ADVANCED BIOLOGY
www.advanced-bio.com

Single-Cell Sequencing to Unveil the Mystery of Embryonic Development

Zida Li, Feng Lin, Chu-Han Zhong, Shue Wang, Xufeng Xue, and Yue Shao**

Embryonic development is a fundamental physiological process that can provide tremendous insights into stem cell biology and regenerative medicine. In this process, cell fate decision is highly heterogeneous and dynamic, and investigations at the single-cell level can greatly facilitate the understanding of the molecular roadmap of embryonic development. Rapid advances in the technology of single-cell sequencing offer a perfectly useful tool to fulfill this purpose. Despite its great promise, single-cell sequencing is highly interdisciplinary, and successful applications in specific biological contexts require a general understanding of its diversity as well as the advantage versus limitations for each of its variants. Here, the technological principles of single-cell sequencing are consolidated and its applications in the study of embryonic development are summarized. First, the technology basics are presented and the available tools for each step including cell isolation, library construction, sequencing, and data analysis are discussed. Then, the works that employed single-cell sequencing are reviewed to investigate the specific processes of embryonic development, including preimplantation, peri-implantation, gastrulation, and organogenesis. Further, insights are provided on existing challenges and future research directions.

1. Introduction

Embryonic development is one of the most fundamental yet mysterious biological processes. Starting from a single-celled zygote, the embryo goes through cleavage, implantation, gastrulation, and organogenesis, transforming from a single-celled zygote to an organism with a large number of cells with drastically different phenotypes.[1] In this process, tremendous incidences on the molecular and cellular levels take place, driving essential developmental activities such as lineage specifications, axis patterning, and organogenesis. Detailed understandings of the molecular mechanisms of this process, such as the transcriptome and epigenome, are critical for fundamental embryology study, management of reproduction-related diseases, and regenerative medicine.[2] Given that embryonic cells are scarce yet highly heterogeneous, analysis with the single-cell resolution is thus essential for a complete knowledge of embryonic development. However, conventional cell analysis, such as DNA microarray, quantitative real-time polymerase chain reaction (PCR), and sequencing, generally requires a sample of pooled cells and thus gauges the ensemble average, masking the cell heterogeneity among the tested sample. By performing analysis on individual cells separately, one could obtain analysis with single-cell resolution.

The single-cell analysis normally includes four steps: 1) single-cell isolation and manipulation, 2) sample preparation, 3) sequencing, and 4) data analysis.[3] In principle, single-cell analysis can be performed following the same testing mechanisms as the bulk assays. However, there are a few challenges

Z. Li
Department of Biomedical Engineering
School of Medicine
Shenzhen University
Shenzhen 518060, China
E-mail: zidali@szu.edu.cn

Z. Li
Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging
Department of Biomedical Engineering
School of Medicine
Shenzhen University
Shenzhen 518060, China

F. Lin
Department of Mechanics and Engineering Science
College of Engineering
Peking University
Beijing 100871, China

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/adbi.202101151.

DOI: 10.1002/adbi.202101151

C.-H. Zhong
International Center for Applied Mechanics
State Key Laboratory for Strength and Vibration of Mechanical Structures
School of Aerospace Engineering
Xi'an Jiaotong University
Xi'an 710049, China

S. Wang
Department of Chemistry, Chemical, and Biomedical Engineering
Tagliatela College of Engineering
University of New Haven
West Haven, CT 06561, USA

X. Xue
Department of Mechanical Engineering
University of Michigan
Ann Arbor, MI 48109, USA

Y. Shao
Institute of Biomechanics and Medical Engineering
Department of Engineering Mechanics
School of Aerospace Engineering
Tsinghua University
Beijing 100084, China
E-mail: yshao@tsinghua.edu.cn

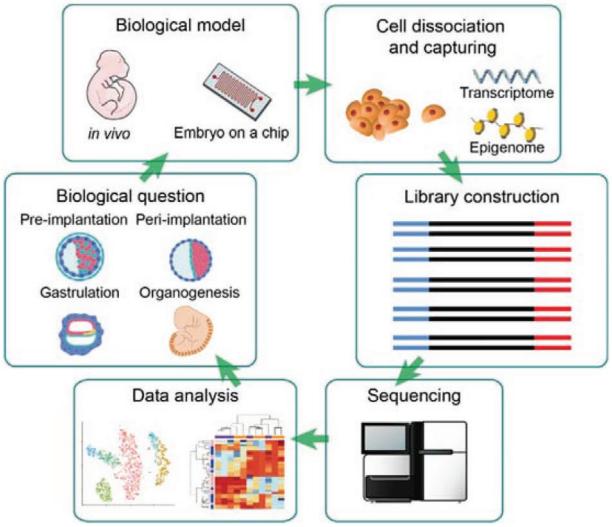


Figure 1. Flow diagram of single-cell analysis for the study of embryo development. In Section 2, we discuss specific steps in the single-cell analysis including cell isolation, library preparation, sequencing, and data analysis. In Section 3, we present representative works on the study of embryo development using different models. Reproduced under the terms of the CC-BY license.[22] Copyright 2021, The Authors, published by Cell Press.

in translating assays from bulk to single cells.[4] First, due to cells' small size and oftentimes vast number, a technique for efficient cell isolation, manipulation, and indexing with high throughput is necessary. Second, since each cell has only a minute amount of analyte, reliable analysis of such small samples requires high-performance analyte enrichment and reactions.[5] Last, high-throughput single-cell analysis generates high-volume data; thus, a high-performance data analysis pipeline is needed.[6,7] As such, the development of single-cell analysis techniques requires collaborative innovations among biologists, biochemists, mechanical engineers, and data analysts.

Recent advances in microengineering, biochemistry, and computer science have greatly facilitated the progress of single-cell analysis.[8,9] For example, microengineered wells and droplets have enabled efficient single-cell isolations, serving as high throughput single-cell manipulation platforms.[10–12] The next-generation sequencing (NGS) technology has drastically lowered the sequencing cost, pushing sequencing toward a common analysis practice.[13,14] Emergent computational hardware and algorithms provided easy access to high-volume sequencing data analysis. Consequently, there has been a constant improvement in the performance of single-cell analysis and its biological and clinical applications are rapidly growing.

Seeking answers to biological questions in embryo development using single-cell sequencing requires the understanding of the technology basics as well as the awareness of successful examples. Given the highly interdisciplinary nature of single-cell sequencing, understanding such a technology could be challenging. A few reviews have summarized the technological aspects of single-cell sequencing.[6,15] However, narrating the technology in the applicational context of embryogenesis could potentially facilitate the understanding and adoption in specific research questions. Indeed, a few reviews have introduced the application of single-cell sequencing in germline cell development,[16] cleavage,[17] lineage specification,[18] early embryo characterization,[19,20] and stem cell biology.[21] These works provided great overviews of the practical applications of single-cell sequencing in embryo development-related research areas. However, given the fast progress of these research fields, reviews that summarize the most recent research progress in the technology of single-cell sequencing, as well as its application in each step of embryonic development, have been relatively inadequate. Given the extremely rich transcriptional and epigenetic modifications happening in embryonic development and the resultant cell heterogeneity, single cell analysis can be a powerful tool in sketching the molecular landscape of this important biological process. Such a review could potentially facilitate the adoption of single cell analysis and help solve these research questions. Therefore, this review consolidates the applications of single cell analysis in embryonic development and aims to provide perspectives on the implementation in the technical and biological aspects. As shown in **Figure 1**, we first review the basics of single cell analysis by discussing the available technologies on cell isolation, library construction, sequencing, and data analysis. We then present representative works of utilizing single-cell analysis in embryonic development in processes including preimplantation, peri-implantation, gastrulation, and organogenesis. We conclude the review with remarks on existing challenges and potential research directions.

2. Single-Cell Sequencing Technology

Single-cell analysis generally consists of four steps, namely, single-cell isolation, preparation of DNA library, sequencing, and data analysis. Tissues are first dissociated into dispersed single cells, and then individual cells are isolated to separate reactors, such as tubes, microwells, or microdroplets, and lysed to release the cellular content. Depending on the analyzed targets, such as mRNA or histone modification, specific chemistry is adopted, and a DNA library is then prepared from the released content. When the cell number is small, such as in the analysis of rare cells (e.g., circulating tumor cells) and the study of early embryo development, samples are prepared individually from each cell. However, for the sequencing of a vast number of cells, given that it is practically not feasible to sequence cells on an individual basis and normally a large number of cells need to be sequenced in one batch on the sequencer, it is necessary to incorporate a labeling process and barcode each cell with a unique short DNA sequence, which can be "scanned" and analyzed from the sequencing results. The prepared library is then sent to a sequencer to read the sequences. Finally, the sequencing results are analyzed using bioinformatics techniques, and biological insights are obtained if the analysis is successful. In this section, we present the existing methods of these four steps and discuss some representative works.

2.1. Cell Isolation

The first step in the single-cell analysis is to break down tissue into disengaged single cells and capture/manipulate these single cells. Though these steps may appear straightforward,

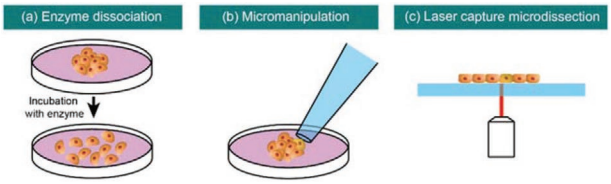


Figure 2. a–c) Cell dissociation techniques adopted in the single-cell analysis as indicated.

cell dissociation is not trivial and may pose a profound effect on the experiment results.[4,23]

2.1.1. Cell Dissociation

Bulk tissues can be broken down into single cells using dissociating enzymes, such as trypsin[24,25] and TrypLE,[26] as shown in **Figure 2**a. These enzymes digest the adhesions or extracellular matrix and detach cells from the tissue microenvironment. However, since the digestion of the extracellular matrix using these enzymes requires incubation at 37 °C, which is also the favorable temperature of other cellular activities, the transcriptome or other molecular signatures of the cells may change upon the enzyme treatment, leading to altered measurement.[4,27]

A few methods have been proposed to alleviate this issue. One method was to use transcription inhibitors (e.g., Actinomycin D and α-amanitin) and preserve the transcriptome landscape before applying the dissociating enzymes.[28] However, the flipside that this method brings about is the prolonged uptake of the inhibitors, during which RNA turnover could happen, altering the RNA content. A second method is to utilize cold-adapted proteolytic enzymes.[29] For example, by using a code active protease from a soil bacterium isolated from Himalayan glaciers, the transcription of cells to be analyzed can be preserved.

In addition to enzymatic methods, single-cell can also be retrieved by physical methods, such as micromanipulation and laser capture microdissection (LCM), as shown in Figure 2b,c. Micromanipulation uses microcapillaries mounted on precision translation stages and applies negative pressure to manipulate single cells under the guidance of an optical microscope. As a standard technique in biological applications such as patch-clamp and in vitro fertilization, micromanipulation provides instant supervision and immediate feedback on cell manipulation, enabling direct monitoring of the cell status. Since a micromanipulator is a commercialized tool and requires little special expertise to operate, it is easy to be set up in different laboratories. Using this technique, the first single-cell sequencing experiment was achieved in 2009.[30,31] Individual blastomeres were picked from four-cell stage mouse embryos by glass capillary pipetting and the single-cell transcriptome was profiled. Other seminal works on single-cell sequencing, such as STRT-Seq,[32] CEL-Seq,[24] SMART-Seq,[33] and scRRBS-Seq,[34] also utilized micromanipulation for cell isolation. In addition, a few other studies on early embryo development, where cell numbers are small, also used this cell isolation method.[35,36]

LCM utilizes a laser beam to cut out the region of interest from a thin layer of solid tissue and extracts the region for downstream analysis. LCM also enables real-time monitoring of the cell retrieving process, which is beneficial for quality control. In addition, LCM is built on an optical microscope, which is standard laboratory equipment, making it easily adoptable. Though LCM has been most commonly used to procure subpopulations of tissue with multiple cells, it has also been adapted to achieve single-cell resolution. For example, LCM-based single-cell RT-PCR was performed on individual CD38+ cells in inflammatory disease in the central nervous system[37] Recently, LCM-seq, an LCM-based full-length mRNA-sequencing technology, was developed to study the transcriptomics of neurons isolated from mouse and human tissues.[38,39] The authors showed that this technology could be adapted to achieve utility down to single captured cells. Despite the laborious experimental process and the resultant low throughput, LCM-seq possessed two major merits: it excise-capture cells without the need for dissociation, which is particularly desirable in application settings where cells are susceptible to enzymes, and it preserves the spatial information of the cells.[40]

Despite the successful implementation of single-cell sequencing based on micromanipulation and LCM, these methods typically could only process tens of cells, which was labor-intensive and limited its wider adoption. Parallel analysis of tens of thousands of cells would enable the characterization of unknown cell types in complex tissues and tremendously facilitate the mechanistic study in cell state-related biology.[41] To this end, high throughput cell capturing techniques, especially those based on microfluidics, have been developed, bringing single-cell sequencing to a new chapter.

2.1.2. Cell Capturing and Indexing

To achieve single-cell sequencing with high throughput, two major challenges must be tackled. First, a large number of cells need to be captured into individual reactors with minimal human intervention. The cell content would then be released into each reactor and utilized for downstream biochemical analysis. Ideally, the cell capturing process should be efficient with minimal cell waste, allowing for the identification of rare cell subpopulations. Second, cells need to be indexed to enable cell identity recovery after pooled sequencing. Though the cost of sequencing has decreased drastically over the last decade, it still scales with the number of sequencing runs. As such, pooling the samples from individual cells and sequencing them in a single batch would significantly reduce the overall cost. In this part, we first present the basics of cell indexing and then discuss the reported cell capturing methods and corresponding indexing techniques for single-cell sequencing.

Cell Indexing: Cell indexing, or cell barcoding, is usually achieved by adding unique DNA sequences (cell barcodes) to the samples during the preparation step of the sequencing samples. After sequencing, the cell barcodes appear in each read and enable the tracing of the cell origin of each read. For the sake of sequencing accuracy, DNA samples are normally amplified through reactions such as PCR to increase the DNA concentration. The amplification process is oftentimes biased,

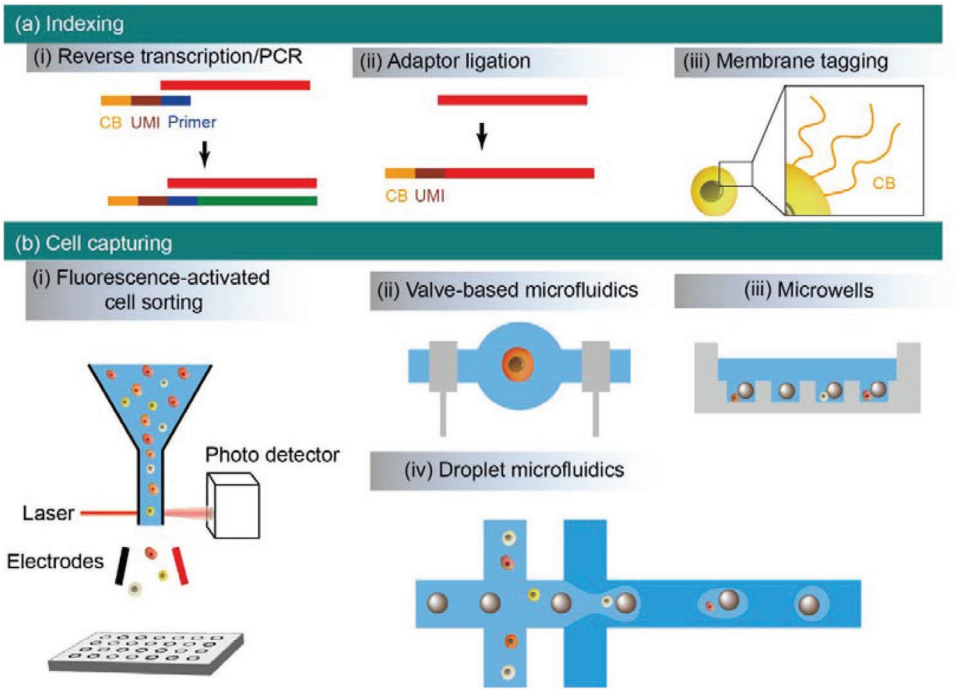


Figure 3. a,b) Techniques for cell indexing and capturing.

which could mask the number of each DNA/RNA in the original samples. As such, unique molecular identifiers (UMIs), which serve as barcodes of each DNA/RNA in the original samples, are oftentimes introduced in the sample preparation steps and facilitate the tracing of each read to the original DNA/RNA. Consequently, in each read of the sequencing data, there exists a cell barcode and a UMI: the cell barcode identifies all the reads from the same cell and the UMI identifies all the reads from the same original DNA/RNA within that cell.

