



# Technical Advances in Single-Cell RNA Sequencing and Applications in Normal and Malignant Hematopoiesis

Xiang-tao Huang, Xi Li, Pei-zhong Qin, Yao Zhu, Shuang-nian Xu\* and Jie-ping Chen\*

Center of Haematology, Southwest Hospital, Third Military Medical University (Army Medical University), Chongqing, China

## OPEN ACCESS

### Edited by:

Massimo Breccia,  
La Sapienza University of Rome, Italy

### Reviewed by:

Michael Diamantidis,  
University Hospital of Larissa, Greece  
Diverio Daniela,  
Policlinico Umberto I, Italy

### \*Correspondence:

Shuang-nian Xu  
xshuangnian@tmmu.edu.cn  
Jie-ping Chen  
Chenjpxn@163.com

### Specialty section:

This article was submitted to  
Hematologic Malignancies,  
a section of the journal  
Frontiers in Oncology

**Received:** 28 August 2018

**Accepted:** 19 November 2018

**Published:** 07 December 2018

### Citation:

Huang X, Li X, Qin P, Zhu Y, Xu S and  
Chen J (2018) Technical Advances in  
Single-Cell RNA Sequencing and  
Applications in Normal and Malignant  
Hematopoiesis. *Front. Oncol.* 8:582.  
doi: 10.3389/fonc.2018.00582

Single-cell RNA sequencing (scRNA-seq) has been tremendously developed in the past decade owing to overcoming challenges associated with isolation of massive quantities of single cells. Previously, cell heterogeneity and low quantities of available biological material posed significant difficulties to scRNA-seq. Cell-to-cell variation and heterogeneity are fundamental and intrinsic characteristics of normal and malignant hematopoietic cells; this heterogeneity has often been ignored in omics studies. The application of scRNA-seq has profoundly changed our comprehension of many biological phenomena, including organ development and carcinogenesis. Hematopoiesis, is actually a maturation process for more than ten distinct blood and immune cells, and is thought to be critically involved in hematological homeostasis and in sustaining the physiological functions. However, aberrant hematopoiesis directly leads to hematological malignancy, and a deeper understanding of malignant hematopoiesis will provide deeper insights into diagnosis and prognosis for patients with hematological malignancies. Here, we aim to review the recent technical progress and future prospects for scRNA-seq, as applied in physiological and malignant hematopoiesis, in efforts to further understand the hematopoietic hierarchy and to illuminate personalized therapy and precision medicine approaches used in the clinical treatment of hematological malignancies.

**Keywords:** single-cell RNA sequencing, normal hematopoiesis, malignant hematopoiesis, hematopoietic hierarchy, hematological malignancy

## INTRODUCTION

Single-cell-based technologies are being rapidly developed and applied in the field of pluripotent and tissue-specific normal and malignant hematopoiesis. With recent advances in single-cell isolation platforms, including the ICELL8™ single-cell system, the Fluidigm C1 single-cell automatic preparation system and the Illumina Bio-Rad single-cell sequencing solution, it is possible to isolate single cells from multiple specimens simultaneously with high throughput and high resolution. Single-cell sequencing (SCS) has rapidly advanced along with developments in genome/transcriptome amplification and next-generation sequencing (NGS). The term SCS consists of single-cell genome sequencing, single-cell transcriptome sequencing, namely single-cell RNA sequencing (scRNA-seq), and single-cell epigenome sequencing (1, 2). There exists remarkable heterogeneity and functional distinction among homogenous cell populations, owing to diverse epigenetic states, including DNA methylation, hydroxymethylation, histone modification,

chromatin configuration, noncoding RNA regulation, and structural and regulatory proteins bound to chromatin. Currently, single-cell epigenome sequencing refers primarily to single-cell DNA methylation sequencing and single-cell assay for transposase-accessible chromatin using sequencing (Single-cell ATAC-seq).

Generally, most interpretations of basic experimental and clinical results *in vivo* or *in vitro* are based on the assumption that all of the cells used are homogeneous. However, this averaging of messages processing method always loses the critical information owing to ignoring the cell-to-cell variation, even within genetically homogenous cell populations, and natural heterogeneity within specimens is not truly reflected. Cellular heterogeneity is present in nearly all organs, tissues, and tumors in multicellular organisms, and presents a challenge to discern cellular functions and roles in normal functioning and in disease states as well. Single-cell transcriptome sequencing, also known as scRNA-seq, can elucidate the composition of heterogeneous cell populations, including the cellular heterogeneity during normal and malignant hematopoiesis. The concept of cellular heterogeneity was first proposed in 1957 (3). Single cells are considered the smallest structural and functional unit of an organism, as each cell represents a unique unit with molecular coding across the DNA, RNA, and protein conversions (4). Thus, relevant studies, especially the omics studies, are expected to be carried out at the single cell level. Here, we aim to discuss the recent technical progress as well as the application of scRNA-seq in normal and malignant hematopoiesis, in efforts to better understand the hematopoietic hierarchy and to illuminate personalized therapy and precision medicine approaches used in the clinical treatment of hematological malignancies.

## SINGLE-CELL ISOLATION APPROACHES

Single-cell sequencing applied in transcriptome was first carried out by Tang's lab in 2009 (5). Undoubtedly, single-cell isolation is the critical step in scRNA-seq. At present, the main approaches for isolating single cells from heterogeneous tissues or cultured cells include serial limited dilution, micromanipulation, fluorescence-activated cell sorting (FACS), laser-capture microdissection (LCM), and microfluidics (6–9). New sophisticated methods with higher accuracy and specificity are also emerging, such as Raman tweezers (10, 11) and transcriptome *in vivo* analysis (12). The above approaches differ among their advantages and disadvantages, and each approach has a unique scope of application (Table 1). Here, we will summarize them briefly in the following.

Serial limited dilution, a classical technology in single-cell separation, has been used by laboratories all over the world for decades. Owing to the stochastic distribution property of suspended cells, the number of cells in a highly diluted specimen can be as low as one single cell per well by multiple serial dilution into 96-well plates. Such method of seeding cells at low concentration is indeed easy to carry out with standard laboratory pipettors with a low reagent cost, but it is time consuming since the probability of achieving a single cell in an aliquot is of statistical nature (9). What's more, serial limited dilution is prone to a high false positive rate and to exclusion of

cells of interest, which hinders the isolation of target cells. Hence, while this approach is appropriate for microbial samples, it is not a preferred approach for sorting single cells from complex specimens (11).

Individual cells can also be sorted by micromanipulation using simple mouth pipettes (13). As an operable and low-cost approach with visualization, micromanipulation is most routinely used in single cell isolation, including endosymbionts from termite gut (14) and crenarchaeota from soil (15). However, micromanipulation is time consuming, has low throughput, and can cause mechanical shearing of cells; what's worse, it is only applicable to cells in suspension.

Based on multiple cellular physical or optical properties, such as the relative size, granularity, and intrinsic or extrinsic fluorescence, fluorescence activated cell sorting (FACS) has been applied as a preferred approach in single-cell separation (11). The primary advantages of FACS-based sorting are the operability, low-cost, high levels of accuracy and high throughput. However, FACS system requires a large number of suspended cells as the initial material, which hinders the isolation of low-abundance cell populations. Worse still, sorted cells may be damaged by the high pressure flow, so cell viability should be taken into consideration if sorted live cells are planned to be used in downstream protocols (7).

Laser-capture microdissection (LCM) is an ideal method to capture single cells from tissue samples. LCM is based on the selective adherence of visually targeted cells and tissue fragments to the thermoplastic membrane activated by a low energy infrared laser pulse (6). LCM enables the capture of cells from intact tissues in a natural state, and the RNA integrity of the captured cells is well retained. LCM is highly efficient and specific, and is suitable for rapid separation of interested tissue specimens. Limitations of LCM include its complicated operation, and preparation of the tissue specimens by fixation or cryopreservation. Moreover, RNA contamination may occur during cell separation and tissue cutting.

Innovative microfluidic devices have opened new horizons in single-cell isolation and analysis (16). Fabricated microfluidic chips make up the core component of microfluidic devices; these chips allow for sample compartmentalization and control the management of nanoliter reactions. Together with accurate construction of low-volume chambers and tubes, microfluidics can serve as an ideal method for single cell separation (7). The advantages inherent to microfluidics are high throughput with minimal effort, cost effectiveness, and high accuracy, and microfluidics are suitable for isolating homogeneous cells and able to sort uncultivated cells from a small quantity of samples. However, the limitations are that microfluidics are only useful for transcriptome analysis and can't readily be used to separate variant cells.

Raman tweezers are a newly advanced sophisticated technology for isolating single cells. Raman microspectroscopy enables the identification of different cell types by biochemical profiling without external labeling. After cells have been identified by Raman microspectroscopy, they are subsequently trapped with a laser (11). Single yeast and bacterial cells have been successfully isolated from an artificial cell mixture according to their Raman spectra (10). Raman tweezers are suitable for

**TABLE 1** | The currently used single-cell isolating methods with respective advantages and disadvantages.

Single-cell isolating methods	Advantages	Disadvantages	References
Serial limited dilution	Suitable for microbial specimens	Complicated operation, high false positive rate	(9)
Micromanipulation	Operability Low-cost Small quantity of cells	Low throughput Time consuming Mechanical shearing damage	(11)
Fluorescence-activated cell sorting	High accuracy High-throughput	High initial cell count Low efficiency Potential cell damage by rapid flow	(7)
Laser-capture microdissection	High efficiency High specificity Suitable for tissue specimen	Complicated operation Need staining	(6)
Microfluidics	High efficiency High throughput Homogeneous cells	Only for transcriptome Unable to sort variant cells	(7, 8)
Raman tweezers	Unculturable microorganisms	Potential contamination Potential physical misplacement	(10, 11)
Transcriptome <i>in vivo</i> analysis	Direct RNA capturing Non-invasive operation	Low throughput One tissue at one time	(12)

unculturable microorganisms without any pre-treatment, but only physiologically distinct cells are permitted and exists potential contamination and physical misplacement.

Transcriptome *in vivo* analysis (TIVA), though not strictly a technique for single-cell isolation, is another innovative method that captures RNA from a single cell directly by light activation using TIVA tags. The TIVA capture tag enables researchers to target and isolate RNAs from living cells in their natural microenvironments without damaging surrounding cells, which preserves the cellular response to the microenvironment and, to a great extent, the location information of cells in their resident tissue. The TIVA approach permits cell-specific transcriptome capture from viable intact heterogeneous tissues with non-invasive operation. However, the low throughput and the limitation that a single tissue can be evaluated in each experiment hinder its widespread usage (12).

To sum up, we have described several rough principles for isolating single cells from heterogeneous cell populations. If the cells of interest are already in suspension with a relatively high abundance, FACS methodology will be the ideal method. Cells that are extremely rare can be obtained by micromanipulations. TIVA enables single-cell transcriptome analysis by directly capturing RNA from a single cell using TIVA tags. If the cells are to be isolated from a whole tissue, collagenase and dispase are routinely used. However, enzymatic digestion has adverse impacts on cells, such as influencing cell yield and immune cell function, and can alter gene transcription, which result in changes in the expression of cell surface molecules (17, 18). When tissues are prepared into cell suspensions, cells can be isolated with specific fluorescence tags. Further acquisition of single cells is a tricky step, and may require additional microfluidic processing. When cells disperse into suspension, they lose their position in the tissue; in order to take into consideration the environment of a single cell and its former neighbors, LCM are recommended. However, LCM requires detailed scanning of tissue sections, and careful attention must be paid to acquire cells

of interest using microdissection. In a word, there are respective advantages and disadvantages among these single-cell isolation techniques, and the most suitable method should be chosen according to the experimental requirements.

