Studying hematopoiesis using single-cell technologies

Fang Ye, Wentao Huang and Guoji Guo*

Abstract
Hematopoiesis is probably the best-understood stem cell differentiation system; hematopoietic stem cell (HSC) transplantation represents the most widely used regenerative therapy. The classical view of lineage hierarchy in hematopoiesis is built on cell type definition system by a group of cell surface markers. However, the traditional model is facing increasing challenges, as many classical cell types are proved to be heterogeneous. Recently, the developments of new technologies allow genome, transcriptome, proteome, and epigenome analysis at the single-cell level. For the first time, we can study hematopoietic system at single-cell resolution on a multi-omic scale. Here, we review recent technical advances in single-cell analysis technology, as well as their current applications. We will also discuss the impact of single-cell technologies on both basic research and clinical application in hematology.

Keywords: Single-cell analysis, Hematopoietic stem cell, Lineage hierarchy, Regulatory network

Background
The common landscape of cellular hierarchy has been depicted as a transitional process through multiple intermediate states. In mammalian systems, many stem and progenitor cell types have been identified with a combination of cell surface markers, analyzed through multicolored fluorescence-assisted cell sorting (FACS). In hematopoietic system, extensive efforts have gone into the characterization of cellular differentiation pathways and genetic regulatory network [1]. The identification of mouse hematopoietic stem cell and progenitor cells (HSPC) [2, 3], along with the separation of hematopoietic stem cells from multipotent progenitors (MPP) [4–7], has indicated that multipotent progenitors (MPP) are generated from self-renewing hematopoietic stem cells. Further identification of common lymphoid progenitors (CLP) and common myeloid progenitors (CMP) [8, 9] suggested that CLP and CMP stand at the bifurcation point of this lineage model. However, other recent findings challenged this view, since they found heterogeneity within these classical progenitors, which implied alternative lineage commitment pathways [10–12].

The signaling pathways and transcriptional networks that regulate hematopoietic stem cell emergence, self-renewal, and differentiation are not well understood [1, 13]. Previous works utilized FACS to separate and purify different progenitor types, and then performed transcriptomic analysis on each cell population. Those works elucidated a complex transcriptional network in the hematopoietic hierarchy [14, 15]. However, these analyses were conducted on bulk samples, which neglected the heterogeneity in the defined population as well as unknown transitional states during the cell fate decision process.

Not until recent years, single-cell analysis has become a powerful tool for studying cellular differentiation pathways. Advances in technology made it possible to track and capture single cells, then analyze its genome, transcriptome, and proteome. For instance, high-throughput single-cell quantitative PCR is highly sensitive in detecting quantitative differences [16–18]. Single-cell RNA-seq [19–22] allows for transcriptome analysis. High-throughput single-cell barcoding and sequencing approaches [23–26], coupled with data analysis methods [27, 28], enable the dissection of heterogeneity in complex cellular systems. In the field of hematopoietic study, single-cell transcriptomic and proteomic analysis provide unprecedented insights into cellular differentiation hierarchy, gene regulatory network, and developmental origin as well as mechanisms for stem cell aging. In this paper, we review the recent technical advances in single-cell methodology and summarize their...
features and contribution to hematopoietic research. In the end, we will discuss the current challenges and future directions in the field.

**Technical advances in single-cell analysis**

The cellular heterogeneity is concealed in the analysis of bulk cells. Cells gathered from the same part of the tissue differ from each other in gene expression and epigenetic status. The single-cell analysis provides a solution to understand the heterogeneity within cell population. Recently, technical developments have been made in the amplification of rare nucleic acid templates (Table 1).

**Single-cell genomic methods**

Several single-cell whole genome amplification (WGA) methods have been developed to amplify the rare genomic DNA. Degenerate-oligonucleotide PCR was used for analyzing copy number variation in cancer cells [29]. Another well-known WGA method was multiple displacement amplification (MDA), which utilized random primers and bacteriophage polymerase to achieve high-coverage single-cell exome sequencing [30, 31]. Multiple annealing and looping-based amplification cycles (MALBAC) have also been developed to reduce the bias in nonlinear genome amplification process. MALBAC achieves both high-coverage and uniform amplification. It can be applied to detect both copy number variations (CNVs) and single nucleotide polymorphisms (SNPs) in single-cell genome [32].

**Single-cell transcriptomic methods**

Single-cell transcriptome analysis remarkably serves as a powerful tool for studying cellular heterogeneity and lineage hierarchy (Fig. 1). There are several available methods: single-cell qPCR [16], single-cell microarray analysis [33], and single-cell RNA-seq [34, 35]. After single-cell isolation from complex tissue, the first challenge is to amplify the small amount of RNA, which is about 10 pg per cell. Four mainstream strategies are used: multiplexed RT-PCR, polyA tailing followed by second-strand synthesis [19], template switching, and in vitro transcription (IVT) [21]. Multiplexed RT-PCR is used in single-cell qPCR experiment [16]. Single-cell qPCR does not need to sequence the sample. It is convenient for detection of dozens of genes. PolyA tailing method was used in single-cell microarray and Tang-seq studies. Smart-seq and Smart-seq2 amplification is a widely used approach for the full-length mRNA analysis of single cells [22, 36, 37]. It uses the template-switching-based protocol to append a primer binding site on the 3’ end of the cDNA. cDNA is then amplified by PCR and sequenced by Illumina sequencing platform. The mRNA coverage of Smart-seq is between 10 and 20%. IVT used in CEL-seq and MARS-seq accomplishes a linear amplification of RNA using T7 promoter and RNA polymerase [21, 38]. The unique molecular identifiers (UMIs) are designed for reducing the amplification bias [39]. They enable the absolute counting of mRNA molecules in the single cell when mRNA capture efficiency and the sequencing depth are good enough. The low coverage of mRNA is a common problem for all existing methods.

