Application of multi-omics in single cells

Xiaolong Kang1,2; Andrew Liu3; George E Liu1*
1Animal Genomics and Improvement Lab, USDA, USA
2Department of Agriculture, Ningxia University, China
3Centennial High School, Ellicott City, MD 21042, USA

*Corresponding Author(s): George E Liu,
Animal Genomics and Improvement Lab, BARC, NEA,
USDA-ARS, USDA, Building 306, BARC-EAST, Beltsville,
MD 20705, USA
Tel: 301-504-6936, Fax: 301-504-8414
Email: George.Liu@ARS.USDA.GOV

Received: Mar 05, 2018
Accepted: Apr 16, 2018
Published Online: Apr 23, 2018
Journal: Annals of Biotechnology
Publisher: MedDocs Publishers LLC
Online edition: http://meddocsonline.org/
Copyright: © Kang X et al (2018). This Article is distributed under the terms of Creative Commons Attribution 4.0 international License

Keywords: Single cell sequencing; Amplification; Genomics; Transcriptomics; Epigenomics; Metabolomics; Proteomics

Abbreviations: WGA: Whole-Genome Amplification; DOP-PCR: Degenerate Oligonucleotide-Primed Polymerase Chain Reaction; MDA: Multiple Displacement Amplification; MALBAC: Multiple Annealing and Looping-Based Amplification Cycle; LIANTI: Linear Amplification Via Transposon Insertion; SC: Single cell; CNV: Copy Number Variant; EMT: Epithelial-to-Mesenchymal Transition; CTCs: Circulating Tumor Cells

Abstract

In recent years, single cell assays have made exciting progresses, overcoming the issue of heterogeneity associated with bulk populations. The fast-developing sequencing methods now enable unbiased, high-throughput and high-resolution view of the heterogeneity from individual cell within a population, in terms of its fate decisions, identity and function. The cell’s state is regulated at different levels, such as DNA, RNA and protein, by complex interplay of intrinsic molecules existing in the organism and extrinsic stimuli such as local environment. Comprehensive profiling of single cell requires a simultaneously dissection from different levels (multi-omics) to avoid incomplete information generated from single cell. In this short review, we first examine the whole genome amplification methods, and then survey the features of the single cell approaches for genome, epigenome, transcriptome, proteome and metabolome profiling. Finally, we briefly analyze advantages of multi-omics measurement from single cells as compared to separate measurement of each molecular type, and discuss opportunities and challenges of combining single cell multi-omics information on resolving phenotype variants.

Introduction

The cell is the basic unit of life, whose phenotypes can vary in response to genotypes and environmental influences. Because the remarkable cell-to-cell heterogeneity exists in single cells, individual cells need to be characterized owing to their stochastic changes or uniqueness [1, 2]. By detecting the behavior and heterogeneity of the individual cells, we could shed lights into the complex biological mechanisms underlying different phenotype variants, such as a developing embryo or a tumor.

The first step of analyzing a single cell is to isolate and capture single cells from bulk populations. Numerous approaches have been developed: mouth pipetting, serial dilution, robotic micromanipulation, laser-capture-micro dissection, flow-assisted cell sorting, and micro fluidic platforms, etc [3]. The advantages and limitations of these methods have been reviewed before [4-6]. In this review, we first examine the whole genome amplification methods and survey each of omics approaches for single cells. We will then briefly discuss the prospects of combining single cell multi-omics information on resolving phenotype variants.

### Whole genome amplification methods from single cells

Obtaining enough amount molecules, including DNA or RNA, from single cell is a great challenge. For example, for single cell sequencing, limited amount of DNA or cDNA molecules need to be amplified with higher fidelity and less bias. Several Whole-Genome Amplification (WGA) methods have been used to obtain sufficient DNA for sequencing (Table 1). Here we quickly summarize a few of these methods and their features.

