

## INVITED REVIEW

# Genetics and immunity in the era of single-cell genomics

Felipe A. Vieira Braga<sup>1,\*</sup>, Sarah A. Teichmann<sup>1,2,3</sup> and Xi Chen<sup>1,\*</sup><sup>1</sup>Wellcome Trust Sanger Institute, <sup>2</sup>European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI) and <sup>3</sup>Cavendish Laboratory, Cambridge University, Cambridge, UK

\*To whom correspondence should be addressed at: Tel: +44 (0)1223 834244; Email: fvb@sanger.ac.uk (FAVB); xc1@sanger.ac.uk (XC)

## Abstract

Recent developments in the field of single-cell genomics (SCG) are changing our understanding of how functional phenotypes of cell populations emerge from the behaviour of individual cells. Some of the applications of SCG include the discovery of new gene networks and novel cell subpopulations, fine mapping of transcription kinetics, and the relationships between cell clonality and their functional phenotypes. Immunology is one of the fields that is benefiting the most from such advancements, providing us with completely new insights into mammalian immunity. In this review, we start by covering new immunological insights originating from the use of single-cell genomic tools, specifically single-cell RNA-sequencing. Furthermore, we discuss how new genetic study designs are starting to explain inter-individual variation in the immune response. We conclude with a perspective on new multi-omics technologies capable of integrating several readouts from the same single cell and how such techniques might push our biological understanding of mammalian immunity to a new level.

## Introduction

Inter-cellular heterogeneity within seemingly homogeneous cell populations has recently emerged as an important source of functional variation within and across samples (1). Over the past few years, technical and methodological developments in the field of single-cell genomics (SCG) have unveiled new biological insights that were previously masked due to measurement approaches that used bulk samples of cells (1,2). By studying gene expression at the single-cell level, one can estimate both the frequency and the strength of transcriptional bursts (3), reflecting the level of noise in gene expression, that is strong but infrequent transcriptional bursts lead to more noise than small but frequent bursts (4–6). Differential burst behaviour can exist for genes with similar mean expressions in bulk populations, so that biological differences are missed when only bulk samples are analyzed (3,7,8). This detailed information about gene expression can be extracted for each allele (maternal versus

paternal) individually, particularly if full-length transcript RNA-seq methods are used (4,6,9).

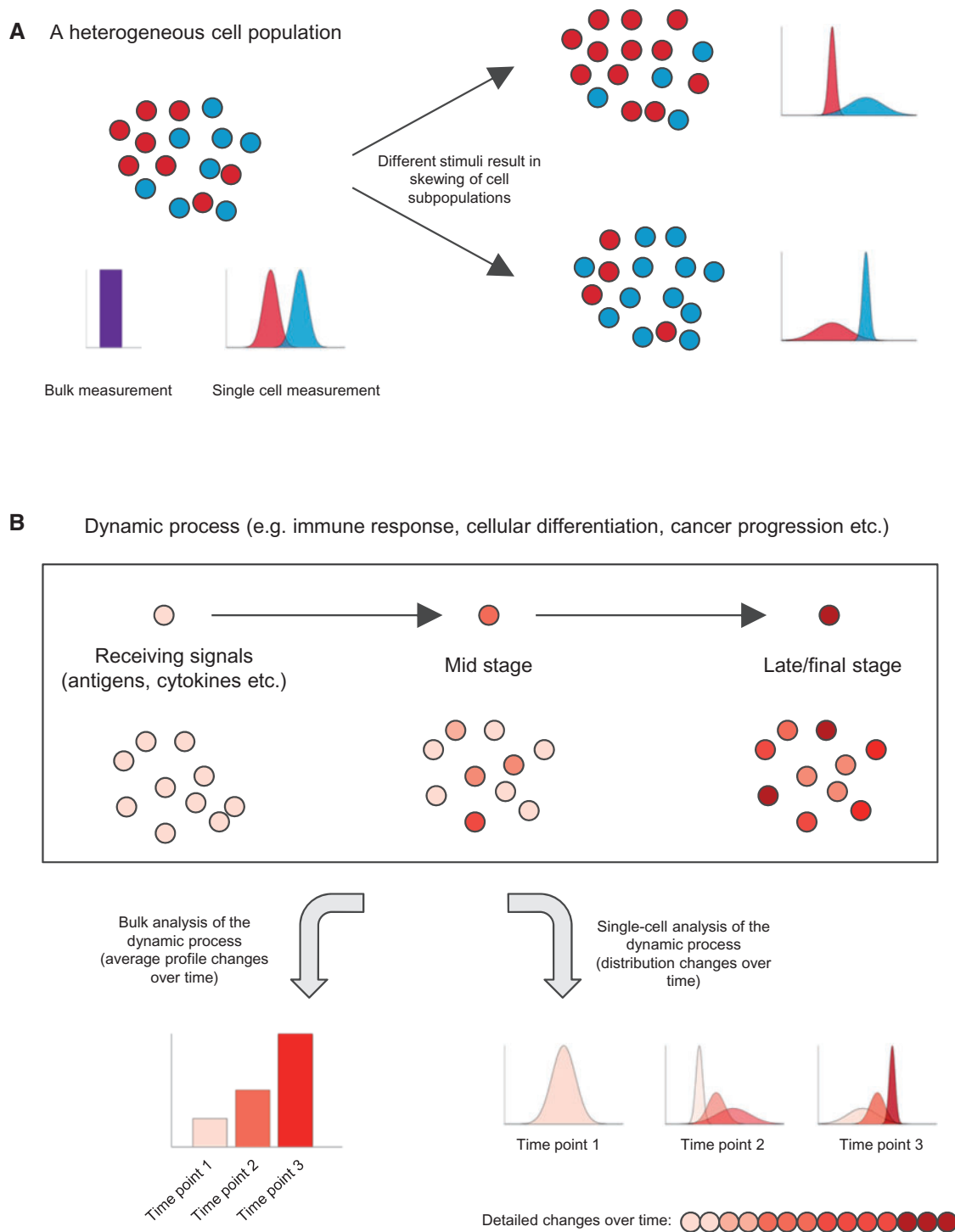
Another level of information inherent within single-cell RNA-sequencing data are gene regulatory interactions and networks, which can be inferred from correlations and clustering of gene expression variability across large numbers of single cells (10,11). Furthermore, single-cell RNA-seq data from individual T or B cells allow one to fine map their clonality and lineage through the somatically recombined T- or B-cell receptor sequences in addition to maintaining the expression information of all the other expressed genes. This reveals direct correlations between their clonal origin and functional phenotypes (12), information that is impossible to obtain by direct bulk analysis.