Recently, application of single-cell transcriptomic analysis has rapidly spread to many areas such as early embryonic development [16, 40–44], cellular reprogramming [18, 45], human breast cancer [46], metastatic melanoma [47], circulating tumor cells [48], olfactory neurogenesis [49], early embryo development [50], neuronal cell heterogeneity, and immune cell pathogenicity [51–53]. These applications demonstrate the broad applicability of single-cell transcriptomic analysis.

**Single-cell proteomic methods**

Traditional single-cell protein analysis depends on fluorescence flow cytometry [54]. The development of mass flow cytometry notably increased multiplexity by isotope label on antibodies [55]. This method resolved the spectral

---

**Table 1 Classification of single-cell analysis methods**

| Method       | Amplification          | Coverage            | References |
|--------------|------------------------|---------------------|------------|
| Genomics     |                        |                     |            |
| MDA          | MDA                    | High coverage       | [30]       |
| MALBAC       | MALBAC                 | High coverage       | [31]       |
| Transcriptomics|                       |                     |            |
| Single-cell qPCR | Multiplexed PCR       | Target gene         | [16]       |
| Tang-seq     | PolyA tailing + second-strand synthesis | Full-length mRNAs | [34]       |
| CEL-seq      | In vitro transcription  | 3’ End of mRNA      | [21]       |
| Smart-seq    | Template switching     | Full-length mRNAs   | [36]       |
| Cyto-seq     | Multiplexed PCR        | 3’ End of mRNA      | [23]       |
| Drop-seq     | Template switching     | 3’ End of mRNA      | [24]       |
| inDrop       | In vitro transcription  | 3’ End of mRNA      | [25]       |
| Proteomics   |                        |                     |            |
| Mass cytometry| NA                    | Target protein      | [55]       |
| Epigenomics  |                        |                     |            |
| scATAC-seq   | Adaptor PCR            | Accessible DNA regions | [64]     |
| scRRBS       | Adaptor PCR            | 1.5 million CpG sites | [59]     |
| scHi-C       | Adaptor PCR            | NA                  | [63]       |
| scChIP-seq   | Adaptor PCR            | About 1000 peaks    | [62]       |

NA not applicable
overlap problem in fluorescence flow cytometry and can detect more than 30 parameters simultaneously. The idea has also been used in multiplexed ion beam imaging (MIBI) [56], which is capable of analyzing up to 100 targets at the same time in the tissue sections. Recent advances in microfluidic chips also enabled multiplexed analyses for quantitative single-cell proteomics [57, 58]. All existed methods only allow detection of limited kinds of protein. A whole proteome analysis approach remains to be developed.

**Single-cell epigenomic methods**

Single-cell epigenomic technologies are becoming more and more accessible. Single-cell reduced representation bisulfite sequencing (scRRBS) and single-cell 5hmC-seq were applied to investigate DNA methylation [59–61]. Single-cell chromatin immunoprecipitation sequencing (ChIP-seq) [62] and single-cell Hi-C [63] have been developed to profile chromatin structure in single cells. Single-cell chromatin accessibility methods, such as single-cell assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq), have been used to investigate cell-to-cell variation in mammalian regulatory elements [64]. Large-scale profiling of single-cell chromatin accessibility landscape can be achieved by combining cellular indexing and ATAC-seq [65].

**Single-cell capture methods for sequencing**

Single-cell capturing is a challenge. However, we have seen the significant progress of platform developments in recent years (Table 2). When samples are rare, mouth pipetting and laser capture microdissection (LCM) [66] are good choices to isolate single cells. But when dealing with large number of cells, throughput becomes the bottleneck. FACS played an importing role in scaling up single-cell collection efficiency, but library generation remains to be labor intensive and costly. Very recently, lots of other convenient methods have been invented (Fig. 2). Some of them are directly linked with combinatorial indexing and sequencing library generations, which greatly facilitated high-throughput single-cell analysis without requirements for expensive instruments.

**Fluidigm C1 system**

Fluidigm C1 system, a commercialized single-cell library preparation platform, uses microfluidic circuits for single-cell capture and mRNA amplification reaction [67]. The whole work flow is highly automatic, and its data shows good stability. It becomes a widely recognized platform for single-cell analysis [68–71]. The high cost for the device and consumables is its main limitation.
DEPArray

DEPArray is another commercialized imaging-based cell sorting platform, which is suitable for selecting rare cells from sample. First, cells are loaded to a microfluidic cartridge, which contains an array of individually controllable electrodes. Then, every individual cell is trapped in a dielectrophoresis (DEP) cage. After imaging, cells of interest can be moved into parking chamber and recovery chamber. This technique allows isolation of tumor cells from the tissue and blood [72] and collection of circulating tumor cells (CTCs) [73, 74]. The ability to image and manipulate individual cell is the main feature of DEPArray. Rare cells can be identified, recovered, and sequenced subsequently.

