Cellular and Molecular Life Sciences

REVIEW

Single-cell analysis of the transcriptome and its application in the characterization of stem cells and early embryos

Na Liu · Lin Liu · Xinghua Pan

Abstract Cellular heterogeneity within a cell population is a common phenomenon in multicellular organisms, tissues, cultured cells, and even FACS-sorted subpopulations. Important information may be masked if the cells are studied as a mass. Transcriptome profiling is a parameter that has been intensively studied, and relatively easier to address than protein composition. To understand the basis and importance of heterogeneity and stochastic aspects of the cell function and its mechanisms, it is essential to examine transcriptomes of a panel of single cells. High-throughput technologies, starting from microarrays and now RNA-seq, provide a full view of the expression of transcriptomes but are limited by the amount of RNA for analysis. Recently, several new approaches for amplification and sequencing the transcriptome of single cells or a limited low number of cells have been developed and applied. In this review, we summarize these major strategies, such as PCR-based methods, IVT-based methods, phi29-DNA polymerase-based methods, and several other methods, including their principles, characteristics, advantages, and limitations, with representative applications in cancer stem cells, early development, and embryonic stem cells. The prospects for development of future technology and application of transcriptome analysis in a single cell are also discussed.

Keywords Single cells · Transcriptome · RNA-seq · Cancer stem cells · Embryonic stem cells

N. Liu (✉) · L. Liu
State Key Laboratory of Medicinal Chemical Biology,
Department of Cell Biology and Genetics, College of Life
Science, Nankai University, Tianjin 300071, China
e-mail: liuna@nankai.edu.cn

N. Liu · X. Pan
Department of Genetics, Yale School of Medicine,
Yale University, New Haven, CT 06520, USA
e-mail: xinghua.pan@yale.edu

Introduction

The past decades have witnessed enormous advances in transcriptome (RNA) analysis, including high-throughput RNA profiling with microarray and deep sequencing (RNA-seq). Although microarray platforms and RNA-seq are extremely powerful, most measurements were performed on populations of cells. The sensitivity and accuracy of single-cell methods are limited by the efficiencies of each sample-processing step. Single-cell RNA-seq requires a procedure that can amplify the very limited starting material sufficiently for large-scale analysis, with high reproducibility, high fidelity, and high coverage, and is simple enough for practical application. A number of methods for amplification and analysis of transcriptomes for low number and single cells have been reported. In recent years, RNA-seq at the single-cell level has unprecedentedly paved the way towards understanding individual cell states, greatly enhancing our understanding of the molecular basis of aberrant cell states and disease development. Here, we present a review of several approaches that have successfully sequenced the transcriptomes at the level of individual cells. We will discuss the amplification of transcriptome, its application to single-cell analysis, and prospective applications.

Cellular heterogeneity and transcriptome analysis

Cellular heterogeneity

Cellular heterogeneity is a widespread nature phenomenon and has been observed in multiple cell and tissue types, ranging from simple unicellular organisms to stem cell populations and complex tissues [1–3]. For example, tumors have been known to be heterogeneous mixtures of cell types

for a long time [4]. Solid tumors are frequently composed of individual, molecularly distinct clones that differ in their proliferation rates and metastatic potential, and most critically, in their sensitivities and responses to drug treatment [5]. The key therapeutic targets in tumors might be cancer stem cells, which represent a small percentage of the total mass of tumors, but could be responsible for tumor repopulation following treatment. Dissecting cell-to-cell variations using single-cell analysis is extremely important in understanding cancer stem cell initiation, progression, metastasis, and therapeutic responses [6]. Furthermore, some stem cell populations, such as embryonic stem cells, adult stem cells, and induced pluripotent stem cells, may be heterogeneous populations [2, 7]. Even within a seeming homogeneous tissue, individual cells show clear heterogeneity. The causes of this heterogeneity include differentiation in different ways, varying stages of the cell cycle, cellular senescence, and non-uniform RNA processing and degradation. Important information may be masked by heterogeneity if the cells are studied as a mass. Therefore, analyzing cells individually will lead to a more accurate representation of cell-to-cell variations instead of the stochastic average masked by bulk measurements [8]. Single-cell analysis (SCA) will be crucial for elucidating cellular diversity and heterogeneity. Examination of transcriptomes of individual cells is essential to understand the basis and importance of heterogeneity and stochastic aspects of gene expression [9]. Further improvement in the breadth of single-cell mRNA analysis has been achieved recently using RNA sequencing [10, 11].

