et al., 2021d). Here we review the applications of ST in three main fields, encompassing the development and homeostasis of healthy tissues, neuroscience, and the tumor microenvironment.

(1) Development and homeostasis of healthy tissue

Most of the studies utilize mouse models to investigate the development of early mammalian embryos. Spatial atlases have been established for several stages of mouse embryonic development. Peng et al. (2019) focused on lineage differentiation and morphogenesis at the post-implantation stages. Geo-seq was applied to profile cell populations at pre-selected positions in all germ layers from pre-gastrulation (embryonic day I5.5) to late gastrulation (E7.5). The study unveiled the dynamic molecular regulation of lineage specification and tissue patterning in time and space. Moreover, they also uncovered the pivotal role of Hippo/Yap signaling during germ-layer development. To further explore the cell fate decisions in the early organogenesis at the end of gastrulation, Lohoff et al. (2022) performed seqFISH on multiple sagittal sections collected from mouse embryos at E8.5–E8.75. Due to the limited number of target genes, they integrated seqFISH with existing single-cell transcriptome atlases to enable genome-wide imputation. By utilizing the generated spatial single-cell map, the authors revealed spatial patterns of gene expression corresponding to dorsal-ventral and rostral-caudal axes in the midbrain and hindbrain region and discovered early dorsal-ventral separation in the gut tube. Recently, Chen et al. (2022) applied high-resolution Stereo-seq to whole mouse embryos at the mid and late-gestation stage spanning from E9.5 to E16.5 and eventually constructed a mouse organogenesis spatiotemporal transcriptomic atlas (MOSTA).

Moving beyond early embryonic development in mice, many researchers have taken advantage of spatial transcriptomics to explore the spatially dependent mechanisms driving the development of organs or tissues in humans. For example, Crosse et al. (2020) utilized the LCM-based RNA sequencing to enable spatially resolved profiling of the developing hematopoietic stem cell (HSC) niche in human embryos at Carnegie stage (CS)16–CS17 (i.e., 39–41 post-conception days). They analyzed the dorsoventral polarized signaling in the aorta and identified ventrally secreted endothelin as an important secreted regulator of early human HSC development. In the study of the developing human heart, Asp et al. (2019) characterized different anatomical regions of human hearts at three developmental stages (4.5–5, 6.5 and 9 post-conception weeks) using the ST technology. With the integration of scRNA-seq and ISS, a comprehensive spatial map was created, providing detailed information about cell subtype localization during human cardiogenesis. Similar strategies were applied to the developmental study of the human intestine ranging from 8 to 22 PCW (Fawkner-Corbett et al., 2021). In addition to generating a spatiotemporal atlas of human intestinal development, they also revealed how morphogen gradients direct cellular differentiation. Spatial transcriptomics has also been applied to the study of cell-type atlas and homeostasis maintenance in healthy tissues of adults, which could serve as a reference to be compared with diseased tissues. Shen et al. (2024) applied the Stereo-seq technology to draw an ST atlas of the human gingiva. By identifying periodontitis-relevant effector cells, genes and pathways, the ST results may aid in the development of new therapeutic strategies for periodontitis. By combining scRNA-seq, snRNA-seq, and 10x Visium ST, Madissoon et al. (2021) created a spatial multi-omics atlas of the human lung and airway, which comprises various novel and known cell types. The spatial lung atlas also revealed specific tissue microenvironments, such as the gland-associated lymphoid niche (GALN), which may play a role in preventing respiratory infections. In another study of the human uterus, Garcia-Alonso et al. (2021) also applied multi-omics technologies to construct a comprehensive cellular map of human endometrium, characterizing the spatiotemporal dynamics across the menstrual cycle. In particular, further spatial interaction analyses revealed the role of NOTCH and WNT signaling pathways in shaping the differentiation of ciliated and secretory cell lineages. With the accumulation of ST data, it is foreseeable that in the near future, integration of multi-source tissue maps will lead to the establishment of a comprehensive reference spatial atlas of the entire human body.

(2) Neuroscience

The explicit layered structures and distinct anatomical regions make the brain an appropriate material to validate newly developed spatial transcriptomics technologies. In return, these ST techniques significantly enhance our understanding of the spatial architecture of brains. Many efforts have been devoted to building reference maps of the brain. Due to the limited size of fields of view and intensive-labor nature of early imaging-based ST techniques, most of the studies focused on specific subregions in the mouse brain. For example, Codeluppi et al. (2018) developed osmFISH and employed the methodology to define the spatial cellular organization of the somatosensory cortex, covering only 33 targeted marker genes and around 5,000 cells. During the same time, Moffitt et al. (2018) generated a spatial molecular map of neurons in the hypothalamic preoptic region by coupling MERFISH with scRNA-seq. Similarly, other subregions of the brain, such as the visual cortex (Wang et al., 2018), the primary motor cortex (Zhang et al., 2021a), the hippocampus (Alon et al., 2021; Shah et al., 2016), and the cerebellum (Kebschull et al., 2020), have been profiled by different imaging-based ST techniques to establish detailed spatial cellular organization maps.

Thanks to the development of high-throughput barcoding-based ST techniques, a molecular atlas of the whole adult mouse brain was established by Ortiz et al. (2020). They utilized the ST technology to profile spatial gene expression of 75 adjacent coronal sections collected from one brain hemisphere along the anteroposterior axis. Through alignment with the Allen mouse brain atlas (ABA), they constructed a complete brain atlas, offering 3D tissue coordinates and detailed ABA neuroanatomical definitions. More importantly, they also defined new area- and layer-specific subregions in the molecular atlas by unsupervised classification. Whether the entire brain or specific subregions are profiled, these atlases together will be of great value to experimental neuroscience, ultimately extending our knowledge about the structure-functional relationships of the brain.

In addition to revealing the spatial organization of cell types in normal brains, spatial transcriptomics can be extended to the study of neurodegenerative or psychiatric diseases, uncovering spatially relevant mechanisms of dysfunction or dysregulation in the nervous system. For example, Chen et al. (2020b) combined the ST technology with ISS to capture the transcriptional changes in the vicinity of amyloid plaques in Alzheimer's disease (AD). In particular, they identified two gene co-expression networks that might be responsive to amyloid plaque deposition in AD. In a study of amyotrophic lateral sclerosis (ALS), Maniatis et al. (2019) employed the ST technology to characterize the spatiotemporal dynamics over the progress of the disease by utilizing murine models of ALS at different stages. Combining with postmortem tissues from ALS patients, they discerned shared spatial patterns of perturbations in transcriptional pathways associated with ALS pathology.

As ST technologies continue to improve in resolution and detection efficiency, we anticipate the establishment of more detailed and comprehensive atlases of the nervous system. These resources will undoubtedly be invaluable for exploring the structure-function relationships of circuits and behaviors.

(3) Tumor microenvironment

Although single-cell transcriptomics has shed light on the cell-type compositions and their functions in the intricate TME, it remains unexplored how these cells are spatially organized to control or promote tumor progression. Spatial transcriptomics makes it possible to study different cell populations and signaling pathways with the spatial context preserved. Generally, tumor microenvironments might include tumor cells, stromal cells, and immune cells. Initial research efforts often concentrate on the interior heterogeneities of tumor regions. In a single-cell and spatial atlas study of human breast cancer (BRCA), Wu et al. (2021d) derived seven gene modules from scRNA-seq to describe the intratumor transcriptional heterogeneity. The enrichment analysis revealed two gene modules mutually exclusive in the tumor regions, which were related to the EMT and proliferation states, respectively. In another study of primary liver cancer, five cancer stem cell (CSC) populations were defined, which showed different distribution patterns in different regions, including the leading edge, tumor, and high-grade portal vein tumor thrombosis (Wu et al., 2021b). Of note, the fraction of PROM1$^+$ CSCs was higher in portal vein tumor thrombosis than in the tumor region, potentially exerting a crucial role in the tumor progression.

Centered on the tumor region, the relative spatial distribution of immune or stromal cell types could be revealed by spatial transcriptomics. In the study of human squamous cell carcinoma (SCC), Ji et al. (2020a) discovered that B cells were infiltrated in the tumor, while regulatory T cells, macrophages, and fibroblasts were abundant at the tumor-stromal border. Conversely, CD8 T cells were notably excluded from the tumor. Similarly, different cell subtypes or states also reveal different spatial patterns. Wu et al. (2021d) identified both inflammatory-like cancer-associated fibroblasts (iCAFs) and myofibroblast-like CAFs (mCAFs) in the TME of breast cancer, yet the two subtypes exhibited markedly distinct spatial distributions. mCAFs were found to be enriched in invasive cancer regions, while iCAFs were dispersedly distributed across invasive cancer, stroma, and lymphocytes-aggregate regions. Some studies are interested in the molecular and cell-type patterns in the tumor-stromal border, namely the invasive fonts of tumors (Hunter et al., 2021). Wu et al. (2021a) characterized the dynamics of cell-type abundance across the invasive fonts and found an immune suppressive microenvironment in the area near the borderline.

The spatial analysis could also recognize some patterned structures and characterize them in the tumor microenvironment. In the abovementioned study of liver cancer (Wu et al., 2021b), unsupervised clustering of ST spots revealed a cluster characterized by high expression of tertiary lymphoid structures (TLS)-related genes, such as *CXCL13*, *CCL19*, *CCL21*, *LTF*, and *LTB*. The pathological examination validated the presence of TLSs. Then, Wu et al. (2021b) defined a TLS-50 signature to locate TLSs in other tissue sections, which was also found to be associated with more favorable prognosis in HCC patients in TCGA. Similarly, Andersson et al. (2021b) also identified TLSs in HER2-positive breast cancer. To further investigate how TLSs influence the response to immunotherapy in cancer, Meylan et al. (2022) used spatial transcriptomics to examine the nature of B cell responses within TLS in renal cell carcinoma (RCC). They discovered that TLSs could generate and propagate anti-tumor antibody-producing plasma cells, which is associated with response to immunotherapy.

Cellular communications are known to play important roles in the immune surveillance or escape of tumors, as well as tumor progression. With the spatial distribution of cell types revealed by cell-type deconvolution or cell mapping analysis, cell-type proximity or colocalization patterns could also be recognized. Moncada et al. (2020) identified the colocalization of inflammatory fibroblasts and stress-response cancer cells by mapping scRNA-seq-defined cell types to ST of pancreatic ductal adenocarcinomas. Similarly, with the integration of scRNA-seq and ST in SCC, a fibrovascular niche was found to surround a tumor-specific keratinocyte population (Ji et al., 2020a). Further interaction analysis revealed the colocalization might be mediated by multiple ligand-receptor pairs. In another study of colorectal cancer, spatial transcriptomics and immunofluorescent staining demonstrated the co-existence of FAP$^+$ fibroblasts and SPP1$^+$ macrophages, which was associated with poor patient survival (Qi et al., 2022).

With the development of spatial multi-omics techniques, additional facets such as cell crosstalk and metabolic states, will be characterized to gain more insights into the complexity of the tumor microenvironment. Understanding the tumor microenvironment will facilitate the identification of therapeutic targets and the design of anti-tumor drugs.

Summary

This chapter provides a comprehensive overview of current advancements in spatial transcriptomics, encompassing technical innovations, computational methods, and diverse applications. Spatial transcriptomics has revolutionized our understanding of tissue organization and cellular heterogeneity, enabling high-resolution visualization of gene expression patterns within intact tissues. The development of computational methods has facilitated the integration and interpretation of spatial transcriptomics data, unveiling spatial regulatory mechanisms and novel molecular interactions. Spatial transcriptomics has been successfully applied in various fields, including developmental biology, neuroscience, cancer research, and immunology, with the potential to accelerate biomarker discovery and personalized medicine approaches. Spatial transcriptomics represents a transformative approach and will continue to be refined to reshape our understanding of complex biological systems. We anticipate it will offer profound insights into tissue homeostasis and disease mechanisms.

Chapter 8 Single-cell CRISPR screening technology

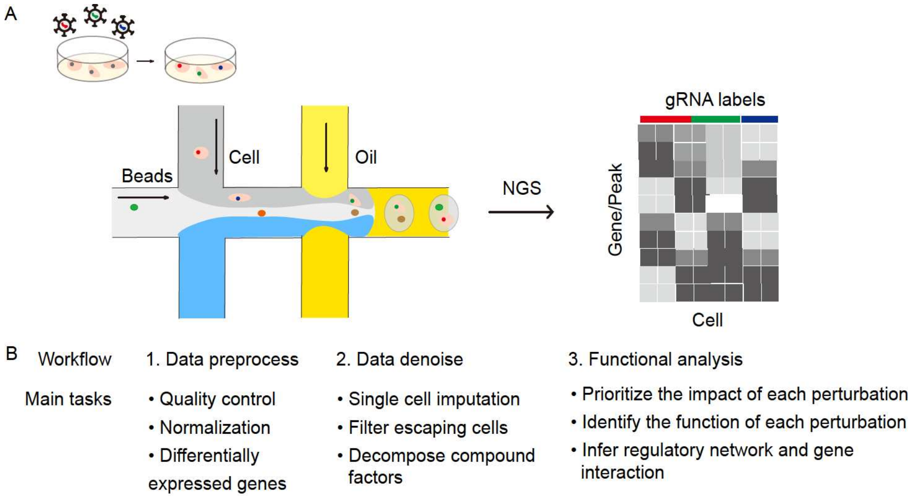
The clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 system is a revolutionary approach to edit the mammalian genome (Cong et al., 2013; Mali et al., 2013). With the development of lentiviral delivery methods, CRISPR screening technology emerged and has enabled genome-wide knockout in a cost-effective manner (Koike-Yusa et al., 2014; Shalem et al., 2014; Wang et al., 2014a). However, CRISPR screening can only analyze genes with very distinct phenotypes, such as those that significantly affect cell growth or those that can be detected directly with antibodies or fluorescent proteins.

In 2016, a new technique, called single-cell CRISPR screening (scCRISPR-seq), was developed that coupled CRISPR perturbations and single-cell sequencing to enable pooled genetic screens at large-scale single-cell resolution (Adamson et al., 2016; Dixit et al., 2016; Jaitin et al., 2016). The key technical innovation of scCRISPR-seq is the creative design of the lentiviral vector, called the Perturb-seq vector, to allow the identification of sgRNA in each cell from sequencing (Figure 16A). scCRISPR-seq can facilitate high-throughput functional dissection of complex regulatory mechanisms and heterogeneous cell populations.

In this chapter, we will comprehensively review scCRISPR-seq in four distinct parts. Firstly, we will introduce representative technologies within each category of scCRISPR-seq. Secondly, we will delve into the primary tools that have been specifically developed for the analysis of scCRISPR-seq data. Thirdly, we will explore notable applications of scCRISPR-seq. Finally, we will draw conclusions and engage in a discussion of the limitations and future trends associated with scCRISPR-seq.

The category of scCRISPR-seq platforms

Currently, numerous alternative scCRISPR-seq platforms have



Workflow	1. Data preprocess	2. Data denoise	3. Functional analysis
Main tasks	• Quality control • Normalization • Differentially expressed genes	• Single cell imputation • Filter escaping cells • Decompose compound factors	• Prioritize the impact of each perturbation • Identify the function of each perturbation • Infer regulatory network and gene interaction

Figure 16. Overview of scCRISPR-seq. A, General schematic of scCRISPR-seq platform. NGS, next generation sequence. B, Bioinformatic analysis of scCRISPR-seq data.

emerged (Table 5). Based on the integrated omics approach of scCRISPR-seq, these platforms can be classified into three primary categories: transcriptome-based scCRISPR-seq, epigenome-based scCRISPR-seq, and multimodal scCRISPR-seq.

Transcriptome-based scCRISPR-seq

The main scCRISPR-seq platforms are transcriptome-based applications that combine CRISPR screens with single-cell RNA-seq. For transcriptome-based scCRISPR-seq, the Perturb-seq vector is generally composed of single-guide RNA (sgRNA), cell barcode (CBC), gene barcode (GBC), and UMI, such as Perturb-seq (Adamson et al., 2016; Dixit et al., 2016) and CRISP-seq (Jaitin et al., 2016). In the Perturb-seq vector, sgRNA is used to direct Cas9 nucleases to induce double-strand breaks at targeted genomic regions and CBC is used to tag each cell, while GBC is used to tag each sgRNA and UMI is used to tag each transcript. Perutrb-seq and CRISP-seq are the first scCRISPR-seq platforms to be developed. These approaches involved complex construction of the Perturb-seq vector, including complex cloning strategy and sometimes the decoupling of the gRNA spacer and its barcode, which limits their versatility. The CROP-seq (Datlinger et al., 2017) optimizes the design of the Perturb-seq vector to allow the detection of sgRNA induced in each cell coupled with mRNA by adding Poly-A tail to the Perturb-seq vector, which greatly reduces the complexity and cost of scCRISPR-seq. However, Hill et al. (2018) demonstrated that the lentivirus swap rate of existing studies was only about 50% because of the Perturb-seq vector designs of these studies. Thus, they optimized CROP-seq vector designs by serving the guide RNA as the barcode to improve the swap rate to 94%. Due to constraints on Perturb-seq vector design, each lentiviral vector of Perturb-seq and CRISP-seq can only deliver a single encoded sgRNA to cells, and CROP-seq enabled the delivery of paired sgRNAs to cells. That is, they are all incompatible with the delivery of multiple sgRNAs. To solve this problem, Replogle et al. (2020) designed direct-capture Perturb-seq, in which expressed sgRNAs are sequenced alongside single-cell transcriptomes and enable the delivery of multiple sgRNAs. Direct-capture Perturb-seq is particularly valuable for the mechanistic dissection of genetic interaction. It further reduces the cost of Perturb-seq experiments. Direct-seq (Song et al., 2020) has similar functions as direct-capture Perturb-seq that enables CRISPR perturbation and its transcriptional readouts profiled together and supports the delivery of multiple sgRNAs. In 2022, the genome-scale Perturb-seq method was introduced by Replogle et al. (2022), enabling unbiased and comprehensive profiling of genome-scale genetic perturbations affecting 9,867 genes. This breakthrough facilitated systematic gene function assignment and the exploration of complex cellular phenotypes. More recently, Li et al. (2023) developed the CRISPR-human organoids-single-cell RNA sequencing (CHOOSE) system. This innovative system enables genetic disruption and single-cell transcriptomics for pooled loss-of-function screening in mosaic organoids.

However, all the aforementioned scCRISPR-seq platforms are limited to *in vitro* applications. In contrast, *in vivo* assays are more attractive due to their closer resemblance to real organic conditions. Therefore, Jin et al. (2020b) developed *in vivo* Perturb-seq, a variation of the Perturb-seq protocol that involves pooled perturbations conducted *in vivo*. Furthermore, PoKI-seq (Roth et al., 2020) has demonstrated the feasibility of *in vivo* investigation of the immunological response of reprogrammed T cells to solid tumors. Recently, Santinha et al. (2023) developed adeno associated virus (AAV)-mediated direct *in vivo* single-cell CRISPR screening, termed AAV-Perturb-seq, a tunable and broadly applicable method for transcriptional linkage analysis and phenotyping of genetic perturbations *in vivo*.

Epigenome-based scCRISPR-seq

In addition to transcriptome applications, there are also epigenetic-based scCRISPR-seq platforms. In 2019, Rubin et al. (2019) developed Perturb-ATAC, a method that combines CRISPR interference or knockout with chromatin accessibility profiling in single cells based on simultaneous detection of CRISPR guide RNAs and open chromatin sites by assay of transposase-accessible chromatin with sequencing (ATAC-seq). They applied this method to determine the roles of a diverse set of trans-regulatory factors, including TFs, chromatin modifiers, and human and viral ncRNAs, which may be useful for dissecting loci where both cis-regulatory elements and ncRNA transcripts have been shown to have effects on gene expression (Cho et al., 2018b; Engreitz et al., 2016; Rubin et al., 2019). Perturb-ATAC expands scCRISPR-seq research into the epigenome field, making

Table 5. The different scCRISPR-seq platforms

Platforms	Omics	In vivi/vitro	Subject	sgRNA sequenced directly without barcode	No. of delivery of sgRNAs	Coupled CRISPR system	References
Perturb-Seq	Transcriptome	*In vitro*	K562	no	Single	CRISPR knockout	(Dixit et al., 2016)
Perturb-seq	Transcriptome	*In vitro*	K562	no	Single	CRISPR interference	(Adamson et al., 2016)
CRISP-seq	Transcriptome	*In vitro*	Mouse BMDCs	no	Single	CRISPR knockout	(Jaitin et al., 2016)
CROP-seq	Transcriptome	*In vitro*	Jurkat	yes	Paired	CRISPR knockout	(Datlinger et al., 2017)
Mosaic-seq	Multimodal (Transcriptome, enhancer)	*In vitro*	K562	no	Single	CRISPR interference	(Xie et al., 2017)
Improved CROP-seq	Transcriptome	*In vitro*	MCF10A	yes	Paired	CRISPR knockout and interference	(Hill et al., 2018)
Perturb-ATAC	Epigenome (Chromatin accessibility)	*In vitro*	Primary human keratinocytes, B lymphoblasts	no	Single	CRISPR knockout and interference	(Rubin et al., 2019)
ECCITE-seq	Multimodal (Transcriptome, proteome, clonotypes)	*In vitro*	Sez4, MyLa, PBMC, NIH-3T3	yes	Single	CRISPR knockout	(Mimitou et al., 2019)
Direct-capture Perturb-seq	Transcriptome	*In vitro*	iPSCs, K562	yes	Multiple	CRISPR knockout, interference and activation	(Replogle et al., 2020)
Direct-seq	Transcriptome	*In vitro*	Jurkat, K562	yes	Multiple	CRISPR knockout and activation	(Song et al., 2020)
In vivo Perturb-seq	Transcriptome	*In vivo*	Progenitor cells of the mouse forebrain	no	Single	CRISPR interference	(Jin et al., 2020b)
PoKI-seq	Transcriptome	*In vivo*	Human primary T cells, NSG mice bearing human melanoma cells	no	Single	CRISPR knock-in	(Roth et al., 2020)
Spear-ATAC	Epigenome (Chromatin accessibility)	*In vitro*	K562, GM12878, MCF7	yes	Single	CRISPR knockout and interference	(Pierce et al., 2021)
CRISPR-sciATAC	Epigenome (Chromatin accessibility)	*In vitro*	NIH-3T3, K562	yes	Single	CRISPR knockout	(Liscovitch-Brauer et al., 2021)
genome-scale Perturb-seq	Transcriptome	*In vitro*	K562, RPE1	yes	Single	CRISPR interference	(Replogle et al., 2022)
Perturb-map	Multimodal (Spatial transcriptome, imaging)	*In vitro, in vivo*	293T, KP, 4T1	no	Single	CRISPR knockout	(Dhainaut et al., 2022)
AAV-Perturb-seq	Transcriptome	*In vivo*	Adult mouse brain prefrontal cortex	yes	Single	CRISPR interference	(Santinha et al., 2023)
CHOOSE screen	Transcriptome	*In vitro* (organoid)	Brain organoid	no	Paired	CRISPR knockout	(Li et al., 2023)

scCRISPR-seq more powerful and broader applications. However, Perturb-ATAC is constrained by high costs and low throughput. In response to this limitation, Spear-ATAC (Pierce et al., 2021) was developed to achieve significantly higher cell throughput and a substantial cost reduction, offering a more practical alternative. Additionally, CRISPR-sciATAC (Liscovitch-Brauer et al., 2021) demonstrated similar cell throughput and cost to Spear-ATAC. However, it exhibited limited sensitivity to subtle changes in chromatin accessibility.

Multimodal scCRISPR-seq

Multimodal single-cell assays provide high-resolution snapshots of heterogeneous cell populations, but the scCRISPR-seq platforms above are all limited to one modality, such as transcriptome or epigenome. Thus, to apply the technique to multi-omics simultaneously, multimodal scCRISPR-seq was developed. Xie et al. (2017) developed mosaic single-cell analysis by indexed CRISPR sequencing (Mosaic-seq) to perturb enhancers and jointly measure each cell's transcriptome and its induced sgRNA. Mosaic-seq provides a novel tool to interrogate the functions of noncoding genes in a perturbation-based manner. In addition, Mimitou et al. (2019) developed expanded CRISPR-compatible cellular indexing of transcriptomes and epitopes by sequencing (ECCITE-seq), which allowed simultaneous detection of transcriptomes, proteins, clonotypes, and CRISPR perturbations from every single cell. By constructing a 49-marker panel of ECCITE-seq antibodies to profile human peripheral blood mononuclear cells (PBMCs), they recovered many important results (Fanok et al., 2018; Stoeckius et al., 2017), demonstrating the power of ECCITE-seq to combine immunophenotype, clonotype, and transcriptome information. Spatial transcriptomics is able to characterize gene expression profiles while retaining information about the spatial tissue context, which provides new insights into different fields of biology, such as neuroscience, developmental biology, and cancer research (Moses and Pachter, 2022) (see the spatial section below). Recently, a new multimodal scCRISPR-seq called Perturb-map (Dhainaut et al., 2022) was developed to enable multimodal phenotyping of CRISPR screens *in situ* by imaging and spatial transcriptomics. Perturb-map is based on a protein bar code (Pro-Code) system that uses triplet combinations of a few linear epitopes to create a higher-order set of unique bar codes (Wroblewska et al., 2018). These unique bar codes can mark cells expressing different CRISPR gRNAs. It should be noted that Perturb-map is the only scCRISPR-seq platform that enables *in vivo* CRISPR screens combined with spatial transcriptome, which is particularly suitable for the identification of genetic determinants of tumor composition, organization, and immunity. Dhainaut et al. (2022) applied Perturb-map to the study of the TME. They knocked out 35 genes in a mouse model of lung cancer and found that knockout of Tgfbr2 can promote TME remodeling and immune exclusion.

The tools to analyze scCRISPR-seq data

scCRISPR-seq data contain rich perturbation information, which is a natural advantage to exploring the association between genotype and phenotype at a single cell level. For example, by applying Perturb-seq to the K562 cell line, Adamson et al. (2016) have shown that perturbation of PERK has a greater impact on the unfolded protein response than ATF6 and IRE1α. Datlinger et al. (2017) perturbed 23 transcription factors in the Jurkat cell line under the condition of T cell receptor (TCR) activation with CROP-seq and found that knockouts of LCK, ZAP70, and LAT have a strong negative effect on TCR activation signaling.

However, the analysis of scCRISPR-seq data is a major challenge due to its inherent noise. Thus, several bioinformatic tools have been developed to help analyze scCRISPR-seq data (Table S12 in Supporting Information). Generally, these scCRISPR-seq data analysis tools focus on three parts (Figure 16B): (i) Data preprocessing, including quality control, normalization, and differentially expressed genes detection, such as MIMOSCA (Dixit et al., 2016), MUSIC (Duan et al., 2019), and SCREE (Wei et al., 2023). (ii) Data denoise, including single-cell imputation, escaping cells filtering, and compound factors decomposing, such as MUSIC, mixscape (Papalexi et al., 2021), and SCREE (Wei et al., 2023). (iii) Functional analysis, including prioritizing the impact of each perturbation, identifying the function of each perturbation, inferring regulatory network and gene interaction, such as MUSIC, Normalisr (Wang, 2021), scMAGeCK (Yang et al., 2020), Pando (Fleck et al., 2023) and GEARS (Roohani et al., 2023). Specifically, LRICA is proposed to decode the driver signal/component of the data by low-rank matrix factorization. MIMOSCA is a computational framework for calculating the relationship between sgRNA and each gene. LRICA and MIMOSCA were developed as prototypes without executable and user-friendly implementations. Thus, Duan et al. (2019) developed MUSIC, a general computational framework to evaluate the impact of each perturbation with topic modeling (Blei and Lafferty, 2007), which was originally presented in the machine learning and natural language processing community for latent topic discovery in a particular set of documents. MUSIC links genotype to phenotype with tolerance to substantial noise and analyzes scCRISPR-seq data from three perspectives, i.e., prioritizing the gene perturbation effect as an overall perturbation effect, in a functional topic-specific manner, and quantifying correlations between different perturbations. scMAGeCK is also a framework for analyzing scCRISPR-seq data, which is extended from MAGeCK (Li et al., 2014). scMAGeCK includes two modules, scMAGeCK-RRA and scMAGECK-LR, where scMAGeCK-RRA is used to identify significantly enriched sgRNAs by the negative binomial distribution, and scMAGeCK-LR is used to assess affected genes by linear regression. scMAGeCK showed a good control of false positives and better sensitivity than other methods. In addition to the general computational framework of scCRISPR-seq, some tools focus on data denoising. For example, SCEPTRE was developed for scCRISPR-seq data calibration using conditional randomization testing. SCEPTRE demonstrated good calibration and sensitivity to scCRISPR-seq data, yielding hundreds of new regulatory relationships supported by orthogonal biological evidence. mixscape was developed to improve the signal-to-noise ratio of scCRISPR-seq data by filtering escaping cells (cells induced sgRNA, but did not exhibit perturbation effect) by mixed discriminant analysis. Normalisr is developed to reconstruct gene regulatory network for scCRISPR-seq data. Wang et al. (2022g) emphasized the significance of identifying clone cells, as they can lead to false positives in scCRISPR-seq data. SCREE serves as a comprehensive pipeline for scCRISPR-seq data analysis. In contrast to the previously mentioned approaches, which primarily concentrated on data denoising and mining in scCRISPR-seq, GEARS was specifically designed to predict transcriptional responses to both single and multigene perturbations. These methodologies have substantially enhanced the analysis of scCRISPR-seq data.

Applications of scCRISPR-seq

scCRISPR-seq is widely applied in various fields due to its powerful capabilities, including linking genotype to phenotype, dissecting genetic regulations, and investigating genetic mechanisms in specific diseases, such as tumor and autism.

Linking genotype to phenotype

Compared with traditional CRISPR screening, which can only identify genes with very distinct phenotypes, scCRISPR-seq has the ability to uncover the functions of any genes. Therefore, scCRISPR-seq is naturally suited for linking genotype to phenotype on a large scale. For example, Jaitin et al. (2016) revealed the effect of 22 TFs on the regulation of antiviral, inflammatory, or developmental processes in lipopolysaccharide (LPS) stimulated born marrow cells (BMCs) by CRISP-seq. Adamson et al. (2016) analyzed systematically the effect of 83 unfolded protein response (UPR) related genes in K562 cells by Perturb-seq. In addition, genome-scale Perturb-seq (Replogle et al., 2022) offers unbiased, comprehensive profiling of genetic perturbations (9,867 genes), facilitating systematical dissection of relationships between genes related to gene translation and ribosome biogenesis.

Dissecting genetic regulations

scCRISPR-seq is also used to dissect complex relationships between genomic elements, including coding genes, transcription factors, chromatin regulators, enhancers, and other noncoding elements. For example, Adamson et al. (2016) discovered the crosstalk between three UPR sensor genes (*ATF6*, *PERK*, and *IRE1*) using Perturb-seq. CROP-seq perturbed TFs regulating TCR activation in Jurkat cells upon LPS stimulation and uncovered the relationship between TFs. In addition, scCRISPR-seq for enhancer perturbation, such as Mosaic-seq (Xie et al., 2017), could discover novel enhancer-gene pairs. In addition, scCRISPR-seq coupled with scATAC-seq, such as Perturb-ATAC, Spear-ATAC, and CRISPR-sciATAC could reveal epigenetic landscape remodelers in human B lymphocytes and leukemia cells (Liscovitch-Brauer et al., 2021; Pierce et al., 2021; Rubin et al., 2019).

Investigating genetic mechanisms

Several *in vivo* scCRISPR-seq platforms are available, enabling studies of genetic mechanisms in specific diseases such as tumors and autism. For instance, Perturb-map (Dhainaut et al., 2022) facilitates the identification of genetic determinants related to tumor composition, organization, and immunity. Using Perturb-map, Dhainaut et al. (2022) discovered that the knockout of tgfbr2 in lung cancer cells promotes tumor microenvironment remodeling and immune exclusion. Roth et al. (2020) conducted a screen for chimeric antigen receptors that enhance T cell anti-tumor functions, improving tumor infiltration and cell killing rates under immunosuppressive conditions in melanoma using PoKI-seq. Furthermore, Jin et al. (2020b) evaluated 35 *de novo* loss-of-function risk genes associated with autism spectrum disorder/neurodevelopmental delay (ASD/ND) using *in vivo* Perturb-seq. They identified cell type-specific and evolutionarily conserved gene modules from both neuronal and glial cell classes. Li et al. (2023) also focused on these high-risk autism spectrum disorder genes, and they uncovered their effects on cell fate determination in mosaic organoids with CHOOSE system. Recently, Santinha et al. (2023) employed AAV-Perturb-seq to systematically analyze the phenotypic landscape associated with 22q11.2 deletion syndrome genes in the prefrontal cortex of adult mouse brains. They identified three 22q11.2-linked genes actively involved in both established and previously unrecognized pathways governing neuronal functions *in vivo*.

Summary

In this chapter, we have presented a comprehensive review of scCRISPR-seq, divided into three distinct parts, which include the categories of scCRISPR-seq, tools for the analysis of scCRISPR-seq data, and notable applications of scCRISPR-seq.

scCRISPR-seq has been a powerful approach for functional genomics research (Bock et al., 2022). In this section, we have categorized scCRISPR-seq into three primary categories based on its integrated omics approach: transcriptome-based scCRISPR-seq, epigenome-based scCRISPR-seq, and multimodal scCRISPR-seq. Given the inherent noise in scCRISPR-seq data, a multitude of bioinformatic tools have been developed to aid in its analysis, resulting in significant improvements.

The versatility of scCRISPR-seq has led to its widespread application across various fields, offering potent capabilities such as connecting genotype to phenotype, dissecting genetic regulations, and exploring genetic mechanisms in specific diseases like tumors and autism. Nevertheless, before its broader adoption in biological research, three key aspects require attention: (i) reducing complexity and cost: efforts should be made to further streamline and reduce the complexity and cost of scCRISPR-seq experiments. This will enhance scalability and accessibility, allowing more laboratories to leverage this technology. (ii) Expanding applicability to complex tissues and *in vivo* settings: while current scCRISPR-seq platforms primarily target cell lines, there is a pressing need to develop more robust scCRISPR-seq platforms that can be applied to more complex tissues, including organoids, and ideally *in vivo* settings. This expansion will enable a broader range of biological investigations. (iii) Noise reduction techniques: As the number of scCRISPR-seq platforms grows, it becomes crucial to develop more powerful methods for deciphering the inherent noise in scCRISPR-seq data. These methods will contribute to the reliability and interpretability of scCRISPR-seq results, further enhancing their utility in diverse research contexts.