Cyto-seq

The Cyto-seq method uses a microwell array to capture a large number of single cells [23]. Individual cell and indexed bead are trapped in the same well. After cell lysis, mRNAs are captured by indexed oligonucleotides on beads. The beads are then pooled, followed by a reverse transcription. The indexed cDNAs are then amplified with primers of target genes using multiplexed PCR. Expression of targeted genes is quantified by unique molecular identifiers (UMIs). The microwell array used by Cyto-seq is very simple to fabricate. The system is relatively convenient for small labs to set up.

Drop-seq and inDrop

One notable high-throughput platform developed very recently is the droplet-based microfluidics [26], which can efficiently sort individual cells into thousands of nanoliter droplet. The nanoliter droplet serves as a tiny reaction chamber, which significantly reduces the usage of reagent and increases the concentration of target molecules. In 2015, two groups simultaneously applied the droplet-based microfluidics single-cell analysis platforms to large-scale transcriptome study, which are termed Drop-seq and inDrop, respectively. Drop-seq puts individual cell and an indexed bead into a nanoliter
RNA-seq, spike-in RNA is recommended as an RNA-seq data with or without UMIs [80]. For single-cell review gave a standard pipeline of handling single-cell normalization, and transcripts quantification. A recent computational tools are developed for preprocessing, In single-cell transcriptomic sequencing analysis, lots of Single-cell sequencing generates huge amount of data. Data preprocessing workflow of single-cell data analysis. have been developed. Here, we briefly introduce the ger suitable. Several single-cell computational methods the data. Traditional computational methods are no lon-
dramatically increases the dimension and complexity of massive data set. The large amount of cell and gene
were together amplified and sequenced [24]. Meanwhile, inDrop demonstrates a very similar strategy, but the reverse transcription reaction is done inside the droplets. It utilizes IVT rather than template switching method for the transcriptomic amplification [25]. Both approaches adopted UMIs to quantify individual mRNAs. However, the dramatic increase of throughput comes with a cost. The detected gene number for single cells drops down due to the limitation of sequencing depth.

Single-cell imaging
Single-cell imaging enables phenotypic characterization of single cells, while preserving their spatial information. DEPArray has image-based single-cell sorting function. Besides, time-lapse microscopy allowed continuous live image on single cells. This method provides important information of the dynamic cell fate decision process during blood cell differentiation [75, 76]. Hardware and software requirements for setting up the single-cell long-term imaging system have also been thoroughly discussed [77].

Single-cell transplantation
Single-cell transplantation is a powerful approach to verify the stem cell identity. Single-cell transplantation of hematopoietic stem cell (HSC) purified by different surface marker proved their multilineage reconstitution function [5, 78]. Recently, Notta et al. revealed CD49f as a human HSC marker. Flow-sorted single cell based on CD49f and mitochondrial dye rhodamine-123 displayed robust chimerism even 20 weeks after transplantation [79].

Single-cell data analysis
The next-generation sequencing platforms generate massive data set. The large amount of cell and gene dramatically increases the dimension and complexity of the data. Traditional computational methods are no longer suitable. Several single-cell computational methods have been developed. Here, we briefly introduce the workflow of single-cell data analysis.

Data preprocessing
Single-cell sequencing generates huge amount of data. In single-cell transcriptomic sequencing analysis, lots of computational tools are developed for preprocessing, normalization, and transcripts quantification. A recent review gave a standard pipeline of handling single-cell RNA-seq data with or without UMIs [80]. For single-cell RNA-seq, spike-in RNA is recommended as an artificially designed internal controller in experiments to estimate technical variation [81, 82]. After filtering out data with low quality, sequencing reads would be aligned to a reference genome or transcriptome. After counting and normalization, the single-cell transcriptomic data will be converted into a digital gene expression (DGE) matrix for further analysis. For single-cell genomic data, people developed normalization algorithms based on channel, genome composition, and recurrent genome artifact corrections to improve the CNV detection in single-cell array CGH data [83]. An optimized protocol was used to correct biases inherent in the WGA procedure for the genome-wide copy number analysis [84]. For single-cell qPCR data processing, the use of a single reference gene is not recommended. In order to address the inherent noise in single-cell gene expression data, normalization by the median Ct value was applied [85]. For mass cytometry, bead based signature and an algorithm was used to determine data quality [86].

Data visualization and clustering
The downstream analysis focuses on visualizing the high-dimensional single-cell gene expression data and clustering the transcriptionally distinct subgroups. One visualization tool for high-dimensional data is principal component analysis (PCA), which has been widely applied in single-cell research [16, 25, 87]. PCA maps the high-dimensional data points into a low-dimensional space. Another visualization tool for reducing high-dimensional data into two or three dimensions is t-distributed stochastic neighbor embedding (t-SNE) [88, 89]. In SNE, nearby data points in high-dimensional space remain their similarity in low-dimensional space. However, SNE is hampered by the crowding problem, which means that the clusters could not be totally separated from each other. To alleviate the crowding problem, the Student’s t distribution method is introduced in t-SNE to compute the similarity between two points. As a powerful visualization tool, it has demonstrated great capacity in recent high-throughput single-cell studies [24, 47]. PCA and t-SNE are usually combined for the visualization of large-scale data. After identification of cell subpopulation, one can extract specific gene markers for each subpopulation. To improve differential gene identification from noisy single-cell data, Kharchenko et al. reported a probabilistic model of expression-magnitude distortions typical of single-cell RNA-sequencing measurements [90]. Other downstream clustering pipelines for high-throughput single-cell gene expression are mostly based on R package or MATLAB. A computational strategy named Seurat integrated these visualization methods and tools into an R package to deal with single-cell RNA-seq data [24, 91].
Pathway and network modeling
During lineage commitment process, the continuity of single-cell gene expression can be used to infer differentiation pathway. Spanning-tree progression analysis of density-normalized events (SPADE) analysis uses this idea to infer cellular hierarchy from large-scale single-cell data set without assigning temporal order [17, 55, 92, 93]. For gene regulatory network, some research groups completed network modeling with STRING and functional NET database [94, 95]. Weighted gene co-expression network analysis (WGCNA) is a R package which is available for conducting weighted gene co-expression network analysis [96]. Gene expression networks could also be integrated with epigenomic data like ChIP-seq binding data sets [17]. Computational tools have strengthened our ability to extract valuable information from large-scale data, thereby playing an indispensable role in the single-cell analysis.