Beyond the insights mentioned above, a major advantage of SCG methods is that they allow the discovery of new cell states or cell types within a sample (Fig. 1A). SCG methods have frequently led to the discovery of new subtypes of cells without *a priori* knowledge about cell type-specific markers. One of the

Received: June 6, 2016. Revised: June 6, 2016. Accepted: June 15, 2016

© The Author 2016. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.



**Figure 1.** Single-cell measurements retain critical cellular heterogeneity information that is lost by bulk genomics assays. (A) Bulk measurements of a cell population cannot distinguish different cellular states. Single-cell analyses can reveal different cell subpopulations and predict/investigate cell skewing upon receiving external stimuli. (B) Single-cell measurements provide higher temporal resolution and a more comprehensive overview of a dynamic process.

most commonly investigated systems by SCG technologies has been the mammalian immune system, which consists of a wide variety of cell types responsible to fight infection and cancer. Early single-cell transcriptomic studies showcased the feasibility of identifying distinct cell types from a complex tissue and revealing potential novel markers for specific cell types (13–15). Recent studies further demonstrated that it was possible to

uncover new hidden cell subpopulations within very similar cells (16–20). Examples include steroidogenic mouse T helper 2 cells (16), mouse Th2 developmental stages (17), different subpopulations within human ILC3 cells (20), mouse Th17 cells (18), the highly divergent subpopulations of mouse invariant natural killer T (iNKT) cells (21), and most recently, three cellular states during mouse CD4+ T-cell activation (22).

In this review, we will address how new developments in SCG are changing our understanding of biology, with a specific focus on the immune system. The immune system displays tremendous genetic and environmentally determined inter-individual variation and has a central role in determining human health, so is a fertile area for the application of SCG methods.

### Massively parallel single-cell sequencing: what it tells us so far?

In recent years, the field of SCG has advanced rapidly and revolutionized our view of many biological processes. Due to the development of both single-cell capture technologies and whole genome/transcriptome amplification methods, it is now feasible to interrogate the genome and quantify gene expression from single cells by next generation sequencing (NGS). The first step in a typical SCG experiment is to capture individual single cells. This can be done by traditional fluorescence-activated cell sorting of single cells into individual wells of 96- or 384-well plates, or by using microfluidic platforms such as the Fluidigm C1 or microdroplet technologies (23–25). Once individual cells are captured, the cells are lysed, and the genomic DNA or mRNA is amplified by specific protocols to make NGS libraries for sequencing. For single-cell genome sequencing, see two recent reviews by Huang *et al* (26) and by Gawad *et al* (2); for single-cell RNA sequencing (scRNA-seq), see a recent review by Kolodziejczyk *et al* (1).

The genome is usually considered stable during development due to the low error rate of DNA replication (27,28). However, errors do occur, and over sufficient cell divisions, genomic heterogeneity within the same organism due to these somatic mutations—also known as genetic mosaicism—are expected. This genomic heterogeneity can have profound influences on both unicellular and multicellular organisms. Although current genome-wide association study (GWAS) has been very successful in identifying genetic variants that are responsible for many human diseases, most of the discoveries are based on studies conducted at the level of the tissue or individual (29,30). The genetic differences among individual cells are largely neglected and cannot be observed when taking the average signal from a bulk population. However, the genetic heterogeneity of different cells is important in many complex traits, such as cancer (31,32). It is likely that single-cell DNA sequencing, as well as other SCG technologies, will shed light on the role of somatic mutations in the coming years.

Single-RNA sequencing has provided many new insights into immunity in the past few years. For instance, Mahata *et al* investigated the gene expression profiles of many individual mouse T helper 2 (Th2) cells and identified a novel subgroup of Th2 cells that produce the steroid pregnenolone, indicating that this specific Th2 subtype contribute to immune homeostasis via steroidogenesis (16). In a more recent study of the same group, Proserpio *et al* pinpoint the relationship between cellular proliferation and differentiation during *ex vivo* Th2 differentiation and *in vivo* malaria infection, revealing three distinct cell states with different cytokine secretion and proliferation rates. Specifically, the cytokine-secreting T cells proliferate twice as fast as activated cells that do not express cytokines (22).

Gaublomme *et al* studied the heterogeneity of gene expression during the differentiation of Th17 cells in mice. They used both *in vivo* Th17 cells from central nervous system (CNS) and lymph nodes (LN) from an experimental autoimmune encephalomyelitis model of autoimmunity along with *in vitro* differentiated Th17 cells. Data from scRNA-seq indicated that Th17 cells

span a spectrum of cellular states with distinct gene expression signatures; furthermore, novel regulatory factors important for Th17 self renewal were identified (18). In a later study, Björklund *et al* investigated the heterogeneity of human innate lymphoid cells (ILCs) by performing scRNA-seq on hundreds of CD127+ ILCs and natural killer (NK) cells from human tonsils (20). Clustering the transcriptomes of 648 cells revealed four distinct populations of cells that correspond to four known ILC populations: ILC1, ILC2, ILC3 and NK cells. More interestingly, the ILC3 group appeared to be very heterogeneous, and further clustering on 1,958 annotated immune genes separated ILC3 cells into three different subpopulations, two out of which have never been described in humans before (20).

Recently, Engel *et al* combined scRNA-seq with bulk H3K27ac ChIP-seq data to investigate the heterogeneity of mouse iNKT subsets (NKT0, NKT1, NKT2 and NKT17). Distinct gene expression programs were identified for different iNKT subsets. More importantly, a substantial difference between cells from within the same subsets was prominent, especially in the NKT2 subset. Interestingly, one particular enhancer, HS V, from within the classical Th2 locus control region (33,34), was identified as a key enhancer for the expression of the cytokine IL-4 in NKT2 cells (21). These studies beautifully illustrate how cells within seemingly homogeneous population can be quite different. Indeed, even among well-defined cell types, new cell subpopulations can often be found.

