



Single Cell Multi-Omics Technology: Methodology and Application

Youjin Hu^{1*}, Qin An², Katherine Sheu², Brandon Trejo², Shuxin Fan¹ and Ying Guo^{3*}

¹ Zhongshan Ophthalmic Center, State Key Laboratory of Ophthalmology, Sun-Ye-Sat University, Guangzhou, China,

² Department of Human Genetics, David Geffen School of Medicine, UCLA, Los Angeles, CA, United States, ³ The Second Affiliated Hospital, Xiangya School of Medicine, Central South University, Changsha, China

OPEN ACCESS

Edited by:

Xinghua Victor Pan,
Yale University, United States

Reviewed by:

Zhibin Wang,
Johns Hopkins University,
United States
Leonard C. Edelstein,
Thomas Jefferson University,
United States
Stephen Clark,
Babraham Institute (BBSRC),
United Kingdom

*Correspondence:

Youjin Hu
huyoujin@gzoc.com
Ying Guo
mytyl.g@hotmail.com

Specialty section:

This article was submitted to
Molecular Medicine,
a section of the journal
Frontiers in Cell and Developmental
Biology

Received: 15 December 2017

Accepted: 08 March 2018

Published: 20 April 2018

Citation:

Hu Y, An Q, Sheu K, Trejo B, Fan S
and Guo Y (2018) Single Cell
Multi-Omics Technology:
Methodology and Application.
Front. Cell Dev. Biol. 6:28.
doi: 10.3389/fcell.2018.00028

In the era of precision medicine, multi-omics approaches enable the integration of data from diverse omics platforms, providing multi-faceted insight into the interrelation of these omics layers on disease processes. Single cell sequencing technology can dissect the genotypic and phenotypic heterogeneity of bulk tissue and promises to deepen our understanding of the underlying mechanisms governing both health and disease. Through modification and combination of single cell assays available for transcriptome, genome, epigenome, and proteome profiling, single cell multi-omics approaches have been developed to simultaneously and comprehensively study not only the unique genotypic and phenotypic characteristics of single cells, but also the combined regulatory mechanisms evident only at single cell resolution. In this review, we summarize the state-of-the-art single cell multi-omics methods and discuss their applications, challenges, and future directions.

Keywords: single cell transcriptome, single cell multi-omics profiling, single cell epigenome, single cell proteome, gene regulation, epigenetics

INTRODUCTION

According to the central dogma, also known as the DNA-RNA-protein axis, DNA provides the code for RNA, which is translated to produce proteins that fulfill biological functions (Crick, 1970). To discover the regulatory mechanisms behind RNA transcription and protein translation, the most straightforward approach is to analyze both DNA and RNA, or both RNA and protein, from the same sample. Despite the complexity of tissues comprised of heterogeneous cell populations, such as cancer, most experimental results to date have been based on analysis of bulk samples, which theoretically read an averaged signal from the population and prevent resolution of cellular variation (Navin et al., 2011; Huang et al., 2015; Gawad et al., 2016). To decipher the mechanism of heterogeneous gene transcriptional regulation, integrated measurement and co-analysis of multiple types of molecules, such as DNA, RNA, and protein, at single cell level is required.

The invention of PCR methods in 1983 made it possible to analyze the picogram amounts of DNA in single cells, although these initial methods could only amplify small, targeted regions of the genome. However, the development of whole genome amplification (WGA) and whole transcriptome amplification (WTA) methods (Tang et al., 2009; Zong et al., 2012; Huang et al., 2015; Wang and Navin, 2015; Gawad et al., 2016) soon allowed quantitative measurement of DNA and RNA for multiple genes in single cells. At the same time, the development of next generation sequencing technology has enabled genome-wide analysis of DNA and RNA in single cells. Inspired by the very first report of single cell DNA sequencing

and single cell RNA sequencing, scientists have developed numerous methods to measure other omics at single cell level, including single cell DNA methylation, single cell chromatin sequencing and single cell proteome analysis [Figure 1, A detailed introduction of single cell sequencing methods has been reviewed elsewhere (Wang and Navin, 2015; Gawad et al., 2016)].

Single cell genome-wide approaches provide a valuable opportunity to measure different molecules, such as DNA, RNA, protein, and chromatin with ultimate resolution. By isolating multiple types of molecules (DNA, RNA, or protein) from a single cell simultaneously, it is feasible to profile different types of molecules in parallel. For example, genomic DNA can be used to assay the single cell genome, methylome or chromatin accessibility, while RNA from the same cell can be used to profile the transcriptome, and protein the proteome. Utilizing these different single cell omics profiling strategies as building blocks, we can construct a multi-omics profile for the same cell. Here, we summarize current single cell multi-omics approaches, such as scG&T-seq (single cell Genome & Transcriptome sequencing), scMT-seq (single cell Methylome and Transcriptome sequencing), scM&T-seq (single cell Methylome & Transcriptome sequencing), scTrio-seq (single-cell triple omics sequencing), and scCOOL-seq (single cell Chromatin Overall Omic-scale Landscape Sequencing) (MacAulay et al., 2015; Angermueller et al., 2016; Hou et al., 2016; Hu et al., 2016), with each of them measuring a different combination of omics data (Figure 2). We also review the bioinformatics advances that have been necessary to understand the large amounts of multi-dimensional data arising from single cell multi-omics profiling, and we examine the potential for this technology to elucidate numerous biological enigmas.