Studying hematopoiesis at single-cell level
Classical knowledge about hematopoiesis is built on cell type definition system using flow cytometry analysis. Such a system is limited with a small number of cell surface markers for cell classification. Recently, taking the advantage of the aforementioned technologies and the related data analysis pipelines, we start to realize that many of these classical cell types are heterogeneous. Single-cell analysis helped to provide unprecedented insights into long-lasting questions in topics of hematopoietic study, such as HSC heterogeneity, differentiation pathway, fate decision, regulatory network, HSC aging, and HSC origin.

HSC heterogeneity
Blood cell production depends on HSC’s self-renewal and multilineage differentiation abilities. However, hematopoietic stem cells are heterogeneous in differentiation behavior. Single-cell transplantation is the most definitive assessment of HSC functional heterogeneity. Classical single-cell transplantation experiment showed that HSC defined by mouse homolog of CD34 reconstituted the lymphohematopoietic system for more than 3 months in mice. Highly purified mouse HSCs based on the expression of CD34 demonstrate variability in self-renewal potential and multilineage differentiation potential [5]. A recent research utilized single-cell transplantation assay to analyze phenotypic long-term HSC systematically. Donor-derived contribution to the circulating white blood cells showed at least four distinct patterns. They provide solid evidence that primitive hematopoietic cells can maintain distinct repopulation properties upon serial transplantation in vivo [97]. Using similar approaches, Morita et al. found that in the HSC subset, single cells behave differently based on their CD150 expression. Decreased expression of CD150 appears to be associated with reduced erythroblast/megakaryocyte differentiation potential. The balanced long-term repopulating cells are enriched in the CD150 intermediate subpopulation [78]. To gain deeper insight into the regulatory program of mouse HSCs, Wilson et al. linked single-cell functional assays with flow cytometric index sorting and single-cell gene expression assays. They identify key molecules that associate with long-term durable self-renewal and provide a single-cell molecular dataset that can be further analyzed regarding HSC heterogeneity [98].

Differentiation pathway
In the classical model of hematopoiesis, an organized hematopoietic lineage tree starts with multipotent HSC, and then followed by oligopotent and unipotent progenitors. However, recent single-cell results challenged the classic model and proposed that traditional hematopoietic progenitor types are very heterogeneous [99, 100]. Guo et al. used 280 multiplexed qPCR assays to analyze over 1500 single mouse hematopoietic cells [17]. The analysis revealed dramatic heterogeneity within all of the classically defined progenitor types, such as HSC, MPP, CMP, and CLP. The comprehensive data revealed a revised hierarchy of hematopoietic stem cell differentiation in which megakaryocytic and erythroid (MegE) lineage was the first to branch from hematopoietic stem cells (Fig. 3b). More recently, Notta et al. combined single-cell gene expression analysis and single-cell functional assay to study human hematopoiesis, and their results challenge the classical human hematopoietic hierarchy model [100] (Fig. 3d). They found that the cell hierarchy differed from fetal stage to adult stage. In fetus, multipotent, oligopotent, and unipotent progenitors are all can be seen, while only multipotent and unipotent progenitor stages were observed in adult. They show that megakaryocytic lineage can derive from HSC and multipotent progenitors in fetus but only branch from HSC in adult. Franziska et al. combined index sorting with MARS-seq to analyze mouse bone marrow CMP. The remarkable data set revealed seven transcriptionally distinct subpopulations within CMP cells. These subpopulations showed unexpected priming towards seven differentiation fates but no progenitors with a mixed state. The findings challenged traditional common myeloid progenitors (CMP) defined by cell surface markers and built a single-cell reference for studying mouse myeloid differentiation [99] (Fig. 3c). Kristiansen et al. analyzed the differentiation process from fetal liver HSCs to B-1a/B-2 B cells and provided novel insights into the B cell lineage development [101]. Macaulay et al. applied single-cell RNA-seq to study thrombocyte lineage commitment in zebra fish. They placed all data points into a continuum to form a refined lineage pathway [102].
Another work combined single-cell transcriptional profile and immune-phenotype to clarify differentiation pathway in megakaryocyte-erythroid progenitors (MEP) [103].

Single-cell mass cytometry is also a widely used method in hematopoietic differentiation pathway study. A recent primer article detailed the development of mass cytometry and related data analysis methods [104]. Bendall et al. measured 34 parameters for human bone marrow samples. Using the data, they constructed a minimal spanning tree-based hematopoietic differentiation pathway. The research also revealed system-wide signaling responses among traditional cell subsets [55]. Single-cell mass cytometry was also used to map cell cycle phases for human hematopoietic cells [105]. The method allowed deep profiling of major phases of the cell cycle simultaneously in single cells. Single-cell mass cytometry data can be aligned onto a unified trajectory. This approach accurately predicts the stepwise human B cell developmental path de novo. The trajectory revealed virtually all the cellular states of early B cells differentiated from hematopoietic stem cell [106].

