Single-cell RNA sequencing to explore immune cell heterogeneity

Efthymia Papalexi[1,2] *and Rahul Satija*[1,2]

Abstract | Advances in single-cell RNA sequencing (scRNA-seq) have allowed for comprehensive analysis of the immune system. In this Review, we briefly describe the available scRNA-seq technologies together with their corresponding strengths and weaknesses. We discuss in depth how scRNA-seq can be used to deconvolve immune system heterogeneity by identifying novel distinct immune cell subsets in health and disease, characterizing stochastic heterogeneity within a cell population and building developmental 'trajectories' for immune cells. Finally, we discuss future directions of the field and present integrated approaches to complement molecular information from a single cell with studies of the environment, epigenetic state and cell lineage.

Flow cytometry
Laser-based technology that allows for simultaneous quantification of the abundance of up to 17 cell surface proteins using fluorescently labelled antibodies.

Mass cytometry
(commercial name CyTOF). Mass spectrometry technique used as an alternative to flow cytometry that allows for the quantification of cellular protein levels by using isotopes that overcome problems associated with the spectral overlap of fluorophores.

[1]*Center for Genomics and Systems Biology, New York University, New York, NY 10003–6688, USA.*
[2]*New York Genome Center, New York, New York 10013, USA.*

Correspondence to R.S.
rsatija@nygenome.org

doi:10.1038/nri.2017.76
Published online 7 Aug 2017

From bacteria to humans, the diverse and adaptable nature of foreign threats has driven the evolution of a powerful and flexible defence response. To maintain its effectiveness, this so-called immune system has produced highly specialized (pathogen-specific) cell types that work together to prevent, retain a memory of and eliminate disease[1–4].

Single-cell resolution is therefore essential to understanding how the immune system gives rise to such a breadth of potential responses against many different pathogens[5]. Recently, new technologies have been developed that enable the profiling of single cells using next-generation sequencing, which offers an unbiased approach to studying immune cell diversity. In this Review, we present an overview of existing single-cell technologies and discuss their strengths and limitations (BOX 1). We also explore ways in which these approaches can deepen our understanding of immunological responses and disease, and we examine cutting-edge trends and potential future innovations in the field.

'Targeted' single-cell profiling technologies

A large number of techniques have leveraged advances in microscopy, cytometry, molecular biology and, most recently, next-generation sequencing to profile single cells. Many of these approaches have been developed and optimized to be used in studies that aim to deconvolve immune cell heterogeneity, but they can differ by orders of magnitude in terms of the number of cells that can be analysed per experiment (the breadth of cellular profiling) and the number of genes per cell that can be detected (the depth of cellular profiling).

'Targeted' technologies can assess a pre-selected set of molecular dimensions (pre-selected genes for mRNA expression studies and protein-level detection) across hundreds to millions of cells using known molecular baits — such as fluorescently labelled oligonucleotide probes, fluorescent or metal-conjugated antibodies, or PCR primers — to profile genes or proteins with single-cell resolution. For example, recent advances in flow cytometry[6] have allowed for the routine and simultaneous profiling of up to 17 proteins per cell using fluorescent antibodies. By using metal-conjugated antibodies to overcome the spectral limits of fluorescent proteins, mass cytometry[7] can further extend profiling to the simultaneous detection of about 40 proteins per cell, with an order of magnitude increase in the number of cells that can be studied at one time[8,9]. These technologies have led to the discovery and characterization of major and minor cell types in the mammalian immune system[10]. However, their application is limited to a small number of parameters that are selected based on prior knowledge or guesswork (such as genes or surface proteins), and the profiling of these parameters depends on the availability of gene sequences for primer design or protein-specific antibodies.

As an alternative to cytometry, gene-specific primers can be used to carry out quantitative PCR (qPCR) on single cells[11], which allows for the fluorescent quantification of single-cell mRNA levels[12,13]. Single-cell qPCR (sc-qPCR) does not require sample library preparation or deep sequencing, and it therefore offers a rapid and highly quantitative assay for single-cell gene expression, particularly in the absence of specific antibodies. Commercial microfluidic approaches have been used to multiplex up to

Box 1 | **Summary of current single-cell profiling technologies**

The available technologies for single-cell RNA sequencing (scRNA-seq) have unique strengths and weaknesses (see table). Before choosing which technology to use for a particular study, it is important to consider the scale of the experiment, the cost and sensitivity of each method and the biological question to be answered. Advances in droplet microfluidics[33–35] now enable routine profiling of thousands of cells in a single experiment. These methods are ideally suited for discovering rare cell types or deconvolving highly heterogeneous populations such as whole tissue or organ samples. However, these technologies have reduced sensitivity per cell, and they may not be able to identify subtle transcriptional differences between cells. Alternative technologies, such as plate-based protocols[29–32] or commercial microfluidics solutions (Fluidigm C1), are capable of deep profiling of single cells but at a substantially increased cost. These technologies are better suited to study stochastic variability between single cells or to deconvolve subtle transcriptomic differences in 'homogeneous' populations. In addition, plate-based methods that use index-sorting for cell isolation enable the recording of cellular immunophenotypes alongside the transcriptome, and the Fluidigm C1 allows for cells to be individually imaged before sequencing.

As these technologies mature, they suggest a powerful complementary approach, whereby complex tissues are first 'atlased' using high-breadth droplet-based technologies to identify new populations of interest and associated markers. Subsequently, these markers can be used for enrichment and deep sequencing using high-depth, plate-based approaches.

	FACS	CyTOF	qPCR	Plate-based protocols (STRT-seq, SMART-seq, SMART-seq2)	Fluidigm C1	Pooled approaches (CEL-seq, MARS-seq, SCRB-seq, CEL-seq2)	Massively parallel approaches (Drop-seq, InDrop)
Cell capture method	Laser	Mass cytometry	Micropipettes	FACS	Microfluidics	FACS	Microdroplets
Number of cells per experiment	Millions	Millions	300–1,000	50–500	48–96	500–2,000	5,000–10,000
Cost	$0.05 per cell	$35 per cell	$1 per cell	$3–6 per well	$35 per cell	$3–6 per well	$0.05 per cell
Sensitivity	Up to 17 markers	Up to 40 markers	10–30 genes per cell	7,000–10,000 genes per cell for cell lines; 2,000–6,000 genes per cell for primary cells	6,000–9,00 genes per cell for cell lines; 1,000–5,000 genes per cell for primary cells	7,000–10,000 genes per cell for cell lines; 2,000–6,000 genes per cell for primary cells	5,000 genes per cell for cell lines; 1,000–3,000 genes per cell for primary cells

CEL-seq, cell expression by linear amplification and sequencing; CyTOF, cytometry by time of flight (mass cytometry); FACS, fluorescence-activated cell sorting; InDrop, indexing droplets sequencing; MARS-seq, massively parallel single-cell RNA sequencing; qPCR, quantitative PCR; SCRB-seq, single-cell RNA barcoding and sequencing; STRT-seq, single-cell tagged reverse transcription sequencing.

96 primer pairs together in a single assay, and indeed, these approaches were shown to be extremely promising in deconvolving the molecular heterogeneity of the developing immune system[14,15]. However, similarly to cytometry-based approaches, qPCR assays also require measurement of a preselected pool of genes, which introduces bias and limits the potential for discovery of new genes and proteins of interest.

As a result, there has been substantial interest around new methods that are capable of unbiased molecular profiling of single cells by leveraging new techniques based on next-generation sequencing. The development of single-cell RNA sequencing (scRNA-seq) approaches has allowed for unbiased single-cell transcriptome profiling to enable the discovery of new cellular states, the profiling of genetic heterogeneity ranging from single nucleotide polymorphisms to diverse immunoglobulin sequences, and the study of the transcriptomes of non-model organisms.

Towards unbiased single-cell profiling

The first protocols for bulk RNA-seq offered an unbiased alternative to microarrays[16–18] but required millions of cells (~1 µg of total mRNA transcripts)[19]. Whereas some of the first immunological studies used abundant leukocyte cell populations[20,21], the need to study rare cell populations and to discover new cellular states necessitated the development of RNA-seq protocols with a lower cell input[22,23]. Particularly in the field of immunology, these new RNA-seq protocols, in combination with microarray data, allowed for the profiling of various rare cell populations with the use of only 1 ng of RNA isolated from 100–1,000 immune cells. This led to the generation of large collaborative databases, including the Illumina Body Map Expression Atlas[24]; the Differentiation Map (DMAP) project[25], which profiled 39 distinct human immune cell types; and the Immunological Genome Project, which profiled murine immune cell subsets. These databases are powerful community resources to identify modules of co-regulated genes across many cell types and conditions for cellular subsets with well-defined markers.