Microscopy, FACS, FISH, or real-time PCR-based methods can provide a preliminary type of single-cell analysis to experiments but are able to assay only a handful of genes or some simple phenotypes at a time. This problem can be overcome by sequence or array analysis of the transcriptome. Furthermore, the analysis of the transcriptome in individual cells offers a number of advantages compared to cell-averaging experiments [12–14]. Recent advances in molecular biology have enabled us to carry out single-cell genomic or transcriptomic analysis using whole-genome or cDNA amplification [15–22]. Single-cell analysis represents one of the novel areas of application for high-throughput sequencing, and is particularly important for the study of populations that have a high degree of intrinsic variation, such as the brain, cancer, immune cells and stem cells. Recent studies have also shown that gene expression is invariably heterogeneous even in evidently similar cell types [3, 23]. Such stochastic variations in the transcriptomes have important implication for the composition of cells.

Transcriptome analysis

Omics is the general term for genome-wide studies, including genomics, epigenomics, transcriptomics, proteomics, metabolomics, lipidomics, and interactomics, and is the foundation for systems biology [24]. Single-cell analysis is the new frontier in omics.

The "transcriptome" is the complete set of transcripts for a certain cell or a population of cells. The term was first used in 1997 [25]. Transcriptome analysis has been widely used in molecular biological research for more than a decade. Understanding the transcriptome is essential for interpreting the functional elements of the genome and revealing the molecular constituents of cells and tissues, and also for understanding individual development and disease progress. Transcriptome profiling has many applications in areas of research where acquisition of small, highly specific tissue or cell samples is required for accurate expression analysis. Several methods have been used for transcriptome analysis, such as microarray (RNA hybridization based), SAGE (serial analysis of gene expression), MPSS (massively parallel signature sequencing), and RNA-seq [11]. Hybridization-based microarray analysis is a high-throughput procedure. However, this method has several limitations, which include: reliance upon existing knowledge about the genome sequence; high background levels owing to cross-hybridization; and a limited dynamic range of detection owing to both background and saturation of signals [11, 26]. Recently, following the dramatically decreasing costs of sequencing, RNA-seq has become the preferred method for transcriptomic analyses [11].

In 2006, the first RNA-Seq paper was published [27]. RNA-seq, also known as RNA-sequencing or transcriptome-sequencing, refers to the use of high-throughput sequencing technologies to sequence cDNA libraries transcribed from all the RNAs in cells, and can be used to quantify, profile, and discover new RNA transcripts by sequence reads. The transcripts can then be mapped on the reference genome to get comprehensive information, such as transcription localization and alternative splicing status. RNA-seq has been widely used in biological, medical, and pharmaceutical research [11]. Besides superior accuracy in the quantification of gene expression, RNA-seq offers other advantages, such as the possibility to detect novel transcripts, splice variants or allele-specific expression [28–30]. By analyzing the transcriptome at unprecedented depth and accuracy, thousands of new transcript variants and isoforms have been shown to be expressed in mammalian tissues or organs. These advances greatly accelerate our understanding of the complexity of gene expression, regulation, and networks for mammalian cells [29, 31, 32].

Transcriptome amplification and RNA-seq for single cells

RNA-seq for a single cell can produce information that cannot be attained by analysis of multi-cell populations, but is limited

by the amount of RNA needed for analysis. RNA-seq usually needs microgram amounts of total RNA for analysis, which corresponds to million or submillion level of mammalian cells. However, there are only around ~10 pg of total RNA and ~0.1 pg of mRNA in single cells. In order to process the RNA-seq at the single-cell level, an amplification step is needed. Commonly used amplification techniques are based on two different approaches, linear isothermal amplification by in vitro transcription (IVT) of the cDNA population into complementary RNA (cRNA) [33, 34], and PCR amplification of the entire population of cDNA following reverse transcription [35, 36]. Recently, other methods have been developed to perform RNA-seq at the single-cell level, which were template switching for preparing full-length cDNA for PCR amplification, and Phi29 DNA polymerase-based RNA amplification [37, 38]. The amplified cDNA can then be analyzed with deep sequencing [15–17, 21, 22]. The representative strategies of these will be described in detail in the following parts.

PCR-based amplification

PCR-based amplification was the first report on the preparation of single-cell cDNAs for single-cell transcriptome analysis, such as cDNA microarray and RNA-seq analysis [35]. With PCR-based RNA amplification, any RNA starting amount can be employed. The advantage of the PCR strategy is the exponential amplification of cDNAs so that single-cell cDNAs can be amplified more than a million-fold in several hours, allowing analysis at the single-cell level. In 2009, Tang firstly reported his RNA-seq research at the single-cell level. In this research, a PCR-based amplification was used in combination with SOLID sequencing [10]. The PCR-based method has been extended to include a multiplexing step for the amplification of multiple cells in parallel, allowing for high-throughput analysis, with the Illumina sequencing platform [18].