Epilogue

scRNA-seq technology has attracted widespread attention from many scientists around the world because it has the advantage of providing an unprecedented method to study cell heterogeneity at the single-cell level. A mere 14 years have elapsed since the establishment of a new era in scRNA-seq research, which was preceded by the initial conceptual and technical breakthrough achieved by Tang et al. (2009) in 2009. The field of scRNA-seq research is currently experiencing a surge in studies, driven by the continuous development of sequencing technology and bioinformatics. The maturity of scRNA-seq technology has greatly facilitated advancements in other single-cell omics studies. At present, single-cell omics detection has been extended to the genome (Dey et al., 2015), epigenome (Muto et al., 2021), spatial transcriptomics (Chen et al., 2022), proteome (Peterson et al., 2017; Specht et al., 2021), metabolome (Shrestha, 2020) and other multiomics levels (Angermueller et al., 2016), providing a more comprehensive, refined and complete analysis strategy for single-cell level research. In this review, we summarize the state-of-the-art developments in single-cell omics technologies, data analyses, and their applications, outlining the landscape of the single-cell sequencing field across multiple layers.

In Chapter 1, we provide a comprehensive overview of the currently available scRNA-seq technologies, experimental methodologies, data analysis procedures, and their applications within the biomedical field. Initially, single-cell sequencing was performed by isolating single cells and independently constructing a sequencing library. These single-cell sequencing technologies can only detect a small number of cells (tens to hundreds), such as the Tang method, STRT-seq, and SMART-seq (Islam et al., 2012; Ramsköld et al., 2012; Tang et al., 2009). However, with the in-depth study of sequencing technology, single-cell identification based on barcode tags has emerged, and with the emergence of new single-cell separation technologies based on microdroplets or microwells such as Drop-Seq and Cyto-Seq (Fan et al., 2015a; Macosko et al., 2015), and single-cell transcriptome sequencing has entered the era of high-throughput. The sequencing cost has been dramatically reduced, while automation and throughput have been significantly increased. ScRNA-seq technology solves the problem of cell heterogeneity, opens new avenues for personalized treatment of clinical diseases, especially tumors, and promotes the development of precision medicine. However, scRNA-seq has limitations of low capture efficiency and high dropouts due to the low amount of starting material. Compared with bulk RNA-seq, scRNA-seq produces noisier and more variable data. Although, researchers have designed a variety of tools to conduct diverse scRNA-seq data analyses, the technical noise and biological variation (e.g., stochastic transcription) still pose huge challenges for computational analysis of scRNA-seq data (Chen et al., 2019a). Therefore, data analysis methods still need to be further optimized and improved.

Compared with the increasingly mature scRNA-seq technology, the other single-cell omics technologies are still budding. In Chapters 2, 3, 4, and 5, we focus on the state-of-the-art tools, computational methods, and applications for single-cell genome, epigenome, proteomic, and metabolomics sequencing over the past ten years. ScWGS has revolutionized our understanding of genetic variation and its impact on human health and disease. The rapid development of it has accelerated genomic research, enabled personalized medicine, and provided valuable insights into the genetic basis of diseases and human genome diversity. Cells exhibit extensive heterogeneity in terms of chromatin accessibility, nucleosome positioning, histone modifications, and DNA methylation. Mapping this epigenomic information in single-cell samples is very important for developmental biology, cancer research, and so on. Advances in single-cell epigenomic sequencing approaches are enabling high-resolution mapping of chromatin states in single cells. However, nowadays, single-cell epigenomic techniques suffer from data loss. As a result, even though individual cell epigenomic data sets are powerful resources for clustering analyses and for revealing cellular heterogeneity based on the collection of a great number of target sites, they have only very limited ability to provide information on single target sites (Carter and Zhao, 2021). Therefore, improving the coverage of chromatin target sites in various individual cell epigenomic assays will be required in future studies, which will contribute to understanding cell heterogeneity at a whole cell level and single specific site. Single-cell proteomics is in the early stage of explosive development due to its complex constituents, low abundance, wide dynamic range, and lack of amplified ability. Just in 2019, analysis of the proteome from single cells was described as a "dream", but today there have been several promising tools developed (Marx, 2019). We believe that with the optimization of accessibility and the further improvement of throughput, the truly large-scale applications of single-cell proteomics in scientific and clinical research, such as organ maps, drug screening, and precise disease classification, are within reach. Single-cell metabolomics is used to identify the composition of metabolites in a single cell, measure their abundance, and study their dynamic changes. Meanwhile, the metabolome represents the downstream products of the genome, transcriptome, and proteome, and provides a more immediate and dynamic snapshot of the functionality (Shrestha, 2020). Overall, single-cell omics technologies are still in the budding stage, and they are going to continue to flourish.

A single cell serves as the fundamental unit of life. Multi-omics analysis of a single cell can offer profound insights into the cell's phenotype, disease state, and environmental impacts. In Chapters 6, 7, and 8, we comprehensively summarize the integrated analysis of multi-omics, the combined application of scRNA-seq and CRISPR screening, and spatial transcriptome. In intricate biological processes, such as tumorigenesis and aging, heterogeneity occurs on different levels, including the genome, transcriptome, proteome, and epigenome. If only one component is analyzed from a single cell at a time, only the local overview of the gene regulatory network can be detected, while the complex global situation cannot be accurately predicted. In this situation, multi-omics technology highlights its unique advantages, which can provide a more complete map of the gene regulatory network in the study of complex tissues. For spatial transcriptomics, it has enabled the measurement of gene expression with spatial information preserved, which will be conducive to investigating intercellular relationships and discovering novel regulation mechanisms in the spatial context. In addition, spatial transcriptomics makes it possible to explore the spatial regulation mechanisms of cell fate determination and the architecture of tissue patterning. Compared with traditional CRISPR hybrid screening, the combination of scRNA-seq and CRISPR can not only screen thousands of gRNAs in a single experiment but also simultaneously capture perturbed full-transcriptome data for the clearest understanding of cell type specific gene function and pathway analysis. Therefore, the combination of these techniques enables a better and deeper understanding of key biological processes and mechanisms, which is an important direction for the development of single-cell technology in the future.

Nowadays, single-cell omics technologies have witnessed significant advancements in terms of both throughput and resolution. Moving forward, the primary trends in single-cell technology development are to improve the efficiency and throughput of single-cell sorting, enhance the sequencing coverage and sensitivity, realize high-throughput multi-omics studies, and develop more automated single-cell technology platforms, which will help reduce the cost and technical threshold of single-cell technology. The single-cell technology promises to be widely used in the field of scientific research and research transformation and will have a great contribution to health monitoring, disease diagnosis, and treatment.

Compliance and ethics

The authors declare that they have no conflict of interest.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (32130020, 32025009, 82030099, 30700397, 31970638, 61572361, 81973701, U23A20513, 32222026, 82373446), the National Key Research and Development Program of China (2021YFF1201200, 2021YFF1200900, 2022YFA1106000), the Shanghai Municipal Science and Technology Major Project (2017SHZDZX01), the Fundamental Research Funds for the Central Universities (20002150110, 22120230292). Beihang University & Capital Medical University Plan (BHME-201904), the Special Fund of the Pediatric Medical Coordinated Development Center of Beijing Hospitals Authority (XTCX201809), the Cooperative Research Fund of the Affiliated Wuhu Hospital of East China Normal University (40500-20104-222400), the Fundamental Research Funds for the Central Universities (226-2024-00001), the Shanghai Municipal Science and Technology Commission "Science and Technology Innovation Action Plan" technical standard project (21DZ2201700), the Shanghai Municipal Science and Technology Commission "Science and Technology Innovation Action Plan" natural science foundation project (23ZR1435800), the Shanghai Natural Science Foundation Program (17ZR1449400), the Shanghai Artificial Intelligence Technology Standard Project (19DZ2200900), the Open Research Fund of Key Laboratory of Advanced Theory and Application in Statistics and Data Science-MOE, ECNU, Key Laboratory of MEA, Ministry of Education, ECNU. We would like to express our sincerest gratitude to the members of the expert panel of the Program on Single-Cell Omics.

Supporting information

The supporting information is available online at https://doi.org/10.1007/s11427-023-2561-0. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

Ascensión, A.M., Ibáñez-Solé, O., Inza, I., Izeta, A., and Araúzo-Bravo, M.J. (2022). Triku: a feature selection method based on nearest neighbors for single-cell data. Gigascience 11, giac017.

Abdelfattah, N., Kumar, P., Wang, C., Leu, J.S., Flynn, W.F., Gao, R., Baskin, D.S., Pichumani, K., Ijare, O.B., Wood, S.L., et al. (2022). Single-cell analysis of human glioma and immune cells identifies S100A4 as an immunotherapy target. Nat Commun 13, 767.

Abid, A., Zhang, M.J., Bagaria, V.K., and Zou, J. (2018). Exploring patterns enriched in a dataset with contrastive principal component analysis. Nat Commun 9, 2134.

Abouleila, Y., Onidani, K., Ali, A., Shoji, H., Kawai, T., Lim, C.T., Kumar, V., Okaya, S., Kato, K., Hiyama, E., et al. (2019). Live single cell mass spectrometry reveals cancer-specific metabolic profiles of circulating tumor cells. Cancer Sci 110, 697–706.

Achim, K., Pettit, J.B., Saraiva, L.R., Gavriouchkina, D., Larsson, T., Arendt, D., and Marioni, J.C. (2015). High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. Nat Biotechnol 33, 503–509.

Adamson, B., Norman, T.M., Jost, M., Cho, M.Y., Nuñez, J.K., Chen, Y., Villalta, J.E., Gilbert, L.A., Horlbeck, M.A., Hein, M.Y., et al. (2016). A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. Cell 167, 1867–1882.e21.

Adossa, N., Khan, S., Rytkönen, K.T., and Elo, L.L. (2021). Computational strategies for single-cell multi-omics integration. Comput Struct Biotechnol J 19, 2588–2596.

Affinati, A.H., Sabatini, P.V., True, C., Tomlinson, A.J., Kirigiti, M., Lindsley, S.R., Li, C., Olson, D.P., Kievit, P., Myers, M.G., et al. (2021). Cross-species analysis defines the conservation of anatomically segregated VMH neuron populations. eLife 10, e69065.

Ai, S., Xiong, H., Li, C.C., Luo, Y., Shi, Q., Liu, Y., Yu, X., Li, C., and He, A. (2019). Profiling chromatin states using single-cell itChIP-seq. Nat Cell Biol 21, 1164–1172.

Aicher, T.P., Carroll, S., Raddi, G., Gierahn, T., Wadsworth, M.H., 2nd, Hughes, T.K., Love, C., and Shalek, A.K. (2019). Seq-Well: a sample-efficient, portable picowell platform for massively parallel single-cell RNA sequencing. In: Proserpio, V., ed. Single Cell Methods. Methods in Molecular Biology. New York: Humana. 111–132.

Aird, D., Ross, M.G., Chen, W.S., Danielsson, M., Fennell, T., Russ, C., Jaffe, D.B., Nusbaum, C., and Gnirke, A. (2011). Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome Biol 12, R18.

Akar-Ghibril, N. (2022). Defects of the innate immune system and related immune deficiencies. Clinic Rev Allerg Immunol 63, 36–54.

Alexovič, M., Sabo, J., and Longuespée, R. (2021). Automation of single-cell proteomic sample preparation. Proteomics 21, 2100198.

Aliverti, E., Lum, K., Johndrow, J.E., and Dunson, D.B. (2021). Removing the influence of group variables in high-dimensional predictive modelling. J R Stat Soc Ser A Stat Soc 184, 791–811.

Aliverti, E., Tilson, J.L., Filer, D.L., Babcock, B., Colaneri, A., Ocasio, J., Gershon, T.R., Wilhelmsen, K.C., and Dunson, D.B. (2020). Projected t-SNE for batch correction. Bioinformatics 36, 3522–3527.

Allen, C., Chang, Y., Ma, Q., and Chung, D. (2022). MAPLE: a hybrid framework for multi-sample spatial transcriptomics data. bioRxiv, doi: 10.1101/2022.02.28.482296.

Alon, S., Goodwin, D.R., Sinha, A., Wassie, A.T., Chen, F., Daugharthy, E.R., Bando, Y., Kajita, A., Xue, A.G., Marrett, K., et al. (2021). Expansion sequencing: spatially precise *in situ* transcriptomics in intact biological systems. Science 371, eaax2656.

Alquicira-Hernandez, J., Sathe, A., Ji, H.P., Nguyen, Q., and Powell, J.E. (2019). scPred: accurate supervised method for cell-type classification from single-cell RNA-seq data. Genome Biol 20, 264.

Alvarez, M., Rahmani, E., Jew, B., Garske, K.M., Miao, Z., Benhammou, J.N., Ye, C.J., Pisegna, J.R., Pietiläinen, K.H., Halperin, E., et al. (2020). Enhancing droplet-based single-nucleus RNA-seq resolution using the semi-supervised machine learning classifier DIEM. Sci Rep 10, 11019.

Amodio, M., van Dijk, D., Srinivasan, K., Chen, W.S., Mohsen, H., Moon, K.R., Campbell, A., Zhao, Y., Wang, X., Venkataswamy, M., et al. (2019). Exploring single-cell data with deep multitasking neural networks. Nat Methods 16, 1139–1145.

Anchang, B., Davis, K.L., Fienberg, H.G., Williamson, B.D., Bendall, S.C., Karacosta, L. G., Tibshirani, R., Nolan, G.P., and Plevritis, S.K. (2018). DRUG-NEM: optimizing drug combinations using single-cell perturbation response to account for intratumoral heterogeneity. Proc Natl Acad Sci USA 115, E4294–E4303.

Andersson, A., Andrusivová, Ž., Czarnewski, P., Li, X., Sundström, E., and Lundeberg, J. (2021a). A Landmark-based Common Coordinate Framework for Spatial Transcriptomics Data. bioRxiv, doi: 10.1101/2021.11.11.468178.

Andersson, A., Bergenstråhle, J., Asp, M., Bergenstråhle, L., Jurek, A., Fernández Navarro, J., and Lundeberg, J. (2020). Single-cell and spatial transcriptomics enables probabilistic inference of cell type topography. Commun Biol 3, 565.

Andersson, A., Larsson, L., Stenbeck, L., Salmén, F., Ehinger, A., Wu, S.Z., Al-Eryani, G., Roden, D., Swarbrick, A., Borg, Å., et al. (2021b). Spatial deconvolution of HER2-positive breast cancer delineates tumor-associated cell type interactions. Nat Commun 12, 6012.

Andor, N., Lau, B.T., Catalanotti, C., Kumar, V., Sathe, A., Belhocine, K., Wheeler, T. D., Price, A.D., Song, M., Džakula, Ž., et al. (2020). Joint single cell DNA-Seq and RNA-Seq of cancer reveals subclonal signatures of genomic instability and gene expression. bioRxiv, doi:10.1101/445932.

Andrews, T.S., and Hemberg, M. (2018). Identifying cell populations with scRNASeq. Mol Aspects Med 59, 114–122.

Angelo, M., Bendall, S.C., Finck, R., Hale, M.B., Hitzman, C., Borowsky, A.D., Levenson, R.M., Lowe, J.B., Liu, S.D., Zhao, S., et al. (2014). Multiplexed ion beam imaging of human breast tumors. Nat Med 20, 436–442.

Angermueller, C., Clark, S.J., Lee, H.J., Macaulay, I.C., Teng, M.J., Hu, T.X., Krueger, F., Smallwood, S.A., Ponting, C.P., Voet, T., et al. (2016). Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. Nat Methods 13, 229–232.

Angermueller, C., Lee, H.J., Reik, W., and Stegle, O. (2017). DeepCpG: accurate prediction of single-cell DNA methylation states using deep learning. Genome Biol 18, 67.

Aran, D., Looney, A.P., Liu, L., Wu, E., Fong, V., Hsu, A., Chak, S., Naikawadi, R.P., Wolters, P.J., Abate, A.R., et al. (2019). Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nat Immunol 20, 163–172.

Argelaguet, R., Arnol, D., Bredikhin, D., Deloro, Y., Velten, B., Marioni, J.C., and Stegle, O. (2020). MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. Genome Biol 21, 111.

Argelaguet, R., Clark, S.J., Mohammed, H., Stapel, L.C., Krueger, C., Kapourani, C.A., Imaz-Rosshandler, I., Lohoff, T., Xiang, Y., Hanna, C.W., et al. (2019). Multi-omics profiling of mouse gastrulation at single-cell resolution. Nature 576, 487–491.

Argelaguet, R., Cuomo, A.S.E., Stegle, O., and Marioni, J.C. (2021). Computational principles and challenges in single-cell data integration. Nat Biotechnol 39, 1202–1215.

Argelaguet, R., Velten, B., Arnol, D., Dietrich, S., Zenz, T., Marioni, J.C., Buettner, F., Huber, W., and Stegle, O. (2018). Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets. Mol Syst Biol 14, e8124.

Arisdakessian, C., Poirion, O., Yunits, B., Zhu, X., and Garmire, L.X. (2019). DeepImpute: an accurate, fast, and scalable deep neural network method to impute single-cell RNA-seq data. Genome Biol 20, 211.

Armingol, E., Baghdassarian, H.M., Martino, C., Perez-Lopez, A., Aamodt, C., Knight, R., and Lewis, N.E. (2022). Context-aware deconvolution of cell–cell communication with Tensor-cell2cell. Nat Commun 13, 3665.

Arnol, D., Schapiro, D., Bodenmiller, B., Saez-Rodriguez, J., and Stegle, O. (2019). Modeling cell-cell interactions from spatial molecular data with spatial variance component analysis. Cell Rep 29, 202–211.e6.

Arrastia, M.V., Jachowicz, J.W., Ollikainen, N., Curtis, M.S., Lai, C., Quinodoz, S.A., Selck, D.A., Ismagilov, R.F., and Guttman, M. (2022). Single-cell measurement of higher-order 3D genome organization with scSPRITE. Nat Biotechnol 40, 64–73.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene Ontology: tool for the unification of biology. Nat Genet 25, 25–29.

Ashuach, T., Gabitto, M.I., Jordan, M.I., and Yosef, N. (2021). MultiVI: deep generative model for the integration of multi-modal data. bioRxiv, doi: 10.1101/2021.08.20.457057.

Ashuach, T., Reidenbach, D.A., Gayoso, A., and Yosef, N. (2022). PeakVI: a deep generative model for single-cell chromatin accessibility analysis. Cell Rep Methods 2, 100182.

Asp, M., Giacomello, S., Larsson, L., Wu, C., Fürth, D., Qian, X., Wärdell, E., Custodio, J., Reimegård, J., Salmén, F., et al. (2019). A spatiotemporal organ-wide gene expression and cell atlas of the developing human heart. Cell 179, 1647–1660.e19.

Assarsson, E., Lundberg, M., Holmquist, G., Björkesten, J., Thorsen, S.B., Ekman, D., Eriksson, A., Rennel Dickens, E., Ohlsson, S., Edfeldt, G., et al. (2014). Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. PLoS ONE 9, e95192.

Azodi, C.B., Zappia, L., Oshlack, A., and McCarthy, D.J. (2021). splatPop: simulating population scale single-cell RNA sequencing data. Genome Biol 22, 341.

Bacher, R., Chu, L.F., Leng, N., Gasch, A.P., Thomson, J.A., Stewart, R.M., Newton, M., and Kendziorski, C. (2017). SCnorm: robust normalization of single-cell RNA-seq data. Nat Methods 14, 584–586.

Bae, S., Choi, H., and Lee, D.S. (2021). Discovery of molecular features underlying the morphological landscape by integrating spatial transcriptomic data with deep features of tissue images. Nucleic Acids Res 49, e55.

Bae, S., Na, K.J., Koh, J., Lee, D.S., Choi, H., and Kim, Y.T. (2022). CellDART: cell type inference by domain adaptation of single-cell and spatial transcriptomic data. Nucleic Acids Res 50, e57.

Bai, L., Liang, J., and Cao, F. (2021). Semi-supervised clustering with constraints of different types from multiple information sources. IEEE Trans Pattern Anal Mach Intell 43, 3247–3258.

Bais, A.S., and Kostka, D. (2020). scds: computational annotation of doublets in single-cell RNA sequencing data. Bioinformatics 36, 1150–1158.

Bandura, D.R., Baranov, V.I., Ornatsky, O.I., Antonov, A., Kinach, R., Lou, X., Pavlov, S., Vorobiev, S., Dick, J.E., and Tanner, S.D. (2009). Mass cytometry: technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry. Anal Chem 81, 6813–6822.

Bao, F., Deng, Y., Wan, S., Shen, S.Q., Wang, B., Dai, Q., Altschuler, S.J., and Wu, L.F. (2022). Integrative spatial analysis of cell morphologies and transcriptional states with MUSE. Nat Biotechnol 40, 1200–1209.

Bar-Joseph, Z., Gitter, A., and Simon, I. (2012). Studying and modelling dynamic biological processes using time-series gene expression data. Nat Rev Genet 13, 552–564.

Barkas, N., Petukhov, V., Nikolaeva, D., Lozinsky, Y., Demharter, S., Khodosevich, K., and Kharchenko, P.V. (2019). Joint analysis of heterogeneous single-cell RNA-seq dataset collections. Nat Methods 16, 695–698.

Bartosovic, M., Kabbe, M., and Castelo-Branco, G. (2021). Single-cell CUT&Tag profiles histone modifications and transcription factors in complex tissues. Nat Biotechnol 39, 825–835.

Barwinska, D., El-Achkar, T.M., Melo Ferreira, R., Syed, F., Cheng, Y.H., Winfree, S., Ferkowicz, M.J., Hato, T., Collins, K.S., Dunn, K.W., et al. (2021). Molecular characterization of the human kidney interstitium in health and disease. Sci Adv 7, eabd3359.

Basile, G., Kahraman, S., Dirice, E., Pan, H., Dreyfuss, J.M., and Kulkarni, R.N. (2021). Using single-nucleus RNA-sequencing to interrogate transcriptomic profiles of archived human pancreatic islets. Genome Med 13, 128.

Baslan, T., Kendall, J., Volyanskyy, K., McNamara, K., Cox, H., D'Italia, S., Ambrosio, F., Riggs, M., Rodgers, L., Leotta, A., et al. (2020). Novel insights into breast cancer copy number genetic heterogeneity revealed by single-cell genome sequencing. eLife 9, e51480.

Becht, E., McInnes, L., Healy, J., Dutertre, C.A., Kwok, I.W.H., Ng, L.G., Ginhoux, F., and Newell, E.W. (2019). Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol 37, 38–44.

Bendall, S.C., Davis, K.L., Amir, E.D., Tadmor, M.D., Simonds, E.F., Chen, T.J., Shenfeld, D.K., Nolan, G.P., and Pe'er, D. (2014). Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. Cell 157, 714–725.

Bendall, S.C., Simonds, E.F., Qiu, P., Amir, E.D., Krutzik, P.O., Finck, R., Bruggner, R. V., Melamed, R., Trejo, A., Ornatsky, O.I., et al. (2011). Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. Science 332, 687–696.

Bergen, V., Lange, M., Peidli, S., Wolf, F.A., and Theis, F.J. (2020). Generalizing RNA velocity to transient cell states through dynamical modeling. Nat Biotechnol 38, 1408–1414.

Bergen, V., Soldatov, R.A., Kharchenko, P.V., and Theis, F.J. (2021). RNA velocity—current challenges and future perspectives. Mol Syst Biol 17, e10282.

Bergenstråhle, L., He, B., Bergenstråhle, J., Abalo, X., Mirzazadeh, R., Thrane, K., Ji, A. L., Andersson, A., Larsson, L., Stakenborg, N., et al. (2022). Super-resolved spatial transcriptomics by deep data fusion. Nat Biotechnol 40, 476–479.

Bergman, H.M., and Lanekoff, I. (2017). Profiling and quantifying endogenous molecules in single cells using nano-DESI MS. Analyst 142, 3639–3647.

Bernstein, N.J., Fong, N.L., Lam, I., Roy, M.A., Hendrickson, D.G., and Kelley, D.R. (2020). Solo: doublet identification in single-cell RNA-seq via semi-supervised deep learning. Cell Syst 11, 95–101.e5.

Biancalani, T., Scalia, G., Buffoni, L., Avasthi, R., Lu, Z., Sanger, A., Tokcan, N., Vanderburg, C.R., Segerstolpe, Å., Zhang, M., et al. (2021). Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. Nat Methods 18, 1352–1362.

BinTayyash, N., Georgaka, S., John, S.T., Ahmed, S., Boukouvalas, A., Hensman, J., and Rattray, M. (2021). Non-parametric modelling of temporal and spatial counts data from RNA-seq experiments. Bioinformatics 37, 3788–3795.

Blackburn, D.M., Lazure, F., Corchado, A.H., Perkins, T.J., Najafabadi, H.S., and Soleimani, V.D. (2019). High-resolution genome-wide expression analysis of single myofibers using SMART-Seq. J Biol Chem 294, 20097–20108.

Blackburn, D.M., Lazure, F., and Soleimani, V.D. (2021). SMART approaches for genome-wide analyses of skeletal muscle stem and niche cells. Crit Rev Biochem Mol Biol 56, 284–300.

Blecher-Gonen, R., Barnett-Itzhaki, Z., Jaitin, D., Amann-Zalcenstein, D., Lara-Astiaso, D., and Amit, I. (2013). High-throughput chromatin immunoprecipitation for genome-wide mapping of *in vivo* protein-DNA interactions and epigenomic states. Nat Protoc 8, 539–554.

Blei, D.M., and Lafferty, J.D. (2007). A correlated topic model of science. Ann Appl Stat 1, 17–35.

Blondel, V.D., Guillaume, J.L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. J Stat Mech 2008, P10008.

Bloom, J.D. (2018). Estimating the frequency of multiplets in single-cell RNA sequencing from cell-mixing experiments. PeerJ 6, e5578.

Bock, C., Datlinger, P., Chardon, F., Coelho, M.A., Dong, M.B., Lawson, K.A., Lu, T., Maroc, L., Norman, T.M., Song, B., et al. (2022). High-content CRISPR screening. Nat Rev Methods Primers 2, 8.

Boileau, P., Hejazi, N.S., and Dudoit, S. (2020). Exploring high-dimensional biological data with sparse contrastive principal component analysis. Bioinformatics 36, 3422–3430.

Boisset, J.C., Vivié, J., Grün, D., Muraro, M.J., Lyubimova, A., and van Oudenaarden, A. (2018). Mapping the physical network of cellular interactions. Nat Methods 15, 547–553.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.

Bolisetty, M.T., Stitzel, M.L., and Robson, P. (2017). CellView: interactive exploration of high dimensional single cell RNA-seq data. bioRxiv, doi:10.1101/123810.

Borella, M., Martello, G., Risso, D., and Romualdi, C. (2021). PsiNorm: a scalable normalization for single-cell RNA-seq data. Bioinformatics 38, 164–172.

Boufea, K., Seth, S., and Batada, N.N. (2020). scID uses discriminant analysis to identify transcriptionally equivalent cell types across single-cell RNA-seq data with batch effect. iScience 23, 100914.

Boyd, D.F., Allen, E.K., Randolph, A.G., Guo, X.J., Weng, Y., Sanders, C.J., Bajracharya, R., Lee, N.K., Guy, C.S., Vogel, P., et al. (2020). Exuberant fibroblast activity compromises lung function via ADAMTS4. Nature 587, 466–471.

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

Bravo González-Blas, C., Minnoye, L., Papasokrati, D., Aibar, S., Hulselmans, G., Christiaens, V., Davie, K., Wouters, J., and Aerts, S. (2019). cisTopic: cis-regulatory topic modeling on single-cell ATAC-seq data. Nat Methods 16, 397–400.

Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol 34, 525–527.

Breda, J., Zavolan, M., and van Nimwegen, E. (2021). Bayesian inference of gene expression states from single-cell RNA-seq data. Nat Biotechnol 39, 1008–1016.

Brennecke, P., Anders, S., Kim, J.K., Kołodziejczyk, A.A., Zhang, X., Proserpio, V., Baying, B., Benes, V., Teichmann, S.A., Marioni, J.C., et al. (2013). Accounting for technical noise in single-cell RNA-seq experiments. Nat Methods 10, 1093–1095.

Brombacher, E., Hackenberg, M., Kreutz, C., Binder, H., and Treppner, M. (2022). The performance of deep generative models for learning joint embeddings of single-cell multi-omics data. bioRxiv, doi: 10.1101/2022.06.06.494951.

Browaeys, R., Saelens, W., and Saeys, Y. (2020). NicheNet: modeling intercellular communication by linking ligands to target genes. Nat Methods 17, 159–162.

Brown, J., Ni, Z., Mohanty, C., Bacher, R., and Kendziorski, C. (2021). Normalization by distributional resampling of high throughput single-cell RNA-sequencing data. Bioinformatics 37, 4123–4128.

Brüning, R.S., Tombor, L., Schulz, M.H., Dimmeler, S., and John, D. (2022). Comparative analysis of common alignment tools for single-cell RNA sequencing. Gigascience 11, giac001.

Brunner, A., Thielert, M., Vasilopoulou, C., Ammar, C., Coscia, F., Mund, A., Hoerning, O.B., Bache, N., Apalategui, A., Lubeck, M., et al. (2022). Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation. Mol Syst Biol 18, e10798.

Budnik, B., Levy, E., Harmange, G., and Slavov, N. (2018). SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. Genome Biol 19, 161.

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

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

Buhler, J. (2001). Efficient large-scale sequence comparison by locality-sensitive hashing. Bioinformatics 17, 419–428.

Bullard, J.H., Purdom, E., Hansen, K.D., and Dudoit, S. (2010). Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinformatics 11, 94.

Burrows, M., and Wheeler, D.J. (1994). A block-sorting lossless data compression algorithm. SRC Research Report. Palo Alto: Systems Research Center.

Cabello-Aguilar, S., Alame, M., Kon-Sun-Tack, F., Fau, C., Lacroix, M., and Colinge, J. (2020). SingleCellSignalR: inference of intercellular networks from single-cell transcriptomics. Nucleic Acids Res 48, e55.

Cable, D.M., Murray, E., Zou, L.S., Goeva, A., Macosko, E.Z., Chen, F., and Irizarry, R. A. (2022). Robust decomposition of cell type mixtures in spatial transcriptomics. Nat Biotechnol 40, 517–526.

Cahill, J.F., and Kertesz, V. (2018). Automated optically guided system for chemical analysis of single plant and algae cells using laser microdissection/liquid vortex capture/mass spectrometry. Front Plant Sci 9, 1211.

Cahill, J.F., and Kertesz, V. (2020). Laser capture microdissection-liquid vortex capture mass spectrometry metabolic profiling of single onion epidermis and microalgae cells. In: Shrestha, B., ed. Single Cell Metabolism. Methods in Molecular Biology. New York: Humana. 89–101.

Cakir, B., Prete, M., Huang, N., van Dongen, S., Pir, P., and Kiselev, V.Y. (2020). Comparison of visualization tools for single-cell RNAseq data. NAR Genomics Bioinf 2, lqaa052.

Candelli, T., Schneider, P., Garrido Castro, P., Jones, L.A., Bodewes, E., Rockx-Brouwer, D., Pieters, R., Holstege, F.C.P., Margaritis, T., and Stam, R.W. (2022). Identification and characterization of relapse-initiating cells in MLL-rearranged infant ALL by single-cell transcriptomics. Leukemia 36, 58–67.

Candès, E., Li, X., Ma, Y., and Wright, J.A. (2011). Robust principal component analysis? JACM 58, 11.

Canete, N.P., Iyengar, S.S., Ormerod, J.T., Baharlou, H., Harman, A.N., and Patrick, E. (2022). spicyR: spatial analysis of *in situ* cytometry data in R. Bioinformatics 38, 3099–3105.

Cang, Z., and Nie, Q. (2020). Inferring spatial and signaling relationships between cells from single cell transcriptomic data. Nat Commun 11, 2084.

Cang, Z., Ning, X., Nie, A., Xu, M., and Zhang, J. (2021). SCAN-IT: domain segmentation of spatial transcriptomics images by graph neural network. BMVC 32, 406.

Cang, Z., Zhao, Y., Almet, A.A., Stabell, A., Ramos, R., Plikus, M.V., Atwood, S.X., and Nie, Q. (2023). Screening cell-cell communication in spatial transcriptomics via collective optimal transport. Nat Methods 20, 218–228.

Cannoodt, R., Saelens, W., Deconinck, L., and Saeys, Y. (2021). Spearheading future omics analyses using dyngen, a multi-modal simulator of single cells. Nat Commun 12, 3942.

Cao, J., Cusanovich, D.A., Ramani, V., Aghamirzaie, D., Pliner, H.A., Hill, A.J., Daza, R.M., McFaline-Figueroa, J.L., Packer, J.S., Christiansen, L., et al. (2018). Joint profiling of chromatin accessibility and gene expression in thousands of single cells. Science 361, 1380–1385.

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. (2017). Comprehensive single-cell transcriptional profiling of a multicellular organism. Science 357, 661–667.

Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A.J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F.J., et al. (2019). The single-cell transcriptional landscape of mammalian organogenesis. Nature 566, 496–502.

Cao, Y., Su, B., Guo, X., Sun, W., Deng, Y., Bao, L., Zhu, Q., Zhang, X., Zheng, Y., Geng, C., et al. (2020a). Potent neutralizing antibodies against SARS-CoV-2 identified by high-throughput single-cell sequencing of convalescent patients' B cells. Cell 182, 73–84.e16.

Cao, Y., Wang, X., and Peng, G. (2020b). SCSA: a cell type annotation tool for single-cell RNA-seq data. Front Genet 11, 490.

Cao, Z.J., and Gao, G. (2022). Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. Nat Biotechnol 40, 1458–1466.

Carter, B., Ku, W.L., Kang, J.Y., Hu, G., Perrie, J., Tang, Q., and Zhao, K. (2019). Mapping histone modifications in low cell number and single cells using antibody-guided chromatin tagmentation (ACT-seq). Nat Commun 10, 3747.

Carter, B., and Zhao, K. (2021). The epigenetic basis of cellular heterogeneity. Nat Rev Genet 22, 235–250.