#### DOP-PCR:

The Degenerate Oligonucleotide-Primed Polymerase Chain Reaction (DOP-PCR), was widely used to amplify genome from single cell in earlier years [7]. DOP has important applications in genome mapping and can be used to identify the origin of markers, measure CNVs, and map translocation breakpoints on a large genomic scale [6,7,19]. Because of the exponential amplification nature of PCR, DOP-PCR has low genome coverage, high amplification biases [10], and high drop-out rate [19-21].

#### MDA:

Multiple Displacement Amplification (MDA), is another common method of DNA amplification in single-cell whole-genome analyses. Using random primers and Phi29 DNA polymerase, circular DNA templates can be amplified 10,000-fold in a few hours [10]. Although offering much higher genome coverage than DOP-PCR, MDA gives rise to chimeric reads and introduces huge amplification bias because of its exponential amplification process [22,23]. Furthermore, such sequence-dependent bias of MDA is not reproducible along the genome from cell to cell.

#### MALBAC:

By incorporating quasi-linear amplification through looping-based amplicon protection into PCR, the Multiple Annealing And Looping-Based Amplification Cycle (MALBAC) method reduces the sequence-dependent bias introduced by exponential amplification [15]. The primers in the initial reaction of MALBAC are designed to share common sequences that form loops and inhibit the repeated (potentially biased) priming from their ends. MALBAC offers high uniformity across the genome. Sequencing DNA amplified with MALBAC can achieve 93% genome coverage ≥1x for a single human cell at 25x mean sequencing depth [16].

#### LIANTI:

To further reduce amplification bias and errors, a new method, Linear Amplification via Transposon Insertion (LIANTI), which combines Tn5 transposition and T7 in vitro transcription for single-cell genomic analyses has been recently developed [17]. During LIANTI, Tn5 transposition first randomly fragments and inserts T7 promoter sequence into genomic DNA. T7 RNA polymerase is then used to generate amplified antisense RNA. After reverse transcription and second strand synthesis, double-stranded LIANTI amplicons are ready for DNA library preparation and high throughput sequencing. Therefore, by replacing PCR with in vitro transcription, LIANTI effectively decreases PCR’s errors and biases induced by nonspecific priming and exponential amplification

#### The ‘omics’ application based on single cell

In the rest parts, we will survey the profiling methods for each molecule type and then briefly discuss the opportunities and challenges offered by measuring them simultaneously at cellular resolution.

### Single cell genomics:

Single cell genome sequencing was used to characterize mutations, structural variations, aneuploidies, and recombination in the genome [14,24]. It has also been used to study the diversity, evolution and role of genetic mosaicism [22,25]. Single cell genome sequencing is crucial for revealing genetic heterogeneity and cell-lineage relationships in normal and diseased tissues [26-28]. As a precise evaluation of prognosis is important in creating an effective treatment strategy for cancers, single-cell technology has allowed many new prognostic factors to be detected and confirmed. For example, it was applied to identify and trace the origin of disseminated tumor cells in breast cancer [29]. In prostate cancer, single cell sequencing analysis has been applied to show that loss of PTEN can predict poor prognosis [30]. Therefore, single-cell technology can provide prognosis more accurately than before [31].

### Single cell transcriptomics:

Temporal and spatial changes in gene transcription drive the development of organism. Single-cell RNA-seq (scRNA-seq) was first reported in 2009 for analyzing the mouse blastomere transcriptome at a single-cell resolution [1]. It can be used for determining gene regulatory networks at whole genome scale in an objective and unbiased way. When combined with over expression, knockout or knockdown of a gene of interest, scRNA-seq can reveal the gene expression network in target cells [1,32,33]. It also has the potential to provide transcriptomic information from intratumoral cells and to identify the subpopulations within a tumor, and to detect putative cancer stem cells. scRNA-seq is regarded as a promising way to improve diagnosis and prognosis, and provide more precise target therapy [34,35]. Although the great potential in detecting heterogeneity between cells from same individual, limitations also exist in single cell transcriptome. For instance, spike-ins are needed as unique molecular identifiers to allow accurate normalization and quality control of the raw data [36]. Reverse transcriptase and subsequent polymerase-based amplification steps often have prone to introduce biases in representation in the data. In scRNA-seq, it is estimated that only 10–40% of the original mRNA molecules from a cell are represented in the final sequencing library [37,38], suggesting that there is still a long way to improving the accuracy of amplification and mRNA library construction.