One complication of PCR amplification is the accumulation of primer dimers and other nonspecific products during amplification, especially during later cycles of PCR [9, 19]. Some concern has also existed over the impact of PCR amplification on the accuracy of gene expression quantification of RNA-seq, as exponential amplification can skew the original quantitative relationships between genes from an initial population. Several modified PCR-based methods of cDNA amplification have been developed, such as global PCR amplification (GA), 3′-end amplification (TPEA), and strand-switch-mediated reverse transcription amplification (SMART) [36, 39–41].

The first effort on single-cell RNA-seq

The first RNA-seq study in single cells appeared in 2009 [10]. This method was based on a widely used single-cell whole-transcriptome amplification method [19, 42]. A single cell is manually picked under a microscope and lysed. Then mRNAs are reverse-transcribed into cDNAs using a polyT primer with anchor sequence and unused primers are digested, after which terminal transferase is used to attach polyA tails to the first-strand cDNAs at the 3′ end, and second-strand cDNAs are synthesized using polyT primers with another anchor sequence. Furthermore, amine-modified primers were used for the second round of PCR to remove the residual primers and primer dimer from the sequencing library to improve throughput. Then, the libraries are amplified, fragmented, and ligated with adaptors. With Applied Biosystems' next-generation sequencing SOLiD system, this method was successfully applied to trace the process of the derivation of embryonic stem cells from the inner cell mass of blastocysts, illustrating that the approach works can be used for the analysis the transcriptome of relatively small-sized individual cells [10, 43].

This method extended the cDNA fragment sizes from 0.85 to 3 kb by extending the incubation time compared with previously reported [19, 42]. This improvement permitted capturing full-length cDNAs for the majority (64 %) of expressed genes [10]. This method treats the cDNA as any double-stranded DNA (dsDNA) for sequencing, and subjects it to a standard sample preparation protocol including fragmentation, adapter ligation, and library amplification. The entire procedure is repeated for each single cell to be analyzed, and the resulting reads were distributed along the entire cDNA length (but strongly biased to the 3′ end), which permitted a partial analysis of alternative splice isoforms [37]. However, this technique has its limitation: for most of the mRNAs longer than 3 kb, the 5′ end of the mRNA will not be detected [10, 21].

SMA

SMA (semirandom primed PCR-based mRNA transcriptome amplification procedure) was another method suitable for single-cell RNA amplification on the bench top. After cDNA was generated, semirandom priming and universal amplification PCR was used to generate a sufficient amount of short (with uniformed sizes) and overlapping cDNA fragments along the entire length of cDNAs for RNA-seq. After the amplification was finished, the semirandom primer was completely removed with a type II restriction enzyme, BciVI, whose recognition sequence was built into the primer [38]. The output DNA fragments can even be directly ligated to sequencing adapter without end-repairing. SMA could be performed in a microfluidic apparatus that has PCR capability. Carrying out the reaction in nanoliter volumes has the potential to substantially improve

single-cell work [44]. Outstandingly, compared with other methods, SMA can cover the full length of any size of transcripts. The full length message of transcripts enables splicing form and other structure message to be detected.

Template-switching-based single-cell RNA-seq

The "template-switching" method is also marketed as the "switching mechanism at the 5′ end of the RNA transcript" (SMART), and it is a specific type of PCR amplification method of RNA. The Moloney Murine Leukemia Virus (MMLV) reversed transcriptase has a property that can add a few non-templated nucleotides (mainly cytosines) to the 3′ end of the first-strand cDNA when the reversed transcription reaches the 5′ end mRNA [37, 45]. When the first-stand cDNA synthesis was finished, the newly added nucleotides at the end of cDNA provide a primer binding site. When an added oligonucleotide contains sequences that can pair with these new nucleotides, the reverse transcriptase switches from the mRNA to the new DNA oligonucleotide [37, 41, 45]. The cDNA then contains primer binding sites at both the 3′ and 5′ end and can be amplified by stretch PCR. The resulting cDNA is enriched for full-length transcripts. This is the advantage of this method, and is also the main disadvantage, as template switching will occur only if reverse transcription successfully reaches the 5′ end; any partially reverse transcribed mRNA will fail to be obtained. Three RNA-seq approaches have been published based on this method: STRT-seq, SMART-seq, and SMART-seq2.

STRT-seq

STRT (SMART-based single-cell tagged reverse transcription) is the second RNA-seq method used at the single-cell level to analyze the transcriptome. STRT method is a PCR-based multiplexed single-cell RNA-seq method using an Illumina platform, and was introduced by Islam et al. [18]. It is the first template-switching method-based single-cell RNA-seq protocol. By using template-switching mechanism, a barcode and an upstream primer-binding sequence are introduced simultaneously with reverse transcription. All cDNAs are pooled and amplified together, then prepared for 5′ end sequencing.