Casey, M.J., Fliege, J., Sánchez-García, R.J., and MacArthur, B.D. (2023). An information-theoretic approach to single cell sequencing analysis. bioRxiv, doi: 10.1101/2020.10.01.322255.

Castro, D.C., Xie, Y.R., Rubakhin, S.S., Romanova, E.V., and Sweedler, J.V. (2021). Image-guided MALDI mass spectrometry for high-throughput single-organelle characterization. Nat Methods 18, 1233–1238.

Chang, Y., He, F., Wang, J., Chen, S., Li, J., Liu, J., Yu, Y., Su, L., Ma, A., Allen, C., et al. (2022). Define and visualize pathological architectures of human tissues from spatially resolved transcriptomics using deep learning. Comput Struct Biotechnol J 20, 4600–4617.

Chatterton, Z., Lamichhane, P., Ahmadi Rastegar, D., Fitzpatrick, L., Lebhar, H., Marquis, C., Halliday, G., and Kwok, J.B. (2023). Single-cell DNA methylation sequencing by combinatorial indexing and enzymatic DNA methylation conversion. Cell Biosci 13, 2.

Chen, A., Liao, S., Cheng, M., Ma, K., Wu, L., Lai, Y., Qiu, X., Yang, J., Xu, J., Hao, S., et al. (2022). Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. Cell 185, 1777–1792.e21.

Chen, C., Wu, C., Wu, L., Wang, X., Deng, M., and Xi, R. (2020a). scRMD: imputation for single cell RNA-seq data via robust matrix decomposition. Bioinformatics 36, 3156–3161.

Chen, C., Xing, D., Tan, L., Li, H., Zhou, G., Huang, L., and Xie, X.S. (2017a). Single-cell whole-genome analyses by Linear Amplification via Transposon Insertion (LIANTI). Science 356, 189–194.

Chen, F., Lin, L., Zhang, J., He, Z., Uchiyama, K., and Lin, J.M. (2016a). Single-cell analysis using drop-on-demand inkjet printing and probe electrospray ionization mass spectrometry. Anal Chem 88, 4354–4360.

Chen, G., Ning, B., and Shi, T. (2019a). Single-cell RNA-seq technologies and related computational data analysis. Front Genet 10, 317.

Chen, H., Albergante, L., Hsu, J.Y., Lareau, C.A., Lo Bosco, G., Guan, J., Zhou, S., Gorban, A.N., Bauer, D.E., Aryee, M.J., et al. (2019b). Single-cell trajectories reconstruction, exploration and mapping of omics data with STREAM. Nat Commun 10, 1903.

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

Chen, H.I.H., Jin, Y., Huang, Y., and Chen, Y. (2016b). Detection of high variability in gene expression from single-cell RNA-seq profiling. BMC Genomics 17, 508.

Chen, J., Suo, S., Tam, P.P., Han, J.D.J., Peng, G., and Jing, N. (2017b). Spatial transcriptomic analysis of cryosectioned tissue samples with Geo-seq. Nat Protoc 12, 566–580.

Chen, K.H., Boettiger, A.N., Moffitt, J.R., Wang, S., and Zhuang, X. (2015a). Spatially resolved, highly multiplexed RNA profiling in single cells. Science 348, aaa6090.

Chen, L., and Zheng, S. (2018). BCseq: accurate single cell RNA-seq quantification with bias correction. Nucleic Acids Res 46, e82.

Chen, M., and Zhou, X. (2018). VIPER: variability-preserving imputation for accurate gene expression recovery in single-cell RNA sequencing studies. Genome Biol 19, 196.

Chen, P.Y., Cokus, S.J., and Pellegrini, M. (2010). BS Seeker: precise mapping for bisulfite sequencing. BMC Bioinformatics 11, 203.

Chen, Q., Yan, G., Gao, M., and Zhang, X. (2015b). Ultrasensitive proteome profiling for 100 living cells by direct cell injection, online digestion and nano-LC-MS/MS analysis. Anal Chem 87, 6674–6680.

Chen, S., Lake, B.B., and Zhang, K. (2019d). High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. Nat Biotechnol 37, 1452–1457.

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34, i884–i890.

Chen, W., Zhao, Y., Chen, X., Yang, Z., Xu, X., Bi, Y., Chen, V., Li, J., Choi, H., Ernest, B., et al. (2021a). A multicenter study benchmarking single-cell RNA sequencing technologies using reference samples. Nat Biotechnol 39, 1103–1114.

Chen, W.T., Lu, A., Craessaerts, K., Pavie, B., Sala Frigerio, C., Corthout, N., Qian, X., Laláková, J., Kühnemund, M., Voytyuk, I., et al. (2020b). Spatial transcriptomics and *in situ* sequencing to study Alzheimer's disease. Cell 182, 976–991.e19.

Chen, Z., Yang, Z., Yuan, X., Zhang, X., and Hao, P. (2021b). scSensitiveGeneDefine: sensitive gene detection in single-cell RNA sequencing data by Shannon entropy. BMC Bioinformatics 22, 211.

Cheng, J., Zhang, J., Wu, Z., and Sun, X. (2021). Inferring microenvironmental regulation of gene expression from single-cell RNA sequencing data using scMLnet with an application to COVID-19. Brief Bioinform 22, 988–1005.

Cheow, L.F., Courtois, E.T., Tan, Y., Viswanathan, R., Xing, Q., Tan, R.Z., Tan, D.S.W., Robson, P., Loh, Y.H., Quake, S.R., et al. (2016). Single-cell multimodal profiling reveals cellular epigenetic heterogeneity. Nat Methods 13, 833–836.

Cheung, T.K., Lee, C.Y., Bayer, F.P., McCoy, A., Kuster, B., and Rose, C.M. (2021). Defining the carrier proteome limit for single-cell proteomics. Nat Methods 18, 76–83.

Chidester, B., Zhou, T., and Ma, J. (2021). SPICEMIX: integrative single-cell spatial modeling for inferring cell identity. bioRxiv, doi: 10.1101/2020.11.29.383067.

Cho, C.S., Xi, J., Si, Y., Park, S.R., Hsu, J.E., Kim, M., Jun, G., Kang, H.M., and Lee, J.H. (2021). Microscopic examination of spatial transcriptome using Seq-Scope. Cell 184, 3559–3572.e22.

Cho, H., Berger, B., and Peng, J. (2018a). Generalizable and scalable visualization of single-cell data using neural networks. Cell Syst 7, 185–191.e4.

Cho, S.W., Xu, J., Sun, R., Mumbach, M.R., Carter, A.C., Chen, Y.G., Yost, K.E., Kim, J., He, J., Nevins, S.A., et al. (2018b). Promoter of lncRNA gene PVT1 is a tumor-suppressor DNA boundary element. Cell 173, 1398–1412.e22.

Chu, W.K., Edge, P., Lee, H.S., Bansal, V., Bafna, V., Huang, X., and Zhang, K. (2017). Ultraaccurate genome sequencing and haplotyping of single human cells. Proc Natl Acad Sci USA 114, 12512–12517.

Chua, R.L., Lukassen, S., Trump, S., Hennig, B.P., Wendisch, D., Pott, F., Debnath, O., Thürmann, L., Kurth, F., Völker, M.T., et al. (2020). COVID-19 severity correlates with airway epithelium–immune cell interactions identified by single-cell analysis. Nat Biotechnol 38, 970–979.

Chung, N.C., and Storey, J.D. (2015). Statistical significance of variables driving systematic variation in high-dimensional data. Bioinformatics 31, 545–554.

Clark, S.J., Argelaguet, R., Kapourani, C.A., Stubbs, T.M., Lee, H.J., Alda-Catalinas, C., Krueger, F., Sanguinetti, G., Kelsey, G., Marioni, J.C., et al. (2018). scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. Nat Commun 9, 781.

Clark, S.J., Smallwood, S.A., Lee, H.J., Krueger, F., Reik, W., and Kelsey, G. (2017). Genome-wide base-resolution mapping of DNA methylation in single cells using single-cell bisulfite sequencing (scBS-seq). Nat Protoc 12, 534–547.

Codeluppi, S., Borm, L.E., Zeisel, A., La Manno, G., van Lunteren, J.A., Svensson, C.I., and Linnarsson, S. (2018). Spatial organization of the somatosensory cortex revealed by osmFISH. Nat Methods 15, 932–935.

Cokus, S.J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C.D., Pradhan, S., Nelson, S.F., Pellegrini, M., and Jacobsen, S.E. (2008). Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature 452, 215–219.

Collins, M., Dasgupta, S., and Schapire, R.E. (2002). A generalization of principal component analysis to the exponential family. In: Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic. Cambridge: MIT Press.

Conde, D., Triozzi, P.M., Balmant, K.M., Doty, A.L., Miranda, M., Boullosa, A., Schmidt, H.W., Pereira, W.J., Dervinis, C., and Kirst, M. (2021). A robust method of nuclei isolation for single-cell RNA sequencing of solid tissues from the plant genus *Populus*. PLoS ONE 16, e0251149.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819–823.

Cong, Y., Motamedchaboki, K., Misal, S.A., Liang, Y., Guise, A.J., Truong, T., Huguet, R., Plowey, E.D., Zhu, Y., Lopez-Ferrer, D., et al. (2020). Ultrasensitive single-cell proteomics workflow identifies >1000 protein groups per mammalian cell. Chem Sci 12, 1001–1006.

Consortium, E.P., Moore, J.E., Purcaro, M.J., Pratt, H.E., Epstein, C.B., Shoresh, N., Adrian, J., Kawli, T., Davis, C.A., Dobin, A., et al. (2020a). Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature 583, 699–710.

Consortium, E.P., Snyder, M.P., Gingeras, T.R., Moore, J.E., Weng, Z., Gerstein, M.B., Ren, B., Hardison, R.C., Stamatoyannopoulos, J.A., Graveley, B.R., et al. (2020b). Perspectives on ENCODE. Nature 583, 693–698.

Corces, M.R., Shcherbina, A., Kundu, S., Gloudemans, M.J., Frésard, L., Granja, J.M., Louie, B.H., Eulalio, T., Shams, S., Bagdatli, S.T., et al. (2020). Single-cell epigenomic analyses implicate candidate causal variants at inherited risk loci for Alzheimer's and Parkinson's diseases. Nat Genet 52, 1158–1168.

Cortal, A., Martignetti, L., Six, E., and Rausell, A. (2021). Gene signature extraction and cell identity recognition at the single-cell level with Cell-ID. Nat Biotechnol 39, 1095–1102.

Creixell, P., Reimand, J., Haider, S., Wu, G., Shibata, T., Vazquez, M., Mustonen, V., Gonzalez-Perez, A., Pearson, J., Sander, C., et al. (2015). Pathway and network analysis of cancer genomes. Nat Methods 12, 615–621.

Crosse, E.I., Gordon-Keylock, S., Rybtsov, S., Binagui-Casas, A., Felchle, H., Nnadi, N. C., Kirschner, K., Chandra, T., Tamagno, S., Webb, D.J., et al. (2020). Multi-layered spatial transcriptomics identify secretory factors promoting human hematopoietic stem cell development. Cell Stem Cell 27, 822–839.e8.

Csardi, G., and Nepusz, T. (2006). The igraph software package for complex network research. InterJournal Complex Systems 1695, 1–9.

Ctortecka, C., Hartlmayr, D., Seth, A., Mendjan, S., Tourniaire, G., and Mechtler, K. (2022a). An automated workflow for multiplexed single-cell proteomics sample preparation at unprecedented sensitivity. bioRxiv, doi: 10.1101/2021.04.14.439828.

Ctortecka, C., Krššáková, G., Stejskal, K., Penninger, J.M., Mendjan, S., Mechtler, K., and Stadlmann, J. (2022b). Comparative proteome signatures of trace samples by multiplexed data-independent acquisition. Mol Cell Proteomics 21, 100177.

Cui, Y., Li, C., Jiang, Z., Zhang, S., Li, Q., Liu, X., Zhou, Y., Li, R., Wei, L., Li, L., et al. (2021). Single-cell transcriptome and genome analyses of pituitary neuroendocrine tumors. Neuro Oncol 23, 1859–1871.

Cui, Y., Zheng, Y., Liu, X., Yan, L., Fan, X., Yong, J., Hu, Y., Dong, J., Li, Q., Wu, X., et al. (2019). Single-cell transcriptome analysis maps the developmental track of the human heart. Cell Rep 26, 1934–1950.e5.

Cusanovich, D.A., Daza, R., Adey, A., Pliner, H.A., Christiansen, L., Gunderson, K.L., Steemers, F.J., Trapnell, C., and Shendure, J. (2015). Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. Science 348, 910–914.

Dago, A.E., Stepansky, A., Carlsson, A., Luttgen, M., Kendall, J., Baslan, T., Kolatkar, A., Wigler, M., Bethel, K., Gross, M.E., et al. (2014). Rapid phenotypic and genomic change in response to therapeutic pressure in prostate cancer inferred by high content analysis of single circulating tumor cells. PLoS ONE 9, e101777.

Daina, G., Ramos, L., Obradors, A., Rius, M., Martinez-Pasarell, O., Polo, A., del Rey, J., Obradors, J., Benet, J., and Navarro, J. (2013). First successful double-factor PGD for Lynch syndrome: monogenic analysis and comprehensive aneuploidy screening. Clin Genet 84, 70–73.

Dal Molin, A., and Di Camillo, B. (2019). How to design a single-cell RNA-sequencing experiment: pitfalls, challenges and perspectives. Brief Bioinform 20, 1384–1394.

Danese, A., Richter, M.L., Chaichoompu, K., Fischer, D.S., Theis, F.J., and Colomé-Tatché, M. (2021). EpiScanpy: integrated single-cell epigenomic analysis. Nat Commun 12, 5228.

Dang, Y., Zhu, L., Yuan, P., Liu, Q., Guo, Q., Chen, X., Gao, S., Liu, X., Ji, S., Yuan, Y., et al. (2023). Functional profiling of stage-specific proteome and translational transition across human pre-implantation embryo development at a single-cell resolution. Cell Discov 9, 10.

Datlinger, P., Rendeiro, A.F., Schmidl, C., Krausgruber, T., Traxler, P., Klughammer, J., Schuster, L.C., Kuchler, A., Alpar, D., and Bock, C. (2017). Pooled CRISPR screening with single-cell transcriptome readout. Nat Methods 14, 297–301.

Davies, D.L., and Bouldin, D.W. (1979). A cluster separation measure. IEEE Trans Pattern Anal Mach Intell PAMI-1, 224–227.

de Boer, C.G., and Regev, A. (2018). BROCKMAN: deciphering variance in epigenomic regulators by k-mer factorization. BMC Bioinformatics 19, 253.

de Kanter, J.K., Lijnzaad, P., Candelli, T., Margaritis, T., and Holstege, F.C.P. (2019). CHETAH: a selective, hierarchical cell type identification method for single-cell RNA sequencing. Nucleic Acids Res 47, e95.

de Souza, N. (2012). The ENCODE project. Nat Methods 9, 1046.

Dean, F.B., Nelson, J.R., Giesler, T.L., and Lasken, R.S. (2001). Rapid amplification of plasmid and phage DNA using Phi29 DNA polymerase and multiply-primed rolling circle amplification. Genome Res 11, 1095–1099.

DeLaughter, D.M. (2018). The use of the fluidigm C1 for RNA expression analyses of single cells. CP Mol Biol 122, e55.

Delcher, A.L., Phillippy, A., Carlton, J., and Salzberg, S.L. (2002). Fast algorithms for large-scale genome alignment and comparison. Nucleic Acids Res 30, 2478–2483.

Delmans, M., and Hemberg, M. (2016). Discrete distributional differential expression (D3E)—a tool for gene expression analysis of single-cell RNA-seq data. BMC Bioinformatics 17, 110.

Deng, Y., Bartosovic, M., Kukanja, P., Zhang, D., Liu, Y., Su, G., Enninful, A., Bai, Z., Castelo-Branco, G., and Fan, R. (2022a). Spatial-CUT&Tag: spatially resolved chromatin modification profiling at the cellular level. Science 375, 681–686.

Deng, Y., Bartosovic, M., Ma, S., Zhang, D., Kukanja, P., Xiao, Y., Su, G., Liu, Y., Qin, X., Rosoklija, G.B., et al. (2022b). Spatial profiling of chromatin accessibility in mouse and human tissues. Nature 609, 375–383.

DePasquale, E.A.K., Schnell, D.J., Van Camp, P.J., Valiente-Alandí, Í., Blaxall, B.C., Grimes, H.L., Singh, H., and Salomonis, N. (2019). DoubletDecon: deconvoluting doublets from single-cell RNA-sequencing data. Cell Rep 29, 1718–1727.e8.

Dephoure, N., and Gygi, S.P. (2012). Hyperplexing: a method for higher-order multiplexed quantitative proteomics provides a map of the dynamic response to rapamycin in yeast. Sci Signal 5, rs2.

Derks, J., Leduc, A., Wallmann, G., Huffman, R.G., Willetts, M., Khan, S., Specht, H., Ralser, M., Demichev, V., and Slavov, N. (2022). Increasing the throughput of sensitive proteomics by plexDIA. bioRxiv, doi: 10.1101/2021.11.03.467007.

DeTomaso, D., and Yosef, N. (2016). FastProject: a tool for low-dimensional analysis of single-cell RNA-Seq data. BMC Bioinformatics 17, 315.

Dey, S.S., Kester, L., Spanjaard, B., Bienko, M., and van Oudenaarden, A. (2015). Integrated genome and transcriptome sequencing of the same cell. Nat Biotechnol 33, 285–289.

Dhainaut, M., Rose, S.A., Akturk, G., Wroblewska, A., Nielsen, S.R., Park, E.S., Buckup, M., Roudko, V., Pia, L., Sweeney, R., et al. (2022). Spatial CRISPR genomics identifies regulators of the tumor microenvironment. Cell 185, 1223–1239.e20.

Dimitrov, D., Türei, D., Garrido-Rodriguez, M., Burmedi, P.L., Nagai, J.S., Boys, C., Ramirez Flores, R.O., Kim, H., Szalai, B., Costa, I.G., et al. (2022). Comparison of methods and resources for cell-cell communication inference from single-cell RNA-Seq data. Nat Commun 13, 3224.

Dimov, I.K., Kijanka, G., Park, Y., Ducrée, J., Kang, T., and Lee, L.P. (2011). Integrated microfluidic array plate (iMAP) for cellular and molecular analysis. Lab Chip 11, 2701.

Ding, B., Zheng, L., Zhu, Y., Li, N., Jia, H., Ai, R., Wildberg, A., and Wang, W. (2015). Normalization and noise reduction for single cell RNA-seq experiments. Bioinformatics 31, 2225–2227.

Ding, J., Shah, S., and Condon, A. (2016). densityCut: an efficient and versatile topological approach for automatic clustering of biological data. Bioinformatics 32, 2567–2576.

Ding, J., Sharon, N., and Bar-Joseph, Z. (2022). Temporal modelling using single-cell transcriptomics. Nat Rev Genet 23, 355–368.

Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C.P., Jerby-Arnon, L., Marjanovic, N.D., Dionne, D., Burks, T., Raychowdhury, R., et al. (2016). Perturb-seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. Cell 167, 1853–1866.e17.

Do, V.H., and Canzar, S. (2021). A generalization of t-SNE and UMAP to single-cell multimodal omics. Genome Biol 22, 130.

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

Dohmen, J., Baranovskii, A., Ronen, J., Uyar, B., Franke, V., and Akalin, A. (2021). Identifying tumor cells at the single cell level. bioRxiv, doi: 10.1101/2021.10.15.463909.

Domínguez Conde, C., Xu, C., Jarvis, L.B., Rainbow, D.B., Wells, S.B., Gomes, T., Howlett, S.K., Suchanek, O., Polanski, K., King, H.W., et al. (2022). Cross-tissue immune cell analysis reveals tissue-specific features in humans. Science 376, eabl5197.

Dong, K., and Zhang, S. (2021). Joint reconstruction of *cis*-regulatory interaction networks across multiple tissues using single-cell chromatin accessibility data. Brief Bioinform 22, bbaa120.

Dong, K., and Zhang, S. (2022). Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. Nat Commun 13, 1739.

Dong, R., and Yuan, G.C. (2020). GiniClust3: a fast and memory-efficient tool for rare cell type identification. BMC Bioinformatics 21, 158.

Dong, R., and Yuan, G.C. (2021). SpatialDWLS: accurate deconvolution of spatial transcriptomic data. Genome Biol 22, 145.

Dong, X., Tang, K., Xu, Y., Wei, H., Han, T., and Wang, C. (2022). Single-cell gene regulation network inference by large-scale data integration. Nucleic Acids Res 50, e126.

Dries, R., Chen, J., del Rossi, N., Khan, M.M., Sistig, A., and Yuan, G.C. (2021a). Advances in spatial transcriptomic data analysis. Genome Res 31, 1706–1718.

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

Du, S., Zhai, L., Ye, S., Wang, L., Liu, M., and Tan, M. (2023). In-depth urinary and exosome proteome profiling analysis identifies novel biomarkers for diabetic kidney disease. Sci China Life Sci 66, 2587–2603.

Duan, B., Zhou, C., Zhu, C., Yu, Y., Li, G., Zhang, S., Zhang, C., Ye, X., Ma, H., Qu, S., et al. (2019). Model-based understanding of single-cell CRISPR screening. Nat Commun 10, 2233.

Duan, H., Li, F., Shang, J., Liu, J., Li, Y., and Liu, X. (2022). scVAEBGM: clustering analysis of single-cell ATAC-seq data using a deep generative model. Interdiscip Sci 14, 917–928.

Dueñas, M.E., Essner, J.J., and Lee, Y.J. (2017). 3D MALDI mass spectrometry imaging of a single cell: spatial mapping of lipids in the embryonic development of zebrafish. Sci Rep 7, 14946.

Duncan, K.D., Bergman, H.M., and Lanekoff, I. (2017). A pneumatically assisted nanospray desorption electrospray ionization source for increased solvent versatility and enhanced metabolite detection from tissue. Analyst 142, 3424–3431.

Duren, Z., Chang, F., Naqing, F., Xin, J., Liu, Q., and Wong, W.H. (2022). Regulatory analysis of single cell multiome gene expression and chromatin accessibility data with scREG. Genome Biol 23, 114.

Duren, Z., Chen, X., Zamanighomi, M., Zeng, W., Satpathy, A.T., Chang, H.Y., Wang, Y., and Wong, W.H. (2018). Integrative analysis of single-cell genomics data by coupled nonnegative matrix factorizations. Proc Natl Acad Sci USA 115, 7723–7728.

Edsgärd, D., Johnsson, P., and Sandberg, R. (2018). Identification of spatial expression trends in single-cell gene expression data. Nat Methods 15, 339–342.

Efremova, M., Vento-Tormo, M., Teichmann, S.A., and Vento-Tormo, R. (2020). CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. Nat Protoc 15, 1484–1506.

Ellis, M.J., Ding, L., Shen, D., Luo, J., Suman, V.J., Wallis, J.W., van Tine, B.A., Hoog, J., Goiffon, R.J., Goldstein, T.C., et al. (2012). Whole-genome analysis informs breast cancer response to aromatase inhibition. Nature 486, 353–360.

Elosua-Bayes, M., Nieto, P., Mereu, E., Gut, I., and Heyn, H. (2021). SPOTlight: seeded NMF regression to deconvolute spatial transcriptomics spots with single-cell transcriptomes. Nucleic Acids Res 49, e50.

Elyanow, R., Dumitrascu, B., Engelhardt, B.E., and Raphael, B.J. (2020). netNMF-sc: leveraging gene-gene interactions for imputation and dimensionality reduction in single-cell expression analysis. Genome Res 30, 195–204.

Emara, S., Amer, S., Ali, A., Abouleila, Y., Oga, A., and Masujima, T. (2017). Single-cell metabolomics. In: Sussulini, A., ed. Metabolomics: From Fundamentals to Clinical Applications. Advances in Experimental Medicine and Biology. Cham: Springer, 323–343.

Emmert-Buck, M.R., Bonner, R.F., Smith, P.D., Chuaqui, R.F., Zhuang, Z., Goldstein, S.R., Weiss, R.A., and Liotta, L.A. (1996). Laser capture microdissection. Science 274, 998–1001.

Eng, C.H.L., Lawson, M., Zhu, Q., Dries, R., Koulena, N., Takei, Y., Yun, J., Cronin, C., Karp, C., Yuan, G.C., et al. (2019). Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. Nature 568, 235–239.

Engreitz, J.M., Haines, J.E., Perez, E.M., Munson, G., Chen, J., Kane, M., McDonel, P.E., Guttman, M., and Lander, E.S. (2016). Local regulation of gene expression by lncRNA promoters, transcription and splicing. Nature 539, 452–455.

Eraslan, G., Drokhlyansky, E., Anand, S., Fiskin, E., Subramanian, A., Slyper, M., Wang, J., Van Wittenberghe, N., Rouhana, J.M., Waldman, J., et al. (2022). Single-nucleus cross-tissue molecular reference maps toward understanding disease gene function. Science 376, eabl4290.

Eraslan, G., Simon, L.M., Mircea, M., Mueller, N.S., and Theis, F.J. (2019). Single-cell RNA-seq denoising using a deep count autoencoder. Nat Commun 10, 390.

Ernst, J., and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. Nat Methods 9, 215–216.

Ester, M., Kriegel, H.P., Sander, J., and Xu, X.W. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. Portland Oregon. 226–231.

Evans, C., Hardin, J., and Stoebel, D.M. (2018). Selecting between-sample RNA-Seq normalization methods from the perspective of their assumptions. Brief Bioinform 19, 776–792.

Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B., Korninger, F., May, B., et al. (2018). The reactome pathway knowledgebase. Nucleic Acids Res 46, D649–D655.

Fan, H.C., Fu, G.K., and Fodor, S.P.A. (2015a). Combinatorial labeling of single cells for gene expression cytometry. Science 347, 1258367.

Fan, X., Zhang, X., Wu, X., Guo, H., Hu, Y., Tang, F., and Huang, Y. (2015b). Single-cell RNA-seq transcriptome analysis of linear and circular RNAs in mouse preimplantation embryos. Genome Biol 16, 148.

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

Fang, X., and Ho, J.W.K. (2021). FlowGrid enables fast clustering of very large single-cell RNA-seq data. Bioinformatics 38, 282–283.

Fanok, M.H., Sun, A., Fogli, L.K., Narendran, V., Eckstein, M., Kannan, K., Dolgalev, I., Lazaris, C., Heguy, A., Laird, M.E., et al. (2018). Role of dysregulated cytokine signaling and bacterial triggers in the pathogenesis of cutaneous T-cell lymphoma. J Investig Dermatol 138, 1116–1125.

Farlik, M., Sheffield, N.C., Nuzzo, A., Datlinger, P., Schönegger, A., Klughammer, J., and Bock, C. (2015). Single-cell DNA methylome sequencing and bioinformatic inference of epigenomic cell-state dynamics. Cell Rep 10, 1386–1397.

Fawkner-Corbett, D., Antanaviciute, A., Parikh, K., Jagielowicz, M., Gerós, A.S., Gupta, T., Ashley, N., Khamis, D., Fowler, D., Morrissey, E., et al. (2021). Spatiotemporal analysis of human intestinal development at single-cell resolution. Cell 184, 810–826.e23.

Femino, A.M., Fay, F.S., Fogarty, K., and Singer, R.H. (1998). Visualization of single RNA transcripts *in situ*. Science 280, 585–590.

Feng, D., Li, H., Xu, T., Zheng, F., Hu, C., Shi, X., and Xu, G. (2022). High-throughput single cell metabolomics and cellular heterogeneity exploration by inertial microfluidics coupled with pulsed electric field-induced electrospray ionization-high resolution mass spectrometry. Anal Chim Acta 1221, 340116.

Feng, D., Whitehurst, C.E., Shan, D., Hill, J.D., and Yue, Y.G. (2019). Single Cell Explorer, collaboration-driven tools to leverage large-scale single cell RNA-seq data. BMC Genomics 20, 676.

Feng, Y., and Li, L.M. (2021). MUREN: a robust and multi-reference approach of RNA-seq transcript normalization. BMC Bioinformatics 22, 386.

Ferguson, C.N., Fowler, J.W.M., Waxer, J.F., Gatti, R.A., and Loo, J.A. (2014). Mass Spectrometry-Based Tissue Imaging of Small Molecules. In: Woods, A., and Darie, C., eds. Advancements of Mass Spectrometry in Biomedical Research. Advances in Experimental Medicine and Biology. Cham: Springer. 283–299.

Ferragina, P., and Manzini, G. (2001). An experimental study of an opportunistic index. In: Proceedings of the Twelfth Annual ACM-SIAM Symposium on Discrete Algorithms. Philadelphia: Society for Industrial and Applied Mathematics. 269–278.

Ferronika, P., van den Bos, H., Taudt, A., Spierings, D.C.J., Saber, A., Hiltermann, T.J. N., Kok, K., Porubsky, D., van der Wekken, A.J., Timens, W., et al. (2017). Copy number alterations assessed at the single-cell level revealed mono- and polyclonal seeding patterns of distant metastasis in a small-cell lung cancer patient. Ann Oncol 28, 1668–1670.

Finak, G., McDavid, A., Yajima, M., Deng, J., Gersuk, V., Shalek, A.K., Slichter, C.K., Miller, H.W., McElrath, M.J., Prlic, M., et al. (2015). MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. Genome Biol 16, 278.

Fischer, D.S., Schaar, A.C., and Theis, F.J. (2021). Learning cell communication from spatial graphs of cells. bioRxiv, doi: 10.1101/2021.07.11.451750.

Fischer, J., and Ayers, T. (2021). Single nucleus RNA-sequencing: how it's done, applications and limitations. Emerg Top Life Sci 5, 687–690.

Fleck, J.S., Jansen, S.M.J., Wollny, D., Zenk, F., Seimiya, M., Jain, A., Okamoto, R., Santel, M., He, Z., Camp, J.G., et al. (2023). Inferring and perturbing cell fate regulomes in human brain organoids. Nature 621, 365–372.

Fleming, S.J., Marioni, J.C., and Babadi, M. (2019). CellBender remove-background: a deep generative model for unsupervised removal of background noise from scRNA-seq datasets. bioRxiv, doi: 10.1101/791699.

Flyamer, I.M., Gassler, J., Imakaev, M., Brandão, H.B., Ulianov, S.V., Abdennur, N., Razin, S.V., Mirny, L.A., and Tachibana-Konwalski, K. (2017). Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. Nature 544, 110–114.

Fortunato, S. (2009). Community detection in graphs. Phys Rep 486, 75–174.

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. (2014). *EGFR* variant heterogeneity in glioblastoma resolved through single-nucleus sequencing. Cancer Discov 4, 956–971.

Franzén, O., Gan, L.M., and Björkegren, J.L.M. (2019). PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. Database 2019, baz046.

Frei, A.P., Bava, F.A., Zunder, E.R., Hsieh, E.W.Y., Chen, S.Y., Nolan, G.P., and Gherardini, P.F. (2016). Highly multiplexed simultaneous detection of RNAs and proteins in single cells. Nat Methods 13, 269–275.

Frommer, M., McDonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Grigg, G.W., Molloy, P.L., and Paul, C.L. (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc Natl Acad Sci USA 89, 1827–1831.

Fu, H., Xu, H., Chong, K., Li, M., Ang, K.S., Lee, H.K., Ling, J., Chen, A., Shao, L., Liu, L., et al. (2021a). Unsupervised spatially embedded deep representation of spatial transcriptomics. bioRxiv, doi: 10.1101/2021.06.15.448542.

Fu, L., Zhang, L., Dollinger, E., Peng, Q., Nie, Q., and Xie, X. (2020). Predicting transcription factor binding in single cells through deep learning. Sci Adv 6, eaba9031.

Fu, X., Sun, L., Chen, J.Y., Dong, R., Lin, Y., Palmiter, R.D., Lin, S., and Gu, L. (2021b). Continuous polony gels for tissue mapping with high resolution and RNA capture efficiency. bioRxiv, doi: 10.1101/2021.03.17.435795.

Fu, Y., Li, C., Lu, S., Zhou, W., Tang, F., Xie, X.S., and Huang, Y. (2015). Uniform and accurate single-cell sequencing based on emulsion whole-genome amplification. Proc Natl Acad Sci USA 112, 11923–11928.

Fujii, T., Matsuda, S., Tejedor, M.L., Esaki, T., Sakane, I., Mizuno, H., Tsuyama, N., and Masujima, T. (2015). Direct metabolomics for plant cells by live single-cell mass spectrometry. Nat Protoc 10, 1445–1456.

Fukunaga, K., and Hostetler, L. (1975). The estimation of the gradient of a density function, with applications in pattern recognition. IEEE Trans Inform Theor 21, 32–40.

Gabitto, M.I., Rasmussen, A., Wapinski, O., Allaway, K., Carriero, N., Fishell, G.J., and Bonneau, R. (2020). Characterizing chromatin landscape from aggregate and single-cell genomic assays using flexible duration modeling. Nat Commun 11, 747.

Galler, K., Bräutigam, K., Große, C., Popp, J., and Neugebauer, U. (2014). Making a big thing of a small cell—recent advances in single cell analysis. Analyst 139, 1237–1273.

Gan, Y., Huang, X., Zou, G., Zhou, S., and Guan, J. (2022). Deep structural clustering for single-cell RNA-seq data jointly through autoencoder and graph neural network. Brief Bioinform 23, bbac018.

Gao, C., Liu, J., Kriebel, A.R., Preissl, S., Luo, C., Castanon, R., Sandoval, J., Rivkin, A., Nery, J.R., Behrens, M.M., et al. (2021a). Iterative single-cell multi-omic integration using online learning. Nat Biotechnol 39, 1000–1007.

Gao, W., Ku, W.L., Pan, L., Perrie, J., Zhao, T., Hu, G., Wu, Y., Zhu, J., Ni, B., and Zhao, K. (2021b). Multiplex indexing approach for the detection of DNase I hypersensitive sites in single cells. Nucleic Acids Res 49, e56.

Gao, X., Hu, D., Gogol, M., and Li, H. (2019). ClusterMap: compare multiple single cell RNA-Seq datasets across different experimental conditions. Bioinformatics 35, 3038–3045.