### Single cell epigenetics:

Epigenomic mechanisms are central to the regulation of gene expression and study of the epigenomes of single cells is essential to understanding cellular identity, cellular function and phenotypes that are not predictable by genotype alone. Epigenetic alterations as a marker for early diagnosis may also become new targets for cancer prevention and treatment [39,40]. For example, the Epithelial-to-Mesenchymal Transition (EMT) is a...
key mechanism enabling epithelial tumor cells to disseminate and metastasize. Pixberg et al. established an assay to simultaneously analyze promoter methylation of three EMT-associated genes (miR-200c/141, miR-200b/a/249 and CDH1) in single cells through a protocol of agarose embedded bisulfite treatment [41]. Their results showed methylation at the promoter of microRNA-200 family was significantly higher in prostate circulating tumor cells (CTCs). These data also revealed an epigenetic heterogeneity among CTCs and indicated tumor-specific active epigenetic regulation of EMT-associated genes during blood-borne dissemination. In another study, Litzenburger et al. identified the cell surface marker CD24 as co-varying with chromatin accessibility changes linked to transcription factor GATA in single cells by using single-cell chromatin accessibility and RNA-seq data in K562 leukemic cells [42]. Their results showed that GATA/CD24+ cells have the capability to rapidly reconstitute the heterogeneity within the entire starting population, suggesting that single-cell chromatin accessibility can guide prospective characterization of cancer heterogeneity. Moreover, studies of genome-wide hydroxymethylation [43], chromatin conformation [44] and DNA adenine methyltransferase identification [45], also provide new insights on how the epigenomics has impact on the gene expression at a single-cell resolution, which will be helpful to detecting the phenotype variants or heterogeneity of cancer cell. Recently, scRRBS and sc-WGBS or scBS-seq (for DNA methylation), scChiP-seq (for transcription factor occupancy and histone codes), scDNAse-seq and scATAC-seq (for chromatin state), scHIC (for chromosome conformation capture), and others are emerging to enable single cell epigenomics studies [46].

Single cell proteomics:

A cell’s proteome ties genotype to phenotype by defining its response to the various internal and external stimuli. For example, tumor suppressor protein p53 is crucial in many cancers. High resolution single-cell analyses revealed that the results of the bulk cell studies failed to uncover p53’s true dynamic response [47]. Instead of decreased magnitude, individual cells display series of equal p53 pulses with fixed amplitude and duration, independent of the intensity of external stimuli. The misleading average results from the bulk cell studies are related to a reduced cell number and loss of synchronization between single cells at later times. Therefore, single-cell proteomics will provide fundamental and valuable understanding of genetic heterogeneity in their responses to drugs and other stimuli, especially in cancer clinic research [48]. Although most of single-cell proteomics approaches are still limited to dozens of proteins, they already demonstrated the feasibility of realizing a more detailed characterization of cellular phenotypes [49-51].

Single cell metabolomics:

Metabolomics, when combined with genomics, transcriptomics and proteomics, offers us a synthetic view to fully understand the functionality of each individual cell. Within a single cell, the transcripts derived from DNA are translated into proteins, which act as enzymes to catalyze intermediate products of metabolism. Therefore, metabolites act as a connection between genotype and phenotype on single cell level, providing a logical view on single cell’s behavior. Single cell metabolomic method was applied to a single isolated CTC from a neuroblastoma patient’s blood for a comprehensive detection of the metabolite and lipid profiles [52]. The metabolic profile of the single CTC was acquired along with detection of vital molecules such as amino acids, catecholamine metabolites, which are specific to neuroblastosoma cancer and drugs from the patient’s treatment regimen. This indicated that single cell metabolomic could be useful for monitoring drug delivery concentration levels to targeted cells.

