REVIEW

doi:10.1038/nature21350

Scaling single-cell genomics from phenomenology to mechanism

Amos Tanay[1,2] & Aviv Regev[3,4,5]

Three of the most fundamental questions in biology are how individual cells differentiate to form tissues, how tissues function in a coordinated and flexible fashion and which gene regulatory mechanisms support these processes. Single-cell genomics is opening up new ways to tackle these questions by combining the comprehensive nature of genomics with the microscopic resolution that is required to describe complex multicellular systems. Initial single-cell genomic studies provided a remarkably rich phenomenology of heterogeneous cellular states, but transforming observational studies into models of dynamics and causal mechanisms in tissues poses fresh challenges and requires stronger integration of theoretical, computational and experimental frameworks.

Multicellular organisms have evolved sophisticated strategies for cooperation between cells, such that a single genome encodes numerous specialized and complementary functional programs that maximize fitness when they work together[1–4]. Compartmentalization at several levels — cells, tissues and organs — leads to functional diversification of cells and systems with the same underlying genome. Physical copies of the genome are embedded in cells to enable them to maintain a semi-autonomous decision-making process through the selective management of small-molecule, RNA and protein concentrations in cytoplasmic and nuclear compartments. Theoretically, this permits genomes to break the inherent symmetry that is imposed by the precise duplication of DNA in multicellular species. In particular, it facilitates cellular differentiation through the progressive acquisition of specific intracellular molecular compositions, enabling epigenetic mechanisms to emerge and implement cellular memory. At a higher level of organization, intercellular signalling, extracellular structures and environmental cues are used to form complex spatial structures in which cells (and their genomes) are physically embedded. This creates further levels of compartmentalization that encode complex and structured tissues.

A Linnaean framework is an important part of the present theoretical foundation of cellular heterogeneity, defining taxonomies of cell types and aiming to map this hierarchical classification to the ontogeny of the organism. In this context, the composition of single cells in tissues and compartments has historically been characterized using microscopy and fluorescence-activated cell sorting (FACS). Although these techniques have high spatial or cellular resolution, they rely necessarily on a limited number of molecular or visual markers, giving rise to classification schemes that are difficult to study globally and quantitatively, even when combined with downstream genome-wide analysis. Nevertheless, extensive molecular taxonomies are being built through the accumulation of large repositories of images and gene-expression data[5–10]. These taxonomies have proved successful in some somatic tissues with minimal cell differentiation in the adult (for example, the retina[11]) and in tissues that dynamically and continuously re-differentiate throughout the lifespan of an organism (for example, the haematopoietic system[10,12]).

Despite continued progress in enhancing the palette of available markers and the throughput of microscopy and molecular tissue profiling, the Linnaean framework may be inherently inadequate when extended to whole tissues and organisms. Specifically, the approach becomes ambiguous whenever complex mixtures of molecular behaviours are coupled tightly, both spatially and functionally, in ways that make it impossible to separate them physically. Hierarchical classification is also not naturally applicable when molecular mechanisms drive continuous effects, such as during a developmental continuum, branching in differentiation, plastic transitions between several cell types, the diffusion of morphogen gradients or the pathological, irregular deterioration of cellular programs. Indeed, in developmental biology, especially in the context of embryonic differentiation, models for continuous change are well developed, and understanding how such processes diverge to form terminally differentiated stable states is a considerable challenge. Even the precise definition of a cell type is far from being agreed, and various fields describe this important concept differently.

The advent of single-cell genomics, which has enabled profiling of the genetic and molecular states of ever-growing numbers of individual cells, could also open the way for new data-driven definitions of cell identity and function that are less encumbered conceptually by strict a priori hierarchies and less dependent on predefined markers. Genomics — a field that emerged by taking advantage of the genomic consistency of individuals — is therefore being rapidly repurposed to handle complex genomic functions that occur ultimately in single cells. This could help to reformulate key axes such as temporal dynamics, spatial organization and molecular mechanisms that inform and control cell identity. The main challenge for the emerging field of single-cell genomics is to combine the power of quantitative and comprehensive genomics with microscopic resolution to replace coarse-grained, stepwise and deterministic models for cell and tissue function with more quantitative and predictive models.

State-of-the-art single-cell genomics

Genome-wide transcriptional profiling and epigenomics opened the way to comprehensive measurements of the molecular state of cells, in lieu of strategies based on selected markers. Until recently, comprehensive genomic analysis had relied either on pooling heterogeneous mixtures of cells or on sorting and then profiling subpopulations. Bulk profiling provides only mixture averages, which enables genome-wide screening for regulated genes and the detection of compositional changes between experiments, but not the direct identification

[1]Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 76100, Israel. [2]Department of Biological Regulation, Weizmann Institute of Science, Rehovot 76100, Israel. [3]Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. [4]Howard Hughes Medical Institute, Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02140, USA. [5]Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

of intrinsic cellular programs that together compose the samples under study. The analysis of sorted populations, meanwhile, is limited to known subpopulations and sorting panels and masks variation in samples.