Altogether, these findings challenged the current step-by-step bifurcation hierarchy in hematopoietic system. Transcriptional heterogeneity among progenitors suggested that cell differentiation may proceed in a more sophisticated way. Single-cell analysis helped to rebuild hematopoietic lineage hierarchy and identify markers of previously unknown subpopulations.

**Fate decision**

How do blood cells make their lineage decisions? Single-cell analysis helped to resolve this question with higher precision. Pina et al. used single-cell qPCR to define primary multipotent self-renewing cells and early erythroid-committed cells [107]. Their data suggested an uncoordinated molecular transition between self-renewal and committed states. They also found dissociation between self-renewal potential and transcriptome-wide activation in lineage program. Multipotent cells are unlikely to change into the committed state under independent activation by individual regulators. A very recent work aimed to capture mixed-lineage states in mouse hematopoietic stem and progenitor cells. With single-cell transcriptomic data, they provided evidence that mixed-lineage intermediates manifested concurrent expression of hematopoietic stem cell/progenitor and myeloid progenitor cell genes [108].

Single-cell imaging and tracking methods are also powerful tools for examining blood generation and lineage determination. Eilken et al. used time-lapse microscopy to achieve continuous long-term single-cell observation and detected hemogenic endothelial cells giving rise to blood cells [75]. Hoppe et al. applied time-lapse imaging and single-cell tracking to explore co-regulation of transcription factors GATA1 and PU.1 in differentiation dynamics of single HSCs [76]. They found that the ratio of PU.1 and GATA1 is not a key regulator for HSPC lineage decision.
which challenged the old view on the early hematopoietic lineage separation.

**Regulatory network**
The explosion of genomic data enabled generation of regulatory networks in various biological systems. However, network build with data from population of cells are intrinsically flawed, because the fundamental unit of gene regulation is single cell, rather than heterogeneous population. Regulatory model at single-cell level starts to emerge in recent years. In 2013, Moignard et al. analyzed 18 key hematopoietic transcription factors in hundreds of blood stem cells and progenitor cells [109]. They revealed factor interaction between Gata2, Gfi1, and Gfi1b. In their model, Gfi1 represses Gata2, whereas Gata2 activates Gfi1b. Gata2 functions in a regulatory loop to modulate Gfi1/Gfi1b cross-antagonism during HSC entry into the lympho-myeloid lineages. In the same year, Guo et al. build a remarkably similar single-cell model to explain early megakaryocyte and erythroid lineage regulation, in which Gata2 primes MegE fate and represses lympho-myeloid fate [17]. In 2015, Moignard et al. introduced another strategy using diffusion maps to analyze single-cell qPCR data. Based on state transition graphs from 3934 cells in the mouse embryo, they generated a transcriptional regulatory network model to explain the whole blood development process [28]. Thanks to the emergence of high-throughput single-cell analysis method, we are able to elucidate gene regulation mechanisms for the first time at the single-cell level.

**HSC aging**
It has been a debate whether intrinsic cell changes or variations in composition contributes to the systematic HSC aging. Single-cell analysis helped to provide a deeper look into this question. Grover et al. analyzed old and young HSCs transcriptomes at the single-cell level. They identified significantly increased molecular platelet priming and functional platelet bias in the aged HSCs [110]. They observed that loss of the FOG-1 transcription factor associated with HSC platelet programming increased lymphoid output. Thus, increased platelet bias is a key process during HSCs ageing. Kowalczyk et al. compared cells from young and old mice by single-cell RNA-seq. They found lower frequency of cells in G1 phase among old long-term HSCs. Moreover, old short-term (ST) HSCs resemble young long-term (LT) HSCs, suggesting that they exist in a less differentiated state [111]. Single-cell analyses of HSC aging process demonstrate that both compositional changes and intrinsic, population-wide changes contribute to phenotypic aging.

**HSC origin**
The origin of HSCs in mammalian systems has long been a mystery. The classical study identified the primitive type of hematopoietic stem cell in yolk sac (YS-HSC). But definitive hematopoiesis is maintained by HSC originated within the aorta-gonad-mesonephros (AGM) region of the embryo [112, 113]. Definite hematopoiesis produces HSCs with multilineage potential and long-term reconstitution ability. A recent work adopted single-cell RNA-seq to analyze endothelial cells and pre-HSCs in the mouse AGM region. Zhou et al. demonstrated that pre-HSCs have unique features in transcription factor network, signaling pathway, and unique cell cycle status at the single-cell level. They identified new surface markers for pre-HSC isolation and revealed the importance of mTOR in regulating the stepwise generation of HSCs in vivo (Fig. 3e). Transplantation of isolated pre-HSCs demonstrates strong self-renewal capability. New marker of pre-HSCs would provide new insights into HSC development and push forward future clinical applications of HSCs [114, 115]. Thus, single-cell analysis show great convenience and precision in tracing the origins of HSCs. Future efforts may focus on the transition of HSC heterogeneity during developmental stage. Besides, the origin of human hematopoietic stem cells also deserves further investigation.

**Conclusions**
Single-cell analyses have achieved remarkable advancement in recent years. Hematological researches also benefited from the rapid progress of single-cell technology. Future application of these approaches will impact the field in many aspects.