The development of low-input RNA-seq protocols paved the way for further optimization down to the single-cell level, culminating in an explosion of new scRNA-seq platforms. With the large number of methods available, each with distinct strengths and weaknesses, it is often unclear which option is most suitable for addressing a specific research question. Here, we review many of the available options and discuss how they differ in terms of workflow (FIG. 1), sensitivity and data quality, in addition to outlining their ideally suited biological applications (BOX 1).

Quantitative PCR
(qPCR). Polymerase chain reaction used to quantify gene expression levels using fluorescently labelled nucleotides and by tracking fluorescence levels during amplification cycles.

Microfluidic approaches
Single-cell RNA-sequencing techniques that use microfluidic devices for single-cell isolation.

Microarrays
Technique used to detect gene expression levels of many genes simultaneously. Microarrays use gene-specific probes that can be hybridized to complementary fluorescently labelled cDNA molecules. The fluorescence intensity is used to quantify gene expression.

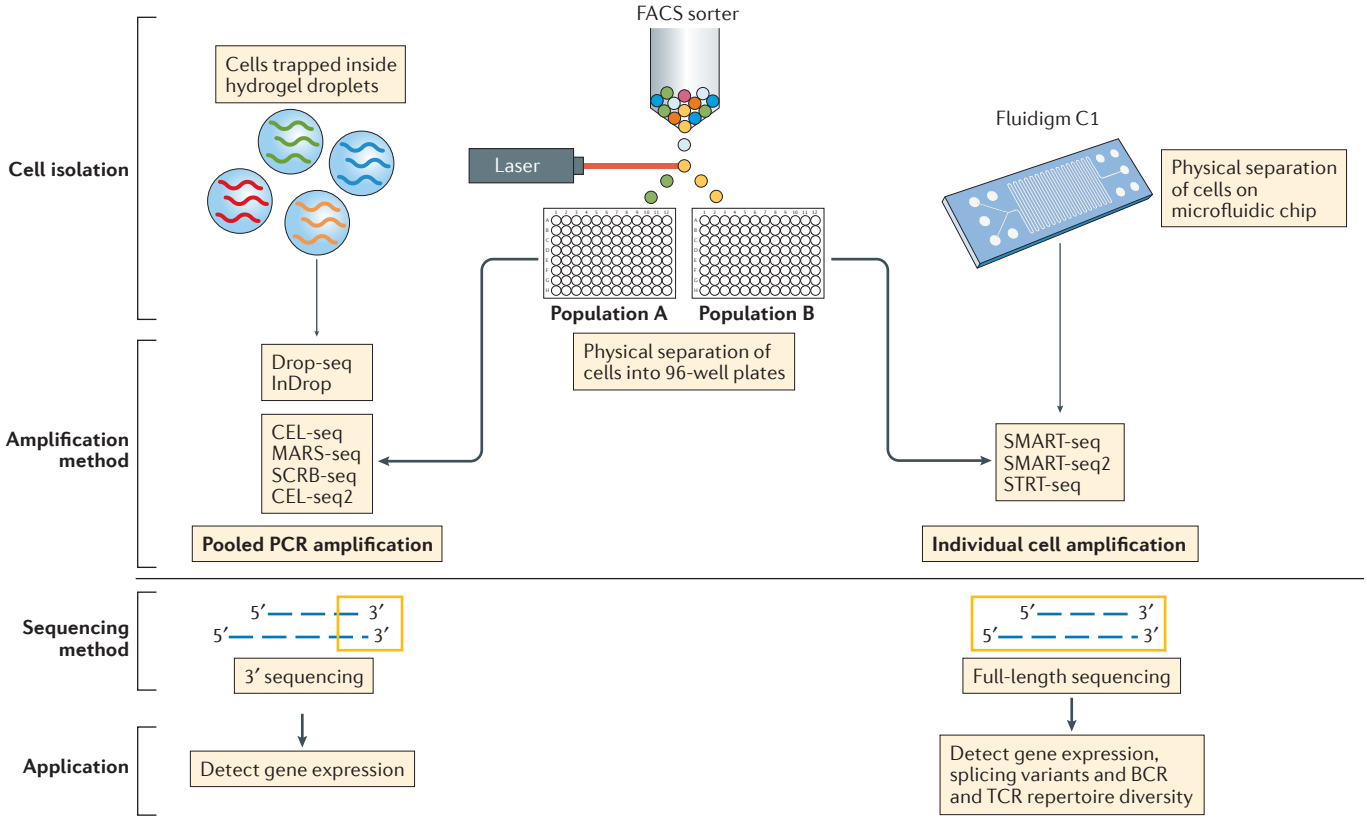

Figure 1 | **Overview of scRNA-seq technologies.** Single-cell RNA sequencing (scRNA-seq) technologies use many different methods for cell isolation and transcript amplification. Whereas some technologies capture cells using microfluidic devices that trap cells inside hydrogel droplets, other technologies rely on methods (such as fluorescence-activated cell sorting (FACS) into 96-well plates and the microfluidic chips used by Fluidigm C1) that physically separate one cell from another in wells. Once cells are lysed, reverse transcription and PCR amplification are carried out. Droplet-based approaches, and some plate-based approaches, allow for pooled PCR amplification using cellular barcoding techniques, which decreases the cost as only one PCR reaction is required per experiment or plate. In other plate-based approaches and for Fluidigm C1, the number of PCR amplification reactions is equal to the number of cells that are being profiled, which makes these approaches expensive. PCR products are further processed to prepare samples for sequencing. Some approaches that use sequencing of the 3′ end of each transcript allow for quantification of expression of each gene within a cell. Other approaches, however, can sequence full-length transcripts, which allows not only for detection of gene expression but also for analysis of splicing variants and B cell receptor (BCR) or T cell receptor (TCR) repertoire diversity. InDrop, indexing droplets sequencing; MARS-seq, massively parallel single-cell RNA sequencing.

Plate-based protocols. Most 96-well protocols, such as single-cell tagged reverse transcription (STRT) sequencing (STRT-seq), SMART-seq and SMART-seq2 (REFS 23,26–28), use micropipettes or fluorescence-activated cell sorting (FACS) to place individual cells into wells containing lysis buffer. These platforms offer a fast and efficient way to analyse 50 to 500 single cells in one experiment. Single cells can be stored in plates long-term before analysis, allowing for a flexible experimental set-up with optional pause points when time is limited. However, reverse transcription is carried out on individual wells, which requires additional pipetting steps that can slow down the process and introduce technical noise in the samples. In addition, the early versions of these platforms had low sensitivity and were quite costly. Subsequent studies optimized this platform to increase accuracy, sensitivity and throughput, as well as to decrease processing time. Moreover, these protocols are amenable to automation with liquid-handling robotics. These methods are generalizable, as they offer the opportunity to profile any cell, independent of size and type, that can pass through a micropipette or FACS sorter machine. Overall, they have high sensitivity and can measure 5,000–10,000 genes per single cell.

Fluidigm C1. In 2012, Fluidigm introduced the C1, an automated microfluidic platform for scRNA-seq that can individually capture up to 96 cells at a time on a single microfluidic chip. Downstream molecular steps are automated and parallelized in nanolitre-sized volumes. In addition, this platform offers the option to evaluate the captured cells under the microscope before the reverse transcription and amplification steps of the protocol. At least 10,000 cells are required as input, which suggests that this platform is not ideal for rare cell populations. To avoid introducing selection bias, it is required that cells be of similar size and shape. The sensitivity of the Fluidigm C1 is similar to that of plate-based protocols,

Reverse transcription
Conversion of a mRNA molecule to complementary DNA (cDNA) using reverse transcriptase enzymes isolated from RNA viruses.

and this method works well with homogeneous cell populations. Although the microfluidic platform enables savings in molecular reagents and labour, the cost of the microfluidic chips is substantial and limits the feasibility of large-scale experiments.

