Molecular Aspects of Medicine xxx (2017) 1–8

Contents lists available at ScienceDirect

Molecular Aspects of Medicine

journal homepage: www.elsevier.com/locate/mam

Single cell epigenome sequencing technologies

Lu Wen*, Fuchou Tang*

ARTICLE INFO

Article history:
Received 1 June 2017
Received in revised form
5 September 2017
Accepted 13 September 2017
Available online xxx

ABSTRACT

Epigenetic regulation plays crucial roles in the development and disease processes. Since different cell types with distinct epigenetic characteristics are always intermingled together in the tissues or organs, the single cell analysis provides a universal resolution for dissecting their intrinsic complexities. In this review, we discuss recent advances of developing single cell epigenome sequencing techniques for profiling DNA modifications (5mC, 5hmC, 5fC and 5caC), chromatin accessibility, histone modifications, protein-DNA interactions and three-dimensional genome architecture (Hi-C) in an individual mammalian cell. Recent progresses of single cell multi-omics sequencing techniques are also reviewed. These techniques are emerging as powerful tools for unravelling the unique epigenomic features of rare cell types and epigenetic heterogeneity within a seemingly homogenous cell population. We also discussed the current limitations and future development trend of single cell epigenome sequencing techniques.

The human body is composed of a wide variety of cell types with interdependent functions. To understand the development or disease process in human, it is important to acquire genome-wide information for decoding the identity, characteristics and functional status of each cell type. Epigenetic regulation is a crucial layer of the molecular mechanisms underlying the appropriate activation or repression of gene expression. In past decades, methods for profiling genome-wide epigenetic information have been effective in identifying gene regulatory elements related to cell identity and status during physiological development or pathological diseases. However, different cell types are always intermingled together in a tissue or organ, and it is often difficult to isolate a single type of cells in an unbiased manner. Moreover, even within a given cell type, there is intrinsic epigenetic heterogeneity due to cell cycle, microenvironments, etc. In recent years, along with the fast progression of single cell transcriptome sequencing technologies (Tang et al., 2009; Wen and Tang, 2016; Yang et al., 2017), single cell epigenome sequencing technologies have also been rapidly developed (Guo et al., 2013; Nagano et al., 2013; Schwartzman and Tanay, 2015; Clark et al., 2016). These technologies tremendously help scientists to explore the epigenomes of rare cell types and to decipher the epigenetic heterogeneities among a population of cells, which are generally inaccessible by bulk analyses.

1. Methodologies for single cell epigenome sequencing techniques

1.1. Single-cell sequencing techniques for profiling DNA methylation

DNA methylation (5mC) is the covalent addition of a methyl group to the 5' carbon of cytosine, which mainly occurs at CpG dinucleotides in the mammalian genome. It plays an essential role in silencing of repeat elements, repression of transcription, inactivation of the X chromosome, and maintenance of gene imprinting, and abnormally alternated in human diseases such as cancer (Baylin and Jones, 2011).

Bisulfite sequencing is known as the gold standard for DNA methylation analysis, given its precision and single-base resolution. Recent studies have established genome-wide DNA methylation sequencing methods at single-cell resolution based on reduced-representation bisulfite sequencing (RRBS) or post-bisulfite adaptor tagging (PBAT) bisulfite sequencing strategies (Guo et al., 2013; Smallwood et al., 2014; Farlik et al., 2015; Gravina et al., 2016). RRBS is a cost-effective genome-wide DNA methylation sequencing method that uses a restriction endonuclease digestion (usually DNA methylation insensitive restriction enzyme *Msp*I) and a size-selection strategy for enrichment of the CpG-dense regions in the genome (Meissner et al., 2008). To adapt this method to single cell analysis, we integrated the first five steps in the standard RRBS protocol before polymerase chain reaction (PCR) amplification into a one-tube reaction to minimize the loss of the tiny amount of the genomic DNA in an individual cell during the DNA purification steps (Guo et al., 2013, 2015a). For an individual mouse

* Corresponding authors. ICG, BIOPIC, School of Life Sciences, Peking University, Beijing, PR China.
E-mail addresses: wenlu@pku.edu.cn (L. Wen), tangfuchou@pku.edu.cn (F. Tang).

http://dx.doi.org/10.1016/j.mam.2017.09.002
0098-2997/

diploid cell, the scRRBS (single-cell RRBS) assay covers, on average, one million CpG sites with approximately 70% of the CpG islands in the genome being captured (see Table 1 for a summary of the single-cell epigenome sequencing technologies).

The PBAT method is theoretically able to cover the entire genome, similar to the standard whole genome bisulfite sequencing (WGBS) method (Miura et al., 2012). It is different from the standard WGBS in that the adaptors are added after, instead of before, bisulfite treatment. This circumvents the drawback of bisulfite-induced fragmentation of the DNA templates and is more suitable for analysis of the trace amounts of genomic DNAs within an individual cell. Three PBAT-based assays have been reported (Smallwood et al., 2014; Farlik et al., 2015; Gravina et al., 2016). The scBS-Seq (single-cell bisulfite sequencing) assay (Smallwood et al., 2014; Clark et al., 2017), which includes five rounds of pre-amplification, covers more CpG sites (an average of 3.7 million CpG sites in an individual mouse diploid cell) than the other two methods (averaging 1 or 2.2 million CpG sites in an individual mouse diploid cell) (Farlik et al., 2015; Gravina et al., 2016). scBS-seq covers more CpG sites compared to scRRBS and merges only a dozen cells reproduce the bulk result. Because of these features, scBS-seq is more advantageous than scRRBS if one wants to obtain methylation information from as many as CpG sites in an individual cell as possible. However, the scRRBS focuses on the restricted but more informative CpG-dense regions in the genome, allowing for a lower sequencing burden, less sequencing cost for each cell, and more overlapped regions covered in each cell for side-by-side analysis of their heterogeneity.