Cell and molecule indexing is normally achieved during DNA or RNA synthesis based on base-pairing in sample preparation steps, as shown in **Figure 3**a. Taking scRNA-seq as an example, the indices can be added through reverse transcription, where each primer is specifically designed to contain poly(T) as a priming site, along with a UMI and a cell barcode. After reverse transcription, the cDNA would contain the UMI and cell barcode. Additionally, the barcode can also be added to the DNA samples through adaptor ligation. With specifically designed adaptor sequences as the indices, DNA samples can be tagged with barcode sequences, which would appear in the sequencing result. Furthermore, another strategy of cell indexing is implemented by tagging plasma membrane with barcoded oligonucleotide sequences, which can be identified in the sequencing data for demultiplexing.[42]

Cell Capturing: Fluorescence-activated cell sorting (FACS): The initial high-throughput single-cell sequencing methods utilized FACS for single-cell capturing, partly due to its wide use and easy access. As shown in Figure 3b, upon cell dissociation, the cell suspension is fed into the FACS instrument and travel in a microchannel, which only allows a single cell to pass through. An electric field is then applied to divert the cells into a designated path based on the fluorescence signal detected. In addition to fluorescence signals, the forward-scattered and side-scattered signals can be simultaneously collected, providing information on the size of the cells and offering the opportunity to exclude clusters of multiple cells. In 2013, Sasagawa et al. reported a scRNA-Seq named Quartz-Seq, which used FACS to sort mouse embryonic stem cells (MESCs) and collected the sorted cells in PCR tubes. Combined with the technical improvement in whole transcriptome amplification, the investigators analyzed up to 96 cells and successfully detected expression heterogeneity between cells with good reproducibility.[43]

Modern FACS can analyze tens of thousands of cells per second, which is suitable for high throughput single-cell analysis. To accommodate the high-speed sorting, the FACS instrument is typically coupled with an automatic translation stage to sequentially direct single cells into individual tubes or wells on microtiter plates. In addition, automated liquid handling platforms are sometimes adopted to reduce the workload of manual pipetting and increase the throughput. For example, Jaitin et al. utilized 384-well plates as reactors for individual cells and Agilent Bravo for the automation of liquid handling and developed a novel scRNA-seq platform named MARS-Seq.[44] The primers for reverse transcription contained UMIs and barcodes, which were then passed to the resultant cDNA and served as cell and molecule index. Based on this platform, they analyzed 1536 single cells from spleens and demonstrated the powerful application of this tool in the investigation of cell diversity. Similar performance was also achieved in other studies based on similar sorting and indexing techniques.[45]

Compared to micromanipulation and LCM, FACS and cell plate-based single-cell sequencing techniques have significantly improved the analysis throughput. However, since each cell requires an individual compartment, the number of cells that can be processed is limited by the number of wells, which could rarely go beyond a few thousand. To address this

problem, Cao et al. developed combinatorial indexing and reported sci-RNA-seq, which showed the capability of analyzing more than 50 000 cells.[46] Cells were sorted into the first cell plate for the first round of barcoding, with each well containing 10–100 cells, before cells were pooled and randomly sorted into the second cell plate for the second round of barcoding. In this way, multiple cells can be processed in the same compartment, which drastically increased the cell numbers. Notably, instead of using FACS to cell dispensing, combinatorial index could be implemented by simply splitting and pooling samples by pipetting, as reported in SPLit-seq by Rosenberg et al.[26] Using this method, more than 156 000 cells were analyzed simultaneously.

Microfluidics is the technology that manipulates fluids and particles on the micron scale, which is also the dimension of most cells, giving it great potential in manipulations on the cellular level.[47] Indeed, tremendous effort has been devoted to developing microfluidics-based technologies for single-cell manipulation in applications such as blood fractionation,[48] circulating tumor cell isolation,[49] single-cell immunophenotyping,[12] etc. Compared to other single-cell manipulation technologies, microfluidics possesses the advantages of automated operation, reduced reagent consumption, and parallel processing with ultrahigh throughput.[50] Among various microfluidic tools, microvalves, microwells, and droplets are relatively successful in the applications of single-cell sequencing.[51]

Microvalves: Microvalves are pneumatic valves integrated into microfluidic devices and offer a means for exquisite fluid control. These microfluidic devices are usually composed of two functional layers, namely a flow-channel layer and a control-channel layer.[52] The channels on the two layers are specifically designed such that flow channels can be shut on/off by pressurizing/releasing certain control channels. A reaction chamber on the flow-channel layer can thus be formed by closing two adjacent valves. By consecutively switching on/off adjacent valves, fluid operations such as peristatic pumping and mixing can be achieved. Using this powerful tool, highly integrated microfluidic devices have been developed for the single-cell genome, transcriptome, and epigenome profiling. In 2011, a microfluidic device was reported to separate and amplify individual chromosomes from a single human cell based on such pneumatic valves.[53] Though the analyzing rate was one cell per device, this work demonstrated the great potential of microfluidic automation in single-cell analysis. Parallelization was later demonstrated with increased complexity in the genomic analysis of single human sperm.[54] In this work, 24 single cells were analyzed in one single device on average, which greatly increased the throughput. With further iteration in both device integration and sample processing,[55] the throughput and sensitivity of this technology were further improved. Despite that, this technology required specialized equipment such as pumps and controllers, and the device fabrication, assembly, and operation required extensive expertise in microfluidics and control, making it very difficult to be adopted by biology laboratories. This problem was alleviated by the commercialization of the technology (Fluidigm), which can analyze up to 800 cells per device with high automation. Applications in single-cell RNA-sequencing[56] and single-cell assays for transposase-accessible chromatin using sequencing of embryonic stem cells with high throughput were also demonstrated,[57] which analyzed 800 and 1632 cells, respectively. Following up on these seminal works, several improvements were made aiming at reducing contamination from cell debris[58] and simplified operation.[59]

The advantage of valve-based microfluidics is that it can simultaneously process hundreds of cells with limited human intervention. Nevertheless, besides the system complexity and high technical barrier, the efficiency of cell capturing is highly dependent on the distribution of cell sizes and studying cell types with different sizes may require different device designs, which increased the fabrication cost. In addition, the valves take up significant portions of the chip space, limiting the number of cell chambers that can be integrated.

Microwells: Microwells were proposed as another cell capturing method to enable single-cell profiling with high throughput. A surface with arrays of microwells, typically with diameters of 30 μm, was fabricated using microfabrication techniques before cell suspension is loaded onto the surface and cells fall into microwells. The incidence of doublets critically depends on the cell density following Poisson distribution. Thus, cell density is normally low to ensure that most wells receive only one cell or no cell at all. Typically, only 10% of the microwells receive single cells. After cell capturing, each microwell with a cell serves as a reactor, and cells inside are lysed to release the molecule to be analyzed. To index the analytes, such as mRNA, two methods have been proposed. Fan et al. developed CytoSeq and loaded barcoded beads into the microwells to capture and index the mRNAs.[60] The beads were functionalized with oligonucleotides which incorporate cell label, UMI, and oligo-dT to capture the polyA tail of mRNA. The diameter of the beads was about 20 μm, which guaranteed that each microwell could only fit in one bead. After mRNAs had hybridized on beads, beads were collected, and reverse transcription, amplification, and sequencing were performed. It was shown that on average 1250 cells could be processed in each experiment, though the throughput could be scaled to 10 000s cells or even more per experiment by simply increasing the size of the microwell array. A similar method, termed Microwell-seq, was developed by Han et al. to map the mouse cell atlas by analyzing more than 400 000 cells from different tissues, showing the great applicability of microwell-based single-cell sequencing technology.[61]

Instead of overlaying the microwell array with cell suspension and having cells randomly sediment into individual wells, another work used a multisample nanodispenser to dispense a droplet of cell suspension into each microwell.[62] After that, microwells were imaged to identify those containing single cells, and further processing would be limited to those wells. Since this method could recognize and actively exclude microwells with doublets, the cell density of the cell suspension could be higher and up to one-third of the microwells contained single cells, which was much higher than methods relying on gravitational sedimentation (≈10%).[60,61] Active dispensing also offered the opportunity to functionalize the bottom of each microwell with a specifically designed oligonucleotide sequence for mRNA capturing and indexing, which eliminated the use of beads. Using this platform, Goldstein et al. characterized more than 1000 human and mouse cells and identified the cell types. Nevertheless, the volume of the microwells in this platform was about 150 nL, which was much larger compared to that of 20 pL

in other works, increasing the reagent consumption and limiting the level of microwell integration.

Droplet Microfluidics: The throughput of valve- and microwell-based platforms were nevertheless limited by the available space on the microfluidic chips. In contrast, droplet microfluidics could generate hundreds to thousands of droplets per second, providing almost unlimited reaction chambers.[63,64] By encapsulating a single cell inside, along with an indexing bead, each droplet serves as a compartment for cell analysis, before the droplets are pooled to perform amplification and sequencing. In 2015, Macosko et al. reported the successful implementation of droplet microfluidics for single-cell transcriptome profiling, named Drop-seq.[41,65] In this work, resin beads were functionalized and linked with oligonucleotide sequences, incorporating PCR handle, cell barcode, UMI, and oligo-dT for mRNA capture. Drop-seq was able to capture 7000 cells for analysis in each experiment on average, which allowed the authors to analyze about 45 000 mouse retinal cells and identify the subpopulations. A major technical hurdle lay in the cell and bead encapsulation process. To obtain reliable analysis results, the encapsulation of cell doublets should be minimal. The encapsulation followed Poisson distribution: to reduce doublets, the cell and bead concentration should be reduced. Consequently, the encapsulation efficiency was compromised and typically less than 10% of the original cells were effectively encapsulated, which could be problematic for experiments where cell utilization is vital. To address this problem, inertial microfluidics was utilized to line up cells and beads before entering the droplet generation region.[66,67] The specifically designed spiral and serpentine channels generated secondary flows, which forced cells and beads to order on a line with regular intervals instead of being distributed across the channel randomly. As a result, the incidence of multiple encapsulations was reduced, and the encapsulation efficiency was improved. Aside from improvement from a hydrodynamic perspective, preindexing cells before encapsulation were shown to enable cell overloading in droplets, thus greatly improving the cell utilization rate.[68] Another strategy was to separately encapsulate cells and beads, pair cell-encapsulated droplets with the bead-encapsulated droplet, and merge them. Droplet merging was commonly achieved by sequentially injecting two types of droplets in a flow and using an electric field to destabilize the interface and induce coalescing. Though this method has been successfully implemented for single-cell analysis,[69] sequencing small samples could be challenging since the droplet synchronization steps required a considerable number of cells to flow by. Chung et al. designed microwells within a microfluidic channel to trap a bead-encapsulated droplet and a cell-encapsulated droplet before perfluorobutanol was flowed into the device and triggered merging.[70,71] The designated shape and dimension of the microwells ensured that exactly one bead-encapsulated droplet and one cell-encapsulated droplet could fit into the microwells. The authors showed this platform was capable of single-cell expression profiling of rare samples at a very high yield.[72] In another study, to lower the equipment demand of Drop-seq, a low-cost microfluidic instrument was reported.[73]

Another seminal work on droplet microfluidics-based single-cell analysis used hydrogel microspheres in lieu of resin bead for cell barcoding.[74] Similarly, hydrogel microspheres were linked with cell barcode, UMI, and oligo-dT, along with a photocleavable spacer for the release of primers. Since the hydrogel microspheres were intrinsically soft and deformable, they could be closely packed in the injecting channel, which allowed for microsphere feeding at a regular interval and ensured that nearly all droplets contained a single microsphere. Using inDrop, the authors achieved cell encapsulation rates up to 12 000 per hour and investigated the cell heterogeneity of mouse embryonic stem cells after the withdrawal of leukemia inhibitory factor by profiling over 10 000 cells. Using a similar principle, Zheng et al. developed a more standardized platform with a much faster gel loading speed and integrated 8 channels in a single chip, allowing for the simultaneous processing of 8 samples and capturing of thousands of cells in ≈6 min in each channel with a capturing efficiency of ≈50%.[75] Nevertheless, these platforms still required the coencapsulation of a bead with a single cell, which could only be achieved by limiting dilution of cells and compromising the throughput. To address this problem, Datlinger et al. adopted combinatorial indexing[46] and preindexed the cells before performing cell/bead encapsulation, which permitted the encapsulation of multiple cells in a single droplet.[68] As a result, cells could be loaded at a much higher concentration and the throughput was improved to 15-fold. Using this method, the authors achieved the analysis of up to 150 000 cells per channel, which demonstrated its great potential in massive-scale single-cell analysis.

In addition to transcriptome analysis, droplet microfluidics has also been applied in genome and epigenome analysis. For example, Lan and colleagues developed single-cell genomic sequencing (SiC-seq), which used droplet microfluidics to first generate barcode droplets and cell-encapsulated hydrogel microspheres, before the two types of droplets were merged along with PCR mix in another microfluidic device.[76] Since the hydrogel microspheres were permeable to a wide range of molecules, this encapsulation strategy permitted more flexible sample processing while keeping each genome confined in a compartment. The authors demonstrated that SiC-seq could process more than 50 000 cells per run in a few hours. In another study, Rotem et al. implemented chromatin immunoprecipitation followed by sequencing (ChIP-seq) using droplet microfluidics for high throughput single-cell chromatin profiling, termed Drop-ChIP.[77] Cells were first encapsulated in droplets along with a solution containing lysis buffer and micrococcal nuclease, which lysed the cells and cut accessible linker DNA. The cell-containing droplets were then merged with droplets containing unique barcodes for labeling, before the combined contents were immunoprecipitated, amplified, and sequenced. Since only 1152 unique barcodes were designed in this study, each collection was limited to 100 cells to ensure that less than 5% of cells shared the same barcode. However, the overall throughput was not compromised since a "collect index" could further be added before sequencing.