Pooled approaches. The approaches described above leverage either automation or microfluidics to reduce costs and improve throughput. However, the even simpler idea of applying a barcode to cells at early stages and then carrying out downstream molecular steps in parallel was first introduced in the cell expression by linear amplification and sequencing[29] (CEL-seq) protocol. This approach markedly decreased reagent and labour costs while simultaneously increasing the scale of each experiment (500–2,000 cells per run). In the CEL-seq protocol, a unique primer (containing a poly T tract, a cell barcode, a 5′ Illumina sequencing adaptor and a T7 promoter) is introduced into each cell during reverse transcription. By introducing these unique cell barcodes, all cDNAs can be pooled together after reverse transcription, and a single amplification reaction can be carried out. In a protocol known as massively parallel scRNA-seq (MARS-seq),[30] this idea has been extended by combining single-cell barcoding with 384-well-plate FACS sorting to increase the scale and lower the associated costs. Combining FACS sorting with single-cell barcoding ensures successful physical separation of single cells into wells (FACS sorting) while lowering costs by allowing reactions to be pooled and processed as one sample in later steps of the protocol. This strategy was quickly adopted by many early forms of plate-based approaches and resulted in further optimized versions with higher sensitivity and lower costs (single-cell RNA barcoding and sequencing (SCRB-seq)[31] and CEL-seq2 (REF. 32)). In summary, these single-cell barcoding strategies offer an unbiased method for isolating various cell types, improving throughput and lowering the costs of experiments.

Massively parallel approaches. The development of microfluidics and reverse emulsions devices allowed for isolation of single cells into droplets containing lysis buffer and cellular barcodes. These methods not only allowed for unbiased cell capture, as with FACS-sorted plate-based approaches, but also used massive parallelization to increase the number of cells that could be profiled in one run to tens of thousands. This pioneering approach was first exemplified by two academically developed technologies, known as Drop-seq[33] and inDrop (indexing droplets sequencing)[34], and has been further developed into commercially available platforms as well[35]. However, the increased breadth of these experiments comes with reduced sensitivity. In contrast to other existing methods, droplet-based methods typically have reduced transcript recovery (3–10% compared with 10–20% for other methods). We anticipate that the sensitivity of Drop-seq and similar methods will increase as protocols continue to be optimized and sequencing costs continue to decrease[36], which will enable an increase in sequencing depth.

General considerations and ongoing limitations. The methods for RNA-seq described here differ widely in associated costs, experimental scale, single-cell isolation methods, and data quality and sensitivity (BOX 1). It is advisable to consider each of these parameters carefully before choosing the appropriate method to use. For example, is it more powerful to sequence a large number of cells (high breadth) at low coverage or to sequence a smaller number of cells very deeply (increase the number of genes recovered per cell)? Studies aiming to identify cell clusters that can be defined by many genes, with an emphasis on finding rare cell populations, should prioritize a breadth-based approach, whereas studies aiming to distinguish stochastic variation in individual genes require a high depth of sequencing. Furthermore, potential phenotypic differences between cells may also drive the choice of technology. For cells with substantial differences in size and shape, FACS sorting or droplet-based methods can be used for cellular profiling owing to the equal probability for collecting different populations, whereas commercial microfluidic approaches such as the Fluidigm C1 may bias the population examined.

Despite improvements in terms of cost and scale for traditional scRNA-seq, molecular limitations remain. The above methods fail to capture non-polyadenylated RNA transcripts, because non-coding RNAs (such as microRNAs, long non-coding RNAs and circular RNAs)[37] and bacterial RNAs[38] are discarded during traditional poly T priming of reverse transcription. In addition, whereas protocols with an individual amplification strategy (FIG. 1) enable sequencing of the full transcripts, high-throughput multiplexed methods sequence only the 3′ end and cannot recover splicing patterns or sequence variants. Moreover, even the most sensitive methods will struggle to detect low-abundance transcripts, which is a limiting factor when exploring more subtle differences between cell subsets[39]. Finally, transcriptomic measurements between cells cannot capture the proteomic or epigenetic heterogeneity that may drive cellular behaviour, and thus, scRNA-seq results describe only a subset of the molecular phenotype of a cell. Although these limitations pose challenges for molecular technology and nanotechnology, rapid developments in the field are beginning to address these concerns, yielding sensitive, global and integrated technologies for single-cell profiling (discussed later).

Applications of scRNA-seq

Before the development of scRNA-seq technologies, the discovery of new cell subsets involved using cell surface markers. Although these approaches were powerful, they required prior guesswork or knowledge in order to discover various immune cell types. The development of high-dimensional, single-cell technologies enabled an unbiased, alternative workflow that allowed sequencing of cells without prior knowledge of genes and proteins of interest and grouping of cells based on their transcriptional signatures. In this section, we present studies that have used scRNA-seq to characterize homogeneous immune cell populations in health

Barcode
A 12–20 nucleotide sequence that is uniquely assigned to a cell during reverse transcription and is used to trace mRNA transcripts back to their cellular origins.

Reverse emulsions devices
Devices that create oil-in-water emulsions, producing droplets that can encapsulate single cells.

Chromatin immunoprecipitation-sequencing (CHIP-seq). A technique that uses crosslinking of protein–DNA interactions and sequencing to identify protein-binding patterns and motifs on DNA.

and disease, discover the variation in stochastic gene expression that drives immunological responses and reconstruct developmental trajectories for immune cells.

Characterizing distinct cell subsets. The technologies described above provide powerful approaches for deconvolving heterogeneity in the immune system, enabling the discovery of pathogenic immune cell populations. Although many exciting studies have emerged recently, here, we discuss in depth three pioneering examples that highlight how the unbiased potential of scRNA-seq can be used for the discovery of cellular states in health and disease across diverse systems.

In 2014, Jaitin *et al.*[30] developed MARS-seq to dissect mouse splenic cellular diversity (FIG. 2a). Using hierarchical clustering and the probabilistic mixture model, they classified splenic cells into transcriptionally distinct groups. They validated these *in silico* predictions by comparing these groups to existing transcriptional profiles of classical haematopoietic cell populations, and found their groups to be transcriptionally similar to B cell, natural killer (NK) cell, macrophage, monocyte and plasmacytoid dendritic cell populations. Using lipopolysaccharide (LPS) stimulation to mimic infection, they studied the immediate responses of various splenic subpopulations. Analysis of their transcriptional profiles revealed groups of cell type-specific response genes, as well as many type I interferon (IFN) response genes that were highly expressed in all subpopulations. Furthermore, they identified dendritic cell (DC) subpopulations with distinct transcriptional profiles, which supports the idea of cellular state diversity within a cell population. Finally, they proposed that, in response to LPS stimulation, many immune cell types can preserve their identity and respond to infection by activating cell type-specific, as well as more generic, transcriptional programmes. Overall, these findings provided an exciting vision of how we can discover and re-annotate cell types without any prior knowledge using high-throughput single-cell sequencing.

Additional studies have continued to uncover previously unknown heterogeneity of CD127$^+$ innate lymphoid cells (ILCs) in human tonsil and small intestine[40] (FIG. 2b). Using scRNA-seq, Bjorklund *et al.* found four distinct ILC clusters with transcriptional profiles corresponding to previously characterized ILC1, ILC2, ILC3 and NK cell populations (based on surface marker expression)[40]. More importantly, they identified previously hidden transcriptional signatures within these populations, which suggests the existence of functionally distinct subpopulations of cells. Whereas Bjorklund *et al.* focused on tonsil-derived ILCs, Gury-BenAri *et al.*[41] focused on helper-like ILCs in the mouse small intestine and tried to assess their heterogeneity using scRNA-seq together with chromatin immunoprecipitation-sequencing (CHIP-seq) and assay for transposase-accessible chromatin-sequencing (ATAC-seq). Transcriptomic