Multi-omics:

Furthermore, because a cell’s state is determined by the complex interaction of different molecules from genome, epigenome, transcriptome, proteome and metabolome, new multi-omics approaches for measuring different types of molecules simultaneously are also recently reported, including DR-seq, G&T-seq, scTrio-seq, scMT-seq, scNMT-seq, and many others (Table 2) [49,53-66]. Based on integrated measurement and co-analysis, simultaneous profiling of distinct types of molecules at single cell level (DNA, RNA, and protein) will enable us comprehensive understanding of cellular function and phenotype variation. Several reviews are available on discussing the features of single-cell multi-omics [46,53,55,58,67-69]. Considering the complexity of cell heterogeneity from the same individual, multi-omics approaches will enhance our power to detect the genotype-phenotype relationships comprehensively and unambiguously. Five complementary strategies for data integration from measuring two or more different molecules in the same cell have been proposed, including “combine”, “separate”, “split”, “convert”, and “predict” approaches [53]. Application of multi-omics in single cells will enable, amongst other things, the generation of mechanistic models relating (epi) genomic variations and transcript/protein expression dynamics, which in turn should allow a more detailed exploration of cellular behavior in health and disease [55]. For example, one recent study reported the scNMT-seq (single-cell nucleosome, methylation and transcription sequencing) [59]. They investigated chromatin accessibility, DNA methylation and transcription simultaneously by applying a GpC methyltransferase to label open chromatin followed by bisulfite and RNA sequencing. Methylated cytosines in a GpC context demarcate accessible DNA (linker regions and nucleosome-free DNA), while methylation is read from conversion events of cytosines in a CpG context. By profiling the mouse embryonic stem cell, they found novel links between all three molecular layers and revealed dynamics coupling between epigenomic layers during differentiation. However, one limitation of scNMT-seq is the need to filter out C-C-G and G-C-G positions from the raw data, which reduced the number of cytosines that can be assayed compared with scBS-seq by ~50%. Additionally, these types of multi-omics studies are also suffering from certain challenges and limitations, for example, many of them are still low throughput with low genome coverage (e.g. scBS-seq data cover less than ~40% of genome) and low mappability rates (~20-30%). Finally, raw data for each omic type must also be separately filtered, processed, mapped to account for low signal-to-noise ratio due to locus dropout, amplification bias, and technical variation.

Application of Third-Generation Sequencing (TGS) technologies:

Compared to the second-generation short read sequencing platforms, the third-generation sequencing technologies, including Pacific Biosciences (PacBio) Single Molecule Real Time (SMRT) sequencing and the Oxford Nanopore Technologies sequencing (ONT), can generate average read lengths over 10,000 bp and some read length up to 1 Mb or more [70,71]. Using bulk cells, the major applications of third-generation technologies range from de novo sequencing, resequencing, to transcriptomics, epigenetics, metagenomics, and others. For example, they
have been used to produce highly accurate and contiguous genome assemblies, avoiding the coverage bias introduced by the whole genome amplification. They have also been applied to resequencing analyses, to create detailed maps of phased structural variants. TGS technologies have also been widely used to study transcriptomes (e.g. Iso-seq), recognizing thousands of novel isoforms and gene fusions that were not found using second-generation short read sequencing [72]. Finally, some of the technologies (like ONT) also allow for direct measurement of epigenetic modifications from single DNA, RNA or protein molecules, such as methylation of DNA using Nanopolish or SignalAlign [71]. Even though the current sequence quality of TGS needs additional improvement, their future potential applications will be exciting in single cell genomics, single cell transcriptomics, single cell epigenetics, and even proteomics.

