Emerging Techniques in Single-Cell Epigenomics and their Applications to Cancer Research
Lo PK and Zhou Q*

Abstract
Epigenomics encompasses studies of the chemical modifications of genomic DNA and associated histones, interactions between genomic DNA sequences and proteins, the dynamics of the chromosomal conformation, the functional relationships between these epigenetic events, and the regulatory impacts of these epigenetic events on gene expression in cells. In comparison to current techniques that are only capable of characterizing average epigenomic features across bulk cell ensembles, single-cell epigenomics methodologies are emerging as powerful new techniques to study cellular plasticity and heterogeneity, as seen in stem cells and cancer. Here we summarize available techniques for studies of single-cell epigenomics, review their current applications to cancer research, and discuss future possibilities. This review also highlights that the full potential of single-cell epigenetic studies will be comprehended through integrating the multi-omics information of genomics, epigenomics and transcriptomics.

Keywords
Epigenomics; Genomic DNA; Oncogenesis; Chemotherapeutic treatments; single-cell techniques; intratumor heterogeneity

Methodologies of Single-Cell Epigenomics

Newly developed single-cell techniques are concisely summarized in this section to provide readers with the overview of advances in single-cell epigenomics. The timeline for the evolution of technologies on single-cell epigenomics is shown in Figure 1. Advantages and disadvantages of these established single-cell epigenomics methods and their comparisons are summarized in Table 1. Comprehensive details of these techniques have been well reviewed elsewhere [2,3].

Analysis of single-cell DNA modifications

Methylation at cytosine residues of CpG dinucleotide in the genome is a major type of DNA modifications, which is involved in silencing gene expression. Among methods used to analyze DNA methylation, bisulfite sequencing (BS-seq) that involves the chemical conversion of unmethylated cytosine into uracil is a powerful tool to measure CpG methylation in a single-base resolution [4,5]. Recent technical advances enable BS-seq to measure DNA methylation at a single-cell level. To reach the single-cell level, single cells are isolated and transferred into separate PCR tubes/wells (e.g. the 96-well platform) by serial dilution, micromanipulation (e.g. mouth pipetting), fluorescence-activated cell sorting (FACS), laser-capture microdissection (LCM) and microfluidic devices [6]. The extracted DNA derived from single cells can be processed either by reduced-representation bisulfite sequencing (RRBS) or by post-bisulfite
adaptor tagging (PBAT) to create indexed sequencing libraries wherein well-specific barcodes that can be traced back to particular single cells are incorporated into bisulfite-converted sequence reads (Figure 2) [7-10]. In the scRRBS method, five DNA processing steps before PCR amplification, including genomic DNA purification, restriction enzyme digestion, end repair/dA tailing, adapter ligation and bisulfite conversion, are prepared in a single tube to avoid loss of single-cell DNA (Figure 2) [7]. To increase the sequencing coverage of methylated DNA, bisulfite treatment in PBAT-based single-cell BS-seq (scBS-seq) is performed before incorporation of indexed sequencing primers into DNA libraries (Figure 2) [8-10]. This new procedure of PBAT prevents the degradation of indexed sequencing libraries caused by bisulfite treatment [8-10]. In addition to these bisulfite-dependent methods, a bisulfite-independent method called genome wide CpG island (CGI) methylation sequencing for single cells (scCGI-seq) has been invented to profile DNA methylation patterns in a single cell, a new technique combining methylation-sensitive restriction enzyme digestion and multiple displacement amplification (MDA) for selective detection of methylated CGIs (Figure 2) [11]. The MDA is an experimental step that selectively amplifies methylated CGI-containing long DNA strands but not short unmethylated CGI fragments. These single-cell methods preserve long-range DNA methylation profiles across a sparse, diploid genomic landscape. By comparing scBS-seq to scRRBS, scBS-seq provides the higher coverage of CpG dinucleotide (~3.7 million CpGs) than that of scRRBS (~one million CpGs) [7,9,12]. Therefore, scBS-seq includes more CpG-sparse regions than scRRBS. Nevertheless, the coverage of CGI from scRRBS analysis is higher than that from scBS-seq analysis [7,12]. Moreover, scRRBS profiles the genome in a relatively consistent and less random manner when compared to scBS-seq. This advantage allows scRRBS to provide the higher overlap between the individual CpG sites covered in different single cells [12]. Due to both methods offering their common and unique DNA methylation information, these two techniques are complementary to each other. Therefore, the choice of which method to use relies on the interest of specific studies. With regard to the most recent scCGI-seq, it offers the higher coverage of CGIs (>70% of all CGIs) than that of scRRBS [11]. This improved coverage from
scCGI-seq represents a 66-fold increase in the fraction of consistently profiled CGIs across individual cells when compared to scRRBS [11]. Owing to this substantial improvement, scCGI-seq has become a powerful and reliable method to profile DNA methylation of CGIs in a single cell.