One important direction for single-cell analysis is to develop cost-effective methods. The present commercialized single-cell analysis equipment is limited by both cost and throughput. More economical and convenient methods are required for broad application of high-throughput single-cell analysis. As technology goes, future throughput is expected to achieve tens of thousands of single cells per experiment, while the cost is going to drop below 0.1 USD per cell.

Another orientation is the large-scale single-cell database. The explosively increasing single-cell data has raised a big challenge for scientists. Data comparison between different studies remains to be difficult. Data management becomes a burden for ordinary labs. Better computational analysis methods are needed. A large-scale database and online data analysis pipeline would be extremely helpful for the integration of various data sets collected from different tissues. An open-access platform may eventually lead to the completion of human single-cell atlas database, which might significantly impact basic research and clinical diagnosis.
In the future, multi-omics will be a trend for single-cell analysis, such as profiling DNA methylation and transcriptome from the same cell simultaneously [116]. Single-cell triple omics sequencing method, including genome, DNA methylation, and transcriptome, is another innovation [117]. A complete circuit of a single cell can be detected by the integration of genomics, epigenomics, transcriptomics, and proteomics. Multi-omics data will provide a more comprehensive understanding to the hematopoiesis and other basic questions of life.

In situ single-cell sequencing methods will preserve spatial information. It should be extremely helpful for studying stem cell microenvironment. Cell niches are playing essential roles for the self-renewal and differentiation of hematopoietic stem cell in vivo. Synthetic microenvironment has been used to generate functional hematopoietic stem cells [118–120]. In situ single-cell genomics would be ideal to analyze hematopoiesis in their local niches. Such information should be important for understanding the behavior of HSC and leukemia. It will provide guidance for promoting HSC expansion and to inhibit leukemia growth.

The progress of basic methodologies will accelerate the diagnosis and treatment for hematological diseases. Single-cell analysis of wild-type blood cells may offer guidance for HSC generation in vitro and improve transplantation-based therapies. Single-cell analysis of leukemia cells will help to find heterogeneity and clonal composition in cancers and guide leukemia treatments [121]. Recently, the grand plan for precision medicine has been put on agenda. Nothing is more precise than single cells. Single-cell analysis will allow for highly precise treatment for different individuals, as well as different cells in one patient. This powerful strategy will have a long-term impact on both basic research and clinical application in the field of hematology.

**Abbreviations**

AGM: Aorta-gonad-mesonephros; ATAC-seq: Assay for transposable-accessible chromatin with high-throughput sequencing; Baso: Basophil; CLP: Common lymphoid progenitor; CMF: Common myeloid progenitor; CNVs: Copy number variations; CTGs: Circulating tumor cells; CyTOF: Cytometry by Time-Of-Flight; DC: Dendritic cells; DEP: Dielectrophoresis; DGE: Digital gene expression; DGE: Digital gene expression matrix; E: Embryonic days; Esos: Eosinophil; Er: Erythrocytes; FACs: Fluorescence-assisted cell sorting; GMP: Granulocyte-monocyte progenitor; Gran: Granulocytes; Hi-C: High-throughput/resolution chromosome conformation capture; HSC: Hematopoietic stem cell; HSFC: Hematopoietic stem and progenitor cells; IVT: In vitro transcription; LCM: Laser capture microdissection; Ly: Lymphoid; MALBAC: Multiple annealing and looping-based amplification cycles; MATLAB: Matrix laboratory; MDA: Multiple displacement amplification; MegE: Megakaryocytic and erythroid; MEP: Megakaryocyte-erythroid progenitors; Mk: Megakaryocytes; Monoc: Monocytes; MPP: Multipotent progenitors; mRNA: Messenger RNA; miTOR: Mechanistic targets of rapamycin; My: Myeloid; Neu: Neutrophil; PCA: Principal component analysis; scRRBS: Single-cell reduced representation bisulfite sequencing; SNPs: Single nucleotide polymorphisms; SPADE: Spanning-tree progression analysis of density-normalized events; t-SNE: t-distributed stochastic neighbor embedding; UMI: Unique molecular identifiers; WGA: Whole genome amplification; WGCNA: Weighted gene co-expression network analysis; YS: Yolk sac.

**Acknowledgements**

Not applicable.

**Funding**

This work was supported by funding from 1000 Youth Talent Plan (G.G), the National Project 985 (G.G), and the Fundamental Research Funds for the Central Universities (G.G).

**Availability of data and materials**

The material supporting the conclusion of this review has been included within the article.

**Authors’ contributions**

WH and FY wrote the manuscripts. FY mainly contributed to the single cell technology part, and WH mainly contribute to the HSC research part. GG gave suggestions and gave revisions. All authors contributed to the writing of the review. All authors read and approved the final manuscript.

**Competing interests**

The authors declare that they have no competing interests.

**Consent for publication**

This is not applicable for this review.

**Ethics approval and consent to participate**

This is not applicable for this review.