STRT quantifies transcripts through reads mapping to 5′ ends of mRNA [18], compared with the initial mRNA-seq method, which preferentially amplified the 3′ ends of mRNAs. Hence, the data could only be used to identify distal splicing events [10]. An important aspect of STRT is its ability to pinpoint the exact location of the 5′ end of transcripts, which is often lost in methods that show 3′ bias. This method could be used to analyze promoter usage in single cells and, in effect, provide a straightforward method

for single-cell CAGE (cap analysis of gene expression). Furthermore, it facilitates quantification, as each mRNA molecule results in a single cDNA molecule, and thus the number of reads observed should be proportional to the number of mRNA molecules. The chief advantage of this approach is the great reduction in cost and time, afforded by the early barcoding strategy. Compared with previous methods [10], it is more suitable for large-scale quantitative of single cell analysis, as well as for the characterization of transcription start sites, but it is unsuitable for the detection of alternatively spliced transcripts [18, 37].

Smart-seq

A new RNA-seq method, called Smart-seq, can be used to study single-cell transcriptome analysis. This Smart-seq also relies on template-switching technology as STRT-seq, which generates full-length transcripts that can then be amplified and sequenced. Smart-seq is a simplification of STRT without barcoding and so it is a lower throughput method. It has improved read coverage across transcripts, which enhances detailed analyses of alternative transcript isoforms and identification of single-nucleotide polymorphisms [46]. The researchers used the method to analyze gene expression of circulating tumor cells isolated from the blood of a melanoma patient. They report hundreds of differentially expressed genes by comparing a few cells per cell type [46]. This protocol, in slightly altered form, was successfully implemented to profile the transcriptome of single neurons [47].

Current methods to examine single-cell transcriptome suffer from coverage limitation, which either favors the 3′ end or the 5′ end [10, 18]. The Smart-seq kit that came out last year is the first one that aims to cover the whole part of the transcript (towards full-length coverage). It gives us more information, not only about the gene expression level, but which variant of a gene is expressed. It should be noted that the full-length cDNA is synthesized, some incomplete synthesized cDNA are lost in the analysis, which may reduce the coverage when it is applied to single cells because in single cell process many transcripts may be partially damaged.

Smart-seq2

Recently, an improved method, Smart-seq2, which is an updated version of Smart-seq, was reported in the journal Nature Methods [48]. Based on the Smart-seq, this team systematically evaluated a large number of variations in reverse transcription, template-switching, and PCR pre-amplification and compared the results to Smart-seq. They generated a new protocol for single-cell RNA-seq, named Smart-seq2 [48]. Smart-seq2 captures an increased number of transcripts from a single cell, along with increases

in transcript coverage and accuracy compared to Smart-seq libraries. RNA sensitivity was increased about threefold. The new procedure consistently captures three to four times as many RNA molecules, which often translates into 2,000 more genes per cell than current methods allow [46]. Such coverage is important when trying to simultaneously track polymorphism or mutation patterns and gene expression with RNA from individual cells. Furthermore, by using off-the-shelf reagents, Smart-seq2 is very cost-effective [48].

IVT-based amplification

Linear RNA amplification is the first strategy that has been used to successfully amplify RNA for molecular profiling studies [34], which actually promoted the birth of the era of single cell analysis. The most commonly used mechanism for linear isothermal RNA amplification is based on T7 RNA polymerase-mediated in vitro transcription (IVT) [34, 49]. IVT-based RNA amplification protocol is more tedious and time-consuming, requiring three rounds of amplification (~5 days work/cell), because each round of IVT can amplify the cDNAs only up to 1,000-fold [33]. These features have prevented the original IVT protocol from being directly used for single-cell RNA-Seq. The main advantage of the IVT strategy is its specificity and ratio fidelity while reducing accumulation of nonspecific products; its drawback is that cDNAs typically less than 1 kb are generated [9, 33], the resulting library is biased towards the 3′ end of genes, the low efficiency, and the time consuming procedure. This disadvantage was then partly overcome by several groups and cDNA sequences up to 3 kb or more were obtained [10, 21, 38, 42].