Garcia-Alonso, L., Handfield, L.F., Roberts, K., Nikolakopoulou, K., Fernando, R.C., Gardner, L., Woodhams, B., Arutyunyan, A., Polanski, K., Hoo, R., et al. (2021). Mapping the temporal and spatial dynamics of the human endometrium *in vivo* and *in vitro*. Nat Genet 53, 1698–1711.

Gauthier, M., Agniel, D., Thiébaut, R., and Hejblum, B.P. (2021). Distribution-free complex hypothesis testing for single-cell RNA-seq differential expression analysis. bioRxiv, doi: 10.1101/2021.05.21.445165.

Gawad, C., Koh, W., and Quake, S.R. (2014). Dissecting the clonal origins of childhood acute lymphoblastic leukemia by single-cell genomics. Proc Natl Acad Sci USA 111, 17947–17952.

Gaydosik, A.M., Tabib, T., Domsic, R., Khanna, D., Lafyatis, R., and Fuschiotti, P. (2021). Single-cell transcriptome analysis identifies skin-specific T-cell responses in systemic sclerosis. Ann Rheum Dis 80, 1453–1460.

Gebreyesus, S.T., Siyal, A.A., Kitata, R.B., Chen, E.S.W., Enkhbayar, B., Angata, T., Lin, K.I., Chen, Y.J., and Tu, H.L. (2022). Streamlined single-cell proteomics by an integrated microfluidic chip and data-independent acquisition mass spectrometry. Nat Commun 13, 37.

Genshaft, A.S., Li, S., Gallant, C.J., Darmanis, S., Prakadan, S.M., Ziegler, C.G.K., Lundberg, M., Fredriksson, S., Hong, J., Regev, A., et al. (2016). Multiplexed, targeted profiling of single-cell proteomes and transcriptomes in a single reaction. Genome Biol 17, 188.

Gerlach, J.P., van Buggenum, J.A.G., Tanis, S.E.J., Hogeweg, M., Heuts, B.M.H., Muraro, M.J., Elze, L., Rivello, F., Rakszewska, A., van Oudenaarden, A., et al. (2019). Combined quantification of intracellular (phospho-)proteins and transcriptomics from fixed single cells. Sci Rep 9, 1469.

Gerniers, A., Bricard, O., and Dupont, P. (2021). MicroCellClust: mining rare and highly specific subpopulations from single-cell expression data. Bioinformatics 37, 3220–3227.

Ghaddar, B., and De, S. (2022). Reconstructing physical cell interaction networks from single-cell data using Neighbor-seq. Nucleic Acids Res 50, e82.

Gierahn, T.M., Wadsworth Ii, M.H., Hughes, T.K., Bryson, B.D., Butler, A., Satija, R., Fortune, S., Love, J.C., and Shalek, A.K. (2017). Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. Nat Methods 14, 395–398.

Giesen, C., Wang, H.A.O., Schapiro, D., Zivanovic, N., Jacobs, A., Hattendorf, B., Schüffler, P.J., Grolimund, D., Buhmann, J.M., Brandt, S., et al. (2014). Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. Nat Methods 11, 417–422.

Giladi, A., Cohen, M., Medaglia, C., Baran, Y., Li, B., Zada, M., Bost, P., Blecher-Gonen, R., Salame, T.M., Mayer, J.U., et al. (2020). Dissecting cellular crosstalk by sequencing physically interacting cells. Nat Biotechnol 38, 629–637.

Gogolewski, K., Sykulski, M., Chung, N.C., and Gambin, A. (2019). Truncated robust principal component analysis and noise reduction for single cell RNA sequencing data. J Comput Biol 26, 782–793.

Goldstein, L.D., Chen, Y.J.J., Dunne, J., Mir, A., Hubschle, H., Guillory, J., Yuan, W., Zhang, J., Stinson, J., Jaiswal, B., et al. (2017). Massively parallel nanowell-based single-cell gene expression profiling. BMC Genomics 18, 519.

Gong, B., Zhou, Y., and Purdom, E. (2021). Cobolt: integrative analysis of multimodal single-cell sequencing data. Genome Biol 22, 351.

Gong, X., Zhao, Y., Cai, S., Fu, S., Yang, C., Zhang, S., and Zhang, X. (2014). Single cell analysis with probe ESI-mass spectrometry: detection of metabolites at cellular and subcellular levels. Anal Chem 86, 3809–3816.

Gong, Y., Srinivasan, S.S., Zhang, R., Kessenbrock, K., and Zhang, J. (2022). scEpiLock: a weakly supervised learning framework for cis-regulatory element localization and variant impact quantification for single-cell epigenetic data. Biomolecules 12, 874.

Gonzalez-Pena, V., Natarajan, S., Xia, Y., Klein, D., Carter, R., Pang, Y., Shaner, B., Annu, K., Putnam, D., Chen, W., et al. (2021). Accurate genomic variant detection in single cells with primary template-directed amplification. Proc Natl Acad Sci USA 118, e2024176118.

Govek, K.W., Yamajala, V.S., and G. Camara, P. (2019). Clustering-independent analysis of genomic data using spectral simplicial theory. PLoS Comput Biol 15, e1007509.

Gralinska, E., Kohl, C., Sokhandan Fadakar, B., and Vingron, M. (2022). Visualizing cluster-specific genes from single-cell transcriptomics data using association plots. J Mol Biol 434, 167525.

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

Gravina, S., Ganapathi, S., and Vijg, J. (2015). Single-cell, locus-specific bisulfite sequencing (SLBS) for direct detection of epimutations in DNA methylation patterns. Nucleic Acids Res 43, e93.

Graving, J.M., and Couzin, I.D. (2020). VAE-SNE: a deep generative model for simultaneous dimensionality reduction and clustering. bioRxiv, doi: 10.1101/2020.07.17.207993.

Greguš, M., Kostas, J.C., Ray, S., Abbatiello, S.E., and Ivanov, A.R. (2020). Improved sensitivity of ultralow flow LC-MS-based proteomic profiling of limited samples using monolithic capillary columns and FAIMS technology. Anal Chem 92, 14702–14712.

Gross, A., Schoendube, J., Zimmermann, S., Steeb, M., Zengerle, R., and Koltay, P. (2015). Technologies for single-cell isolation. Int J Mol Sci 16, 16897–16919.

Grosselin, K., Durand, A., Marsolier, J., Poitou, A., Marangoni, E., Nemati, F., Dahmani, A., Lameiras, S., Reyal, F., Frenoy, O., et al. (2019). High-throughput single-cell ChIP-seq identifies heterogeneity of chromatin states in breast cancer. Nat Genet 51, 1060–1066.

Grothues, D., Cantor, C.R., and Smith, C.L. (1993). PCR amplification of megabase

DNA with tagged random primers (T-PCR). Nucl Acids Res 21, 1321–1322.

Grün, D., Lyubimova, A., Kester, L., Wiebrands, K., Basak, O., Sasaki, N., Clevers, H., and van Oudenaarden, A. (2015). Single-cell messenger RNA sequencing reveals rare intestinal cell types. Nature 525, 251–255.

Grün, D., and van Oudenaarden, A. (2015). Design and analysis of single-cell sequencing experiments. Cell 163, 799–810.

Grunau, C., Clark, S.J., and Rosenthal, A. (2001). Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. Nucleic Acids Res 29, 65e–65.

Gu, C., and Liu, Z. (2021). A network regularized linear model to infer spatial expression pattern for single cells. bioRxiv, doi: 10.1101/2021.03.07.434296.

Gu, H., Bock, C., Mikkelsen, T.S., Jäger, N., Smith, Z.D., Tomazou, E., Gnirke, A., Lander, E.S., and Meissner, A. (2010). Genome-scale DNA methylation mapping of clinical samples at single-nucleotide resolution. Nat Methods 7, 133–136.

Gu, L., Li, X., Li, Z., Wang, Q., Zheng, K., Yu, G., Dai, C., Li, J., Zhao, B., Zhang, H., et al. (2022a). Increasing the sensitivity, recovery, and integrality of spatially resolved proteomics by LCM-MTA. bioRxiv, doi: 10.1101/2022.08.21.504675.

Gu, L., Li, Z., Wang, Q., Zhang, H., Sun, Y., Li, C., and Wang, H. (2022b). An ultra-sensitive and easy-to-use multiplexed single-cell proteomic analysis. bioRxiv, doi: 10.1101/2022.01.02.474723.

Guo, H., and Li, J. (2021). scSorter: assigning cells to known cell types according to marker genes. Genome Biol 22, 69.

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

Guo, M., Bao, E.L., Wagner, M., Whitsett, J.A., and Xu, Y. (2017). SLICE: determining cell differentiation and lineage based on single cell entropy. Nucleic Acids Res 45, e54.

Guo, M., Wang, H., Potter, S.S., Whitsett, J.A., and Xu, Y. (2015). SINCERA: a pipeline for single-cell RNA-seq profiling analysis. PLoS Comput Biol 11, e1004575.

Guo, X., Su, J., Zhou, H., Liu, C., Cao, J., and Li, L. (2019). Community detection based on genetic algorithm using local structural similarity. IEEE Access 7, 134583–134600.

Gupta, I., Collier, P.G., Haase, B., Mahfouz, A., Joglekar, A., Floyd, T., Koopmans, F., Barres, B., Smit, A.B., Sloan, S.A., et al. (2018). Single-cell isoform RNA sequencing characterizes isoforms in thousands of cerebellar cells. Nat Biotechnol 36, 1197–1202.

Habib, N., Avraham-Davidi, I., Basu, A., Burks, T., Shekhar, K., Hofree, M., Choudhury, S.R., Aguet, F., Gelfand, E., Ardlie, K., et al. (2017). Massively parallel single-nucleus RNA-seq with DroNc-seq. Nat Methods 14, 955–958.

Habib, N., Li, Y., Heidenreich, M., Swiech, L., Avraham-Davidi, I., Trombetta, J.J., Hession, C., Zhang, F., and Regev, A. (2016). Div-Seq: single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons. Science 353, 925–928.

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

Hagège, H., Klous, P., Braem, C., Splinter, E., Dekker, J., Cathala, G., de Laat, W., and Forné, T. (2007). Quantitative analysis of chromosome conformation capture assays (3C-qPCR). Nat Protoc 2, 1722–1733.

Hagemann-Jensen, M., Ziegenhain, C., Chen, P., Ramsköld, D., Hendriks, G.J., Larsson, A.J.M., Faridani, O.R., and Sandberg, R. (2020). Single-cell RNA counting at allele and isoform resolution using Smart-seq3. Nat Biotechnol 38, 708–714.

Haghverdi, L., Büttner, M., Wolf, F.A., Buettner, F., and Theis, F.J. (2016). Diffusion pseudotime robustly reconstructs lineage branching. Nat Methods 13, 845–848.

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

Hahaut, V., Pavlinic, D., Carbone, W., Schuierer, S., Balmer, P., Quinodoz, M., Renner, M., Roma, G., Cowan, C.S., and Picelli, S. (2022). Fast and highly sensitive full-length single-cell RNA sequencing using FLASH-seq. Nat Biotechnol 40, 1447–1451.

Haigis, K.M., Cichowski, K., and Elledge, S.J. (2019). Tissue-specificity in cancer: the rule, not the exception. Science 363, 1150–1151.

Han, K.Y., Kim, K.T., Joung, J.G., Son, D.S., Kim, Y.J., Jo, A., Jeon, H.J., Moon, H.S., Yoo, C.E., Chung, W., et al. (2018a). SIDR: simultaneous isolation and parallel sequencing of genomic DNA and total RNA from single cells. Genome Res 28, 75–87.

Han, L., Wu, H.J., Zhu, H., Kim, K.Y., Marjani, S.L., Riester, M., Euskirchen, G., Zi, X., Yang, J., Han, J., et al. (2017). Bisulfite-independent analysis of CpG island methylation enables genome-scale stratification of single cells. Nucleic Acids Res 45, e77.

Han, Q., Bradshaw, E.M., Nilsson, B., Hafler, D.A., and Love, J.C. (2010). Multidimensional analysis of the frequencies and rates of cytokine secretion from single cells by quantitative microengraving. Lab Chip 10, 1391.

Han, W., Cheng, Y., Chen, J., Zhong, H., Hu, Z., Chen, S., Zong, L., Hong, L., Chan, T. F., King, I., et al. (2022). Self-supervised contrastive learning for integrative single cell RNA-seq data analysis. Brief Bioinform 23, bbac377.

Han, X., Wang, R., Zhou, Y., Fei, L., Sun, H., Lai, S., Saadatpour, A., Zhou, Z., Chen, H., Ye, F., et al. (2018b). Mapping the mouse cell atlas by microwell-seq. Cell 172, 1091–1107.e17.

Han, X., Zhou, Z., Fei, L., Sun, H., Wang, R., Chen, Y., Chen, H., Wang, J., Tang, H., Ge, W., et al. (2020). Construction of a human cell landscape at single-cell level. Nature 581, 303–309.

Hao, M., Hua, K., and Zhang, X. (2021a). SOMDE: a scalable method for identifying spatially variable genes with self-organizing map. Bioinformatics 37, 4392–4398.

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck Iii, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021b). Integrated analysis of multimodal single-cell data. Cell 184, 3573–3587.e29.

Harada, A., Maehara, K., Handa, T., Arimura, Y., Nogami, J., Hayashi-Takanaka, Y., Shirahige, K., Kurumizaka, H., Kimura, H., and Ohkawa, Y. (2019). A chromatin integration labelling method enables epigenomic profiling with lower input. Nat Cell Biol 21, 287–296.

Harris, R.A., Wang, T., Coarfa, C., Nagarajan, R.P., Hong, C., Downey, S.L., Johnson, B.E., Fouse, S.D., Delaney, A., Zhao, Y., et al. (2010). Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. Nat Biotechnol 28, 1097–1105.

Hasanaj, E., Wang, J., Sarathi, A., Ding, J., and Bar-Joseph, Z. (2022). Interactive single-cell data analysis using Cellar. Nat Commun 13, 1998.

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. (2016). CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. Genome Biol 17, 77.

Hashimshony, T., Wagner, F., Sher, N., and Yanai, I. (2012). CEL-seq: single-cell RNA-seq by multiplexed linear amplification. Cell Rep 2, 666–673.

He, B., Bergenstråhle, L., Stenbeck, L., Abid, A., Andersson, A., Borg, Å., Maaskola, J., Lundeberg, J., and Zou, J. (2020a). Integrating spatial gene expression and breast tumour morphology via deep learning. Nat Biomed Eng 4, 827–834.

He, D., Zakeri, M., Sarkar, H., Soneson, C., Srivastava, A., and Patro, R. (2022). Alevin-fry unlocks rapid, accurate and memory-frugal quantification of single-cell RNA-seq data. Nat Methods 19, 316–322.

He, J., Zhao, F., Chen, B., Cui, N., Li, Z., Qin, J., Luo, L., Zhao, C., and Li, L. (2023). Alterations in immune cell heterogeneities in the brain of aged zebrafish using single-cell resolution. Sci China Life Sci 66, 1358–1378.

He, S., Bhatt, R., Birditt, B., Brown, C., Brown, E., Chantranuvatana, K., Danaher, P., Dunaway, D., Filanoski, B., Garrison, R.G., et al. (2021a). High-plex multiomic analysis in FFPE tissue at single-cellular and subcellular resolution by spatial molecular imaging. bioRxiv, doi: 10.1101/2021.11.03.467020.

He, X., Cai, D., and Niyogi, P. (2005). Laplacian score for feature selection. In: Proceedings of the 18th International Conference on Neural Information Processing Systems. Cambridge: MIT Press. 507–514.

He, Y., Hariharan, M., Gorkin, D.U., Dickel, D.E., Luo, C., Castanon, R.G., Nery, J.R., Lee, A.Y., Zhao, Y., Huang, H., et al. (2020b). Spatiotemporal DNA methylome dynamics of the developing mouse fetus. Nature 583, 752–759.

He, Y., Tang, X., Huang, J., Ren, J., Zhou, H., Chen, K., Liu, A., Shi, H., Lin, Z., Li, Q., et al. (2021b). ClusterMap for multi-scale clustering analysis of spatial gene expression. Nat Commun 12, 5909.

Hebenstreit, D. (2012). Methods, challenges and potentials of single cell RNA-seq. Biology 1, 658–667.

Hedlund, E., and Deng, Q. (2018). Single-cell RNA sequencing: technical advancements and biological applications. Mol Aspects Med 59, 36–46.

Hercus, C. (2009). Novocraft short read alignment package. Available from URL: http://www.novocraft.com.

Herring, C.A., Banerjee, A., McKinley, E.T., Simmons, A.J., Ping, J., Roland, J.T., Franklin, J.L., Liu, Q., Gerdes, M.J., Coffey, R.J., et al. (2018). Unsupervised trajectory analysis of single-cell RNA-seq and imaging data reveals alternative tuft cell origins in the gut. Cell Syst 6, 37–51.e9.

Hicks, S.C., Liu, R., Ni, Y., Purdom, E., and Risso, D. (2021). mbkmeans: fast clustering for single cell data using mini-batch k-means. PLoS Comput Biol 17, e1008625.

Hie, B., Bryson, B., and Berger, B. (2019). Efficient integration of heterogeneous single-cell transcriptomes using Scanorama. Nat Biotechnol 37, 685–691.

Hill, A.J., McFaline-Figueroa, J.L., Starita, L.M., Gasperini, M.J., Matreyek, K.A., Packer, J., Jackson, D., Shendure, J., and Trapnell, C. (2018). On the design of CRISPR-based single-cell molecular screens. Nat Methods 15, 271–274.

Hoadley, K.A., Yau, C., Hinoue, T., Wolf, D.M., Lazar, A.J., Drill, E., Shen, R., Taylor, A.M., Cherniack, A.D., Thorsson, V., et al. (2018). Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. Cell 173, 291–304.e6.

Hochgerner, H., Lönnerberg, P., Hodge, R., Mikes, J., Heskol, A., Hubschle, H., Lin, P., Picelli, S., La Manno, G., Ratz, M., et al. (2017). STRT-seq-2i: dual-index 5′ single cell and nucleus RNA-seq on an addressable microwell array. Sci Rep 7, 16327.

Hou, R., Denisenko, E., and Forrest, A.R.R. (2019). scMatch: a single-cell gene expression profile annotation tool using reference datasets. Bioinformatics 35, 4688–4695.

Hou, R., Denisenko, E., Ong, H.T., Ramilowski, J.A., and Forrest, A.R.R. (2020). Predicting cell-to-cell communication networks using NATMI. Nat Commun 11, 5011.

Hou, Y., Fan, W., Yan, L., Li, R., Lian, Y., Huang, J., Li, J., Xu, L., Tang, F., Xie, X.S., et al. (2013). Genome analyses of single human oocytes. Cell 155, 1492–1506.

Hou, Y., Guo, H., Cao, C., Li, X., Hu, B., Zhu, P., Wu, X., Wen, L., Tang, F., Huang, Y., et al. (2016). Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. Cell Res 26, 304–319.

Hu, H., Li, Z., Li, X., Yu, M., and Pan, X. (2022a). ScCAEs: deep clustering of single-cell RNA-seq via convolutional autoencoder embedding and soft K-means. Brief Bioinform 23, bbab321.

Hu, J., Li, X., Coleman, K., Schroeder, A., Ma, N., Irwin, D.J., Lee, E.B., Shinohara, R. T., and Li, M. (2021a). SpaGCN: integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. Nat Methods 18, 1342–1351.

Hu, J., Li, X., Hu, G., Lyu, Y., Susztak, K., and Li, M. (2020a). Iterative transfer learning with neural network for clustering and cell type classification in single-cell RNA-seq analysis. Nat Mach Intell 2, 607–618.

Hu, J., Zhong, Y., and Shang, X. (2022b). A versatile and scalable single-cell data integration algorithm based on domain-adversarial and variational approximation. Brief Bioinform 23, bbab400.

Hu, P., Zhang, W., Xin, H., and Deng, G. (2016a). Single cell isolation and analysis. Front Cell Dev Biol 4, 116.

Hu, Y., Huang, K., An, Q., Du, G., Hu, G., Xue, J., Zhu, X., Wang, C.Y., Xue, Z., and Fan, G. (2016b). Simultaneous profiling of transcriptome and DNA methylome from a single cell. Genome Biol 17, 88.

Hu, Y., Peng, T., Gao, L., and Tan, K. (2021b). CytoTalk: *de novo* construction of signal transduction networks using single-cell transcriptomic data. Sci Adv 7, eabf1356.

Hu, Z., Artibani, M., Alsaadi, A., Wietek, N., Morotti, M., Shi, T., Zhong, Z., Santana Gonzalez, L., El-Sahhar, S., Carrami, E.M., et al. (2020b). The repertoire of serous ovarian cancer non-genetic heterogeneity revealed by single-cell sequencing of normal fallopian tube epithelial cells. Cancer Cell 37, 226–242.e7.

Hua, J., Liu, H., Zhang, B., and Jin, S. (2020). LAK: lasso and K-means based single-cell RNA-seq data clustering analysis. IEEE Access 8, 129679–129688.

Huang, G., Li, G., and Cooks, R.G. (2011). Induced nanoelectrospray ionization for matrix-tolerant and high-throughput mass spectrometry. Angew Chem Int Ed 50, 9907–9910.

Huang, J., Yan, L., Fan, W., Zhao, N., Zhang, Y., Tang, F., Xie, X.S., and Qiao, J. (2014). Validation of multiple annealing and looping-based amplification cycle sequencing for 24-chromosome aneuploidy screening of cleavage-stage embryos. Fertil Steril 102, 1685–1691.

Huang, M., Wang, J., Torre, E., Dueck, H., Shaffer, S., Bonasio, R., Murray, J.I., Raj, A., Li, M., and Zhang, N.R. (2018a). SAVER: gene expression recovery for single-cell RNA sequencing. Nat Methods 15, 539–542.

Huang, Q., Mao, S., Khan, M., Zhou, L., and Lin, J.M. (2018b). Dean flow assisted cell ordering system for lipid profiling in single-cells using mass spectrometry. Chem Commun 54, 2595–2598.

Huang, Y., and Sanguinetti, G. (2021). BRIE2: computational identification of splicing phenotypes from single-cell transcriptomic experiments. Genome Biol 22, 251.

Huang, Y., and Zhang, P. (2021). Evaluation of machine learning approaches for cell-type identification from single-cell transcriptomics data. Brief Bioinform 22, bbab035.

Hughes, A.E., Magrini, V., Demeter, R., Miller, C.A., Fulton, R., Fulton, L.L., Eades, W. C., Elliott, K., Heath, S., Westervelt, P., et al. (2014a). Clonal architecture of secondary acute myeloid leukemia defined by single-cell sequencing. PLoS Genet 10, e1004462.

Hughes, A.J., Spelke, D.P., Xu, Z., Kang, C.C., Schaffer, D.V., and Herr, A.E. (2014b). Single-cell western blotting. Nat Methods 11, 749–755.

Hunter, M.V., Moncada, R., Weiss, J.M., Yanai, I., and White, R.M. (2021). Spatially resolved transcriptomics reveals the architecture of the tumor-microenvironment interface. Nat Commun 12, 6278.

Hunter, R.L., Actor, J.K., Hwang, S.A., Karev, V., and Jagannath, C. (2014). Pathogenesis of post primary tuberculosis: immunity and hypersensitivity in the development of cavities. Ann Clin Lab Sci 44, 365–387.

Hutter, C., and Zenklusen, J.C. (2018). The Cancer Genome Atlas: creating lasting value beyond its data. Cell 173, 283–285.

Hwang, B., Lee, J.H., and Bang, D. (2018). Single-cell RNA sequencing technologies and bioinformatics pipelines. Exp Mol Med 50, 1–14.

Ianevski, A., Giri, A.K., and Aittokallio, T. (2022). Fully-automated and ultra-fast cell-type identification using specific marker combinations from single-cell transcriptomic data. Nat Commun 13, 1246.

Ibáñez, A.J., Fagerer, S.R., Schmidt, A.M., Urban, P.L., Jefimovs, K., Geiger, P., Dechant, R., Heinemann, M., and Zenobi, R. (2013). Mass spectrometry-based metabolomics of single yeast cells. Proc Natl Acad Sci USA 110, 8790–8794.

Ibáñez, A.J., and Svatos, A. (2020). Applications of microarrays for mass spectrometry (MAMS) in single-cell metabolomics. In: Shrestha, B., ed. Single Cell Metabolism. Methods in Molecular Biology. New York: Humana. 73–88.

Ilicic, T., Kim, J.K., Kolodziejczyk, A.A., Bagger, F.O., McCarthy, D.J., Marioni, J.C., and Teichmann, S.A. (2016). Classification of low quality cells from single-cell RNA-seq data. Genome Biol 17, 29.

Isakova, A., Neff, N., and Quake, S.R. (2021). Single-cell quantification of a broad RNA spectrum reveals unique noncoding patterns associated with cell types and states. Proc Natl Acad Sci USA 118, e2113568118.

Islam, S., Kjällquist, U., Moliner, A., Zajac, P., Fan, J.B., Lönnerberg, P., and Linnarsson, S. (2011). Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. Genome Res 21, 1160–1167.

Islam, S., Kjällquist, U., Moliner, A., Zajac, P., Fan, J.B., Lönnerberg, P., and Linnarsson, S. (2012). Highly multiplexed and strand-specific single-cell RNA 5′ end sequencing. Nat Protoc 7, 813–828.

Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., Lönnerberg, P., and Linnarsson, S. (2014). Quantitative single-cell RNA-seq with unique molecular identifiers. Nat Methods 11, 163–166.

Jackson, C.A., and Vogel, C. (2022). New horizons in the stormy sea of multimodal single-cell data integration. Mol Cell 82, 248–259.

Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Elefant, N., Paul, F., Zaretsky, I., Mildner, A., Cohen, N., Jung, S., Tanay, A., et al. (2014). Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. Science 343, 776–779.

Jaitin, D.A., Weiner, A., Yofe, I., Lara-Astiaso, D., Keren-Shaul, H., David, E., Salame, T.M., Tanay, A., van Oudenaarden, A., and Amit, I. (2016). Dissecting immune circuits by linking CRISPR-pooled screens with single-cell RNA-seq. Cell 167, 1883–1896.e15.

Jerby-Arnon, L., Shah, P., Cuoco, M.S., Rodman, C., Su, M.J., Melms, J.C., Leeson, R., Kanodia, A., Mei, S., Lin, J.R., et al. (2018). A cancer cell program promotes T cell exclusion and resistance to checkpoint blockade. Cell 175, 984–997.e24.

Ji, A.L., Rubin, A.J., Thrane, K., Jiang, S., Reynolds, D.L., Meyers, R.M., Guo, M.G., George, B.M., Mollbrink, A., Bergenstråhle, J., et al. (2020a). Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma. Cell 182, 497–514.e22.

Ji, Z., and Ji, H. (2016). TSCAN: pseudo-time reconstruction and evaluation in single-cell RNA-seq analysis. Nucleic Acids Res 44, e117.

Ji, Z., Zhou, W., Hou, W., and Ji, H. (2020b). Single-cell ATAC-seq signal extraction and enhancement with SCATE. Genome Biol 21, 161.

Ji, Z., Zhou, W., and Ji, H. (2017). Single-cell regulome data analysis by SCRAT. Bioinformatics 33, 2930–2932.

Jia, C., Hu, Y., Kelly, D., Kim, J., Li, M., and Zhang, N.R. (2017). Accounting for technical noise in differential expression analysis of single-cell RNA sequencing data. Nucleic Acids Res 45, 10978–10988.

Jiang, H., Goulbourne, C.N., Tatar, A., Turlo, K., Wu, D., Beigneux, A.P., Grovenor, C. R.M., Fong, L.G., and Young, S.G. (2014). High-resolution imaging of dietary lipids in cells and tissues by NanoSIMS analysis. J Lipid Res 55, 2156–2166.

Jiang, L., Chen, H., Pinello, L., and Yuan, G.C. (2016). GiniClust: detecting rare cell types from single-cell gene expression data with Gini index. Genome Biol 17, 144.

Jiang, R., Sun, T., Song, D., and Li, J.J. (2022a). Statistics or biology: the zero-inflation controversy about scRNA-seq data. Genome Biol 23, 31.

Jiang, Y., Gao, X., Liu, Y., Yan, X., Shi, H., Zhao, R., Chen, Z.J., Gao, F., Zhao, H., and Zhao, S. (2024). Cellular atlases of ovarian microenvironment alterations by diet and genetically-induced obesity. Sci China Life Sci 67, 51–66.

Jiang, Y., Harigaya, Y., Zhang, Z., Zhang, H., Zang, C., and Zhang, N.R. (2022b). Nonparametric single-cell multiomic characterization of trio relationships between transcription factors, target genes, and cis-regulatory regions. Cell Syst 13, 737–751.e4.

Jin, K., Li, B., Yan, H., and Zhang, X.F. (2022a). Imputing dropouts for single-cell RNA sequencing based on multi-objective optimization. Bioinformatics 38, 3222–3230.

Jin, S., Guerrero-Juarez, C.F., Zhang, L., Chang, I., Ramos, R., Kuan, C.H., Myung, P., Plikus, M.V., and Nie, Q. (2021). Inference and analysis of cell-cell communication using CellChat. Nat Commun 12, 1088.

Jin, S., Zhang, L., and Nie, Q. (2020a). scAI: an unsupervised approach for the integrative analysis of parallel single-cell transcriptomic and epigenomic profiles. Genome Biol 21, 25.

Jin, W., Tang, Q., Wan, M., Cui, K., Zhang, Y., Ren, G., Ni, B., Sklar, J., Przytycka, T. M., Childs, R., et al. (2015). Genome-wide detection of DNase I hypersensitive sites in single cells and FFPE tissue samples. Nature 528, 142–146.

Jin, X., Simmons, S.K., Guo, A., Shetty, A.S., Ko, M., Nguyen, L., Jokhi, V., Robinson, E., Oyler, P., Curry, N., et al. (2020b). *In vivo* Perturb-Seq reveals neuronal and glial abnormalities associated with autism risk genes. Science 370, eaaz6063.

Jin, Z., Zhang, X., Dai, X., Huang, J., Hu, X., Zhang, J., and Shi, L. (2022b). InterCellDB: a user-defined database for inferring intercellular networks. Adv Sci 9, e2200045.

Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. (2007). Genome-wide mapping of *in vivo* protein-DNA interactions. Science 316, 1497–1502.

Jones, A., Townes, F.W., Li, D., and Engelhardt, B.E. (2022a). Alignment of spatial genomics and histology data using deep Gaussian processes. bioRxiv, doi: 10.1101/2022.01.10.475692.

Jones, R.C., Karkanias, J., Krasnow, M.A., Pisco, A.O., Quake, S.R., Salzman, J., Yosef, N., Bulthaup, B., Brown, P., Harper, W., et al. (2022b). The Tabula Sapiens: a multiple-organ, single-cell transcriptomic atlas of humans. Science 376, eabl4896.

Junker, J.P., Noël, E.S., Guryev, V., Peterson, K.A., Shah, G., Huisken, J., McMahon, A. P., Berezikov, E., Bakkers, J., and van Oudenaarden, A. (2014). Genome-wide RNA tomography in the zebrafish embryo. Cell 159, 662–675.

Junttila, M.R., and de Sauvage, F.J. (2013). Influence of tumour micro-environment heterogeneity on therapeutic response. Nature 501, 346–354.

Kaminow, B., Yunusov, D., and Dobin, A. (2021). STARsolo: accurate, fast and versatile mapping/quantification of single-cell and single-nucleus RNA-seq data. bioRxiv, doi: 10.1101/2021.05.05.442755.

Kang, H.M., Subramaniam, M., Targ, S., Nguyen, M., Maliskova, L., McCarthy, E., Wan, E., Wong, S., Byrnes, L., Lanata, C.M., et al. (2018). Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. Nat Biotechnol 36, 89–94.

Kang, J.B., Nathan, A., Weinand, K., Zhang, F., Millard, N., Rumker, L., Moody, D.B., Korsunsky, I., and Raychaudhuri, S. (2021). Efficient and precise single-cell reference atlas mapping with symphony. Nat Commun 12, 5890.

Kantlehner, M., Kirchner, R., Hartmann, P., Ellwart, J.W., Alunni-Fabbroni, M., and Schumacher, A. (2011). A high-throughput DNA methylation analysis of a single cell. Nucleic Acids Res 39, e44.

Kapourani, C.A., Argelaguet, R., Sanguinetti, G., and Vallejos, C.A. (2021). scMET: Bayesian modeling of DNA methylation heterogeneity at single-cell resolution. Genome Biol 22, 114.

Kapourani, C.A., and Sanguinetti, G. (2019). Melissa: Bayesian clustering and imputation of single-cell methylomes. Genome Biol 20, 61.

Karaiskos, N., Wahle, P., Alles, J., Boltengagen, A., Ayoub, S., Kipar, C., Kocks, C., Rajewsky, N., and Zinzen, R.P. (2017). The *Drosophila* embryo at single-cell transcriptome resolution. Science 358, 194–199.

Karas, M., Bahr, U., and Dülcks, T. (2000). Nano-electrospray ionization mass spectrometry: addressing analytical problems beyond routine. Fresenius J Anal Chem 366, 669–676.

Karayel, O., Virreira Winter, S., Padmanabhan, S., Kuras, Y.I., Vu, D.T., Tuncali, I., Merchant, K., Wills, A.M., Scherzer, C.R., and Mann, M. (2022). Proteome profiling of cerebrospinal fluid reveals biomarker candidates for Parkinson's disease. Cell Rep Med 3, 100661.

Kats, I., Vento-Tormo, R., and Stegle, O. (2021). SpatialDE2: fast and localized variance component analysis of spatial transcriptomics. bioRxiv, doi: 10.1101/2021.10.27.466045.

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

Kaymaz, Y., Ganglberger, F., Tang, M., Haslinger, C., Fernandez-Albert, F., Lawless, N., and Sackton, T.B. (2021). HieRFIT: a hierarchical cell type classification tool for projections from complex single-cell atlas datasets. Bioinformatics 37, 4431–4436.

Ke, R., Mignardi, M., Pacureanu, A., Svedlund, J., Botling, J., Wählby, C., and Nilsson, M. (2013). *In situ* sequencing for RNA analysis in preserved tissue and cells. Nat Methods 10, 857–860.

Kebschull, J.M., Richman, E.B., Ringach, N., Friedmann, D., Albarran, E., Kolluru, S. S., Jones, R.C., Allen, W.E., Wang, Y., Cho, S.W., et al. (2020). Cerebellar nuclei evolved by repeatedly duplicating a conserved cell-type set. Science 370, eabd5059.