Given that the outcome of DNA methylation is to alter gene expression, concurrent profiling analysis of both DNA methylome and RNA expression profiles in a single cell is critical for understanding the relationship between DNA methylation patterns and RNA expression profiles in a single cell. Recently, two techniques, called scMT-seq and scM&T-seq, have been established for parallel single-cell sequencing analysis of both methylome and transcriptome [13,14]. scMT-seq is a combination of both scRRBS and scRNA-seq (Smart2-seq) and scM&T-seq is a technique integrating scBS-seq with scRNA-seq [13,14]. Both single-cell techniques have taken the advantage of the method capable of simultaneously isolating cytosolic RNA for RNA-seq and nuclear genomic DNA for DNA methylene profiling [13,14]. These multi-omics methods have opened a window of opportunity for dissecting the mechanisms of epigenetic gene regulation.

In addition to methylated cytosine (5mC), the other minor types of DNA modifications have been identified, such as hydroxymethylated cytosine (5hmC) that is generated by oxidation of 5-methylcytosine, a reaction mediated by ten-eleven translocation (TET) methylcytosine dioxygenase. 5hmC has been known to contribute to the active demethylation of the paternal chromosomes during early embryonic stages [15]. The currently established methods for measuring these minor DNA modifications, such as TET-assisted bisulfite sequencing (TAB-seq) and Aba-seq, could potentially be compatible with single-cell approaches. Indeed, Mooijman and colleagues presented a single-cell Aba-seq technology, based on 5hmC glucosylation and glucosylation-dependent digestion of DNA, to analyze genome-wide cell-to-cell variability in 5hmC profiles [16-18].

### Analysis of single-cell histone modifications

The various covalent modifications of histones have been discovered and they are functionally involved in regulating genomic characteristics and transcriptional states [19]. These histone marks are routinely mapped using chromatin immunoprecipitation sequencing (ChIP-seq), which involves chromatin immunoprecipitation with antibodies specific to these histone marks. Although this method is powerful, high background noise is problematic for performing ChIP-seq at the single-cell level. To overcome this issue, genomic
DNA from a pool of single cells is processed through micrococcal nuclease (Mnase) digestion and barcoding steps before the immunoprecipitation step. This improved approach enables pull-down to be effectively performed on thousands of cells. Although the low coverage of sequencing reads is the disadvantage of scChIP-seq, this limited coverage potentially enables this technique to concurrently assess a large number of single cells. Indeed, by combination with a droplet-based microfluidics setup, this approach can process large numbers of single cells in parallel [20].

In addition to single-cell ChIP-seq (scChIP-seq), another method called DamID has been developed to map protein-DNA interactions at a single-cell level. This new method involves the expression of a fusion protein of *Escherichia coli* deoxyadenosine methylase (Dam) and the interested protein in cells. This design allows Dam to methylate DNA on adenine residues in close proximity to the sites of protein binding. Methylated sites are then digested by the damylation-sensitive restriction enzyme DpnI, followed by ligation with sequencing adapters to generate sequencing libraries. This new method has been successfully applied to the study of genomic DNA interactions with the nuclear lamina in single cells [21]. Moreover, single-cell DamID could be an alternative for genome-wide analysis of histone modifications by fusing Dam with specific histone readers or modifiers. Due to the current sequencing resolution of single-cell DamID limited to the order of 100 kb, this disadvantage to some extent restricts its application. Therefore, future optimizations and improvements are needed for enabling this technique to map transcription factor binding sites in single cells.

**Analysis of single-cell chromatin accessibility and conformation**

Transcriptional activation is known to result in disruption of nucleosome organization at promoters, enhancers, silencers, insulators and locus control regions due to transcription factor binding. Therefore, these regulatory DNA regions coincide with open/accessible genomic sites of remodeled chromatin. Mapping of these accessible chromatin regions is relevant to understanding of gene expression, cell proliferation, cell differentiation, functional diversification and disease development. Several techniques have been developed to map accessible chromatin regions in a single cell. The first mapping method called the assay for transposase-accessible chromatin using sequencing (ATAC-seq) is based on the employment of a Tn5 transposase enzyme to simultaneously cleave DNA within accessible open chromatin regions and the followed ligation of fragmented DNA with adapter sequences (a process called tagmentation) [22,23]. To achieve the single-cell level, ATAC-seq adopts a “combinatorial indexing” strategy in which the tagmentation is performed on 96 pools of a few thousand nuclei with the incorporation of a unique barcode into each pool and then these 96 pools are mixed and split again before a second barcode is incorporated into each new pool via polymerase chain reaction (PCR) (Figure 3). Optimized two-round barcode indexing enables a particular barcode combination to be specific to a single cell (Figure 3) [24]. Therefore, this approach allows the ATAC-seq data to represent the epigenomics information of a single cell. Another modified single-cell ATAC-seq method employs a microfluidics device to generate single-cell pools instead of using the combinatorial indexing method [25]. Although its throughput is substantially lower, this modified technique has substantially improved the sequencing resolution and coverage, encompassing 70,000 reads per cell compared with 3000 reads from the prior method.