### Human genetic variation of the healthy immune response

Despite living in an environment, full of life threatening microorganisms, our immune system ensures that humans thrive on our planet. The essential role of the immune system is illustrated by the devastating effects of primary and acquired immunodeficiencies that can lead to persistent infection, cancer susceptibility and even death (35). Uncontrolled immune responses against self (autoimmunity) can lead to extensive tissue destruction and complex clinical manifestations (36). Primary immunodeficiencies are normally associated with one (or a few) well-defined genetic mutations (37), while autoimmune disorders are complex multifactorial diseases where both environment and genetics play a role (36). Despite several single-nucleotide polymorphisms (SNPs) associated with autoimmunity, only a small fraction of carriers develop autoimmunity (38).

The genotypic and phenotypic diversity present in autoimmune disorders is also present in the immune response of 'healthy' individuals (39–41). Using a cohort of twin siblings, Brodin *et al* (42) have estimated that the inter-individual variation of CD8 T-cell and B-cell subpopulations is mainly driven by the environment, while CD4 T cells have shown a high degree of heritability (42). Eosinophils, neutrophils and some NK cell features have also shown a high degree of heritability, while monocytes have shown a low heritability. Orru *et al* (43) genotyped a non-twin cohort and analysed more than 200 immune traits, identifying a variable degree of heritability depending on the analysed trait. An especially high degree of heritability was found in CD4 T regulatory cells (Tregs) (43). Tregs are responsible for regulating the extent of the immune response and so play a major role in diseases like cancer and autoimmunity. Therefore, understanding the genetic basis of immune system variation in healthy individuals can improve our understanding of how diseases develop (44).

The studies investigating human immune variation mentioned above have primarily made use of flow cytometry to

define populations and analyse specific protein expression in thousands of single cells. Despite their tremendous contribution to our understanding of immune variation, there are limitations to this technology. The number of parameters analysed tends to be small, and they must be selected *a priori*, generating an ascertainment bias. Genomic technologies offer an alternative to overcome such biases. Dimas *et al* (45) analysed primary fibroblasts, Epstein-Barr virus-immortalized B cells (lymphoblastoid cell lines or LCLs), and T cells from the same individual and identified cell-type specific regulatory mechanisms. By combining genotyping with unbiased whole cell transcriptome analysis, they were able to identify expression quantitative trait loci (eQTLs) that controlled gene expression in a cell-type specific way (45), establishing a new paradigm for the study of genetic variation in the immune system.

By analysing peripheral blood total CD4<sup>+</sup> T cells, Murphy *et al* (46) identified eQTLs regulating expression of IL23R and IL12R, two cytokine receptors essential for an appropriate immune response, and believed to be involved in autoimmune disorders. To minimize the effects related to the environment and the life history of the cells, Raj *et al* (47) analysed naive CD4 T cells, and identified several eQTLs in loci associated with autoimmune diseases, such as rheumatoid arthritis and multiple sclerosis, meaning that the relevant genes are differentially regulated by distinct genetic variants, mostly in non-coding regions (47). While resting naive CD4 T cells display a large number of eQTLs, analysis of effector memory CD4 T cells has identified only a few additional eQTLs (48), including variants in the IL23R locus. Regulatory CD4 T cells also display several eQTLs for genes like ENTPD1, FCRL1, and CD52 (49). Despite our knowledge of the role of many of the genes identified in these studies, such as IL23R (50) and CD52 (51), our overall understanding of the immune cell functions of several of these genes is fairly limited. Further investigations are needed to fully clarify the impact of such genetic variants.

Monocytes play an essential role in orchestrating the immune response: they are capable of presenting antigens and secreting cytokines fundamental for effector function and immune cell regulation (52). Several eQTLs have been identified in monocytes (53,54). The variant rs10784774 at the 12q15 locus was shown to regulate a series of genes, with an especially strong effect detected for *LYZ* and *CREB1*, genes with important functions for bacterial defense and transcriptional regulation of monocytes (55,56). Neutrophils are another subset of cells essential for the direct bacterial destruction and immune cell regulation via secretion of cytokines (57). Hundreds of eQTLs have been identified to regulate gene expression in human neutrophils (58), including several associated with neutrophil function. One example is rs933222, which was shown to be associated with the expression of *RAC2*, a gene involved in the NADPH oxidase complex, which is essential for generating the oxidative burst necessary for pathogen killing (58,48).

The studies discussed above suggested the dominance of cell-type specific eQTLs, with very few eQTLs shared across multiple cell types. However, a recent study from Peters *et al* (59) analysed monocytes, CD4 T cells, CD8 T cells, B cells and neutrophils from the same donor. Instead of first computing eQTLs for each cell type and then comparing the lists with each other, they used a Bayesian model (eQTL Bayesian Model Averaging, or 'eQTLBMA') to integrate the analysis of eQTLs from the five different cell types within the same model (59). They identified around 45% of the detected eQTLs in all five of the cell types analysed. This result is in marked contrast to previous studies where, for instance, only 21.8% of the detected

eQTLs were shared between monocytes and B cells (54). This suggests that appropriate study designs and data analysis approaches are crucial for establishing the real impact of genetic variants in immune cells.