Kelly, R.T. (2020). Single-cell proteomics: progress and prospects. Mol Cell Proteomics 19, 1739–1748.

Kelly, T.K., Liu, Y., Lay, F.D., Liang, G., Berman, B.P., and Jones, P.A. (2012). Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. Genome Res 22, 2497–2506.

Keren-Shaul, H., Kenigsberg, E., Jaitin, D.A., David, E., Paul, F., Tanay, A., and Amit, I. (2019). MARS-seq2.0: an experimental and analytical pipeline for indexed sorting combined with single-cell RNA sequencing. Nat Protoc 14, 1841–1862.

Khan, S.A., Lehmann, R., Martinez-de-Morentin, X., Ruiz, A.M., Lagani, V., Kiani, N. A., Gomez-Cabrero, D., and Tegner, J. (2022). scAEGAN: unification of single-cell genomics data by adversarial learning of latent space correspondences. bioRxiv, doi: 10.1101/2022.04.19.488745.

Kharchenko, P.V., Silberstein, L., and Scadden, D.T. (2014). Bayesian approach to single-cell differential expression analysis. Nat Methods 11, 740–742.

Khatri, P., Sirota, M., and Butte, A.J. (2012). Ten years of pathway analysis: current approaches and outstanding challenges. PLoS Comput Biol 8, e1002375.

Kim, C., Lee, H., Jeong, J., Jung, K., and Han, B. (2022). MarcoPolo: a method to discover differentially expressed genes in single-cell RNA-seq data without depending on prior clustering. Nucleic Acids Res 50, e71.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat Methods 12, 357–360.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14, R36.

Kim, H.J., Lin, Y., Geddes, T.A., Yang, J.Y.H., and Yang, P. (2020a). CiteFuse enables multi-modal analysis of CITE-seq data. Bioinformatics 36, 4137–4143.

Kim, T.H., and Ren, B. (2006). Genome-wide analysis of protein-DNA interactions. Annu Rev Genom Hum Genet 7, 81–102.

Kim, T.H., Zhou, X., and Chen, M. (2020b). Demystifying "drop-outs" in single-cell UMI data. Genome Biol 21, 196.

Kind, J., Pagie, L., de Vries, S.S., Nahidiazar, L., Dey, S.S., Bienko, M., Zhan, Y., Lajoie, B., de Graaf, C.A., Amendola, M., et al. (2015). Genome-wide maps of nuclear lamina interactions in single human cells. Cell 163, 134–147.

Kiselev, V.Y., Kirschner, K., Schaub, M.T., Andrews, T., Yiu, A., Chandra, T., Natarajan, K.N., Reik, W., Barahona, M., Green, A.R., et al. (2017). SC3: consensus clustering of single-cell RNA-seq data. Nat Methods 14, 483–486.

Kiselev, V.Y., Yiu, A., and Hemberg, M. (2018). scmap: projection of single-cell RNA-seq data across data sets. Nat Methods 15, 359–362.

Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell 161, 1187–1201.

Klein, C.A., Schmidt-Kittler, O., Schardt, J.A., Pantel, K., Speicher, M.R., and Riethmüller, G. (1999). Comparative genomic hybridization, loss of heterozygosity, and DNA sequence analysis of single cells. Proc Natl Acad Sci USA 96, 4494–4499.

Kleshchevnikov, V., Shmatko, A., Dann, E., Aivazidis, A., King, H.W., Li, T., Elmentaite, R., Lomakin, A., Kedlian, V., Gayoso, A., et al. (2022). Cell2location maps fine-grained cell types in spatial transcriptomics. Nat Biotechnol 40, 661–671.

Knight, P., Gauthier, M.L., Pardo, C.E., Darst, R.P., Kapadia, K., Browder, H., Morton, E., Riva, A., Kladde, M.P., and Bacher, R. (2021). Methylscaper: an R/Shiny app for joint visualization of DNA methylation and nucleosome occupancy in single-molecule and single-cell data. Bioinformatics 37, 4857–4859.

Knouse, K.A., Wu, J., Whittaker, C.A., and Amon, A. (2014). Single cell sequencing reveals low levels of aneuploidy across mammalian tissues. Proc Natl Acad Sci USA 111, 13409–13414.

Kobak, D., and Linderman, G.C. (2021). Initialization is critical for preserving global data structure in both t-SNE and UMAP. Nat Biotechnol 39, 156–157.

Kobayashi, H., Koike, T., Sakashita, A., Tanaka, K., Kumamoto, S., and Kono, T. (2016). Repetitive DNA methylome analysis by small-scale and single-cell shotgun bisulfite sequencing. Genes Cells 21, 1209–1222.

Koenig, A.L., Shchukina, I., Amrute, J., Andhey, P.S., Zaitsev, K., Lai, L., Bajpai, G., Bredemeyer, A., Smith, G., Jones, C., et al. (2022). Single-cell transcriptomics reveals cell-type-specific diversification in human heart failure. Nat Cardiovasc Res 1, 263–280.

Koike-Yusa, H., Li, Y., Tan, E.P., Velasco-Herrera, M.D.C., and Yusa, K. (2014). Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. Nat Biotechnol 32, 267–273.

Kolodziejczyk, A.A., Kim, J.K., Svensson, V., Marioni, J.C., and Teichmann, S.A. (2015). The technology and biology of single-cell RNA sequencing. Mol Cell 58, 610–620.

Kompauer, M., Heiles, S., and Spengler, B. (2017). Atmospheric pressure MALDI mass spectrometry imaging of tissues and cells at 1.4-μm lateral resolution. Nat Methods 14, 90–96.

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

Korthauer, K.D., Chu, L.F., Newton, M.A., Li, Y., Thomson, J., Stewart, R., and Kendziorski, C. (2016). A statistical approach for identifying differential distributions in single-cell RNA-seq experiments. Genome Biol 17, 222.

Kriebel, A.R., and Welch, J.D. (2022). UINMF performs mosaic integration of single-cell multi-omic datasets using nonnegative matrix factorization. Nat Commun 13, 780.

Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics 27, 1571–1572.

Ku, W.L., Nakamura, K., Gao, W., Cui, K., Hu, G., Tang, Q., Ni, B., and Zhao, K. (2019). Single-cell chromatin immunocleavage sequencing (scChIC-seq) to profile histone modification. Nat Methods 16, 323–325.

Ku, W.L., Pan, L., Cao, Y., Gao, W., and Zhao, K. (2021). Profiling single-cell histone modifications using indexing chromatin immunocleavage sequencing. Genome Res 31, 1831–1842.

Kuchroo, M., Miyagishima, D.F., Steach, H.R., Godavarthi, A., Takeo, Y., Duy, P.Q., Barak, T., Erson-Omay, E.Z., Youlten, S., Mishra-Gorur, K., et al. (2022). spARC recovers human glioma spatial signaling networks with graph filtering. bioRxiv, doi: 10.1101/2022.08.24.505139.

Kueckelhaus, J., von Ehr, J., Ravi, V.M., Will, P., Joseph, K., Beck, J., Hofmann, U.G., Delev, D., Schnell, O., and Heiland, D.H. (2020). Inferring spatially transient gene expression pattern from spatial transcriptomic studies. bioRxiv, doi: 10.1101/2020.10.20.346544.

Kumar, A., Ryan, A., Kitzman, J.O., Wemmer, N., Snyder, M.W., Sigurjonsson, S., Lee, C., Banjevic, M., Zarutskie, P.W., Lewis, A.P., et al. (2015). Whole genome prediction for preimplantation genetic diagnosis. Genome Med 7, 35.

Kumar, V.S. (2021). Seq-Well: seeking a simpler way to profile RNA from single cells. Clin Chem 67, 454–456.

Kuppe, C., Ramirez Flores, R.O., Li, Z., Hayat, S., Levinson, R.T., Liao, X., Hannani, M. T., Tanevski, J., Wünnemann, F., Nagai, J.S., et al. (2022). Spatial multi-omic map of human myocardial infarction. Nature 608, 766–777.

Kurtenbach, S., Dollar, J.J., Cruz, A.M., Durante, M.A., Decatur, C.L., and Harbour, J. W. (2021). PieParty: visualizing cells from scRNA-seq data as pie charts. Life Sci Alliance 4, e202000986.

La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M.E., Lönnerberg, P., Furlan, A., et al. (2018). RNA velocity of single cells. Nature 560, 494–498.

Labib, M., and Kelley, S.O. (2020). Single-cell analysis targeting the proteome. Nat Rev Chem 4, 143–158.

Lai, B., Gao, W., Cui, K., Xie, W., Tang, Q., Jin, W., Hu, G., Ni, B., and Zhao, K. (2018). Principles of nucleosome organization revealed by single-cell micrococcal nuclease sequencing. Nature 562, 281–285.

Lake, B.B., Chen, S., Hoshi, M., Plongthongkum, N., Salamon, D., Knoten, A., Vijayan, A., Venkatesh, R., Kim, E.H., Gao, D., et al. (2019). A single-nucleus RNA-sequencing pipeline to decipher the molecular anatomy and pathophysiology of human kidneys. Nat Commun 10, 2832.

Lake, B.B., Chen, S., Sos, B.C., Fan, J., Kaeser, G.E., Yung, Y.C., Duong, T.E., Gao, D., Chun, J., Kharchenko, P.V., et al. (2018). Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. Nat Biotechnol 36, 70–80.

Laks, E., McPherson, A., Zahn, H., Lai, D., Steif, A., Brimhall, J., Biele, J., Wang, B., Masud, T., Ting, J., et al. (2019). Clonal decomposition and DNA replication states defined by scaled single-cell genome sequencing. Cell 179, 1207–1221.e22.

Lal, A., Chiang, Z.D., Yakovenko, N., Duarte, F.M., Israeli, J., and Buenrostro, J.D. (2021). Deep learning-based enhancement of epigenomics data with AtacWorks. Nat Commun 12, 1507.

Lall, S., Ray, S., and Bandyopadhyay, S. (2021). RgCop-A regularized copula based method for gene selection in single-cell RNA-seq data. PLoS Comput Biol 17, e1009464.

Lan, F., Demaree, B., Ahmed, N., and Abate, A.R. (2017). Single-cell genome sequencing at ultra-high-throughput with microfluidic droplet barcoding. Nat Biotechnol 35, 640–646.

Lance, C., Luecken, M.D., Burkhardt, D.B., Cannoodt, R., Rautenstrauch, P., Laddach, A., Ubingazhibov, A., Cao, Z.J., Deng, K., Khan, S., et al. (2022). Multimodal single cell data integration challenge: results and lessons learned. bioRxiv, doi: 10.1101/2022.04.11.487796.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods 9, 357–359.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10, R25.

Langmead, B., Wilks, C., Antonescu, V., and Charles, R. (2019). Scaling read aligners to hundreds of threads on general-purpose processors. Bioinformatics 35, 421–432.

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. (2019). Droplet-based combinatorial indexing for massive-scale single-cell chromatin accessibility. Nat Biotechnol 37, 916–924.

Laurens, V.D.M., and Hinton, G. (2008). Visualizing data using t-SNE. J Machine Learn Res 9, 2579–2605.

Lebrigand, K., Magnone, V., Barbry, P., and Waldmann, R. (2020). High throughput error corrected Nanopore single cell transcriptome sequencing. Nat Commun 11, 4025.

Leduc, A., Huffman, R.G., Cantlon, J., Khan, S., and Slavov, N. (2022). Exploring functional protein covariation across single cells using nPOP. Genome Biol 23, 261.

Lee, B., Namkoong, H., Yang, Y., Huang, H., Heller, D., Szot, G.L., Davis, M.M., Husain, S.Z., Pandol, S.J., Bellin, M.D., et al. (2022). Single-cell sequencing unveils distinct immune microenvironments with CCR6-CCL20 crosstalk in human chronic pancreatitis. Gut 71, 1831–1842.

Lee, J.H., Daugharthy, E.R., Scheiman, J., Kalhor, R., Ferrante, T.C., Terry, R., Turczyk, B.M., Yang, J.L., Lee, H.S., Aach, J., et al. (2015). Fluorescent *in situ* sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. Nat Protoc 10, 442–458.

Legetth, O., Rodhe, J., Lang, S., Dhapola, P., Wallergård, M., and Soneji, S. (2021). CellexalVR: a virtual reality platform to visualize and analyze single-cell omics data. iScience 24, 103251.

Lescroart, F., Wang, X., Lin, X., Swedlund, B., Gargouri, S., Sànchez-Dànes, A., Moignard, V., Dubois, C., Paulissen, C., Kinston, S., et al. (2018). Defining the earliest step of cardiovascular lineage segregation by single-cell RNA-seq. Science 359, 1177–1181.

Leung, K., Zahn, H., Leaver, T., Konwar, K.M., Hanson, N.W., Pagé, A.P., Lo, C.C., Chain, P.S., Hallam, S.J., and Hansen, C.L. (2012). A programmable droplet-based microfluidic device applied to multiparameter analysis of single microbes and microbial communities. Proc Natl Acad Sci USA 109, 7665–7670.

Leung, M.L., Davis, A., Gao, R., Casasent, A., Wang, Y., Sei, E., Vilar, E., Maru, D., Kopetz, S., and Navin, N.E. (2017). Single-cell DNA sequencing reveals a late-dissemination model in metastatic colorectal cancer. Genome Res 27, 1287–1299.

Levy, E., and Slavov, N. (2018). Single cell protein analysis for systems biology. Essays Biochem 62, 595–605.

Li, B., Li, Y., Li, K., Zhu, L., Yu, Q., Cai, P., Fang, J., Zhang, W., Du, P., Jiang, C., et al. (2020a). APEC: an accesson-based method for single-cell chromatin accessibility analysis. Genome Biol 21, 116.

Li, C., Fleck, J.S., Martins-Costa, C., Burkard, T.R., Themann, J., Stuempflen, M., Peer, A.M., Vertesy, Á., Littleboy, J.B., Esk, C., et al. (2023). Single-cell brain organoid screening identifies developmental defects in autism. Nature 621, 373–380.

Li, C., Liu, B., Kang, B., Liu, Z., Liu, Y., Chen, C., Ren, X., and Zhang, Z. (2020b). SciBet as a portable and fast single cell type identifier. Nat Commun 11, 1818.

Li, D., Velazquez, J.J., Ding, J., Hislop, J., Ebrahimkhani, M.R., and Bar-Joseph, Z. (2022a). TraSig: inferring cell-cell interactions from pseudotime ordering of scRNA-Seq data. Genome Biol 23, 73.

Li, G., Fu, S., Wang, S., Zhu, C., Duan, B., Tang, C., Chen, X., Chuai, G., Wang, P., and Liu, Q. (2022b). A deep generative model for multi-view profiling of single-cell RNA-seq and ATAC-seq data. Genome Biol 23, 20.

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34, 3094–3100.

Li, H., Brouwer, C.R., and Luo, W. (2022c). A universal deep neural network for in-depth cleaning of single-cell RNA-Seq data. Nat Commun 13, 1901.

Li, H., Sun, Y., Hong, H., Huang, X., Tao, H., Huang, Q., Wang, L., Xu, K., Gan, J., Chen, H., et al. (2022d). Inferring transcription factor regulatory networks from single-cell ATAC-seq data based on graph neural networks. Nat Mach Intell 4, 389–400.

Li, H., Courtois, E.T., Sengupta, D., Tan, Y., Chen, K.H., Goh, J.J.L., Kong, S.L., Chua, C., Hon, L.K., Tan, W.S., et al. (2017). Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. Nat Genet 49, 708–718.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760.

Li, H., Ruan, J., and Durbin, R. (2008a). Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res 18, 1851–1858.

Li, H., Zhu, B., Xu, Z., Adams, T., Kaminski, N., and Zhao, H. (2021a). A Markov random field model for network-based differential expression analysis of single-cell RNA-seq data. BMC Bioinformatics 22, 524.

Li, J., Sheng, Q., Shyr, Y., and Liu, Q. (2022e). scMRMA: single cell multiresolution marker-based annotation. Nucleic Acids Res 50, e7.

Li, K., Ouyang, Z., Chen, Y., Gagnon, J., Lin, D., Mingueneau, M., Chen, W., Sexton, D., and Zhang, B. (2022f). Cellxgene VIP unleashes full power of interactive visualization and integrative analysis of scRNA-seq, spatial transcriptomics, and multiome data. bioRxiv, doi: 10.1101/2020.08.28.270652.

Li, L., Guo, F., Gao, Y., Ren, Y., Yuan, P., Yan, L., Li, R., Lian, Y., Li, J., Hu, B., et al. (2018a). Single-cell multi-omics sequencing of human early embryos. Nat Cell Biol 20, 847–858.

Li, L., Tang, H., Xia, R., Dai, H., Liu, R., and Chen, L. (2022g). Intrinsic entropy model for feature selection of scRNA-seq data. J Mol Cell Biol 14, mjac008.

Li, Q., Zhang, M., Xie, Y., and Xiao, G. (2021b). Bayesian modeling of spatial molecular profiling data via gaussian process. Bioinformatics 37, 4129–4136.

Li, R., Li, Y., Kristiansen, K., and Wang, J. (2008b). SOAP: short oligonucleotide alignment program. Bioinformatics 24, 713–714.

Li, R., and Yang, X. (2022). *De novo* reconstruction of cell interaction landscapes from single-cell spatial transcriptome data with DeepLinc. Genome Biol 23, 124.

Li, R., Yu, C., Li, Y., Lam, T.W., Yiu, S.M., Kristiansen, K., and Wang, J. (2009).

SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics 25, 1966–1967.

Li, S., Plouffe, B.D., Belov, A.M., Ray, S., Wang, X., Murthy, S.K., Karger, B.L., and Ivanov, A.R. (2015). An integrated platform for isolation, processing, and mass spectrometry-based proteomic profiling of rare cells in whole blood. Mol Cell Proteomics 14, 1672–1683.

Li, S., Su, K., Zhuang, Z., Qin, Q., Gao, L., Deng, Y., Liu, X., Hou, G., Wang, L., Hao, P., et al. (2022h). A simple, rapid, and practical method for single-cell proteomics based on mass-adaptive coating of synthetic peptides. Sci Bull 67, 581–584.

Li, W., Xu, H., Xiao, T., Cong, L., Love, M.I., Zhang, F., Irizarry, R.A., Liu, J.S., Brown, M., and Liu, X.S. (2014). MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. Genome Biol 15, 554.

Li, W.V., and Li, J.J. (2018). An accurate and robust imputation method scImpute for single-cell RNA-seq data. Nat Commun 9, 997.

Li, X., Lee, L., Abnousi, A., Yu, M., Liu, W., Huang, L., Li, Y., and Hu, M. (2022i). SnapHiC2: a computationally efficient loop caller for single cell Hi-C data. Comput Struct Biotechnol J 20, 2778–2783.

Li, X., Wang, K., Lyu, Y., Pan, H., Zhang, J., Stambolian, D., Susztak, K., Reilly, M.P., Hu, G., and Li, M. (2020c). Deep learning enables accurate clustering with batch effect removal in single-cell RNA-seq analysis. Nat Commun 11, 2338.

Li, X., Zeng, G., Li, A., and Zhang, Z. (2021c). DeTOKI identifies and characterizes the dynamics of chromatin TAD-like domains in a single cell. Genome Biol 22, 217.

Li, Y., Ge, X., Peng, F., Li, W., and Li, J.J. (2022j). Exaggerated false positives by popular differential expression methods when analyzing human population samples. Genome Biol 23, 79.

Li, Y., Hu, X., Lin, R., Zhou, G., Zhao, L., Zhao, D., Zhang, Y., Li, W., Zhang, Y., Ma, P., et al. (2022k). Single-cell landscape reveals active cell subtypes and their interaction in the tumor microenvironment of gastric cancer. Theranostics 12, 3818–3833.

Li, Y., Xu, X., Song, L., Hou, Y., Li, Z., Tsang, S., Li, F., Im, K.M.G., Wu, K., Wu, H., et al. (2012). Single-cell sequencing analysis characterizes common and cell-lineage-specific mutations in a muscle-invasive bladder cancer. Gigascience 1, 12.

Li, Z.Y., Huang, M., Wang, X.K., Zhu, Y., Li, J.S., Wong, C.C.L., and Fang, Q. (2018b). Nanoliter-scale oil-air-droplet chip-based single cell proteomic analysis. Anal Chem 90, 5430–5438.

Li, Z., Cheng, S., Lin, Q., Cao, W., Yang, J., Zhang, M., Shen, A., Zhang, W., Xia, Y., Ma, X., et al. (2021d). Single-cell lipidomics with high structural specificity by mass spectrometry. Nat Commun 12, 2869.

Li, Z., Meisner, J., and Albrechtsen, A. (2022l). PCAone: fast and accurate out-of-core PCA framework for large scale biobank data. bioRxiv, doi: 10.1101/2022.05.25.493261.

Li, Z., Sun, C., Wang, F., Wang, X., Zhu, J., Luo, L., Ding, X., Zhang, Y., Ding, P., Wang, H., et al. (2022m). Molecular mechanisms governing circulating immune cell heterogeneity across different species revealed by single-cell sequencing. Clin Transl Med 12, e689.

Li, Z., and Zhou, X. (2022). BASS: multi-scale and multi-sample analysis enables accurate cell type clustering and spatial domain detection in spatial transcriptomic studies. Genome Biol 23, 168.

Liang, L., Yu, J., Li, J., Li, N., Liu, J., Xiu, L., Zeng, J., Wang, T., and Wu, L. (2021a). Integration of scRNA-seq and bulk RNA-seq to analyse the heterogeneity of ovarian cancer immune cells and establish a molecular risk model. Front Oncol 11, 711020.

Liang, S., Mohanty, V., Dou, J., Miao, Q., Huang, Y., Müftüoğlu, M., Ding, L., Peng, W., and Chen, K. (2021b). Single-cell manifold-preserving feature selection for detecting rare cell populations. Nat Comput Sci 1, 374–384.

Liang, Y., Acor, H., McCown, M.A., Nwosu, A.J., Boekweg, H., Axtell, N.B., Truong, T., Cong, Y., Payne, S.H., and Kelly, R.T. (2021c). Fully automated sample processing and analysis workflow for low-input proteome profiling. Anal Chem 93, 1658–1666.

Liao, Y., Smyth, G.K., and Shi, W. (2013). The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res 41, e108.

Liao, Y., Smyth, G.K., and Shi, W. (2019). The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. Nucleic Acids Res 47, e47.

Lin, C., and Bar-Joseph, Z. (2019). Continuous-state HMMs for modeling time-series single-cell RNA-Seq data. Bioinformatics 35, 4707–4715.

Lin, G.L., and Hankenson, K.D. (2011). Integration of BMP, Wnt, and notch signaling pathways in osteoblast differentiation. J Cell Biochem 112, 3491–3501.

Lin, H., Zhang, Z., Zhang, M.Q., Ma, B., and Li, M. (2008). ZOOM! Zillions of oligos mapped. Bioinformatics 24, 2431–2437.

Lin, L., and Zhang, L. (2022). Joint analysis of scATAC-seq datasets using epiConv. BMC Bioinformatics 23, 309.

Lin, Y., Wu, T.Y., Wan, S., Yang, J.Y.H., Wong, W.H., and Wang, Y.X.R. (2022). scJoint integrates atlas-scale single-cell RNA-seq and ATAC-seq data with transfer learning. Nat Biotechnol 40, 703–710.

Linderman, G.C., Rachh, M., Hoskins, J.G., Steinerberger, S., and Kluger, Y. (2019). Fast interpolation-based t-SNE for improved visualization of single-cell RNA-seq data. Nat Methods 16, 243–245.

Linderman, G.C., Zhao, J., Roulis, M., Bielecki, P., Flavell, R.A., Nadler, B., and Kluger, Y. (2022). Zero-preserving imputation of single-cell RNA-seq data. Nat Commun 13, 192.

Liscovitch-Brauer, N., Montalbano, A., Deng, J., Méndez-Mancilla, A., Wessels, H.H., Moss, N.G., Kung, C.Y., Sookdeo, A., Guo, X., Geller, E., et al. (2021). Profiling the genetic determinants of chromatin accessibility with scalable single-cell CRISPR screens. Nat Biotechnol 39, 1270–1277.

Littman, R., Hemminger, Z., Foreman, R., Arneson, D., Zhang, G., Gómez-Pinilla, F., Yang, X., and Wollman, R. (2021). Joint cell segmentation and cell type annotation for spatial transcriptomics. Mol Syst Biol 17, e10108.

Liu, H., Zhou, J., Tian, W., Luo, C., Bartlett, A., Aldridge, A., Lucero, J., Osteen, J.K., Nery, J.R., Chen, H., et al. (2021). DNA methylation atlas of the mouse brain at single-cell resolution. Nature 598, 120–128.

Liu, J., Lichtenberg, T., Hoadley, K.A., Poisson, L.M., Lazar, A.J., Cherniack, A.D., Kovatich, A.J., Benz, C.C., Levine, D.A., Lee, A.V., et al. (2018a). An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. Cell 173, 400–416.e11.

Liu, L., Chen, D., Wang, J., and Chen, J. (2020a). Advances of single-cell protein analysis. Cells 9, 1271.

Liu, M., Liu, Y., Di, J., Su, Z., Yang, H., Jiang, B., Wang, Z., Zhuang, M., Bai, F., and Su, X. (2017). Multi-region and single-cell sequencing reveal variable genomic heterogeneity in rectal cancer. BMC Cancer 17, 787.

Liu, N., Liu, L., and Pan, X. (2014). Single-cell analysis of the transcriptome and its application in the characterization of stem cells and early embryos. Cell Mol Life Sci 71, 2707–2715.

Liu, R., Pan, N., Zhu, Y., and Yang, Z. (2018b). T-Probe: an integrated microscale device for online *In Situ* single cell analysis and metabolic profiling using mass spectrometry. Anal Chem 90, 11078–11085.

Liu, R., and Yang, Z. (2021). Single cell metabolomics using mass spectrometry: techniques and data analysis. Anal Chim Acta 1143, 124–134.

Liu, T., and Wang, Z. (2022). scHiCEmbed: bin-specific embeddings of single-cell Hi-C data using graph auto-encoders. Genes 13, 1048.

Liu, W., Liao, X., Luo, Z., Yang, Y., Lau, M.C., Jiao, Y., Shi, X., Zhai, W., Ji, H., Yeong, J., et al. (2022). Probabilistic embedding, clustering, and alignment for integrating spatial transcriptomics data with PRECAST. bioRxiv, doi: 10.1101/2022.06.26.497672.

Liu, Y., Chen, X., Zhang, Y., and Liu, J. (2019). Advancing single-cell proteomics and metabolomics with microfluidic technologies. Analyst 144, 846–858.

Liu, Y., Li, H., Xu, Y., Liu, Y., Peng, X., and Wang, J. (2023). IsoCell: an approach to enhance single cell clustering by integrating isoform-level expression through orthogonal projection. IEEE ACM Trans Comput Biol Bioinf 20, 465–475.

Liu, Y., Yang, M., Deng, Y., Su, G., Enninful, A., Guo, C.C., Tebaldi, T., Zhang, D., Kim, D., Bai, Z., et al. (2020b). High-spatial-resolution multi-omics sequencing via deterministic barcoding in tissue. Cell 183, 1665–1681.e18.

Lodato, M.A., Woodworth, M.B., Lee, S., Evrony, G.D., Mehta, B.K., Karger, A., Lee, S., Chittenden, T.W., D'Gama, A.M., Cai, X., et al. (2015). Somatic mutation in single human neurons tracks developmental and transcriptional history. Science 350, 94–98.

Lohoff, T., Ghazanfar, S., Missarova, A., Koulena, N., Pierson, N., Griffiths, J.A., Bardot, E.S., Eng, C.H.L., Tyser, R.C.V., Argelaguet, R., et al. (2022). Integration of spatial and single-cell transcriptomic data elucidates mouse organogenesis. Nat Biotechnol 40, 74–85.

Lombard-Banek, C., Moody, S.A., Manzini, M.C., and Nemes, P. (2019). Microsampling capillary electrophoresis mass spectrometry enables single-cell proteomics in complex tissues: developing cell clones in live *Xenopus laevis* and zebrafish embryos. Anal Chem 91, 4797–4805.

Lombard-Banek, C., Moody, S.A., and Nemes, P. (2016). Single-cell mass spectrometry for discovery proteomics: quantifying translational cell heterogeneity in the 16-cell frog (*Xenopus*) embryo. Angew Chem Int Ed 55, 2454–2458.

Long, Y., Ang, K.S., Li, M., Chong, K.L.K., Sethi, R., Zhong, C., Xu, H., Ong, Z., Sachaphibulkij, K., Chen, A., et al. (2023). Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with GraphST. Nat Commun 14, 1155.

Lopez-Delisle, L., and Delisle, J.B. (2022). baredSC: Bayesian approach to retrieve expression distribution of single-cell data. BMC Bioinformatics 23, 36.

Lopez, R., Li, B., Keren-Shaul, H., Boyeau, P., Kedmi, M., Pilzer, D., Jelinski, A., Yofe, I., David, E., Wagner, A., et al. (2022). DestVI identifies continuums of cell types in spatial transcriptomics data. Nat Biotechnol 40, 1360–1369.

Lorthongpanich, C., Cheow, L.F., Balu, S., Quake, S.R., Knowles, B.B., Burkholder, W. F., Solter, D., and Messerschmidt, D.M. (2013). Single-cell DNA-methylation analysis reveals epigenetic chimerism in preimplantation embryos. Science 341, 1110–1112.

Lotfollahi, M., Litinetskaya, A., and Theis, F.J. (2022). Multigrate: single-cell multi-omic data integration. bioRxiv, doi: 10.1101/2022.03.16.484643.

Lovatt, D., Ruble, B.K., Lee, J., Dueck, H., Kim, T.K., Fisher, S., Francis, C., Spaethling, J.M., Wolf, J.A., Grady, M.S., et al. (2014). Transcriptome *in vivo* analysis (TIVA) of spatially defined single cells in live tissue. Nat Methods 11, 190–196.

Love, J.C., Ronan, J.L., Grotenbreg, G.M., van der Veen, A.G., and Ploegh, H.L. (2006). A microengraving method for rapid selection of single cells producing antigen-specific antibodies. Nat Biotechnol 24, 703–707.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 550.

Loza, M., Teraguchi, S., Standley, D.M., and Diez, D. (2022). Unbiased integration of single cell transcriptome replicates. NAR Genomics Bioinf 4, lqac022.

Lu, S., Conn, D.J., Chen, S., Johnson, K.D., Bresnick, E.H., and Keleş, S. (2021). MLG: multilayer graph clustering for multi-condition scRNA-seq data. Nucleic Acids Res 49, e127.

Lubeck, E., Coskun, A.F., Zhiyentayev, T., Ahmad, M., and Cai, L. (2014). Single-cell *in situ* RNA profiling by sequential hybridization. Nat Methods 11, 360–361.

Luecken, M.D., Büttner, M., Chaichoompu, K., Danese, A., Interlandi, M., Mueller, M. F., Strobl, D.C., Zappia, L., Dugas, M., Colomé-Tatché, M., et al. (2022). Benchmarking atlas-level data integration in single-cell genomics. Nat Methods 19, 41–50.

Luecken, M.D., and Theis, F.J. (2019). Current best practices in single-cell RNA-seq analysis: a tutorial. Mol Syst Biol 15, e8746.

Lun, A.T.L., Riesenfeld, S., Andrews, T., Dao, T.P., Gomes, T., and Marioni, J.C. (2019). EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. Genome Biol 20, 63.

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. (2017). Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. Science 357, 600–604.

Lynch, A.W., Theodoris, C.V., Long, H.W., Brown, M., Liu, X.S., and Meyer, C.A. (2022). MIRA: joint regulatory modeling of multimodal expression and chromatin accessibility in single cells. Nat Methods 19, 1097–1108.

Ma, A., Wang, X., Wang, C., Li, J., Xiao, T., Wang, J., Li, Y., Liu, Y., Chang, Y., Wang, D., et al. (2021a). DeepMAPS: single-cell biological network inference using heterogeneous graph transformer. bioRxiv, doi: 10.1101/2021.10.31.466658.

Ma, F., and Pellegrini, M. (2020). ACTINN: automated identification of cell types in single cell RNA sequencing. Bioinformatics 36, 533–538.

Ma, Q., Li, S., Zhuang, W., Li, S., Wang, J., and Zeng, D. (2021b). Self-supervised time series clustering with model-based dynamics. IEEE Trans Neural Netw Learn Syst 32, 3942–3955.

Ma, S., de la Fuente Revenga, M., Sun, Z., Sun, C., Murphy, T.W., Xie, H., González-Maeso, J., and Lu, C. (2018). Cell-type-specific brain methylomes profiled via ultralow-input microfluidics. Nat Biomed Eng 2, 183–194.

Ma, S., Zhang, B., LaFave, L.M., Earl, A.S., Chiang, Z., Hu, Y., Ding, J., Brack, A., Kartha, V.K., Tay, T., et al. (2020). Chromatin potential identified by shared single-cell profiling of RNA and chromatin. Cell 183, 1103–1116.e20.

Ma, W., Su, K., and Wu, H. (2021c). Evaluation of some aspects in supervised cell type identification for single-cell RNA-seq: classifier, feature selection, and reference construction. Genome Biol 22, 264.

Ma, Y., and Zhou, X. (2022). Spatially informed cell-type deconvolution for spatial transcriptomics. Nat Biotechnol 40, 1349–1359.

Macaulay, I.C., Haerty, W., Kumar, P., Li, Y.I., Hu, T.X., Teng, M.J., Goolam, M., Saurat, N., Coupland, P., Shirley, L.M., et al. (2015). G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. Nat Methods 12, 519–522.

Macaulay, I.C., and Voet, T. (2014). Single cell genomics: advances and future perspectives. PLoS Genet 10, e1004126.

Macnair, W., and Robinson, M. (2023). SampleQC: robust multivariate, multi-cell type, multi-sample quality control for single-cell data. Genome Biol 24, 23.

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. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell 161, 1202–1214.

Macqueen, J. (1967). Some methods for classification and analysis of multivariate observations. In: Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability. 281–297.

Madissoon, E., Oliver, A.J., Kleshchevnikov, V., Wilbrey-Clark, A., Polanski, K., Orsi, A.R., Mamanova, L., Bolt, L., Richoz, N., Elmentaite, R., et al. (2021). A spatial multi-omics atlas of the human lung reveals a novel immune cell survival niche. bioRxiv, doi: 10.1101/2021.11.26.470108.