2.2. Library Preparation

After cells have been isolated, cells are lysed and DNA libraries can be prepared for the subsequent step of sequencing. Depending on the experimental goal, libraries of different

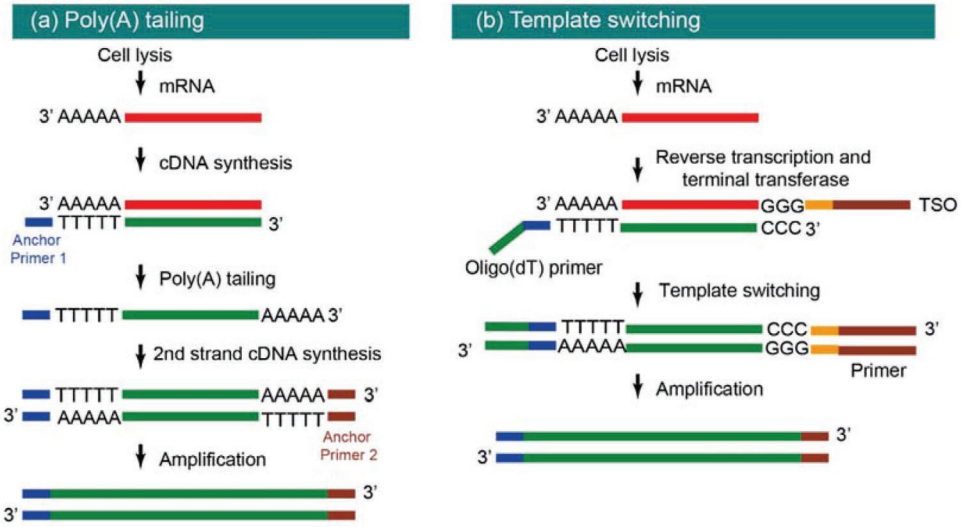


Figure 4. a,b) Library preparation of transcriptome profiling. Library preparation techniques for single-cell transcriptome profiling, including poly(A) tailing and template switching, as indicated. a) Reproduced with permission.[30,81] Copyright 2009, Springer Nature. b) Reproduced with permission.[30,81] Copyright 2014, Springer Nature.

cellular aspects can be prepared following specific protocols.[78] In the study of embryonic development, transcriptome and epigenome are commonly assessed to reveal the status of gene expression. Transcriptome reveals the status of gene expression and thus provides critical knowledge on tissue differentiation. Epigenome unveils the chemical changes to the chromosome and thus the underlying machinery of gene expression. Epigenome mainly includes the status of histone modification, DNA methylation, and chromatin accessibility. In this section, we discuss the protocols that have been reported for single-cell transcriptomic and epigenomic analysis, respectively.

2.2.1. Transcriptome

The library of single-cell transcriptome can be prepared using the technology of single-cell RNA sequencing (scRNA-seq).[79] The library preparation in scRNA-seq usually consists of three steps, namely, mRNA capturing, cDNA synthesis, and cDNA amplification.

In mammalian cells, mRNA makes up only 5% of the total RNA, with the majority being rRNA and tRNA, making the enrichment of mRNA very challenging. Fortunately, in eukaryotes, mRNA usually consists of a poly(A) tail, which can be targeted for mRNA capture. Indeed, most scRNA-seq chose poly(A) as the priming site for mRNA capturing and subsequent reverse transcription. However, given the low abundance of mRNA, the transcript capturing follows Poisson sampling, leading to the observation that only 10–20% of the transcripts are captured by the primers and reverse transcribed into first-strand cDNA.[80] cDNA can be synthesized using two different approaches, namely, poly(A) tailing and template switching. In poly(A) tailing, after reverse transcription, poly(A) tails are added to the 3′ end of the first-strand cDNA, which serves as the priming site for the second strand cDNA (**Figure 4**a).[30,44] The shortcoming of poly(A) tailing is that the polyadenylation sometimes happens before synthesis of the first-strand cDNA has reached the end of the mRNA, resulting in 3′ bias. To achieve full-length cDNA synthesis, template switching has been adopted. Template switching utilizes a special reverse transcriptase, namely, Moloney murine leukemia virus (MMLV) reverse transcriptase, and a template-switching oligonucleotide (TSO), as shown in Figure 4b. MMLV reverse transcriptase adds a few nucleotides as it reaches the end of the mRNA during the synthesis of first-strand cDNA, and TSO uses these newly added nucleotides as the anchoring site and serves as the template to synthesize the additional sequences on the first-strand cDNA, as shown in Figure 4b. Consequently, the cDNA contains the full length of the mRNA, and it avoids the underrepresentation of the 5′ end of mRNA.

The resultant cDNA needs to be amplified before being sequenced. The polymerase chain reaction is commonly adopted for cDNA amplification. However, PCR is a nonlinear amplification process, and the amplification bias is drastically amplified after tens of amplification cycles. This issue can be alleviated by using a different amplification strategy of in vitro transcription (IVT). IVT generates multiple copies of mRNAs based on the synthesized cDNA linearly and avoids bias amplification. The disadvantage is that IVT requires an additional step of reverse transcription, leading to an extra 3′ bias.

2.2.2. Epigenome

Histone modification can both promote and repress gene expression. For example, trimethylation of H3K4 marks the start sites of active genes, while trimethylated H3K9 indicates inactive genes.[82] Histone modifications are commonly detected using chromatin precipitation (ChIP) and have been adapted to achieve single-cell epigenomic profiling by combining droplet microfluidics (**Figure 5**a)[77] and combinatorial indexing.[83] In addition, a method named Cleavage Under Targets & Tagmentation (CUT&Tag) was reported to perform single-cell epigenomic profiling.[84] Briefly, antibodies were used

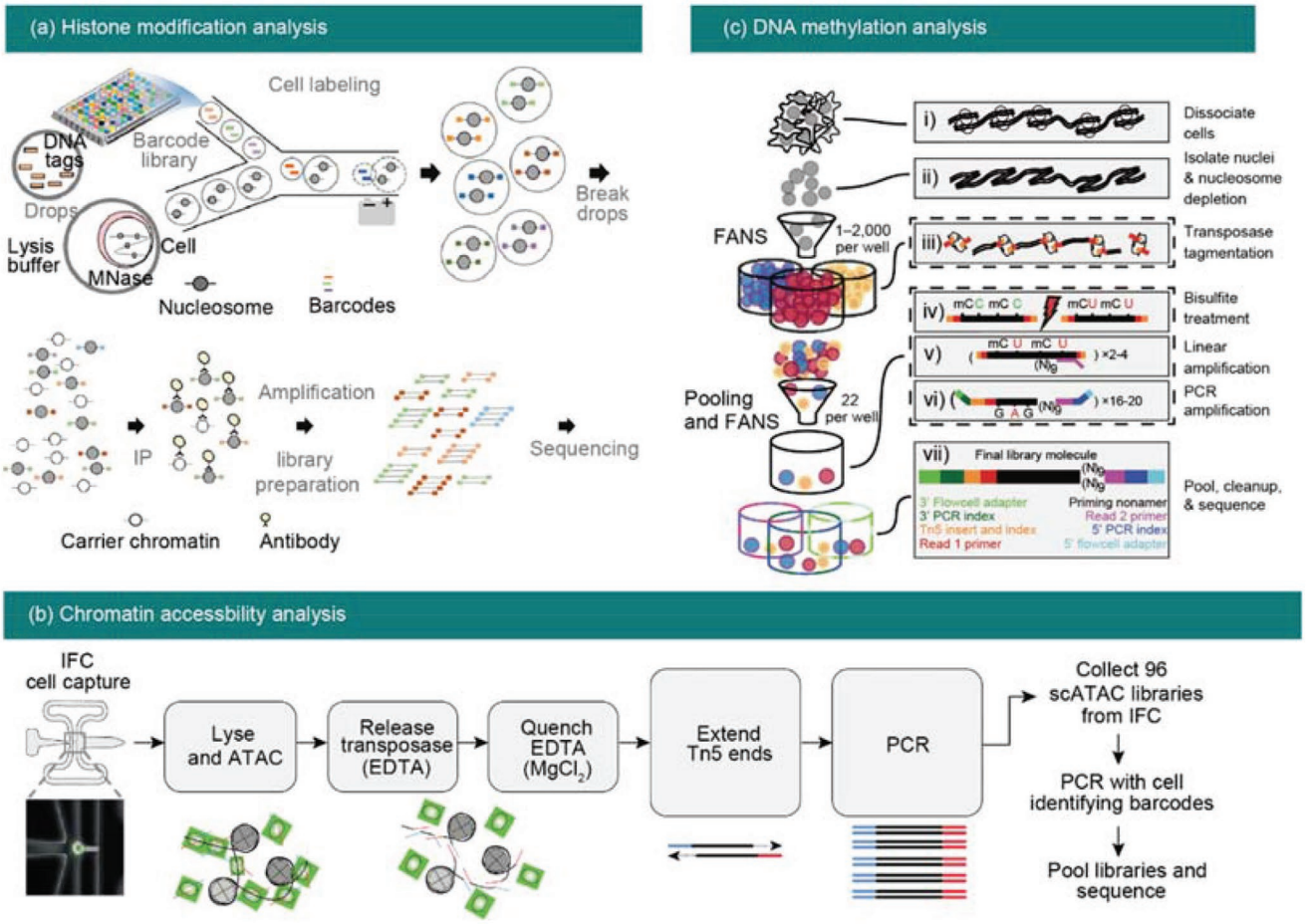


Figure 5. a–c) Library preparation of epigenome profiling. Library preparation techniques for single-cell epigenome profiling. a) Reproduced with permission.[57,77,86] Copyright 2015, Springer Nature. b) Reproduced with permission.[57,77,86] Copyright 2015, Springer Nature. c) Reproduced with permission.[57,77,86] Copyright 2018, Springer Nature.

to target specific chromatin protein, and Protein A-Tn5 transposase fusion protein was subsequently supplemented to cleave the chromatin. Other methods based on a similar principle but different indexing strategies have also been reported.[85]

Chromatin accessibility is usually measured based on the differential susceptibility to enzymatic cleavage or methylation in accessible and inaccessible regions.[87] Enzymes, including DNase, hyperactive transposase (Tn5), and Micrococcal nuclease (MNase), preferentially cleave the accessible regions, generating a DNA library predominantly composed of accessible regions. The library is subsequently amplified and sequenced. Methods based on these enzymes were named DNase-seq, ATAC-seq, and MNase-seq, respectively, and single-cell accessibility has been successfully measured by integrating barcoding strategies, as shown in Figure 5b.[57,88,89] Another method, named nucleosome occupancy and methylome sequencing (NOMe-seq), uses GpC methyltransferase to methylate accessible GpC sites without cleaving the open regions, and the subsequent bisulfite conversion leaves marks in the sequencing results. This method has also been adapted to achieve single-cell resolution.[90]

DNA methylation plays a significant role during embryonic development.[91] DNA methylation is usually detected by bisulfite sequencing (BS-seq), where nonmethylated Cs are converted to U by sodium bisulfite and subsequently sequenced as T. The methylome can then be analyzed from the sequencing data. To achieve single-cell methylome analysis, high throughput cell sorting and combinatorial indexing have been implemented with BS-seq, as reported in a few recent works, as shown in Figure 5c.[34,86,92,93]

2.3. Sequencing

The rapid progress in the study of single-cell analysis was also attributed to the dramatic progress that has been made in the sequencing technologies, with the throughput constantly improving and the cost dropping down. As reported by the National Human Genome Research Institute, the cost of sequencing a human genome was 100 million US dollars in 2001 and dropped to 10 million US dollars in 2007.[14] After that, owing to the emergence and wide application of next-generation sequencing technologies, the cost per human genome has dropped drastically. In 2019, the cost per human genome was as low as 1000 US dollars, enabling DNA sequencing as a routine laboratory and clinical testing tool.[94]

The initial sequencing technology adopted chain-termination sequencing, which is also known as Sanger sequencing, as shown in **Figure 6**a. Briefly, di-deoxynucleotides (ddNTPs),

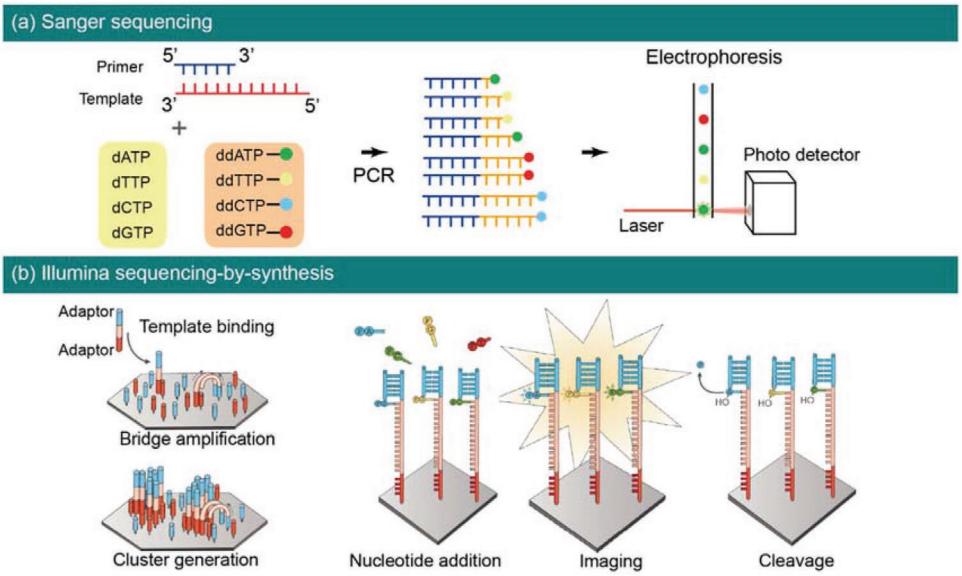


Figure 6. a,b) Sequencing methods. Representative sequencing techniques as indicated. Reproduced with permission.[13] Copyright 2016, Springer Nature.

which are modified deoxynucleotides that terminate DNA strand elongation in chain polymerase reaction, are utilized in the amplification of single-stranded DNA template and randomly incorporated in the elongation, resulting in multiple copies of the single-stranded DNA template with different lengths. The four ddNTPs, namely, ddATP, ddTTP, ddCTP, or ddGTP, are usually labeled with different fluorescent dyes. The resultant DNA fragments are then separated based on length using gel electrophoresis in a single lane capillary gel, and the fluorescent signals were detected and converted to sequence data. Though Sanger sequencing has made a significant contribution in the initial works related to DNA sequencing, a typical throughput was only about 0.032 Mb h^{-1},[95] which hindered the wide application of sequencing in fundamental research and clinical diagnosis.

NGS performs sequencing in a massively parallel manner, wherein millions to billions of DNA fragments can be read in a single run, significantly increasing the throughput and lowering the cost. NGS technologies normally start with the fragmentation of the sample DNA into appropriate length, typically tens to hundreds of base pairs depending on the particular technologies, before adaptors are ligated to the DNA fragments. After that, there are two major steps followed, namely, template amplification and sequencing. The aim of template amplification is to locally generate thousands of identical copies of DNA fragments and increase the signal in the following steps, and strategies include emulsion PCR, bridge amplification, template walking, and rolling circle amplification. For sequencing, there are mainly two categories, namely, sequencing-by-ligation (SBL) and sequencing-by-synthesis (SBS). SBL utilizes DNA ligase and fluorescently labeled DNA probes with known bases to hybrid the template. Upon a complimentary binding and ligation, imaging is performed, and the bases are identified. In contrast, SBS uses polymerase and fluorescently labeled dNTP to synthesize a complimentary strand, wherein the addition of dNTP is imaged, and the base calls are made. Many NGS platforms, such as SOLiD, BGISEQ, GeneReader, and Illumina, have shown success in sequencing with their own pros and cons. Here, we briefly introduce the technology of Illumina as an example, given that Illumina is the most commonly used platform.