More recently, methods capable of high-throughput, single-cell methylome sequencing have been reported. In the snmC-seq (single-nucleus methylcytosine sequencing) strategies, including fluorescence-activated cell sorting (FACS), plate-based bisulfite treatment, DNA purification and multiplex processing were applied to increase the throughput of a PBAT-like assay (Luo et al., 2017). In the sci-MET (single-cell combinatorial indexing for methylation analysis) method, a combinatorial tagging strategy, which will be described in more detail later in this article, was used (Mulqueenet al, 2017).

In their first applications, these single cell DNA methylome sequencing methods allowed for charting of the genome-wide and base-resolution maps of mammalian oocytes, early embryos and primordial germ cells (PGCs), as well as uncovering dynamic whole-genome DNA demethylation and re-methylation waves during epigenomic reprogramming of preimplantation and PGC development (Guo et al, 2013, 2014a, 2015a; 2014b; Guo et al, 2015b). In addition, it has been shown that neuronal sub-populations can be identified by the gene body non-CpG (CpH) methylation patterns of individual cells, and furthermore, regulatory elements of the neuronal subpopulations can be identified through differential CpG methylation patterns (Luo et al., 2017). Moreover, DNA methylation heterogeneity can be observed in mouse embryonic stem cells (ESCs) (Smallwood et al., 2014).

1.2. Single-cell sequencing techniques for profiling other DNA modifications

The oxidization derivatives of 5mC, e.g., 5-hydroxylmethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC), are produced by the ten-eleven translocation methylcytosine dioxygenase (TET) proteins and are the critical intermediate products of the active DNA demethylation pathway (Shen et al., 2014; Xu and Walsh, 2014). Identifying these DNA modifications provides crucial clues about active DNA demethylation and their potential functions for gene expression regulation. Traditional bisulfite sequencing is not able to distinguish among these cytosine modifications: both 5mC

Table 1
Summary of single-cell epigenome sequencing technologies.

Assays	Epigenetic markers	Strategies	Coverage per cell	Throughput	References
scRRBS	5mC	RRBS	1.5 million CpG sites	1–100	(Guo et al., 2013; Guo et al., 2015a)
scBS-seq	5mC	PBAT (five rounds of preamplification)	3.7 million CpG sites	1–100	(Smallwood et al., 2014; Clark et al., 2017)
scWGBS	5mC	PBAT (one round of preamplification) with 3′ tagging	1.0 million CpG sites	1–100	(Farlik et al., 2015)
scWGBS (Gravina)	5mC	PBAT (two rounds of preamplification)	2.2 million CpG sites	1–100	(Gravina et al., 2016)
snmC-seq	5mC	PBAT (one round of preamplification) with 3′ tagging	5% of the genome	100–1000	(Luo et al., 2017)
sci-MET	5mC	in nuclei Tn5 transposition for combinatorial cellular indexing	0.8% of the genome	>1000	(Mulqueenet al, 2017)
scAba-seq	5hmC	Aba-seq	44,000 5hmC sites	100–1000	(Mooijman et al., 2016)
CLEVER-seq	5fC	Selectively labeling of 5fC that induces C-to-T conversion in PCR amplification, combined with MALBAC	60-80% of the genome	1–100	(Zhu et al., 2017)
scMAB-seq	5fC/5caC	Combining MAB-seq with scRRBS or scBS-seq	Depending on scRRBS or scBS-seq used	1–100	(Wu et al., 2017)
sciATAC-seq	Chromatin accessibility	Combining ATAC-seq with combinatorial cellular indexing	450 DHSs	>1000	(Cusanovich et al., 2015)
scATAC-seq	Chromatin accessibility	Combining ATAC-seq with microfluidics	5,000 DHSs	100–1000	(Buenrostro et al., 2015)
scDNase-seq	Chromatin accessibility	DNase-seq	38,000 DHSs	1–100	(Jin et al., 2015)
Drop-ChIP	Histone modification	Combining drop-based microfluidics and DNA barcoding with ChIP-seq	1,000 H3K4me2 peaks	>1000	(Rotem et al., 2015)
scDamID	Nuclear-lamina interactions	DamID	450,000 Dam/*DpnI* sites	1–100	(Kind et al., 2015)
scHi-C	3D chromatin architecture	Combining in nuclei Hi-C with manipulation of individual nuclei	1,000 contacts	1–100	(Nagano et al., 2013)
snHi-C	3D chromatin architecture	Phi29 amplification after proximal ligation	400,000 contacts	1–100	(Flyamer et al., 2017)
sciHi-C	3D chromatin architecture	Combining in nuclei Hi-C with combinatorial cellular indexing	9,000 contacts	>1000	(Ramani et al., 2017)

Please cite this article in press as: Wen, L., Tang, F., Single cell epigenome sequencing technologies, Molecular Aspects of Medicine (2017), http://dx.doi.org/10.1016/j.mam.2017.09.002