## SINGLE-CELL RNA AMPLIFICATION AND SEQUENCING METHODS

Generally speaking, a single mammalian cell only has ~10 pg of total RNA and less than 0.1 pg of mRNA. However, RNA-seq usually requires microgram amounts of total RNA for analysis, which equals the total content of millions of cells. Therefore, mRNA is reverse transcribed into cDNA, and amplification of the cDNA library is required prior to deep sequencing. Currently, there are three commonly used RNA amplification strategies in scRNA-seq: PCR-based amplification, linear isothermal amplification by T7-based *in vitro* transcription (IVT) and Phi29 DNA polymerase-based amplification. These strategies are reviewed below.

### PCR-Based Amplification and Sequencing

The PCR-based amplification is a nonlinear amplification process. The use of PCR-based amplification for preparing cDNA from single cells for scRNA-seq analysis was first reported in 1993 (19). This method was modified and applied to the first RNA-seq research in single cells in 2009 by Tang et al. (5). Based on the synthesis pattern of the second chain of cDNA, PCR-based amplification can be divided into homopolymer tailing (known as the Brady/Tang protocol), template-switching and random priming.

For the Brady/Tang protocol, after single-cell isolation and lysis, mRNAs are reverse transcribed into cDNAs using a polyT primer with a special anchor sequence. Afterwards, ~30 nt polyA tails are attached to the first-strand cDNA at the 3'-end through terminal deoxynucleotidyl transferase. Finally, the second-strand

cDNA is synthesized using polyT primers with another anchor sequence. For template-switching, the second-strand cDNA is synthesized by the terminal transferase activity and the template-switching activity of reverse transcriptase, which derives three strategies for scRNA-seq, namely single-cell tagged reverse transcription sequencing (STRT-seq) (20), switching mechanism at the 5'-end of the RNA transcript sequencing (Smart-seq) (21) and Smart-seq2 (22). Random priming, another approach for amplification, involves semi-random primed PCR-based mRNA transcriptome amplification procedure. In this method, the second-strand cDNA synthesis is initiated from semi-random sequences of oligonucleotide primers. After amplification, the semi-random primers are immediately removed by BcivI, a type II restriction enzyme. Otherwise, what's different from other methods is that generated cDNA fragments are directly ligated to sequencing adaptor with no end-repairing (23). Importantly, compared to other strategies, the random priming covers full-length of any size of transcripts.

### STRT-seq

STRT-seq is the first Smart-based scRNA-seq approach based on template-switching mechanism (20). In STRT, mRNA is reverse-transcribed into cDNA by a tailed oligo(dT) primer with a barcode and an upstream primer-binding sequence, and the barcode is introduced into the 5'-end of the transcripts by Moloney Murine Leukemia Virus reverse- transcriptase (MMLV RT).

The 3'-ends of mRNAs are preferentially amplified, and transcript quantification is accomplished through reads mapping to 5'-ends of mRNAs (20). The greatest advantage of this method is the highly accurate positioning of the 5'-end of transcripts, which diminishes the 3'-bias. Owing to the barcode strategy, this technique shows low cost and short time. Furthermore, as each mRNA relates to a single cDNA, so the number of reads observed should be directly proportional to the number of mRNA, thus it is believed that quantification can be achieved (24). However, STRT is unsuitable for applications seeking to detect alternatively spliced transcripts (25).

### Smart-seq

Smart-seq is a robust and reproducible method for sequencing the transcriptomes of single mammalian cells with improved read coverage across transcripts and relative high sensitivity and quantitative accuracy. This strategy also relies on the terminal transferase and template-switching activities of MMLV RT (26). Briefly, polyadenylated RNA is reverse transcribed through tailed oligo(dT) priming by MMLV RT. Once the reverse transcription reaction reaches the 5'-end of an RNA molecule, a few non-templated nucleotides are added to the 3'-end of the cDNA by the terminal transferase activity of MMLV RT. This newly synthesized cDNA serves as a new template by the template-switching activity of MMLV RT, and then continues to transcribe, thus completing the amplification of transcriptome. Smart-seq greatly enriches transcripts with intact 5'-ends and eliminates the second-strand synthesis (21). The transcript amplification part of Smart-seq is now marketed as SMARTer Ultra Low RNA Kit for Illumina Sequencing.

### Smart-seq2

Smart-seq2 was recently described as an updated version of Smart-seq (22). In this protocol, a relatively mild (hypotonic) lysis buffer containing free dNTPs and tailed oligo(dT) oligonucleotides [30-nt poly(dT) stretch and a 25-nt universal 5' anchor sequence] is used to lyse cells without interfering with or inhibiting reverse transcription. Exchanging only a single guanylate for a locked nucleic acid (LNA) guanylate at the template-switching oligonucleotides (TSOs) 3'-end (rGrG+G) led to a two-fold increase in cDNA library yield, which seems to be a consequence of increased thermal stability of LNA:DNA base pairs. Additionally, the presence of betaine, a methyl group donor, in combination with high (9–12 mM) MgCl<sub>2</sub> concentrations significantly increased cDNA yield by 2 to 4-fold. What's more, the sensitivity and accuracy was increased by the use of KAPA HiFi Hot Start (KAPA) DNA polymerase, which efficiently amplified the first-strand cDNA directly after reverse transcription with no reduction in cDNA yield, but increased average cDNA length by 450-nt (22, 27). Therefore, compared to Smart-seq, optimized Smart-seq2 with improved reverse transcription, template switching, and pre-amplification increases cDNA library yield and full-length coverage, increases sensitivity and quantitative accuracy, and decreases technical bias.

### T7-Based IVT Amplification and Sequencing

The use of the T7 RNA polymerase to perform linear isothermal amplification of cDNA by IVT, requiring three rounds of amplification, was first reported by Van Gelder in 1990 (28). This strategy has been extensively applied to scRNA-seq in recent years, and helped promote the birth of the era of single-cell analysis (29–31). T7-based IVT amplification was developed for the synthesis of the second-chain cDNA sequences with no need for an anchor sequence. In this method, when the first-chain cDNA is synthesized, the RNA polymerase promoter sequence is led into and cDNA amplification is initiated by an IVT reaction. Currently, there are several scRNA-seq strategies based on T7 RNA polymerase, including cell expression by linear amplification and sequencing (CEL-seq) (29), CEL-seq2 (32), Quartz-seq (33), massively parallel single-cell RNA sequencing (MARS-seq) (30), and indexing droplets RNA sequencing (inDrops-seq) (34, 35). In what follows, we will review them in detail.

### CEL-seq

CEL-seq is the first method to use IVT that allowed the 3'-end of mRNAs to be detected. Considering that the application of high throughput sequencing was limited by the small initial amounts of RNA, CEL-seq was developed to overcome the limitation by barcoding and pooling samples before linearly amplifying mRNA with the use of one round of IVT (29). This technique is initiated by a single cell reverse transcription reaction using a CEL-seq primer, which is designed with an anchored poly(dT) stretch, a characteristic barcode, an Illumina 5' sequencing adaptor, and a T7 promoter (29). RNAs from individual cells are uniquely barcoded with a CEL-seq primer for reverse transcription. After

the second-strand synthesis, multiple samples are pooled for IVT reactions. Then, the amplified RNAs are fragmented and the reactions are terminated by placing on ice with the addition of EDTA, followed by RNA cleanup before entry into a modified version of the Illumina directional RNA protocol. After 12 cycles of PCR amplification, the fragments with both 5'- and 3'-Illumina adaptors are selected to construct the cDNA libraries, which are sequenced on an Illumina HiSeq2000 via paired-end reads strategy, where the first read recovers the barcode, while the second read identifies the mRNA transcript (24, 29). Compared to Smart-based scRNA-seq, CEL-seq gives rise to more reproducible, linear, and more sensitive outcomes than a PCR-based amplification; these advantages rely on the linear pattern of amplification. What's more, unlike PCR amplification, CEL-seq does not exponentially deplete sequences which are unfavorable to the PCR process (24, 29, 36).

### CEL-seq2

Based on the technique of CEL-seq, Hashimshony et al. developed a modified version, called CEL-seq2, with higher sensitivity, lower cost, and less hands-on time (32). In this protocol, cell lysis, reverse transcription, the second-strand synthesis, and IVT were performed on-chip. Early barcoding enabled highly-multiplexed analysis, and 3'-end tagging was used, which enables accurate estimation of expression levels without having to account for gene length and with fewer sequencing reads required (32, 37). For CEL-seq2 primers, a 6-base unique molecular identifier (UMI) was integrated into the CEL-seq primer which is located between the barcode and the Illumina 5' sequencing adaptor. The length of the barcode was reduced from eight to six nucleotides, and the T7 promoter and the Illumina 5' sequencing adaptor were shortened. The procedures for CEL-seq2 are nearly identical to CEL-seq, except that the sequencing is performed on the Fluidigm C1 system.

The advantages of CEL-seq2 are derived from the use of optimized primers, reagents, RNA cleanup and purification, and library preparation steps. These modifications greatly improved the quality of the data and make CEL-seq2 more time- and cost-efficient. What's more, the Fluidigm C1 system-enabled cell-barcoding allows for a single library construction, instead of setting up multiple library preparations individually (32).

### Quartz-seq

Emerging Quartz-seq is another scRNA-seq technique that allows highly quantitative gene expression analysis. Quartz-seq is based on polyA tailing, PCR amplification, and IVT, and follows a simpler procedure resulting in higher reproducibility and greater sensitivity surveying the mRNA content of single cells than previously developed methods. Moreover, comprehensive and quantitative detection of gene expression heterogeneity can be achieved by this approach. Quartz-seq was first reported by Sasagawa et al. in 2013 (33). First of all, polyadenylated RNA is reverse transcribed to generate the first strand cDNA using reverse transcription primers that contain oligo(dT), the T7 promoter and PCR target region sequences. After the first-strand cDNA synthesis, the majority of the reverse transcription primer is digested by exonuclease I. Then, a polyA tail is added to the

3'-ends of the first-strand cDNA and any surviving RT primer, and the synthesis of the second strand is initiated with the tagging primer. The resulting cDNA and the byproducts from the surviving primers contain the whole-transcript amplification adaptor sequences, which include the reverse transcription primer sequence and the tagging primer sequence. Subsequently, these DNAs are used for suppressing PCR with the suppression PCR primer. After fragmenting the amplified cDNA using the Covaris shearing system and enriching the short cDNA fragments, the high-quality cDNA libraries are obtained. Finally, the cDNA sequencing libraries are analyzed by an Illumina sequencer (33). In this protocol, only six reaction steps are required and all steps are completed in a single PCR tube without any purification, which dramatically simplifies the Quartz-seq method.