In addition to the transposase-based method, a DNase I-based single-cell sequencing (scDNase-seq) has been applied to analysis of chromatin accessibility in a single cell [26]. The design of this method is based on the fact that open/accessible chromatin regions are vulnerable to DNase I digestion. This technique offers a high sequencing resolution (300,000 mapped reads per single cell), but low mapping efficiency and lower throughput are its disadvantages. In particular, scDNase-seq and aforementioned scATAC-seq could be potentially conducted in parallel with scRNA-seq through the effective separation of DNA from RNA or parallel amplification of both materials [14,27]. Moreover, an attractive technique called nucleosome occupancy and methylome sequencing (NOMe-seq) is capable of simultaneously analyzing nucleosome positioning, accessible chromatin and DNA methylation on a genome-wide scale [28]. NOMe-seq exploits a methyltransferase enzyme to methylate GpC dinucleotides within open/accessible genomic regions, but not those within nucleosome-bound DNA regions. After the GpC methylation reaction, DNA is treated with bisulfite and processed as BS-seq for next-generation sequencing analysis. Due to the difference between enzymatic (GpC) and endogenous (GpG) DNA methylation, NOMe-seq is able to map both methylation patterns in a single-base resolution. By integrating scNOMe-seq with PBAT-based scBS-seq, a novel multi-omics method called single-cell COOL-seq has been developed for parallel analysis of the chromatin state/nucleosome positioning, DNA methylation, copy number variation and ploidy in single cells (Figure 4) [29]. scCOOL-seq has been utilized to analyze the reprogramming of the chromatin state and DNA methylation in mouse preimplantation embryos [29]. This new multi-omics methodology offers the first single-cell and parental allele-specific analysis of the global chromatin state and DNA methylation dynamics at single-base resolution in early mouse embryos and provides new insights into this embryonic process involving the heterogeneous yet highly ordered features of epigenomics reprogramming.

Due to no selectivity for open chromatin, the high depth of sequencing may be necessary for scNOMe-seq to guarantee coverage of genomic regions of interest. Moreover, given the need of filtering out some ambiguous nucleotide positions (e.g. C-C-G and G-C-G), this reduces the coverage of genome-wide cytosines from scNOMe-seq analysis by ~50% when compared to that from scBS-seq analysis. However, despite this reduced coverage, scNOMe-seq is still able to profile a large proportion of the genomic loci with important regulatory roles (e.g. promoters and enhancers). In addition to characterizing linear chromatin organization, studies of chromosomal conformation have become important as the three-dimensional topology of chromosomes determines the interactions of gene promoters with enhancers, silencers, and insulators. A genome-scale technique called Hi-C has been developed to define the conformation of chromosomes. In Hi-C, cellular chromatin DNA is fixed and then digested by sequence-specific restriction enzymes while maintaining intact protein–protein and protein-DNA interactions. Re-ligation of digested DNA ends generates DNA chimaeras, which preserve spatial proximity interactions between different genomic loci. Given that Hi-C analysis of bulk cell populations may result in some ambiguity in interpretation of the results, single-cell Hi-C (scHi-C) has been developed as a cutting-edge technique to assess chromosomal conformation in a single cell [30,31]. scHi-C is currently limited in its resolution but still able to define the individual chromosome organization, compartmentalization, and interchromosomal interactions.
Single-cell epigenomics and cancer

The study of intratumor heterogeneity is important as it relates to tumor microenvironment, genetic as well as epigenetic diversity resulting from clonal evolution of tumor cells, and cancer progression to invasive/metastatic disease (Figure 5) [32]. To determine whether intratumor heterogeneity can be defined by their genomic, epigenomic and transcriptomic profiles, single-cell sequencing technologies are mandatory. Indeed, single-cell DNA sequencing (scGenome-seq) was successfully employed to reveal that multiple cancer types can undergo clonal evolution, and to identify founder mutations and subclonal mutations that are implicated in cancer progression (Figure 5) [33,34]. Similarly, single-cell transcriptome profiling has been exploited to discover stem-cell-like subpopulations within cancers (Figure 5) [35,36]. These single-cell-based studies have provided new insights into how cancer progresses and whether single-cell profiling can be used to predict disease outcome.

