Studying epigenomics in single cells: what is feasible and what can we learn?

“It seems most likely that advances in understanding epigenetic processes at the level of single cells will come from improved or entirely new imaging techniques to accurately visualize chromatin and its modifications at single loci in single cells.”

First draft submitted: 9 September 2015; Accepted for publication: 16 September 2015; Published online: 8 December 2015

Keywords: DNA sequencing • epigenetics • RNA sequencing • single-cell

Technological developments over the past few years in fluorescence-activated cell sorting (FACS), imaging, laser dissection and microfluidics, together with improved methods of DNA amplification, bar-coding and sequencing have enabled researchers to start to apply well-established approaches in molecular biology to analyze individual cells [1]. This raises the very reasonable question of what single cell analysis might tell us that could not be learnt from the analysis of populations of homogeneous purified cells? Maybe the simplest, unsurprising answer is that, even within the most highly purified population of cells it appears that, at the molecular level, every cell is different. This cellular heterogeneity may be structured: some cells may express a particular combination of genes while another subset expresses a different combination. In turn, this distinct gene expression might correlate with specific cellular phenotypes and functional properties. As such substructures will be averaged in a population of cells, the ability to detect subpopulations expressing specific combinations of genes will be lost unless analyzed at the single cell level.

A further advantage of single cell analysis relates to the observation that gene expression is a fundamentally stochastic process [2]. For example, *in situ* RNA FISH and reporter gene analysis of cells residing in the same, genetically homogeneous population often shows that nascent RNA transcripts are only present in a proportion of the cells and at widely

varying levels [3]. Furthermore, live imaging of such cells shows that transcription occurs in bursts rather than in a continuous fashion. This stochastic gene expression is driven by a combination of random intrinsic transcriptional ‘noise’ and multiple extrinsic factors that regulate gene expression. It is clear that the modulation of this stochastic gene expression has fundamental consequences for cellular function [2], however, this information is lost at the cell population level.

Before the development of next generation sequencing, analysis of RNA at the single cell level was largely limited to RNA FISH and nonquantitative PCR. These techniques could be used to study chromosomal or genetic rearrangements, the distribution of specific classes of nucleic acids within single cells, and nascent and stable RNA from single, or small numbers of genes. However, it is now feasible to analyze the entire transcriptome from a single cell with a variety of RNA-sequencing techniques [4]. Using these data, it is possible to disassemble a population of cells into its individual components, reassemble the data and, via model-driven and model-free clustering approaches, or exploratory ordination techniques such as principal components analysis (PCA), identify subpopulations of cells with recurrent combinations of RNA transcripts that were previously merged and missed by conventional population analysis. These new techniques have been harnessed to unravel previously unrecognized cellular heterogeneity in certain tis-

Quin F Wills

Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford, OX3 9DS, UK and Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford, OX3 7BN, UK

Douglas R Higgs

Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford, OX3 9DS, UK

Adam J Mead

Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford, OX3 9DS, UK and NIHR Biomedical Research Centre, University of Oxford, Churchill Hospital, Oxford, OX3 7LE, UK
Author for correspondence: adam.mead@imm.ox.ac.uk

sues and to thereby identify new cell populations in a series of landmark studies [5–9].

The above studies have unequivocally established the importance of single cell genomic approaches to describe both the transcriptional heterogeneity and combinatorial gene expression structures that underlie cell-to-cell phenotypic variation. The field now faces the next challenge: to gain a better understanding of the mechanistic basis for this heterogeneity of gene expression. It is clear that variations in DNA sequence can impact on stochastic gene expression [10,11]. Importantly, new protocols have been developed that enable both DNA and RNA to be sequenced from the same single cell thereby allowing the relationship between variation in DNA sequence and heterogeneity of gene expression to be analyzed at an unprecedented scale [12,13]. However, a major challenge for such single cell genomic analysis relates to incompleteness of the data both for DNA and RNA analysis. For example, current RNA-sequencing techniques probably capture only approximately 10% of the mRNA content of a cell [14]. This means that it is not possible to achieve a complete picture of the genomic landscape of any single cell using the current approaches. For lower level expressed transcripts (<10 copies per cell), this is a particular problem [14]. Innovative bioinformatic approaches have been used to help control for this ‘technical noise’ [15], nevertheless, incomplete data remain a major challenge for single cell genomic datasets that computational approaches can only partially overcome.