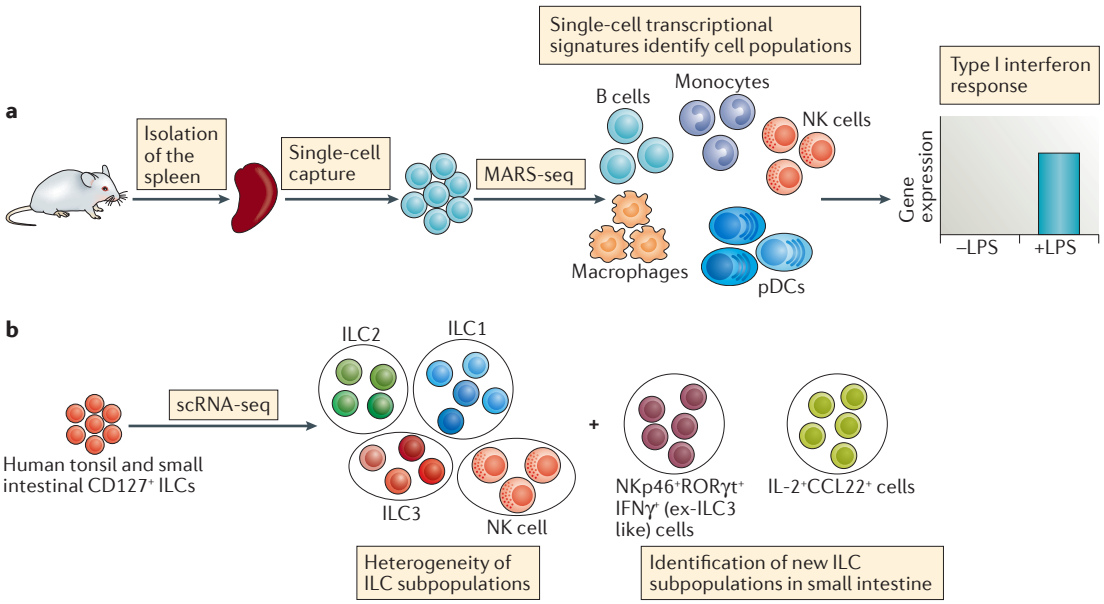

Figure 2 | **scRNA-seq uncovers distinct cell subsets in the healthy immune system. a** | Mouse splenic cellular diversity was dissected using the massively parallel single-cell RNA sequencing (MARS-seq) protocol. B cell, monocyte, natural killer (NK) cell, macrophage and plasmacytoid dendritic cell (pDC) populations were identified based on single-cell transcriptional signatures. Further analysis of the transcriptional profiles of pDCs showed that there is heterogeneity within the population. Finally, stimulation with lipopolysaccharide (LPS) to mimic viral infection induced the expression of type I interferon (IFN) response genes in all cell types, which suggests that these splenic cell populations respond to viral infection by upregulating antiviral genes of the type I IFN response[30]. **b** | Human tonsil and small intestine CD127$^+$ innate lymphoid cells (ILCs) include all of the previously characterized ILC subpopulations (ILC1, ILC2 and ILC3) and NK cells. Single-cell RNA sequencing (scRNA-seq) allowed for the detection of further heterogeneity of ILC subpopulations in human tonsil cells[40] as well as for the identification of two new small intestine ILC subpopulations marked by high levels of expression of NKp46, retinoic acid receptor-related orphan receptor-γt (RORγt) and interferon-γ (IFNγ) (ex-ILC3 like cells) or IL-2 and CCL22 (REF. 41).

analysis of CD127$^+$ ILCs revealed heterogeneity within known ILC subsets (ILC1, ILC2 and ILC3) as well as the existence of two previously unidentified ILC subsets that expressed NKp46, retinoic acid receptor-related orphan receptor-γt and IFNγ (ex-ILC3–like cells) or IL-2 and CCL22. Notably, these results highlighted the importance of the microbiome in shaping the cellular diversity of ILCs in the small intestine by showing that all ILC subsets in the intestine of antibiotic-treated and germ-free mice acquired ILC3-like expression profiles.

These preliminary studies in healthy tissue paved the way for profiling of such tissues in a disease context, which allowed for the identification of molecular drivers of disease in pathogenic cell subsets. For example, Gaublomme *et al.*[42] used scRNA-seq to identify the T helper 17 (T_H17) cell states that drive the pathogenesis of experimental autoimmune encephalomyelitis (EAE, a model for human multiple sclerosis) in mice[42] using *in vivo* and *in vitro* models (FIG. 3a). They profiled T_H17 cells after *in vivo* and *in vitro* EAE induction and discovered that these cells were highly heterogeneous. Comparative analysis of *in vivo*- and *in vitro*-isolated pathogenic T_H17 cells revealed wide spectrums of pathogenicity that were similar but not identical to each other. They identified a transcriptional signature that highly correlated with the most pathogenic T_H17 cells, and further computational analysis revealed the candidate genes that most likely drive T_H17 cell pathogenicity (*Gpr65*, *Plzp*, *Toso* and *Cd5l*). The discovery and validation of these candidate genes *in vivo* has opened a new window for the potential development of more effective therapeutic agents for the treatment and cure of multiple sclerosis.

In summary, these studies have established that surface phenotypes are not sufficient to define cellular states in disease and have proposed new scRNA-seq methods to study innate immunological processes as well as disease pathogenesis and progression at high resolution. Uncovering the key regulators of immune responses and pathogenicity can markedly contribute to the discovery and development of new therapeutic agents targeting immunological diseases such as multiple sclerosis. We anticipate that in the near future scRNA-seq will be used for the discovery of novel haematopoietic progenitor cell populations, the identification of additional distinct immune cell subsets that drive disease and the development of an 'atlas' of immune cell types and states.

Characterizing the heterogeneity of a population. Stochastic patterns of gene expression among cells within a 'homogeneous' population might be at the core of how the immune system can produce such a breadth of responses to maintain homeostasis and battle infections[43,44]. Evidence that stochastic heterogeneity provides response breadth has been previously provided by analysing surface marker expression[45]. The development of single-cell genomics methods raises the exciting possibility of carrying out these types of studies in a genome-wide manner[46] to uncover unexpected and potentially stochastic variability within immune cell populations.

Molecular stochasticity within a cell type is particularly relevant for B and T cells, which use V(D)J recombination to generate diverse B cell receptor (BCR) and T cell receptor (TCR) chains that allow them to recognize a wide variety of peptide–MHC combinations

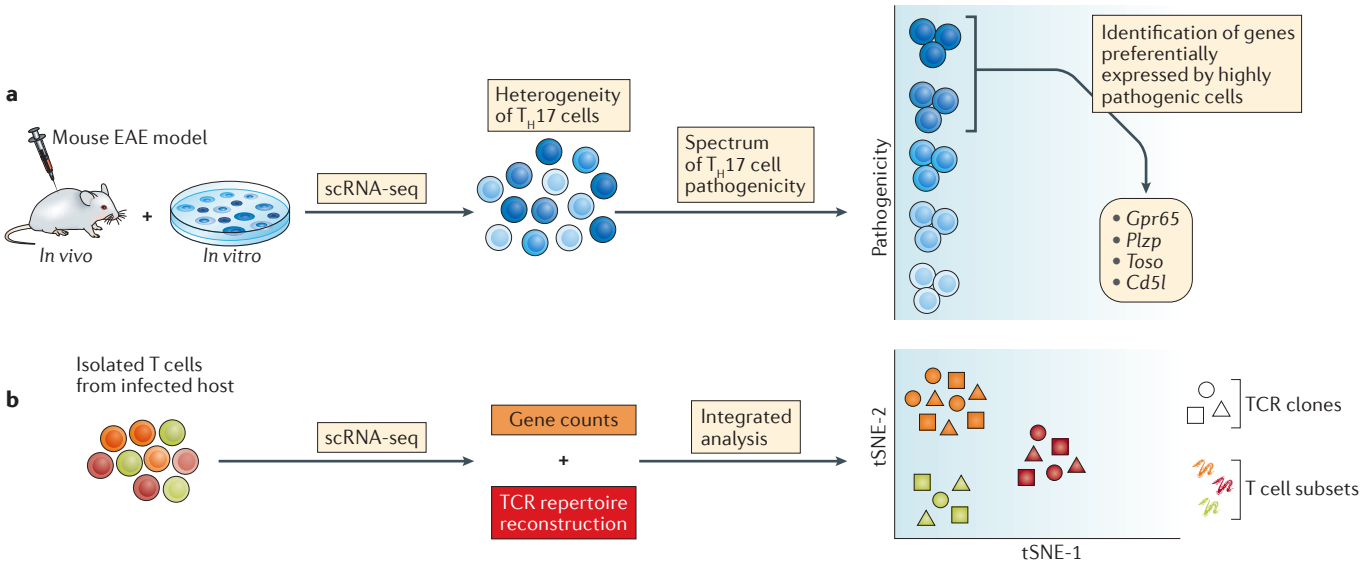

Figure 3 | **Single-cell profiling uncovers distinct cell subsets in disease. a** | Mouse *in vitro* and *in vivo* models of experimental autoimmune encephalomyelitis (EAE) recapitulate features of the human disease multiple sclerosis. T helper 17 (T_H17) cells have an important role in EAE pathogenesis. Single-cell RNA sequencing (scRNA-seq) analysis of T_H17 cells in EAE showed that there is a spectrum of T_H17 cell pathogenicity ranging from non-pathogenic to highly pathogenic cells. *Grpr65*, *Plzp*, *Toso* and *Cd5l* are expressed in highly pathogenic cells, and these markers might therefore be used for diagnosis and to design new therapeutic approaches for multiple sclerosis[42]. **b** | scRNA-seq and T cell receptor (TCR) repertoire reconstruction can be used to infer changes in T cell clonality and transcriptional profiles in response to various infections[47,49,50]. tSNE, t-distributed stochastic neighbour embedding.