### MARS-seq

To construct single cell libraries from polyadenylated RNA, a new method, termed MARS-seq was developed by Jaitin et al. in 2014 (30). Briefly, single cells are isolated by FACS into 384-well plates containing hypotonic cell lysis buffer and reverse transcription primers. The reverse transcription primer consists of anchored poly(dT), UMI, cell barcode, partial Illumina paired-end primer sequence, and a T7 RNA polymerase promoter. The cell barcode is used for subsequent de-multiplexing, and UMI for correction of amplification biases (30, 37). After the first-strand cDNA synthesis, exonuclease I is added and samples are pooled for the second-strand synthesis. Subsequently, samples are linearly amplified by IVT and chemically fragmented into a median size of ~200 nucleotides. Then, a partial Illumina Read1 sequencing adapter including a pool barcode is single strand ligated to the fragmented RNA using a T4 RNA ligase I. The ligated product is reverse transcribed and the cDNA is purified. Next, the sequencing library is completed and amplified through a nested PCR reaction, with the forward primer containing the Illumina P5-Read1 sequences and the reverse primer containing the P7-Read2 sequences. Finally, MARS-Seq libraries are paired-end sequenced by an Illumina HiSeq 2500 (30). This protocol is based on FACS-sorting of single cells into 384-well plates and subsequent automated processing on a Bravo robot station, and is generally performed on pooled and labeled material, leading to significant increases in throughput and reproducibility.

### InDrops-seq

Droplet microfluidics have been used to develop a technique for indexing thousands of individual cells for RNA sequencing, termed indexing droplets (inDrops) RNA sequencing; this technique has the capability to index >15,000 cells per hour, boasts low technical noise, and is adaptable to other sequencing-based assays (34, 35). Briefly, isolated individual cells are encapsulated into nanoliter droplets containing lysis buffer, reverse-transcription mix, and reverse transcription oligonucleotide primers consisting of poly(dT), UMIs, cell barcodes, sequencing adaptors, T7 RNA polymerase promoters, and photocleavable spacers linked to hydrogel microspheres (35). After encapsulation primers are released by photocleavage (i.e., 8-min exposure to >350-nm UV radiation), cDNA synthesis is

initiated, and the synthesized cDNAs remaining trapped in each droplet are tagged with a barcode during reverse transcription. After synthesis of the second-strand cDNA, linear amplification is carried out by IVT. Subsequently, encapsulated droplets are broken and the amplified antisense RNA is fragmented using zinc-ion-mediated cleavage. Cleaved RNA is converted into a cDNA library for next-generation sequencing by a second reverse transcription reaction and a few cycles of PCR. Finally, the cDNA libraries are sequenced on Illumina sequencing platforms (34, 35). This general strategy used follows that developed by Hashimshony et al. (29) and Jaitin et al. (30), but several steps have been added or revised. For example, in this methodology, the 3'-5' exonuclease ExoI and dsDNA-specific restriction endonuclease HinfI are used to digest unused ssDNA primers and cleaves the T7 promoter sequence in primer dimers and hairpins, respectively (35).

Recently, a novel method was developed that combines exponential amplification of a limited number of PCR cycles with linear amplification of T7-based IVT for RNA amplification (38). The novelty of this technique lies in that it allows for high-fidelity gene expression profiling of individual cells and for significant reduction of 3'-bias; this enables detection of all regions of RNA transcripts. To achieve this objective, some protocol modifications have been made. These modifications include: (1) the introduction of "extending primers" that contain a Kozak sequence at the 3'-ends during PCR which allows capture of 5'-ends of a gene's coding sequence; (2) the combination of modified polyT and modified random primers to make the PCR step more efficient and to improve the preservation of relative gene abundance, which secures full-length RNA coverage, and eliminating carryover of reverse transcriptase that greatly diminishes reverse transcriptase inhibitory effect on the following PCR steps; (3) using a SmartSpec 3000 spectrophotometer (Bio-Rad), the RNA yield was evaluated, and the yield of amplified RNA came up to 200–250 µg from a single cell, which is sufficient to apply to any RNA sequencing technique (38). In a word, this new technique has great potential for future applications.

## Phi29 DNA Polymerase-Based Amplification

Derived from the *Bacillus subtilis* phage Phi29, the Phi29 DNA replicative polymerase has strong strand displacement activity (39). Phi29 DNA polymerase based RNA amplification has been successfully applied to construct cDNA libraries from single prokaryotic (40) and eukaryotic cells (23). Reverse transcribed single strand cDNA tends to self-ligate and to amplify like a rolling circle in the presence of Phi29 DNA polymerase, so Phi29 DNA polymerase-based amplification is also referred to as rolling circle amplification.

### PMA

Phi29 DNA polymerase-based mRNA transcriptome amplification (PMA), an approach for single and low quantities cell cDNA amplification, is the first strategy using Phi29 DNA polymerase to construct cDNA libraries for the sequencing of whole mRNA transcriptomes. The PMA method allows for full-length transcript coverage to be obtained. Methods for

multiple displacement amplification of circularized cDNA have been discussed in detail by Liu and her colleagues (24).

### TTA

Another single-cell RNA amplification method based on Phi29 DNA polymerase is total transcript amplification (TTA) (40). This transcript amplification method involves multiple rounds of PCR and/or linear amplification of cDNA and is challenged by low amounts of RNA and lack of polyA-tails for easy tagging and mRNA specific amplification (5, 33, 40). However, the TTA strategy yields reproducible data, low fold-change bias, and a high number of genes efficiently amplified (40).

## MALBAC-RNA

In addition to the above methods, multiple annealing and looping based amplification cycles (MALBAC) is also applied to single-cell transcriptome amplification (termed MALBAC-RNA). MALBAC-RNA was developed and first used in whole genome amplification by Xie et al. (41, 42). This strategy boasts high detection efficiency, accuracy, and reproducibility. After cell lysis and reverse transcription, primers with 7 random nucleotides are annealed to the synthesized first-strand cDNA and are then extended by DNA polymerase with strand displacement activity. Amplicons are then melted off the original template after extension, and are looped to protect themselves from being further amplified. This MALBAC-RNA step includes a total of 10 cycles of quasilinear amplification (MALBAC-RNA pre-amplification), followed by another 19 cycles of PCR. Finally, these MALBAC amplified DNA products are directly used for sequencing library preparation and are then sequenced on an Illumina HiSeq 2000 instrument (42).

Recently, several other strategies have been exploited for single-cell transcriptome amplification and sequencing, including multiple annealing and dC-tailing based quantitative single-cell RNA sequencing (MATQ-seq) (43), CRISPR droplet sequencing (CROP-seq) (44), and geographical position sequencing (Geo-seq) (45).

## MATQ-seq

MATQ-seq is a highly sensitive and quantitative method for single-cell sequencing of total RNA, including noncoding and nonpolyadenylated RNA, and was inspired by the primers used in MALBAC (43). In MATQ-seq, reverse transcription is performed with a primer that consisting of the MALBAC primer, three consecutive G or T, and 20 consecutive T (dT<sub>20</sub>). This primer design differs from MALBAC, in that it mainly contains G, A, and T bases. After 9 cycles of reaction, primers and RNAs are digested by T4 DNA polymerase and RNase H and RNase I, respectively. Then, polyC tailing is performed on the cDNA, and the second-strand synthesis initiates with G-enriched MALBAC primers, containing a random hexamer UMI sequence (referred to as amplicon indexes); the use of these primers decreases the 3'- or 5'-end bias to a great extent. After amplification, Illumina Truseq adaptors are ligated to the samples, purified cDNA libraries are pooled as instructed by Illumina protocol, and libraries are sequenced on the Illumina Nextseq 500 platform (43).

## CROP-seq

Combining pooled CRISPR screening with scRNA-seq, a new method called CROP-seq was developed. CROP-seq involves directly linking guide-RNA (gRNA) expression to transcriptome responses in thousands of individual cells (44). In this strategy, a gRNA vector (LentiGuide-Puro) is introduced, because CRISPR gRNA lacks a polyA tail and is not detectable with current scRNA-seq assays, which makes CROP-seq compatible with previously developed scRNA-seq methods. A detailed protocol was reported in the *Journal of Nature Methods* (44). Briefly, single cells are co-encapsulated with barcoded beads and resuspended in Drop-seq lysis buffer. Reverse transcription and exonuclease I treatment are then performed, and an additional ten enrichment cycles using the Illumina Nextera XT i7 primers, along with the Drop-seq New-P5 SMART-PCR hybrid oligonucleotides, are performed to enrich the cDNA library. Finally, cDNA libraries are sequenced with paired-end SBS chemistry on using an Illumina HiSeq 3000/4000 instrument. Using sophisticated molecular technologies, Datlinger's team constructed a viral vector that can help researchers see the CRISPR gRNA in single-cell sequencing (44). With this technique, thousands of genome editing events in single cell were analyzed in a high throughput manner; this will facilitate high-throughput functional dissection of complex regulatory mechanisms and heterogeneous cell populations.

## Geo-seq

*In situ* position information of cells is lost by current single-cell sequencing strategies. However, the native spatial information is critical to some research fields, such as developmental biology and tumor biology. Driven by this necessity, a novel single-cell transcriptome sequencing technique called Geo-seq was exploited. Geo-seq combines LCM and scRNA-seq technology (45). In this approach, LCM-sorted single cells are lysed for first-strand cDNA synthesis, and a modified Smart-seq2 protocol is performed to amplify the cDNA. Of the many scRNA-seq methods, Smart-seq2 was chosen for the ability of achieving full-length cDNA synthesis with a low amplification bias in a short time period (27, 46). Geo-seq is also compatible with Brady/Tang's method (5, 19). cDNAs that pass quality control are then used to construct the library for next-generation sequencing (45). As an alternative method, Tomo-seq has been devised for spatial gene profiling (47). Geo-seq is a high-efficiency, high-resolution strategy for spatial transcriptome analysis, which can be used in 3D reconstruction of the transcription profile, and can also be used to study the transcriptome information of a small amount of tissues or cells with special structures. The Geo-seq technique shows wide application potentials to address biological and pathological questions of cellular properties such as prospective cell fates, biological functions and gene regulatory networks (45).

## Split-Pool Ligation-Based Transcriptome Sequencing

Amazingly, a latest article published in *Science* reported a low-cost scRNA-seq method, called "split-pool ligation-based transcriptome sequencing" (SPLiT-seq) (48). In SPLiT-seq, four

rounds of combinatorial barcoding are applied to an individual transcriptome. In the first round, formaldehyde fixed cells are added into a well-plate, and in-cell reverse transcription is accomplished using well-specific barcoded primers. Next, all cells are collected and redistributed into a new well-plate, where an in-cell ligation reaction appends a second well-specific barcode to the transcribed cDNA. The third round of barcoding applies barcodes containing unique molecular identifiers (UMIs). After that, the cells are pooled, divided into sublibraries, and sequencing barcodes are introduced by PCR; meanwhile, the fourth barcodes are led into. Finally, after paired-end sequencing on a MiSeq or NextSeq system (Illumina), each individual transcriptome is assembled by combinatorial reads that combine the same four barcodes (48).