Mah, C.K., Ahmed, N., Lam, D., Monell, A., Kern, C., Han, Y., Cesnik, A.J., Lundberg, E., Zhu, Q., Carter, H., et al. (2022). Bento: a toolkit for subcellular analysis of spatial transcriptomics data. bioRxiv, doi: 10.1101/2022.06.10.495510.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science 339, 823–826.

Malta, T.M., Sokolov, A., Gentles, A.J., Burzykowski, T., Poisson, L., Weinstein, J.N., Kamińska, B., Huelsken, J., Omberg, L., Gevaert, O., et al. (2018). Machine learning identifies stemness features associated with oncogenic dedifferentiation. Cell 173, 338–354.e15.

Maniatis, S., Äijö, T., Vickovic, S., Braine, C., Kang, K., Mollbrink, A., Fagegaltier, D., Andrusivová, Ž., Saarenpää, S., Saiz-Castro, G., et al. (2019). Spatiotemporal dynamics of molecular pathology in amyotrophic lateral sclerosis. Science 364, 89–93.

Marco, E., Karp, R.L., Guo, G., Robson, P., Hart, A.H., Trippa, L., and Yuan, G.C. (2014). Bifurcation analysis of single-cell gene expression data reveals epigenetic landscape. Proc Natl Acad Sci USA 111, E5643–5650.

Marinov, G.K., Williams, B.A., McCue, K., Schroth, G.P., Gertz, J., Myers, R.M., and Wold, B.J. (2014). From single-cell to cell-pool transcriptomes: stochasticity in gene expression and RNA splicing. Genome Res 24, 496–510.

Martin, P.C.N., Kim, H., Lövkvist, C., Hong, B., and Won, K.J. (2022). Vesalius: high-resolution *in silico* anatomization of spatial transcriptomic data using image analysis. Mol Syst Biol 18, e11080.

Marx, V. (2019). A dream of single-cell proteomics. Nat Methods 16, 809–812.

Maseda, F., Cang, Z., and Nie, Q. (2021). DEEPsc: a deep learning-based map connecting single-cell transcriptomics and spatial imaging data. Front Genet 12, 636743.

Masuda, T., Inamori, Y., Furukawa, A., Yamahiro, M., Momosaki, K., Chang, C.H., Kobayashi, D., Ohguchi, H., Kawano, Y., Ito, S., et al. (2022). Water droplet-in-oil digestion method for single-cell proteomics. Anal Chem 94, 10329–10336.

McConnell, M.J., Lindberg, M.R., Brennand, K.J., Piper, J.C., Voet, T., Cowing-Zitron, C., Shumilina, S., Lasken, R.S., Vermeesch, J.R., Hall, I.M., et al. (2013). Mosaic copy number variation in human neurons. Science 342, 632–637.

McGinnis, C.S., Murrow, L.M., and Gartner, Z.J. (2019a). DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. Cell Syst 8, 329–337.e4.

McGinnis, C.S., Patterson, D.M., Winkler, J., Conrad, D.N., Hein, M.Y., Srivastava, V., Hu, J.L., Murrow, L.M., Weissman, J.S., Werb, Z., et al. (2019b). MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. Nat Methods 16, 619–626.

McInnes, L., Healy, J., Saul, N., and Großberger, L. (2018). UMAP: uniform manifold approximation and projection. J Open Source Software 3, 861.

McPherson, A., Roth, A., Laks, E., Masud, T., Bashashati, A., Zhang, A.W., Ha, G., Biele, J., Yap, D., Wan, A., et al. (2016). Divergent modes of clonal spread and intraperitoneal mixing in high-grade serous ovarian cancer. Nat Genet 48, 758–767.

Medaglia, C., Giladi, A., Stoler-Barak, L., De Giovanni, M., Salame, T.M., Biram, A., David, E., Li, H., Iannacone, M., Shulman, Z., et al. (2017). Spatial reconstruction of immune niches by combining photoactivatable reporters and scRNA-seq. Science 358, 1622–1626.

Mei, S., Qin, Q., Wu, Q., Sun, H., Zheng, R., Zang, C., Zhu, M., Wu, J., Shi, X., Taing, L., et al. (2017). Cistrome Data Browser: a data portal for ChIP-Seq and chromatin accessibility data in human and mouse. Nucleic Acids Res 45, D658–D662.

Meier, F., Brunner, A.D., Koch, S., Koch, H., Lubeck, M., Krause, M., Goedecke, N., Decker, J., Kosinski, T., Park, M.A., et al. (2018). Online parallel accumulation-serial fragmentation (PASEF) with a novel trapped ion mobility mass spectrometer. Mol Cell Proteomics 17, 2534–2545.

Meir, Z., Mukamel, Z., Chomsky, E., Lifshitz, A., and Tanay, A. (2020). Single-cell analysis of clonal maintenance of transcriptional and epigenetic states in cancer cells. Nat Genet 52, 709–718.

Melsted, P., Booeshaghi, A.S., Liu, L., Gao, F., Lu, L., Min, K.H., da Veiga Beltrame, E., Hjörleifsson, K.E., Gehring, J., and Pachter, L. (2021). Modular, efficient and constant-memory single-cell RNA-seq preprocessing. Nat Biotechnol 39, 813–818.

Melsted, P., Ntranos, V., and Pachter, L. (2019). The barcode, UMI, set format and BUStools. Bioinformatics 35, 4472–4473.

Meng, L., Wang, C., Shi, Y., and Luo, Q. (2021). Si-C is a method for inferring super-resolution intact genome structure from single-cell Hi-C data. Nat Commun 12, 4369.

Merritt, C.R., Ong, G.T., Church, S.E., Barker, K., Danaher, P., Geiss, G., Hoang, M., Jung, J., Liang, Y., McKay-Fleisch, J., et al. (2020). Multiplex digital spatial profiling of proteins and RNA in fixed tissue. Nat Biotechnol 38, 586–599.

Meylan, M., Petitprez, F., Becht, E., Bougoüin, A., Pupier, G., Calvez, A., Giglioli, I., Verkarre, V., Lacroix, G., Verneau, J., et al. (2022). Tertiary lymphoid structures generate and propagate anti-tumor antibody-producing plasma cells in renal cell cancer. Immunity 55, 527–541.e5.

Mezger, A., Klemm, S., Mann, I., Brower, K., Mir, A., Bostick, M., Farmer, A., Fordyce, P., Linnarsson, S., and Greenleaf, W. (2018). High-throughput chromatin accessibility profiling at single-cell resolution. Nat Commun 9, 3647.

Miao, Z., Deng, K., Wang, X., and Zhang, X. (2018). DEsingle for detecting three types of differential expression in single-cell RNA-seq data. Bioinformatics 34, 3223–3224.

Michielsen, L., Reinders, M.J.T., and Mahfouz, A. (2021). Hierarchical progressive learning of cell identities in single-cell data. Nat Commun 12, 2799.

Miller, B.F., Bambah-Mukku, D., Dulac, C., Zhuang, X., and Fan, J. (2021). Characterizing spatial gene expression heterogeneity in spatially resolved single-cell transcriptomic data with nonuniform cellular densities. Genome Res 31, 1843–1855.

Miller, B.F., Huang, F., Atta, L., Sahoo, A., and Fan, J. (2022). Reference-free cell type deconvolution of multi-cellular pixel-resolution spatially resolved transcriptomics data. Nat Commun 13, 2339.

Miltenyi, S., Müller, W., Weichel, W., and Radbruch, A. (1990). High gradient magnetic cell separation with MACS. Cytometry 11, 231–238.

Mimitou, E.P., Cheng, A., Montalbano, A., Hao, S., Stoeckius, M., Legut, M., Roush, T., Herrera, A., Papalexi, E., Ouyang, Z., et al. (2019). Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells. Nat Methods 16, 409–412.

Minakshi, P., Ghosh, M., Kumar, R., Patki, H.S., Saini, H.M., Ranjan, K., Brar, B., and Prasad, G. (2019). Chapter 15—Single-cell metabolomics: technology and applications. Single-Cell Omics. New York: Academic Press. 319–353.

Ming, J., Lin, Z., Zhao, J., Wan, X., Consortium, T.T.M., Ezran, C., Liu, S., Yang, C., and Wu, A.R. (2022). FIRM: flexible integration of single-cell RNA-sequencing data for large-scale multi-tissue cell atlas datasets. Brief Bioinform 23, bbac167.

Minoura, K., Abe, K., Nam, H., Nishikawa, H., and Shimamura, T. (2021). A mixture-of-experts deep generative model for integrated analysis of single-cell multiomics data. Cell Rep Methods 1, 100071.

Minussi, D.C., Nicholson, M.D., Ye, H., Davis, A., Wang, K., Baker, T., Tarabichi, M., Sei, E., Du, H., Rabbani, M., et al. (2021). Breast tumours maintain a reservoir of subclonal diversity during expansion. Nature 592, 302–308.

Misra, B.B. (2020). Open-Source Software Tools, Databases, and Resources for Single-Cell and Single-Cell-Type Metabolomics. In: Shrestha, B., ed. Single Cell Metabolism. Methods in Molecular Biology. New York: Humana. 191–217.

Mizuno, H., Tsuyama, N., Harada, T., and Masujima, T. (2008). Live single-cell video-mass spectrometry for cellular and subcellular molecular detection and cell classification. J Mass Spectrom 43, 1692–1700.

Moehlin, J., Mollet, B., Colombo, B.M., and Mendoza-Parra, M.A. (2021). Inferring biologically relevant molecular tissue substructures by agglomerative clustering of digitized spatial transcriptomes with multilayer. Cell Syst 12, 694–705.e3.

Moffitt, J.R., Bambah-Mukku, D., Eichhorn, S.W., Vaughn, E., Shekhar, K., Perez, J.D., Rubinstein, N.D., Hao, J., Regev, A., Dulac, C., et al. (2018). Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. Science 362, eaau5324.

Moncada, R., Barkley, D., Wagner, F., Chiodin, M., Devlin, J.C., Baron, M., Hajdu, C. H., Simeone, D.M., and Yanai, I. (2020). Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. Nat Biotechnol 38, 333–342.

Mooijman, D., Dey, S.S., Boisset, J.C., Crosetto, N., and van Oudenaarden, A. (2016). Single-cell 5hmC sequencing reveals chromosome-wide cell-to-cell variability and enables lineage reconstruction. Nat Biotechnol 34, 852–856.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5, 621–628.

Moses, L., and Pachter, L. (2022). Museum of spatial transcriptomics. Nat Methods 19, 534–546.

Mu, Q., Chen, Y., and Wang, J. (2019). Deciphering brain complexity using single-cell sequencing. Genomics Proteomics Bioinf 17, 344–366.

Mulqueen, R.M., Pokholok, D., Norberg, S.J., Torkenczy, K.A., Fields, A.J., Sun, D., Sinnamon, J.R., Shendure, J., Trapnell, C., O'Roak, B.J., et al. (2018). Highly scalable generation of DNA methylation profiles in single cells. Nat Biotechnol 36, 428–431.

Mund, A., Brunner, A.D., and Mann, M. (2022a). Unbiased spatial proteomics with single-cell resolution in tissues. Mol Cell 82, 2335–2349.

Mund, A., Coscia, F., Kriston, A., Hollandi, R., Kovács, F., Brunner, A.D., Migh, E., Schweizer, L., Santos, A., Bzorek, M., et al. (2022b). Deep Visual Proteomics defines single-cell identity and heterogeneity. Nat Biotechnol 40, 1231–1240.

Muntel, J., Kirkpatrick, J., Bruderer, R., Huang, T., Vitek, O., Ori, A., and Reiter, L. (2019). Comparison of protein quantification in a complex background by DIA and TMT workflows with fixed instrument time. J Proteome Res 18, 1340–1351.

Muskovic, W., and Powell, J.E. (2021). DropletQC: improved identification of empty droplets and damaged cells in single-cell RNA-seq data. Genome Biol 22, 329.

Muto, Y., Wilson, P.C., Ledru, N., Wu, H., Dimke, H., Waikar, S.S., and Humphreys, B. D. (2021). Single cell transcriptional and chromatin accessibility profiling redefine cellular heterogeneity in the adult human kidney. Nat Commun 12, 2190.

Nagano, T., Lubling, Y., Stevens, T.J., Schoenfelder, S., Yaffe, E., Dean, W., Laue, E.D., Tanay, A., and Fraser, P. (2013). Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. Nature 502, 59–64.

Narayan, A., Berger, B., and Cho, H. (2021). Assessing single-cell transcriptomic variability through density-preserving data visualization. Nat Biotechnol 39, 765–774.

Nault, R., Saha, S., Bhattacharya, S., Dodson, J., Sinha, S., Maiti, T., and Zacharewski, T. (2022). Benchmarking of a Bayesian single cell RNAseq differential gene expression test for dose-response study designs. Nucleic Acids Res 50, e48.

Navin, N., Kendall, J., Troge, J., Andrews, P., Rodgers, L., McIndoo, J., Cook, K., Stepansky, A., Levy, D., Esposito, D., et al. (2011). Tumour evolution inferred by single-cell sequencing. Nature 472, 90–94.

Nehar-Belaid, D., Hong, S., Marches, R., Chen, G., Bolisetty, M., Baisch, J., Walters, L., Punaro, M., Rossi, R.J., Chung, C.H., et al. (2020). Mapping systemic lupus erythematosus heterogeneity at the single-cell level. Nat Immunol 21, 1094–1106.

Nemes, P., and Vertes, A. (2007). Laser ablation electrospray ionization for atmospheric pressure, *in vivo*, and imaging mass spectrometry. Anal Chem 79, 8098–8106.

Nestorowa, S., Hamey, F.K., Pijuan Sala, B., Diamanti, E., Shepherd, M., Laurenti, E., Wilson, N.K., Kent, D.G., and Göttgens, B. (2016). A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. Blood 128, e20–e31.

Network, C.G.A. (2012). Comprehensive molecular portraits of human breast tumours. Nature 490, 61–70.

Nguyen, S.N., Liyu, A.V., Chu, R.K., Anderton, C.R., and Laskin, J. (2017). Constant-distance mode nanospray desorption electrospray ionization mass spectrometry imaging of biological samples with complex topography. Anal Chem 89, 1131–1137.

Ni, X., Zhuo, M., Su, Z., Duan, J., Gao, Y., Wang, Z., Zong, C., Bai, H., Chapman, A.R., Zhao, J., et al. (2013). Reproducible copy number variation patterns among single circulating tumor cells of lung cancer patients. Proc Natl Acad Sci USA 110, 21083–21088.

Ni, Z., Prasad, A., Chen, S., Halberg, R.B., Arkin, L.M., Drolet, B.A., Newton, M.A., and Kendziorski, C. (2022). SpotClean adjusts for spot swapping in spatial transcriptomics data. Nat Commun 13, 2971.

Nicin, L., Abplanalp, W.T., Schänzer, A., Sprengel, A., John, D., Mellentin, H., Tombor, L., Keuper, M., Ullrich, E., Klingel, K., et al. (2021). Single nuclei sequencing reveals novel insights into the regulation of cellular signatures in children with dilated cardiomyopathy. Circulation 143, 1704–1719.

Niebler, S., Müller, A., Hankeln, T., and Schmidt, B. (2020). RainDrop: rapid activation matrix computation for droplet-based single-cell RNA-seq reads. BMC Bioinformatics 21, 274.

Nitzan, M., Karaiskos, N., Friedman, N., and Rajewsky, N. (2019). Gene expression cartography. Nature 576, 132–137.

Noël, F., Massenet-Regad, L., Carmi-Levy, I., Cappuccio, A., Grandclaudon, M., Trichot, C., Kieffer, Y., Mechta-Grigoriou, F., and Soumelis, V. (2021). Dissection of intercellular communication using the transcriptome-based framework ICELLNET. Nat Commun 12, 1089.

Nwosu, A.J., Misal, S.A., Truong, T., Carson, R.H., Webber, K.G.I., Axtell, N.B., Liang, Y., Johnston, S.M., Virgin, K.L., Smith, E.G., et al. (2022). In-depth mass spectrometry-based proteomics of formalin-fixed, paraffin-embedded tissues with a spatial resolution of 50–200 μm. J Proteome Res 21, 2237–2245.

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., and Kanehisa, M. (1999). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 27, 29–34.

Ortiz, C., Navarro, J.F., Jurek, A., Märtin, A., Lundeberg, J., and Meletis, K. (2020). Molecular atlas of the adult mouse brain. Sci Adv 6, eabb3446.

Padron-Manrique, C., Vázquez-Jiménez, A., Esquivel-Hernandez, D.A., Lopez, Y.E.M., Neri-Rosario, D., Sánchez-Castañeda, J.P., Giron-Villalobos, D., and Resendis-Antonio, O. (2022). Diffusion on PCA-UMAP manifold captures a well-balance of local, global, and continuum structure to denoise single-cell RNA sequencing data. bioRxiv, doi: 10.1101/2022.06.09.495525.

Paik, D.T., Tian, L., Williams, I.M., Rhee, S., Zhang, H., Liu, C., Mishra, R., Wu, S.M., Red-Horse, K., and Wu, J.C. (2020). Single-cell RNA sequencing unveils unique transcriptomic signatures of organ-specific endothelial cells. Circulation 142, 1848–1862.

Pan, N., Rao, W., Kothapalli, N.R., Liu, R., Burgett, A.W.G., and Yang, Z. (2014). The single-probe: a miniaturized multifunctional device for single cell mass spectrometry analysis. Anal Chem 86, 9376–9380.

Pan, N., Rao, W., and Yang, Z. (2020). Single-probe mass spectrometry analysis of metabolites in single cells. In: Shrestha, B., ed. Single Cell Metabolism. Methods in Molecular Biology. New York: Humana. 61–71.

Pan, N., Standke, S.J., Kothapalli, N.R., Sun, M., Bensen, R.C., Burgett, A.W.G., and Yang, Z. (2019). Quantification of drug molecules in live single cells using the single-probe mass spectrometry technique. Anal Chem 91, 9018–9024.

Pang, M., Su, K., and Li, M. (2021). Leveraging information in spatial transcriptomics to predict super-resolution gene expression from histology images in tumors. bioRxiv, doi: 10.1101/2021.11.28.470212.

Papalexi, E., Mimitou, E.P., Butler, A.W., Foster, S., Bracken, B., Mauck Iii, W.M., Wessels, H.H., Hao, Y., Yeung, B.Z., Smibert, P., et al. (2021). Characterizing the molecular regulation of inhibitory immune checkpoints with multimodal single-cell screens. Nat Genet 53, 322–331.

Parekh, S., Ziegenhain, C., Vieth, B., Enard, W., and Hellmann, I. (2018). zUMIs—A fast and flexible pipeline to process RNA sequencing data with UMIs. Gigascience 7, giy059.

Park, J., Choi, W., Tiesmeyer, S., Long, B., Borm, L.E., Garren, E., Nguyen, T.N., Tasic, B., Codeluppi, S., Graf, T., et al. (2021). Cell segmentation-free inference of cell types from *in situ* transcriptomics data. Nat Commun 12, 3545.

Passarelli, M.K., Pirkl, A., Moellers, R., Grinfeld, D., Kollmer, F., Havelund, R., Newman, C.F., Marshall, P.S., Arlinghaus, H., Alexander, M.R., et al. (2017). The 3D OrbiSIMS—label-free metabolic imaging with subcellular lateral resolution and high mass-resolving power. Nat Methods 14, 1175–1183.

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods 14, 417–419.

Patro, R., Mount, S.M., and Kingsford, C. (2014). Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. Nat Biotechnol 32, 462–464.

Patty, B.J., and Hainer, S.J. (2021). Transcription factor chromatin profiling genome-wide using uliCUT&RUN in single cells and individual blastocysts. Nat Protoc 16, 2633–2666.

P. E. de Souza, C., Andronescu, M., Masud, T., Kabeer, F., Biele, J., Laks, E., Lai, D., Ye, P., Brimhall, J., Wang, B., et al. (2020). Epiclomal: probabilistic clustering of sparse single-cell DNA methylation data. PLoS Comput Biol 16, e1008270.

Pei, G., Yan, F., Simon, L.M., Dai, Y., Jia, P., and Zhao, Z. (2023). *deCS*: a tool for systematic cell type annotations of single-cell RNA sequencing data among human tissues. Genomics Proteomics Bioinf 21, 370–384.

Peng, G., Suo, S., Cui, G., Yu, F., Wang, R., Chen, J., Chen, S., Liu, Z., Chen, G., Qian, Y., et al. (2019). Molecular architecture of lineage allocation and tissue organization in early mouse embryo. Nature 572, 528–532.

Peng, M., Wamsley, B., Elkins, A.G., Geschwind, D.H., Wei, Y., and Roeder, K. (2021a). Cell type hierarchy reconstruction via reconciliation of multi-resolution cluster tree. Nucleic Acids Res 49, e91.

Peng, T., Chen, G.M., and Tan, K. (2021b). GLUER: integrative analysis of single-cell omics and imaging data by deep neural network. bioRxiv, doi: 10.1101/2021.01.25.427845.

Peres-Neto, P.R., Jackson, D.A., and Somers, K.M. (2005). How many principal components? stopping rules for determining the number of non-trivial axes revisited. Comput Stat Data Anal 49, 974–997.

Peterson, V.M., Zhang, K.X., Kumar, N., Wong, J., Li, L., Wilson, D.C., Moore, R., McClanahan, T.K., Sadekova, S., and Klappenbach, J.A. (2017). Multiplexed quantification of proteins and transcripts in single cells. Nat Biotechnol 35, 936–939.

Petrany, M.J., Swoboda, C.O., Sun, C., Chetal, K., Chen, X., Weirauch, M.T., Salomonis, N., and Millay, D.P. (2020). Single-nucleus RNA-seq identifies transcriptional heterogeneity in multinucleated skeletal myofibers. Nat Commun 11, 6374.

Petukhov, V., Guo, J., Baryawno, N., Severe, N., Scadden, D.T., Samsonova, M.G., and Kharchenko, P.V. (2018). dropEst: pipeline for accurate estimation of molecular counts in droplet-based single-cell RNA-seq experiments. Genome Biol 19, 78.

Petukhov, V., Xu, R.J., Soldatov, R.A., Cadinu, P., Khodosevich, K., Moffitt, J.R., and Kharchenko, P.V. (2022). Cell segmentation in imaging-based spatial transcriptomics. Nat Biotechnol 40, 345–354.

Pham, D., Tan, X., Xu, J., Grice, L.F., Lam, P.Y., Raghubar, A., Vukovic, J., Ruitenberg, M.J., and Nguyen, Q. (2020). stLearn: integrating spatial location, tissue morphology and gene expression to find cell types, cell-cell interactions and spatial trajectories within undissociated tissues. bioRxiv, doi: 10.1101/2020.05.31.125658.

Picelli, S. (2017). Single-cell RNA-sequencing: the future of genome biology is now. RNA Biol 14, 637–650.

Picelli, S., Björklund, Å.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat Methods 10, 1096–1098.

Picelli, S., Faridani, O.R., Björklund, Å.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nat Protoc 9, 171–181.

Picher, Á.J., Budeus, B., Wafzig, O., Krüger, C., García-Gómez, S., Martínez-Jiménez, M.I., Díaz-Talavera, A., Weber, D., Blanco, L., and Schneider, A. (2016). TruePrime is a novel method for whole-genome amplification from single cells based on TthPrimPol. Nat Commun 7, 13296.

Pierce, S.E., Granja, J.M., and Greenleaf, W.J. (2021). High-throughput single-cell chromatin accessibility CRISPR screens enable unbiased identification of regulatory networks in cancer. Nat Commun 12, 2969.

Pieters, R., De Lorenzo, P., Ancliffe, P., Aversa, L.A., Brethon, B., Biondi, A., Campbell, M., Escherich, G., Ferster, A., Gardner, R.A., et al. (2019). Outcome of infants younger than 1 year with acute lymphoblastic leukemia treated with the interfant-06 protocol: results from an international phase III randomized study. J Clin Oncol 37, 2246–2256.

Pliner, H.A., Packer, J.S., McFaline-Figueroa, J.L., Cusanovich, D.A., Daza, R.M., Aghamirzaie, D., Srivatsan, S., Qiu, X., Jackson, D., Minkina, A., et al. (2018). Cicero predicts cis-regulatory DNA interactions from single-cell chromatin accessibility data. Mol Cell 71, 858–871.e8.

Pliner, H.A., Shendure, J., and Trapnell, C. (2019). Supervised classification enables rapid annotation of cell atlases. Nat Methods 16, 983–986.

Polański, K., Young, M.D., Miao, Z., Meyer, K.B., Teichmann, S.A., and Park, J.E. (2020). BBKNN: fast batch alignment of single cell transcriptomes. Bioinformatics 36, 964–965.

Prabhakaran, S. (2022). Sparcle: assigning transcripts to cells in multiplexed images. Bioinform Adv 2, vbac048.

Preissl, S., Fang, R., Huang, H., Zhao, Y., Raviram, R., Gorkin, D.U., Zhang, Y., Sos, B. C., Afzal, V., Dickel, D.E., et al. (2018). Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. Nat Neurosci 21, 432–439.

Prieto-Vila, M., Usuba, W., Takahashi, R., Shimomura, I., Sasaki, H., Ochiya, T., and Yamamoto, Y. (2019). Single-cell analysis reveals a preexisting drug-resistant subpopulation in the luminal breast cancer subtype. Cancer Res 79, 4412–4425.

Przytycki, P.F., and Pollard, K.S. (2022). CellWalkR: an R package for integrating and visualizing single-cell and bulk data to resolve regulatory elements. Bioinformatics 38, 2621–2623.

Qi, J., Sun, H., Zhang, Y., Wang, Z., Xun, Z., Li, Z., Ding, X., Bao, R., Hong, L., Jia, W., et al. (2022). Single-cell and spatial analysis reveal interaction of FAP+ fibroblasts and SPP1+ macrophages in colorectal cancer. Nat Commun 13, 1742.

Qi, R., Wu, J., Guo, F., Xu, L., and Zou, Q. (2021). A spectral clustering with self-weighted multiple kernel learning method for single-cell RNA-seq data. Brief Bioinform 22, bbaa216.

Qian, X., Harris, K.D., Hauling, T., Nicoloutsopoulos, D., Muñoz-Manchado, A.B., Skene, N., Hjerling-Leffler, J., and Nilsson, M. (2020). Probabilistic cell typing enables fine mapping of closely related cell types *in situ*. Nat Methods 17, 101–106.

Qiao, C., and Huang, Y. (2021). Representation learning of RNA velocity reveals robust cell transitions. Proc Natl Acad Sci USA 118, e2105859118.

Qiu, P., Simonds, E.F., Bendall, S.C., Gibbs Jr, K.D., Bruggner, R.V., Linderman, M.D., Sachs, K., Nolan, G.P., and Plevritis, S.K. (2011). Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. Nat Biotechnol 29, 886–891.

Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H.A., and Trapnell, C. (2017). Reversed graph embedding resolves complex single-cell trajectories. Nat Methods 14, 979–982.

Ramani, V., Deng, X., Qiu, R., Gunderson, K.L., Steemers, F.J., Disteche, C.M., Noble, W.S., Duan, Z., and Shendure, J. (2017). Massively multiplex single-cell Hi-C. Nat Methods 14, 263–266.

Ramilowski, J.A., Goldberg, T., Harshbarger, J., Kloppmann, E., Lizio, M., Satagopam, V.P., Itoh, M., Kawaji, H., Carninci, P., Rost, B., et al. (2015). A draft network of ligand-receptor-mediated multicellular signalling in human. Nat Commun 6, 7866.

Ramsköld, 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. (2012). Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nat Biotechnol 30, 777–782.

Ran, D., Zhang, S., Lytal, N., and An, L. (2020). scDoc: correcting drop-out events in single-cell RNA-seq data. Bioinformatics 36, 4233–4239.

Rao, W., Pan, N., and Yang, Z. (2015). High resolution tissue imaging using the single-probe mass spectrometry under ambient conditions. J Am Soc Mass Spectrom 26, 986–993.

Rappez, L., Stadler, M., Triana, S., Gathungu, R.M., Ovchinnikova, K., Phapale, P., Heikenwalder, M., and Alexandrov, T. (2021). SpaceM reveals metabolic states of single cells. Nat Methods 18, 799–805.

Reed, E.R., and Monti, S. (2021). Multi-resolution characterization of molecular taxonomies in bulk and single-cell transcriptomics data. Nucleic Acids Res 49, e98.

Ren, H., Walker, B.L., Cang, Z., and Nie, Q. (2022). Identifying multicellular spatiotemporal organization of cells with SpaceFlow. Nat Commun 13, 4076.

Replogle, J.M., Norman, T.M., Xu, A., Hussmann, J.A., Chen, J., Cogan, J.Z., Meer, E.J., Terry, J.M., Riordan, D.P., Srinivas, N., et al. (2020). Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. Nat Biotechnol 38, 954–961.

Replogle, J.M., Saunders, R.A., Pogson, A.N., Hussmann, J.A., Lenail, A., Guna, A., Mascibroda, L., Wagner, E.J., Adelman, K., Lithwick-Yanai, G., et al. (2022). Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. Cell 185, 2559–2575.e28.

Riba, A., Oravecz, A., Durik, M., Jiménez, S., Alunni, V., Cerciat, M., Jung, M., Keime, C., Keyes, W.M., and Molina, N. (2022). Cell cycle gene regulation dynamics revealed by RNA velocity and deep-learning. Nat Commun 13, 2865.

Richman, L.P., Goyal, Y., Jiang, C.L., and Raj, A. (2023). ClonoCluster: a method for using clonal origin to inform transcriptome clustering. Cell Genomics 3, 100247.

Riebensahm, C., Joosse, S.A., Mohme, M., Hanssen, A., Matschke, J., Goy, Y., Witzel, I., Lamszus, K., Kropidlowski, J., Petersen, C., et al. (2019). Clonality of circulating tumor cells in breast cancer brain metastasis patients. Breast Cancer Res 21, 101.

Risso, D., Perraudeau, F., Gribkova, S., Dudoit, S., and Vert, J.P. (2018). A general and flexible method for signal extraction from single-cell RNA-seq data. Nat Commun 9, 284.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43, e47.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140.

Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol 11, R25.

Rodriguez-Meira, A., Buck, G., Clark, S.A., Povinelli, B.J., Alcolea, V., Louka, E., McGowan, S., Hamblin, A., Sousos, N., Barkas, N., et al. (2019). Unravelling intratumoral heterogeneity through high-sensitivity single-cell mutational analysis and parallel RNA sequencing. Mol Cell 73, 1292–1305.e8.

Rodriques, S.G., Stickels, R.R., Goeva, A., Martin, C.A., Murray, E., Vanderburg, C.R., Welch, J., Chen, L.M., Chen, F., and Macosko, E.Z. (2019). Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. Science 363, 1463–1467.

Roohani, Y., Huang, K., and Leskovec, J. (2023). Predicting transcriptional outcomes of novel multigene perturbations with GEARS. Nat Biotechnol doi: 10.1038/s41587-023-01905-6.

Rooijers, K., Markodimitraki, C.M., Rang, F.J., de Vries, S.S., Chialastri, A., de Luca, K. L., Mooijman, D., Dey, S.S., and Kind, J. (2019). Simultaneous quantification of protein-DNA contacts and transcriptomes in single cells. Nat Biotechnol 37, 766–772.

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

Rosenthal, M., Bryner, D., Huffer, F., Evans, S., Srivastava, A., and Neretti, N. (2019). Bayesian estimation of three-dimensional chromosomal structure from single-cell Hi-C data. J Comput Biol 26, 1191–1202.

Rotem, A., Ram, O., Shoresh, N., Sperling, R.A., Goren, A., Weitz, D.A., and Bernstein, B.E. (2015). Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. Nat Biotechnol 33, 1165–1172.

Roth, T.L., Li, P.J., Blaeschke, F., Nies, J.F., Apathy, R., Mowery, C., Yu, R., Nguyen, M.L.T., Lee, Y., Truong, A., et al. (2020). Pooled knockin targeting for genome engineering of cellular immunotherapies. Cell 181, 728–744.e21.

Rubakhin, S.S., Lanni, E.J., and Sweedler, J.V. (2013). Progress toward single cell metabolomics. Curr Opin Biotechnol 24, 95–104.

Rubin, A.J., Parker, K.R., Satpathy, A.T., Qi, Y., Wu, B., Ong, A.J., Mumbach, M.R., Ji, A.L., Kim, D.S., Cho, S.W., et al. (2019). Coupled single-cell CRISPR screening and epigenomic profiling reveals causal gene regulatory networks. Cell 176, 361–376.e17.

Rubio, C., Rodrigo, L., Mir, P., Mateu, E., Peinado, V., Milán, M., Al-Asmar, N., Campos-Galindo, I., Garcia, S., and Simón, C. (2013). Use of array comparative genomic hybridization (array-CGH) for embryo assessment: clinical results. Fertil Steril 99, 1044–1048.

Ruepp, A., Waegele, B., Lechner, M., Brauner, B., Dunger-Kaltenbach, I., Fobo, G., Frishman, G., Montrone, C., and Mewes, H.W. (2010). CORUM: the comprehensive resource of mammalian protein complexes—2009. Nucleic Acids Res 38, D497–D501.

Santinha, A.J., Klingler, E., Kuhn, M., Farouni, R., Lagler, S., Kalamakis, G., Lischetti, U., Jabaudon, D., and Platt, R.J. (2023). Transcriptional linkage analysis with *in vivo* AAV-Perturb-seq. Nature 622, 367–375.

Sasagawa, Y., Danno, H., Takada, H., Ebisawa, M., Tanaka, K., Hayashi, T., Kurisaki, A., and Nikaido, I. (2018). Quartz-Seq2: a high-throughput single-cell RNA-sequencing method that effectively uses limited sequence reads. Genome Biol 19, 29.

Sasagawa, Y., Nikaido, I., Hayashi, T., Danno, H., Uno, K.D., Imai, T., and Ueda, H.R. (2013). Quartz-Seq: a highly reproducible and sensitive single-cell RNA sequencing method, reveals non-genetic gene-expression heterogeneity. Genome Biol 14, R31.

Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nat Biotechnol 33, 495–502.