on antigen-presenting cells. Intriguingly, paired and full-length TCR and BCR sequences can be read from full-length scRNA-seq data, as these genes are highly expressed (FIG. 3b). Stubbington *et al.*[47] developed a computational method known as TRaCer to detect TCR heterodimeric diversity from full-length mRNA scRNA-seq data. TRaCer can be used to extract TCR-derived sequencing reads for individual cells and map them against a TCR reference pool that contains all possible combinations of V and J segments. Combining transcriptome sequencing with TCR reconstruction has allowed multiple groups to make associations between lymphocyte clonality and heterogeneous responses to infection. For example, Stubbington *et al.* used this approach to identify changes in T cell clones by comparing samples from before, during and after infection with *Salmonella enterica* subsp. *enterica* serovar Typhimurium. They found a clonotype expansion of $CD4^+$ T cells during infection, with each clonotype carrying TCR sequences that are likely to be specific for an *S.* Typhimurium antigen. Lönnberg *et al.*[48] used a similar approach to reconstruct the bifurcation of mouse T_H1 cell versus T follicular helper cell fates in response to malaria, and the results demonstrated that individual clones populated both fates. Indeed, multiple computational approaches now exist for TCR repertoire reconstruction, including single-cell TCR sequencing (scTCR-seq)[49] and the TCR reconstruction algorithm for paired-end single cell (TRAPeS)[50], and similar approaches should work for BCR reconstruction as well. Although this strategy is currently limited to those profiling methods that sequence full-length RNA such as SMART-seq2 (FIG. 1), we anticipate that new molecular methods will soon enable paired transcriptomic and immune repertoire profiling in massively parallel and multiplexed assays.

Shalek *et al.*[51] carried out scRNA-seq analysis of bone marrow-derived DCs (BMDCs) after LPS stimulation to study variations in gene expression and splicing patterns among BMDCs in response to infection (FIG. 4a). The results showed significant differences in mRNA abundance of LPS-response pathway genes between cells. This finding is important, as it suggests that the observed heterogeneity has functional consequences for each cell. This transcriptional heterogeneity may give BMDCs the breadth or flexibility to respond appropriately to numerous types and levels of infection. In addition, this was the first report of heterogeneity in the splicing patterns of mRNAs between single cells. Further scRNA-seq data analysis revealed a cluster of approximately 100 genes — including many anti-viral genes, among which were the antiviral master regulator genes *Irf7* and *Stat2* — that are bimodally expressed in BMDCs in response to LPS stimulation. This finding suggests that LPS stimulation promotes variable *Irf7* and *Stat2* activation, which in turn induces bimodal expression of numerous antiviral genes. A follow-up study from the same group[52] showed that only a small subset of BMDCs expresses antiviral genes during the early stages of infection, whereas during the late stages of infection, these genes are uniformly expressed by all BMDCs. The 'early responder' BMDCs are responsible for sensing the infection and then

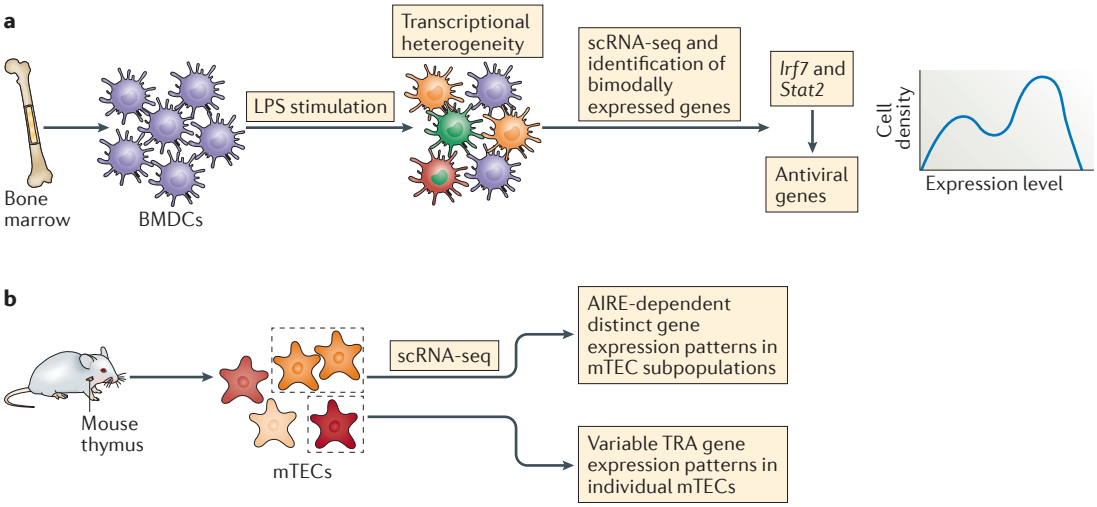

Figure 4 | **Characterizing heterogeneity within one immune cell population using scRNA-seq.** **a** | Bone marrow-derived dendritic cells (BMDCs) respond to infections and help the immune system recruit other cell types to combat these infections and stop them from spreading. Lipopolysaccharide (LPS) stimulation is used as a technique to mimic infections *in vivo*. Single-cell RNA sequencing (scRNA-seq) analysis of LPS-stimulated BMDCs revealed variation in antiviral gene expression and mRNA splicing patterns of single BMDCs. Upon stimulation, BMDCs have bimodal expression of the antiviral master regulator genes *Irf7* and *Stat2*, which in turn promotes the bimodal expression of many other antiviral genes[51,52]. **b** | Medullary thymic epithelial cells (mTECs) stochastically express tissue-specific self-antigens (tissue-restricted antigens, TRAs) to mediate immune system self-tolerance. Single-cell analysis of mTECs revealed distinct TRA expression patterns. In addition, it allowed for the identification of distinct autoimmune regulator (AIRE)-dependent gene expression patterns in mTEC subpopulations. This variability in AIRE-dependent genes and TRA expression patterns might be the mechanism by which mTECs achieve self-tolerance to multiple tissues[54].

signalling to the rest of the BMDCs to react similarly. Based on this result, Shalek *et al.* proposed that dynamic regulation of the activation of signalling circuits in response to infection (for example, transcription factors that activate specific antiviral genes based on the type of infection) is the general mechanism that the immune system uses to respond appropriately to multiple threats. Overall, these studies provide proof that single-cell technologies can be used to discover networks of cells that influence each other through intercellular circuits and paracrine signalling.

Stochastic expression of self-antigens in medullary thymic epithelial cells (mTECs) is a strategy used by the immune system to achieve the broad diversity of antigen expression that is required to mediate self-tolerance in the thymus. Two independent groups have used single-cell profiling to understand how mTECs regulate antigen expression to maintain immunological self-tolerance. Focusing on different aspects of this phenomenon, they discovered that mTECs have distinct ectopic tissue-restricted antigen (TRA) expression patterns and that the transcription factor known as autoimmune regulator (AIRE) induces the expression of distinct TRA gene clusters in mTEC subpopulations, which might account for the observed variation in self-antigen expression patterns (FIG. 4b). Specifically, Brennecke *et al.*[53] used scRNA-seq to characterize the transcriptomes of mTECs, and the results showed that these cells are highly heterogeneous and have variable TRA gene expression patterns. Closer analysis revealed the presence of distinct TRA co-expression patterns in single cells and led to the hypothesis that regulation of TRA genes follows discernible patterns. Finally, ATAC-seq data showed that co-expressed genes are found in close proximity in the genome and that TRA-expression loci show increased chromatin accessibility. Meredith *et al.*[54] carried out scRNA-seq on wild-type and AIRE-deficient mice to show that AIRE regulates the expression of a specific set of genes. They also presented additional evidence supporting the idea that AIRE-dependent target genes are expressed at low frequencies and that AIRE-induced TRA-related genes are located in close proximity in the genome, in agreement with a prior report by Brennecke *et al.*[53]. Dimensional reduction analysis revealed that the newly discovered AIRE-dependent gene clusters are expressed in distinct mTEC subpopulations in wild-type mice. Moreover, by looking at the correlation between DNA methylation at CpG dinucleotides and AIRE expression, Meredith *et al.* found that wild-type and AIRE-deficient mice have highly correlated methylation patterns, suggesting that AIRE does not alter DNA methylation at CpG dinucleotides near the AIRE-dependent gene clusters in mTECs. In summary, these two studies showed the existence of extremely variable TRA expression patterns in single mTECs and highlighted the role of AIRE in controlling the expression of distinct TRA-related gene clusters through an unknown mechanism.