**Conclusion**

The application of single-cell omics has already provided great insight into our understandings of diverse biological processes with broad implications for both basic and clinical research that have previously been difficult to resolve from bulk population cells. On the other hand, certain challenges remain in the procedure of single cell isolation, whole genome amplification, library construction, sequencing, bioinformatics analysis, and data integration. For instance, depending on platforms or methods, single cell approaches suffer from low coverage, bias, errors, when compared to those for bulk cells. In conclusion, even with its challenges, we are confident that single cell multi-omics will provide us new opportunities for future research. As multi-omics technologies become more widely accessible and improved, they will lead to the unprecedented full-dimension discoveries about single cells.

**Acknowledgements**

The project was supported in part by AFRI grants No. 2013-67015-20951 from the USDA NIFA, grant No. US-4997-17 from the US-Israel Binational Agricultural Research and Development Fund. Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the US Department of Agriculture. The USDA is an equal opportunity provider and employer.

**Tables**

<table>
<thead>
<tr>
<th>Table 1: Characteristics of commonly used whole genome amplification methods.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Method</strong></td>
</tr>
<tr>
<td>DNA yield (per 100 μl reaction)</td>
</tr>
<tr>
<td>DNA product length (bp)</td>
</tr>
<tr>
<td>Amplification bias range between loci</td>
</tr>
<tr>
<td>DNA polymerase</td>
</tr>
<tr>
<td>DNA polymerase error rate</td>
</tr>
<tr>
<td>Amplification method</td>
</tr>
<tr>
<td>Coverage</td>
</tr>
<tr>
<td>False positive rate (amplification error rate)</td>
</tr>
<tr>
<td>False-negative rate (coverage and allelic dropout)</td>
</tr>
<tr>
<td>Uniformity</td>
</tr>
<tr>
<td>References</td>
</tr>
</tbody>
</table>
### Table 2: Selected multi-omics studies in single cells.