METHODS FOR ISOLATING MULTIPLE TYPES OF MOLECULES FROM A SINGLE CELL

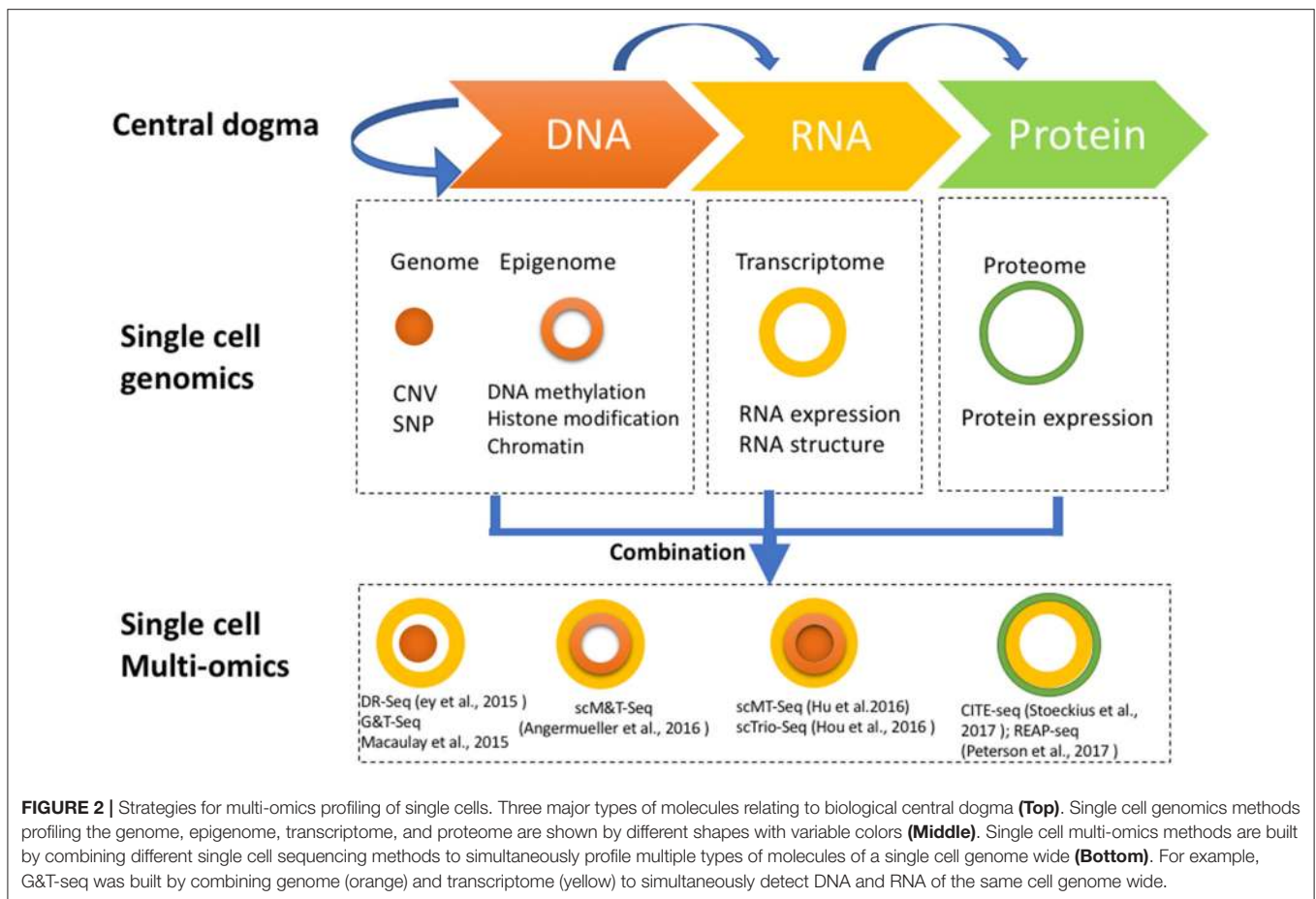
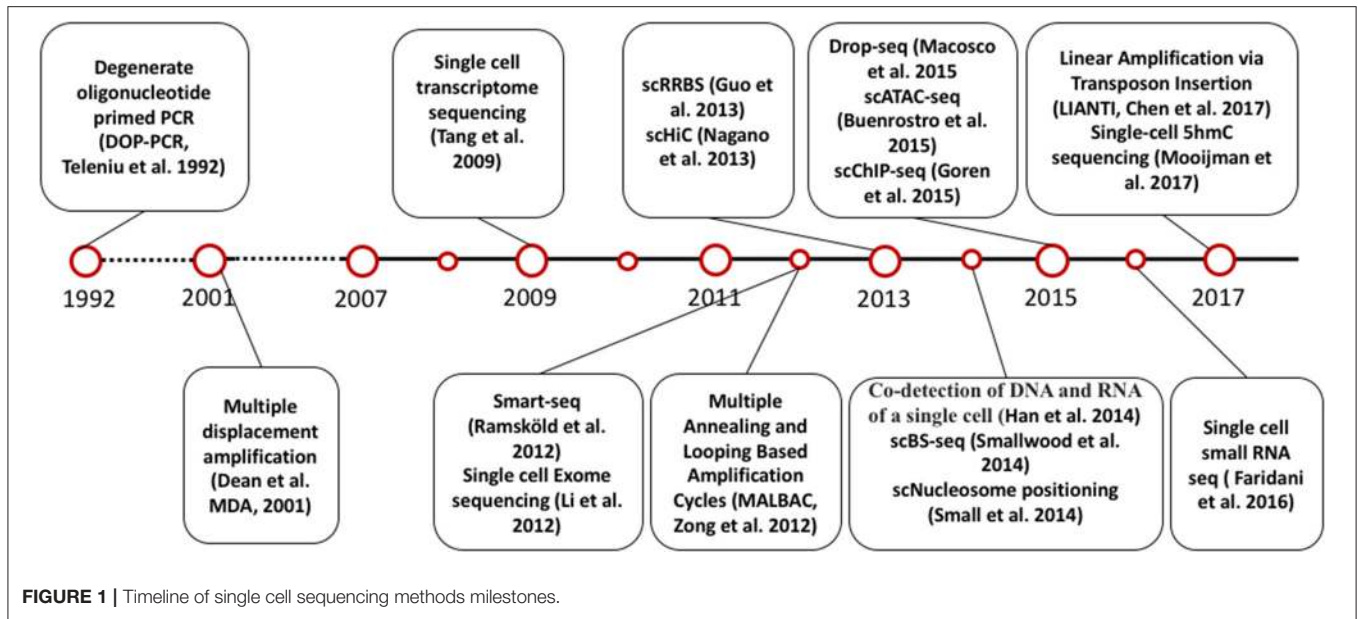
Isolating multiple types of molecules from a single cell is the starting point for single cell multi-omics measurement, and generally can be divided into two steps.

The first step is to collect a single cell randomly from a population with heterogeneity. The standard protocol is to get viable, intact cells by mechanical or enzymatic dissociation and then capture single cells from the dissociated cell suspension. Several approaches can be used, including mouth pipetting, serial dilution, robotic micromanipulation, flow-assisted cell sorting (FACS), and microfluidic platforms (Wang and Navin, 2015). Although these collection approaches are borrowed from methods developed for single cell mono-omics sequencing, additional considerations must be taken for multi-omics to ensure that multiple types of molecules can be viably measured in the same cell. The success of this first collection step is critical for preserving an accurate representation of the DNA, RNA, and protein within the cell for downstream measurements. The method used for the initial dissociation of tissues into single cells—mechanical or enzymatic—needs to be selected with consideration for both the nature of the

starting material and the types of sequencing to be performed. Clinical samples such as solid tumors are often obtained flash frozen or embedded in paraffin (FFPE), making multi-omics measurements that include cytoplasmic RNA or protein more challenging. However, because this type of freezing process perturbs the cytoplasmic membrane while keeping the nuclear membrane intact, multi-omics measurements that involve the genome, epigenome, and chromatin-associated RNA are still possible after creation of nuclear suspensions (Navin, 2015). For fresh tissues, choice of mechanical or enzymatic dissociation reflects the need for both cell integrity and dissociation quality. Prolonged exposure to common dissociation enzymes such as papain, collagenase, dispase, and neutral protease can result in degradation of RNA and proteins, or generation of cell debris that aberrantly activate cell signaling pathways and cell surface proteins (Autengruber et al., 2012; Volovitz et al., 2016). Mechanical mincing of the starting material through trituration or nanofiltration may also disrupt accurate representation of the proteome or transcriptome in cells that contain long projections such as neurons. These pitfalls in turn can complicate the subsequent computational analyses performed on the data, which often involve identification of correlative relationships among the different layers of multi-omics data obtained. Thus, both tissue-specific and measurement-specific aspects of obtaining multi-omics measurements need to be considered in order to achieve optimized single cell suspensions.

Next, the technique used to select single cells after separation of bulk tissues also has an impact on the feasibility of combinatorial multi-omics measurements. The advantages of techniques such as mouth pipetting and serial dilution include the simplicity and rapidity of moving single cells from the cell suspension to individual reaction chambers. This helps limit the degradation of more volatile molecules such as RNA or protein and may reduce the possibility of non-physiologic changes in chromatin accessibility and chromatin conformation (Wang and Navin, 2015; Svensson et al., 2017). Robotic manipulation, FACS, and microfluidic capture platforms have the advantage of the ability to sort through subpopulations by cell labeling, but require more extensive manipulation of single cells using expensive equipment (Ortega et al., 2017). Of the numerous options, selection of a protocol for isolating single cells for multi-omics data collection will ultimately depend on the molecules that need to be preserved, the type of tissue obtained, and the cost.

The second step is to isolate multiple types of molecules from the same cell, for which there are four main strategies: To isolate DNA and RNA of a single cell, the first strategy is physical separation, including separation of nucleus from cytosol, as genomic DNA is contained in the nucleus and the majority of mRNAs are located in the cytosol. Single cells are treated with a membrane-selective lysis buffer, through which the cell membrane is broken down while the nucleus is kept intact. Then, single nuclei are separated from cytoplasm by micropipetting, centrifugation, or antibody-conjugated magnetic microbeads (Hou et al., 2016; Hu et al., 2016; Han et al., 2018; Table 1). This method has been demonstrated to be highly efficient by several research groups, including our lab. Our data indicates that profiling of cytosolic RNA can resemble the transcriptome of the



whole cell. However, this method is low throughput (Hu et al., 2016), as the nucleus-picking procedure is manual and cannot be automated easily. Methods based on centrifugation (Hou et al.,

2016) or antibody conjugated magnetic microbeads (Han et al., 2018) can achieve relatively higher throughput in isolating DNA and RNA from single cells.