The above-described findings highlight the importance of immune cell variability as a mechanism for coping with different types of infection and regulating immunological self-tolerance. Further applications of single-cell genomics in additional cell populations of the immune system will broaden our understanding of systemic responses to infection and the pathogenesis of many autoimmune diseases.

Dissecting cell fate branch points. Developmental processes are driven by a series of transcriptional changes that allow for cell differentiation and commitment to a specific lineage and eventual cell type. Making use of the ability to detect discrete cell subtypes using single-cell analysis, studies have shown that developmental processes can be represented as a continuum of transitional cell states. Therefore, capturing cells in an unbiased way across multiple developmental stages and then reconstructing their developmental progression provides a unique methodology to study cellular decision-making and differentiation. Such methodology was proposed even before the emergence of single-cell genomics. For example, Bendall *et al.*[55] developed an algorithm known as Wanderlust to reconstruct the B cell developmental trajectory at extremely high resolution from CyTOF (mass cytometry) data, which uncovered coordination points along this trajectory where rewiring of major signalling pathways and changes in the expression of surface proteins mark the transition from one cell state to the next. Similar ideas have been powerfully applied to sc-qPCR data on haematopoietic stem cells (HSCs), enabling reconstruction of the HSC differentiation hierarchy, identification of the earliest HSC differentiation events, and detection of a distinct cellular hierarchy in MLL-AF9 type acute myeloid leukaemia.

scRNA-seq is an exciting extension of this methodology, providing access to rich data on molecular phenotypes that extend beyond surface markers. Trapnell *et al.*[56] developed Monocle, an algorithm for single-cell trajectory reconstruction from RNA-seq data, and showed the potential of this approach for understanding how molecular heterogeneity influences cell fate, in particular in relation to muscle development. Here, we discuss a growing field of exciting studies that apply scRNA-seq to studying mammalian immune system development and haematopoiesis, together with complementary approaches.

Schlitzer *et al.*[57] focused on dissecting the cell fate decision that is made by DC progenitors when committing to either the cDC1 or cDC2 lineage. The authors profiled a total of 250 FACS-sorted single cells belonging to three different DC precursor groups — macrophage and DC precursors, common DC precursors (CDPs) and pre-DCs — then reconstructed a developmental trajectory from the transcriptomic data and identified a bifurcation point that corresponds with the emergence of transcriptomically 'primed' progenitors. They next identified a set of genes whose expression level changed with fate choice. In particular, Siglec-H and Ly6C were identified as markers of cDC1-primed (Siglec-H$^-$Ly6C$^-$) and cDC2-primed (Siglec-H$^-$Ly6C$^+$) pre-DC subpopulations, respectively, and these predictions were validated with *in vitro* experiments. Indeed, individual fate decisions and developmental

cDC1 or cDC2 lineage
Functionally distinct conventional dendritic cell subgroups characterized by high levels of expression of the surface markers CD8α and CD103 (cDC1) or CD4 and CD11b (cDC2).

trajectories are now routinely profiled with scRNA-seq, and such studies have led to a deeper understanding of the regulators of early myeloid[58], lymphoid[59] and megakaryocytic differentiation[60], as well as T cell commitment to helper[48], cytotoxic or effector states[61].

Whereas the studies described above focused primarily on individual fate decisions, other studies have described the pioneering use of scRNA-seq for the broad profiling of haematopoietic progenitors. Paul *et al.*[62] applied scRNA-seq to 2,730 bone marrow cells ($KIT^+SCA1^-Lineage^-$). In contrast to the expected homogenous populations of common myeloid progenitors (CMPs), megakaryocyte–erythroid progenitors (MEPs) and granulocyte–macrophage precursors, the transcriptome data showed 19 cell clusters that represented either distinct or transitional cellular states; specifically, the data showed that individual CMPs are largely transcriptionally committed to a distinct myeloid fate. In addition, the results demonstrated the involvement of the transcription factors *Cebpa* and *Cebpe* in determining granulocyte–monocyte and neutrophil specification, respectively, and that perturbations of *Cebpa* and *Cebpe* lead to a haematopoietic block. Nestorowa *et al.*[63] used a similar approach to profile 1,600 haematopoietic stem and progenitor cells in mouse bone, and identified dynamic changes in gene expression among single cells. This allowed the authors to reconstruct a map showing the differentiation trajectories of HSCs and progenitor cells. Grover *et al.*[64] found that as HSCs age, they become highly biased towards megakaryocyte and platelet differentiation, with limited potential to give rise to other lineages. Moreover, this phenotype can be rescued by deletion of the platelet transcription factor FOG1.

Overall, the above studies show the power of scRNA-seq in reconstructing lineage trajectories and branching points and identifying previously unknown transcription factors that control transitions from one cellular state to the next in immune system development (FIG. 5). This methodology can be easily extended from reconstructing developmental trajectories in haematopoiesis to reconstructing developmental trajectories of immune cells in any organism in both healthy and disease states, which will broaden our understanding of the true intermediates in blood differentiation, the lineage relationships between different subpopulations, and the developmental checkpoints and blockades that accompany disease. In addition, as these methods become more widely used to dissect developmental processes, we will start to shed light on broader biological questions about the nature of cell differentiation. Some researchers view differentiation as a series of discrete stages leading to lineage commitment and cell type specification, whereas others view this process as a continuum of cell states that gradually lose stem cell identity while deciding on their ultimate fate. The studies discussed above provide evidence for both of these ideas. For example, scRNA-seq studies of HSCs have described previously uncharacterized transitional developmental cell states[62–64], whereas other studies have identified distinct CDP, HSC, MEP and myeloid subpopulations with varying differentiation potential[57–60]. These findings suggest a more complex model whereby the continuum of cellular states functions as a bridge to connect discrete differentiation stages and ensures a smooth transition from one stage to the next.

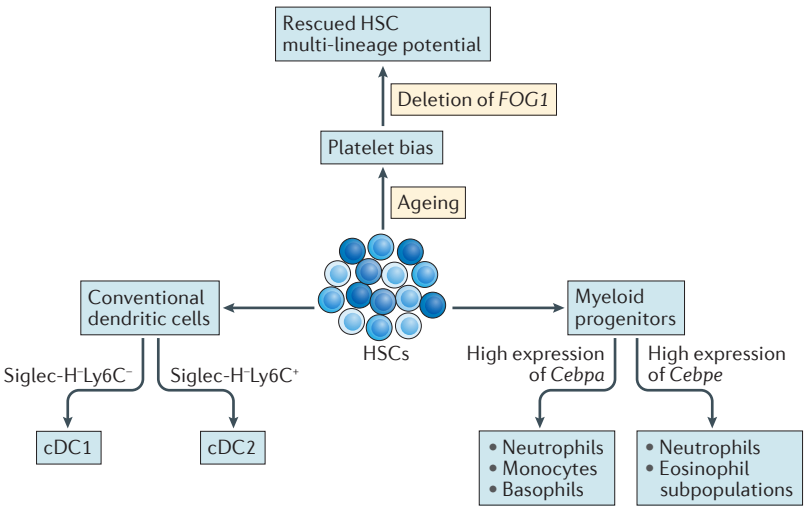

Figure 5 | **scRNA-seq helps identify cell fate branch points during HSC differentiation.** The expression of transcription factors and abundance of surface proteins determine cell fate specification during differentiation. As haematopoietic stem cells (HSCs) age, they become increasingly platelet biased; however, this phenotype can be rescued *in vitro* by deletion of the major platelet transcription factor FOG1 (REF. 64). Myeloid progenitor potential to generate mast cells and eosinophils or monocytes and macrophages relies on the presence or absence of GATA1, respectively[85]. In addition, high levels of expression of the transcription factor *Cebpa* direct myeloid progenitors towards neutrophil, monocyte and basophil lineages, whereas high levels of expression of *Cebpe* seem to be found primarily in neutrophil and eosinophil subpopulations[62]. Finally, conventional dendritic cells (cDCs) rely on Siglec-H and Ly6C abundance to determine whether they will become cDC type 1 (cDC1) or cDC type 2 (cDC2)[57].

Emerging directions for single-cell profiling

Specialized cell types allow the immune system to achieve a wide range of responses in health and disease. We have described how scRNA-seq analysis can be used as a tool for unbiased discovery of unidentified cell types, cell states and biologically meaningful cellular heterogeneity, as well as for reconstructing lineage progression during various developmental processes of the immune system. Such discoveries are now not only possible but routinely made. We anticipate that cutting-edge advances in single-cell technologies, allowing for the integrated analysis of immune repertoires and molecular state, will deepen our understanding of lymphocyte behaviour, particularly as these approaches can be scaled to larger datasets. Moreover, although cell culture and mouse models have been extremely useful in helping us understand how the immune system operates, tools are now in place for the profiling of human tissues, which will allow for analysis of both the healthy human immune system as well as the immune response in many poorly characterized autoimmune and inflammatory diseases (such as rheumatoid arthritis, Crohn's disease and psoriasis).