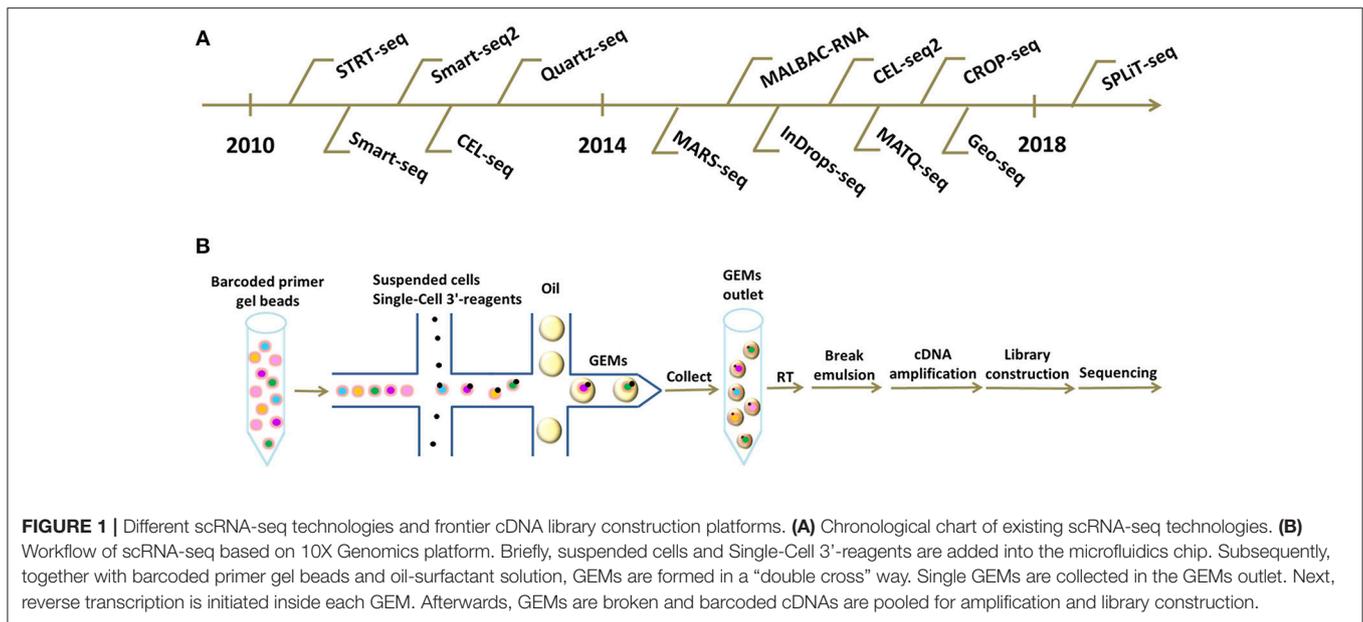
SPLiT-seq is compatible with fixed cells or nuclei, and requires no partitioning of single cells into individual compartments, but completely relies on the cells themselves as compartments. This novel technique labels the cellular origin of RNA through combinatorial DNA barcodes only with pipetting steps and requires no complex apparatus, which greatly simplify the workflow, decreases the cost, and widely facilitates the scalable profiling of single cells (48). As this method emerges as a viable sequencing technique, we believe that the cost of scRNA-seq will be greatly reduced and the application of scRNA-seq will become more efficient and widespread. A chronological chart of all these existing scRNA-seq technologies listed above is shown in **Figure 1A**.

## Frontier cDNA Library Preparation Platforms in Single-cell RNA-seq

For years, Illumina systems have been regarded as the leaders in the field of cDNA library construction in scRNA-seq. However, at the Advances in Genome Biology and Technology (AGBT) meeting in 2015, the GemCode platform based on 10X Genomics was officially launched, which immediately raised great interest in the research community. A year later, Chromium<sup>TM</sup> single cell 3'-solution, based on 10X Genomics platform as an updated version of GemCode platform, was launched. The Chromium<sup>TM</sup> methodology greatly accelerates the applications of scRNA-seq and assists digital single-cell transcriptome sequencing analysis. Within 3 years, scRNA-seq based on the 10X Genomics platforms has exploded in popularity (49–60).

Single-cell transcriptome sequencing, using the 10X Genomics platforms, can obtain expression information from 1,000 to 10,000 cells simultaneously, and can analyze the grouping of cells and detect gene expression differences between cell populations. This method is currently regarded as the "golden method" to evaluate tumor heterogeneity, immune cell populations, and embryonic development.

The workflow on 10X Genomics platform is summarized below (**Figure 1B**). Initially, suspended cells are combined with GemCode Single-Cell 3' reagents in one channel of an 8-channel microfluidics chip. Subsequently, millions of barcoded primer gel beads and oil-surfactant solution from other channels are introduced to form GEMs (Gel Bead-In-Emulsions) in a "Double



Cross" way, GEMs containing single cells are then collected in the GEM outlet. Next, reverse transcriptions are initiated inside each GEM with primers containing Illumina adaptors, 10X barcodes, UMIs, and oligo(dT) sequences. Afterwards, the GEMs are broken and barcoded cDNAs are pooled for amplification and library construction (59). GEM formation is the core of this technology and is vital to the success of scRNA-seq.

10X Genomics platforms are high throughput technologies that agglomerate single-cell isolation, cDNA amplification and library construction into one process, which greatly reduces sample manipulation and saves cost and time. Moreover, the 10X Genomics platform boasts up to 65% single cell capture rate, and cell suspension preparation, single cell capturing, cDNA amplification, and library construction can be achieved within a single day. Taking these advantages into consideration, we believe that 10X Genomics platforms will have wide application prospects in the field of scRNA-seq.

## SINGLE-CELL RNA SEQUENCING APPLIED IN NORMAL HEMATOPOIESIS

The hematopoietic system is well established as a paradigm for studying cellular hierarchies. Traditional approaches to study hematopoiesis involve purification of cell populations based on a series of surface markers and functional assays; these assessments obscure underlying heterogeneity contained within any phenotypically defined cell population (61). In recent years, advances in scRNA-seq have opened new avenues to characterize heterogeneity in a large variety of cellular systems, and will lead to the discovery of novel subpopulations (62–65).

### Single-Cell RNA Sequencing in HSPCs

Tang et al. demonstrated that by applying both Smart-seq2 and InDrops scRNA-seq approaches to define cell heterogeneity, they were able to identify a group of classically defined and erythroid-primed hematopoietic stem and progenitor cells (HSPCs) (62). What's more, Paul et al. identified 18 unrecognized subpopulations with distinct lineages priming toward specific lineage fates, as compared to the traditional definition of myeloid progenitors by analyzing HSPCs by scRNA-seq (63). They discovered that *Cebpa* and *Cebpe* are two cell fate decision transcription factors found in myeloid progenitors. High levels of *Cebpa* expression direct myeloid progenitors to neutrophil, monocyte, and basophil lineages, whereas high levels of *Cebpe* expression seem to be found primarily in neutrophil and eosinophil subpopulations (63). Thus, scRNA-seq was recommended as an ideal tool for demarcating developmentally plastic and highly dynamic cell populations in hematopoiesis. Additionally, applying scRNA-seq to both standard HSC and their progeny multipotent progenitor 1 (MPP 1), three main clusters with distinct features were separated, including active, quiescent, and a novel, undefined fraction of cells (64). Furthermore, Kowalczyk et al. found extensive transcriptome variability among HSCs assessed by scRNA-seq (66). Additionally, scRNA-seq is able to determine lineage heterogeneity among stem cell compartments. Tsang et al. applied scRNA-seq to 180 highly purified *Bcl11a*<sup>-/-</sup> and *Bcl11a*<sup>+/+</sup> HSCs, and found that *Bcl11a*<sup>-/-</sup> HSCs consist of two distinct myeloerythroid-restricted subpopulations (65). Moreover, scRNA-seq was used to distinguish aneuploid cells from diploid cells within the hematopoietic stem and progenitor cells. Zhao et al. performed scRNA-seq on bone marrow-derived CD34<sup>+</sup> cells from patients with bone marrow failure and cytogenetic abnormalities, distinguishing

aneuploidy hematopoietic stem and progenitor cells from diploid cells (67).

In addition, scRNA-seq was successfully applied to tracing the formation of HSCs (68, 69). In addition to haemogenic endothelial cells, HSCs are derived from pre-HSCs, which consist of at least two HSC-competent intermediates around embryonic day 11 (E11), namely, CD45<sup>+</sup> pre-HSCs and CD45<sup>-</sup> pre-HSCs. Applying scRNA-seq to CD45<sup>+</sup> and CD45<sup>-</sup> pre-HSCs in E11 mouse aorta-gonad-mesonephros, Zhou *et al.* revealed that there existed a remarkable heterogeneity in relevant populations during the stepwise development of HSC from the pre-HSC stage, and showed a continuous development from endothelial cells to HSCs through CD45<sup>+</sup> and CD45<sup>-</sup> pre-HSCs (68). Autophagy plays an essential role in self-renewal and differentiation in embryonic hematopoiesis. Recently, Hu *et al.* applied scRNA-seq to five cell populations related to HSC formation and reported that the transcription activity of autophagy-related genes was substantially increased when endothelial cells committed to pre-HSCs, suggesting that autophagy plays a critical role in HSC formation (69).

scRNA-seq has also been widely applied to hematopoietic cell differentiation (70–74). HSCs sit at the very top of a differentiation hierarchy, which produces the full spectrum of mature blood cells via intermediate progenitors (70). Using scRNA-seq, Nestorowa *et al.* profiled 1,656 single HSPCs to reconstruct differentiation trajectories, and revealed the dynamic expression changes associated with early lymphoid, erythroid, and granulocyte-macrophage differentiation (70). This enabled the reconstruction of an atlas showing the differentiation trajectories of HSPCs. Additionally, with phenotypic and functional assays, at least four subpopulations were revealed within the B220<sup>+</sup>CD117<sup>int</sup>CD19<sup>-</sup>NK1.1<sup>-</sup> uncommitted hematopoietic progenitor, of which the Ly6D<sup>+</sup>SiglecH<sup>-</sup>CD11c<sup>-</sup> subtype exhibited strong B-cell potential, whereas the Ly6D<sup>-</sup>SiglecH<sup>-</sup>CD11c<sup>-</sup> subtype showed mixed lympho-myeloid potential (71). Applying scRNA-seq to the Ly6D<sup>+</sup>SiglecH<sup>-</sup>CD11c<sup>-</sup> subtype, Alberti-Servera *et al.* identified a new subset, which was thought to be the direct precursor of the first B-cell committed stage (71). Additionally, combining index FACS sorting with computational reconstruction of thrombocyte developmental chronology from scRNA-seq data, Macaulay *et al.* revealed the continuous nature of thrombocyte lineage commitment from stem cells to mature cells (72). Similarly, Velten *et al.* applied scRNA-seq to human CD34<sup>+</sup> HSPCs, and showed that human HSC lineage commitment is a continuous process (73). In another study, 2000 HSCs from young and old *Vwf-EGFP* mice were isolated, respectively, and 61-young and 74-old HSCs were captured and prepared for scRNA-seq. Researchers identified increased molecular platelet priming and functional platelet bias as the predominant age-dependent change to HSCs, including a significant increase in a previously unrecognized class of HSCs, a *Vwf*<sup>+</sup> platelet-biased HSC subpopulation, that exclusively produce platelets, which demonstrated that there exists molecular and functional platelet bias in aged HSCs (74). Additionally, they found

that FOG-1 is an intrinsic regulator of *Vwf*<sup>+</sup> platelet-biased HSC subset for the first time. Data shows that FOG-1 is critical to the maintenance of the *Vwf*<sup>+</sup> platelet-biased HSC subset (74).

scRNA-seq has also been applied to evaluate the regulation of signaling pathways in HSCs (65, 74, 75). For example, Cabezas-Wallscheid *et al.* applied scRNA-seq to determine vitamin A in the maintenance and function of HSCs, they uncovered a role for vitamin-A retinoic acid signaling in the regulation of HSC dormancy, and they reported that the transition from dormancy toward cell-cycle entry is a continuous developmental path associated with upregulation of biosynthetic processes rather than a stepwise progression (75). Furthermore, applying scRNA-seq to *Bcl11a*<sup>-/-</sup> and *Bcl11a*<sup>+/+</sup> HSCs, Tsang *et al.* discovered that transcription rates periodically increase from G1 to S/G2/M phase, suggesting that *Bcl11a* deficiency is associated with increased proliferation in the HSC compartment (74). In these cases, it was demonstrated that scRNA-seq is a powerful approach to identify novel intrinsic or extrinsic regulators of HSC function.

scRNA-seq has been applied to identify distinct myeloid progenitor differentiation pathways, and has the potential to simultaneously resolve cellular heterogeneity and to identify the molecular determinants that are selectively and uniformly expressed in novel subgroups (76). Lymphoid-primed multi-potent progenitors (LMPPs) and common myeloid progenitors (CMPs) are regarded as the first branch point for separating lineage commitment pathways from HSCs. However, it has been proposed that both differentiation pathways generate myeloid innate immune cells through the same myeloid restricted pre-granulocyte macrophage progenitor (pre-GMP). Applying scRNA-seq to pre-GMPs, Drissen *et al.* identified two distinct myeloid differentiation pathways, namely the *Gata1*-expressing pathway, which generates mast cells, eosinophils, megakaryocytes, and erythroid cells, and the *Gata1*-negative pathway, which generates monocytes, neutrophils, and lymphocytes (76).

scRNA-seq can provide a detailed molecular characterization of single cells, and is highly complementary to traditional differentiation or FACS-based phenotyping approaches (77). To individually profile 20,000 hematopoietic progenitor cells, without prior enrichment or depletion of individual lineages based on surface markers from human cord blood, a massively parallel scRNA-seq approach, Drop-seq, was applied; this approach identified intermediate stages that simultaneously co-express “primed” programs for multiple downstream lineages and revealed striking heterogeneity in the early molecular transitions between myeloid subsets during human cord blood hematopoiesis (78).