<table>
<thead>
<tr>
<th>Omics</th>
<th>Cell type</th>
<th>Method</th>
<th>Measurement and description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genomics &amp; Epigenetics</td>
<td>Mouse preimplantation embryos</td>
<td>scCOOL-seq</td>
<td>Genome, DNA methylome, Chromatin accessibility: The reprogramming of the chromatin state and DNA methylation, copy number variation and ploidy were detected simultaneously from the same preimplantation embryos cell</td>
<td>Guo et al., 2017 [54]</td>
</tr>
<tr>
<td>Genomics &amp; Transcriptomics</td>
<td>Mouse embryos, human breast cancer cells; human lymphoblastoid cells</td>
<td>G&amp;T-seq</td>
<td>Genome, Transcriptome: Genomic DNA and full-length mRNA from single cells were separated (polyadenylated - poly(A) RNA is divided from its genomic DNA using a biotinylated oligo-dT primer in adaptation), then amplified in parallel and sequenced</td>
<td>Macaulay et al., 2015 [56]</td>
</tr>
<tr>
<td>Genomics, Epigenetics &amp; Transcriptomics</td>
<td>Mouse embryonic stem cell</td>
<td>DR-seq</td>
<td>Genome, Transcriptome: This method can be used to analyze genomic variation and transcriptome heterogeneity from the same cell without physical separation of the nucleic acids prior to amplification</td>
<td>Dey et al., 2015 [57]</td>
</tr>
<tr>
<td>Genomics, Epigenetics &amp; Transcriptomics</td>
<td>Human hepatocellular carcinomas (cancer cells)</td>
<td>scTrio-seq</td>
<td>Genome, DNA methylome, Transcriptome: Detect genome (CNVs), epigenome (DNA methylome), and transcriptome information of the single cells simultaneously, which allow for distinction between subpopulations within the same single cancer cell</td>
<td>Hou et al. 2016 [58]</td>
</tr>
<tr>
<td>Transcriptomics &amp; Epigenetics</td>
<td>Mouse embryonic stem cells</td>
<td>scNMT-seq</td>
<td>DNA methylome, transcriptome, chromatin accessibility: Chromatin accessibility, DNA methylation and transcriptome profiling were investigated parallelly by applying a GpC methyltransferase to label open chromatin followed by bisulfite and RNA sequencing</td>
<td>Clark et al., 2018 [59]</td>
</tr>
<tr>
<td>Transcriptomics &amp; Epigenetics</td>
<td>Human brain cells</td>
<td>scDrop-seq &amp; scTHS-seq</td>
<td>Transcriptome, Chromatin accessibility: Integration data generated from nuclear transcriptomic and DNA accessibility maps allow for the cell-type distinctions in human brain tissues</td>
<td>Lake et al., 2018 [60]</td>
</tr>
<tr>
<td>Transcriptomics &amp; Epigenetics</td>
<td>Mouse embryonic stem cell</td>
<td>scM&amp;T-seq</td>
<td>Transcriptome, DNA methylome: DNA methylome and transcriptome profile were investigated simultaneously by applying scBS-seq to genomic DNA purified according to the G&amp;T-seq protocol</td>
<td>Angermueller et al., 2016 [61]</td>
</tr>
<tr>
<td>Transcriptomics &amp; Epigenetics</td>
<td>Human primordial germ cells</td>
<td>scrRNA-seq &amp; WGBS</td>
<td>Transcriptome, DNA methylome: Characterized the developmental-stage-specific features of global DNA methylation and transcriptome of human PGCs at single-cell and single-base resolutions</td>
<td>Guo et al., 2015 [62]</td>
</tr>
<tr>
<td>Transcriptomics &amp; Proteomics</td>
<td>Human glioblastoma cells</td>
<td>PEA &amp; Gene Expression Assay</td>
<td>Transcriptome, Proteome: Large sets of both RNA and protein targets in single-cell lysates was surveyed simultaneously for cell functions and responses</td>
<td>Darmanis et al., 2016 [49]</td>
</tr>
<tr>
<td>Transcriptomics &amp; Proteomics</td>
<td>Mouse macrophage cells</td>
<td>Microfluidics &amp; scRNA-seq</td>
<td>Transcriptome, Proteome: The genome-wide transcriptome can be acquired from the same cell with protein profile which detected by a splittable single-cell microchip integrating high-density antibody array</td>
<td>Georg and Wang 2016 [63]</td>
</tr>
<tr>
<td>Transcriptomics &amp; Proteomics</td>
<td>Human peripheral blood mononuclear cells</td>
<td>Branched DNA assay</td>
<td>Transcriptome, Proteome: mRNA content is quantified with increased fluorescence signal detection and reduced background, as well as low hybridization temperatures, and leaving intact assayed cells</td>
<td>Soh et al., 2016 [64]</td>
</tr>
<tr>
<td>Proteomics &amp; Metabolomics</td>
<td>Human glioblastoma cancer cells</td>
<td>Integrated SCBC chip (Single-cell barcode chip)</td>
<td>Proteome, Metabolome: Metabolic heterogeneity and interactions between metabolites and signaling proteins can be resolved readily and accurately by using SCBS chip which integrated with surface-competitive binding assays with fluorescence readouts and functional protein immunoassays</td>
<td>Xue et al., 2015 [65]</td>
</tr>
<tr>
<td>Proteomics &amp; Metabolomics</td>
<td>Human primary naive T cell</td>
<td>High-resolution MS (mass spectrometry)</td>
<td>Proteome, Metabolome: Dynamic metabolome and proteome profiles of T cells were surveyed by using high-resolution mass spectrometry</td>
<td>Geiger et al., 2016 [66]</td>
</tr>
</tbody>
</table>
References


8. Cheung VG, Nelson SF. Whole genome amplification using a degenerate oligonucleotide primer allows hundreds of genotypes to be performed on less than one nanogram of genomic DNA. Proceedings of the National Academy of Sciences. 1996; 93: 14676-14679.