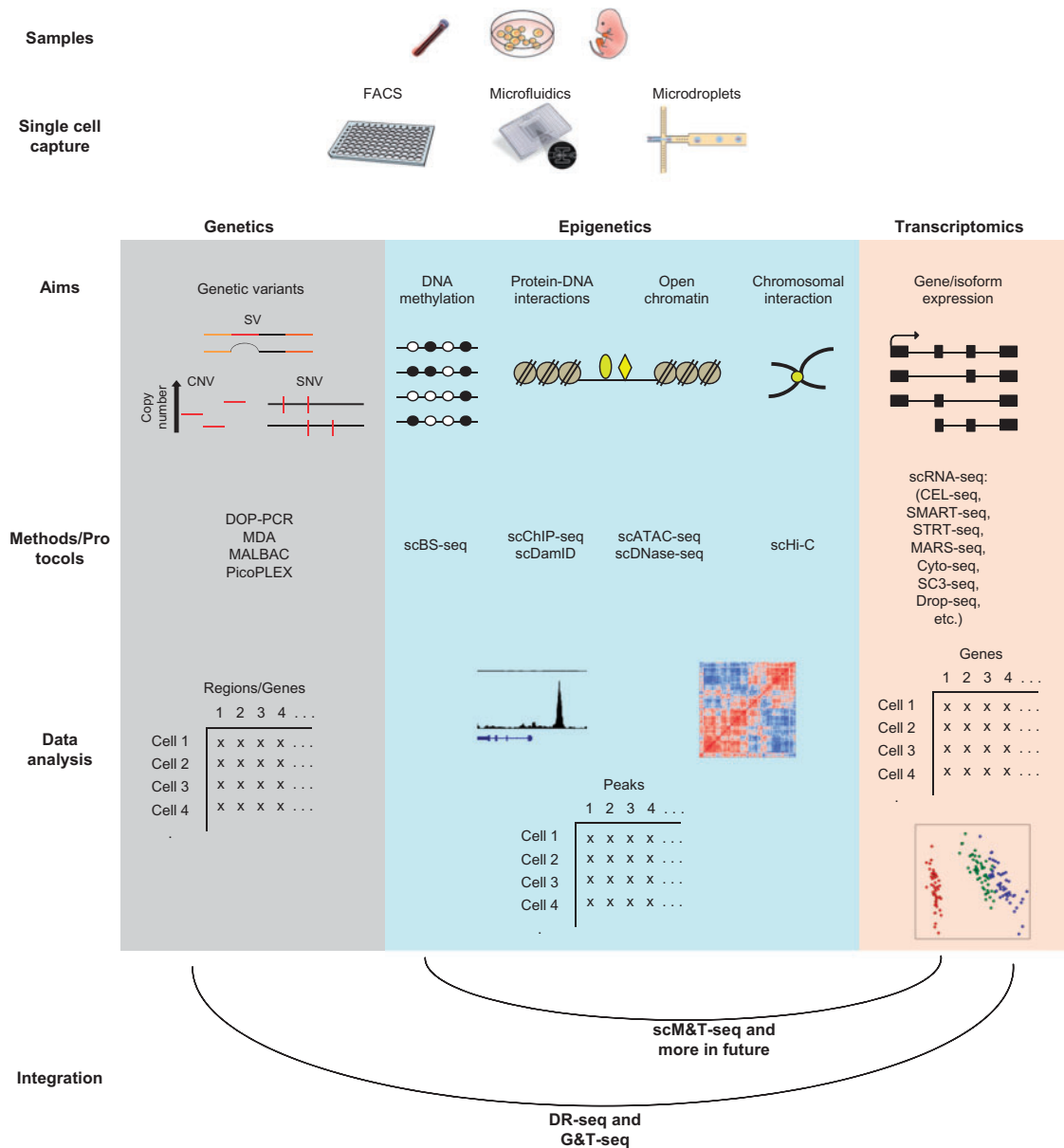
#### Four dimensions in genomic data analysis

Most genomic studies analyse cells in their steady state. However, many biological processes such as cancer progression, cellular differentiation and immune response are dynamic. In the context of a dynamic process, different cells may behave differently or respond/progress at different speeds (Fig. 1B). One classic example is the development of the nematode *Caenorhabditis elegans*. By exploiting the variation and asynchrony in worm development, Francesconi and Lehner identified thousands of eQTLs during a 12-hour developmental period. This provides a deep characterization of the genetic architecture of the worm's regulation of development (60).

T-cell activation is another dynamic process in which antigen and inflammatory cytokines mediate extensive transcriptional changes. Ye *et al* (61) analysed gene expression in CD4 T cells during polyclonal T-cell receptor activation under neutral or T helper 17 (Th17) polarising conditions. Despite high inter-individual variability in their cytokine responses, this variability did not have high heritability. Cytokine receptors displayed less inter-individual gene expression variability, yet several eQTLs appear to regulate their expression level (61). Importantly, more than half of the eQTLs identified were not detected when only resting CD4 T cells from the exact same donors were analysed—in other words, they were only revealed upon activation/polarisation (47). Hu *et al* (48) also analysed CD4 T cells after polyclonal activation and in addition to several eQTLs associated with gene expression; they identified rs389862, a variant that directly influences the direct proliferative capacity.

To date, thousands of genetic variants that are associated with complex traits have been identified by GWAS and deep sequencing (62). The majority of the genetic variants are located outside protein-coding regions (63,64), indicating that they are involved in the regulation of the gene expression. Hawkins *et al* (65) performed ChIP-seq to analyse the global chromatin state of Th1 and Th2 CD4 T cells, identifying lineage-specific enhancers. Strikingly, they found several SNPs associated with autoimmune diseases overlapping with these enhancers. For example, they discovered that rs10774213, an SNP associated with type 1 diabetes, resides within an enhancer region of the gene *CCND2*, and has a BACH2 binding motif. BACH2 is a transcription factor that regulates effector T-cell differentiation (66), and the enhancer is one of the most prominent super enhancer regions in mouse T cells (67). In fact, autoimmune disease-associated SNPs tend to be enriched in super enhancer regions (67).

Looking forward to the potential of SCG to make contributions to immune genetics, one could imagine that SNPs located in non-coding regions (such as enhancers and super enhancers) (68) might directly affect transcription kinetics, influencing either burst frequency or intensity. The use of single-cell RNA-seq data will be an invaluable tool to address such questions (3). In a typical SCG experiment, many single cells are captured and analysed so that the temporal information is retained within data from a 'snap shot' cell sample (69). By ordering the single cells in 'pseudotime' based on their gene expression profiles, one can improve temporal resolution of a dynamic biological process such as differentiation and infection (Fig. 1B), without *a priori* knowledge of marker genes (70).