### Single-Cell RNA Sequencing in Lymphoid Cells

Lymphoid cells are derived from HSCs, and are mainly responsible for body immunity. scRNA-seq has rarely been applied to studying lymphoid cells in hematopoiesis. In one example, Yu *et al.* applied scRNA-seq to bone marrow innate lymphoid cells (ILCs); they identified distinct ILC progenitor

subsets, including a novel programmed death 1 high expression (PD-1<sup>hi</sup>) ILC progenitor subset, and delineated distinct ILC developmental states and pathways (79). In another example, Patil et al. applied scRNA-seq to more than 9000 CD4<sup>+</sup> T cells from the peripheral blood of healthy donors to unravel heterogeneity among CD4<sup>+</sup> cytotoxic T lymphocytes (CD4-CTLs), and to evaluate the transcriptional profiles and clonality within the population; they identified that CD4-CTL is derived from a subset of effector memory T cells expressing CD45RA(T<sub>EMRA</sub>) (80). Identification of the CD4-CTL precursor population may enable further research of how CD4-CTLs arise, and could elucidate the mechanisms that may be applied to generate durable and effective CD4-CTL immunity (80).

## Single-Cell RNA Sequencing in Myeloid Cells

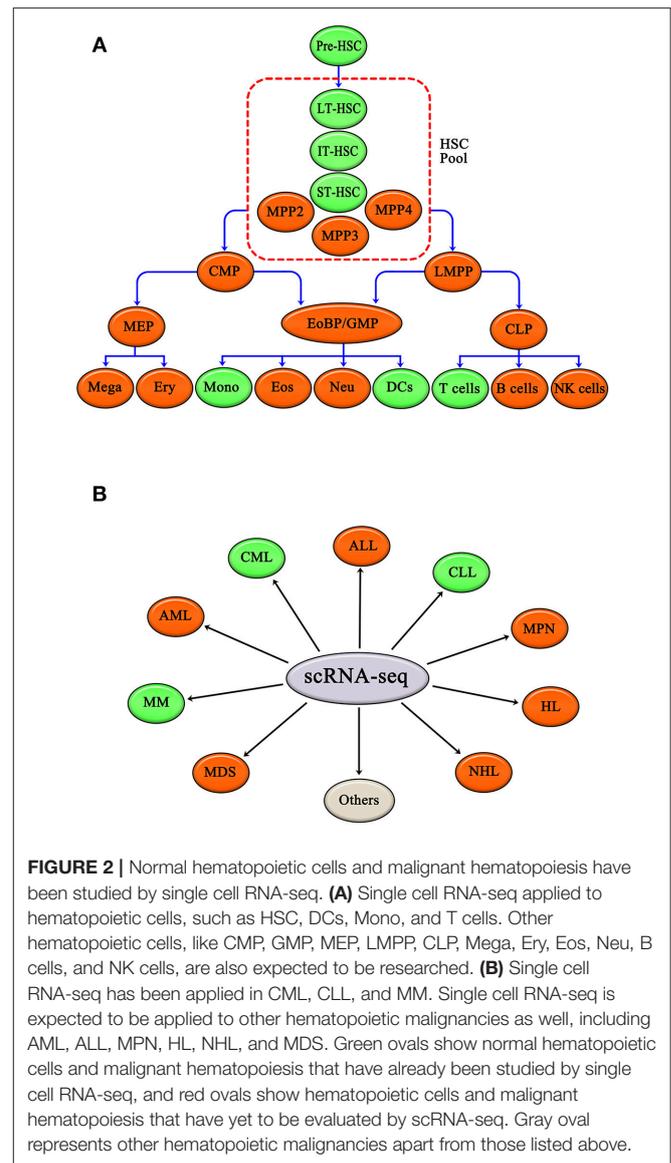
### Dendritic Cells

Dendritic cells (DCs) are immune cells that are differentiated from granulocyte monocyte progenitors (GMPs). The main functions for DCs are antigen sensing and presentation. Little is known about the transcriptional heterogeneity within cells in responses to antigens. scRNA-seq was used to investigate heterogeneity in the response of bone-marrow derived DCs to lipopolysaccharide (LPS), and revealed an extensive, bimodal variation in mRNA abundance and splicing patterns, which demonstrated previously unobserved levels of heterogeneity between DCs (81). What's more, Shalek et al. used scRNA-seq to show that bone-marrow derived DCs, stimulated under different conditions *in vitro*, demonstrate variable responses that are mediated by interferon paracrine signaling (82). Additionally, using scRNA-seq, Villani et al. identified six human DC clusters, including two clusters mapping closely to the well-established DC subsets and four new clusters. These newly identified DC subsets included one that shares properties with plasmacytoid dendritic cells (pDCs) but potently activates T cells, redefining the roles for pDCs (83).

Generally speaking, the expression of transcription factors and abundance of surface proteins determine cell fate specification during differentiation. Applying scRNA-seq to mouse bone marrow conventional dendritic cells (cDCs), Papalexis et al. identified that Siglec-H and Ly6C abundance determines whether cDCs will become cDC type 1 (cDC1, Siglec-H<sup>-</sup>Ly6C<sup>-</sup>) or cDC type 2 (cDC2, Siglec-H<sup>-</sup>Ly6C<sup>+</sup>) (84).

### Monocyte

scRNA-seq has also been applied to identifying heterogeneity within monocyte populations. Villani et al. sequenced 372 single blood monocytes by scRNA-seq, and identified four monocyte subsets, including two known and two new clusters (83). In addition, 1,700 suspended human monocytes were sequenced by 10X Genomics Chromium, revealing differences in monocyte differentiation fate; specifically, IRF4 and MAFB were found to be essential for differentiation of monocyte-derived dendritic cells (mo-DCs) and differentiation of monocyte-derived macrophages (mo-Macs), respectively. Moreover, aryl hydrocarbon receptor activation promotes differentiation of mo-DCs and inhibits differentiation of mo-Macs through BLIMP-1 (52).



**FIGURE 2** | Normal hematopoietic cells and malignant hematopoiesis have been studied by single cell RNA-seq. **(A)** Single cell RNA-seq applied to hematopoietic cells, such as HSC, DCs, Mono, and T cells. Other hematopoietic cells, like CMP, GMP, MEP, LMPP, CLP, Mega, Ery, Eos, B cells, and NK cells, are also expected to be researched. **(B)** Single cell RNA-seq has been applied in CML, CLL, and MM. Single cell RNA-seq is expected to be applied to other hematopoietic malignancies as well, including AML, ALL, MPN, HL, NHL, and MDS. Green ovals show normal hematopoietic cells and malignant hematopoiesis that have already been studied by single cell RNA-seq, and red ovals show hematopoietic cells and malignant hematopoiesis that have yet to be evaluated by scRNA-seq. Gray oval represents other hematopoietic malignancies apart from those listed above.

To sum up, scRNA-seq has been applied to study various hematopoietic cells, including HSCs, DCs, monocytes, and T cells. Other hematopoietic cells, such as common myeloid progenitors (CMP), granulocyte-monocyte progenitors (GMP), megakaryocyte-erythroid progenitors (MEP), lymphoid-primed multipotential progenitors (LMPP), common lymphoid progenitors (CLP), megakaryocytes, erythrocytes, eosinophils, neutrophils, B cells and natural killer cells (NK cells), are also expected to be studied using scRNA-seq (Figure 2A).

## SINGLE-CELL RNA SEQUENCING APPLIED IN MALIGNANT HEMATOPOIESIS

Apart from its broad use in studying physiological hematopoiesis, scRNA-seq has also been extensively applied to malignant hematopoiesis. Leukemia, lymphoma, multiple myeloma (MM),

myeloproliferative neoplasm (MPN), and myelodysplastic syndrome (MDS) are the common hematopoietic malignancies which seriously threaten health and even the survival of patients. Currently, scRNA-seq has been used to study leukemia (85, 86) and multiple myeloma (87) in order to better understand these diseases and find more effective therapeutic strategies.

Combining high sensitivity mutation detection with scRNA-seq, Giustacchini et al. analyzed more than 2000 stem cells (SCs) from patients with chronic myeloid leukemia (CML), revealing previously unrecognized heterogeneity of CML-SCs, including the identification of a subset of CML-SCs with a distinct molecular signature that selectively persisted during prolonged therapy. Furthermore, a blast-crisis-specific SC population was identified, which was also present in a subclone of CML-SCs during the chronic phase (85).

scRNA-seq has also revealed transcriptional heterogeneity in chronic lymphoid leukemia (CLL). Wang et al. performed single-cell whole-transcriptome sequencing of CLL samples, and identified significant transcriptional heterogeneity representing a variety of cellular processes, including cell cycle, immune signaling, antigen presentation, phospholipid binding, and protein folding (86). Furthermore, they combined scRNA-seq with targeted mutation detection and identified *LCPI* and *WNKI* mutations as novel CLL driving events (86).

Moreover, scRNA-seq has been applied in MM to accurately identify chromosomal translocations and to facilitate differentiation of normal plasma cells from malignant cells in an unbiased fashion (87). Unsupervised hierarchical clustering of MM cells between two patients revealed different subtypes of MM cells. Chromosomal translocations, such as the t(11;14) translocation and t(6;14) translocation, are important markers for stratifying the clinical risks of MM (88). Using single circulation tumor cell (CTC) RNA-seq to identify chromosomal translocations, Lohr et al. identified which genes are the most differentially expressed between single MM CTCs and normal B lymphocytes, suggesting that single CTC gene expression profiling may be applied for the diagnosis and classification of MM (87). scRNA-seq has the potential for great clinical utility and can elucidate precision medicine approaches for the treatment of relapsed/refractory MM.

Additionally, scRNA-seq could also be used to identify putative leukemia stem cell populations based on “stemness” signatures, using available data (89), or to predict *de novo* stem cell identities from scRNA-seq data, using recent evolutionary algorithms (90).

scRNA-seq has been applied to several hematopoietic malignancies, including CML, CLL, and MM; and other malignancies, including acute myeloid leukemia (AML), acute lymphoid leukemia (ALL), MPN, Hodgkin’s lymphoma (HL), non-Hodgkin’s lymphoma (NHL), and MDS, are likely to be evaluated by scRNA-seq in the future (**Figure 2B**).