In Illumina sequencing, purified DNA samples are first cut into short segments using transposases and tagged with adaptors on both ends, before additional motifs, including sequencing primer binding sites, indices, and terminal sequences, are added. The prepared samples are then introduced into the flow cell, and the tagged DNA fragments were attached to the oligos coated on the flow cells (Figure 6b). DNA fragments are then clonally amplified through bridge amplification, in a process called cluster generation. After that, DNA fragments are sequenced with SBS technology as aforementioned. After primer binding, fluorescently labeled dNTPs are added to the flow cell and pair with the template. dNTP additions lead to the detection of a characteristic fluorescent signal, which is then detected by imaging and used for the base call. The fluorescent dye is then cleaved, and the cycle is repeated. The sequences of the fragmented DNA are then further processed to generate the sequence of the original DNA samples.

2.4. Data Analysis

To make sense of the sequencing data, a few steps are usually followed, including quality check (QC), read alignment, normalization, and interpretation, as shown in **Figure 7**.[96] We talk about each step in the following section.

2.4.1. Quality Check

Raw data files from typical sequencers, including NextSeq, HiSeq, and NovaSeq, are in binary base call (BCL) format and then converted to FASTQ format files using conversion software, such as Illumina bcl2fastq and bcl2fastq2 Conversion

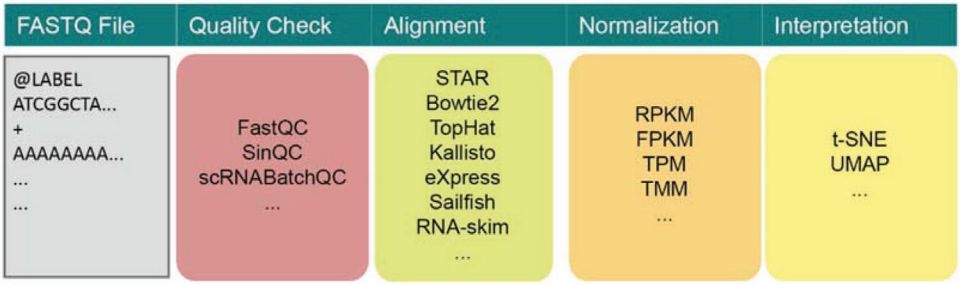


Figure 7. Typical data analysis pipelines for single-cell sequencing and available tools.

Software. These FASTQ files will then go through a QC to remove low-quality bases (usually at the 3′ end) and sequencing adapters. Among QC tools, FastQC is a widely used one for the assessment of quality distributions over the entire reads. It provides a simple way to check raw sequencing data coming from high throughput sequencing pipelines. Jiang et al. developed a single-cell RNA-seq Quality Control (SinQC) method and software tool to detect artifacts in sequencing samples by integrating both gene expression patterns and data quality information.[97] In another work, Liu et al. developed scRNABatchQC and enabled the simultaneous comparison of multiple sample sets over different biological and technical features. The authors also demonstrated that scRNABatchQC was capable of identifying and characterizing sources of variability in single-cell transcriptome data.[98]

2.4.2. Read Alignment

After the quality check, the resultant clean reads will undergo alignment. Read alignment is the process of mapping reads to a reference genome or transcriptome. STAR is a conventional RNA-seq aligner that finds the longest possible sequence matching one or more sequences in the reference genome.[99] Bowtie2 is another full-text minute index alignment approach, which combines the strength of full-length index and dynamic programming algorithms to achieve high speed, sensitivity, and accuracy to find long, gapped alignments.[100] It is particularly efficient at aligning to relatively long genomes and aligning reads of about 50 to 1000 characters. TopHat and TopHat2 use Bowtie2 to align spliced reads, and they look into the unmapped reads and try to align using the information of splicing junctions.[101,102] These alignment-based methods are conceptually simple, but the read-alignment step can be time-consuming and computationally intensive despite recent advancements in fast read aligners. Kallisto is a pseudoaligner that breaks reads into chunks of sequences called K-mers, which are then mapped to a reference transcriptome.[103] Compared to the traditional aligner, pseudoaligners are generally faster, making them conducive for single-cell studies. Compared to Kallisto, eXpress uses a streaming algorithm with linear run time to determine abundances of sequencing data in real time.[104] Patro et al. developed a lightweight method for quantifying transcript abundance from RNA-seq read, namely, Sailfish[105] and Salmon.[106] Since Sailfish avoids read mapping, which is a time-consuming step, it provides much faster quantification estimates without losing accuracy. Salmon is another k-mer counting software that learns and corrects sequence-specific and GC biases.[106] It combines a new dual-phase parallel inference algorithm and feature-rich bias models with an ultrafast read mapping procedure. RNA-Skim, another RNA-seq quantification method, partitions the transcriptome into disjoint transcript clusters based on sequence similarity and introduces the notion of sig-mers, which are a special type of k-mers.[107] Among all these methods, traditional alignment-based and alignment-free quantifications methods perform similarly for common gene targets, though alignment-free quantifications methods are much faster especially for long RNAs.

2.4.3. Normalization

After alignment, clean reads with high mapping quality will be considered for the generation of the gene expression matrix. Normalization is an essential step to remove cell-specific bias, which enables accurate comparisons of expression levels between and within samples. In protocols where (UMIs are utilized, UMIs are first used to correct amplification-related bias, since reads with the same UMI sequence are most likely from the same RNA molecule. Therefore, after the read alignment, UMIs provide a means for molecule counting and quantification of gene expression.[80,108] In addition, there are two widely used normalization approaches, namely, within-sample normalization (WSN) and between-sample normalization (BSN). WSN allows the quantification of expression levels of each gene relative to other genes in the sample. The most widely used metrics include reads per kilobase million (RPKM),[109] fragments per kilobase million (FPKM),[109] and transcripts per kilobase million (TPM).[110,111] RPKM is defined to compare experiments or different samples so that additional normalization is integrated into the denominator term, which is expressed in million. However, these WSN approaches are not sufficient when detecting differentially expressed genes.

To address this problem, BSN categorizes the sequencing results based on sequencing depth instead of all the reads. The most commonly used BSN approaches include trimmed mean of M-values (TMM),[112] relative log expression (RLE),[113] and upper quartile (UQ).[114,115] TMM is a scaling normalization method for differential expression analysis of RNA-seq data, and it is based on the hypothesis that most genes are not differentially expressed. First, TMM picks a reference sample, and the other samples are considered test samples. For each

test sample, the most expressed genes and the genes with the largest log ratios are first excluded, and then TMM is computed as the weighted mean of log ratios between the test and reference samples. RLE is another normalization method similar to TMM based on the same hypothesis regarding differential expression. The RLE scaling factor is calculated as the median of the ratio of its read counts to its geometric mean across for each gene in all the samples. UQ is another normalization method that is calculated by applying the upper quartile of 0.75 to the gene counts of all runs. In this method, it first removes the genes that have zero read counts for all samples, and the remaining gene counts are divided by the upper quartile of counts and multiplied by the mean upper quartile across all samples of the dataset.[115] Among all the different RNA-seq normalization methods, the commonly used ones perform well in simple normalization tasks. However, since the selection of variable genes is highly sensitive to normalization methods, it may affect the downstream data analysis, including dimensionality reduction and clustering analysis. The potential for combining SWN and BSN methods has recently been explored by Risso et al.,[116] recommending the use of within-sample GC-content normalization in combination with BSN.

2.4.4. Interpretation

After preprocessing of the raw data, the normalized data is ready for further analysis. However, interpreting single-cell RNA sequencing data sets can be challenging given that each cell is represented by over 30 000 genes. To identify cell types, cell states, or development lineages, each data point need to be compared with others. Since single-cell datasets are normally high dimensional in a large number of individual cells, dimensionality reduction is typically the next step after the raw count matrix is normalized. Researchers have been developing different algorithms to reduce the dimension by projecting high-dimensional data into low-dimensional space. Principal component analysis (PCA) is a widely used linear projection method. By projecting the high-dimensional datasets into subspace with fewer columns, it is easier to visualize the samples with interpretations. The disadvantage of linear dimensionality reduction methods is that they normally cannot reflect the complex structures of single-cell sequencing data in low-dimensional spaces. Compared to linear projection methods, nonlinear methods, such as t-distributed stochastic neighbor embedding algorithm (t-SNE),[117,118] showed better results in many aspects including maintaining similarities in local neighborhoods of data, which may be important to the overall data structure. t-SNE is a manifold-based dimensionality reduction algorithm, and it projects the data sets based on similarities, which were measured in their gene expressions. After t-SNE, similar cells are plotted close to each other and dissimilar cells are far apart in the lower-dimensional space. Several additional projection methods, such as PHATE,[119] scvis,[120] uniform manifold approximation and projection (UMAP),[121] and single-cell interpretation via multi-kernel learning (SIMLR),[122] are also utilized for dimensionality reduction. It is worth noting that dimensionality reduction may result in the loss of important biological information, such as the spatial organization of cells and genome networks.

Clustering is the process of grouping cells into different clusters based on the cell types. The goal of clustering is to categorize individuals into different subsets based on their similarities or distance between the data points. Once clustering is completed, the next step is to identify marker genes that are expressed in different clusters. However, the clustering results should be carefully interpreted since the grouped clusters may not be biologically related. Further statistical analysis and validations should be performed to confirm the biological information of the groups.

3. Applications of Single-Cell Analysis

The identity and behavior of a cell are determined by its gene expression network, which is regulated by genetic and epigenetic mechanisms. Hence, deciphering the temporal and spatial patterns of gene expression in embryogenesis is a crucial step toward understanding the early developmental process. In this section, we present studies that used single-cell analysis to extend our knowledge of the transcriptional and epigenetic landscape for early embryogenesis such as preimplantation, gastrulation, and organogenesis.

3.1. Preimplantation

During the first 7 days of human development, the zygote undergoes cellular division and establishes the first three distinct cell types of the mature blastocyst: trophectoderm (TE), primitive endoderm (PE), and epiblast (EPI) (**Figure 8**).[22] In 2013, Yan et al. applied scRNA-Seq to human oocytes and human preimplantation embryos at different stages.[35] They have identified 2733 potential novel long noncoding RNAs (lncRNAs). Many lncRNAs are variably expressed among cells of the same embryo but consistently present in different embryos of the same stage, and this suggests that the lncRNA in human embryos is potentially functional. Also, they found a large set of lineage-specific genes that can discriminate EPI, PE, and TE lineage cells in human blastocysts, which provides insight into the lineage separation of EPI, PE, and TE.[35]

Recently, Petropoulos et al. mapped the transcriptional landscape of human preimplantation and studied lineage specification between EPI, PE, and TE using scRNA-seq (**Figure 9**a).[123] They found that cellular transcriptomes primarily segregated according to the embryonic stage, followed by segregations into lineages (TE–EPI–PE), embryo-to-embryo variability, and subpopulations. To study when and how the divergence of TE/inner cell mass (ICM) and EPI/PE occurs in humans. The authors applied map dimensionality reduction on all cells using lineage-specific genes. Interestingly, blastocyst forms three distinct transcriptional states corresponding to TE, EPI, and PE during E5, after which the segregation (based on lineage-specific genes) did not further increase. During early E5, the cells had activated about half of the TE genes while still maintaining the expression of early EPI genes, indicating an intermediate stage of coexpression of lineage markers and the plasticity of the cells during early E5. This study demonstrated that the segregation of all three lineages occurs simultaneously

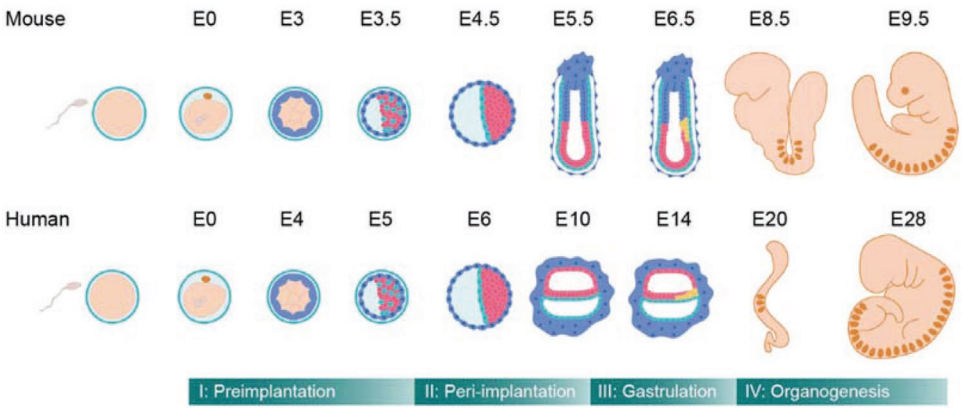


Figure 8. Overview of mouse and human embryo development. Reproduced under the terms of the CC-BY license.[22] Copyright 2021, The Authors, published by Cell Press.

and coincides with blastocyst formation at E5.[123] These results highlight a unique development scheme in human preimplantation embryogenesis, as the model developed from mouse studies suggests that TE and ICM fate is initiated in a positional and cell polarization-dependent manner within the morula, followed by a subsequent progressive maturation of EPI and PE in the blastocyst.