**Figure 2.** The current workflow of single-cell genomic methods in genetics, epigenetics and transcriptomics, and emerging combined multi-omic technologies from the same single cell. A typical workflow starts with single cell capture through either traditional FACS, microfluidics capture or microdroplets. Then DNA or RNA is prepared for sequencing by different protocols based on the aims of the study and the area of interest (genetics, epigenetics or transcriptomics). Computational analyses are used later on to extract interesting biological information. Recent emerging technologies allow combined genetics-transcriptomics and epigenetics-transcriptomics investigation from the same cell.

In an effort to use pseudotemporal ordering of single cells to study myoblast differentiation, Trapnell *et al* developed an unsupervised algorithm called Monocle (70). This was one of the first methods to order single cells in 'pseudotime' to quantitatively measure the progress of single cells through a biological process. The algorithm revealed interesting findings relating to skeletal myoblast differentiation, including the switch-like inactivation of the key regulator ID1, a sequential wave of transcriptional reconfiguration and eight novel transcription factors that repress differentiation (70). These examples demonstrate the power of single-cell technologies to integrate the analysis of complex cellular states involved in dynamic temporal processes. We anticipate that these techniques will give us further insights into processes such as cell fate decisions during immune responses.

### Future outlook: multi-omic single-cell methods for genetic and transcriptional analysis

Gene expression heterogeneity is commonly observed in many single-cell transcriptome studies, especially in a dynamic process like the immune response where different cells respond to stimuli in slightly different ways or at different speeds. The extent to which this heterogeneity is due to stochastic differences or genetic differences is not clear at the moment. Therefore, integrating genetic and transcriptomic data at the single-cell level is extremely appealing and will help us to link individual genetic variants to individual cell phenotypes. One early study has revealed scQTLs for seven genes where no eQTLs had been identified by whole-tissue experiments (3). In addition, computational

efforts are under way to improve the genetic dissection of certain common traits using data from SCG methods (71).

At the moment, the development of new methods is a very active area of research in the field of SCG (1,2,72). Many traditional genome-wide assays are now possible at the single-cell level (73–81) (Fig. 2), and combined omic methods from the same cell are now becoming available. Dey *et al* developed a combined genomic DNA and mRNA sequencing (DR-Seq) method (82), and Macaulay *et al* developed a ‘genome and transcriptome’ sequencing (G&T-seq) method (83). Both methods combine whole genome amplification and whole transcriptome amplification to investigate genomic DNA and mRNA from the same cell, which makes the integration of genetic and transcriptional data at the single-cell level feasible. In a more recent study, Angermueller *et al* developed the scM&T-seq method (84), which was based on G&T-seq where a bisulphite convention was performed after the physical separation of genomic DNA from mRNA. This enables the investigation of DNA methylation and gene expression from the same cell.

Multi-omic profiling from the same single cell will usher in an exciting new era of single-cell systems biology, such that hitherto intractable biological questions can be addressed. We predict that more multi-omic methods from the same single cell will be developed in future, providing fundamental insights of a new nature, and revolutionizing the way we investigate basic biological questions.

## Acknowledgements

We thank Dr. Michael Stubbington, Dr. Gozde Kar and Dr. Tapio Lönnberg from the Teichmann group for the critical reading of the manuscript.

*Conflict of Interest Statement.* None declared.

## Funding

This work is supported by Open Targets MFG-GRAMMAR 664918 (F.A.V.B and S.A.T), MRG-GRAMMAR (X.C. and S.A.T) and the Wellcome Trust Sanger Institute (X.C and S.A.T). Funding to pay the Open Access publication charges for this article was provided The Wellcome Trust Sanger Institute.