CEL-seq

CEL-seq (cell expression by linear amplification and sequencing), a protocol that barcodes and pools multiple samples before linearly amplifying mRNA, partially meets the demand of linear amplification by IVT for sufficient material by pooling barcoded samples, therefore allowing the efficient linear amplification of RNA from single cells and their analysis by sequencing [50]. The CEL-seq method begins with a single-cell reverse-transcription reaction using a primer designed with an anchored polyT, a unique barcode, the 5′ Illumina sequencing adaptor, and a T7 promoter. Second-strand synthesis is performed and then the cDNA samples are pooled and comprise sufficient template material for IVT reaction. The RNA is fragmented to a size distribution appropriate for sequencing, the Illumina 3′ adaptor is added, RNA is reverse transcribed to DNA, and the 3′-most fragments that contain both Illumina adaptors and a barcode are selected. The resulting library undergoes paired-end sequencing, where the first read recovers the barcode, whereas the second read identifies the mRNA transcript [50]. CEL-seq method was successfully used to analyze early *C. elegans* embryonic development at a single-cell resolution. The robust transcriptome quantifications enabled by CEL-seq will be useful for transcriptomic analyses of complex tissues containing populations of diverse cell types [50]. When compared to the STRT method introduced above [18], CEL-seq gives more reproducible, linear, and sensitive results than a PCR-based amplification method [50]. Its advantage lies in the linear mode of amplification, which unlike PCR amplification, does not exponentially deplete sequences that are unfavorable to PCR process [45].

Quartz-seq

Quartz-seq is another single-cell RNA-seq method based on a poly-A tailing, PCR amplification and in-vitro transcription, which claims a simpler protocol and higher reproducibility and sensitivity surveying the mRNA content of individual cells than existing methods. This method has certain advantages: byproduct synthesis was suppressed obviously; the whole amplification process can be finished in a single tube. The Quartz-Seq approach successfully detected gene expression heterogeneity between ES cells and more developmentally mature "primitive endoderm" cells based on gene expression. Quartz-seq is also able to measure clear differences in ES cells gene expression between different cell-cycle phases. The reproducibility and sensitivity of Quartz-seq are high. Compared with the Smart-seq method, Quartz-seq has a higher Pearson correlation coefficient (PPC) (the PCC of Quartz-Seq was approximately 0.93, whereas the PCC of Smart-Seq was approximately 0.72). Quartz-seq can detect 81.4 % transcripts, more than Smart-seq (only 63.1 %). Higher reproducibility and sensitivity were also detected when compared with the CEL-seq method [51]. The Quartz-seq proves valuable for exploring genetic differences among individual cells.

Phi29 DNA polymerase-based RNA amplification

Several approaches have been proposed for obtaining transcriptome data from single cells as illustrated above. These approaches may yield biased representations of sequences along the mRNA and fail to give complete sequences for long mRNAs because long DNA templates are discriminated against by either PCR or IVT based methods [10, 50]. Two single-cell RNA amplification methods were raised recently, based on the Phi29 DNA polymerase [52, 53]. The Phi29 DNA polymerase is the replicative polymerase from the *Bacillus subtilis* phage phi29 (Φ29) [52]. This polymerase is a highly processive polymerase with strong strand displacement activity that allows for highly efficient isothermal DNA amplification. This enzyme was first

successfully used in whole genome amplification (WGA) [53], and now it can also be used in RNA amplification.

TTA

A novel method of TTA (total transcript amplification), using Phi29 polymerase multiple displacement amplification (MDA) of circularized cDNA, was developed for a single prokaryotic cell transcriptome analysis [54]. These existing methods of transcript amplification described above involve multiple rounds of PCR and/or linear amplification of cDNA [10, 50], but would encounter many challenges when these methods were used in single bacterium. These challenges included the low amount of RNA, the lack of polyA-tails for easy tagging and mRNA specific amplification [54]. The TTA method yielded reproducible data, low fold-change bias, and a high number of genes efficiently amplified from a single prokaryotic cell with low drop-outs as detectable by microarray [54].

PMA

Phi29 DNA polymerase-based mRNA transcriptome amplification (PMA) is an approach for single and low quantities (LQ) cell cDNA amplification. It is the 1st method using phi29 DNA polymerase to generate material for sequencing the whole mRNA transcriptome. Firstly, the mRNA derived cDNA was selectively generated from the total RNA by poly-dT primed reverse transcription. Following it, the cDNA was circularized by intermolecular ligation. Then the phi29 DNA polymerase-based rolling cycle amplification (RCA) was applied. Therefore, PMA capture both the 5′ end and 3′ end sequences [38]. PMA is suitable for sequencing the transcriptome covering full length of all sizes of transcripts. These full-length cDNAs would be important for resolution of ambiguities in assigning splice isoforms. Compared with existed methods, PMA is a method most suitable for full-length RNA sequencing. Furthermore, PMA has a particular advantage for application to microfluidic systems: it is relatively simple in operation with few steps of manipulations and isothermal reaction. This would allow a large number of single cells to be amplified in parallel [38]. It needs to note that PMA requires the gDNA in sample to be removed before the amplification procedure is applied. This is then rewarded with very clean RNA data without any genomic DNA signal.