**Received**: 25 November 2016 **Accepted**: 13 January 2017

**Published online**: 21 January 2017

**References**

1. Orkin SH, Zon LL. Hematopoiesis: an evolving paradigm for stem cell biology. Cell. 2008;132:631–44.
2. Muller-Seiberg CE, Whitlock CA, Weissman IL. Isolation of two early B lymphocyte progenitors from mouse marrow: a committed pre-pre-B cell and a clonogenic Thy-1−b hematopoietic stem cell. Cell. 1986;44:653–62.
3. Visser JW, Bauman JG, Mulder AH, Elisson JF, de Leeuw AM. Isolation of murine pluripotent hematopoietic stem cells. J Exp Med. 1984;159:1576–90.
4. Morrison SJ, Weissman IL. The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatble by phenotype. Immunity. 1994;1:661–73.
5. Osawa M, Hanada K, Hamada H, Nakashita H. Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell. Science. 1996;273:242–5.
6. Morrison SJ, Wandycz AM, Hemmati HD, Wright DE, Weissman IL. Identification of a lineage of multipotent hematopoietic progenitors. Development. 1997;124:1929–39.
7. Kiel MJ, Yilmaz OH, Ishizaka T, Terhorst C, Morrison SJ. SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. Cell. 2005;121:1109–21.
8. Kondo M, Weissman IL, Akashi K. Identification of clonogenic lymphoid progenitors in mouse bone marrow. Cell. 1997;91:661–72.
9. Akashi K, Traver D, Miyamoto T, Weissman IL. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. Nature. 2000;404:193–7.
10. Adolfsson J, Månsson R, Buza-Vidas N, Hultquist A, Liuba K, Jensen CT, Bryder D, Yang L, Borge OJ, Thoren LA, et al. Identification of Flt3− lymphomyeloid stem cells lacking erythroid-megakaryocytic potential a revised road map for adult blood lineage commitment. Cell. 2005;121:295–306.
11. Arenou Y, Mizuno S, Chong Y, Shiogamatsu H, Iino T, Iwasaki H, Graf T, Mayfield R, Chan S, Kastner P, Akashi K. Reciprocal activation of GATA-1 and PU.1 marks initial specification of hematopoietic stem cells into myeloid/Megakaryocyte and myelolymphoid lineages. Cell Stem Cell. 2007;1:416–27.
12. Pronk CJ, Rossi DJ, Månsson R, Attema JL, Nordinh GL, Chan CK, Sigvardsson M, Weissman IL, Bryder D. Elucidation of the phenotypic, functional, and molecular topography of a myeloid/lymphoid progenitor cell hierarchy, Cell Stem Cell. 2007;1:428–42.
13. Doulavos S, Notta F, Laurenti E, Dick JE. Hematopoiesis: a human perspective. Cell Stem Cell. 2012;10:20–36.
14. Novershtern N, Subramanian A, Lawton LN, Mak RH, Haining WN, McConkey ME, Habib N, Yosef N, Chang CY, Shay T, et al. Densely interconnected...
transcriptional control cell state changes in mouse hematopoiesis. Cell. 2011;144:296–309.
15. Jovicic V, Shay T, Sylvia K, Zuk O, Sun X, Kang J, Regev A, Koller D, Best AJ, Knell J, et al. Identification of transcriptional regulators in the mouse immune system. Nat Immunol. 2013;14:633–43.
16. Guo G, Huss M, Tong QG, Wang C, Li Sun L, Clarke ND, Robson P. Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. Dev Cell. 2010;18:675–85.
17. Guo G, Luc S, Marco E, Lin TW, Peng C, Kerényi MA, Beye S, Kim W, Xu J, et al. Mapping cellular hierarchy by single-cell analysis of the cell surface repertoire. Cell Stem Cell. 2013;13:492–505.
18. Buganim Y, Faddah DA, Cheng AW, Istitovich E, Markoulaki S, Ganz K, Klemm SL, van Oudenaarden A, Jaenisch R. Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchical phase. Cell. 2012;150:1209–22.
19. Tang F, Barbacioru C, Yuan W, Nordman E, Lee C, Xu N, Wang X, Bodeau J, Tuch BB, Siddiqui A, et al. miRNA-seq whole-transcriptome analysis of a single cell. Nat Methods. 2009;6:377–82.
20. Tang F, Lao K, Surani MA. Development and applications of single-cell transcriptome analysis. Nat Methods. 2011;8:52–11.
21. Hashimshony T, Wagner F, Steiner Y, Nanou I. CEL-seq: single-cell RNA-seq by multiplexed linear amplification. Cell Rep. 2012;2:666–73.
22. Picelli S, Björklund A, 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.
23. Fan HC, Fu FH, Fedor SP. Expression profiling. Combinatorial labeling of single cells for gene expression cytometry. Science. 2015;347:1258367.
24. Macosko EZ, Basu A, Satij A, Nemesh J, Shekhar K, Goldman M, Tirich S, Bialas AR, Kamitaki N, Marsterreck EM, et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell. 2015;161:1202–14.
25. Klein AM, Mazutis L, Akaruna I, Tallapragada N, Verheyen L, Visel D, Kirschner MW. Droplet barcoding for applied gene expression profiling to embryonic stem cells applied to embryonic stem cells. Cell. 2015;161:1187–201.
26. Mazutis L, Gilbert J, Ung W, Weitz DA, Griffiths AD, Heyman JA. Single-cell analysis and sorting using droplet-based microfluidics. Nat Protoc. 2013;8:870–91.
27. Amir e-A, Davis KL, Tadmor MD, Simonds EF, Levine JH, Bendall SC, Shenfeld DK, Krishnaswamy S, Natarajan GP, Perez D. viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. Nat Biotechnol. 2013;31:545–52.
28. Moignard V, Woodhouse S, Haghverdi L, Alix J, Tanaka Y, Wilkinson AC, Buettner F, Macaulay IC, Jayward W, Diamanti E, et al. Decoding the regulatory network of early blood development from single-cell gene expression measurements. Nat Biotechnol. 2015;33:269–76.
29. Navin N, Kendall J, Troge J, Andrews P, Rodgers L, Mc淖o D, Cook K, Stepankova O, KISIBLE D, Esposto D, et al. Tumour evolution inferred by single-cell sequencing. Nature. 2011;472:294–9.
30. Hou Y, Song L, Zhu P, Zhang B, Tao Y, Xu L, Li F, Wu K, Liang J, Shao D, et al. Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm. Cell. 2012;148:873–85.
31. Xu H, Hou Y, Yin X, Bao L, Tang A, Song L, Li F, Tang S, Wu K, Wu H, et al. Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. Cell. 2012;148:886–95.
32. Zong C, Lu S, Chapman AR, Xie XS. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. Science. 2013;338:1522–6.
33. Kamme F, Salunga R, Yu J, Tran DT, Zhou J, Luo L, Bittner A, Guo HQ, Miller JN, Ran J, et al. Single-cell microarray analysis in hippocampus CA1: demonstration and validation of cellular heterogeneity. J Neurosci. 2003;23:3607–15.
34. Tang F, Barbacioru C, Nordman E, Li B, Xu N, Bashkirov V, Lao K, Surani MA. RNA-seq analysis to capture the transcriptome landscape of a single cell. Nat Protoc. 2010;5:516–35.
35. Tang F, Barbacioru C, Bao S, Lee C, Nordman E, Wang X, Lao K, Surani MA. Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-seq analysis. Cell Stem Cell. 2010;6:466–78.
36. 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.
37. Islam S, Kjällquist U, Mollner A, Zajac P, Fan JB, Lönnerberg P, Linnarsson S. Highly multiplexed and strand-specific single-cell RNA’ end sequencing. Nat Protoc. 2012;7:813–28.
107. Pina C, Fugazza C, Tipping AJ, Brown J, Soneji S, Teles J, Peterson C, Enver T. Inferring rules of lineage commitment in haematopoiesis. Nat Cell Biol. 2012;14:287–94.