## References

- Kolodziejczyk, A.A., Kim, J.K., Svensson, V., Marioni, J.C. and Teichmann, S.A. (2015) The technology and biology of single-cell RNA sequencing. *Mol. Cell*, **58**, 610–620.
- Gawad, C., Koh, W. and Quake, S.R. (2016) Single-cell genome sequencing: current state of the science. *Nat. Rev. Genet.*, **17**, 175–188.
- Wills, Q.F., Livak, K.J., Tipping, A.J., Enver, T., Goldson, A.J., Sexton, D.W. and Holmes, C. (2013) Single-cell gene expression analysis reveals genetic associations masked in whole-tissue experiments. *Nat. Biotechnol.*, **31**, 748–752.
- Borel, C., Ferreira, P.G., Santoni, F., Delaneau, O., Fort, A., Popadin, K.Y., Garieri, M., Falconnet, E., Ribaux, P., Guipponi, M., *et al.* (2015) Biased allelic expression in human primary fibroblast single cells. *Am. J. Hum. Genet.*, **96**, 70–80.
- Kim, J.K. and Marioni, J.C. (2013) Inferring the kinetics of stochastic gene expression from single-cell RNA-sequencing data. *Genome Biol.*, **14**, R7.
- Ginart, P., Kalish, J.M., Jiang, C.L., Yu, A.C., Bartolomei, M.S. and Raj, A. (2016) Visualizing allele-specific expression in single cells reveals epigenetic mosaicism in an H19 loss-of-imprinting mutant. *Genes Dev.*, **30**, 567–578.
- Avraham, R., Haseley, N., Brown, D., Penaranda, C., Jijon, H.B., Trombetta, J.J., Satija, R., Shalek, A.K., Xavier, R.J., Regev, A., *et al.* (2015) Pathogen Cell-to-Cell Variability Drives Heterogeneity in Host Immune Responses. *Cell*, **162**, 1309–1321.
- Shalek, A.K., Satija, R., Shuga, J., Trombetta, J.J., Gennert, D., Lu, D., Chen, P., Gertner, R.S., Gaublomme, J.T., Yosef, N., *et al.* (2014) Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature*, **510**, 363–369.
- Kim, J.K., Kolodziejczyk, A.A., Illicic, T., Teichmann, S.A. and Marioni, J.C. (2015) Characterizing noise structure in single-cell RNA-seq distinguishes genuine from technical stochastic allelic expression. *Nat. Commun.*, **6**, 8687.
- Padovan-Merhar, O. and Raj, A. (2013) Using variability in gene expression as a tool for studying gene regulation. *Wiley Interdiscip. Rev. Syst. Biol. Med.*, **5**, 751–759.
- Xue, Z., Huang, K., Cai, C., Cai, L., Jiang, C.Y., Feng, Y., Liu, Z., Zeng, Q., Cheng, L., Sun, Y.E., *et al.* (2013) Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature*, **500**, 593–597.
- Stubbington, M.J.T., Lönnberg, T., Proserpio, V., Clare, S., Speak, A.O., Dougan, G. and Teichmann, S.A. (2016) T cell fate and clonality inference from single-cell transcriptomes. *Nat. Methods*, **13**, 329–332.
- Islam, S., Kjällquist, U., Moliner, A., Zajac, P., Fan, J.B., Lönnberg, P. and Linnarsson, S. (2011) Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res.*, **21**, 1160–1167.
- Ramsköld, D., Luo, S., Wang, Y.C., Li, R., Deng, Q., Faridani, O.R., Daniels, G.A., Khrebtkova, I., Loring, J.F., Laurent, L.C., *et al.* (2012) Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.*, **30**, 777–782.
- Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Elefant, N., Paul, F., Zaretsky, I., Mildner, A., Cohen, N., Jung, S., Tanay, A., *et al.* (2014) Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science*, **343**, 776–779.
- Mahata, B., Zhang, X., Kolodziejczyk, A.A., Proserpio, V., Haim-Vilmovsky, L., Taylor, A.E., Hebenstreit, D., Dingler, F.A., Moignard, V., Göttgens, B., *et al.* (2014) Single-cell RNA sequencing reveals T helper cells synthesizing steroids de novo to contribute to immune homeostasis. *Cell Rep*, **7**, 1130–1142.
- Buettner, F., Natarajan, K.N., Casale, F.P., Proserpio, V., Scialdone, A., Theis, F.J., Teichmann, S.A., Marioni, J.C. and Stegle, O. (2015) Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat. Biotechnol.*, **33**, 155–160.
- Gaublomme, J.T., Yosef, N., Lee, Y., Gertner, R.S., Yang, L.V., Wu, C., Pandolfi, P.P., Mak, T., Satija, R., Shalek, A.K., *et al.* (2015) Single-Cell Genomics Unveils Critical Regulators of Th17 Cell Pathogenicity. *Cell*, **163**, 1400–1412.
- Grün, D., Lyubimova, A., Kester, L., Wiebrands, K., Basak, O., Sasaki, N., Clevers, H. and van Oudenaarden, A. (2015) Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature*, **525**, 251–255.
- Björklund, Å.K., Forkel, M., Picelli, S., Konya, V., Theorell, J., Friberg, D., Sandberg, R. and Mjösberg, J. (2016) The heterogeneity of human CD127+ innate lymphoid cells revealed by single-cell RNA sequencing. *Nat. Immunol.*, **17**, 451–460.
- Engel, I., Seumois, G., Chavez, L., Samaniego-Castruita, D., White, B., Chawla, A., Mock, D., Vijayanand, P. and