As the field progresses, we envision important advances in the ability to integrate diverse phenotypic parameters of a cell together with its transcriptome.

For example, in the bone marrow, cellular localization has an important role in downstream fate choice[65–68], but information about cellular positioning and the microenvironment is lost when carrying out scRNA-seq analyses. To address this issue, the development of new computational and experimental methods will allow us to reconstruct the spatial organization of cells within an embryo or a tissue. For example, computational strategies[69,70] can infer cellular localization in zebrafish embryos and annelid brains by integrating scRNA-seq data with *in situ* RNA expression patterns. Similarly, fluorescent *in situ* sequencing (FISSEQ[71]) of mRNA isolated from different cell types and tissues allows for enrichment of context-specific transcripts while preserving tissue architecture and enabling detailed analysis of RNA localization. In addition, combining scRNA-seq with HSC lentiviral barcoding strategies[72,73] can be used to integrate lineage and transcriptome information from the same cells.

Lastly, the development of new technologies is extending single-cell profiling beyond the transcriptome, with particular advances in genomic[74], chromatin[75,76], methylation[77–80] and proteomic[81] assays. Particularly exciting are the strategies being developed to multiplex these measurements together, which enables joint profiling of multiple molecular modalities from the same cell (for example, genome plus transcriptome[82,83] or immunophenotype plus transcriptome[84]). These integrated strategies will continue to allow us to discover and define rich cellular phenotypes and to explore their function in the immune system in both health and disease.

1. Germain, R. N. Maintaining system homeostasis: the third law of Newtonian immunology. *Nat. Immunol.* **13**, 902–906 (2012).
2. Littman, D. R. & Rudensky, A. Y. Th17 and regulatory T cells in mediating and restraining inflammation. *Cell* **140**, 845–858 (2010).
3. Wing, K. & Sakaguchi, S. Regulatory T cells exert checks and balances on self tolerance and autoimmunity. *Nat. Immunol.* **11**, 7–13 (2010).
4. Sakaguchi, S., Miyara, M., Costantino, C. M. & Hafler, D. A. FOXP3$^+$ regulatory T cells in the human immune system. *Nat. Rev. Immunol.* **10**, 490–500 (2010).
5. Proserpio, V. & Lönnberg, T. Single cell technologies are revolutionizing the approach to rare cells. *Immunol. Cell Biol.* **94**, 225–229 (2015).
6. Perfetto, S. P., Chattopadhyay, P. K. & Roederer, M. Seventeen-colour flow cytometry: unravelling the immune system. *Nat. Rev. Immunol.* **4**, 648–655 (2004).
7. Spitzer, M. H. & Nolan, G. P. Mass cytometry: single cells, many features. *Cell* **165**, 780–791 (2016).
8. Bendall, S. C., Nolan, G. P., Roederer, M. & Chattopadhyay, P. K. A deep profiler's guide to cytometry. *Trends Immunol.* **33**, 323–332 (2012).
9. Bendall, S. C. *et al.* Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science* **332**, 687–696 (2011).
This manuscript introduces CyTOF technology, which enables profiling of up to 17 markers across millions of cells. Here, CyTOF is used to study immune signalling in healthy human haematopoiesis.
10. Chattopadhyay, P. K., Gierahn, T. M., Roederer, M. & Love, J. C. Single-cell technologies for monitoring immune systems. *Nat. Immunol.* **15**, 128–135 (2014).
11. Ståhlberg, A., Kubista, M. & Aman, P. Single-cell gene-expression profiling and its potential diagnostic applications. *Expert Rev. Mol. Diagn.* **11**, 735–740 (2011).
12. Heid, C. A., Stevens, J., Livak, K. J. & Williams, P. M. Real time quantitative PCR. *Genome Res.* **6**, 986–994 (1996).
13. Bengtsson, M., Hemberg, M., Rorsman, P. & Ståhlberg, A. Quantification of mRNA in single cells and modelling of RT-qPCR induced noise. *BMC Mol. Biol.* **9**, 63 (2008).
14. Moignard, V. *et al.* Characterization of transcriptional networks in blood stem and progenitor cells using high-throughput single-cell gene expression analysis. *Nat. Cell Biol.* **15**, 363–372 (2013).
15. Moignard, V. *et al.* Decoding the regulatory network of early blood development from single-cell gene expression measurements. *Nat. Biotechnol.* **33**, 269–276 (2015).
16. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621–628 (2008).
17. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
18. Tang, F. *et al.* RNA-Seq analysis to capture the transcriptome landscape of a single cell. *Nat. Protoc.* **5**, 516–535 (2010).
The first paper to carry out scRNA-seq.
19. Wilhelm, B. T. & Landry, J. R. RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. *Methods* **48**, 249–257 (2009).
20. Soreq, L. *et al.* Whole transcriptome RNA sequencing data from blood leukocytes derived from Parkinson's disease patients prior to and following deep brain stimulation treatment. *Genom. Data* **3**, 57–60 (2014).
21. McLoughlin, K. E. *et al.* RNA-seq transcriptional profiling of peripheral blood leukocytes from cattle infected with *Mycobacterium bovis*. *Front. Immunol.* **5**, 396 (2014).
22. Bhargava, V., Ko, P., Willems, E., Mercola, M. & Subramaniam, S. Quantitative transcriptomics using designed primer-based amplification. *Sci. Rep.* **3**, 1740 (2013).
23. Ramsköld, D. *et al.* Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.* **30**, 777–782 (2012).
24. Kapushesky, M. *et al.* Gene Expression Atlas update — a value-added database of microarray and sequencing-based functional genomics experiments. *Nucleic Acids Res.* **40**, D1077–D1081 (2012).
25. Novershtern, N. *et al.* Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell* **144**, 296–309 (2011).
26. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377–382 (2009).
27. Islam, S. *et al.* Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res.* **21**, 1160–1167 (2011).
28. Picelli, S. *et al.* Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10**, 1096–1098 (2013).
29. Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep.* **2**, 666–673 (2012).
30. Jaitin, D. A. *et al.* Massively parallel single-cell RNA-Seq for marker-free decomposition of tissues into cell types. *Science* **343**, 776–779 (2014).
31. Soumillon, M., Cacchiarelli, D., Semrau, S., van Oudenaarden, A. & Mikkelsen, T. S. Characterization of directed differentiation by high-throughput single-cell RNA-Seq. Preprint at *bioRxiv* http://dx.doi.org/10.1101/003236 (2014).
32. Hashimshony, T. *et al.* CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol.* **17**, 77 (2016).
33. Macosko, E. Z. *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
34. Klein, A. M. *et al.* Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201 (2015).
References 33 and 34 introduce the concept of droplet-based barcoding for massively parallel single-cell transcriptomics.
35. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. Preprint at *bioRxiv* http://dx.doi.org/10.1101/065912 (2016).
36. Hayden, E. C. The $1,000 genome. *Nature* **507**, 295 (2014).
37. Yang, L., Duff, M., Graveley, B., Carmichael, G. & Chen, L.-L. Genomewide characterization of non-polyadenylated RNAs. *Genome Biol.* **12**, R16 (2011).
38. Kang, Y. *et al.* Transcript amplification from single bacterium for transcriptome analysis. *Genome Res.* **21**, 925–935 (2011).
39. Saliba, A. E., Westermann, A. J., Gorski, S. A. & Vogel, J. Single-cell RNA-seq: advances and future challenges. *Nucleic Acids Res.* **42**, 8845–8860 (2014).
40. Bjorklund, A. K. *et al.* The heterogeneity of human CD127+ innate lymphoid cells revealed by single-cell RNA sequencing. *Nat. Immunol.* **17**, 451–460 (2016).
41. Gury-BenAri, M. *et al.* The spectrum and regulatory landscape of intestinal innate lymphoid cells are shaped by the microbiome. *Cell* **166**, 1231–1246. e13 (2016).
References 40 and 41 report heterogeneity in tonsil and intestine-derived CD127$^+$ ILCs; they describe distinct ILC subpopulations and previously unknown ILC types.
42. Gaublomme, J. T. *et al.* Single-cell genomics unveils critical regulators of Th17 cell pathogenicity. *Cell* **163**, 1400–1412 (2015).
This study identifies highly pathogenic subsets of T_H17 cells that are characterized by increased expression of *Gpr65*, *Plzp*, *Toso* and *Cd5l*.
43. Lipniacki, T., Paszek, P., Brasier, A. R., Luxon, B. A & Kimmel, M. Stochastic regulation in early immune response. *Biophys. J.* **90**, 725–742 (2006).
44. Lipniacki, T., Hat, B., Faeder, J. R. & Hlavacek, W. S. Stochastic effects and bistability in T cell receptor signaling. *J. Theor. Biol.* **254**, 110–122 (2008).
45. Feinerman, O. *et al.* Single-cell quantification of IL-2 response by effector and regulatory T cells reveals critical plasticity in immune response. *Mol. Syst. Biol.* **6**, 437 (2010).
46. Satija, R. & Shalek, A. K. Heterogeneity in immune responses: from populations to single cells. *Trends Immunol.* **35**, 219–229 (2014).
47. Stubbington, M. J. T. *et al.* T cell fate and clonality inference from single-cell transcriptomes. *Nat. Methods* **13**, 329–332 (2016).
48. Lönnberg, T. *et al.* Single-cell RNA-seq and computational analysis using temporal mixture modeling resolves TH1/TFH fate bifurcation in malaria. *Sci. Immunol.* **2**, eaal2192 (2017).
49. Redmond, D., Poran, A. & Elemento, O. Single-cell TCRseq: paired recovery of entire T-cell alpha and beta chain transcripts in T-cell receptors from single-cell RNAseq. *Genome Med.* **8**, 80 (2016).
50. Afik, S. *et al.* Targeted reconstruction of T cell receptor sequence from single cell RNA-sequencing links CDR3 length to T cell differentiation state. Preprint at *bioRxiv* http://dx.doi.org/10.1101/072744 (2016).
References 47–50 introduce computational approaches for TCR repertoire reconstruction from full-length scRNA-seq data.
51. Shalek, A. K. *et al.* Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* **498**, 236–240 (2013).