Application of single-cell RNA-seq in stem cells and early development characterization

Single-cell RNA-seq can be used to determine the gene expression regulation networks at the whole-genome scale, especially for stem cell identification, including cancer stem cells, pluripotent stem cells, and adult stem cells. Furthermore, this technique is relevant for the analysis of cells during early development of the embryos. All of these cells possess highly dynamic character with a high level of heterogeneous subpopulations. Single-cell transcriptome analysis is very important for analysis of rare cell types.

Identifying circulating cancer stem cells

As introduced above, cell heterogeneity between cells in tumors has been known for a long time [4]. Such cellular heterogeneity could be studied by robust techniques for single-cell analysis. Single-cell transcriptome analysis is a feasible strategy to elucidate the heterogeneity and somatic mutation-based evolution of tumors, to identify the subpopulations in a tumor, and to detect putative cancer stem cells [55]. Circulating tumor cells (CTCs) are cells that depart from a primary tumor or metastasis and enter the blood stream via the leaky vasculature that arises around a growing tumor. CTCs may be defined as circulating cancer stem cells, for its capacity of self-renewal and for initiation of distant metastases; some of these cells are also resistant to traditional chemotherapy. These circulating tumor cells are clinically significant. They have important roles in diagnostic, prognostics, and therapeutic implications for patients with cancer [56]. Transcriptomic analysis of individual CTCs might provide useful information to make personalized medical decisions for cancer therapy and provide insights into the biological processes involved in metastasis. However, CTCs transcriptome analysis was limited by the low amount of cells in blood. The progress in single-cell RNA-seq offers an excellent opportunity to advance the understanding of gene expression in individual CTCs and test whether CTC transcriptomic information can be used in clinical therapy for tumor patients. Smart-seq was successfully used to putative circulating tumor cells captured from the blood of a melanoma patient. Distinct gene expression patterns were identified for melanoma circulating tumor cells [46]. The gene expression profile of CTCs was also identified in prostate cancer using single-cell mRNA-seq [57].

Monitoring the change of early embryo development and embryonic stem cells

Every mammalian individual is developed from the totipotent zygote. Mammalian pre-implantation development is a complex process involving dramatic changes in the transcriptional architecture, such as zygotic genome activation process (ZGA). It is common opinion that this step playing important role in the early development. However, previous gene expression profiling of human pre-implantation has

been limited by blastomere size. Recent advances in single-cell RNA-seq technology have provided the unprecedented opportunity to study gene regulation in the ZGA process [58]. In this research, they identified novel stage-specific genes during the pre-implantation embryo development using the single-cell RNA-seq technology. This finding extends our knowledge of the transcriptional architecture, sequential order of gene activation, and genetic programming for early embryogenesis [58].

Differential gene expression in individual cells is a key determinant of cellular differentiation, functions, and physiology. Deciphering the gene expression in embryos at single cell level is a crucial step towards understanding early developmental processes. Pioneering work in transcriptional profiling has been carried out by transcriptional technology using bulk of cells, and has given some important information [59, 60]. However, as this information obtained from bulk of cells, some specific cell fate determined gene expression which differs from one cell to others will be masked, and some cell types are transient. To well characterize the genetic basis of cell fate determination and coordination among the embryonic stem cells, transcriptional profiling should be done at the single-cell level. Studying gene regulation at the single-cell level is necessary and effective for understanding normal development [61]. Single cell level mRNA-seq has been used to analyze the digital transcriptome of individual blastomeres in early mouse embryos [10, 62]. A recent research successfully profiles a comprehensive set of transcriptome landscapes of human pre-implantation embryos and human ES cells [63]. The researchers applied single-cell RNA-Seq analysis with global gene expression analysis for the gene expression characteristics of three lineages at human late blastocyst, which is trophectoderm (TE), epiblast (EPI), and primitive endoderm (PE). They have also historically documented the dynamics of global gene expression during the process of human ES cells derivation and found significant differences in transcriptomes between that EPI cells and primary human ES cells outgrowth [63]. These studies provide a comprehensive framework of the transcriptome landscapes of early development and embryonic stem cells.

Embryonic stem cells (ES cells) are derived from the inner cell mass (ICM) of the blastocysts, and have indefinitely self-renewal ability and pluripotency [64]. There have been intensive studies on ES cells in recent years, but these have usually been on bulk cells by RNA-seq and microarray [65]. The precise changes accompanying the process of ES cells derivation is not clear, due to the limitation in technology. In order to investigate the early development, the single cell transcriptome analysis provides us an excellent tool. Now, several groups have successfully gained insight into this process by the single cell RNA-seq technology [43]. Furthermore, it is feasible to address allele-specific expression (ASE) during early development using single-cell RNA-seq analysis [62].