- Kronenberg, M. (2016) Innate-like functions of natural killer T cell subsets result from highly divergent gene programs. *Nat. Immunol.*, 10.1038/ni.3437.
22. Proserpio, V., Piccolo, A., Haim-Vilmovsky, L., Kar, G., Lönnberg, T., Svensson, V., Pramanik, J., Natarajan, K.N., Zhai, W., Zhang, X., et al. (2016) Single-cell analysis of CD4+ T-cell differentiation reveals three major cell states and progressive acceleration of proliferation. *Genome Biol.*, 17, 1–15.
  23. Mazutis, L., Gilbert, J., Ung, W.L., Weitz, D.A., Griffiths, A.D. and Heyman, J.A. (2013) Single-cell analysis and sorting using droplet-based microfluidics. *Nat. Protoc.*, 8, 870–891.
  24. Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A. and Kirschner, M.W. (2015) Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell*, 161, 1187–1201.
  25. Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldaman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015) Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell*, 161, 1202–1214.
  26. 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.
  27. Loeb, L.A. (1991) Mutator phenotype may be required for multistage carcinogenesis. *Cancer Res.*, 51, 3075–3079.
  28. Kunkel, T.A. (2004) DNA replication fidelity. *J. Biol. Chem.*, 279, 16895–16898.
  29. Amberger, J., Bocchini, C.A., Scott, A.F. and Hamosh, A. (2009) McKusick's Online Mendelian Inheritance in Man (OMIM). *Nucleic Acids Res.*, 37, D793–D796.
  30. Tringe, S.G., von Mering, C., Kobayashi, A., Salamov, A.A., Chen, K., Chang, H.W., Podar, M., Short, J.M., Mathur, E.J., Detter, J.C., et al. (2005) Comparative Metagenomics of Microbial Communities. *Science*, 308, 554–557.
  31. Hanahan, D. and Weinberg, R.A. (2011) Hallmarks of cancer: the next generation. *Cell*, 144, 646–674.
  32. Navin, N.E. (2015) The first five years of single-cell cancer genomics and beyond. *Genome Res.*, 25, 1499–1507.
  33. Lee, D.U., Agarwal, S. and Rao, A. (2002) Th2 lineage commitment and efficient IL-4 production involves extended demethylation of the IL-4 gene. *Immunity*, 16, 649–660.
  34. Lee, G.R., Fields, P.E., Griffin, T.J. and Flavell, R.A. (2003) Regulation of the Th2 cytokine locus by a locus control region. *Immunity*, 19, 145–153.
  35. Lim, M.S. and Elenitoba-Johnson, K.S.J. (2004) The Molecular Pathology of Primary Immunodeficiencies. *J. Mol. Diagn.*, 6, 59.
  36. Ramos, P.S., Shedlock, A.M. and Langefeld, C.D. (2015) Genetics of autoimmune diseases: insights from population genetics. *J. Hum. Genet.*, 60, 657–664.
  37. Milner, J.D. and Holland, S.M. (2013) The cup runneth over: lessons from the ever-expanding pool of primary immunodeficiency diseases. *Nat. Rev. Immunol.*, 13, 635–648.
  38. Mori, M., Yamada, R., Kobayashi, K., Kawaida, R. and Yamamoto, K. (2005) Ethnic differences in allele frequency of autoimmune-disease-associated SNPs. *J. Hum. Genet.*, 50, 264–266.
  39. Carr, E.J., Dooley, J., Garcia-Perez, J.E., Lagou, V., Lee, J.C., Wouters, C., Meyts, I., Goris, A., Boeckstaens, G., Linterman, M.A., et al. (2016) The cellular composition of the human immune system is shaped by age and cohabitation. *Nat. Immunol.*, 17, 461–468.
  40. Longo, D.M., Louie, B., Wang, E., Pos, Z., Marincola, F.M., Hawtin, R.E. and Cesano, A. (2012) Inter-donor variation in cell subset specific immune signaling responses in healthy individuals. *Am. J. Clin. Exp. Immunol.*, 1, 1–11.
  41. De Jager, P.L., Nir, H., Diane, M., Aviv, R., Stranger, B.E. and Christophe, B. (2015) ImmVar project: Insights and design considerations for future studies of 'healthy' immune variation. *Semin. Immunol.*, 27, 51–57.
  42. Brodin, P., Jovic, V., Gao, T., Bhattacharya, S., Angel, C.J.L., Furman, D., Shen-Orr, S., Dekker, C.L., Swan, G.E., Butte, A.J., et al. (2015) Variation in the human immune system is largely driven by non-heritable influences. *Cell*, 160, 37–47.
  43. Orrù, V., Valeria, O., Maristella, S., Gabriella, S., Carlo, S., Francesca, V., Mariano, D., Sandra, L., Magdalena, Z., Fabio, B., et al. (2013) Genetic Variants Regulating Immune Cell Levels in Health and Disease. *Cell*, 155, 242–256.
  44. Vignali, D.A.A., Collison, L.W. and Workman, C.J. (2008) How regulatory T cells work. *Nat. Rev. Immunol.*, 8, 523–532.
  45. Dimas, A.S., Deutsch, S., Stranger, B.E., Montgomery, S.B., Borel, C., Attar-Cohen, H., Ingle, C., Beazley, C., Arcelus, M.G., Sekowska, M., et al. (2009) Common Regulatory Variation Impacts Gene Expression in a Cell Type-Dependent Manner. *Science*, 325, 1246–1250.
  46. Murphy, A., J.-H.C., Xu, M., Carey, V.J., Lazarus, R., Liu, A., Szeffler, S.J., Strunk, R., DeMuth, K., Castro, M., et al. (2010) Mapping of numerous disease-associated expression polymorphisms in primary peripheral blood CD4 lymphocytes. *Hum. Mol. Genet.*, 19, 4745–4757.
  47. Raj, T., Rothamel, K., Mostafavi, S., Ye, C., Lee, M.N., Replogle, J.M., Feng, T., Lee, M., Asinovski, N., Frohlich, I., et al. (2014) Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. *Science*, 344, 519–523.
  48. Hu, X., Kim, H., Raj, T., Brennan, P.J., Trynka, G., Teslovich, N., Slowikowski, K., Chen, W.M., Onengut, S., Baecher-Allan, C., et al. (2014) Regulation of gene expression in autoimmune disease loci and the genetic basis of proliferation in CD4+ effector memory T cells. *PLoS Genet.*, 10, e1004404.
  49. Ferraro, A., D'Alise, A.M., Raj, T., Asinovski, N., Phillips, R., Ergun, A., Replogle, J.M., Bernier, A., Laffel, L., Stranger, B.E., et al. (2014) Interindividual variation in human T regulatory cells. *Proceedings of the National Academy of Sciences*, 111, E1111–E1120.
  50. Kastirr, I., Crosti, M., Maglie, S., Paroni, M., Steckel, B., Moro, M., Pagani, M., Abrignani, S. and Geginat, J. (2015) Signal Strength and Metabolic Requirements Control Cytokine-Induced Th17 Differentiation of Uncommitted Human T Cells. *The Journal of Immunology*, 195, 3617–3627.
  51. Watanabe, T., Tomoko, W., Jun-ichi, M., Yoshiaki, S., Hiroko, I., Kaori, H., Kumiko, K., Yasunori, U., Yumi, A., Shuji, K., et al. (2006) CD52 is a novel costimulatory molecule for induction of CD4 regulatory T cells. *Clin. Immunol.*, 120, 247–259.
  52. Randolph, G.J., Claudia, J. and Chunfeng, Q. (2008) Antigen presentation by monocytes and monocyte-derived cells. *Curr. Opin. Immunol.*, 20, 52–60.
  53. Fairfax, B.P., Humburg, P., Makino, S., Naranbhai, V., Wong, D., Lau, E., Jostins, L., Plant, K., Andrews, R., McGee, C., et al. (2014) Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science*, 343, 1246949.
  54. Fairfax, B.P., Seiko, M., Jayachandran, R., Katharine, P., Stephen, L., Alexander, D., Peter, E., Cordelia, L., Vannberg, F.O. and Knight, J.C. (2012) Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nat. Genet.*, 44, 502–510.
  55. Thompson, H.L. and Wilton, J.M. (1992) Interaction and intracellular killing of *Candida albicans* blastospores by