The cancer epigenome has been known to be substantially remodeled during tumorigenesis, metastasis and drug resistance (Figure 5). As therapeutic drugs targeting epigenetic enzymes (e.g. DNA methyltransferases and histone deacetylases) have been demonstrated to be effective in treatment of several cancer types, these indicate that epigenomic alterations play a critical role in disease progression [37]. It has been shown that DNA hypomethylation occurs on a global scale in cancer. In contrast, DNA hypermethylation occurs at specific genomic loci, which are associated with aberrant nucleosome positioning and chromatin modifications. Recently, genome-wide profiling of methylome in triple-negative breast cancers (TNBCs) has stratified TNBCs into three distinct methylation clusters, which are associated with patients’ prognosis [38]. Nevertheless, the important question regarding whether intratumor heterogeneity can be reflected by epigenetic heterogeneity can only be addressed by single-cell epigenomic analyses. Two recent studies have reported the potential of single-cell epigenomic techniques in addressing the role of epigenetic heterogeneity in cancer. By employing a single-cell triple omics sequencing technique called scTrio-seq that can simultaneously assess the genomic copy-number variations (CNVs), DNA methylome, and transcriptome of 25 single cancer cells, Hou and colleagues have identified two subpopulations present within cells isolated from a human hepatocellular carcinoma tissue sample [39]. Another study by Litzenburger and colleagues employed scATAC-seq combined with scRNA-seq to reveal that in K562 leukemic cells, the levels of the cell surface marker CD24 correlated with chromatin accessibility changes linked to GATA transcription factors in single cells [40]. Their finding is clinically relevant as the status of GATA/CD24 in leukemic cells was related to drug sensitivity.
Figure 4: The diagram of the scNOMe-seq technology in the scCOOL-seq method. The GpC methyltransferase adds the methyl group to cytosine residues of GpC dinucleotides in the open chromatin regions, but not in the closed chromatin regions. After the GpC methylation reaction, genomic DNA is subjected to post-bisulfite adaptor-tagging sequencing (PBAT-seq) analysis. DNA methylation and chromatin accessibility can be profiled simultaneously according to CpG and GpC methylation patterns, respectively.

Figure 5: The diagram for illustrating the roles of genomic, epigenomics and transcriptomic alterations in tumorigenesis and metastasis. The current established single-cell genomics, epigenomics, transcriptomics and their derived multi-omics methods are also shown in the diagram to indicate their applications.
and differential self-renewal capacity. Their study suggests that GATA factors are critically involved in determining cell phenotypes through modulating epigenomic plasticity. Evidence from these single-cell epigenomic studies demonstrates that single-cell epigenomic features can be utilized to characterize intratumor heterogeneity and its roles in drug sensitivity and the clonal evolution of cancer cells.

Perspectives

Numerous single-cell epigenomic technologies have been developed to profile the distinct types of epigenomic alterations occurring at a single-cell level in heterogeneous tumor tissue (Figure 5). By combining these single-cell epigenomic techniques with spatial transcriptomics technologies, several multi-omics studies (e.g. scMT-seq, scM&T-seq, scATACT-seq, scCOOL-seq, scTrios-seq) have been established to analyze two to three different omics in a same single cell (Figure 5). It can be imagined that new multi-omics technologies with ability to analyze genomic, epigenomic and transcriptomic profiles in parallel will be rapidly developed in the future to identify rare subpopulations of cancer cells within a tumor. Some of aforementioned studies have demonstrated this possibility. Studies of cancer stem cells (CSCs) are clinically important as CSCs have been considered to contribute to intratumor heterogeneity, invasive progression, metastasis, and drug resistance. Due to dynamic self-renewal, differentiation and dormancy of CSCs in the tumor microenvironment, single-cell epigenomic techniques will be powerful tools to effectively trace CSC plasticity in a tumor. Single-cell technologies will be also very useful to analyze multi-omic profiles of circulating tumor cells (CTCs), representing metastatic cancer cells, in comparison to those from primary tumors. Multi-omic analyses of CTCs are crucial for understanding the mechanisms underlying cancer metastasis. Moreover, single-cell epigenomic studies may potentially guide the development of novel therapeutic strategies with clinical significance in cancer. Particularly, single-cell analysis of DNA methylation is an attractive strategy for cancer screening as the DNA methylation landscapes of mammalian cells with single-cell resolved-representation bisulfite sequencing. Nature Protoc 10: 845-859.


References