While DNA sequence variations are important, it is also necessary to understand how heterogeneity of gene expression is generated from identical DNA sequences, and how the resulting cell-to-cell variability of gene expression is determined. In order to explore this, it is necessary to apply the techniques we use to assess the epigenetic landscape in populations of cells to single cells. These include analyzing chromosome conformation, nucleosome phasing, regions of nuclease or transposase accessibility, chromatin modifications and the interaction of transcription factors and cofactors with chromatin. In addition, an important component of the epigenetic profile is to analyze covalent modifications of DNA, by methylation for example. Although the final readout for each of these approaches comes in the form of DNA that can be sequenced, all of these analyses require some manipulation of the cell to produce the DNA fragments to be analyzed. Clearly, the more steps involved before producing the DNA readout, the less efficient the process becomes and the less complete the dataset for each cell analyzed. For example, ChIP-seq is not possible using fewer than 10^3 – 10^4 cells. Even using the most efficient protocol for

analyzing DNA methylation (the most tractable of the epigenetic phenomena using genomics) only 10–50% of CpG sites of each cell are covered [16–18]. Although the data are incomplete, some pioneering studies have developed various protocols to analyze epigenetic phenomena, genome-wide, at the single cell level and preliminary data have been obtained [16–21]. However, in all cases, meaningful biological data are only obtained by the reassembly of single cell data and computationally inferring what might be happening in single cells rather than obtaining a true picture of what is happening in any individual cell [18,20]. As for the primary analysis of DNA and RNA, it is possible to use such information, via clustering or ordination techniques, to identify subpopulations of cells that differ from each other with respect to their epigenetic landscapes; for example, by identifying particular patterns of transposase accessibility using ATAC seq [20,21], which is already producing new and interesting perspectives. For example, in a study of 15,000 cells [21], a median of 1685 ATAC seq reads per cell was enough to identify different human cell lines. Others [20] have obtained a mean of 73,000 reads per cell; sufficient density to study the biology of regulatory domains. Nevertheless, since all epigenetic phenomena can only produce two (both alleles), one (a single allele) or no (neither allele) signals, it seems unlikely that it would ever be possible to attain a sufficiently discriminating signal/technical-noise ratio at a genome-wide scale to confidently determine the epigenetic state at a single locus, in a single cell, using current genomics approaches.

A further major limitation for the above technical approaches is that they only provide a ‘snapshot’ of the genetic and epigenetic landscape at any given time, as it is necessary to destroy the cell in order to carry out the analysis. As most epigenetic signals are transient during development and differentiation, if the goal is to correlate these findings with assays of cellular function, stated simply, we need to know how they are generated, how they are inherited from one generation of cells to their progeny, and when and how they are removed. Addressing the order of events cannot be achieved at the population level and it will be necessary to develop new tools and new approaches to single cell analysis that will probably rely on imaging cells and, ultimately, imaging and tracing live cells. Solving this issue lies at the heart of the mechanistic questions we want to address in the field of epigenetics. Work is underway to develop such techniques. In the past, immunofluorescence microscopy has been used to visualize the nuclear localization and movement of chromosomes, subchromosomal structures (e.g., telomeres) and artificial loci. In addition, immunofluorescence has been used to look at global patterns of chromatin modifications

and the distribution of transcription factors, cofactors, chromatin associated factors and RNA. The challenge is to visualize these phenomena at a single locus, and steps towards this may be to use fluorescently labeled specific antigen binding fragments (Fabs) [22], single chain antibodies or histacs all of which can directly attach to specific chromatin modifications, in live cells. Another developing technology that maybe applied to the analysis of epigenetics in single cells is referred to as the ^{m6}A-Tracer system. In this, proteins of interest can be fused to Dam methylase, which methylates A residues in DNA. Then, using a fragment of the restriction enzyme DpnII (which recognizes ^{m6}A) fused to eGFP, it is possible to see where the protein of interest has interacted with chromatin via the modified ^{m6}A that has been imposed by the interacting fusion protein. Originally used to visualize interactions between chromatin and lamin A, this system promises to be very useful in tracing ephemeral epigenetic interactions [23].

Another newly described and ingenious method to analyze fixed cells uses a combination of a biotin-labeled oligonucleotide that anneals to a DNA sequence of interest (e.g., a promoter), which can then be recognized by an antibiotin antibody. A second antibody directed against a specific chromatin modification (e.g. H3K4me3) is then used to detect chromatin modified in this way. Each antibody is engineered to carry a DNA oligo, and when the two oligos are in close proximity it is possible to generate a signal that can be visualized in a fixed cell by rolling circle PCR. This so-called proximity ligation assay (PLA) provides a method to investigate the epigenetic state of a single locus in a single cell [24].