## OUTLOOK

Hematopoiesis is the process of development of normal hematopoietic cells, which is an ordered multi-step process

originated from a rare population of multipotent and self-renewing HSCs within the bone marrow. During malignant hematopoiesis, including leukemias, MDS, and MM, et al., the uncontrolled proliferation and blocked differentiation of precursors sustains malignancy growth at the expense of normal blood cell production in bone marrow. Normal and malignant hematopoiesis is perhaps the best-defined hierarchy model and has led the way in the application of new single-cell approaches to unravel cellular heterogeneity.

Before the advent of scRNA-seq, discovering novel cell subsets in normal and malignant hematopoiesis mainly relies on cell surface markers via FACS-sorting, which required prior knowledge or informed guesswork. The advances in high-dimensional, single-cell techniques have since enabled unbiased, alternative workflows to sequence cells with no need for prior knowledge of genes and proteins, and grouping of cells is accomplished based on their transcriptional signatures (84). So far, scRNA-seq has shown great effectiveness in unraveling complex cell populations, reconstructing developmental trajectories, and modeling transcriptional dynamics in but not restricted to normal and malignant hematopoiesis (91).

Cellular heterogeneity is present in nearly every organ and tissue, as well as tumors, which hinders efforts to unravel cellular functions and the roles of different cell populations, and presents a challenge to the development of precision medicine. scRNA-seq is a powerful tool to reveal cell-to-cell variation and cellular heterogeneity, discover new cell types and characterize malignant hematopoiesis and other tumorigenesis (92). Over the years, cellular heterogeneity has been evaluated in many tissues and cells using scRNA-seq (93–103). scRNA-seq has been applied to solve heterogeneity within hematopoietic cells as well (63, 64, 79, 81, 83). Additionally, RNA sequencing combined with single-cell network analysis has identified *DDIT3* as a nodal lineage regulator in hematopoiesis (104), indicating that scRNA-seq and network analysis will advance our knowledge of hematopoiesis. Subsequent bioinformatics analysis of scRNA-seq data can be used to identify key genes involved in embryonic development, immune responses, and carcinogenesis. The maturation of scRNA-seq techniques has made it easier to evaluate gene expression profiles at single-cell and single-base resolution (21, 105). More recently, Athanasiadis et al. applied scRNA-seq to uncover transcriptional states and fate decisions in hematopoiesis, to order cells along their differentiation hierarchy, and to computationally reconstruct the blood lineage tree (106). scRNA-seq has been applied to guide the diagnosis and treatment of the patients with malignant hematological diseases in recent years. scRNA-seq of CML-SCs at diagnosis of patients predicts molecular response to tyrosine kinase inhibitor treatment (85). In addition, CLL scRNA-seq reveals transcriptional heterogeneity and uncovers mutated *LCPI* and *WNKI* as novel CLL drivers (86), they are likely to be novel drug targets for the treatment of CLL. Furthermore, scRNA-seq enables classification of MM (87), which is vital to future personalized therapy and precision medicine.

Besides scRNA-seq, single-cell qPCR has also been used to study hematopoiesis. Moore et al. performed transcriptional profiling of single hematopoietic cells by single-cell qPCR, and

**TABLE 2** | The advantages and disadvantages of three main single-cell cDNA amplification methods.

Amplification methods	Advantages	Disadvantages	References
PCR-based amplification	Full-length transcript	Introduce variability Introduce discrepancy	(21)
T7-based IVT amplification	Low amplification bias	High 3'-end bias Unable to detect full transcript	(29)
Phi29-based amplification	Low quantity of cDNA Full-length transcript	Remove genomic DNA firstly	(24)

analyzed normal, aberrant, and malignant hematopoiesis in zebrafish (107). Compared to scRNA-seq, single-cell qPCR has no need for special and expensive reagents and instruments. However, it relies on specific primer design and is limited to detect known target genes, which introduces a source of subjectivity. Additionally, limited by the amount of RNA, only a dozen of highly expressed genes can be detected by single-cell qPCR. Worse still, it is prone to false positive findings and only suitable for a small number of cells. Compared to scRNA-seq, traditional RNA-sequencing ignores the cell-to-cell variation and cannot truly reflect natural heterogeneity, but it largely promotes our understanding of hematopoiesis before single-cell technique emerged.

As reviewed above, traditional views of the hematopoietic hierarchy have been challenged, modified, enriched, and redefined by scRNA-seq. This new information is leading us to major conceptual shifts and more precise and unambiguous understanding of hematopoietic system, including, but not limited to, hematopoietic cells fate decisions. In addition, gene expression network regulation of hematopoiesis has been elucidated using scRNA-seq (64, 74, 75). Deeper knowledge of regulation of hematopoiesis is helpful to better understand hematopoietic malignancy, and will facilitate new drug research and development. Additionally, scRNA-seq can be applied to other hematopoietic progenitors and mature blood cells, including CMPs, GMPs, MEPs, LMPP, CLPs, megakaryocytes (Mega), erythrocytes (Ery), Eosinophils (Eos), Neutrophils (Neu), B cells, and NK cells (**Figure 2A**). In addition, scRNA-seq is expected to be applied to other malignant hematological system diseases as well, including AML, ALL, HL, NHL, MPN, and MDS (**Figure 2B**).

Nevertheless, it still remains a challenge to hone and refine scRNA-seq techniques. A critical step in scRNA-seq is whole transcriptome amplification (WTA), which comprises reverse transcription of mRNA into cDNA, followed by cDNA amplification. Currently, there are three main cDNA amplification approaches: PCR-based amplification, T7-based IVT amplification, and Phi29 DNA polymerase-based amplification. The advantages and disadvantages of these techniques are summarized in **Table 2**. Briefly, PCR-based amplification of cDNA, being exponential in nature, is prone to introduce variability and discrepancies, especially among low-abundance transcripts (21). T7-based IVT cDNA amplification is a linear amplification method, which will theoretically reduce amplification bias compared to PCR-based methods (29). However, IVT amplification shows a very high 3'-end bias and fails to detect the full spectrum of transcripts, whereas

PCR-based protocols are capable of amplifying full-length cDNA. Phi29-based amplification strategies are suitable for low cDNA amplification, and can amplify full-length transcripts, but genomic DNA (gDNA) must be removed before amplification (24). The most appropriate scRNA-seq method should be selected according to the specific experimental requirements.

Researchers are exploring how best to analyze the vast quantities of sequencing data that these methods produce, and which algorithm is the most useful. Challenges remain regarding comparing gene expression sequencing data between different experimental conditions or different organisms, and integrating different kinds of omics data. While data from thousands of single cells can be tricky to analyze, software advances are making it easier. It is expected that more sophisticated and useful analytical frameworks and advanced software will be rapidly developed to handle these challenges. One frustration that accompanies scRNA-seq is a high instance of noise, and subtle changes in amplification and capture efficiency can result in huge differences between cells, in the case of such small amounts of starting RNA material. We must be vigilant against batch effects and introducing variation in samples due to different processing dates or use of different reagent batches.

In conclusion, scRNA-seq is a powerful technique that can enable better understanding of normal and malignant hematopoiesis although some challenges remain to be resolved. The advent of novel technologies, such as the 10X Genomics based GemCode platform and the Chromium single-cell 3'-solution, further drives the application of scRNA-seq. We envision that the wide application of scRNA-seq to study the hematopoietic microenvironment will be very helpful for understanding the complexity of hematopoiesis and will provide insights into the diagnosis and prognosis of patients with hematological diseases.

## AUTHOR CONTRIBUTIONS

XH wrote this review, PQ and YZ searched some of the literature, SX and XL revised the manuscript, and SX and JC designed the framework. SX and JC are co-corresponding authors.

## FUNDING

This work has been supported by the Military Logistics Major Research Project (No. AMS17J007), Youth Cultivation Project of Medical Science and Technology of Chinese PLA (No. 15QNP058), and Southwest Hospital's Major Innovative Plan for WIT120 (No. SWH2017ZDCX4208).