During mouse preimplantation development, distinct cell lineage patterning is first observed based on morphology during the fourth cleavage to generate a 16-cell embryo. At this stage, some blastomeres divide symmetrically, contributing two daughter cells to the outside region of the embryo (TE), whereas others divide asymmetrically and contribute one daughter cell to the outside (TE) and another to the inside region (ICM). Recent studies using scRNA-Seq revealed an early transcriptional symmetry breaking during the first cleavage to generate a two-cell embryo. Biase et al. revealed reproducible interblastomere differences in two-cell and four-cell mouse

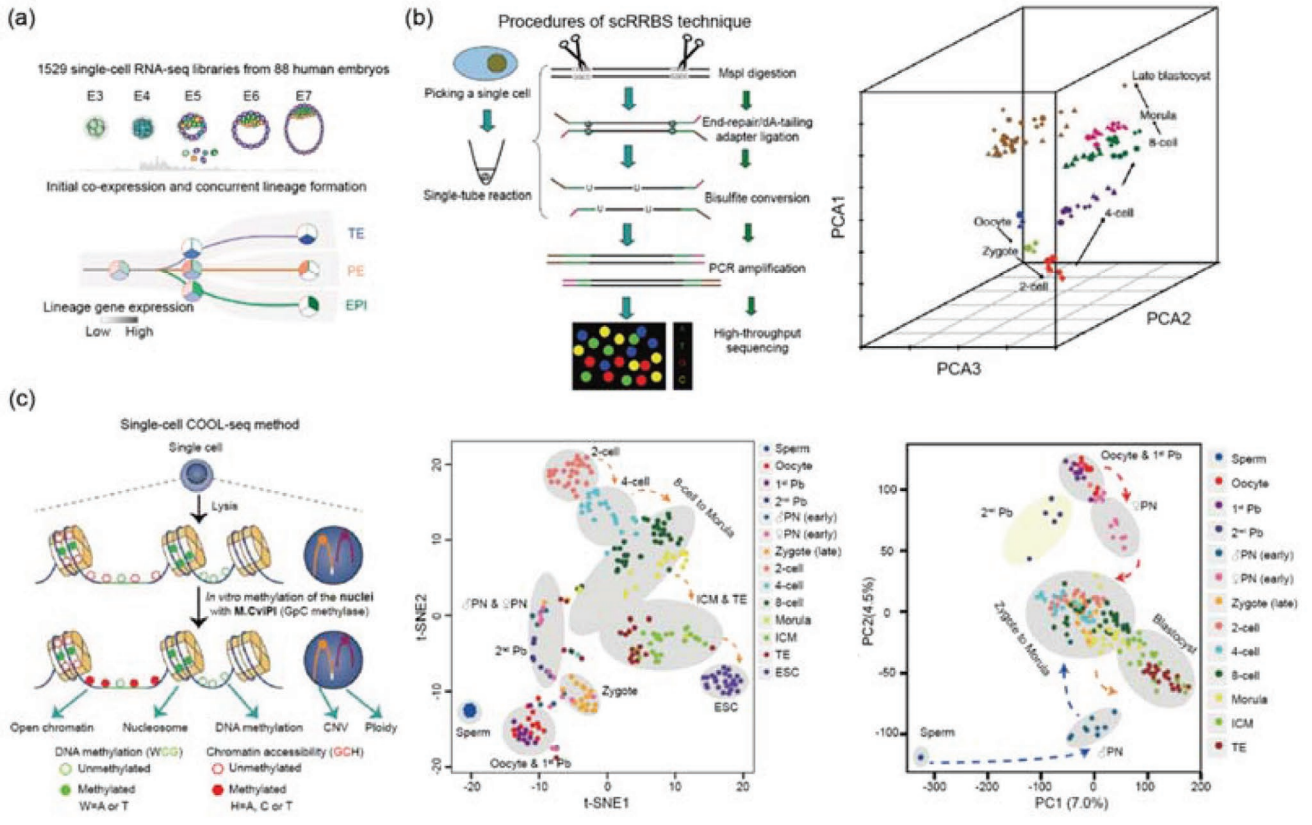


Figure 9. a–c) Application of single-cell RNA-seq technology in revealing the emergency of different cell phenotypes during the preimplantation period. a) Reproduced under the terms of the CC-BY license.[123] Copyright 2016, The Authors. b) Reproduced with permission.[124] Copyright 2015, Springer Nature. c) Reproduced under the terms of the CC-BY license.[125] Copyright 2017, The Authors, published by Springer Nature.

embryos, which is significantly larger than interembryo differences.[126] And the gene pairs with consistent directions of interblastomere differences at the four-cell-stage would exhibit the consistent directions of differences between ICM and TE, suggesting the differential gene expression may be involved in directing ICM/TE lineage specification. Shi et al. showed that the initial blastomere-to-blastomere difference was due to partition error and then zygotic transcriptional activation further elevated this difference. They also demonstrated such transcriptional symmetry breaking also occurs during the first cleavage in human preimplantation embryos.

During embryo development, it is generally believed that two major rounds of epigenetic reprograming including DNA methylation and histone modification occur during the formation of primordial germ cells (PGCs) and in preimplantation development. Different methods have been developed to characterize the epigenetic landscape of preimplantation embryos in single-cell resolution. Burton et al. applied scRNA-Seq to mouse preimplantation embryos and focused their analysis on chromatin modifiers.[127] They found that dramatic changes in expression of chromatin modifiers took place in the early stages of embryo development, from oocyte to four-cell-stage, suggesting drastic epigenetic reprograming occurs earlier. Importantly, they identified Prdm14 as an important modifier that regulates TF/ICM bifurcation. Prdm14 expresses heterogeneously in four-cell-stage embryos and mechanistic studies suggest that expression of prdm14 promotes H3R26me2 and leads the cells to differentiate to ICM.[127]

In 2013, Tang and co-workers developed to map DNA methylation in single-cell resolution based on reduced representation bisulfite sequencing (scRRBS).[34] They have applied scRRBS to analyze mESCs, sperm, oocytes, and zygotes. They found during zygotes development, both male and female pronuclei underwent significant demethylation, with demethylation of male pronuclei more drastic than female pronuclei.[34] Recently, the same group used an improved method, single-cell bisulfite sequencing (scBS-seq) to perform DNA methylome analysis of preimplantation human embryos (Figure 9b).[124] In this method, sequencing adaptors are ligated after bisulfite treatment to minimize DNA degradation. They have identified three major waves of global demethylation and two major waves of de novo DNA methylation. By comparing the genome of the embryo with the sperm donor, they determined the parental origin of each blastomere and found that demethylation is faster in paternal genome than maternal genome at zygote stage, which is consistent with mouse zygote development. By tracing cell division of four-cell state embryo using fluorescence labeling, they found DNA methylome for the two daughter cells showed negative correlation, indicating passive dilution of DNA methylation during cell division and suggesting DNA methylation pattern could be used to trace cell lineage during early development.[124]

To measure all the different layers of epigenetic information from the same cells, Tang's group developed a single-cell multiomics sequencing technology (scCOOL-seq) to simultaneously measure the chromatin state, nucleosome positioning, DNA methylation, and copy number variation (CNV) from the same single cells (Figure 9c).[125] They have applied scCOOL-seq to mouse and human preimplantation embryos. They found distinct chromatin accessibility in parental genomes during preimplantation development: the chromatin of human embryos was more open than mouse embryos. Importantly, by comparing results from scCOOL-seq and scRNA-seq, they found a positive correlation between the chromatin accessibility of promoters and the expression levels of corresponding genes and a negative correlation between DNA methylation and chromatin accessibility of promoters during human preimplantation development. By examining the transcription factor motifs in nucleosome-depleted regions (NDRs), they found, in both mouse and human embryos, the bind motifs of pluripotency and early embryonic regulators showed strong stage-specific enrichment at distal NDRs and some were enriched much earlier than cell fate specification.[125]

3.2. Peri-Implantation

During day 8 to day 12 of human embryo development, the mature blastocyst implants into the uterine wall. This period typically is called peri-implantation. Implantation is one of the most mysterious developmental milestones during mammalian embryogenesis. However, the implantation of the entire conceptus into the maternal endometrium is less understood due to the limited access and observation technology to the embryo early after implantation in vivo. Failure of implantation is a substantial cause of early pregnancy loss. In humans, it is assessed that about 40–60% of conceptions fail, with the majority of the losses occurring implantation. So it is pretty important to examine the lineage specification and corresponding patterns of the transcriptome and DNA methylome during implantation to reveal the underlying mechanisms driving human embryo implantation.

In 2019, Tang and co-workers mimicked the implantation of human embryos and applied scCOOL-seq to examine the gene-expression network and lineage-specific DNA methylation patterns of human peri-implantation embryos at single-cell resolution (**Figure 10**a).[128] They analyzed more than 8000 individual cells from 65 human peri-implantation embryos. They used STRT-Seq to construct single-cell RNA-seq libraries, MALBAC amplification technique to perform single-cell whole-genome sequencing, and single-cell whole-genome bisulfite sequencing (scBS-seq) to analyze DNA methylome, finding there is robust preparation for the establishment of a mother-to-offspring connection during implantation. They also identified the specific genes which are expressed in the EPI, PE, and TE cells. Moreover, according to the analysis of parental allele-specific expression of X-chromosome-linked genes, they found that the ratio of X chromosomes to autosomes in female cells showed a bit higher than that in male cells during implantation. Using the single-cell Trio-seq2 strategy, they analyzed the lineage-specific DNA methylation dynamics during implantation and found that remethylation of the genome in PE lineage was much slower than that in both EPI and TE lineages, suggesting that the embryos initiated DNA remethylation shortly after the blastocyst stage and that PE, EPI, and TE presented obviously distinct and asynchronous DNA remethylation patterns.[128]

Similarly, in 2019, Yuan and co-workers used single-cell RNA sequencing techniques to define the transcriptomic landscape of

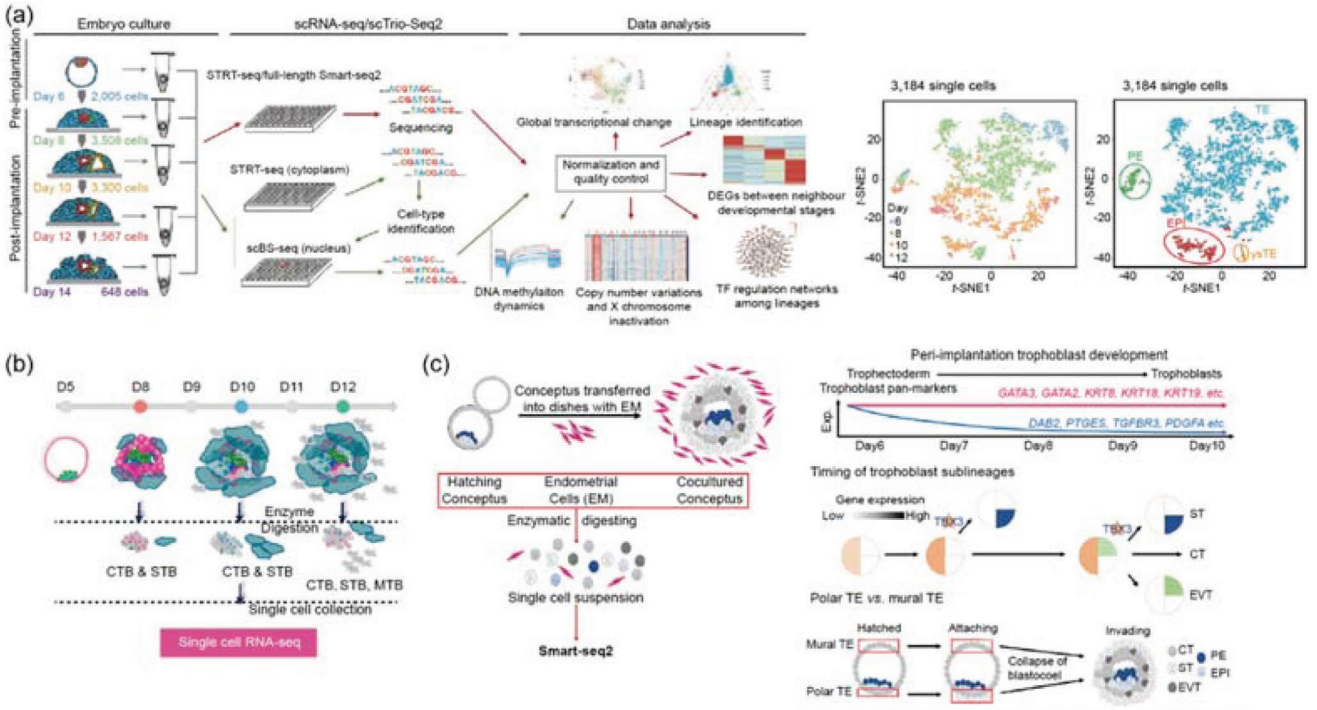


Figure 10. a–c) Single-cell RNA-sequencing transcriptome profiling and trophoblast development during peri-implantation of the embryo. a) Reproduced with permission.[128] Copyright 2019, Springer Nature. b) Reproduced under the terms of the CC-BY license.[129] Copyright 2019, The Authors, published by Nature. c) Reproduced under the terms of the CC-BY license.[130] Copyright 2019, The Authors, published by National Academy of Sciences.

placental trophoblast (TB) from cultured human blastocysts during the implantation period.[129] They shed light on the events driving early placental emergence by examining the transcriptomes of cytoTB (CTB), syncytioTB (STB), and migratory TB (MTB) picked from embryos at increasing days in culture. They found that these three kinds of TB cells emerge in sequencing, with CTB with some properties of STB appearing at D8, while mature STB peaked at D10, and MTB with a mixed MTB/CTB phenotype arose around D10. Subsequently, at around D12, the generation of STB is decreased, CTB institutes a new stage of proliferation, and mature MTB initiates to migrate from the main body of the conceptus (Figure 10b).[129] Recently, the other group (Lv et al.) revealed the underlying regulatory mechanism driving trophoblast fate divergence during the human peri-implantation period using single-cell RNA sequencing.[130] They performed SMART-seq2 assay to analyze the transcriptome of trophoblast cells and identified that T-box transcription factor 3 (TBX3) is essential for mediating the differentiation of cytotrophoblast into syncytiotrophoblast (Figure 10c).[130] Tan et al. found the underlying transcriptional landscape upon ZIKV infection during the pre- and peri-implantation period using single-cell RNA sequencing.[131] They reported that ZIKV infection caused miscarriage and congenital malformations could be ascribed to the ZIKV susceptibility of trophectoderm and neural progenitor cell death.[131]

3.3. Gastrulation

Gastrulation represents a process of embryogenesis, which is defined by the initiation of the primitive streak and subsequent generation of primary germ layers, including definitive endoderm (DE), mesoderm, and (neuro-) ectoderm. During the gastrulation process, the germ layers are coordinated cell division, movement, and rearrangement, guiding mesodermal and/or endodermal progenitors on the inside of the embryo and the ectoderm on the outside, proceeding to fate patterning. Germ layer patterning formation in vivo is synergistically regulated by different developmental signaling pathways, including Nodal/TGFβ, BMP, WNT, and FGF. These signaling pathways are tightly interconnected to dictate cell fate specification, tissue patterning, and rearrangements.

In 2020, Moskowitz and co-workers used single-cell RNA sequencing techniques to reveal the underlying mechanism driving the formation of anterior mesoderm (AM) patterning during gastrulation, founding Hedgehog (Hh)-fibroblast growth factor (FGF) signaling axis is required.[132] They performed transcriptional profiling and drop-seq to interrogate anterior–posterior axis patterning in a mesoderm-specific Hh pathway mutant, observing selective anterior mesoderm defects in mouse embryo development. They concluded that Hh signaling is required for FGF pathway activity in nascent mesoderm during gastrulation, revealing a previously uncharacterized role of Hh signaling in the development of specific anterior mesoderm lineages. In 2019, Meissner's and Weissman's groups reported a flexible, high-information, multichannel molecular recorder with a single-cell readout approach to assemble mouse cell-fate maps. They revealed the developmental relationship between different tissue types through integrating lineage information and scRNA-seq.[133] In the same year, Göttgens and co-workers reported the transcriptional profiles of 116 312

single cells with a median of 3436 genes detected per cell from mouse embryos and constructed a molecular map of cellular differentiation from pluripotency toward endoderm and hematoendothelial lineages. They also highlighted where TAL1 is critical for progression into the blood lineage.[134] Similarly, in 2021, Arnold and colleagues combined scRNA-seq, genetic fate labeling, and imaging approaches to reveal the precise spatiotemporal pattern of the emergence of AM and DE progenitors during germ layer formation of mouse embryo gastrulation.[135] They found the separation of AM and DE lineages from *Eomes*-expressing cells. In 2016, Sun and co-workers utilized scRNA-seq, conventional bisulfite sequencing, and Tet-assisted bisulfite sequencing (TAB-seq) technologies to analyze the epigenetic modification, including cytosine methylation by DNA methyltransferases (DNMTs) and demethylation caused by oxidation of 5-methylcytosine by the Ten-eleven translocation (Tet) family, revealing that TET-mediated demethylation and methylation regulate Lefty-Nodal signaling to dictate primitive streak patterning during mice embryo gastrulation.[136] Furthermore, in 2020, Meissner and co-workers performed scRNA-seq to simultaneously recover robust morphological and transcriptional information of many mutant mouse embryos during the gastrulation period.[137] They found polycomb repressive complex (PRC) 1 and 2 components are considerable cooperativities, while PRC2 dominates in restricting the germline. In 2017, Reik and co-workers described regulatory processes associated with lineage commitment during mouse embryonic development from implantation to early gastrulation at single-cell resolution (**Figure 11**a).[138]

In primates, including humans and nonhuman, gastrulation remains a mystery since accessing primate embryos is difficult at this stage. In 2020, Li and co-workers developed a 3D in vitro human blastocyst-culture system to mimic developmental milestone and 3D architectures of the embryonic disc, amnion, and formation of primitive streak anlage (Figure 11b).[139] They characterized the regulatory network driving the segregation of epiblast, primitive endoderm, and trophoblast using single-cell transcriptome profiling. In parallel, in 2019, Wang and co-workers also established an in vitro culture system to support the development of the cynomolgus monkey embryo beyond early gastrulation.[141] They recapitulated the segregation of epiblast and hypoblast, the emergence of the primordial germ cells, and formation of the anterior–posterior axis, and explored the characteristics and mechanisms driving lineage specification during embryo postimplantation.