Prospects

As reviewed above, for the heterozygous cell population, functional transcriptomics of a single cell can produce a wealth of information at resolutions that cannot be attained by analysis of multi-cell populations or communities, which will be more helpful for us in understanding the relationship between cells. Such advances hinge on the development of innovative methods for single-cell isolation and transcript amplification from a minute amount of starting material with low gene expression bias. Despite these approaches on single-cell transcriptomic analysis, there is still much work to be done, both in enabling new modes and algorithm of analysis, and improving the robustness speed, sensitivity, precision, and economy of those that exist. A major problem is that although we can process several analyses at the single-cell level, such as DNA sequencing, RNA-seq, telomere length measurement, epigenomic analysis, etc [14, 66, 67], at present we can not perform several analyses of individual cells simultaneously. It is very important for the investigators to understand the genotype and phenotype of individual cells, so new methods are expected to analyze a single cell at the genomics, epigenomics, transcriptomics and even proteomics level simultaneously.

References

1. Irish JM, Kotecha N, Nolan GP (2006) Mapping normal and cancer cell signalling networks: towards single-cell proteomics. Nat Rev Cancer 6(2):146–155
2. Graf T, Stadtfeld M (2008) Heterogeneity of embryonic and adult stem cells. Cell Stem Cell 3(5):480–483
3. Huang S (2009) Non-genetic heterogeneity of cells in development: more than just noise. Development 136(23):3853–3862
4. Shackleton M et al (2009) Heterogeneity in cancer: cancer stem cells versus clonal evolution. Cell 138(5):822–829
5. Turner NC, Reis-Filho JS (2012) Genetic heterogeneity and cancer drug resistance. Lancet Oncol 13(4):e178–e185
6. Clarke MF et al (2006) Cancer stem cells—perspectives on current status and future directions: AACR workshop on cancer stem cells. Cancer Res 66(19):9339–9344
7. Buganim Y et al (2012) Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. Cell 150(6):1209–1222