TABLE 1 | Current multi-omics methods.

| Methods (time) | Cell type | Measurement | Approach | Single cell isolation | Major discovery | References |
|--|---|---|--|---------------------------------------|--|---------------------------|
| DR-seq (single cell gDNA and mRNA sequencing) (2015) | Mouse embryonic stem cell line (E14) and breast cancer cell line (SK-BR-3). | Genome, mRNA transcriptome | Amplify gDNA and synthesize cDNA without physically separating the nucleic acids. Product is then split for scWGS (single cell Whole Genome Sequencing) and scRNA-seq. | Mouth pipet | Genome copy number variation could drive transcriptome variability. | Dey et al., 2015 |
| G&T-seq (single cell genome & Transcriptome sequencing) (2016) | HCC38, HCC38-BL and iPSCs carrying trisomy 21. | Genome, mRNA transcriptome | Cell is lysed, genomic DNA and poly(A)+ mRNA is separated by magnetic beads for scWGS and scRNA-seq. | Flow cytometry | Identified transcriptional consequences of chromosomal aneuploidies and inter-chromosomal fusions. | MacAulay et al., 2015 |
| scMT-seq (single cell Methylation and Transcriptome sequencing) (2016) | Mouse dorsal root ganglion neurons. | DNA methylation, mRNA transcriptome | Cell is lysed, cell nucleus containing genomic DNA and cell lysis poly(A)+ mRNA is separated for scRRBS (single cell Reduced Representation Bisulfite Sequencing) and scRNA-seq. | Mouth pipet | Methylation of non-CGI promoters is better anti-correlated with gene transcription, gene body methylation of CGI promoter genes has higher correlation with transcription, potentially reveal allelic specific methylation and allelic expression in single cells. | Hu et al., 2016 |
| scM&T-seq (single cell Methylation & Transcriptome sequencing) (2016) | Mouse embryonic stem cell line (E14), in serum and 2i conditions. | DNA methylation, mRNA transcriptome | Cell is lysed, genomic DNA and poly(A)+ mRNA is separated by magnetic beads for scWGBS (single cell Whole Genome Bisulfite Sequencing) and scRNA-seq (single cell RNA sequencing). | Flow cytometry | Non-CGI promoter methylation and transcription in single cell is negatively correlated; methylation and transcription can be both positively and negatively correlated in distal regulatory regions. | Angermueller et al., 2016 |
| sc-GEM (genotype single cells genotype, gene expression, DNA methylation) (2016) | Human fibroblast, hiPSC, hESC and NSCLC sample. | Genotype single cells while simultaneously interrogating gene expression and DNA methylation at multiple loci | Cell is captured and lysed on C1 Fluidigm chip. RNA is measured using single cell RT-qPCR and methylation is measured using Single Cell Restriction Analysis of Methylation. | Microfluidic device | Tight coupling between the timing of DNA methylation changes and transcription in individual cells; cells have EGFR mutations show a distinct epigenetic signature. | Cheow et al., 2016 |
| scTrio-seq (single-cell triple omics sequencing) (2017) | HepG2 cell line, mESCs and hepatocellular carcinoma tissue sample. | DNA methylation, CNV (copy number variation), mRNA transcriptome | Cell is lysed. DNA and RNA are separated using centrifuge. mRNA is measured using scRNA-seq, methylation is measured using scRRBS. CNV is computationally inferred from scRRBS coverage. | Mouth pipet | Detected subpopulations of cancer cells according to the large-scale CNV, and detected relationships between CNV, methylation and transcription. | Hou et al., 2016 |
| Simultaneous multiplexed measurement of RNA and proteins in single cells (2016) | Cell culture from glioblastoma multiforme patient sample. | Protein and multiple mRNA | Cells are sorted and lysed. Protein is detected by homogeneous affinity-based proximity extension assay (PEA), and RNA is measured by microfluidic qPCR. | Cell sorting and microfluidic devices | RNA and protein data provide complementary information in defining cell states, and significant heterogeneity in cell culture derived from a glioblastoma multiforme patient. | Darmanis et al., 2016 |

(Continued)

TABLE 1 | Continued

| Methods (time) | Cell type | Measurement | Approach | Single cell isolation | Major discovery | References |
|--|--|---|--|--|--|------------------------|
| scCOOL-seq (single cell Chromatin Overall Omic-scale Landscape Sequencing) (2017) | Mouse preimplantation embryos at different developmental stages. | Chromatin state, DNA methylation, and CNV | Cell is lysed. Chromatin is treated by GpC methyltransferase, and treated DNA is sequenced by scWGBS. <i>In vivo</i> methylation is detected as CpG methylation, and chromatin accessibility is computationally inferred by GpC methylation level. | Mouth pipet | DNA methylation is different between paternal and maternal alleles, but their chromatin accessibility states are similar. | Guo et al., 2017 |
| CITE-seq (cellular indexing of transcriptomes and epitopes by sequencing) (2017) | Cord blood mononuclear cells. | Protein and mRNA transcriptome | mRNA is sequenced using 10X genomics platform. Protein is detected by oligo-labeled antibody, which can be read out during sequencing. | Compatible with 10X genomics, adaptable to other platforms | Multimodal data enable to reveal phenotypes that could not be discovered by using scRNA-seq alone. | Stoeckius et al., 2017 |
| REAP-seq (RNA expression and protein sequencing assay) | human lymphocytes | Protein and mRNA transcriptome | mRNA is sequenced using 10X genomics platform. Protein is detected by oligo-labeled antibody, which can be read out during sequencing. | Flow cytometry | assess the costimulatory effects of a CD27 agonist on human CD8+ lymphocytes and to identify and characterize an unknown cell type | Peterson et al., 2017 |
| scNMT-seq (single-cell nucleosome, methylation and transcription sequencing) (2018) | Mouse embryonic stem cells | Nucleosome status, DNA methylation and mRNA transcription | Similar with scM&T methods, DNA and mRNA were isolated. DNA was cut with GpC methyltransferase M.CviPI before bisulfite treatment. | FACS | Novel links between all three molecular layers and revealing dynamic coupling between epigenomic layers during differentiation | Clark et al., 2018 |
| SIDR-seq simultaneous isolation of genomic DNA and total RNA (SIDR) and sequencing. (2018) | Human lung cancer and breast cancer cells, MCF7, HCC827, and SKBR3 cell lines. | Genome, mRNA transcriptome | Nucleus and cytosol of a single cell were separated by antibody-conjugated magnetic microbeads. mRNA is measured using smart-seq2, gDNA is measured using single-cell whole-genome amplification (Repli-g single cell kit) | Manually diluted to 48-well | copy-number variations positively correlated with the corresponding gene expression levels | Han et al., 2018 |

The second strategy uses oligo-dT primer coated magnetic beads to bind and separate polyadenylated mRNA from DNA (MacAulay et al., 2015; Angermueller et al., 2016). Genome wide sequencing of single cell DNA and RNA purified by this method indicated that breadth of genome coverage and number of genes were not affected by the process of separation, indicating high efficiency in the recovery of DNA and RNA. Since this strategy is adaptable to liquid-handling robots or automated work stations, higher throughput can be achieved. However, coverage of isolated DNA was less evenly distributed across the genome compared to that of the whole single cell sequencing, which may result in less accuracy for copy number analysis of certain genomic regions at a suboptimized sequencing depth.

Besides direct physical isolation of DNA and RNA at the beginning, the third strategy is to preamplify DNA and RNA simultaneously, followed by separation into two parts (Dey et al., 2015). Whole transcriptome sequencing of preamplified RNA of one part showed a similar number of genes covered compared to that of whole single cells. However, as the amplified DNA does not retain methylation states, this method is not suitable for methylome analysis.

The fourth strategy is to split the material of a single cell into two parts directly. For example, a recent report used the splitting strategy to split a single cell into two parts and simultaneously analyze the RNA and protein of the same cell (Darmanis et al., 2016). This splitting strategy is not an ideal method to isolate substrates such as DNA because some material will inevitably be lost due to the uneven split. However, for RNA and protein molecules with high copy number in the single cells, this method is feasible as long as the split is even between the two parts.

INTEGRATION OF GENOME AND TRANSCRIPTOME

The first single cell transcriptome analysis was reported in 2009 (Tang et al., 2009), and many additional single cell RNA sequencing methods have been developed since, such as Quartz-seq (Sasagawa et al., 2013), smart-seq (Switching mechanism at 5' end of the RNA transcript) (Goetz and Trimarchi, 2012; Picelli et al., 2014), Cel-seq (Cell expression by linear amplification and sequencing) (Hashimshony et al., 2012) etc., which were developed using different strategies for different purposes. For example, Quartz-seq detects the 3' end of transcripts, while Smart-seq detects full length transcripts. Cel-seq barcodes and pools samples before linearly amplifying mRNA to multiplex single cell samples. In parallel, due to the development of single-cell whole-genome amplification (WGA) methods, single cell genome sequencing technologies have also been established. At present, four major WGA methods have been reported: DOP (degenerate oligonucleotide-primed polymerase chain reaction) (Telenius et al., 1992), MDA (Multiple Displacement Amplification) (Dean et al., 2001), MALBAC (Multiple Annealing and Looping Based Amplification Cycles) (Zong et al., 2012) and PicoPLEX (Rubicon Genomics