## REFERENCES

- Liang J, Cai W, Sun Z. Single-cell sequencing technologies: current and future. *J Genet Genomics* (2014) 41:513–28. doi: 10.1016/j.jgg.2014.09.005
- Wen L, Tang F. Single cell epigenome sequencing technologies. *Mol Aspects Med.* (2018) 59:62–9. doi: 10.1016/j.mam.2017.09.002
- Novick A, Weiner M. Enzyme induction as an all-or-none phenomenon. *Proc Natl Acad Sci USA.* (1957) 43:553–66. doi: 10.1073/pnas.43.7.553
- Coskun AF, Eser U, Islam S. Cellular identity at the single-cell level. *Mol Bio Syst.* (2016) 12:2965–79. doi: 10.1039/C6MB00388E
- Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, et al. mRNA-seq whole-transcriptome analysis of a single cell. *Nat Methods* (2009) 6:377–82. doi: 10.1038/nmeth.1315
- Fend F, Raffeld M. Laser capture microdissection in pathology. *J Clin Pathol.* (2000) 53:666–72. doi: 10.1136/jcp.53.9.666
- Shapiro E, Biezuner T, Linnarsson S. Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat Rev Genet.* (2013) 14:618–30. doi: 10.1038/nrg3542
- Yu ZT, Aw Yong KM, Fu J. Microfluidic blood cell sorting: now and beyond. *Small* (2014) 10:1687–703. doi: 10.1002/smll.201302907
- Gross A, Schoendube J, Zimmermann S, Steeb M, Zengerle R, Koltay P. Technologies for single-cell isolation. *Int J Mol Sci.* (2015) 16:16897–919. doi: 10.3390/ijms160816897
- Huang WE, Ward AD, Whiteley AS. Raman tweezers sorting of single microbial cells. *Environ Microbiol Rep.* (2009) 1:44–9. doi: 10.1111/j.1758-2229.2008.00002.x
- Yilmaz S, Singh AK. Single cell genome sequencing. *Curr Opin Biotechnol.* (2012) 23:437–43. doi: 10.1016/j.copbio.2011.11.018
- Bell TJ, Eberwine J. Live cell genomics: cell-specific transcriptome capture in live tissues and cells. *Methods Mol Biol.* (2015) 1324:447–56. doi: 10.1007/978-1-4939-2806-4\_30
- Kurimoto K, Yabuta Y, Ohinata Y, Saitou M. Global single-cell cDNA amplification to provide a template for representative high-density oligonucleotide microarray analysis. *Nat Protoc.* (2007) 2:739–52. doi: 10.1038/nprot.2007.79
- Hongoh Y, Sharma VK, Prakash T, Noda S, Taylor TD, Kudo T, et al. Complete genome of the uncultured termite group 1 bacteria in a single host protist cell. *Proc Natl Acad Sci USA.* (2008) 105:5555–60. doi: 10.1073/pnas.0801389105
- Kvist T, Ahring BK, Lasken RS, Westermann P. Specific single-cell isolation and genomic amplification of uncultured microorganisms. *Appl Microbiol Biotechnol.* (2007) 74:926–35. doi: 10.1007/s00253-006-0725-7
- Guo MT, Rotem A, Heyman JA, Weitz DA. Droplet microfluidics for high-throughput biological assays. *Lab Chip* (2012) 12:2146–55. doi: 10.1039/c2lc21147e
- Autengruber A, Gereke M, Hansen G, Hennig C, Bruder D. Impact of enzymatic tissue disintegration on the level of surface molecule expression and immune cell function. *Eur J Microbiol Immunol.* (2012) 2:112–20. doi: 10.1556/EuJMI.2.2012.2.3
- Sanchez-Adams J, Athanasiou KA. Regional effects of enzymatic digestion on knee meniscus cell yield and phenotype for tissue engineering. *Tissue Eng Part C Methods* (2012) 18:235–43. doi: 10.1089/ten.tec.2011.0383
- Brady G, Iscove NN. Construction of cDNA libraries from single cells. *Methods Enzymol* (1993) 225:611–23. doi: 10.1016/0076-6879(93)25039-5
- Islam S, Kjällquist U, Moliner A, Zajac P, Fan JB, Lönnerberg P, et al. Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res.* (2011) 21:1160–7. doi: 10.1101/gr.110882.110
- Ramskold D, Luo S, Wang YC, Li R, Deng Q, Faridani OR, et al. Full-length mRNA-seq from single cell levels of RNA and individual circulating tumor cells. *Nat Biotechnol.* (2012) 30:777–82. doi: 10.1038/nbt.2282
- Picelli S, Björklund AK, Faridani OR, Sagasser S, Winberg G, Sandberg R. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat Methods* (2013) 10:1096–8. doi: 10.1038/nmeth.2639
- Pan X, Durrett RE, Zhu H, Tanaka Y, Li Y, Zi X, et al. Two methods for full-length RNA sequencing for low quantities of cells and single cells. *Proc Natl Acad Sci USA.* (2013) 110:594–9. doi: 10.1073/pnas.1217322109
- Liu N, Liu L, Pan X. Single-cell analysis of the transcriptome and its application in the characterization of stem cells and early embryos. *Cell Mol Life Sci.* (2014) 71:2707–15. doi: 10.1007/s00018-014-1601-8
- Islam S, Kjällquist U, Moliner A, Zajac P, Fan JB, Lönnerberg P, et al. Highly multiplexed and strand-specific single-cell RNA 5'-end sequencing. *Nat Protoc.* (2012) 7:813–28. doi: 10.1038/nprot.2012.022
- Zhu YY, Machleder EM, Chenchik A, Li R, Siebert PD. Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. *Biotechniques* (2001) 30:892–7. doi: 10.2144/01304pf02
- Picelli S, Faridani OR, Björklund AK, Winberg G, Sagasser S, Sandberg R. Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc.* (2014) 9:171–81. doi: 10.1038/nprot.2014.006
- Van Gelder RN, von Zastrow ME, Yool A, Dement WC, Barchas JD, Eberwine JH. Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc Natl Acad Sci USA.* (1990) 87:1663–7. doi: 10.1073/pnas.87.5.1663
- Hashimshony T, Wagner F, Sher N, Yanai I. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep.* (2012) 2:666–73. doi: 10.1016/j.celrep.2012.08.003
- Jaitin DA, Kenigsberg E, Keren-Shaul H, Elefant N, Paul F, Zaretzky I, et al. Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* (2014) 343:776–9. doi: 10.1126/science.1247651
- Lovatt D, Ruble BK, Lee J, Dueck H, Kim TK, Fisher S, et al. Transcriptome in vivo analysis (TIVA) of spatially defined single cells in live tissue. *Nat Methods* (2014) 11:190–6. doi: 10.1038/nmeth.2804
- Hashimshony T, Senderovich N, Avital G, Klochendler A, de Leeuw Y, Anavy L, et al. CEL-Seq2: sensitive highly-multiplexed single-cell RNA-seq. *Genome Biol.* (2016) 17:77–83. doi: 10.1186/s13059-016-0938-8
- Sasagawa Y, Nikaido I, Hayashi T. Quartz-Seq: a highly reproducible and sensitive single-cell RNA sequencing method, reveals non-genetic gene-expression heterogeneity. *Genome Biol.* (2013) 14:R31. doi: 10.1186/gb-2013-14-4-r31
- Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, et al. Droplet barcoding for single cell transcriptomics applied to embryonic stem cells. *Cell* (2015) 161:1187–201. doi: 10.1016/j.cell.2015.04.044
- Zilionis R, Nainys J, Veres A, Savova V, Zemmour D, Klein AM, et al. Single-cell barcoding and sequencing using droplet microfluidics. *Nat Protoc.* (2017) 12:44–73. doi: 10.1038/nprot.2016.154
- Hebenstreit D. Methods, challenges and potentials of single cell RNA-seq. *Biology* (2012) 1:658–67. doi: 10.3390/biology1030658
- Ziegenhain C, Vieth B, Parekh S, Reinius B, Guillaumet-Adkins A, Smets M, et al. Comparative analysis of single-cell RNA sequencing methods. *Mol Cell* (2017) 65:631–43. doi: 10.1016/j.molcel.2017.01.023
- Suslov O, Silver DJ, Siebzehnrubl FA, Orjalo A, Ptitsyn A, Steindler DA. Application of an RNA amplification method for reliable single-cell transcriptome analysis. *Biotechniques* (2015) 59:137–48. doi: 10.2144/000114331
- Blanco L, Salas M. Characterization and purification of a phage phi 29-encoded DNA polymerase required for the initiation of replication. *Proc Natl Acad Sci USA.* (1984) 81:5325–9. doi: 10.1073/pnas.81.1.75325
- Kang Y, Norris MH, Zarzycki-Siek J, Nierman WC, Donachie SP, Hoang TT. Transcript amplification from single bacterium for transcriptome analysis. *Genome Res.* (2011) 21:925–35. doi: 10.1101/gr.116103.110
- Zong C, Lu S, Chapman AR, Xie XS. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science* (2012) 338:1622–6. doi: 10.1126/science.1229164
- Chapman AR, He Z, Lu S, Yong J, Tan L, Tang F, et al. Single cell transcriptome amplification with MALBAC. *PLoS ONE* (2015) 10:e0120889. doi: 10.1371/journal.pone.0120889
- Sheng K, Cao W, Niu Y, Deng Q, Zong C. Effective detection of variation in single-cell transcriptomes using MATQ-seq. *Nat Methods* (2017) 14:267–70. doi: 10.1038/nmeth.4145
- Datlinger P, Rendeiro AF, Schmidl C, Krausgruber T, Traxler P, Klughammer J, et al. Pooled CRISPR screening with single-cell transcriptome readout. *Nat Methods* (2017) 14:297–301. doi: 10.1038/nmeth.4177