Recently, Neveu and co-workers designed a computational framework, i.e., MorphoSeq to generate a spatiotemporal atlas of gene expression at the single-cell level and classify the single cell of the individual embryo into cell types without prior knowledge (Figure 11c).[140] They reconstructed the genome-wide gene

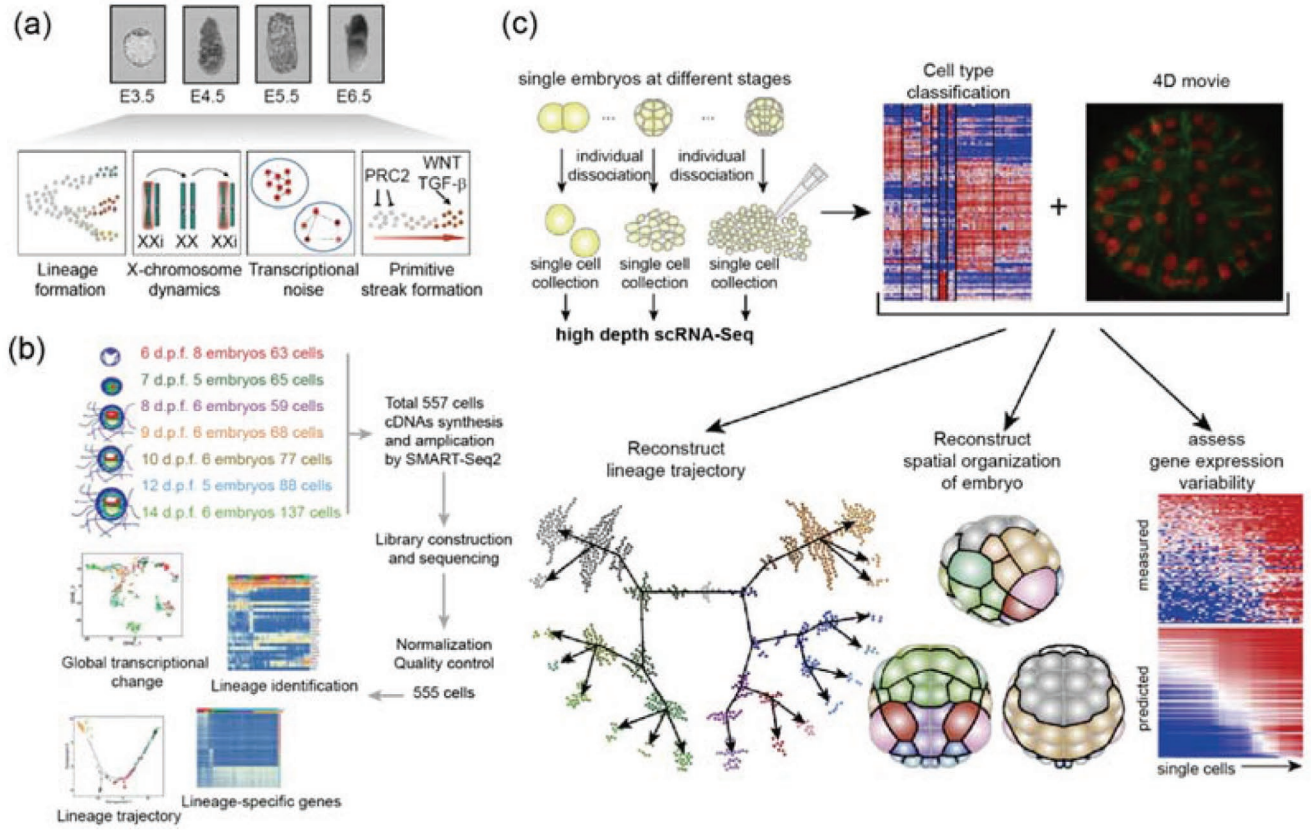


Figure 11. a–c) Lineage delineation by transcriptome using single-cell RNA-seq during gastrulation period. a) Reproduced under the terms of the CC-BY license.[138] Copyright 2017, The Authors. b) Reproduced with permission.[139] Copyright 2019, Springer Nature. c) Reproduced under the terms of the CC-BY license.[140] Copyright 2020, The Authors, published by Nature.

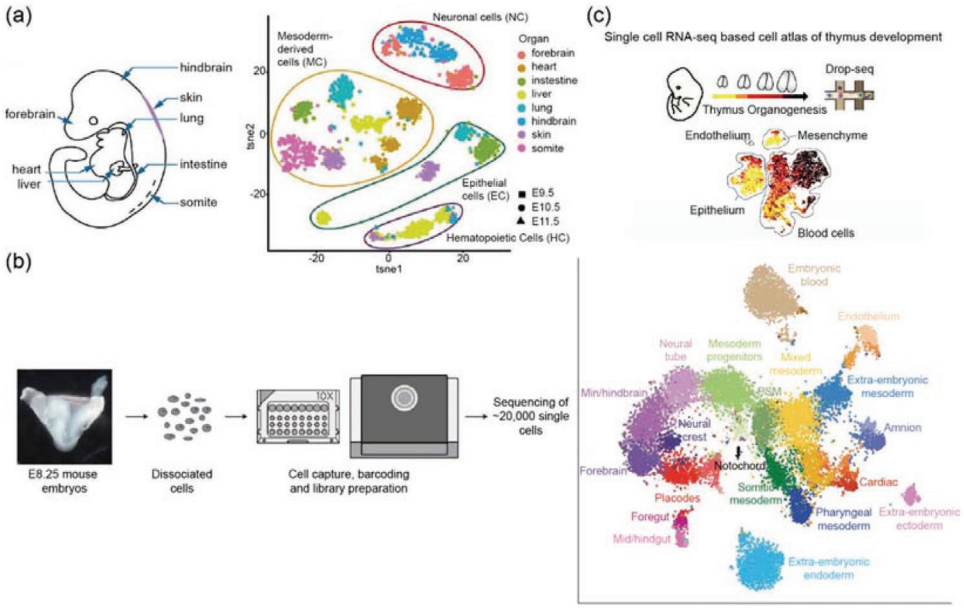


Figure 12. a–c) Application of single-cell RNA-seq technology on identifying cell types during organogenesis. a) Reproduced under the terms of the CC-BY 4.0 license.[146] Copyright 2018, The Authors, published by Nature. b) Reproduced with permission.[147] Copyright 2018, Springer Nature. c) Reproduced with permission.[148] Copyright 2018, Elsevier.

expression trajectory of every single cell in the 18 lineages up to gastrulation in the ascidian *Phallusia mammillata*, mapping physical position of each cell and lineage history.

3.4. Neurulation

Neurulation begins soon after the gastrulation of the embryo, which is essential for the formation of the neural tube and subsequent development of central nervous system. After neural development, the neural plate folds and fuses to form the neural tube, which is directly dictated by the morphogens secreted from surrounding tissues and biophysical cues from extracellular matrix.[142] These extracellular inductive cues are sensed and transduced through intracellular signaling pathways to mediate genetic networks activation, including activation of transcriptional factors, DNA methylation and demethylation, which guides the fate specification of progenitor cells within the neural tube.

Recently, Brivanlou and co-workers generated a neuruloids structure, which recapitulates early human neurulation, through micropattern technology.[143] Then, they uncovered the precise identities and timing of fate specification during neurulation using scRNA-seq technology. They tested over 5105 single cells across two independent experiments, revealed gene expression landmarks of neural, neural crest, sensory placodes, and epidermis, and identified over 100 genes with an expression pattern specific to each population. Furthermore, they unraveled the molecular mechanisms driving the morphogenetic defect of Huntington's disease (HD) through analyzing scRNA-seq data of neuruloids in wild type and 56CAG background, founding the downregulation of WNT/PCP signaling pathway and dramatically decrease in the expression of cytoskeleton-associated genes and actin-myosin contraction.[143]

In 2018, Xu and co-workers investigated the genome-wide transcriptome profile of single cells to uncover cell fate decision and subpopulation specification during human neurulation using scRNA-seq and ATAC-seq.[144] They identified putative novel transcription factors and dissected the dynamics of chromatin accessibility during neural differentiation stages. Together, all these researches pave the way for a deeper understanding of the underlying mechanisms of neurulation, cell fate specification, and neuronal disease.

3.5. Organogenesis

After gastrulation, three germ layers further develop into different organs, and embryos expand from hundreds of cells to millions of cells during organogenesis. scRNA-seq provides a powerful strategy to map the transcriptional landscape at the single-cell level during early organ development. Trapnell's and Shendure's groups developed a high-throughput sequencing method, single-cell combinatorial-indexing RNA-sequencing analysis (sci-RNA-seqs), and analyzed over 2 million cells from different organogenesis-stage mouse embryos and recovered a median of 671 UMIs.[145] By subjecting single-cell transcriptome to Louvain clustering, they identified 39 major clusters. They developed and used monocle 3 to identify major developmental trajectories along mouse organogenesis. They found many trajectories are more complicated than a simple linear path, such as branching path and certain cell types to be generated from multiple origins, highlighting the complexity of organogenesis.[145]

In 2018, Tang and co-workers utilized scRNA-seq technology to examine the transcriptomic features of ≈2000 individual cells from eight organs and tissues from seven mouse embryos between E9.5 to E11.5, investigating the evolutionary

and developmental relationships among various organs and cell types during mouse organogenesis (**Figure 12**a).[146] They applied the SCENIC algorithm to map gene regulatory networks, construct regulon matrix, and perform hierarchy clustering. They obtained four major groups with epithelial, mesodermal, hematopoietic, and neuronal features. Moreover, they uncovered a hybrid epithelial/mesenchymal (E/M) state in epithelial cells, founding E/M hybrid state is a common process within endodermal organs. Another research from the Göttgens group reported the role of the leukotriene pathway in mediating blood progenitor formation during murine organogenesis using scRNA-seq.[147] They profiled the transcriptomic landscape of more than 20 000 individual cells from mouse embryos during the gastrulation to the organogenesis period. They identified around 20 major cell types and dynamic waves of transcription and candidate regulators in the ordering of somatic progenitor cells. Furthermore, they figured out the leukotriene biosynthesis pathway as a regulator in mediating early blood development through analyzing the full transcriptomes of the hemogenic endothelial cells and the blood progenitors (Figure 12b).[147]

In humans, Maehr and co-workers used scRNA-sequencing based on the drop-seq technique to study 8 days of thymus organogenesis. They profiled ≈25 000 cells and revealed cellular heterogeneity and interrogate developmental dynamics, and pinpointed the cell-specific expression patterns in stromal and blood populations. By combining the genome-wide association study data and autoimmune-disease-related genes with cell atlas, they found embryonic thymus-resident cells possibly participate in autoimmune disease etiologies (Figure 12c).[148] Similarly, in 2019, Hu and co-workers profiled the transcriptional landscape of human early T lymphopoiesis from multiple hemogenic and hematopoietic sites spanning embryonic and fetal stages through droplet-based and well-based (STRT-seq).[149] They found a subtype in early thymic progenitors that shared similar properties with a subset of lymphoid progenitors in the fetal liver, and identified a new subset of prethymic lymphoid progenitor in the aorta-gonad-mesonephros region. These researches provided useful information to better understand human early T lymphopoiesis and T lymphocyte regeneration.

4. Conclusion and Perspectives

In past years, fundamental research on stem cells and developmental biology has witnessed rapid progress via the assistance of single-cell sequencing. Despite the fruitful history, the marriage of developmental biology and single-cell sequencing still faces several critical challenges. By far, we mainly rely on pseudotime analysis to reconstruct the developmental trajectory of cell lineages, which could only infer the native process indirectly. Directly probing the temporal progression of cell fate decisions at the single-cell level in live embryos remains a desired target. The recent development of new techniques, such as the Live-seq,[54] might help address such challenges in the near future. In addition, due to extremely rare access to human embryos at the postimplantation and gastrulation stage, single-cell analysis for such precious samples should rely on highly sensitive techniques that could perform single-cell multiomics with a small amount of starting materials (such as that from just one cell). To this end, enhancement of the detection sensitivity and target versatility of single-cell sequencing becomes a fertile ground receiving increasing attention in recent years. Furthermore, the loss of original spatial information has been a major caveat for single-cell sequencing. Therefore, the combination of single-cell sequencing with stereo-sequencing should open tremendous opportunities for resolving the spatiotemporal trajectory of individual cells throughout embryogenesis, which might provide the most comprehensive roadmap for the development of humans in the near future. Once the technology of spatial-omics becomes mature and cost-effective, we believe many of the current mysteries in developmental biology will be unveiled.

Acknowledgements

Z.L. and F.L. contributed equally to this work. The authors acknowledge the support from the National Natural Science Foundation of China (Grant Nos. 11921002 and 11902007), the China Postdoctoral Science Foundation (Grant No. 2019M660313), the Natural Science Foundation of Guangdong Province (Grant No. 2019A1515012010), the Shenzhen Overseas Talent Program, and Tsinghua University Startup Funding.

Conflict of Interest

The authors declare no conflict of interest.

Keywords

developmental biology, embryo development, microfluidics, single cell

Received: August 25, 2021
Revised: November 5, 2021
Published online: December 22, 2021

[1] G. C. Schoenwolf, S. B. Bleyl, P. R. Brauer, P. H. Francis-West, *Larsen's Human Embryology*, Elsevier, Amsterdam **2020**.
[2] Y. Shao, J. Fu, *Curr. Opin. Genet. Dev.* **2020**, *63*, 30.
[3] A. A. Kolodziejczyk, J. K. Kim, V. Svensson, J. C. Marioni, S. A. Teichmann, *Mol. Cell* **2015**, *58*, 610.
[4] S. S. Potter, *Nat. Rev. Nephrol.* **2018**, *14*, 479.
[5] H. Huang, M. Goto, H. Tsunoda, L. Sun, K. Taniguchi, H. Matsunaga, H. Kambara, *Nucleic Acids Res.* **2013**, *42*, e12.
[6] B. Hwang, J. H. Lee, D. Bang, *Exp. Mol. Med.* **2018**, *50*, 1.
[7] O. Stegle, S. A. Teichmann, J. C. Marioni, *Nat. Rev. Genet.* **2015**, *16*, 133.
[8] Y. Chen, J. Song, Q. Ruan, X. Zeng, L. Wu, L. Cai, X. Wang, C. Yang, *Small Methods* **2021**, *5*, 2100111.
[9] A. R. Wu, J. Wang, A. M. Streets, Y. Huang, *Annu. Rev. Anal. Chem.* **2017**, *10*, 439.
[10] J. C. Love, J. L. Ronan, G. M. Grotenbreg, A. G. Van Der Veen, H. L. Ploegh, *Nat. Biotechnol.* **2006**, *24*, 703.
[11] R. Fan, O. Vermesh, A. Srivastava, B. K. H. Yen, L. Qin, H. Ahmad, G. A. Kwong, C.-C. Liu, J. Gould, L. Hood, J. R. Heath, *Nat. Biotechnol.* **2008**, *26*, 1373.
[12] J. He, A. T. Brimmo, M. A. Qasaimeh, P. Chen, W. Chen, *Small Methods* **2017**, *1*, 1700192.