PicoPLEX Kit). In 2013, Han et al. first reported a co-detection of DNA and RNA from the same single cell (Han et al., 2014), which was achieved by physical isolation of cytoplasm (containing cytoplasm RNAs) from nucleus (containing the intact genome) from the same single cells, followed by separate amplification of the transcriptome and genome, and further by respective sequencing of both. Although the initial report showed only the data of the whole transcriptome but not the whole genome, instead of Sanger sequencing of a selected set of genomic sequences, it paved a way to establish multi-omic profiling methods. Later, experimental protocols that simultaneously sequenced the genome and transcriptome were developed by elegantly integrating existing single cell sequencing methods, namely DR-seq (gDNA and mRNA sequencing) (Dey et al., 2015) and G&T-seq (Genome & Transcriptome sequencing) (MacAulay et al., 2015). In DR-seq, a cell is lysed completely, releasing its DNA and RNA into the same reaction system. Genomic DNA and cDNA initially being amplified at the same time is split into two halves: one for RNA-seq using the CEL-seq protocol, and the other half for genome sequencing using MALBAC (Dey et al., 2015). Different from DR-seq, G&T-seq separated poly-A tailed mRNAs from DNA by using oligo-dT-coated magnetic beads. Separated mRNA and DNA were then sequenced using SMART-seq2 and various WGA protocols (MDA or PicoPLEX), respectively (MacAulay et al., 2015). Most recently, Han et al. reported a novel method for simultaneous isolation of genomic DNA and total RNA (SIDR) from single cells by using hypotonic lysis to preserve nuclear lamina integrity and subsequently capturing the cell lysate using antibody-conjugated magnetic microbeads. They found that copy-number variations positively correlated with the corresponding gene expression levels (Han et al., 2018). In summary, using DR-seq, G&T-seq and SIDR, researchers were able to directly determine the correlation between large-scale copy number variation and transcription levels in the CNV regions.

As discussed previously by MacAulay et al. (2017), a substantial advantage of direct measurement of multiple molecular types from the same single cell over separate measurement of each type of molecule from different cells is that genotype-phenotype correlation can be determined unambiguously. First, the genomic variation can be directly linked to the transcriptional variation without being confounded by cell heterogeneity, enabling the dissection of potential molecular mechanisms underlying variable phenotypes among single cells. Second, coupled with lineage record technology, simultaneous sequencing of the genome and transcriptome can be used for reconstruction of lineage trees. Genomic profiling of single cells can divulge the lineage relationship among single cells, based on inherited mutations. The transcriptome profiling of the same single cells can in parallel provide information about the cell's phenotype and function. One intriguing application of this method is to dissect the mechanism of heterogeneity of tumor cells to inform our knowledge of tumor formation and potential therapeutic targets (Shapiro et al., 2013). Third, simultaneous sequencing of DNA and RNA of the same cell can detect DNA mutations with higher accuracy, as the mutations found in DNA or RNA can be verified by each other. This strategy can be

very helpful in situations where highly accurate mutation calling from a single cell is required, such as genetic diagnosis screening during *in vitro* fertilization, when only 1–2 single blastomeres are available (Vermeesch et al., 2016). Of note, post-transcriptional modification such as RNA editing (Tan et al., 2017) which may affect the concordance of variations in both DNA and RNA, should be taken into consideration to precisely call the mutations.

INTEGRATION OF EPIGENOME WITH TRANSCRIPTOME

Based on the development of technologies for single cell epigenome and transcriptome profiling, the methods for the integrated analysis of the epigenome and transcriptome were developed (Angermueller et al., 2016; Hou et al., 2016; Hu et al., 2016). DNA methylation has been demonstrated to have key regulatory functions on gene expression in many biological processes, so the relationship between the DNA methylome and transcriptome from the same single cell is of great interest. Two major methods for single cell methylome analysis are single cell reduced representative bisulfite sequencing (scRRBS) (Guo et al., 2013) and single cell whole genome bisulfite sequencing (scWGBS) (Smallwood et al., 2014). The first reported combined DNA methylome and transcriptome profiling method is scM&T-seq (single cell methylome and transcriptome sequencing), which is developed using the procedure of G&T-seq to isolate DNA and RNA from the same single cell. The protocols for mRNA capture, amplification and sequencing are the same as those in G&T-seq. In parallel, the genomic DNA is subjected to bisulfite treatment and sequencing, allowing the simultaneous profiling of the DNA methylome and RNA transcriptome from the same single cell (Angermueller et al., 2016). Subsequently, scMT-seq (Hu et al., 2016) and scTrio-seq (Hou et al., 2016) were reported using a different strategy to isolate DNA and RNA from a single cell, in which cell membrane but not nucleus was selectively lysed to release RNA, and then intact nucleus was physically separated from the cell lysate (Hou et al., 2016; Hu et al., 2016; Guo et al., 2017). In the scMT-seq method, the single cell nucleus is collected by micropipette and subjected to scRRBS, and mRNA in the lysate is amplified by a modified Smart-seq2 protocol. In the scTrio-seq, the nucleus and cytosol are separated by centrifugation, and genomic DNA contained in the nucleus is sequenced by scRRBS while mRNA is amplified by the scRNA-seq protocol reported by Tang et al. (2009).

The simultaneous profiling of methylome and transcriptome of a single cell provides a unique opportunity to directly measure DNA methylation and gene transcription within the same single cell, and to study the correlation of DNA methylation differences with gene transcription variance across single cells. For example, scM&T-seq investigated the relationship between the transcriptome and DNA methylome, and found that low methylated regions (LMR) showed high variance in methylation level, which is consistent with their role as distal regulatory elements that control gene expression (Angermueller et al., 2016). Our results using scMT-seq found that variable CpG sites were significantly enriched at non-CGI (non-CpG island) promoters

but depleted at CGI (CpG island) promoters, suggesting that non-CGI promoters could be the major region contributing to methylome heterogeneity among dorsal root ganglion single cells. We also found that transcription level was positively correlated with genebody methylation, but negatively correlated with promoter methylation. In addition, by integrating the genomic SNP information, we found a correlation between allelic gene body methylation and allelic expression at single cell level. Thus, scMT-seq allows us to profile genome, DNA methylome and transcriptome in parallel within a single cell (Hu et al., 2016). Similarly, scTrio-seq enables profiling of DNA methylome, genome (CNV) and transcriptome at the same time, in which the copy number variation is computationally inferred from the scRRBS (Hou et al., 2016). Most recently, Guo et al. from the same group reported another single cell multi-omics sequencing method called single-cell COOL-seq that can profile DNA methylation and chromatin state/nucleosome positioning, copy number variation and ploidy simultaneously from the same cell (Guo et al., 2017). Although they did not incorporate the RNA sequencing in this protocol (which is theoretically possible), this method provided new insights into the comprehensive study of genome-wide gene regulation at single cell level. Most recently, Clark et al. reported the scNMT-seq (single-cell nucleosome, methylation, and transcription sequencing), which can simultaneously profile single cell nucleosome, DNA methylation and transcription. By profiling the mouse embryonic stem cell, they found novel links between all three molecular layers and revealed dynamic coupling between epigenomic layers during differentiation (Clark et al., 2018).

PARALLEL PROFILING OF RNA AND PROTEIN

RNA and protein have distinctive biochemical properties. Compared to genomic sequencing methods, the throughput in terms of the number of proteins that can be detected by the single cell proteome profiling is limited. Until now, a few single cell proteomic methods have been developed based on different strategies, including fluorescence-activated cell sorting (FACS), western blot, metal-tagged antibodies followed by mass cytometry, and oligonucleotide labeled antibodies. Although the multiplexing of these approaches were still limited to tens of proteins for a single cell, they still demonstrated the feasibility of detection of protein and RNA expression, paving a way to discover the dynamics of RNA and protein within the same cell. Darmanis et al. developed a method based on homogeneous affinity-based proximity extension assay that converts protein abundance into tag-oligo levels (Darmanis et al., 2016), and both transcript level and protein level were quantified by qPCR. This method has succeeded in capturing parallel profiles of protein and RNA for up to 96 genes (Darmanis et al., 2016). Another approach to simultaneously detect the RNA and protein of the same cell is PLAYR (proximity ligation assay for RNA). Briefly, the RNA transcripts are bound by and ligated to isotope labeled probes. Transcript levels are converted into isotope label levels

that can be easily measured together with elemental isotope-labeled protein using mass cytometry (Frei et al., 2016). With this method, simultaneous quantification of more than 40 different mRNAs and proteins can be achieved, although improvement is required to achieve genome-wide measurement with higher throughput. Most recently, two methods named REAP-seq and CITE-seq with higher throughput have been reported, in which oligonucleotide-labeled antibodies are used to integrate cellular protein and transcriptome measurements into an efficient, single-cell readout (Peterson et al., 2017; Stoeckius et al., 2017). Quantified proteins with 82 barcoded antibodies and more than 20,000 genes can be detected in a single workflow.