8. Wang D, Bodovitz S (2010) Single cell analysis: the new frontier in 'omics'. Trends Biotechnol 28(6):281–290
9. Tang F, Lao K, Surani MA (2011) Development and applications of single-cell transcriptome analysis. Nat Methods 8(4 Suppl):S6–11
10. Tang F et al (2009) mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods 6(5):377–382
11. Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10(1):57–63
12. Warren L et al (2006) Transcription factor profiling in individual hematopoietic progenitors by digital RT-PCR. Proc Natl Acad Sci USA 103(47):17807–17812
13. Raj A et al (2008) Imaging individual mRNA molecules using multiple singly labeled probes. Nat Methods 5(10):877–879
14. Wang F et al (2013) Robust measurement of telomere length in single cells. Proc Natl Acad Sci USA 110(21):E1906–E1912
15. Hou Y et al (2012) Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm. Cell 148(5):873–885
16. Navin N et al (2011) Tumour evolution inferred by single-cell sequencing. Nature 472(7341):90–94
17. Xu X et al (2012) Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. Cell 148(5):886–895
18. Islam S et al (2011) Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. Genome Res 21(7):1160–1167
19. Kurimoto K et al (2006) An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis. Nucleic Acids Res 34(5):e42
20. Saitou M, Barton SC, Surani MA (2002) A molecular programme for the specification of germ cell fate in mice. Nature 418(6895):293–300
21. Tang F et al (2010) RNA-Seq analysis to capture the transcriptome landscape of a single cell. Nat Protoc 5(3):516–535
22. Tougan T, Okuzaki D, Nojima H (2008) Chum-RNA allows preparation of a high-quality cDNA library from a single-cell quantity of mRNA without PCR amplification. Nucleic Acids Res 36(15):e92
23. Li L, Clevers H (2010) Coexistence of quiescent and active adult stem cells in mammals. Science 327(5965):542–545
24. Hood L et al (2004) Systems biology and new technologies enable predictive and preventative medicine. Science 306(5696):640–643
25. Velculescu VE et al (1997) Characterization of the yeast transcriptome. Cell 88(2):243–251
26. Royce TE, Rozowsky JS, Gerstein MB (2007) Toward a universal microarray: prediction of gene expression through nearest-neighbor probe sequence identification. Nucleic Acids Res 35(15):e99
27. Bainbridge MN et al (2006) Analysis of the prostate cancer cell line LNCaP transcriptome using a sequencing-by-synthesis approach. BMC Genom 7:246
28. Morin R et al (2008) Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. Biotechniques 45(1):81–94
29. Sultan M et al (2008) A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. Science 321(5891):956–960
30. Ozsolak F, Milos PM (2011) RNA sequencing: advances, challenges and opportunities. Nat Rev Genet 12(2):87–98
31. Mortazavi A et al (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5(7):621–628
32. Wang ET et al (2008) Alternative isoform regulation in human tissue transcriptomes. Nature 456(7221):470–476
33. Eberwine J et al (1992) Analysis of gene expression in single live neurons. Proc Natl Acad Sci USA 89(7):3010–3014
34. Van Gelder RN et al (1990) Amplified RNA synthesized from limited quantities of heterogeneous cDNA. Proc Natl Acad Sci USA 87(5):1663–1667
35. Brady G, Iscove NN (1993) Construction of cDNA libraries from single cells. Methods Enzymol 225:611–623
36. Dixon AK et al (1998) Expression profiling of single cells using 3 prime end amplification (TPEA) PCR. Nucleic Acids Res 26(19):4426–4431
37. Islam S et al (2012) Highly multiplexed and strand-specific single-cell RNA 5′ end sequencing. Nat Protoc 7(5):813–828
38. Pan X et al (2013) Two methods for full-length RNA sequencing for low quantities of cells and single cells. Proc Natl Acad Sci USA 110(2):594–599
39. Iscove NN et al (2002) Representation is faithfully preserved in global cDNA amplified exponentially from sub-picogram quantities of mRNA. Nat Biotechnol 20(9):940–943
40. Matz M et al (1999) Amplification of cDNA ends based on template-switching effect and step-out PCR. Nucleic Acids Res 27(6):1558–1560
41. Zhu YY et al (2001) Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. Biotechniques 30(4):892–897
42. Kurimoto K et al (2007) Global single-cell cDNA amplification to provide a template for representative high-density oligonucleotide microarray analysis. Nat Protoc 2(3):739–752
43. Tang F et al (2010) Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-Seq analysis. Cell Stem Cell 6(5):468–478
44. Lecault V et al (2012) Microfluidic single cell analysis: from promise to practice. Curr Opin Chem Biol 16(3–4):381–390
45. Hebenstreit D (2012) Methods, challenges and potentials of single Cell RNA-seq. Biology 1(3):658–667
46. Ramskold D et al (2012) Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nat Biotechnol 30(8):777–782
47. Qiu S et al (2012) Single-neuron RNA-Seq: technical feasibility and reproducibility. Front Genet 3:124
48. Picelli S et al (2013) Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat Methods 10:1096–1098
49. Patel OV et al (2005) Validation and application of a high-fidelity mRNA linear amplification procedure for profiling gene expression. Vet Immunol Immunopathol 105(3–4):331–342
50. Hashimshony T et al (2012) CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. Cell Rep 2(3):666–673
51. Sasagawa Y et al (2013) Quartz-Seq: a highly reproducible and sensitive single-cell RNA sequencing method, reveals non-genetic gene-expression heterogeneity. Genome Biol 14(4):R31
52. Blanco L, Salas M (1984) Characterization and purification of a phage phi 29-encoded DNA polymerase required for the initiation of replication. Proc Natl Acad Sci USA 81(17):5325–5329
53. Dean FB et al (2002) Comprehensive human genome amplification using multiple displacement amplification. Proc Natl Acad Sci USA 99(8):5261–5266
54. Kang Y et al (2011) Transcript amplification from single bacterium for transcriptome analysis. Genome Res 21(6):925–935
55. Dalerba P et al (2011) Single-cell dissection of transcriptional heterogeneity in human colon tumors. Nat Biotechnol 29(12):1120–1127
56. Toloudi M et al (2011) Correlation between cancer stem cells and circulating tumor cells and their value. Case Rep Oncol 4(1):44–54
57. Cann GM et al (2012) mRNA-Seq of single prostate cancer circulating tumor cells reveals recapitulation of gene expression and pathways found in prostate cancer. PLoS One 7(11):e49144
58. Xue Z et al (2013) Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. Nature 500(7464):593–597

59. Aghajanova L et al (2012) Comparative transcriptome analysis of human trophectoderm and embryonic stem cell-derived trophoblasts reveal key participants in early implantation. Biol Reprod 86(1):1–21

60. Dobson AT et al (2004) The unique transcriptome through day 3 of human preimplantation development. Hum Mol Genet 13(14):1461–1470

61. Guo G et al (2010) Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. Dev Cell 18(4):675–685

62. Tang F et al (2011) Deterministic and stochastic allele-specific gene expression in single mouse blastomeres. PLoS One 6(6):e21208

63. Yan L et al (2013) Single-cell RNA-Seq profiling of human pre-implantation embryos and embryonic stem cells. Nat Struct Mol Biol 20(9):1131–1139

64. Evans MJ, Kaufman MH (1981) Establishment in culture of pluri-potential cells from mouse embryos. Nature 292(5819):154–156

65. Niwa H (2007) How is pluripotency determined and maintained? Development 134(4):635–646

66. Zong C et al (2012) Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. Science 338(6114):1622–1626

67. Moroz LL, Kohn AB (2013) Single-neuron transcriptome and methylome sequencing for epigenomic analysis of aging. Methods Mol Biol 1048:323–352