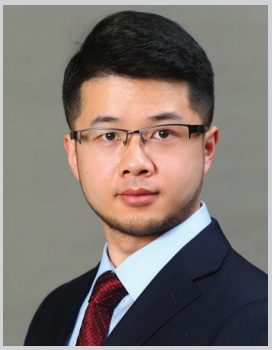
[13] S. Goodwin, J. D. Mcpherson, W. R. Mccombie, *Nat. Rev. Genet.* **2016**, *17*, 333.
[14] K. Wetterstrand, DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP), www.genome.gov/sequencingcostsdata (accessed: July 2021).
[15] M. D. Luecken, F. J. Theis, *Mol. Syst. Biol.* **2019**, *15*, e8746.
[16] L. Wen, F. Tang, *Mol. Cell* **2019**, *76*, 3208.
[17] Y. Wei, H.-L. Zhang, Q. Wang, C. Zhang, in *Single Cell Biomedicine* (Eds: J. Gu, X. Wang), Springer, Singapore **2018**, p. 103.
[18] G. Peng, G. Cui, J. Ke, N. Jing, *Annu. Rev. Genomics Hum. Genet.* **2020**, *21*, 163.
[19] N. Liu, L. Liu, X. Pan, *Cell. Mol. Life Sci.* **2014**, *71*, 2707.
[20] S. Petropoulos, S. P. Panula, J. P. Schell, F. Lanner, *J. Intern. Med.* **2016**, *280*, 252.
[21] P. Kumar, Y. Tan, Cahan, *Development* **2017**, *144*, 17.
[22] A. Gupta, M. P. Lutolf, A. J. Hughes, K. F. Sonnen, *Stem Cell Rep.* **2021**, *16*, 1104.
[23] F. A. V. Braga, R. J. Miragaia, *Single Cell Methods*, Springer, New York **2019**, pp. 9–21.
[24] T. Hashimshony, F. Wagner, N. Sher, I. Yanai, *Cell Rep.* **2012**, *2*, 666.
[25] J. W. Bagnoli, C. Ziegenhain, A. Janjic, L. E. Wange, B. Vieth, S. Parekh, J. Geuder, I. Hellmann, W. Enard, *Nat. Commun.* **2018**, *9*, 2937.
[26] A. B. Rosenberg, C. M. Roco, R. A. Muscat, A. Kuchina, P. Sample, Z. Yao, L. T. Graybuck, D. J. Peeler, S. Mukherjee, W. Chen, S. H. Pun, D. L. Sellers, B. Tasic, G. Seelig, *Science* **2018**, *360*, 176.
[27] S. C. Van Den Brink, F. Sage, Á. Vértesy, B. Spanjaard, J. Peterson-Maduro, C. S. Baron, C. Robin, A. Van Oudenaarden, *Nat. Methods* **2017**, *14*, 935.
[28] C. T. J. Van Velthoven, A. De Morree, I. M. Egner, J. O. Brett, T. A. Rando, *Cell Rep.* **2017**, *21*, 1994.
[29] M. Adam, A. S. Potter, S. S. Potter, *Development* **2017**, *144*, 3625.
[30] F. Tang, C. Barbacioru, Y. Wang, E. Nordman, C. Lee, N. Xu, X. Wang, J. Bodeau, B. B. Tuch, A. Siddiqui, K. Lao, M. A. Surani, *Nat. Methods* **2009**, *6*, 377.
[31] F. Tang, C. Barbacioru, E. Nordman, B. Li, N. Xu, V. I. Bashkirov, K. Lao, M. A. Surani, *Nat. Protoc.* **2010**, *5*, 516.
[32] S. Islam, U. Kjällquist, A. Moliner, P. Zajac, J.-B. Fan, P. Lönnerberg, S. Linnarsson, *Genome Res.* **2011**, *21*, 1160.
[33] D. Ramsköld, S. Luo, Y.-C. Wang, R. Li, Q. Deng, O. R. Faridani, G. A. Daniels, I. Khrebtukova, J. F. Loring, L. C. Laurent, G. P. Schroth, R. Sandberg, *Nat. Biotechnol.* **2012**, *30*, 777.
[34] H. Guo, P. Zhu, X. Wu, X. Li, L. Wen, F. Tang, *Genome Res.* **2013**, *23*, 2126.
[35] L. Yan, M. Yang, H. Guo, L. Yang, J. Wu, R. Li, P. Liu, Y. Lian, X. Zheng, J. Yan, J. Huang, M. Li, X. Wu, L. Wen, K. Lao, R. Li, J. Qiao, F. Tang, *Nat. Struct. Mol. Biol.* **2013**, *20*, 1131.
[36] Z. Xue, K. Huang, C. Cai, L. Cai, C.-Y. Jiang, Y. Feng, Z. Liu, Q. Zeng, L. Cheng, Y. E. Sun, J.-Y. Liu, S. Horvath, G. Fan, *Nature* **2013**, *500*, 593.
[37] K. M. Keays, G. P. Owens, A. M. Ritchie, D. H. Gilden, M. P. Burgoon, *J. Immunol. Methods* **2005**, *302*, 90.
[38] S. Nichterwitz, G. Chen, J. Aguila Benitez, M. Yilmaz, H. Storvall, M. Cao, R. Sandberg, Q. Deng, E. Hedlund, *Nat. Commun.* **2016**, *7*, 12139.
[39] S. Nichterwitz, J. A. Benitez, R. Hoogstraaten, Q. Deng, E. Hedlund, *Methods Mol. Biol.* **2018**, *1649*, 950.
[40] M. Ackers-Johnson, W. L. W. Tan, R. S.-Y. Foo, *Nat. Commun.* **2018**, *9*, 4434.
[41] E. Z. Macosko, A. Basu, R. Satija, J. Nemesh, K. Shekhar, M. Goldman, I. Tirosh, A. R. Bialas, N. Kamitaki, E. M. Martersteck, J. J. Trombetta, D. A. Weitz, J. R. Sanes, A. K. Shalek, A. Regev, S. A. Mccarroll, *Cell* **2015**, *161*, 1202.
[42] C. S. Mcginnis, D. M. Patterson, J. Winkler, D. N. Conrad, M. Y. Hein, V. Srivastava, J. L. Hu, L. M. Murrow, J. S. Weissman, Z. Werb, E. D. Chow, Z. J. Gartner, *Nat. Methods* **2019**, *16*, 619.
[43] Y. Sasagawa, I. Nikaido, T. Hayashi, H. Danno, K. D. Uno, T. Imai, H. R. Ueda, *Genome Biol.* **2013**, *14*, 3097.
[44] D. A. Jaitin, E. Kenigsberg, H. Keren-Shaul, N. Elefant, F. Paul, I. Zaretsky, A. Mildner, N. Cohen, S. Jung, A. Tanay, I. Amit, *Science* **2014**, *343*, 776.
[45] Y. Sasagawa, H. Danno, H. Takada, M. Ebisawa, K. Tanaka, T. Hayashi, A. Kurisaki, I. Nikaido, *Genome Biol.* **2018**, *19*, 29.
[46] J. Cao, J. S. Packer, V. Ramani, D. A. Cusanovich, C. Huynh, R. Daza, X. Qiu, C. Lee, S. N. Furlan, F. J. Steemers, A. Adey, R. H. Waterston, C. Trapnell, J. Shendure, *Science* **2017**, *357*, 661.
[47] H. Yin, D. Marshall, *Curr. Opin. Biotechnol.* **2012**, *23*, 110.
[48] Z. T. F. Yu, K. M. Aw Yong, J. Fu, *Small* **2014**, *10*, 1687.
[49] N. M. Karabacak, P. S. Spuhler, F. Fachin, E. J. Lim, V. Pai, E. Ozkumur, J. M. Martel, N. Kojic, K. Smith, P.-I. Chen, J. Yang, H. Hwang, B. Morgan, J. Trautwein, T. A. Barber, S. L. Stott, S. Maheswaran, R. Kapur, D. A. Haber, M. Toner, *Nat. Protoc.* **2014**, *9*, 694.
[50] Y. Deng, A. Finck, R. Fan, *Annu. Rev. Biomed. Eng.* **2019**, *21*, 365.
[51] S. M. Prakadan, A. K. Shalek, D. A. Weitz, *Nat. Rev. Genet.* **2017**, *18*, 345.
[52] J. Melin, S. R. Quake, *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 213.
[53] H. C. Fan, J. Wang, A. Potanina, S. R. Quake, *Nat. Biotechnol.* **2011**, *29*, 51.
[54] J. Wang, H. C. Fan, B. Behr, S. R. Quake, *Cell* **2012**, *150*, 402.
[55] A. M. Streets, X. Zhang, C. Cao, Y. Pang, X. Wu, L. Xiong, L. Yang, Y. Fu, L. Zhao, F. Tang, Y. Huang, *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 7048.
[56] A. R. Wu, N. F. Neff, T. Kalisky, P. Dalerba, B. Treutlein, M. E. Rothenberg, F. M. Mburu, G. L. Mantalas, S. Sim, M. F. Clarke, S. R. Quake, *Nat. Methods* **2013**, *11*, 41.
[57] J. D. Buenrostro, B. Wu, U. M. Litzenburger, D. Ruff, M. L. Gonzales, M. P. Snyder, H. Y. Chang, W. J. Greenleaf, *Nature* **2015**, *523*, 486.
[58] M. Zhang, Y. Zou, X. Xu, X. Zhang, M. Gao, J. Song, P. Huang, Q. Chen, Z. Zhu, W. Lin, R. N. Zare, C. Yang, *Nat. Commun.* **2020**, *11*, 2118.
[59] Y.-H. Cheng, Y.-C. Chen, E. Lin, R. Brien, S. Jung, Y.-T. Chen, W. Lee, Z. Hao, S. Sahoo, H. Min Kang, J. Cong, M. Burness, S. Nagrath, M. S. Wicha, E. Yoon, *Nat. Commun.* **2019**, *10*, 2163.
[60] H. C. Fan, G. K. Fu, S. P. A. Fodor, *Science* **2015**, *347*, 1258367.
[61] X. Han, R. Wang, Y. Zhou, L. Fei, H. Sun, S. Lai, A. Saadatpour, Z. Zhou, H. Chen, F. Ye, D. Huang, Y. Xu, W. Huang, M. Jiang, X. Jiang, J. Mao, Y. Chen, C. Lu, J. Xie, Q. Fang, Y. Wang, R. Yue, T. Li, H. Huang, S. H. Orkin, G.-C. Yuan, M. Chen, G. Guo, *Cell* **2018**, *172*, 1091.
[62] L. D. Goldstein, Y.-J. J. Chen, J. Dunne, A. Mir, H. Hubschle, J. Guillory, W. Yuan, J. Zhang, J. Stinson, B. Jaiswal, K. B. Pahuja, I. Mann, T. Schaal, L. Chan, S. Anandakrishnan, C.-W. Lin, P. Espinoza, S. Husain, H. Shapiro, K. Swaminathan, S. Wei, M. Srinivasan, S. Seshagiri, Z. Modrusan, *BMC Genomics* **2017**, *18*, 519.
[63] X. Zhang, T. Li, F. Liu, Y. Chen, J. Yao, Z. Li, Y. Huang, J. Wang, *Mol. Cell* **2019**, *73*, 130.
[64] Z. Zhu, C. J. Yang, *Acc. Chem. Res.* **2017**, *50*, 22.
[65] R. Zilionis, J. Nainys, A. Veres, V. Savova, D. Zemmour, A. M. Klein, L. Mazutis, *Nat. Protoc.* **2016**, *12*, 44.
[66] H.-S. Moon, K. Je, J.-W. Min, D. Park, K.-Y. Han, S.-H. Shin, W.-Y. Park, C. E. Yoo, S.-H. Kim, *Lab Chip* **2018**, *18*, 775.
[67] L. Li, P. Wu, Z. Luo, L. Wang, W. Ding, T. Wu, J. Chen, J. He, Y. He, H. Wang, Y. Chen, G. Li, Z. Li, L. He, *ACS Sens.* **2019**, *4*, 1299.
[68] P. Datlinger, A. F. Rendeiro, T. Boenke, M. Senekowitsch, T. Krausgruber, D. Barreca, C. Bock, *Nat. Methods* **2021**, *18*, 635.