45. Chen J, Suo S, Tam PP, Han JJ, Peng G, Jing N. Spatial transcriptomic analysis of cryosectioned tissue samples with Geo-seq. *Nat Protoc.* (2017) 12:566–80. doi: 10.1038/nprot.2017.003
46. Trombetta JJ, Gennert D, Lu D, Satija R, Shalek AK, Regev A. Preparation of single-cell RNA-seq libraries for next generation sequencing. *Curr Protoc Mol Biol.* (2014) 107:4.22.1–17. doi: 10.1002/0471142727.mb0422s107
47. Junker JP, Noël ES, Guryev V, Peterson KA, Shah G, Huisken J, et al. Genome-wide RNA tomography in the zebrafish embryo. *Cell* (2014) 159:662–75. doi: 10.1016/j.cell.2014.09.038
48. Rosenberg AB, Roco CM, Muscat RA, Kuchina A, Sample P, Yao Z, et al. Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* (2018) 360:176–82. doi: 10.1126/science.aam8999
49. Wallrapp A, Riesenfeld SJ, Burkett PR, Abdunnour RE, Nyman J, Dionne D, et al. The neuropeptide NMU amplifies ILC2-driven allergic lung inflammation. *Nature* (2017) 549: 351–6. doi: 10.1038/nature24029
50. Cole CB, Russler-Germain DA, Ketkar S, Verdoni AM, Smith AM, Bangert CV, et al. Haploinsufficiency for DNA methyltransferase 3A predisposes hematopoietic cells to myeloid malignancies. *J Clin Invest.* (2017) 127:3657–74. doi: 10.1172/JCI93041
51. Farmer DT, Nathan S, Finley JK, Shengyang Yu K, Emmerson E, Byrnes LE, et al. Defining epithelial cell dynamics and lineage relationships in the developing lacrimal gland. *Development* (2017) 144:2517–28. doi: 10.1242/dev.150789
52. Goudot C, Coillard A, Villani AC, Gueguen P, Cros A, Sarkizova S, et al. Aryl hydrocarbon receptor controls monocyte differentiation into dendritic cells versus macrophages. *Immunity* (2017) 47:582–96. doi: 10.1016/j.immuni.2017.08.016
53. Pal B, Chen Y, Vaillant F, Jamieson P, Gordon L, Rios AC, et al. Construction of developmental lineage relationships in the mouse mammary gland by single-cell RNA profiling. *Nat Commun.* (2017) 8:1627. doi: 10.1038/s41467-017-01560-x
54. Quan N, Samuel L, Han C, Anne S, Timothy B, Angelika C, et al. *Single-Cell Transcriptome Sequencing of Human Induced Pluripotent Stem Cells Identifies Differentially Primed Subpopulations.* bioRxiv 119255 (2018).
55. Tsang JCH, Vong JSL, Ji L, Poon LCY, Jiang P, Lui KO, et al. Integrative single-cell and cell-free plasma RNA transcriptomics elucidates placental cellular dynamics. *Proc Natl Acad Sci USA.* (2017) 114:E7786–95. doi: 10.1073/pnas.1710470114
56. Yan KS, Gevaert O, Zheng GXY, Anchang B, Probert CS, Larkin KA, et al. Intestinal enteroendocrine lineage cells possess homeostatic and injury-inducible stem cell activity. *Cell Stem Cell* (2017) 21:78–90. doi: 10.1016/j.stem.2017.06.014
57. Yan KS, Janda CY, Chang J, Zheng GXY, Larkin KA, Luca VC, et al. Non-equivalence of Wnt and R-spondin ligands during Lgr5+ intestinal stem cell self-renewal. *Nature* (2017) 545:238–42. doi: 10.1038/nature22313
58. Zepp JA, Zacharias WJ, Frank DB, Cavanaugh CA, Zhou S, Morley MP, et al. Distinct mesenchymal lineages and niches promote epithelial self-renewal and myofibrogenesis in the lung. *Cell* (2017) 170:1134–48. doi: 10.1016/j.cell.2017.07.034
59. Zheng GX, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel digital transcriptional profiling of single cells. *Nat Commun.* (2017) 8:14049. doi: 10.1038/ncomms14049
60. Magella B, Adam M, Potter AS, Venkatasubramanian M, Chetal K, Hay SB, et al. Cross-platform single cell analysis of kidney development shows stromal cells express Gdnf. *Dev Biol.* (2018) 434:36–47. doi: 10.1016/j.ydbio.2017.11.006
61. Povinelli BJ, Rodriguez-Meira A, Mead AJ. Single cell analysis of normal and leukemic hematopoiesis. *Mol Aspects Med.* (2018) 59:85–94. doi: 10.1016/j.mam.2017.08.006
62. Tang Q, Iyer S, Lobbardi R, Moore JC, Chen H, Lareau C, et al. Dissecting hematopoietic and renal cell heterogeneity in adult zebrafish at single-cell resolution using RNA sequencing. *J Exp Med.* (2017) 214:2875–87. doi: 10.1084/jem.20170976
63. Paul F, Arkin Y, Giladi A, Jaitin DA, Kenigsberg E, Keren-Shaul H, et al. Transcriptional heterogeneity and lineage commitment in myeloid progenitors. *Cell* (2015) 163:1663–77. doi: 10.1016/j.cell.2015.11.013
64. Yang J, Tanaka Y, Seay M, Li Z, Jin J, Garmire LX, et al. Single cell transcriptomics reveals unanticipated features of early hematopoietic precursors. *Nucleic Acids Res.* (2017) 45:1281–96. doi: 10.1093/nar/gkw1214
65. Tsang JC, Yu Y, Burke S, Buettner F, Wang C, Kolodziejczyk AA, et al. Single-cell transcriptomic reconstruction reveals cell cycle and multi-lineage differentiation defects in Bcl11a-deficient hematopoietic stem cells. *Genome Biol.* (2015) 16:178. doi: 10.1186/s13059-015-0739-5
66. Kowalczyk MS, Tirosh I, Heckl D, Rao TN, Dixit A, Haas BJ, et al. Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. *Genome Res.* (2015) 25:1860–72. doi: 10.1101/gr.192237.115
67. Zhao X, Gao S, Wu Z, Kajigaya S, Feng X, Liu Q, et al. Single-cell RNA-seq reveals a distinct transcriptome signature of aneuploid hematopoietic cells. *Blood* (2017) 130:2762–73. doi: 10.1182/blood-2017-08-803353
68. Zhou F, Li X, Wang W, Zhu P, Zhou J, He W, et al. Tracing haematopoietic stem cell formation at single-cell resolution. *Nature* (2016) 533:487–92. doi: 10.1038/nature17997
69. Hu Y, Huang Y, Yi Y, Wang H, Liu B, Yu J, et al. Single-cell RNA sequencing highlights transcription activity of autophagy-related genes during hematopoietic stem cell formation in mouse embryos. *Autophagy* (2017) 13:770–1. doi: 10.1080/15548627.2016.1278093
70. Nestorowa S, Hamey FK, Pijuan Sala B, Diamanti E, Shepherd M, Laurenti E, et al. A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. *Blood* (2016) 128:e20–31. doi: 10.1182/blood-2016-05-716480
71. Alberti-Servera L, von Muenchow L, Tsapogas P, Capoferri G, Eschbach K, Beisel C, et al. Single-cell RNA sequencing reveals developmental heterogeneity among early lymphoid progenitors. *EMBO J.* (2017) 36:3619–33. doi: 10.15252/embj.201797105
72. Macaulay IC, Svensson V, Labalette C, Ferreira L, Hamey F, Voet T, et al. Single-cell RNA-sequencing reveals a continuous spectrum of differentiation in hematopoietic cells. *Cell Rep.* (2016) 14:966–77. doi: 10.1016/j.celrep.2015.12.082
73. Velten L, Haas SF, Raffel S, Blaszkiewicz S, Islam S, Hennig BP, et al. Human haematopoietic stem cell lineage commitment is a continuous process. *Nat Cell Biol.* (2017) 19:271–81. doi: 10.1038/ncb3493
74. Grover A, Sanjuan-Pla A, Thongjuea S, Carrelha J, Giustacchini A, Gambardella A, et al. Single-cell RNA sequencing reveals molecular and functional platelet bias of aged haematopoietic stem cells. *Nat Commun.* (2016) 7:11075. doi: 10.1038/ncomms11075
75. Cabezas-Wallscheid N, Buettner F, Sommerkamp P, Klimmeck D, Ladel L, Thalheimer FB, et al. Vitamin A-retinoic acid signaling regulates hematopoietic stem cell dormancy. *Cell* (2017) 169:807–23. doi: 10.1016/j.cell.2017.04.018
76. Drissen R, Buza-Vidas N, Woll P, Thongjuea S, Gambardella A, Giustacchini A, et al. Distinct myeloid progenitor differentiation pathways identified through single cell RNA sequencing. *Nat Immunol.* (2016) 17:666–76. doi: 10.1038/ni.3412
77. Chattopadhyay PK, Gierahn TM, Roederer M, Love JC. Single-cell technologies for monitoring immune systems. *Nat Immunol.* (2014) 15:128–35. doi: 10.1038/ni.2796
78. Zheng S, Papalexi E, Butler A, Stephenson W, Satija R. Molecular transitions in early progenitors during human cord blood hematopoiesis. *Mol Syst Biol.* (2018) 14:e8041. doi: 10.15252/msb.20178041
79. Yu Y, Tsang JC, Wang C, Clare S, Wang J, Chen X, et al. Single-cell RNA-seq identifies a PD-1hi ILC progenitor and defines its development pathway. *Nature* (2016) 539:102–6. doi: 10.1038/nature20105
80. Patil VS, Madrigal A, Schmiedel BJ, Clarke J, O'Rourke P, de Silva AD, et al. Precursors of human CD4+ cytotoxic T lymphocytes identified by single-cell transcriptome analysis. *Sci Immunol.* (2018) 3:eaan8664. doi: 10.1126/sciimmunol.aan8664
81. Shalek AK, Satija R, Adiconis X, Gertner RS, Gaubomme JT, Raychowdhury R, et al. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* (2013) 498:236–40. doi: 10.1038/nature12172
82. Shalek AK, Satija R, Shuga J, Trombetta JJ, Gennert D, Lu D, et al. Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature* (2014) 510:363–9. doi: 10.1038/nature13437

83. Villani AC, Satija R, Reynolds G, Sarkizova S, Shekhar K, Fletcher J, et al. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* (2017) 356:eaa4573. doi: 10.1126/science.aah4573
84. Papalexi E, Satija R. Single-cell RNA sequencing to explore immune cell heterogeneity. *Nat Rev Immunol.* (2018) 18:35–45. doi: 10.1038/nri.2017.76
85. Giustacchini A, Thongjuea S, Barkas N, Woll PS, Povinelli BJ, Booth CAG, et al. Single-cell transcriptomics uncovers distinct molecular signatures of stem cells in chronic myeloid leukemia. *Nat Med.* (2017) 23:692–702. doi: 10.1038/nm.4336
86. Wang L, Fan J, Francis JM, Georghiou G, Hergert S, Li S, et al. Integrated single-cell genetic and transcriptional analysis suggests novel drivers of chronic lymphocytic leukemia. *Genome Res.* (2017) 27:1300–11. doi: 10.1101/gr.217331.116
87. Lohr JG, Kim S, Gould J, Knoechel B, Drier Y, Cotton MJ, et al. Genetic interrogation of circulating multiple myeloma cells at single cell resolution. *Sci Transl Med.* (2016) 8:363ra147. doi: 10.1126/scitranslmed.aac7037
88. Chesi M, Bergsagel PL. Molecular pathogenesis of multiple myeloma: basic and clinical updates. *Int J Hematol.* (2013) 97:313–23. doi: 10.1007/s12185-013-1291-2
89. Ng SW, Mitchell A, Kennedy JA, Chen WC, McLeod J, Ibrahimova N, et al. A 17-gene stemness score for rapid determination of risk in acute leukaemia. *Nature* (2016) 540:433–7. doi: 10.1038/nature20598
90. Grun D, Muraro MJ, Boisset JC, Wiebrands K, Lyubimova A, Dharmadhikari G, et al. *De novo* prediction of stem cell identity using single-cell transcriptome data. *Cell Stem Cell* (2016) 19:266–77. doi: 10.1016/j.stem.2016.05.010
91. Liu S, Trapnell C. Single-cell transcriptome sequencing: recent advances and remaining challenges. *F1000Res* (2016) 5:182. doi: 10.12688/f1000research.7223.1
92. Islam S, Zeisel A, Joost S, La Manno G, Zajac P, Kasper M, et al. Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat Methods* (2014) 11:163–6. doi: 10.1038/nmeth.2772
93. Darmanis S, Sloan SA, Zhang Y, Enge M, Caneda C, Shuer LM, et al. A survey of human brain transcriptome diversity at the single cell level. *Proc Natl Acad Sci USA.* (2015) 112:7285–90. doi: 10.1073/pnas.1507125112
94. DeLaughter DM, Bick AG, Wakimoto H, McKean D, Gorham JM, Kathiriya IS, et al. Single-cell resolution of temporal gene expression during heart development. *Dev Cell* (2016) 39:480–90. doi: 10.1016/j.devcel.2016.10.001
95. Wang Y, Tang Z, Huang H, Li J, Wang Z, Yu Y, et al. Pulmonary alveolar type I cell population consists of two distinct subtypes that differ in cell fate. *Proc Natl Acad Sci USA.* (2018) 115:2407–12. doi: 10.1073/pnas.1719474115
96. Dulken BW, Leeman DS, Boutet SC, Hebestreit K, Brunet A. Single-cell transcriptomic analysis defines heterogeneity and transcriptional dynamics in the adult neural stem cell lineage. *Cell Rep.* (2017) 18:777–90. doi: 10.1016/j.celrep.2016.12.060
97. Mutoji K, Singh A, Nguyen T, Gildersleeve H, Kaucher AV, Oatley MJ, et al. TSPAN8 expression distinguishes spermatogonial stem cells in the prepubertal mouse testis. *Biol Reprod.* (2016) 95:117. doi: 10.1095/biolreprod.116.144220
98. Treutlein B, Brownfield DG, Wu AR, Neff NF, Mantalas GL, Espinoza FH, et al. Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* (2014) 509:371–5. doi: 10.1038/nature13173
99. Björklund AK, Forkel M, Picelli S, Konya V, Theorell J, Friberg D, et al. The heterogeneity of human CD127+ innate lymphoid cells revealed by single-cell RNA sequencing. *Nat Immunol.* (2016) 17:451–60. doi: 10.1038/ni.3368
100. Stanescu DE, Yu R, Won KJ, Stoffers DA. Single cell transcriptomic profiling of mouse pancreatic progenitors. *Physiol Genomics* (2017) 49:105–14. doi: 10.1152/physiolgenomics.00114.2016
101. Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H, et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* (2014) 344:1396–401. doi: 10.1126/science.1254257
102. Min JW, Kim WJ, Han JA, Jung YJ, Kim KT, Park WY, et al. Identification of distinct tumor subpopulations in lung adenocarcinoma via single-cell RNA-seq. *PLoS ONE* (2015) 10:e0135817. doi: 10.1371/journal.pone.0135817
103. Li H, Courtois ET, Sengupta D, Tan Y, Chen KH, Goh JLL, et al. Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. *Nat Genet.* (2017) 49:708–18. doi: 10.1038/ng.3818
104. Pina C, Teles J, Fugazza C, May G, Wang D, Guo Y, et al. Single-cell network analysis identifies DDIT3 as a nodal lineage regulator in hematopoiesis. *Cell Rep.* (2015) 11:1503–10. doi: 10.1016/j.celrep.2015.05.016
105. Tang F, Lao K, Surani MA. Development and applications of single-cell transcriptome analysis. *Nat Methods* (2011) 8(4 Suppl.):S6–11. doi: 10.1038/nmeth.1557
106. Athanasiadis EI, Botthof JG, Andres H, Ferreira L, Lio P, Cvejic A. Single-cell RNA-sequencing uncovers transcriptional states and fate decisions in haematopoiesis. *Nat Commun.* (2017) 8:2045. doi: 10.1038/s41467-017-02305-6
107. Moore FE, Garcia EG, Lobbardi R, Jain E, Tang Q, Moore JC, et al. Single-cell transcriptional analysis of normal, aberrant, and malignant hematopoiesis in zebrafish. *J Exp Med.* (2016) 213:979–92. doi: 10.1084/jem.20152013

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Huang, Li, Qin, Zhu, Xu and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.