Satpathy, A.T., Granja, J.M., Yost, K.E., Qi, Y., Meschi, F., McDermott, G.P., Olsen, B. N., Mumbach, M.R., Pierce, S.E., Corces, M.R., et al. (2019). Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. Nat Biotechnol 37, 925–936.

Savas, P., Virassamy, B., Ye, C., Salim, A., Mintoff, C.P., Caramia, F., Salgado, R., Byrne, D.J., Teo, Z.L., Dushyanthen, S., et al. (2018). Single-cell profiling of breast cancer T cells reveals a tissue-resident memory subset associated with improved prognosis. Nat Med 24, 986–993.

Schatz, M.C. (2009). CloudBurst: highly sensitive read mapping with MapReduce. Bioinformatics 25, 1363–1369.

Schede, H.H., Schneider, C.G., Stergiadou, J., Borm, L.E., Ranjak, A., Yamawaki, T.M., David, F.P.A., Lönnerberg, P., Tosches, M.A., Codeluppi, S., et al. (2021). Spatial tissue profiling by imaging-free molecular tomography. Nat Biotechnol 39, 968–977.

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

Schoof, E.M., Furtwängler, B., Üresin, N., Rapin, N., Savickas, S., Gentil, C., Lechman, E., Keller, U., Dick, J.E., and Porse, B.T. (2021). Quantitative single-cell proteomics as a tool to characterize cellular hierarchies. Nat Commun 12, 3341.

Schubert, S.M., Walter, S.R., Manesse, M., and Walt, D.R. (2016). Protein counting in single cancer cells. Anal Chem 88, 2952–2957.

Scialdone, A., Tanaka, Y., Jawaid, W., Moignard, V., Wilson, N.K., Macaulay, I.C., Marioni, J.C., and Göttgens, B. (2016). Resolving early mesoderm diversification through single-cell expression profiling. Nature 535, 289–293.

See, P., Lum, J., Chen, J., and Ginhoux, F. (2018). A single-cell sequencing guide for immunologists. Front Immunol 9, 2425.

Senabouth, A., Lukowski, S.W., Hernandez, J.A., Andersen, S.B., Mei, X., Nguyen, Q. H., and Powell, J.E. (2019). ascend: R package for analysis of single-cell RNA-seq data. Gigascience 8, giz087.

Sethi, A., Gu, M., Gumusgoz, E., Chan, L., Yan, K.K., Rozowsky, J., Barozzi, I., Afzal, V., Akiyama, J.A., Plajzer-Frick, I., et al. (2020). Supervised enhancer prediction with epigenetic pattern recognition and targeted validation. Nat Methods 17, 807–814.

Setty, M., Tadmor, M.D., Reich-Zeliger, S., Angel, O., Salame, T.M., Kathail, P., Choi, K., Bendall, S., Friedman, N., and Pe'er, D. (2016). Wishbone identifies bifurcating developmental trajectories from single-cell data. Nat Biotechnol 34, 637–645.

Shah, S., Lubeck, E., Zhou, W., and Cai, L. (2016). *In situ* transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus. Neuron 92, 342–357.

Shahryary, Y., Hazarika, R.R., and Johannes, F. (2020). MethylStar: a fast and robust pre-processing pipeline for bulk or single-cell whole-genome bisulfite sequencing data. BMC Genomics 21, 479.

Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelsen, T.S., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G., et al. (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. Science 343, 84–87.

Shang, W., Zhang, Y., Shu, M., Wang, W., Ren, L., Chen, F., Shao, L., Lu, S., Bo, S., Ma, S., et al. (2018). Comprehensive chromosomal and mitochondrial copy number profiling in human IVF embryos. Reprod Biomed Online 36, 67–74.

Shao, X., Li, C., Yang, H., Lu, X., Liao, J., Qian, J., Wang, K., Cheng, J., Yang, P., Chen, H., et al. (2022a). Knowledge-graph-based cell-cell communication inference for spatially resolved transcriptomic data with SpaTalk. Nat Commun 13, 4429.

Shao, X., Liao, J., Li, C., Lu, X., Cheng, J., and Fan, X. (2021a). CellTalkDB: a manually curated database of ligand-receptor interactions in humans and mice. Brief Bioinform 22, bbaa269.

Shao, X., Liao, J., Lu, X., Xue, R., Ai, N., and Fan, X. (2020a). scCATCH: automatic annotation on cell types of clusters from single-cell RNA sequencing data. iScience 23, 100882.

Shao, X., Lu, X., Liao, J., Chen, H., and Fan, X. (2020b). New avenues for systematically inferring cell-cell communication: through single-cell transcriptomics data. Protein Cell 11, 866–880.

Shao, X., Wang, X., Guan, S., Lin, H., Yan, G., Gao, M., Deng, C., and Zhang, X. (2018). Integrated proteome analysis device for fast single-cell protein profiling. Anal Chem 90, 14003–14010.

Shao, X., Wang, Z., Wang, K., Lu, X., Zhang, P., Guo, R., Liao, J., Yang, P., Xu, X., and Fan, X. (2024). A single-cell landscape of human liver transplantation reveals a pathogenic immune niche associated with early allograft dysfunction. Engineering doi: 10.1016/j.eng.2023.12.004.

Shao, X., Yang, H., Zhuang, X., Liao, J., Yang, P., Cheng, J., Lu, X., Chen, H., and Fan, X. (2021b). scDeepSort: a pre-trained cell-type annotation method for single-cell transcriptomics using deep learning with a weighted graph neural network. Nucleic Acids Res 49, e122.

Shao, Y., Zhou, Y., Liu, Y., Zhang, W., Zhu, G., Zhao, Y., Zhang, Q., Yao, H., Zhao, H., Guo, G., et al. (2022b). Intact living-cell electrolaunching ionization mass spectrometry for single-cell metabolomics. Chem Sci 13, 8065–8073.

Shareef, S.J., Bevill, S.M., Raman, A.T., Aryee, M.J., van Galen, P., Hovestadt, V., and Bernstein, B.E. (2021). Extended-representation bisulfite sequencing of gene regulatory elements in multiplexed samples and single cells. Nat Biotechnol 39, 1086–1094.

Shen, Z., Zhang, R., Huang, Y., Chen, J., Yu, M., Li, C., Zhang, Y., Chen, L., Huang, X., Yang, J., et al. (2024). The spatial transcriptomic landscape of human gingiva in health and periodontitis. Sci China Life Sci 67, 720–732.

Sheng, K., Cao, W., Niu, Y., Deng, Q., and Zong, C. (2017). Effective detection of variation in single-cell transcriptomes using MATQ-seq. Nat Methods 14, 267–270.

Shomroni, O., Sitte, M., Schmidt, J., Parbin, S., Ludewig, F., Yigit, G., Zelarayan, L.C., Streckfuss-Bömeke, K., Wollnik, B., and Salinas, G. (2022). A novel single-cell RNA-sequencing approach and its applicability connecting genotype to phenotype in ageing disease. Sci Rep 12, 4091.

Shrestha, B. (2020). Single-cell metabolomics by mass spectrometry. In: Shrestha, B., ed. Single Cell Metabolism. Methods in Molecular Biology. New York: Humana. 1–8.

Shrestha, B., and Vertes, A. (2009). *In situ* metabolic profiling of single cells by laser ablation electrospray ionization mass spectrometry. Anal Chem 81, 8265–8271.

Shvartsburg, A.A., Li, F., Tang, K., and Smith, R.D. (2006). High-resolution field asymmetric waveform ion mobility spectrometry using new planar geometry analyzers. Anal Chem 78, 3706–3714.

Sidore, A.M., Lan, F., Lim, S.W., and Abate, A.R. (2016). Enhanced sequencing coverage with digital droplet multiple displacement amplification. Nucleic Acids Res 44, e66.

Simon, L.M., Wang, Y.Y., and Zhao, Z. (2021). Integration of millions of transcriptomes using batch-aware triplet neural networks. Nat Mach Intell 3, 705–715.

Singer, S.J. (1992). Intercellular communication and cell-cell adhesion. Science 255, 1671–1677.

Sinha, D., Kumar, A., Kumar, H., Bandyopadhyay, S., and Sengupta, D. (2018). dropClust: efficient clustering of ultra-large scRNA-seq data. Nucleic Acids Res 46, e36.

Sinjab, A., Han, G., Treekitkarnmongkol, W., Hara, K., Brennan, P.M., Dang, M., Hao, D., Wang, R., Dai, E., Dejima, H., et al. (2021). Resolving the spatial and cellular architecture of lung adenocarcinoma by multiregion single-cell sequencing. Cancer Discov 11, 2506–2523.

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., and Adey, A.C. (2019). The accessible chromatin landscape of the murine hippocampus at single-cell resolution. Genome Res 29, 857–869.

Skene, P.J., and Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. eLife 6, e21856.

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

Smieja, M., Wolczyk, M., Tabor, J., and Geiger, B.C. (2021). SeGMA: semi-supervised gaussian mixture autoencoder. IEEE Trans Neural Netw Learn Syst 32, 3930–3941.

Smith, T., Heger, A., and Sudbery, I. (2017). UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. Genome Res 27, 491–499.

Soneson, C., and Robinson, M.D. (2018). Bias, robustness and scalability in single-cell differential expression analysis. Nat Methods 15, 255–261.

Song, D., and Li, J.J. (2021). PseudotimeDE: inference of differential gene expression along cell pseudotime with well-calibrated p-values from single-cell RNA sequencing data. Genome Biol 22, 124.

Song, D., Li, K., Ge, X., and Li, J.J. (2023a). ClusterDE: a post-clustering differential expression (DE) method robust to false-positive inflation caused by double dipping. bioRxiv, doi: 10.1101/2023.07.21.550107.

Song, D., Li, K., Hemminger, Z., Wollman, R., and Li, J.J. (2021a). scPNMF: sparse gene encoding of single cells to facilitate gene selection for targeted gene profiling. Bioinformatics 37, i358–i366.

Song, D., Wang, Q., Yan, G., Liu, T., Sun, T., and Li, J.J. (2023b). scDesign3 generates realistic in silico data for multimodal single-cell and spatial omics. Nat Biotechnol 42, 247–252.

Song, J., Liu, Y., Zhang, X., Wu, Q., Gao, J., Wang, W., Li, J., Song, Y., and Yang, C. (2021b). Entropy subspace separation-based clustering for noise reduction (ENCORE) of scRNA-seq data. Nucleic Acids Res 49, e18.

Song, K., Yang, X., An, G., Xia, X., Zhao, J., Xu, X., Wan, C., Liu, T., Zheng, Y., Ren, S., et al. (2022a). Targeting APLN/APJ restores blood-testis barrier and improves spermatogenesis in murine and human diabetic models. Nat Commun 13, 7335.

Song, L., and Crawford, G.E. (2010). DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. Cold Spring Harb Protoc 2010, pdb.prot5384.

Song, Q., Ni, K., Liu, M., Li, Y., Wang, L., Wang, Y., Liu, Y., Yu, Z., Qi, Y., Lu, Z., et al. (2020). Direct-seq: programmed gRNA scaffold for streamlined scRNA-seq in CRISPR screen. Genome Biol 21, 136.

Song, Q., and Su, J. (2021). DSTG: deconvoluting spatial transcriptomics data through graph-based artificial intelligence. Brief Bioinform 22, bbaa414.

Song, Q., Su, J., and Zhang, W. (2021c). scGCN is a graph convolutional networks algorithm for knowledge transfer in single cell omics. Nat Commun 12, 3826.

Song, Q., Zhu, X., Jin, L., Chen, M., Zhang, W., and Su, J. (2022b). SMGR: a joint statistical method for integrative analysis of single-cell multi-omics data. NAR Genomics Bioinf 4, lqac056.

Soumillon, M., Cacchiarelli, D., Semrau, S., Oudenaarden, A.v. and Mikkelsen, T.S. (2014). Characterization of directed differentiation by high-throughput single-cell RNA-Seq. bioRxiv, doi: 10.1101/003236, 003236.

Specht, H., Emmott, E., Petelski, A.A., Huffman, R.G., Perlman, D.H., Serra, M., Kharchenko, P., Koller, A., and Slavov, N. (2021). Single-cell proteomic and transcriptomic analysis of macrophage heterogeneity using SCoPE2. Genome Biol 22, 50.

Specht, H., Harmange, G., Dh, P., Emmott, E., Niziolek, Z., Budnik, B., and Slavov, N. (2018). Automated sample preparation for high-throughput single-cell proteomics. bioRxiv, doi: 10.1101/399774.

Srivastava, A., Sarkar, H., Gupta, N., and Patro, R. (2016). RapMap: a rapid, sensitive and accurate tool for mapping RNA-seq reads to transcriptomes. Bioinformatics 32, i192–i200.

Stadlmann, J., Hudecz, O., Krššáková, G., Ctortecka, C., Van Raemdonck, G., Op De Beeck, J., Desmet, G., Penninger, J.M., Jacobs, P., and Mechtler, K. (2019). Improved sensitivity in low-input proteomics using micropillar array-based chromatography. Anal Chem 91, 14203–14207.

Ståhl, P.L., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J.F., Magnusson, J., Giacomello, S., Asp, M., Westholm, J.O., Huss, M., et al. (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Science 353, 78–82.

Stanojevic, S., Li, Y., Ristivojevic, A., and Garmire, L.X. (2022). Computational methods for single-cell multi-omics integration and alignment. Genomics Proteomics Bioinf 20, 836–849.

Stark, S.G., Ficek, J., Locatello, F., Bonilla, X., Chevrier, S., Singer, F., Aebersold, R., Al-Quaddoomi, F.S., Albinus, J., Alborelli, I., et al. (2020). SCIM: universal single-cell matching with unpaired feature sets. Bioinformatics 36, i919–i927.

Stegle, O., Teichmann, S.A., and Marioni, J.C. (2015). Computational and analytical challenges in single-cell transcriptomics. Nat Rev Genet 16, 133–145.

Stelzer, Y., Shivalila, C.S., Soldner, F., Markoulaki, S., and Jaenisch, R. (2015). Tracing dynamic changes of DNA methylation at single-cell resolution. Cell 163, 218–229.

Stickels, R.R., Murray, E., Kumar, P., Li, J., Marshall, J.L., Di Bella, D.J., Arlotta, P., Macosko, E.Z., and Chen, F. (2021). Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. Nat Biotechnol 39, 313–319.

Stirparo, G.G., Boroviak, T., Guo, G., Nichols, J., Smith, A., and Bertone, P. (2018). Integrated analysis of single-cell embryo data yields a unified transcriptome signature for the human preimplantation epiblast. Development 145, dev158501.

Stoeckius, M., Hafemeister, C., Stephenson, W., Houck-Loomis, B., Chattopadhyay, P. K., Swerdlow, H., Satija, R., and Smibert, P. (2017). Simultaneous epitope and transcriptome measurement in single cells. Nat Methods 14, 865–868.

Stopka, S.A., Khattar, R., Agtuca, B.J., Anderton, C.R., Paša-Tolić, L., Stacey, G., and Vertes, A. (2018). Metabolic noise and distinct subpopulations observed by single cell LAESI mass spectrometry of plant cells *in situ*. Front Plant Sci 9, 1646.

Storrs, E.P., Zhou, D.C., Wendl, M.C., Wyczalkowski, M.A., Karpova, A., Wang, L.B., Li, Y., Southard-Smith, A., Jayasinghe, R.G., Yao, L., et al. (2022). Pollock: fishing for cell states. Bioinform Adv 2, vbac028.

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

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck Iii, W.M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive integration of single-cell data. Cell 177, 1888–1902.e21.

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

Su, J.H., Zheng, P., Kinrot, S.S., Bintu, B., and Zhuang, X. (2020). Genome-scale imaging of the 3d organization and transcriptional activity of chromatin. Cell 182, 1641–1659.e26.

Su, K., Yu, T., and Wu, H. (2021). Accurate feature selection improves single-cell RNA-seq cell clustering. Brief Bioinform 22, bbab034.

Subramanian, A., Alperovich, M., Yang, Y., and Li, B. (2022). Biology-inspired data-driven quality control for scientific discovery in single-cell transcriptomics. Genome Biol 23, 267.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA 102, 15545–15550.

Sun, B., and Kumar, S. (2022). Protein adsorption loss—the bottleneck of single-cell proteomics. J Proteome Res 21, 1808–1815.

Sun, D., Liu, Z., Li, T., Wu, Q., and Wang, C. (2022). STRIDE: accurately decomposing and integrating spatial transcriptomics using single-cell RNA sequencing. Nucleic Acids Res 50, e42.

Sun, S., Zhu, J., and Zhou, X. (2020a). Statistical analysis of spatial expression patterns for spatially resolved transcriptomic studies. Nat Methods 17, 193–200.

Sun, T., Song, D., Li, W.V., and Li, J.J. (2021). scDesign2: a transparent simulator that generates high-fidelity single-cell gene expression count data with gene correlations captured. Genome Biol 22, 163.

Sun, W., Dong, H., Balaz, M., Slyper, M., Drokhlyansky, E., Colleluori, G., Giordano, A., Kovanicova, Z., Stefanicka, P., Balazova, L., et al. (2020b). snRNA-seq reveals a subpopulation of adipocytes that regulates thermogenesis. Nature 587, 98–102.

Sun, X., Liu, Y., and An, L. (2020c). Ensemble dimensionality reduction and feature gene extraction for single-cell RNA-seq data. Nat Commun 11, 5853.

Suo, C., Dann, E., Goh, I., Jardine, L., Kleshchevnikov, V., Park, J.E., Botting, R.A., Stephenson, E., Engelbert, J., Tuong, Z.K., et al. (2022). Mapping the developing human immune system across organs. Science 376, eabo0510.

Svensson, V., Natarajan, K.N., Ly, L.H., Miragaia, R.J., Labalette, C., Macaulay, I.C., Cvejic, A., and Teichmann, S.A. (2017). Power analysis of single-cell RNA-sequencing experiments. Nat Methods 14, 381–387.

Svensson, V., Teichmann, S.A., and Stegle, O. (2018). SpatialDE: identification of spatially variable genes. Nat Methods 15, 343–346.

Tajik, M., Baharfar, M., and Donald, W.A. (2022). Single-cell mass spectrometry. Trends Biotechnol 40, 1374–1392.

Talwar, D., Mongia, A., Sengupta, D., and Majumdar, A. (2018). AutoImpute: autoencoder based imputation of single-cell RNA-seq data. Sci Rep 8, 16329.

Tan, L., Xing, D., Chang, C.H., Li, H., and Xie, X.S. (2018). Three-dimensional genome structures of single diploid human cells. Science 361, 924–928.

Tanevski, J., Flores, R.O.R., Gabor, A., Schapiro, D., and Saez-Rodriguez, J. (2022). Explainable multiview framework for dissecting spatial relationships from highly multiplexed data. Genome Biol 23, 97.

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B.B., Siddiqui, A., et al. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods 6, 377–382.

Tang, L., Peng, S., Bi, Y., Shan, P., and Hu, X. (2014). A new method combining LDA and PLS for dimension reduction. PLoS ONE 9, e96944.

Tang, W., Bertaux, F., Thomas, P., Stefanelli, C., Saint, M., Marguerat, S., and Shahrezaei, V. (2020). bayNorm: Bayesian gene expression recovery, imputation and normalization for single-cell RNA-sequencing data. Bioinformatics 36, 1174–1181.

Tang, Z., Zhang, T., Yang, B., Su, J., and Song, Q. (2022). SiGra: single-cell spatial elucidation through image-augmented graph transformer. bioRxiv, doi: 10.1101/2022.08.18.504464.

Taukulis, I.A., Olszewski, R.T., Korrapati, S., Fernandez, K.A., Boger, E.T., Fitzgerald, T.S., Morell, R.J., Cunningham, L.L., and Hoa, M. (2021). Single-cell RNA-seq of cisplatin-treated adult stria vascularis identifies cell type-specific regulatory networks and novel therapeutic gene targets. Front Mol Neurosci 14, 718241.

Taylor, M.J., Lukowski, J.K., and Anderton, C.R. (2021). Spatially resolved mass spectrometry at the single cell: recent innovations in proteomics and metabolomics. J Am Soc Mass Spectrom 32, 872–894.

Tegowski, M., Flamand, M.N., and Meyer, K.D. (2022). scDART-seq reveals distinct m^6A signatures and mRNA methylation heterogeneity in single cells. Mol Cell 82, 868–878.e10.

Telenius, H.˚., Carter, N.P., Bebb, C.E., Nordenskjo¨ld, M., Ponder, B.A.J., and Tunnacliffe, A. (1992). Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. Genomics 13, 718–725.

Teng, H., Yuan, Y., and Bar-Joseph, Z. (2022). Clustering spatial transcriptomics data. Bioinformatics 38, 997–1004.

Thorner, K., Zorn, A.M., and Chaturvedi, P. (2021). ELeFHAnt: a supervised machine learning approach for label harmonization and annotation of single cell RNA-seq data. bioRxiv, doi: 10.1101/2021.09.07.459342.

Tian, Y., Li, Q., Yang, Z., Zhang, S., Xu, J., Wang, Z., Bai, H., Duan, J., Zheng, B., Li, W., et al. (2022). Single-cell transcriptomic profiling reveals the tumor heterogeneity of small-cell lung cancer. Sig Transduct Target Ther 7, 346.

Tiberi, S., Crowell, H.L., Samartsidis, P., Weber, L.M., and Robinson, M.D. (2022). *distinct*: a novel approach to differential distribution analyses. bioRxiv, doi: 10.1101/2020.11.24.394213.

Tirosh, I., Izar, B., Prakadan, S.M., Wadsworth Ii, M.H., Treacy, D., Trombetta, J.J., Rotem, A., Rodman, C., Lian, C., Murphy, G., et al. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science 352, 189–196.

Tobler, K.J., Brezina, P.R., Benner, A.T., Du, L., Xu, X., and Kearns, W.G. (2014). Two different microarray technologies for preimplantation genetic diagnosis and screening, due to reciprocal translocation imbalances, demonstrate equivalent euploidy and clinical pregnancy rates. J Assist Reprod Genet 31, 843–850.

Torres, A.J., Hill, A.S., and Love, J.C. (2014). Nanowell-based immunoassays for measuring single-cell secretion: characterization of transport and surface binding. Anal Chem 86, 11562–11569.

Townes, F.W., Hicks, S.C., Aryee, M.J., and Irizarry, R.A. (2019). Feature selection and dimension reduction for single-cell RNA-seq based on a multinomial model. Genome Biol 20, 295.

Tracy, S., Yuan, G.C., and Dries, R. (2019). RESCUE: imputing dropout events in single-cell RNA-sequencing data. BMC Bioinformatics 20, 388.

Tran, H.T.N., Ang, K.S., Chevrier, M., Zhang, X., Lee, N.Y.S., Goh, M., and Chen, J. (2020). A benchmark of batch-effect correction methods for single-cell RNA sequencing data. Genome Biol 21, 12.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol 32, 381–386.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25, 1105–1111.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28, 511–515.

Treff, N.R., Fedick, A., Tao, X., Devkota, B., Taylor, D., and Scott Jr., R.T. (2013). Evaluation of targeted next-generation sequencing-based preimplantation genetic diagnosis of monogenic disease. Fertil Steril 99, 1377–1384.e6.

Treff, N.R., Tao, X., Ferry, K.M., Su, J., Taylor, D., and Scott Jr., R.T. (2012). Development and validation of an accurate quantitative real-time polymerase chain reaction-based assay for human blastocyst comprehensive chromosomal aneuploidy screening. Fertil Steril 97, 819–824.e2.

Tsai, C.F., Zhang, P., Scholten, D., Martin, K., Wang, Y.T., Zhao, R., Chrisler, W.B., Patel, D.B., Dou, M., Jia, Y., et al. (2021). Surfactant-assisted one-pot sample preparation for label-free single-cell proteomics. Commun Biol 4, 265.

Tsai, C.F., Zhao, R., Williams, S.M., Moore, R.J., Schultz, K., Chrisler, W.B., Pasa-Tolic, L., Rodland, K.D., Smith, R.D., Shi, T., et al. (2020). An improved boosting to amplify signal with isobaric labeling (iBASIL) strategy for precise quantitative single-cell proteomics. Mol Cell Proteomics 19, 828–838.

Tu, Q., Cameron, R.A., Worley, K.C., Gibbs, R.A., and Davidson, E.H. (2012). Gene structure in the sea urchin *Strongylocentrotus purpuratus* based on transcriptome analysis. Genome Res 22, 2079–2087.

Tung, P.Y., Blischak, J.D., Hsiao, C.J., Knowles, D.A., Burnett, J.E., Pritchard, J.K., and Gilad, Y. (2017). Batch effects and the effective design of single-cell gene expression studies. Sci Rep 7, 39921.

Ulirsch, J.C., Lareau, C.A., Bao, E.L., Ludwig, L.S., Guo, M.H., Benner, C., Satpathy, A. T., Kartha, V.K., Salem, R.M., Hirschhorn, J.N., et al. (2019). Interrogation of human hematopoiesis at single-cell and single-variant resolution. Nat Genet 51, 683–693.

Unger, M.A., Chou, H.P., Thorsen, T., Scherer, A., and Quake, S.R. (2000). Monolithic microfabricated valves and pumps by multilayer soft lithography. Science 288, 113–116.

Uzun, Y., Wu, H., and Tan, K. (2021). Predictive modeling of single-cell DNA methylome data enhances integration with transcriptome data. Genome Res 31, 101–109.

Vaisvila, R., Ponnaluri, V.K.C., Sun, Z., Langhorst, B.W., Saleh, L., Guan, S., Dai, N., Campbell, M.A., Sexton, B.S., Marks, K., et al. (2021). Enzymatic methyl sequencing detects DNA methylation at single-base resolution from picograms of DNA. Genome Res 31, 1280–1289.

Vallejos, C.A., Marioni, J.C., and Richardson, S. (2015). BASiCS: Bayesian analysis of single-cell sequencing data. PLoS Comput Biol 11, e1004333.

van Dijk, D., Sharma, R., Nainys, J., Yim, K., Kathail, P., Carr, A.J., Burdziak, C., Moon, K.R., Chaffer, C.L., Pattabiraman, D., et al. (2018). Recovering gene interactions from single-cell data using data diffusion. Cell 174, 716–729.e27.

Vandenbon, A., and Diez, D. (2020). A clustering-independent method for finding differentially expressed genes in single-cell transcriptome data. Nat Commun 11, 4318.

Vickovic, S., Eraslan, G., Salmén, F., Klughammer, J., Stenbeck, L., Schapiro, D., Äijö, T., Bonneau, R., Bergenstråhle, L., Navarro, J.F., et al. (2019). High-definition spatial transcriptomics for *in situ* tissue profiling. Nat Methods 16, 987–990.

Villas-Bôas, S.G., Mas, S., Åkesson, M., Smedsgaard, J., and Nielsen, J. (2005). Mass spectrometry in metabolome analysis. Mass Spectrometry Rev 24, 613–646.

Vitak, S.A., Torkenczy, K.A., Rosenkrantz, J.L., Fields, A.J., Christiansen, L., Wong, M. H., Carbone, L., Steemers, F.J., and Adey, A. (2017). Sequencing thousands of single-cell genomes with combinatorial indexing. Nat Methods 14, 302–308.

Wagner, F. (2020). Monet: an open-source Python package for analyzing and integrating scRNA-Seq data using PCA-based latent spaces. bioRxiv, doi: 10.1101/2020.06.08.140673.

Wang, B., Zhu, J., Pierson, E., Ramazzotti, D., and Batzoglou, S. (2017a). Visualization and analysis of single-cell RNA-seq data by kernel-based similarity learning. Nat Methods 14, 414–416.

Wang, C., Sun, D., Huang, X., Wan, C., Li, Z., Han, Y., Qin, Q., Fan, J., Qiu, X., Xie, Y., et al. (2020). Integrative analyses of single-cell transcriptome and regulome using MAESTRO. Genome Biol 21, 198.

Wang, C.X., Zhang, L., and Wang, B. (2022a). One Cell At a Time (OCAT): a unified framework to integrate and analyze single-cell RNA-seq data. Genome Biol 23, 102.

Wang, D., Hou, S., Zhang, L., Wang, X., Liu, B., and Zhang, Z. (2021a). iMAP: integration of multiple single-cell datasets by adversarial paired transfer networks. Genome Biol 22, 63.

Wang, J., Agarwal, D., Huang, M., Hu, G., Zhou, Z., Ye, C., and Zhang, N.R. (2019a). Data denoising with transfer learning in single-cell transcriptomics. Nat Methods 16, 875–878.

Wang, J., Fan, H.C., Behr, B., and Quake, S.R. (2012). Genome-wide single-cell analysis of recombination activity and *de novo* mutation rates in human sperm. Cell 150, 402–412.

Wang, J., Qian, J., Hoeksema, M.D., Zou, Y., Espinosa, A.V., Rahman, S.M.J., Zhang, B., and Massion, P.P. (2013). Integrative genomics analysis identifies candidate drivers at 3q26-29 amplicon in squamous cell carcinoma of the lung. Clin Cancer Res 19, 5580–5590.

Wang, J., Vasaikar, S., Shi, Z., Greer, M., and Zhang, B. (2017b). WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. Nucleic Acids Res 45, W130–W137.

Wang, K., Kumar, T., Wang, J., Minussi, D.C., Sei, E., Li, J., Tran, T.M., Thennavan, A., Hu, M., Casasent, A.K., et al. (2023a). Archival single-cell genomics reveals persistent subclones during DCIS progression. Cell 186, 3968–3982.e15.

Wang, K., Li, X., Dong, S., Liang, J., Mao, F., Zeng, C., Wu, H., Wu, J., Cai, W., and Sun, Z.S. (2015). Q-RRBS: a quantitative reduced representation bisulfite sequencing method for single-cell methylome analyses. Epigenetics 10, 775–783.

Wang, L. (2021). Single-cell normalization and association testing unifying CRISPR screen and gene co-expression analyses with Normalisr. Nat Commun 12, 6395.

Wang, L., Ma, H., Wen, Z., Niu, L., Chen, X., Liu, H., Zhang, S., Xu, J., Zhu, Y., Li, H., et al. (2023b). Single-cell RNA-sequencing reveals heterogeneity and intercellular crosstalk in human tuberculosis lung. J Infect 87, 373–384.

Wang, Q., Xiong, H., Ai, S., Yu, X., Liu, Y., Zhang, J., and He, A. (2019b). CoBATCH for high-throughput single-cell epigenomic profiling. Mol Cell 76, 206–216.e7.

Wang, R., Zhao, H., Zhang, X., Zhao, X., Song, Z., and Ouyang, J. (2019c). Metabolic discrimination of breast cancer subtypes at the single-cell level by multiple microextraction coupled with mass spectrometry. Anal Chem 91, 3667–3674.

Wang, S., Karikomi, M., MacLean, A.L., and Nie, Q. (2019d). Cell lineage and communication network inference via optimization for single-cell transcriptomics. Nucleic Acids Res 47, e66.

Wang, S., Sun, S.T., Zhang, X.Y., Ding, H.R., Yuan, Y., He, J.J., Wang, M.S., Yang, B., and Li, Y.B. (2023c). The evolution of single-cell RNA sequencing technology and application: progress and perspectives. Int J Mol Sci 24, 2943.

Wang, T., Johnson, T.S., Shao, W., Lu, Z., Helm, B.R., Zhang, J., and Huang, K. (2019e). BERMUDA: a novel deep transfer learning method for single-cell RNA sequencing batch correction reveals hidden high-resolution cellular subtypes. Genome Biol 20, 165.

Wang, T., Wei, J.J., Sabatini, D.M., and Lander, E.S. (2014a). Genetic screens in human cells using the CRISPR-Cas9 system. Science 343, 80–84.

Wang, X., Allen, W.E., Wright, M.A., Sylwestrak, E.L., Samusik, N., Vesuna, S., Evans, K., Liu, C., Ramakrishnan, C., Liu, J., et al. (2018). Three-dimensional intact-tissue sequencing of single-cell transcriptional states. Science 361, eaat5691.

Wang, Y., Chen, L., Jo, J., and Wang, Y. (2022b). Joint *t*-SNE for comparable projections of multiple high-dimensional datasets. IEEE Trans Vis Comput Graph 28, 623–632.

Wang, Y., Gao, J., Xuan, C., Guan, T., Wang, Y., Zhou, G., and Ding, T. (2022c). FSCAM: CAM-based feature selection for clustering scRNA-seq. Interdiscip Sci 14, 394–408.

Wang, Y., Guan, Z.Y., Shi, S.W., Jiang, Y.R., Wu, Q., Wu, J., Chen, J.B., Ying, W.X., Xu, Q.Q., Fan, Q.X., et al. (2022d). Pick-up single-cell proteomic analysis for quantifying up to 3000 proteins in a tumor cell. bioRxiv, doi: 10.1101/2022.06.28.498038.

Wang, Y., Liu, T., and Zhao, H. (2022e). ResPAN: a powerful batch correction model for scRNA-seq data through residual adversarial networks. Bioinformatics 38, 3942–3949.

Wang, Y., Song, B., Wang, S., Chen, M., Xie, Y., Xiao, G., Wang, L., and Wang, T. (2022f). Sprod for de-noising spatially resolved transcriptomics data based on position and image information. Nat Methods 19, 950–958.

Wang, Y., Waters, J., Leung, M.L., Unruh, A., Roh, W., Shi, X., Chen, K., Scheet, P., Vattathil, S., Liang, H., et al. (2014b). Clonal evolution in breast cancer revealed by single nucleus genome sequencing. Nature 512, 155–160.

Wang, Y., Xie, S., Armendariz, D., and Hon, G.C. (2022g). Computational identification of clonal cells in single-cell CRISPR screens. BMC Genomics 23, 135.

Wang, Y., Yuan, P., Yan, Z., Yang, M., Huo, Y., Nie, Y., Zhu, X., Qiao, J., and Yan, L. (2021b). Single-cell multiomics sequencing reveals the functional regulatory landscape of early embryos. Nat Commun 12, 1247.

Wangwu, J., Sun, Z., and Lin, Z. (2021). scAMACE: model-based approach to the joint analysis of single-cell data on chromatin accessibility, gene expression and methylation. Bioinformatics 37, 3874–3880.

Wei, H., Han, T., Li, T., Wu, Q., and Wang, C. (2023). SCREE: a comprehensive pipeline for single-cell multi-modal CRISPR screen data processing and analysis. Brief Bioinform 24, bbad123.