[69] A. Rotem, O. Ram, N. Shoresh, R. A. Sperling, M. Schnall-Levin, H. Zhang, A. Basu, B. E. Bernstein, D. A. Weitz, *PLoS One* **2015**, *10*, e0116328.
[70] M. T. Chung, K. Kurabayashi, D. Cai, *Lab Chip* **2019**, *19*, 2425.
[71] M. T. Chung, D. Núñez, D. Cai, K. Kurabayashi, *Lab Chip* **2017**, *17*, 3664.
[72] M. T. Chung, *Ph.D. Thesis*, University of Michigan **2019**.
[73] W. Stephenson, L. T. Donlin, A. Butler, C. Rozo, B. Bracken, A. Rashidfarrokhi, S. M. Goodman, L. B. Ivashkiv, V. P. Bykerk, D. E. Orange, R. B. Darnell, H. P. Swerdlow, R. Satija, *Nat. Commun.* **2018**, *9*, 791.
[74] A. M. Klein, L. Mazutis, I. Akartuna, N. Tallapragada, A. Veres, V. Li, L. Peshkin, D. A. Weitz, M. W. Kirschner, *Cell* **2015**, *161*, 1187.
[75] G. X. Y. Zheng, J. M. Terry, P. Belgrader, P. Ryvkin, Z. W. Bent, R. Wilson, S. B. Ziraldo, T. D. Wheeler, G. P. Mcdermott, J. Zhu, M. T. Gregory, J. Shuga, L. Montesclaros, J. G. Underwood, D. A. Masquelier, S. Y. Nishimura, M. Schnall-Levin, P. W. Wyatt, C. M. Hindson, R. Bharadwaj, A. Wong, K. D. Ness, L. W. Beppu, H. J. Deeg, C. Mcfarland, K. R. Loeb, W. J. Valente, N. G. Ericson, E. A. Stevens, J. P. Radich, et al., *Nat. Commun.* **2017**, *8*, 14049.
[76] F. Lan, B. Demaree, N. Ahmed, A. R. Abate, *Nat. Biotechnol.* **2017**, *35*, 640.
[77] A. Rotem, O. Ram, N. Shoresh, R. A. Sperling, A. Goren, D. A. Weitz, B. E. Bernstein, *Nat. Biotechnol.* **2015**, *33*, 1165.
[78] T. Stuart, R. Satija, *Nat. Rev. Genet.* **2019**, *20*, 257.
[79] C. Ziegenhain, B. Vieth, S. Parekh, B. Reinius, A. Guillaumet-Adkins, M. Smets, H. Leonhardt, H. Heyn, I. Hellmann, W. Enard, *Mol. Cell* **2017**, *65*, 631.
[80] S. Islam, A. Zeisel, S. Joost, G. La Manno, P. Zajac, M. Kasper, P. Lönnerberg, S. Linnarsson, *Nat. Methods* **2014**, *11*, 163.
[81] S. Picelli, O. R. Faridani, Å. K. Björklund, G. Winberg, S. Sagasser, R. Sandberg, *Nat. Protoc.* **2014**, *9*, 171.
[82] H. Kimura, *J. Hum. Genet.* **2013**, *58*, 439.
[83] S. Ai, H. Xiong, C. C. Li, Y. Luo, Q. Shi, Y. Liu, X. Yu, C. Li, A. He, *Nat. Cell Biol.* **2019**, *21*, 1164.
[84] H. S. Kaya-Okur, H. S. Kaya-Okur, S. J. Wu, C. A. Codomo, E. S. Pledger, T. D. Bryson, J. G. Henikoff, K. Ahmad, S. Henikoff, *Nat. Commun.* **2019**, *10*, 1930.
[85] Q. Wang, H. Xiong, S. Ai, X. Yu, Y. Liu, J. Zhang, A. He, *Mol. Cell* **2019**, *76*, 206.
[86] R. Mulqueen, D. Pokholok, S. J. Norberg, K. A. Torkenczy, A. J. Fields, D. Sun, J. R. Sinnamon, J. Shendure, C. Trapnell, B. J. O'Roak, Z. Xia, F. J. Steemers, A. C. Adey, *Nat. Biotechnol.* **2018**, *36*, 428.
[87] S. L. Klemm, Z. Shipony, W. J. Greenleaf, *Nat. Rev. Genet.* **2019**, *20*, 207.
[88] D. Cusanovich, R. Daza, A. Adey, H. A. Pliner, L. Christiansen, K. L. Gunderson, F. J. Steemers, C. Trapnell, J. Shendure, *Science* **2015**, *348*, 910.
[89] B. B. Lake, S. Chen, B. C. Sos, J. Fan, G. E. Kaeser, Y. C. Yung, T. E. Duong, D. Gao, J. Chun, P. V. Kharchenko, K. Zhang, *Nat. Biotechnol.* **2018**, *36*, 70.
[90] S. J. Clark, R. Argelaguet, C.-A. Kapourani, T. M. Stubbs, H. J. Lee, C. Alda-Catalinas, F. Krueger, G. Sanguinetti, G. Kelsey, J. C. Marioni, O. Stegle, W. Reik, *Nat. Commun.* **2018**, *9*, 781.
[91] M. V. C. Greenberg, D. Bourc'his, *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 590.
[92] C. Luo, C. L. Keown, L. Kurihara, J. Zhou, Y. He, J. Li, R. Castanon, J. Lucero, J. R. Nery, J. P. Sandoval, B. Bui, T. J. Sejnowski, T. T. Harkins, E. A. Mukamel, M. M. Behrens, J. R. Ecker, *Science* **2017**, *357*, 600.
[93] S. Smallwood, H. J. Lee, C. Angermueller, F. Krueger, H. Saadeh, J. Peat, S. R. Andrews, O. Stegle, W. Reik, G. Kelsey, *Nat. Methods* **2014**, *11*, 817.
[94] Y. Shevchenko, S. Bale, *Cold Spring Harbor Perspect. Med.* **2016**, *6*, a025809.
[95] O. Morozova, M. A. Marra, *Genomics* **2008**, *92*, 255.
[96] B. Vieth, S. Parekh, C. Ziegenhain, W. Enard, I. Hellmann, *Nat. Commun.* **2019**, *10*, 4667.
[97] P. Jiang, J. A. Thomson, R. Stewart, *Bioinformatics* **2016**, *32*, 2514.
[98] Q. Liu, Q. Sheng, J. Ping, M. A. Ramirez, K. S. Lau, R. J. Coffey, Y. Shyr, *Bioinformatics* **2019**, *35*, 5306.
[99] A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, *Bioinformatics* **2013**, *29*, 15.
[100] B. Langmead, S. L. Salzberg, *Nat. Methods* **2012**, *9*, 357.
[101] C. Trapnell, L. Pachter, S. L. Salzberg, *Bioinformatics* **2009**, *25*, 1105.
[102] D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, S. L. Salzberg, *Genome Biol.* **2013**, *14*, R36.
[103] N. L. Bray, H. Pimentel, P. Melsted, L. Pachter, *Nat. Biotechnol.* **2016**, *34*, 525.
[104] A. Roberts, L. Pachter, *Nat. Methods* **2013**, *10*, 71.
[105] R. Patro, S. M. Mount, C. Kingsford, *Nat. Biotechnol.* **2014**, *32*, 462.
[106] R. Patro, G. Duggal, M. I. Love, R. A. Irizarry, C. Kingsford, *Nat. Methods* **2017**, *14*, 417.
[107] Z. Zhang, W. Wang, *Bioinformatics* **2014**, *30*, i283.
[108] C. A. Vallejos, C. A. Vallejos, D. Risso, A. Scialdone, S. Dudoit, J. C. Marioni, *Nat. Methods* **2017**, *14*, 565.
[109] A. Mortazavi, B. A. Williams, K. Mccue, L. Schaeffer, B. Wold, *Nat. Methods* **2008**, *5*, 621.
[110] C. Soneson, M. Delorenzi, *BMC Bioinf.* **2013**, *14*, 91.
[111] B. Li, V. Ruotti, R. M. Stewart, J. A. Thomson, C. N. Dewey, *Bioinformatics* **2009**, *26*, 493.
[112] M. D. Robinson, A. Oshlack, *Genome Biol.* **2010**, *11*, R25.
[113] S. Anders, W. Huber, *Genome Biol.* **2010**, *11*, R106.
[114] F. Abbas-Aghababazadeh, Q. Li, B. L. Fridley, *PLoS One* **2018**, *13*, e0206312.
[115] J. H. Bullard, E. Purdom, K. D. Hansen, S. Dudoit, *BMC Bioinf.* **2010**, *11*, 94.
[116] D. Risso, K. Schwartz, G. Sherlock, S. Dudoit, *BMC Bioinf.* **2011**, *12*, 480.
[117] L. van der Maaten, G. Hinton, *J. Mach. Learn. Res.* **2008**, *9*, 2579.
[118] D. Kobak, P. Berens, *Nat. Commun.* **2019**, *10*, 5416.
[119] K. R. Moon, D. van Dijk, Z. Wang, W. Chen, M. J. Hirn, R. R. Coifman, N. B. Ivanova, G. Wolf, S. Krishnaswamy, *Nat. Biotechnol.* **2019**, *37*, 1482.
[120] J. Ding, A. Condon, S. P. Shah, *Nat. Commun.* **2018**, *9*, 2002.
[121] E. Becht, L. Mcinnes, J. Healy, C.-A. Dutertre, I. W. H. Kwok, L. G. Ng, F. Ginhoux, E. W. Newell, *Nat. Biotechnol.* **2019**, *37*, 38.
[122] B. Wang, J. Zhu, E. Pierson, D. Ramazzotti, S. Batzoglou, *Nat. Methods* **2017**, *14*, 414.
[123] S. Petropoulos, D. Edsgärd, B. Reinius, Q. Deng, S. P. Panula, S. Codeluppi, A. P. Reyes, S. Linnarsson, R. Sandberg, F. Lanner, *Cell* **2016**, *167*, 285.
[124] H. Guo, P. Zhu, F. Guo, X. Li, X. Wu, X. Fan, L. Wen, F. Tang, *Nat. Protoc.* **2015**, *10*, 645.
[125] F. Guo, L. Li, J. Li, X. Wu, B. Hu, P. Zhu, L. Wen, F. Tang, *Cell Res.* **2017**, *27*, 967.
[126] F. H. Biase, X. Cao, S. Zhong, *Genome Res.* **2014**, *24*, 1787.
[127] A. Burton, J. Muller, S. Tu, P. Padilla-Longoria, E. Guccione, M.-E. Torres-Padilla, *Cell Rep.* **2013**, *5*, 687.
[128] F. Zhou, R. Wang, P. Yuan, Y. Ren, Y. Mao, R. Li, Y. Lian, J. Li, L. Wen, L. Yan, J. Qiao, F. Tang, *Nature* **2019**, *572*, 660.
[129] R. C. West, H. Ming, D. M. Logsdon, J. Sun, S. K. Rajput, R. A. Kile, W. B. Schoolcraft, R. M. Roberts, R. L. Krisher, Z. Jiang, Y. Yuan, *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 22635.
[130] B. Lv, Q. An, Q. Zeng, X. Zhang, P. Lu, Y. Wang, X. Zhu, Y. Ji, G. Fan, Z. Xue, *PLoS Biol.* **2019**, *17*, e3000187.
[131] L. Tan, L. A. Lacko, T. Zhou, D. Tomoiaga, R. Hurtado, T. Zhang, A. Sevilla, A. Zhong, C. E. Mason, S. Noggle, T. Evans, H. Stuhlmann, R. E. Schwartz, S. Chen, *Nat. Commun.* **2019**, *10*, 4155.

[132] A. Guzzetta, M. Koska, M. Rowton, K. R. Sullivan, J. Jacobs-Li, J. Kweon, H. Hidalgo, H. Eckart, A. D. Hoffmann, R. Back, S. Lozano, A. M. Moon, A. Basu, M. Bressan, S. Pott, I. P. Moskowitz, *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 15712.
[133] M. M. Chan, Z. D. Smith, S. Grosswendt, H. Kretzmer, T. M. Norman, B. Adamson, M. Jost, J. J. Quinn, D. Yang, M. G. Jones, A. Khodaverdian, N. Yosef, A. Meissner, J. S. Weissman, *Nature* **2019**, *570*, 77.
[134] B. Pijuan-Sala, J. A. Griffiths, C. Guibentif, T. W. Hiscock, W. Jawaid, F. J. Calero-Nieto, C. Mulas, X. Ibarra-Soria, R. C. V. Tyser, D. L. L. Ho, W. Reik, S. Srinivas, B. D. Simons, J. Nichols, J. C. Marioni, B. Göttgens, *Nature* **2019**, *566*, 490.
[135] S. Probst, Sagar, J. Tosic, C. Schwan, D. Grün, S. J. Arnold, *Development* **2021**, *148*, dev193789.
[136] H.-Q. Dai, B.-A. Wang, L. Yang, J.-J. Chen, G.-C. Zhu, M.-L. Sun, H. Ge, R. Wang, D. L. Chapman, F. Tang, X. Sun, G.-L. Xu, *Nature* **2016**, *538*, 528
[137] S. Grosswendt, H. Kretzmer, Z. D. Smith, A. S. Kumar, S. Hetzel, L. Wittler, S. Klages, B. Timmermann, S. Mukherji, A. Meissner, *Nature* **2020**, *584*, 102.
[138] H. Mohammed, I. Hernando-Herraez, A. Savino, A. Scialdone, I. Macaulay, C. Mulas, T. Chandra, T. Voet, W. Dean, J. Nichols, J. C. Marioni, W. Reik, *Cell Rep.* **2017**, *20*, 1215.
[139] L. Xiang, Y. Yin, Y. Zheng, Y. Ma, Y. Li, Z. Zhao, J. Guo, Z. Ai, Y. Niu, K. Duan, J. He, S. Ren, D. Wu, Y. Bai, Z. Shang, X. Dai, W. Ji, T. Li, *Nature* **2020**, *577*, 537.
[140] H. L. Sladitschek, U.-M. Fiuza, D. Pavlinic, V. Benes, L. Hufnagel, P. A. Neveu, *Cell* **2020**, *181*, 922.
[141] H. Ma, J. Zhai, H. Wan, X. Jiang, X. Wang, L. Wang, Y. Xiang, X. He, Z.-A. Zhao, B. Zhao, P. Zheng, L. Li, H. Wang, *Science* **2019**, *366*, 836.
[142] X. F. Xue, R. P. Wang, J. P. Fu, *Curr. Opin. Biomed. Eng.* **2020**, *13*, 127.
[143] T. Haremaki, J. J. Metzger, T. Rito, M. Z. Ozair, F. Etoc, A. H. Brivanlou, *Nat. Biotechnol.* **2019**, *37*, 1198.
[144] Z. C. Shang, D. Chen, Q. Wang, S. Wang, Q. Deng, L. Wu, C. Liu, X. Ding, S. Wang, J. Zhong, D. Zhang, X. Cai, S. Zhu, H. Yang, L. Liu, J. L. Fink, F. Chen, X. Liu, Z. Gao, X. Xu, *Gigascience* **2018**, *7*, giy117.
[145] J. Cao, M. Spielmann, X. Qiu, X. Huang, D. M. Ibrahim, A. J. Hill, F. Zhang, S. Mundlos, L. Christiansen, F. J. Steemers, C. Trapnell, J. Shendure, *Nature* **2019**, *566*, 496.
[146] J. Dong, Y. Hu, X. Fan, X. Wu, Y. Mao, B. Hu, H. Guo, L. Wen, F. Tang, *Genome Biol.* **2018**, *19*, 31.
[147] X. Ibarra-Soria, W. Jawaid, B. Pijuan-Sala, V. Ladopoulos, A. Scialdone, D. J. Jörg, R. C. V. Tyser, F. J. Calero-Nieto, C. Mulas, J. Nichols, L. Vallier, S. Srinivas, B. D. Simons, B. Göttgens, J. C. Marioni, *Nat. Cell Biol.* **2018**, *20*, 127.
[148] E. M. Kernfeld, R. M. J. Genga, K. Neherin, M. E. Magaletta, P. Xu, R. Maehr, *Immunity* **2018**, *48*, 1258.
[149] Y. Zeng, C. Liu, Y. Gong, Z. Bai, S. Hou, J. He, Z. Bian, Z. Li, Y. Ni, J. Yan, T. Huang, H. Shi, C. Ma, X. Chen, J. Wang, L. Bian, Y. Lan, B. Liu, H. Hu, *Immunity* **2019**, *51*, 930.
[150] W. Chen, O. Guillaume-Gentil, R. Dainese, P. Y. de Rainer, M. Zachara, C. G. Gäbelein, J. A. Vorholt, B. Deplancke, *bioRxiv*:2021.03.24.436752, https://doi.org/10.1101/2021.03.24.436752.

Zida Li is an assistant professor in biomedical engineering at Shenzhen University. He received his B.S. degree in mechanical engineering from the University of Science and Technology of China, and Ph.D. degree from the University of Michigan, Ann Arbor, working on the nexus of micro/nanoengineering and biology. His research interests include microfluidic point-of-care testing and single-cell analysis.

Feng Lin is a postdoctoral fellow in biomedical engineering at Peking University. He received his B.S. degree in bioengineering from the Chongqing University, and then received his M.S. degree in biomedical engineering from Chongqing University as well. He received his Ph.D. degree from Peking University, working on the nexus of cell–cell and cell–microenvironment mechanical interaction and mechanobiology. His research interests include tissue engineering and cell mechanics.



Yue Shao is an associate professor in the Institute of Biomechanics and Medical Engineering at Tsinghua University. He received his B.S. and M.S. degrees in engineering mechanics from Tsinghua University, and his Ph.D. degree from the University of Michigan, Ann Arbor. His research interests include stem cells, human development, and synthetic embryo and organ models.