108. Olsson A, Venkatasubramanian M, Chaudhri VK, Aronow BJ, Salomonis N, Singh H, Grimes HL. Single-cell analysis of mixed-lineage states leading to a binary cell fate choice. Nature. 2016;537:698–702.

109. Moligned V, Macaulay K, Swiers G, Buetnot F, Shlütte J, Calero-Nieto FJ, Kinston S, Joshi A, Hannah R, Theis FJ, et al. Characterization of transcriptional networks in blood stem and progenitor cells using high-throughput single-cell gene expression analysis. Nat Cell Biol. 2013;15:363–72.

110. Grover A, Sanjuan-Pla A, Thongjuea S, Carrelia J, Giustacchini A, Gambardella A, Macaulay I, Mancini E, Lui TC, Mead A, et al. Single-cell RNA sequencing reveals molecular and functional platelet bias of aged haematopoietic stem cells. Nat Commun. 2016;7:11075.

111. Kowalczyk MS, Tirol I, Heckl D, Rao TN, Dixit A, Haas BJ, Schneider RK, Wagers AJ, Ebert BL, Regev A. Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. Genome Res. 2015;25:1860–72.

112. Medvinsky A, Dzierzak E. Definitive hematopoiesis is autonomously initiated by the AGM region. Cell. 1996;86:897–906.

113. Zovein AC, Hofmann JJ, Lynch M, French WJ, Turlo KA, Yang Y, Becker MS, Zanetta L, Dejana E, Gasson JC, et al. Fate tracing reveals the endothelial origin of hematopoietic stem cells. Cell Stem Cell. 2008;3:362–36.

114. Zhou F, Li X, Wang W, Zhu P, Zhou J, He W, Ding M, Xiong F, Zheng X, Li Z, et al. Tracing hematopoietic stem cell formation at single-cell resolution. Nature. 2016;533:487–92.

115. Wang X, Gong Y, Erna H. Chasing the precurso of functional hematopoietic stem cells at the single cell levels in mouse embryos. J Hematol Oncol. 2016;9:38.

116. Hou Y, Guo H, Cao C, Li X, Hu B, Zhu P, Wu X, Wen L, Tang F, Huang Y, Peng J. Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. Cell Res. 2016;26:304–19.

117. Hu Y, Huang K, An Q, Du G, Hu G, Xue J, Zhu X, Wang CY, Xue Z, Fan G. Simultaneous profiling of transcriptome and DNA methylome from a single cell. Genome Biol. 2016;17:88.

118. Xu Y, Shan W, Li X, Wang B, Liu S, Wang Y, Long Y, Tie R, Wang L, Cai S, et al. A synthetic three-dimensional niche system facilitates generation of functional hematopoietic cells from human-induced pluripotent stem cells. J Hematol Oncol. 2016;9:102.

119. Ventura Ferreira MS, Bergmann C, Bodensiek I, Peukert K, Abert J, Kramann R, Kachel P, Rutten S, Knuchel R, et al. An engineered multicomponent bone marrow niche for the recapitulation of hematopoiesis at ectopic transplantation sites. J Hematol Oncol. 2016;9:4.

120. Zhao K, Liu Q. The clinical application of mesenchymal stromal cells in hematopoietic stem cell transplantation. J Hematol Oncol. 2016;9:24.

121. Al-Mawali A, Gillis D, Lewis I. Immunoprofiling of leukemic stem cells CD34+/CD19/CD123+ delineate FLT3/ITD-positive clones. J Hematol Oncol. 2016;9:61.