STRATEGIES FOR BIOINFORMATICS ANALYSIS OF SINGLE CELL SEQUENCING DATA

Single cell sequencing technologies for genome wide profiling of DNA and RNA, as well as the subsequent integrative computational analysis methods, are central to the interpretation of single cell multi-omics data. The prelude to this type of analysis hinges first on the development of bioinformatics approaches for single cell single-omics sequencing data for various individual types of molecular measurements. Because technical characteristics of various single cell sequencing protocols are different, the bioinformatics methods involved must also be customized to correctly analyze each data type. The need to address the specific characteristics of different single cell sequencing approaches has inspired many computational methods that allow us to better analyze sequencing datasets involving multiple layers.

Single Cell Genome Sequencing

Two major purposes of single-cell genome sequencing are identifying copy number variation and identifying point mutations/SNPs. Both these questions have been addressed in bulk WGS, and the methods developed for bulk WGS data have provided guidance for single cell WGS analysis.

Copy number variation can be robustly identified using Hidden Markov Model (HMM) or Circular Binary Segmentation (CBS), and these methods have proved effective for scWGS data (Knouse et al., 2016). Although these two methods perform similarly in many situations, user-defined parameter adjustments within the algorithms can affect the sensitivity and specificity of copy number calls. For example, comparison of these two methods on scWGS data with a range of parameters indicated that CBS was more sensitive in calling copy number losses, while HMM was more sensitive in calling gains (Knouse et al., 2016). In the context of single cell CNV analysis, one strategy to reconcile the two approaches has been to take the overlap of CNVs identified by CBS and HMM to increase confidence (Knouse et al., 2016). Considerations in choosing between the methods involve the biological properties of the samples, such as the expected sizes of the CNVs, which could range from whole-arm changes seen in aneuploid tumors to dinucleotide changes observed in inherited polymorphisms or in

microsatellite instability. CBS is more flexible than HMM in that the algorithm recursively searches for segmentation points in an unsupervised approach, while HMM depends on the assumption that segmentation points follow a homogenous Poisson process, which is not always the case and may therefore compromise flexibility (Wineinger et al., 2008).

Many tools have been developed for detecting variations in bulk WGS data (Depristo et al., 2011; Koboldt et al., 2012), and these methods, in principle, should perform well in scWGS data. However, scWGS data suffers from high allele coverage bias and high PCR amplification error, which could impair the performance of variant calling methods if not corrected. Recently, with increased understanding of coverage bias in scWGS data (Zhang et al., 2015), Dong et al. reported a computational method that can correct amplification bias to reduce false positive SNPs resulting from PCR or sequencing errors (Dong et al., 2017). Although this new method still partially relies on GATK to identify new variants, it achieved better accuracy by removing false positive variants resulting from PCR error.

Single Cell Transcriptome Sequencing

Single cell RNA-seq data enables the discovery of exciting and new biological phenomena while presenting new challenges for data analysis. For example, single-cell RNA-seq can help us identify cell subtypes with unprecedented resolution, and reconstruct continuous cell lineages. Some early studies showed that identification of cell subtypes or reconstruction of cell lineage could be done manually by experts with sufficient biological prior knowledge using basic statistical methods (Xue et al., 2013; Treutlein et al., 2014). However, recently, huge datasets with extremely heterogeneous cell populations have precluded the feasibility of manual annotation, and many computational pipelines have been developed. For example, tools based on different theoretical frameworks have been developed to cluster cells based on their gene expression similarity, such as SINCERA (Guo et al., 2015), pcaReduce (Žurauskienė and Yau, 2016), SC3 (Kiselev et al., 2017), and SNN-Cliq (Xu and Su, 2015). Additional tools have been developed to reconstruct cell lineage by ordering cells according to computationally inferred pseudo-time (Trapnell et al., 2014; Cannoodt et al., 2016; Qiu et al., 2017). However, despite the availability of myriad computational software packages for clustering and lineage inference, few benchmarking studies have been done to compare their performance.

In addition to those two classical biological questions, the technical problem of imputation of missing values in single-cell RNA-seq data has recently attracted increasing attention. Single-cell RNA-seq, especially for cells captured by droplet-based methods, is often plagued by missing values due to drop-out events, leading to an exceedingly sparse depiction of the single cell transcriptome. Simply removing genes containing missing values restricts the analysis to only highly expressed genes. To overcome this problem, much effort has been made to impute missing values (Kiselev et al., 2017; Lin et al., 2017). These imputation methods can not only enable us to investigate lowly expressed genes but can also improve the performance of

existing computational methods for other purposes by reducing noise from drop-out events.

Single Cell Methylome Analysis

Compared to bulk WGBS (whole genome bisulfite sequencing) data, the analysis of single cell WGBS requires distinct bioinformatics techniques due to the sparse and uneven coverage of scWGBS (single cell WGBS) libraries across the genome. Although many tools have been developed for bulk WGBS data analysis, these methods will fail if applied to scWGBS data directly. To make scWGBS data analysis possible, the first strategy is to merge data from single cells and analyze the merged data as a sample (Farlik et al., 2016). By combining data from many single cells (usually hundreds), the data coverage becomes high, and the bias from allele dropout is averaged out. However, this strategy cannot be used to address the heterogeneity of methylation among different single cells, because methylation data are merged and averaged among the cell population.

Aside from adapting scWGBS data to existing computational pipelines by merging data, the second strategy is to develop new methods specifically for scWGBS data, and many of these methods aim to aggregate methylation levels from adjacent CpG sites or regions with similar biological properties to overcome the sparseness of scWGBS data. For example, Smallwood et al. segment the genome into 5-kbp, non-overlapping bins and use average methylation level among bins as the feature for subsequent analysis (Smallwood et al., 2014). Similarly, by aggregating methylation signal on regulatory elements, we can reveal regulatory mechanisms behind the changes in the DNA methylome (Farlik et al., 2015). In these methods, each single cell is treated as a sample separately, thus enabling the discovery of DNA methylome heterogeneity among single cells.

Interestingly, besides aggregating existing methylation information to reduce noise, a method based on the deep neuronal network was recently developed, which infers missing methylation information from sequencing motifs (Angermueller et al., 2017). Although this method achieved high prediction accuracy for whole genome, its performance on low-methylated regions, the regulatory regions where methylation level influences gene expression greatly, were not satisfying. However, we believe that the prediction accuracy on LMRs can be further improved by incorporating more features into the same deep learning framework.

Single Cell Sequencing for Chromatin Status Analysis

Success in single cell genome and transcriptome sequencing inspired the development of single cell epigenome sequencing. So far, single cell ChIP-seq (Rotem et al., 2015) (Chromatin Immunoprecipitation Sequencing), DNase-seq, and ATAC-seq (Buenrostro et al., 2015) (Assay for Transposase-Accessible Chromatin using sequencing) has been reported from different groups. Since this type of single cell epigenome data has just begun to emerge, the related computational analysis methods are still in their infancy and only a few methods have been developed specifically for single cell data. For example, scChIP-seq and scATAC-seq have been developed to investigate histone

modification and chromatin accessibility landscapes at single cell level (Buenrostro et al., 2015; Rotem et al., 2015; Corces et al., 2016), and the reads from one single cell are extremely sparse due to the low amount of DNA in a cell. To identify the regions that have histone modification or regions with open chromatin, reads from several dozen to hundred single cell libraries were pooled together, and only this “pooled library” has enough reads for conventional peak calling methods. In the subsequent analysis, these putative peaks will be used as guidance to aggregate sparse signal and remove background signal. Although this method enables the meaningful analysis of scChIP-seq and scATAC-seq without requirement of any new computational methods, concerns have been raised about the sensitivity of this strategy (Zamanighomi et al., 2017). Interestingly, methods designed for scATAC-seq analysis are emerging, such as chromVAR (Schep et al., 2017) and scABC (Zamanighomi et al., 2017). We believe these pipelines will also inspire the development of effective pipelines for scChIP-seq data.

APPLICATION OF SINGLE CELL MULTI-OMICS METHODS

As described above, single cell multi-omics analysis integrates multiple data sets from the genome, epigenome, transcriptome, proteome, providing a unique chance to uncover novel biological processes. By extending and integrating methods developed for single-omics analysis, we can obtain a multi-channel molecular readout and utilize these features from multiple omics types to achieve a more comprehensive depiction of the state of a single cell. In combination with continuously advancing bioinformatic algorithms and computational resources, experimental collection of multi-omics data has allowed us to uncover increasingly important and complex insights.