52. Shalek, A. K. *et al.* Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature* **510**, 363–369 (2014).
53. Brennecke, P. *et al.* Single-cell transcriptome analysis reveals coordinated ectopic gene-expression patterns in medullary thymic epithelial cells. *Nat. Immunol.* **16**, 933–941 (2015).
54. Meredith, M., Zemmour, D., Mathis, D. & Benoist, C. Aire controls gene expression in the thymic epithelium with ordered stochasticity. *Nat. Immunol.* **16**, 942–949 (2015).
References 53 and 54 show how single-cell methods can be used to discover both coordinated and stochastic sources of heterogeneity in medullary thymic epithelial cells.
55. Bendall, S. C. *et al.* Single-cell trajectory detection uncovers progression and regulatory coordination in human b cell development. *Cell* **157**, 714–725 (2014).
56. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).
References 55 and 56 show how single-cell profiling can be used to reconstruct developmental processes and identify regulators of cell fate decisions.
57. Schlitzer, A. *et al.* Identification of cDC1- and cDC2-committed DC progenitors reveals early lineage priming at the common DC progenitor stage in the bone marrow. *Nat. Immunol.* **16**, 718–728 (2015).
58. Drissen, R. *et al.* Distinct myeloid progenitor–differentiation pathways identified through single-cell RNA sequencing. *Nat. Immunol.* **17**, 666–676 (2016).
59. Tsang, J. C. H. *et al.* Single-cell transcriptomic reconstruction reveals cell cycle and multi-lineage differentiation defects in Bcl11a-deficient hematopoietic stem cells. *Genome Biol.* **16**, 178 (2015).
60. Psaila, B. *et al.* Single-cell profiling of human megakaryocyte-erythroid progenitors identifies distinct megakaryocyte and erythroid differentiation pathways. *Genome Biol.* **17**, 83 (2016).
61. Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
62. Paul, F. *et al.* Transcriptional heterogeneity and lineage commitment in myeloid progenitors. *Cell* **163**, 1663–1677 (2015).
63. Nestorowa, S. *et al.* A single cell resolution map of mouse haematopoietic stem and progenitor cell differentiation running title: single cell map of HSPC differentiation. *Blood* **128**, 20–32 (2016).
References 58, 62 and 63 report unbiased profiling of the mouse HSC compartment, which reveals unexpected levels of progenitor cell heterogeneity, particularly in myeloid cells.
64. Grover, A. *et al.* Single-cell RNA sequencing reveals molecular and functional platelet bias of aged haematopoietic stem cells. *Nat. Commun.* **7**, 11075 (2016).
65. Kollet, O. *et al.* Osteoclasts degrade endosteal components and promote mobilization of hematopoietic progenitor cells. *Nat. Med.* **12**, 657–664 (2006).
66. Calvi, L. M. *et al.* Osteoblastic cells regulate the haematopoietic stem cell niche. *Nature* **425**, 841–846 (2003).
67. Winkler, I. G. *et al.* Bone marrow macrophages maintain hematopoietic stem cell (HSC) niches and their depletion mobilizes HSCs. *Blood* **116**, 4815–4828 (2010).
68. Kobayashi, H. *et al.* Angiocrine factors from Akt-activated endothelial cells balance self-renewal and differentiation of haematopoietic stem cells. *Nat. Cell Biol.* **12**, 1046–1056 (2010).
69. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
70. Achim, K. *et al.* High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nat. Biotechnol.* **33**, 503–509 (2015).
71. Lee, J. H. *et al.* Fluorescent *in situ* sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. *Nat. Protoc.* **10**, 442–458 (2015).
72. Wu, C. *et al.* Clonal tracking of rhesus macaque hematopoiesis highlights a distinct lineage origin for natural killer cells. *Cell Stem Cell* **14**, 486–499 (2014).
73. Perié, L. *et al.* Determining lineage pathways from cellular barcoding experiments. *Cell Rep.* **6**, 617–624 (2014).
74. Chen, C. *et al.* Single-cell whole-genome analyses by Linear Amplification via Transposon Insertion (LIANTI). *Science* **356**, 189–194 (2017).
75. Buenrostro, J. D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).
76. Cusanovich, D. A. *et al.* Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910–914 (2015).
77. Farlik, M. *et al.* Single-cell DNA methylome sequencing and bioinformatic inference of epigenomic cell-state dynamics. *Cell Rep.* **10**, 1386–1397 (2015).
78. Gravina, S., Ganapathi, S. & Vijg, J. Single-cell, locus-specific bisulfite sequencing (SLBS) for direct detection of epimutations in DNA methylation patterns. *Nucleic Acids Res.* **43**, e93 (2015).
79. Guo, H. *et al.* Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing. *Genome Res.* **23**, 2126–2135 (2013).
80. Smallwood, S. A. *et al.* Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat. Methods* **11**, 817–820 (2014).
81. Genshaft, A. S. *et al.* Multiplexed, targeted profiling of single-cell proteomes and transcriptomes in a single reaction. *Genome Biol.* **17**, 188 (2016).
82. Dey, S. S., Kester, L., Spanjaard, B., Bienko, M. & van Oudenaarden, A. Integrated genome and transcriptome sequencing of the same cell. *Nat. Biotechnol.* **33**, 285–289 (2015).
83. Macaulay, I. C. *et al.* G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. *Nat. Methods* **12**, 519–522 (2015).
84. Stoeckius, M. *et al.* Large-scale simultaneous measurement of epitopes and transcriptomes in single cells. Preprint at *bioRxiv* http://dx.doi.org/10.1101/113068 (2017).
85. Pang, W. W. *et al.* Human bone marrow hematopoietic stem cells are increased in frequency and myeloid-biased with age. *Proc. Natl Acad. Sci. USA* **108**, 20012–20017 (2011).

Acknowledgements
The authors thank members of the Satija and Littman laboratories for helpful discussions and the anonymous referees for insightful critiques. R.S. is supported by a National Institutes of Health Director's New Innovator Award Program (DP2-HG-009623).

Author contributions
E.P. and R.S. wrote the article and reviewed and edited the manuscript before submission.

Competing interests statement
The authors declare no competing interests.

DATABASES
Differentiation Map (DMAP) project: https://software.broadinstitute.org/software/cprg/?q=node/48
Haematopoietic stem and progenitor cell atlas: http://blood.stemcells.cam.ac.uk/single_cell_atlas.html
Illumina Body Map Expression Atlas: https://www.ebi.ac.uk/gxa/experiments/E-MTAB-513/Results
Immunological Genome Project: https://www.immgen.org/
Portal for multiple scRNA-seq data: https://portals.broadinstitute.org/single_cell

ALL LINKS ARE ACTIVE IN THE ONLINE PDF