Wei, X., Li, Z., Ji, H., and Wu, H. (2022). EDClust: an EM-MM hybrid method for cell clustering in multiple-subject single-cell RNA sequencing. Bioinformatics 38, 2692–2699.

Wei, Z., Zhang, X., Si, X., Gong, X., Zhang, S., and Zhang, X. (2020). Development of Pico-ESI-MS for single-cell metabolomics analysis. In: Shrestha, B., ed. Single Cell Metabolism. Methods in Molecular Biology. New York: Humana. 31–59.

Weibel, K.E., Mor, J.R., and Fiechter, A. (1974). Rapid sampling of yeast cells and automated assays of adenylate, citrate, pyruvate and glucose-6-phosphate pools. Anal Biochem 58, 208–216.

Welch, J.D., Hartemink, A.J., and Prins, J.F. (2017). MATCHER: manifold alignment reveals correspondence between single cell transcriptome and epigenome dynamics. Genome Biol 18, 138.

Welch, J.D., Kozareva, V., Ferreira, A., Vanderburg, C., Martin, C., and Macosko, E.Z. (2019). Single-cell multi-omic integration compares and contrasts features of brain cell identity. Cell 177, 1873–1887.e17.

Wells, D., Escudero, T., Levy, B., Hirschhorn, K., Delhanty, J.D.A., and Munné, S. (2002). First clinical application of comparative genomic hybridization and polar body testing for preimplantation genetic diagnosis of aneuploidy. Fertil Steril 78, 543–549.

Wells, D., Kaur, K., Grifo, J., Glassner, M., Taylor, J.C., Fragouli, E., and Munne, S. (2014). Clinical utilisation of a rapid low-pass whole genome sequencing technique for the diagnosis of aneuploidy in human embryos prior to implantation. J Med Genet 51, 553–562.

Wen, Y., Zhao, J., Zhang, R., Liu, F., Chen, X., Wu, D., Wang, M., Liu, C., Su, P., Meng, P., et al. (2024). Identification and characterization of human hematopoietic mesoderm. Sci China Life Sci 67, 320–331.

Wen, Z.H., Langsam, J.L., Zhang, L., Shen, W., and Zhou, X. (2022). A Bayesian factorization method to recover single-cell RNA sequencing data. Cell Rep Methods 2, 100133.

Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis. London: Springer.

Wilk, A.J., Shalek, A.K., Holmes, S., and Blish, C.A. (2024). Comparative analysis of cell-cell communication at single-cell resolution. Nat Biotechnol 42, 470–483.

Williams, S.M., Liyu, A.V., Tsai, C.F., Moore, R.J., Orton, D.J., Chrisler, W.B., Gaffrey, M.J., Liu, T., Smith, R.D., Kelly, R.T., et al. (2020). Automated coupling of nanodroplet sample preparation with liquid chromatography-mass spectrometry for high-throughput single-cell proteomics. Anal Chem 92, 10588–10596.

Wilson, P.C., Wu, H., Kirita, Y., Uchimura, K., Ledru, N., Rennke, H.G., Welling, P.A., Waikar, S.S., and Humphreys, B.D. (2019). The single-cell transcriptomic landscape of early human diabetic nephropathy. Proc Natl Acad Sci USA 116, 19619–19625.

Wilson, S.B., Howden, S.E., Vanslambrouck, J.M., Dorison, A., Alquicira-Hernandez, J., Powell, J.E., and Little, M.H. (2022). DevKidCC allows for robust classification and direct comparisons of kidney organoid datasets. Genome Med 14, 19.

Wishart, D.S., Tzur, D., Knox, C., Eisner, R., Guo, A.C., Young, N., Cheng, D., Jewell, K., Arndt, D., Sawhney, S., et al. (2007). HMDB: the Human Metabolome Database. Nucleic Acids Res 35, D521–D526.

Wiśniewski, J.R., Hein, M.Y., Cox, J., and Mann, M. (2014). A "Proteomic Ruler" for protein copy number and concentration estimation without spike-in standards. Mol Cell Proteomics 13, 3497–3506.

Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biol 19, 15.

Wolf, F.A., Hamey, F.K., Plass, M., Solana, J., Dahlin, J.S., Göttgens, B., Rajewsky, N., Simon, L., and Theis, F.J. (2019). PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. Genome Biol 20, 59.

Wolock, S.L., Lopez, R., and Klein, A.M. (2019). Scrublet: computational identification of cell doublets in single-cell transcriptomic data. Cell Syst 8, 281–291.e9.

Woo, J., Williams, S.M., Markillie, L.M., Feng, S., Tsai, C.F., Aguilera-Vazquez, V., Sontag, R.L., Moore, R.J., Hu, D., Mehta, H.S., et al. (2021). High-throughput and high-efficiency sample preparation for single-cell proteomics using a nested nanowell chip. Nat Commun 12, 6246.

Wroblewska, A., Dhainaut, M., Ben-Zvi, B., Rose, S.A., Park, E.S., Amir, E.A.D., Bektesevic, A., Baccarini, A., Merad, M., Rahman, A.H., et al. (2018). Protein barcodes enable high-dimensional single-cell CRISPR screens. Cell 175, 1141–1155.e16.

Wu, B., Wang, Y., Yan, J., Liu, M., Li, X., Tang, F., and Bao, S. (2024). Blastoids generated purely from embryonic stem cells both in mice and humans. Sci China Life Sci 67, 418–420.

Wu, H., Wu, Y., Jiang, Y., Zhou, B., Zhou, H., Chen, Z., Xiong, Y., Liu, Q., and Zhang, H. (2022). scHiCStackL: a stacking ensemble learning-based method for single-cell Hi-C classification using cell embedding. Brief Bioinform 23, bbab396.

Wu, K., Jia, F., Zheng, W., Luo, Q., Zhao, Y., and Wang, F. (2017a). Visualization of metallodrugs in single cells by secondary ion mass spectrometry imaging. J Biol Inorg Chem 22, 653–661.

Wu, L., Yan, J., Bai, Y., Chen, F., Xu, J., Zou, X., Huang, A., Hou, L., Zhong, Y., Jing, Z., et al. (2021a). Spatially-resolved transcriptomics analyses of invasive fronts in solid tumors. bioRxiv, doi: 10.1101/2021.10.21.465135.

Wu, P., Gao, Y., Guo, W., and Zhu, P. (2019a). Using local alignment to enhance single-cell bisulfite sequencing data efficiency. Bioinformatics 35, 3273–3278.

Wu, P.H., Gilkes, D.M., Phillip, J.M., Narkar, A., Cheng, T.W.T., Marchand, J., Lee, M. H., Li, R., and Wirtz, D. (2020). Single-cell morphology encodes metastatic potential. Sci Adv 6, eaaw6938.

Wu, R., Guo, W., Qiu, X., Wang, S., Sui, C., Lian, Q., Wu, J., Shan, Y., Yang, Z., Yang, S., et al. (2021b). Comprehensive analysis of spatial architecture in primary liver cancer. Sci Adv 7, eabg3750.

Wu, R., Xing, S., Badv, M., Didar, T.F., and Lu, Y. (2019b). Step-wise assessment and optimization of sample handling recovery yield for nanoproteomic analysis of 1000 mammalian cells. Anal Chem 91, 10395–10400.

Wu, S.J., Furlan, S.N., Mihalas, A.B., Kaya-Okur, H.S., Feroze, A.H., Emerson, S.N., Zheng, Y., Carson, K., Cimino, P.J., Keene, C.D., et al. (2021c). Single-cell CUT&Tag analysis of chromatin modifications in differentiation and tumor progression. Nat Biotechnol 39, 819–824.

Wu, S.Z., Al-Eryani, G., Roden, D.L., Junankar, S., Harvey, K., Andersson, A., Thennavan, A., Wang, C., Torpy, J.R., Bartonicek, N., et al. (2021d). A single-cell and spatially resolved atlas of human breast cancers. Nat Genet 53, 1334–1347.

Wu, W., Liu, Y., Dai, Q., Yan, X., and Wang, Z. (2021e). G2S3: a gene graph-based imputation method for single-cell RNA sequencing data. PLoS Comput Biol 17, e1009029.

Wu, X., Inoue, A., Suzuki, T., and Zhang, Y. (2017b). Simultaneous mapping of active DNA demethylation and sister chromatid exchange in single cells. Genes Dev 31, 511–523.

Xi, Y., and Li, W. (2009). BSMAP: whole genome bisulfite sequence MAPping program. BMC Bioinformatics 10, 232.

Xia, C., Fan, J., Emanuel, G., Hao, J., and Zhuang, X. (2019). Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. Proc Natl Acad Sci USA 116, 19490–19499.

Xie, H., and Ding, X. (2022). The intriguing landscape of single-cell protein analysis. Adv Sci 9, e2105932.

Xie, Q., Han, C., Jin, V., and Lin, S. (2022). HiCImpute: a Bayesian hierarchical model for identifying structural zeros and enhancing single cell Hi-C data. PLoS Comput Biol 18, e1010129.

Xie, S., Duan, J., Li, B., Zhou, P., and Hon, G.C. (2017). Multiplexed engineering and analysis of combinatorial enhancer activity in single cells. Mol Cell 66, 285–299. e5.

Xin, Y., Lyu, P., Jiang, J., Zhou, F., Wang, J., Blackshaw, S., and Qian, J. (2022). LRLoop: feedback loops as a design principle of cell-cell communication. bioRxiv, doi: 10.1101/2022.02.04.479174.

Xing, D., Tan, L., Chang, C.H., Li, H., and Xie, X.S. (2021). Accurate SNV detection in single cells by transposon-based whole-genome amplification of complementary strands. Proc Natl Acad Sci USA 118, e2013106118.

Xiong, K.X., Zhou, H.L., Yin, J.H., Kristiansen, K., Yang, H.M., and Li, G.B. (2021a). Chord: identifying doublets in single-cell RNA sequencing data by an ensemble machine learning algorithm. bioRxiv, doi: 10.1101/2021.05.07.442884.

Xiong, L., Tian, K., Li, Y., and Zhang, Q.C. (2021b). Online single-cell data integration through projecting heterogeneous datasets into a common cell-embedding space. bioRxiv, doi: 10.1101/2021.04.06.438536.

Xu, C., and Su, Z. (2015). Identification of cell types from single-cell transcriptomes using a novel clustering method. Bioinformatics 31, 1974–1980.

Xu, J., Cai, L., Liao, B., Zhu, W., and Yang, J. (2020a). CMF-Impute: an accurate imputation tool for single-cell RNA-seq data. Bioinformatics 36, 3139–3147.

Xu, S., Liu, M., Bai, Y., and Liu, H. (2021a). Multi-dimensional organic mass cytometry: simultaneous analysis of proteins and metabolites on single cells. Angew Chem Int Ed 60, 1806–1812.

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

Xu, Y., Begoli, E., and McCord, R.P. (2021b). sciCAN: single-cell chromatin accessibility and gene expression data integration via Cycle-consistent Adversarial Network. bioRxiv, doi: 10.1101/2021.11.30.470677.

Xu, Y., and McCord, R.P. (2022). Diagonal integration of multimodal single-cell data: potential pitfalls and paths forward. Nat Commun 13, 3505.

Xu, Y., Zhang, Z., You, L., Liu, J., Fan, Z., and Zhou, X. (2020b). scIGANs: single-cell RNA-seq imputation using generative adversarial networks. Nucleic Acids Res 48, e85.

Xu, Z., Chen, D., Hu, Y., Jiang, K., Huang, H., Du, Y., Wu, W., Wang, J., Sui, J., Wang, W., et al. (2022). Anatomically distinct fibroblast subsets determine skin autoimmune patterns. Nature 601, 118–124.

Yan, F., Zhao, Z., and Simon, L.M. (2021). EmptyNN: a neural network based on positive and unlabeled learning to remove cell-free droplets and recover lost cells in scRNA-seq data. Patterns 2, 100311.

Yang, L., Liu, J., Lu, Q., Riggs, A.D., and Wu, X. (2017). SAIC: an iterative clustering approach for analysis of single cell RNA-seq data. BMC Genomics 18, 689.

Yang, L., Wang, Z., Deng, Y., Li, Y., Wei, W., and Shi, Q. (2016). Single-cell, multiplexed protein detection of rare tumor cells based on a beads-on-barcode antibody microarray. Anal Chem 88, 11077–11083.

Yang, L., Zhu, Y., Yu, H., Cheng, X., Chen, S., Chu, Y., Huang, H., Zhang, J., and Li, W. (2020). scMAGeCK links genotypes with multiple phenotypes in single-cell CRISPR screens. Genome Biol 21, 19.

Yang, S., Chen, D., Xie, L., Zou, X., Xiao, Y., Rao, L., Yao, T., Zhang, Q., Cai, L., Huang, F., et al. (2023a). Developmental dynamics of the single nucleus regulatory landscape of pig hippocampus. Sci China Life Sci 66, 2614–2628.

Yang, X., Yan, J., Cheng, Y., and Zhang, Y. (2023b). Learning deep generative clustering via mutual information maximization. IEEE Trans Neural Netw Learn Syst 34, 6263–6275.

Yang, Y., Li, G., Qian, H., Wilhelmsen, K.C., Shen, Y., and Li, Y. (2021a). SMNN: batch effect correction for single-cell RNA-seq data via supervised mutual nearest neighbor detection. Brief Bioinform 22, bbaa097.

Yang, Y., Li, G., Xie, Y., Wang, L., Lagler, T.M., Yang, Y., Liu, J., Qian, L., and Li, Y. (2021b). iSMNN: batch effect correction for single-cell RNA-seq data via iterative supervised mutual nearest neighbor refinement. Brief Bioinform 22, bbab122.

Yang, Y., Shi, X., Liu, W., Zhou, Q., Chan Lau, M., Chun Tatt Lim, J., Sun, L., Ng, C.C. Y., Yeong, J., and Liu, J. (2022). SC-MEB: spatial clustering with hidden Markov random field using empirical Bayes. Brief Bioinform 23, bbab466.

Yao, H., Zhao, H., Zhao, X., Pan, X., Feng, J., Xu, F., Zhang, S., and Zhang, X. (2019). Label-free mass cytometry for unveiling cellular metabolic heterogeneity. Anal Chem 91, 9777–9783.

Yao, Y.X., La, Y.F., Di, R., Liu, Q.Y., Hu, W.P., Wang, X.Y., and Chu, M.X. (2018). Comparison of different single cell whole genome amplification methods and MALBAC applications in assisted reproduction (in Chinese). Hereditas 40, 620–631.

Yasen, A., Aini, A., Wang, H., Li, W., Zhang, C., Ran, B., Tuxun, T., Maimaitinijiati, Y., Shao, Y., Aji, T., et al. (2020). Progress and applications of single-cell sequencing techniques. Infect Genet Evol 80, 104198.

Yin, L., Zhang, Z., Liu, Y., Gao, Y., and Gu, J. (2019a). Recent advances in single-cell analysis by mass spectrometry. Analyst 144, 824–845.

Yin, R., Burnum-Johnson, K.E., Sun, X., Dey, S.K., and Laskin, J. (2019b). High spatial resolution imaging of biological tissues using nanospray desorption electrospray ionization mass spectrometry. Nat Protoc 14, 3445–3470.

Yin, R., Prabhakaran, V., and Laskin, J. (2018). Quantitative extraction and mass spectrometry analysis at a single-cell level. Anal Chem 90, 7937–7945.

Yin, Y., Jiang, Y., Lam, K.W.G., Berletch, J.B., Disteche, C.M., Noble, W.S., Steemers, F. J., Camerini-Otero, R.D., Adey, A.C., and Shendure, J. (2019c). High-throughput single-cell sequencing with linear amplification. Mol Cell 76, 676–690.e10.

You, Y., Tian, L., Su, S., Dong, X., Jabbari, J.S., Hickey, P.F., and Ritchie, M.E. (2021). Benchmarking UMI-based single-cell RNA-seq preprocessing workflows. Genome Biol 22, 339.

Yu, L., Cao, Y., Yang, J.Y.H., and Yang, P. (2022). Benchmarking clustering algorithms on estimating the number of cell types from single-cell RNA-sequencing data. Genome Biol 23, 49.

Yu, M., Abnousi, A., Zhang, Y., Li, G., Lee, L., Chen, Z., Fang, R., Lagler, T.M., Yang, Y., Wen, J., et al. (2021). SnapHiC: a computational pipeline to identify chromatin loops from single-cell Hi-C data. Nat Methods 18, 1056–1059.

Yuan, H., and Kelley, D.R. (2022). scBasset: sequence-based modeling of single-cell ATAC-seq using convolutional neural networks. Nat Methods 19, 1088–1096.

Yuan, P., Xu, C., He, N., Lu, X., Zhang, X., Shang, J., Zhu, H., Gong, C., Kuang, H., Tang, T., et al. (2023). Watermelon domestication was shaped by stepwise selection and regulation of the metabolome. Sci China Life Sci 66, 579–594.

Yuan, Y., and Bar-Joseph, Z. (2020). GCNG: graph convolutional networks for inferring gene interaction from spatial transcriptomics data. Genome Biol 21, 300.

Zahn, H., Steif, A., Laks, E., Eirew, P., VanInsberghe, M., Shah, S.P., Aparicio, S., and Hansen, C.L. (2017). Scalable whole-genome single-cell library preparation without preamplification. Nat Methods 14, 167–173.

Zamanighomi, M., Lin, Z., Daley, T., Chen, X., Duren, Z., Schep, A., Greenleaf, W.J., and Wong, W.H. (2018). Unsupervised clustering and epigenetic classification of single cells. Nat Commun 9, 2410.

Zappia, L., Phipson, B., and Oshlack, A. (2017). Splatter: simulation of single-cell RNA sequencing data. Genome Biol 18, 174.

Zeira, R., Land, M., Strzalkowski, A., and Raphael, B.J. (2022). Alignment and integration of spatial transcriptomics data. Nat Methods 19, 567–575.

Zelig, A., and Kaplan, N. (2020). KMD clustering: robust generic clustering of biological data. bioRxiv, doi: 10.1101/2020.10.04.325233.

Zeng, P., and Lin, Z. (2021). coupleCoC+: an information-theoretic co-clustering-based transfer learning framework for the integrative analysis of single-cell genomic data. PLoS Comput Biol 17, e1009064.

Zeng, Y., Chen, X., Luo, Y., Li, X., and Peng, D. (2021). Deep drug-target binding affinity prediction with multiple attention blocks. Brief Bioinform 22, bbaa347.

Zeng, Y., Wei, Z., Pan, Z., Lu, Y., and Yang, Y. (2022a). A robust and scalable graph neural network for accurate single-cell classification. Brief Bioinform 23, bbab570.

Zeng, Y., Wei, Z., Yu, W., Yin, R., Yuan, Y., Li, B., Tang, Z., Lu, Y., and Yang, Y. (2022b). Spatial transcriptomics prediction from histology jointly through Transformer and graph neural networks. Brief Bioinform 23, bbac297.

Zeng, Y., Wei, Z., Zhong, F., Pan, Z., Lu, Y., and Yang, Y. (2022c). A parameter-free deep embedded clustering method for single-cell RNA-seq data. Brief Bioinform 23, bbac172.

Zenobi, R. (2013). Single-Cell metabolomics: analytical and biological perspectives. Science 342, 1243259.

Zhang, A.W., O'Flanagan, C., Chavez, E.A., Lim, J.L.P., Ceglia, N., McPherson, A., Wiens, M., Walters, P., Chan, T., Hewitson, B., et al. (2019a). Probabilistic cell-type assignment of single-cell RNA-seq for tumor microenvironment profiling. Nat Methods 16, 1007–1015.

Zhang, B., Kirov, S., and Snoddy, J. (2005). WebGestalt: an integrated system for exploring gene sets in various biological contexts. Nucleic Acids Res 33, W741–W748.

Zhang, B., Srivastava, A., Mimitou, E., Stuart, T., Raimondi, I., Hao, Y., Smibert, P., and Satija, R. (2022a). Characterizing cellular heterogeneity in chromatin state with scCUT&Tag-pro. Nat Biotechnol 40, 1220–1230.

Zhang, F., Wu, Y., and Tian, W. (2019b). A novel approach to remove the batch effect of single-cell data. Cell Discov 5, 46.

Zhang, K., Feng, W., and Wang, P. (2022b). Identification of spatially variable genes with graph cuts. Nat Commun 13, 5488.

Zhang, L., Cui, X., Schmitt, K., Hubert, R., Navidi, W., and Arnheim, N. (1992). Whole genome amplification from a single cell: implications for genetic analysis. Proc Natl Acad Sci USA 89, 5847–5851.

Zhang, L., and Nie, Q. (2021). scMC learns biological variation through the alignment of multiple single-cell genomics datasets. Genome Biol 22, 10.

Zhang, L., Sevinsky, C.J., Davis, B.M., and Vertes, A. (2018). Single-cell mass spectrometry of subpopulations selected by fluorescence microscopy. Anal Chem 90, 4626–4634.

Zhang, L., and Vertes, A. (2015). Energy charge, redox state, and metabolite turnover in single human hepatocytes revealed by capillary microsampling mass spectrometry. Anal Chem 87, 10397–10405.

Zhang, L., and Vertes, A. (2018). Single-cell mass spectrometry approaches to explore cellular heterogeneity. Angew Chem Int Ed 57, 4466–4477.

Zhang, L., Zhang, J., and Nie, Q. (2022c). DIRECT-NET: an efficient method to discover cis-regulatory elements and construct regulatory networks from single-cell multiomics data. Sci Adv 8, eabl7393.

Zhang, M., Eichhorn, S.W., Zingg, B., Yao, Z., Cotter, K., Zeng, H., Dong, H., and Zhuang, X. (2021a). Spatially resolved cell atlas of the mouse primary motor cortex by MERFISH. Nature 598, 137–143.

Zhang, M., Hu, S., Min, M., Ni, Y., Lu, Z., Sun, X., Wu, J., Liu, B., Ying, X., and Liu, Y. (2021b). Dissecting transcriptional heterogeneity in primary gastric adenocarcinoma by single cell RNA sequencing. Gut 70, 464–475.

Zhang, M., Liu, S., Miao, Z., Han, F., Gottardo, R., and Sun, W. (2022d). IDEAS: individual level differential expression analysis for single-cell RNA-seq data. Genome Biol 23, 33.

Zhang, R., Meng-Papaxanthos, L., Vert, J.P., and Noble, W.S. (2021c). Semi-supervised single-cell cross-modality translation using Polarbear. bioRxiv, doi: 10.1101/2021.11.18.467517.

Zhang, R., Zhou, T., and Ma, J. (2022e). Multiscale and integrative single-cell Hi-C analysis with Higashi. Nat Biotechnol 40, 254–261.

Zhang, X., Lan, Y., Xu, J., Quan, F., Zhao, E., Deng, C., Luo, T., Xu, L., Liao, G., Yan, M., et al. (2019c). CellMarker: a manually curated resource of cell markers in human and mouse. Nucleic Acids Res 47, D721–D728.

Zhang, X.F., Ou-Yang, L., Yang, S., Zhao, X.M., Hu, X., and Yan, H. (2019d). EnImpute: imputing dropout events in single-cell RNA-sequencing data via ensemble learning. Bioinformatics 35, 4827–4829.

Zhang, Y., Fonslow, B.R., Shan, B., Baek, M.C., and Yates Iii, J.R. (2013). Protein analysis by shotgun/bottom-up proteomics. Chem Rev 113, 2343–2394.

Zhang, Y., Li, Q., Jiang, N., Su, Z., Yuan, Q., Lv, L., Sang, X., Chen, R., Feng, Y., and Chen, Q. (2022f). Dihydroartemisinin beneficially regulates splenic immune cell heterogeneity through the SOD3-JNK-AP-1 axis. Sci China Life Sci 65, 1636–1654.

Zhang, Y., Liu, T., Hu, X., Wang, M., Wang, J., Zou, B., Tan, P., Cui, T., Dou, Y., Ning, L., et al. (2021d). CellCall: integrating paired ligand-receptor and transcription factor activities for cell-cell communication. Nucleic Acids Res 49, 8520–8534.

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

Zhang, Y., and Wang, F. (2021). SSBER: removing batch effect for single-cell RNA sequencing data. BMC Bioinformatics 22, 249.

Zhang, Y., Xie, X., Wu, P., and Zhu, P. (2021e). SIEVE: identifying robust single cell variable genes for single-cell RNA sequencing data. Blood Sci 3, 35–39.

Zhang, Y., Zhang, F., Wang, Z., Wu, S., and Tian, W. (2022g). scMAGIC: accurately annotating single cells using two rounds of reference-based classification. Nucleic Acids Res 50, e43.

Zhao, C., Hu, S., Huo, X., and Zhang, Y. (2017). Dr.seq2: a quality control and analysis pipeline for parallel single cell transcriptome and epigenome data. PLoS ONE 12, e0180583.

Zhao, E., Stone, M.R., Ren, X., Guenthoer, J., Smythe, K.S., Pulliam, T., Williams, S.R., Uytingco, C.R., Taylor, S.E.B., Nghiem, P., et al. (2021a). Spatial transcriptomics at subspot resolution with BayesSpace. Nat Biotechnol 39, 1375–1384.

Zhao, J., Wang, G., Ming, J., Lin, Z., Wang, Y., Wu, A.R., and Yang, C. (2022a). Adversarial domain translation networks for fast and accurate integration of large-scale atlas-level single-cell datasets. bioRxiv, doi: 10.1101/2021.11.16.468892.

Zhao, M., Jiang, J., Zhao, M., Chang, C., Wu, H., and Lu, Q. (2021b). The application of single-cell RNA sequencing in studies of autoimmune diseases: a comprehensive review. Clinic Rev Allerg Immunol 60, 68–86.

Zhao, T., Chiang, Z.D., Morriss, J.W., LaFave, L.M., Murray, E.M., Del Priore, I., Meli, K., Lareau, C.A., Nadaf, N.M., Li, J., et al. (2022b). Spatial genomics enables multi-modal study of clonal heterogeneity in tissues. Nature 601, 85–91.

Zhao, X., Guo, J., Nie, F., Chen, L., Li, Z., and Zhang, H. (2020). Joint principal component and discriminant analysis for dimensionality reduction. IEEE Trans Neural Netw Learn Syst 31, 433–444.

Zhao, Y., Wang, T., Liu, Z., Ke, Y., Li, R., Chen, H., You, Y., Wu, G., Cao, S., Du, Z., et al. (2023). Single-cell transcriptomics of immune cells in lymph nodes reveals their composition and alterations in functional dynamics during the early stages of bubonic plague. Sci China Life Sci 66, 110–126.

Zhao, Z., Zhu, H., Li, Q., Liao, W., Chen, K., Yang, M., Long, D., He, Z., Zhao, M., Wu, H., et al. (2022c). Skin CD4$^+$ Trm cells distinguish acute cutaneous lupus erythematosus from localized discoid lupus erythematosus/subacute cutaneous lupus erythematosus and other skin diseases. J Autoimmun 128, 102811.

Zheng, G.X.Y., 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. (2017). Massively parallel digital transcriptional profiling of single cells. Nat Commun 8, 14049.

Zheng, M., Hu, Z., Mei, X., Ouyang, L., Song, Y., Zhou, W., Kong, Y., Wu, R., Rao, S., Long, H., et al. (2022a). Single-cell sequencing shows cellular heterogeneity of cutaneous lesions in lupus erythematosus. Nat Commun 13, 7489.

Zheng, R., Dong, X., Wan, C., Shi, X., Zhang, X., and Meyer, C.A. (2020). Cistrome Data Browser and Toolkit: analyzing human and mouse genomic data using compendia of ChIP-seq and chromatin accessibility data. Quant Biol 8, 267–276.

Zheng, R., Wan, C., Mei, S., Qin, Q., Wu, Q., Sun, H., Chen, C.H., Brown, M., Zhang, X., Meyer, C.A., et al. (2019). Cistrome Data Browser: expanded datasets and new tools for gene regulatory analysis. Nucleic Acids Res 47, D729–D735.

Zheng, R., Zhang, Y., Tsuji, T., Gao, X., Wagner, A., Yosef, N., Chen, H., Zhang, L., Tseng, Y.H., and Chen, K. (2022b). MEBOCOST: metabolite-mediated cell communication modeling by single cell transcriptome. bioRxiv, doi: 10.1101/2022.05.30.494067.

Zheng, X.T., and Li, C.M. (2012). Single cell analysis at the nanoscale. Chem Soc Rev 41, 2061–2071.

Zhong, L., Yang, X., Zhou, Y., Xiao, J., Li, H., Tao, J., Xi, Q., Chu, C., Li, C., Yang, X., et al. (2022). Exploring the R-ISS stage-specific regular networks in the progression of multiple myeloma at single-cell resolution. Sci China Life Sci 65, 1811–1823.

Zhou, J., Ma, J., Chen, Y., Cheng, C., Bao, B., Peng, J., Sejnowski, T.J., Dixon, J.R., and Ecker, J.R. (2019a). Robust single-cell Hi-C clustering by convolution- and random-walk-based imputation. Proc Natl Acad Sci USA 116, 14011–14018.

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A.H., Tanaseichuk, O., Benner, C., and Chanda, S.K. (2019b). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun 10, 1523.

Zhu, A., Srivastava, A., Ibrahim, J.G., Patro, R., and Love, M.I. (2019a). Nonparametric expression analysis using inferential replicate counts. Nucleic Acids Res 47, e105.

Zhu, C., Yu, M., Huang, H., Juric, I., Abnousi, A., Hu, R., Lucero, J., Behrens, M.M., Hu, M., and Ren, B. (2019b). An ultra high-throughput method for single-cell joint analysis of open chromatin and transcriptome. Nat Struct Mol Biol 26, 1063–1070.

Zhu, C., Zhang, Y., Li, Y.E., Lucero, J., Behrens, M.M., and Ren, B. (2021a). Joint profiling of histone modifications and transcriptome in single cells from mouse brain. Nat Methods 18, 283–292.

Zhu, H., and Wang, Z. (2019). SCL: a lattice-based approach to infer 3D chromosome structures from single-cell Hi-C data. Bioinformatics 35, 3981–3988.

Zhu, H., Zou, G., Wang, N., Zhuang, M., Xiong, W., and Huang, G. (2017). Single-neuron identification of chemical constituents, physiological changes, and metabolism using mass spectrometry. Proc Natl Acad Sci USA 114, 2586–2591.

Zhu, J., and Sabatti, C. (2020). Integrative spatial single-cell analysis with graph-based feature learning. bioRxiv, doi: 10.1101/2020.08.12.248971.

Zhu, J., Sun, S., and Zhou, X. (2021b). SPARK-X: non-parametric modeling enables scalable and robust detection of spatial expression patterns for large spatial transcriptomic studies. Genome Biol 22, 184.

Zhu, P., Guo, H., Ren, Y., Hou, Y., Dong, J., Li, R., Lian, Y., Fan, X., Hu, B., Gao, Y., et al. (2018a). Single-cell DNA methylome sequencing of human preimplantation embryos. Nat Genet 50, 12–19.

Zhu, Q., Shah, S., Dries, R., Cai, L., and Yuan, G.C. (2018b). Identification of spatially associated subpopulations by combining scRNAseq and sequential fluorescence *in situ* hybridization data. Nat Biotechnol 36, 1183–1190.

Zhu, Y., Clair, G., Chrisler, W.B., Shen, Y., Zhao, R., Shukla, A.K., Moore, R.J., Misra, R.S., Pryhuber, G.S., Smith, R.D., et al. (2018c). Proteomic analysis of single mammalian cells enabled by microfluidic nanodroplet sample preparation and ultrasensitive nanoLC-MS. Angew Chem Int Ed 57, 12370–12374.

Zhu, Y., Piehowski, P.D., Zhao, R., Chen, J., Shen, Y., Moore, R.J., Shukla, A.K., Petyuk, V.A., Campbell-Thompson, M., Mathews, C.E., et al. (2018d). Nanodroplet processing platform for deep and quantitative proteome profiling of 10–100 mammalian cells. Nat Commun 9, 882.

Zhu, Y., Podolak, J., Zhao, R., Shukla, A.K., Moore, R.J., Thomas, G.V., and Kelly, R.T. (2018e). Proteome profiling of 1 to 5 spiked circulating tumor cells isolated from whole blood using immunodensity enrichment, laser capture microdissection, nanodroplet sample processing, and ultrasensitive nanoLC-MS. Anal Chem 90, 11756–11759.

Zhu, Y., Zhang, Y.X., Cai, L.F., and Fang, Q. (2013). Sequential operation droplet array: an automated microfluidic platform for picoliter-scale liquid handling, analysis, and screening. Anal Chem 85, 6723–6731.

Zhu, Y., Zhao, R., Piehowski, P.D., Moore, R.J., Lim, S., Orphan, V.J., Paša-Tolić, L., Qian, W.J., Smith, R.D., and Kelly, R.T. (2018f). Subnanogram proteomics: impact of LC column selection, MS instrumentation and data analysis strategy on proteome coverage for trace samples. Int J Mass Spectrom 427, 4–10.

Zhuang, X. (2021). Spatially resolved single-cell genomics and transcriptomics by imaging. Nat Methods 18, 18–22.

Zong, C., Lu, S., Chapman, A.R., and Xie, X.S. (2012). Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. Science 338, 1622–1626.

Zong, Y., Yu, T., Wang, X., Wang, Y., Hu, Z., and Li, Y. (2022). conST: an interpretable multi-modal contrastive learning framework for spatial transcriptomics. bioRxiv, doi: 10.1101/2022.01.14.476408.

Zou, H., Hastie, T., and Tibshirani, R. (2006). Sparse principal component analysis. J Comput Graph Stat 15, 265–286.

Zou, J., Deng, F., Wang, M., Zhang, Z., Liu, Z., Zhang, X., Hua, R., Chen, K., Zou, X., and Hao, J. (2022). scCODE: an R package for data-specific differentially expressed gene detection on single-cell RNA-sequencing data. Brief Bioinform 23, bbac180.

Zou, Z., Hua, K., and Zhang, X. (2021). HGC: fast hierarchical clustering for large-scale single-cell data. Bioinformatics 37, 3964–3965.

Zuo, C., Zhang, Y., Cao, C., Feng, J., Jiao, M., and Chen, L. (2022). Elucidating tumor heterogeneity from spatially resolved transcriptomics data by multi-view graph collaborative learning. Nat Commun 13, 5962.

Žurauskienė, J., and Yau, C. (2016). pcaReduce: hierarchical clustering of single cell transcriptional profiles. BMC Bioinformatics 17, 140.