The first application of single cell multi-omics methods is to identify cell subtypes from a heterogeneous cell population. Previously, for example, single cell RNA-seq approaches were shown to be effective in identifying cell subtypes such as human blood dendritic cells, monocytes, and neurons in human brain cortex (MacOsko et al., 2015; Ofengeim et al., 2017; Villani et al., 2017). Recently, single cell DNA methylation sequencing was also applied to study human brain cortex. By examining non-CpG methylation among single cells, they identified novel cell subtypes that were masked in scRNA-seq analysis (Luo et al., 2017). Epigenetic modifications such as DNA methylation are developmentally regulated and cell type-specific, yet stable over the life span, and therefore profiling the epigenome and transcriptome simultaneously can compensate for the limitation of single cell RNA-seq, which mainly yields information about highly expressed transcripts. Thus, different omics measurements can provide non-redundant information about cell identity and enable more detailed and more accurate dissection of complicated tissues.

Second, single cell multi-omics can be used to reconstruct cell lineage trajectories. Understanding cell lineage trajectories during the complete time course of multicellular animal development is the holy grail of developmental biology. DNA

mutations, as well as epigenetic modifications gained during the cell division and passed to the daughter cells, can be used for lineage tracing, while the transcriptome of the matching single cells can reveal the concomitant alteration of gene expression and transcriptional cell fate change during cell proliferation and differentiation. For example, cancer cells have extremely unstable genomes, and understanding cancer genome evolution is crucial for revealing “driver” mutations or copy number changes that cause carcinogenesis. Single cell multi-omics can not only help us determine the occurrence order of different mutations during cancer evolution, but can also reveal their functional consequences, such as alteration in gene expression, which will eventually help us identify the causal mutations that induce the transition from normal cell to cancer cell.

Lastly but most importantly, single cell multi-omics data provides the resolution to definitively reveal the relationship between different omics readouts. Correlation analysis between different omics is a prevailing approach to generate regulatory hypotheses between two omics data types. For example, cytosine methylation is among the best-studied epigenetic modifications and has been shown to regulate many critical biological processes. With both DNA and RNA sequencing data, DR-seq and G&T-seq have allowed us the ability to reveal correlation between copy number variation and gene expression level at a single cell scale. Further, scTrio-seq showed that large-scale CNVs caused proportional changes in RNA expression of genes within the gained or lost genomic regions, whereas these CNVs generally do not affect DNA methylation in these regions. Our work using scMT-seq not only showed allele-specific expression patterns based on SNV information, but also showed correlation of DNA methylation with allele-specific expression, providing new insight into the study of imprinting and its underlying mechanism. In the near future, multi-omics methods may be helpful for understanding the correlation between DNA mutations with epigenetic modifications and their effects on gene expression to reveal the mechanisms underlying interesting biological questions such as dosage compensation and X-inactivation, among others (Livernois et al., 2012; Graves, 2016). Inevitably, even with single cell multi-omics technology, we are still limited to identifying correlation but not causality. We therefore believe that single cell multi-omics, once combined together with experimental perturbation, will be effective in allowing us to understand causal relationships among omics data types.

Essential to all these applications is the development of computational approaches that help to integrate multiple data layers and to recover information lost due to the sequencing of minute amounts of biological material. Bioinformatic and computational techniques have advanced single cell multi-omics technology in several arenas, such as (1) imputation of “dropped-out” single cell measurements, (2) indirect measurement of another omics layer from a measured one (Farlik et al., 2015; Bock et al., 2016), and (3) mathematical and statistical quantification of multi-dimensional associations (Lane et al., 2017). Imputation methods pull information from groups of similar cells to help to restore measurements for molecules originally in very low abundance, such as lowly expressed

RNA transcripts, filling in sparse data matrices for better representations of the original relationships (Van Dijk et al., 2017; Li and Li, 2018). Furthermore, as our knowledge of biological regulatory relationships increases, one data type may be able to serve as proxy for inference of another omics layer. For example, transcription factor binding or copy number alterations have been indirectly inferred from single cell methylation data (Farlik et al., 2015; Hou et al., 2016). Likewise, copy number information can be inferred from the single cell transcriptome (Tirosh et al., 2016), and chromatin state from the methylome (Guo et al., 2017). In addition, as single cell multi-omics technology becomes progressively high throughput, computational resources and time needed for processing of the raw data will be an important aspect in the flexibility of data analysis. Pipelines and new algorithms that streamline and shorten the computational time needed for data processing will be important for increasingly complex, multi-dimensional experiments. Raw files for each omics type must be separately processed, aligned, filtered, and quality-controlled in a manner that accounts for complications inherent in single cell measurements, such as low signal-to-noise ratio, technical amplification artifacts, and technical variation (Bock et al., 2016). Each omics layer of processed data is then assigned back to the single cell and co-analyzed with both mathematical and statistical models to reveal patterns of regulation. These new computational methods, while still nascent, allow us the capacity to bypass experimental limitations and expose excitingly novel relationships.

CONCLUSIONS AND FUTURE DIRECTIONS

Single cell multi-omics methods have provided countless opportunities to systematically understand biological diversity, and to identify rare cell types and their characteristics with unprecedented accuracy through integration of information from multiple omics levels, including DNA, RNA, and protein. These single cell multi-omics methods will play an important role in many diverse fields, and their applications are rapidly expanding, including (1) delineating cellular diversity, (2) lineage tracing, (3) identifying new cell types, and (4) deciphering the regulatory mechanisms between omics. Although some of the applications have been reported in initial studies, there are still many avenues open for exploration, and the further development of new multi-omics methods will also facilitate their increasing utility. It is anticipated that better performance of multi-omics methods will be generated based on the optimization of current single cell sequencing methods. There are currently several main challenges and thus opportunities for further development of single cell multi-omics technology: (1) Overcoming the limitations of current single cell sequencing methods will facilitate the development of more types of omics measurements on single cells. For example, outside of single cell DNA methylome analysis, there are other single cell epigenome sequencing methods such as scAba-seq (DNA hydroxymethylation) (Mooijman et al., 2016), single cell ATAC-seq (open chromatin) (Buenrostro et al., 2015), single cell Hi-C

(chromatin conformation) (Nagano et al., 2013), and single cell ChIP-seq (histone modifications) (Rotem et al., 2015). However, due to limitations such as low genome coverage and high noise signals derived from locus dropout and PCR amplification, no reliable multi-omics approach based on these methods has been reported yet. Optimization of the existing single cell sequencing methods as well as newly developed methods will provide more opportunities to integrate diverse methods with transcriptomic analysis to reveal the relationship between epigenetic states and RNA transcription variation. (2) New approaches to isolate and label multiple types of molecules of the same single cell will help to increase the number of omics profiled in parallel, from dual-omics to triple-omics or more. Even multiple functional parameters of single cells could be included, such as with the development of patch-seq, which combined whole-cell electrophysiological patch-clamp recordings, single-cell RNA-sequencing, and morphological characterization to identify new cell types in the nervous system (Cadwell et al., 2016, 2017). (3) In contrast to the rich resources of experimental protocols, computational methods

for single cell multi-omics data analysis have just started to emerge. New computational approaches tailored to the analysis of single cell multi-omics data will also substantially facilitate the application of the methods (Yan et al., 2017). In summary, with further development of multi-omics methods, the future will witness an even wider application of single cell multi-omics technology that will result in meaningful findings never before achieved.

AUTHOR CONTRIBUTIONS

YH and YG: Conceived the structure of the manuscript; YH, YG, and QA: Wrote the manuscript; QA, BT, KS, and SF: Read and edited the manuscript.

ACKNOWLEDGMENTS

The work was supported by National Key R&D Program of China (2017YFA0104100, 2017YFC1001300), National Natural Science Foundation of China (31700900).