To conclude, the development of approaches to analyze the molecular biology of single cells is a very

exciting new area of biological research that promises to reveal mechanisms which could never be addressed by analyzing cell populations. At present, although we can infer what is happening inside a single cell, the incomplete information we have does not allow us to determine this with confidence. We rely on re-integrating single cell data for most functional genomic approaches. It is certain that improvements in genomics technology will increase the amount of data that can be obtained by sequencing. However, it is a moot point whether it will ever be possible to have sufficiently complete, robust information to know precisely what the cellular machinery will do next or whether this will always be defined in terms of probability. It seems most likely that advances in understanding epigenetic processes at the level of single cells will come from improved or entirely new imaging techniques to accurately visualize chromatin and its modifications at single loci in single cells. Whatever the approaches, answering the exacting questions we are asking in the field of epigenetics in single cells will critically depend on the efficiency and robustness of any new technique that is developed.

Financial & competing interests disclosure

The Oxford Consortium for Single Cell Biology is supported by a UK Medical Research Council Clinical Research Capabilities and Technologies Initiative award. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

References

- Junker JP, van Oudenaarden A. Every cell is special: genome-wide studies add a new dimension to single-cell biology. *Cell* 157(1), 8–11 (2014).
- Raj A, van Oudenaarden A. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* 135(2), 216–226 (2008).
- Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol.* 4(10), e309 (2006).
- Wills QF, Mead AJ. Application of single-cell genomics in cancer: promise and challenges. *Hum. Mol. Genet.* 24(R1), R74–R84 (2015).
- Jaitin DA, Kenigsberg E, Keren-Shaul H *et al.* Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* 343(6172), 776–779 (2014).
- Durruthy-Durruthy R, Gottlieb A, Hartman BH *et al.* Reconstruction of the mouse otocyst and early neuroblast lineage at single-cell resolution. *Cell* 157(4), 964–978 (2014).
- Treutlein B, Brownfield DG, Wu AR *et al.* Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* 509(7500), 371–375 (2014).
- Wen L, Tang F. Reconstructing complex tissues from single-cell analyses. *Cell* 157(4), 771–773 (2014).
- Grun D, Lyubimova A, Kester L *et al.* Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature* 525, 251–255 (2015).
- Raj A, Rifkin SA, Andersen E, van Oudenaarden A. Variability in gene expression underlies incomplete penetrance. *Nature* 463(7283), 913–918 (2010).
- Raser JM, O'Shea EK. Control of stochasticity in eukaryotic gene expression. *Science* 304(5678), 1811–1814 (2004).
- Dey SS, Kester L, Spanjaard B, Bienko M, van Oudenaarden A. Integrated genome and transcriptome sequencing of the same cell. *Nat. Biotechnol.* 33(3), 285–289 (2015).
- Macaulay IC, Haerty W, Kumar P *et al.* G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. *Nat. Methods* 12(6), 519–522 (2015).

- 14 Islam S, Kjallquist U, Moliner A *et al.* Highly multiplexed and strand-specific single-cell RNA 5' end sequencing. *Nat. Protocols* 7(5), 813–828 (2012).
- 15 Brennecke P, Anders S, Kim JK *et al.* Accounting for technical noise in single-cell RNA-seq experiments. *Nat. Methods* 10(11), 1093–1095 (2013).
- 16 Smallwood SA, Lee HJ, Angermueller C *et al.* Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat. Methods* 11(8), 817–820 (2014).
- 17 Lorthongpanich C, Cheow LF, Balu S *et al.* Single-cell DNA-methylation analysis reveals epigenetic chimerism in preimplantation embryos. *Science* 341(6150), 1110–1112 (2013).
- 18 Farlik M, Sheffield NC, Nuzzo A *et al.* Single-cell DNA methylome sequencing and bioinformatic inference of epigenomic cell-state dynamics. *Cell Rep.* 10(8), 1386–1397 (2015).
- 19 Nagano T, Lubling Y, Stevens TJ *et al.* Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* 502(7469), 59–64 (2013).
- 20 Buenrostro JD, Wu B, Litzenburger UM *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* 523(7561), 486–490 (2015).
- 21 Cusanovich DA, Daza R, Adey A *et al.* Epigenetics. Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* 348(6237), 910–914 (2015).
- 22 Hayashi-Takanaka Y, Yamagata K, Wakayama T *et al.* Tracking epigenetic histone modifications in single cells using Fab-based live endogenous modification labeling. *Nucleic Acids Res.* 39(15), 6475–6488 (2011).
- 23 Kind J, Pagie L, Ortobozkoyun H *et al.* Single-cell dynamics of genome–nuclear lamina interactions. *Cell* 153(1), 178–192 (2013).
- 24 Gomez D, Shankman LS, Nguyen AT, Owens GK. Detection of histone modifications at specific gene loci in single cells in histological sections. *Nat. Methods* 10(2), 171–177 (2013).