- human polymorphonuclear leucocytes, monocytes and monocyte-derived macrophages in aerobic and anaerobic conditions. *Clin. Exp. Immunol.*, **87**, 316–321.
56. Wen, A.Y., Sakamoto, K.M. and Miller, L.S. (2010) The Role of the Transcription Factor CREB in Immune Function. *J Immunol.*, **185**, 6413–6419.
  57. Nathan, C. and Carl, N. (2006) Neutrophils and immunity: challenges and opportunities. *Nat. Rev. Immunol.*, **6**, 173–182.
  58. Naranbhai, V., Fairfax, B.P., Makino, S., Humburg, P., Wong, D., Ng, E., Hill, A.V.S. and Knight, J.C. (2015) Genomic modulators of gene expression in human neutrophils. *Nat. Commun.*, **6**, 7545.
  59. Peters, J.E., Lyons, P.A., Lee, J.C., Richard, A.C., Fortune, M.D., Newcombe, P.J., Richardson, S. and Smith, K.G.C. (2016) Insight into Genotype-Phenotype Associations through eQTL Mapping in Multiple Cell Types in Health and Immune-Mediated Disease. *PLoS Genet.*, **12**, e1005908.
  60. Francesconi, M. and Lehner, B. (2013) The effects of genetic variation on gene expression dynamics during development. *Nature*, **505**, 208–211.
  61. Ye, C.J., Feng, T., H.-K, K., Raj, T., Wilson, M.T., Asinovski, N., McCabe, C., Lee, M.H., Frohlich, I., H.-i, P., et al. (2014) Intersection of population variation and autoimmunity genetics in human T cell activation. *Science*, **345**, 1254665–1254665.
  62. Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorf, L., et al. (2014) The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.*, **42**, D1001–D1006.
  63. Pennisi, E. (2011) The Biology of Genomes. Disease risk links to gene regulation. *Science*, **332**, 1031.
  64. Hindorf, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S. and Manolio, T.A. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. U. S. A.*, **106**, 9362–9367.
  65. Hawkins, R.D., Larjo, A., Tripathi, S.K., Wagner, U., Luu, Y., Lönnberg, T., Raghav, S.K., Lee, L.K., Lund, R., Ren, B., et al. (2013) Global chromatin state analysis reveals lineage-specific enhancers during the initiation of human T helper 1 and T helper 2 cell polarization. *Immunity*, **38**, 1271–1284.
  66. Roychoudhuri, R., Rahul, R., David, C., Peng, L., Yoshiyuki, W., Quinn, K.M., Klebanoff, C.A., Yun, J., Madhusudhanan, S., Eil, R.L., et al. (2016) BACH2 regulates CD8 T cell differentiation by controlling access of AP-1 factors to enhancers. *Nat. Immunol.*, **10**.1038/ni.3441.
  67. Vahedi, G., Kanno, Y., Furumoto, Y., Jiang, K., Parker, S.C.J., Erdos, M.R., Davis, S.R., Roychoudhuri, R., Restifo, N.P., Gadina, M., et al. (2015) Super-enhancers delineate disease-associated regulatory nodes in T cells. *Nature*, **520**, 558–562.
  68. Tak, Y.G. and Farnham, P.J. (2015) Making sense of GWAS: using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome. *Epigenetics Chromatin*, **8**, 57.
  69. Ocone, A., Haghverdi, L., Mueller, N.S. and Theis, F.J. (2015) Reconstructing gene regulatory dynamics from high-dimensional single-cell snapshot data. *Bioinformatics*, **31**, i89–i96.
  70. Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S. and Rinn, J.L. (2014) The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.*, **32**, 381–386.
  71. Chuffart, F., Richard, M., Jost, D., Duplus-Bottin, H., Ohya, Y. and Yvert, G. (2016) Exploiting single-cell quantitative data to map genetic variants having probabilistic effects. *bioRxiv*, 10.1101/040113.
  72. Clark, S.J., Lee, H.J., Smallwood, S.A., Kelsey, G. and Reik, W. (2016) Single-cell epigenomics: powerful new methods for understanding gene regulation and cell identity. *Genome Biol.*, **17**, 72.
  73. 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.
  74. Smallwood, S.A., Lee, H.J., Angermueller, C., Krueger, F., Saadeh, H., Peat, J., Andrews, S.R., Stegle, O., Reik, W. and Kelsey, G. (2014) Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat. Methods*, **11**, 817–820.
  75. Farlik, M., Sheffield, N.C., Nuzzo, A., Datlinger, P., Schönegger, A., Klughammer, J. and Bock, C. (2015) Single-cell DNA methylome sequencing and bioinformatic inference of epigenomic cell-state dynamics. *Cell Rep.*, **10**, 1386–1397.
  76. Nagano, T., Lubling, Y., Stevens, T.J., Schoenfelder, S., Yaffe, E., Dean, W., Laue, E.D., Tanay, A. and Fraser, P. (2013) Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature*, **502**, 59–64.
  77. Cusanovich, D.A., Daza, R., Adey, A., Pliner, H.A., Christiansen, L., Gunderson, K.L., Steemers, F.J., Trapnell, C. and Shendure, J. (2015) Epigenetics. Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science*, **348**, 910–914.
  78. Buenrostro, J.D., Wu, B., Litzenburger, U.M., Ruff, D., Gonzales, M.L., Snyder, M.P., Chang, H.Y. and Greenleaf, W.J. (2015) Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature*, **523**, 486–490.
  79. Kind, J., Pagie, L., de Vries, S.S., Nahidiazar, L., Dey, S.S., Bienko, M., Zhan, Y., Lajoie, B., de Graaf, C.A., Amendola, M., et al. (2015) Genome-wide maps of nuclear lamina interactions in single human cells. *Cell*, **163**, 134–147.
  80. Jin, W., Tang, Q., Wan, M., Cui, K., Zhang, Y., Ren, G., Ni, B., Sklar, J., Przytycka, T.M., Childs, R., et al. (2015) Genome-wide detection of DNase I hypersensitive sites in single cells and FFPE tissue samples. *Nature*, **528**, 142–146.
  81. Rotem, A., Ram, O., Shosh, N., Sperling, R.A., Goren, A., Weitz, D.A. and Bernstein, B.E. (2015) Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nat. Biotechnol.*, **33**, 1165–1172.
  82. 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.
  83. Macaulay, I.C., Haerty, W., Kumar, P., Li, Y.I., Hu, T.X., Teng, M.J., Goolam, M., Saurat, N., Coupland, P., Shirley, L.M., et al. (2015) G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. *Nat. Methods*, **12**, 519–522.
  84. Angermueller, C., Clark, S.J., Lee, H.J., Macaulay, I.C., Teng, M.J., Hu, T.X., Krueger, F., Smallwood, S.A., Ponting, C.P., Voet, T., et al. (2016) Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nat. Methods*, **13**, 229–232.