REFERENCES

- Angermueller, C., Clark, S. J., Lee, H. J., MacAulay, I. C., Teng, M. J., Hu, T. X., et al. (2016). Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nat. Methods* 13, 229–232. doi: 10.1038/nmeth.3728
- Angermueller, C., Lee, H. J., Reik, W., and Stegle, O. (2017). DeepCpG: accurate prediction of single-cell DNA methylation states using deep learning. *Genome Biol.* 18:67. doi: 10.1186/s13059-017-1189-z
- Autengruber, A., Gereke, M., Hansen, G., Hennig, C., and Bruder, D. (2012). Impact of enzymatic tissue disintegration on the level of surface molecule expression and immune cell function. *Eur. J. Microbiol. Immunol.* 2, 112–120. doi: 10.1556/EuJMI.2.2012.2.3
- Bock, C., Farlik, M., and Sheffield, N. C. (2016). Multi-omics of single cells: strategies and applications. *Trends Biotechnol.* 34, 605–608. doi: 10.1016/j.tibtech.2016.04.004
- Buenrostro, J. D., Wu, B., Litzenburger, U. M., Ruff, D., Gonzales, M. L., Snyder, M. P., et al. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* 523, 486–490. doi: 10.1038/nature14590
- Cadwell, C. R., Palasantza, A., Jiang, X., Berens, P., Deng, Q., Yilmaz, M., et al. (2016). Electrophysiological, transcriptomic and morphologic profiling of single neurons using Patch-seq. *Nat. Biotechnol.* 34, 199–203. doi: 10.1038/nbt.3445
- Cadwell, C. R., Scala, F., Li, S., Livrizzi, G., Shen, S., Sandberg, R., et al. (2017). Multimodal profiling of single-cell morphology, electrophysiology, and gene expression using Patch-seq. *Nat. Protoc.* 12, 2531–2553. doi: 10.1038/nprot.2017.120
- Cannoodt, R., Saelens, W., and Saeys, Y. (2016). Computational methods for trajectory inference from single-cell transcriptomics. *Eur. J. Immunol.* 46, 2496–2506. doi: 10.1002/eji.201646347
- Cheow, L. F., Courtois, E. T., Tan, Y., Viswanathan, R., Xing, Q., Tan, R. Z., et al. (2016). Single-cell multimodal profiling reveals cellular epigenetic heterogeneity. *Nat. Methods* 13, 833–836. doi: 10.1038/nmeth.3961
- Clark, S. J., Argelaguet, R., Kapourani, C. A., Stubbs, T. M., Lee, H. J., Alda-Catalinas, C., et al. (2018). scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. *Nat. Commun.* 9:781. doi: 10.1038/s41467-018-03149-4
- Corces, M. R., Buenrostro, J. D., Wu, B., Greenside, P. G., Chan, S. M., Koenig, J. L., et al. (2016). Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat. Genet.* 48, 1193–1203. doi: 10.1038/ng.3646
- Crick, F. (1970). Central dogma of molecular biology. *Nature* 227, 561–563. doi: 10.1038/227561a0
- Darmanis, S., Gallant, C. J., Marinescu, V. D., Niklasson, M., Segerman, A., Flamourakis, G., et al. (2016). Simultaneous multiplexed measurement of RNA and proteins in single cells. *Cell Rep.* 14, 380–389. doi: 10.1016/j.celrep.2015.12.021
- Dean, F. B., Nelson, J. R., Giesler, T. L., and Lasken, R. S. (2001). Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Res.* 11, 1095–1099. doi: 10.1101/gr.180501
- DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498. doi: 10.1038/ng.806
- Dey, S. S., Kester, L., Spanjaard, B., Bienko, M., and van Oudenaarden, A. (2015). Integrated genome and transcriptome sequencing of the same cell. *Nat. Biotechnol.* 33, 285–289. doi: 10.1038/nbt.3129
- Dong, X., Zhang, L., Milholland, B., Lee, M., Maslov, A. Y., Wang, T., et al. (2017). Accurate identification of single-nucleotide variants in whole-genome-amplified single cells. *Nat. Methods* 14, 491–493. doi: 10.1038/nmeth.4227
- Farlik, M., Halbritter, F., Müller, F., Choudry, F. A., Ebert, P., Klughammer, J., et al. (2016). DNA methylation dynamics of human hematopoietic stem cell differentiation. *Cell Stem Cell* 19, 808–822. doi: 10.1016/j.stem.2016.10.019
- Farlik, M., Sheffield, N. C., Nuzzo, A., Datlinger, P., Schönegger, A., Klughammer, J., et al. (2015). Single-cell DNA methylome sequencing and bioinformatic inference of epigenomic cell-state dynamics. *Cell Rep.* 10, 1386–1397. doi: 10.1016/j.celrep.2015.02.001
- Frei, A. P., Bava, F. A., Zunder, E. R., Hsieh, E. W., Chen, S. Y., Nolan, G. P., et al. (2016). Highly multiplexed simultaneous detection of RNAs and proteins in single cells. *Nat. Methods* 13, 269–275. doi: 10.1038/nmeth.3742
- Gawad, C., Koh, W., and Quake, S. R. (2016). Single-cell genome sequencing: current state of the science. *Nat. Rev. Genet.* 17, 175–188. doi: 10.1038/nrg.2015.16
- Goetz, J. J., and Trimarchi, J. M. (2012). Transcriptome sequencing of single cells with Smart-Seq. *Nat. Biotechnol.* 30, 763–765. doi: 10.1038/nbt.2325
- Graves, J. A. (2016). Evolution of vertebrate sex chromosomes and dosage compensation. *Nat. Rev. Genet.* 17, 33–46. doi: 10.1038/nrg.2015.2
- Guo, F., Li, L., Li, J., Wu, X., Hu, B., Zhu, P., et al. (2017). Single-cell multi-omics sequencing of mouse early embryos and embryonic stem cells. *Cell Res.* 27, 967–988. doi: 10.1038/cr.2017.82
- Guo, H., Zhu, P., Wu, X., Li, X., Wen, L., and Tang, F. (2013). Single-cell methylome landscapes of mouse embryonic stem cells and early embryos

- analyzed using reduced representation bisulfite sequencing. *Genome Res.* 23, 2126–2135. doi: 10.1101/gr.161679.113
- Guo, M., Wang, H., Potter, S. S., Whittsett, J. A., and Xu, Y. (2015). SINCERA: a pipeline for single-cell rna-seq profiling analysis. *PLoS Comput. Biol.* 11:e1004575. doi: 10.1371/journal.pcbi.1004575
- Han, K. Y., Kim, K. T., Joung, J. G., Son, D. S., Kim, Y. J., Jo, A., et al. (2018). SIDR: simultaneous isolation and parallel sequencing of genomic DNA and total RNA from single cells. *Genome Res.* 28, 75–87. doi: 10.1101/gr.223263.117
- Han, L., Zi, X., Garmire, L. X., Wu, Y., Weissman, S. M., Pan, X., et al. (2014). Co-detection and sequencing of genes and transcripts from the same single cells facilitated by a microfluidics platform. *Sci. Rep.* 4:6485. doi: 10.1038/srep06485
- Hashimshony, T., Wagner, F., Sher, N., and Yanai, I. (2012). CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep.* 2, 666–673. doi: 10.1016/j.celrep.2012.08.003
- Hou, Y., Guo, H., Cao, C., Li, X., Hu, B., Zhu, P., et al. (2016). Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. *Cell Res.* 26, 304–319. doi: 10.1038/cr.2016.23
- Hu, Y., Huang, K., An, Q., Du, G., Hu, G., Xue, J., et al. (2016). Simultaneous profiling of transcriptome and DNA methylome from a single cell. *Genome Biol.* 17:88. doi: 10.1186/s13059-016-0950-z
- Huang, L., Ma, F., Chapman, A., Lu, S., and Xie, X. S. (2015). Single-cell whole-genome amplification and sequencing: methodology and applications. *Annu. Rev. Genomics Hum. Genet.* 16, 79–102. doi: 10.1146/annurev-genom-090413-025352
- Kiselev, V. Y., Kirschner, K., Schaub, M. T., Andrews, T., Yiu, A., Chandra, T., et al. (2017). SC3: consensus clustering of single-cell RNA-seq data. *Nat. Methods* 14, 483–486. doi: 10.1038/nmeth.4236
- Knouse, K. A., Wu, J., and Amon, A. (2016). Assessment of megabase-scale somatic copy number variation using single-cell sequencing. *Genome Res.* 26, 376–384. doi: 10.1101/gr.198937.115
- Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., McLellan, M. D., Lin, L., et al. (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* 22, 568–576. doi: 10.1101/gr.129684.111
- Lane, K., Van Valen, D., Defelice, M. M., MacKlin, D. N., Kudo, T., Jaimovich, A., et al. (2017). Measuring signaling and RNA-Seq in the same cell links gene expression to dynamic patterns of NF-kappaB activation. *Cell Syst.* 4, 458.e455–469.e455. doi: 10.1016/j.cels.2017.03.010
- Li, W. V., and Li, J. J. (2018). An accurate and robust imputation method scImpute for single-cell RNA-seq data. *Nat. Commun.* 9:997. doi: 10.1038/s41467-018-03405-7
- Lin, P., Troup, M., and Ho, J. W. (2017). CIDR: ultrafast and accurate clustering through imputation for single-cell RNA-seq data. *Genome Biol.* 18:59. doi: 10.1186/s13059-017-1188-0
- Livernois, A. M., Graves, J. A., and Waters, P. D. (2012). The origin and evolution of vertebrate sex chromosomes and dosage compensation. *Heredity* 108, 50–58. doi: 10.1038/hdy.2011.106
- Luo, C., Keown, C. L., Kurihara, L., Zhou, J., He, Y., Li, J., et al. (2017). Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. *Science* 357, 600–604. doi: 10.1126/science.aan3351
- MacAulay, I. C., Haerty, W., Kumar, P., Li, Y. I., Hu, T. X., Teng, M. J., et al. (2015). G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. *Nat. Methods* 12, 519–522. doi: 10.1038/nmeth.3370
- MacAulay, I. C., Ponting, C. P., and Voet, T. (2017). Single-Cell multiomics: multiple measurements from single cells. *Trends Genet.* 33, 155–168. doi: 10.1016/j.tig.2016.12.003
- MacOsko, E. Z., Basu, A., Satija, R., Nemes, J., Shekhar, K., Goldman, M., et al. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161, 1202–1214. doi: 10.1016/j.cell.2015.05.002
- Mooijman, D., Dey, S. S., Boisset, J. C., Crosetto, N., and van Oudenaarden, A. (2016). Single-cell 5hmC sequencing reveals chromosome-wide cell-to-cell variability and enables lineage reconstruction. *Nat. Biotechnol.* 34, 852–856. doi: 10.1038/nbt.3598
- Nagano, T., Lubling, Y., Stevens, T. J., Schoenfelder, S., Yaffe, E., Dean, W., et al. (2013). Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* 502, 59–64. doi: 10.1038/nature12593
- Navin, N. E. (2015). The first five years of single-cell cancer genomics and beyond. *Genome Res.* 25, 1499–1507. doi: 10.1101/gr.191098.115
- Navin, N., Kendall, J., Troge, J., Andrews, P., Rodgers, L., McIndoo, J., et al. (2011). Tumour evolution inferred by single-cell sequencing. *Nature* 472, 90–94. doi: 10.1038/nature09807
- Ofengeim, D., Giagtzoglou, N., Huh, D., Zou, C., and Yuan, J. (2017). Single-Cell RNA sequencing: unraveling the brain one cell at a time. *Trends Mol. Med.* 23, 563–576. doi: 10.1016/j.molmed.2017.04.006
- Ortega, M. A., Poirion, O., Zhu, X., Huang, S., Wolfgruber, T. K., Sebra, R., et al. (2017). Using single-cell multiple omics approaches to resolve tumor heterogeneity. *Clin. Transl. Med.* 6:46. doi: 10.1186/s40169-017-0177-y
- Peterson, V. M., Zhang, K. X., Kumar, N., Wong, J., Li, L., Wilson, D. C., et al. (2017). Multiplexed quantification of proteins and transcripts in single cells. *Nat. Biotechnol.* 35, 936–939. doi: 10.1038/nbt.3973
- Picelli, S., Faridani, O. R., Björklund, A. K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* 9, 171–181. doi: 10.1038/nprot.2014.006
- Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H. A., et al. (2017). Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* 14, 979–982. doi: 10.1038/nmeth.4402
- Rotem, A., Ram, O., Shores, N., Sperling, R. A., Goren, A., Weitz, D. A., et al. (2015). Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nat. Biotechnol.* 33, 1165–1172. doi: 10.1038/nbt.3383
- Sasagawa, Y., Nikaïdo, I., Hayashi, T., Danno, H., Uno, K. D., Imai, T., et al. (2013). Quartz-Seq: a highly reproducible and sensitive single-cell RNA sequencing method, reveals non-genetic gene-expression heterogeneity. *Genome Biol.* 14:R31. doi: 10.1186/gb-2013-14-4-r31
- Schep, A. N., Wu, B., Buenrostro, J. D., and Greenleaf, W. J. (2017). chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* 14, 975–978. doi: 10.1038/nmeth.4401
- Shapiro, E., Biezuner, T., and Linnarsson, S. (2013). Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat. Rev. Genet.* 14, 618–630. doi: 10.1038/nrg3542
- Smallwood, S. A., Lee, H. J., Angermueller, C., Krueger, F., Saadeh, H., Peat, J., et al. (2014). Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat. Methods* 11, 817–820. doi: 10.1038/nmeth.3035
- Stoeckli, M., Hafemeister, C., Stephenson, W., Houck-Loomis, B., Chattopadhyay, P. K., Swerdlow, H., et al. (2017). Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* 14, 865–868. doi: 10.1038/nmeth.4380
- Svensson, V., Natarajan, K. N., Ly, L. H., Miragaia, R. J., Labalette, C., MacAulay, I. C., et al. (2017). Power analysis of single-cell RNA-sequencing experiments. *Nat. Methods* 14, 381–387. doi: 10.1038/nmeth.4220
- Tan, M. H., Li, Q., Shanmugam, R., Piskol, R., Kohler, J., Young, A. N., et al. (2017). Dynamic landscape and regulation of RNA editing in mammals. *Nature* 550, 249–254. doi: 10.1038/nature24041
- Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., et al. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* 6, 377–382. doi: 10.1038/nmeth.1315
- Telenius, H., Carter, N. P., Bebb, C. E., Nordenskjöld, M., Ponder, B. A., and Tunnacliffe, A. (1992). Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. *Genomics* 13, 718–725. doi: 10.1016/0888-7543(92)90147-K
- Tirosh, I., Izar, B., Prakadan, S. M., Wadsworth, M. H. II., Treacy, D., Trombetta, J. J., et al. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* 352, 189–196. doi: 10.1126/science.aad0501
- Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., et al. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* 32, 381–386. doi: 10.1038/nbt.2859
- Treutlein, B., Brownfield, D. G., Wu, A. R., Neff, N. F., Mantalas, G. L., Espinoza, F. H., et al. (2014). Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* 509, 371–375. doi: 10.1038/nature13173
- Van Dijk, D., Nainys, J., Sharma, R., Kathail, P., Carr, A. J., Moon, K. R., et al. (2017). MAGIC: a diffusion-based imputation method reveals gene-gene interactions in single-cell RNA-sequencing data. *bioRxiv*. doi: 10.1101/111591
- Vermeesch, J. R., Voet, T., and Devriendt, K. (2016). Prenatal and pre-implantation genetic diagnosis. *Nat. Rev. Genet.* 17, 643–656. doi: 10.1038/nrg.2016.97

- Villani, A. C., Satija, R., Reynolds, G., Sarkizova, S., Shekhar, K., Fletcher, J., et al. (2017). Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* 356:eaah4573. doi: 10.1126/science.aah4573
- Volovitz, I., Shapira, N., Ezer, H., Gafni, A., Lustgarten, M., Alter, T., et al. (2016). A non-aggressive, highly efficient, enzymatic method for dissociation of human brain-tumors and brain-tissues to viable single-cells. *BMC Neurosci.* 17:30. doi: 10.1186/s12868-016-0262-y
- Wang, Y., and Navin, N. E. (2015). Advances and applications of single-cell sequencing technologies. *Mol. Cell* 58, 598–609. doi: 10.1016/j.molcel.2015.05.005
- Wineinger, N. E., Kennedy, R. E., Erickson, S. W., Wojczynski, M. K., Bruder, C. E., and Tiwari, H. K. (2008). Statistical issues in the analysis of DNA copy number variations. *Int. J. Comput. Biol. Drug Des.* 1, 368–395. doi: 10.1504/IJCBDD.2008.022208
- Xu, C., and Su, Z. (2015). Identification of cell types from single-cell transcriptomes using a novel clustering method. *Bioinformatics* 31, 1974–1980. doi: 10.1093/bioinformatics/btv088
- Xue, Z., Huang, K., Cai, C., Cai, L., Jiang, C. Y., Feng, Y., et al. (2013). Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature* 500, 593–597. doi: 10.1038/nature12364
- Yan, J., Risacher, S. L., Shen, L., and Saykin, A. J. (2017). Network approaches to systems biology analysis of complex disease: integrative methods for multi-omics data. *Brief. Bioinform.* 2017, 1–12. doi: 10.1093/bib/bbx066
- Zamanighomi, M., Lin, Z., Daley, T., Schep, A., Greenleaf, W. J., and Wong, W. H. (2017). Unsupervised clustering and epigenetic classification of single cells. *bioRxiv*. doi: 10.1101/143701
- Zhang, C. Z., Adalsteinsson, V. A., Francis, J., Cornils, H., Jung, J., Maire, C., et al. (2015). Calibrating genomic and allelic coverage bias in single-cell sequencing. *Nat. Commun.* 6:6822. doi: 10.1038/ncomms7822
- Zong, C., Lu, S., Chapman, A. R., and Xie, X. S. (2012). Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science* 338, 1622–1626. doi: 10.1126/science.1229164
- Žurauskiene, J., and Yau, C. (2016). pcaReduce: hierarchical clustering of single cell transcriptional profiles. *BMC Bioinformatics* 17:140. doi: 10.1186/s12859-016-0984-y

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

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