In the past few years, several genomics, transcriptomics and epigenomics assays have been reconfigured to enable the analysis of single cells. It is now possible to collect genome-wide profiles of RNA[13–17], DNA[18–23], histone modifications[24,25], chromatin accessibility[26,27], DNA methylation[28–31], nuclear lamina interactions[32] and chromosomal contacts[33], as well as the protein signatures of single cells[34]. Initial studies focused on improving the throughput (number of cells), robustness (performance in samples of varying quality), complexity (the number of distinct molecules that can be captured from each cell) and accuracy (level of noise) of the assays. Single-cell genomics has already described cellular heterogeneity at unprecedented detail in several systems[16,17,26,35–54].

Single-cell RNA sequencing (RNA-seq) is at the forefront of these methods, particularly in terms of throughput. Whereas initial studies analysed between a handful and 100 cells[13,14,17,55], a series of technological advances, including robotics[16,56], microfluidics[39,57] and reverse emulsion and hydrogel droplets, increased the throughput of assays to tens of thousands or hundreds of thousands of cells per experiment[58,59]. There were also advances in techniques for acquiring cells and processing minuscule amounts of RNA[14,60], which improved the robustness of single-cell RNA-seq on small samples such as biopsies[54,61] and fixed cells[62] and even helped to avoid tissue dissociation entirely by isolating cell nuclei[48,63]. Although it is still not routine to apply single-cell RNA-seq to clinical samples, this target is certainly in reach[38,54].

In contrast to throughput and robustness, the complexity and noise of single-cell RNA-seq remains difficult to characterize and optimize. At a basic level, single-cell RNA-seq aims to take a complete census of the messenger RNA molecules in each cell with minimal errors. But because individual cells have a variable mRNA content that is unknown a priori, and because other cellular features may affect the recovery of mRNA from cells, it is not possible at present to estimate the performance of an assay using replicate experiments. Instead, estimates of complexity are based on independent assumptions of the mRNA content of cells, which is predicted to be 10^5–10^6 molecules per cell, when rRNA is excluded from estimates of total RNA[64,65]. Noise estimations are performed following the analysis of variance of transcripts with presumed lower variability, or using spike-in controls (reviewed in ref. 66). Unique molecular identifiers[16,17,56,58,59,67] and techniques for error correction (also reviewed in ref. 66) greatly reduce the level of technical noise by addressing PCR duplicates, detecting cross-cell contamination and defining molecule counts for downstream statistical models. Nevertheless, even when controlling for biases effectively, single-cell RNA-seq remains a sampling strategy that directly measures only some of the cells in a population, and only a fraction of the RNA molecules in each cell.

Sampling, which initially seems to be a limitation, can be a powerful approach to designing efficient experiments, when applied correctly. The optimal single-cell RNA-seq sampling strategy for a cell population or tissue depends on the question under investigation. Various analyses[16,17,39,52,58,59,68] suggest that because the marginal utility of sequencing single-cell RNA-seq libraries decreases rapidly with sequencing depth, aiming to study a greater number of cells with fewer reads per cell may be better suited to the identification and classification of cell types. When a large number of cells are profiled, clusters of cells with similar RNA distributions can be identified and pooled to form idealized models of single-cell transcription with a resolution that is limited only by the number of cells. Conversely, grouping sampled cells into fine granularity clusters cannot be achieved unless a minimal amount of RNA is captured from each cell. Greater sampling depths may also be required for analysing the regulatory relationships between genes in single cells.

Throughput, robustness and complexity are also being optimized for single-cell epigenomics assays. At present, such assays analyse from dozens to hundreds of cells, with partial automation. Complexity is particularly challenging for single-cell epigenomics assays, which must target single-copy molecules and, unlike single-cell RNA-seq, cannot buffer partial sampling through the analysis of high-copy-number molecules. Two pooling strategies have been applied to circumvent low-complexity data. The first strategy pools single cells[25–27,47] to enable effective analysis even when the molecule recovery rate is only 1–10%. The second strategy pools signals in the same cell across multiple related loci[25,26] (for example, those known to be bound by the same transcription factor), which can help to recover the epigenetic state of a cell, even when coverage per locus is sparse.

Importantly, DNA methylation, histone modifications, chromatin accessibility and 3D chromosome organization all carry unique information that is not available from single-cell RNA-seq, even at maximal complexity and throughput. For example, changes in chromatin organization may foreshadow differentiation events, before they are recognizable at the level of RNA expression[47,69], and may be a more reliable fingerprint of a cell's type and stability. DNA methylation landscapes may reflect the cell's developmental potential and the activity of regulatory elements in ways that cannot be inferred from the instantaneous RNA levels of the cell. Emerging strategies can simultaneously measure several types of profiles in the same cell, therefore helping to relate how regulatory events occur from DNA to RNA, and on to protein and cellular phenotype. A number of studies have already combined the measurement of DNA and RNA[70–72], RNA and various protein signatures[73–76] and RNA and DNA methylation[72], and further pairwise and multiway combinations will probably arise soon.

Single-cell genomics data can therefore now include tens of thousands of high-quality single-cell RNA-seq profiles together with hundreds of single-cell epigenomic profiles, possibly for the same cells. These data provide the ultimate Linnaean toolkit, supporting the unbiased and comprehensive classification of cells into subpopulations while simultaneously defining genome-wide transcription and epigenomic states for each detected cell types. But to go beyond such data and their classification into mechanistic models of genome and tissue regulation, single cells must first be measured and interpreted in their temporal and spatial contexts.

The temporal axis and inferring dynamics

Biological processes are dynamic at several timescales, including minutes to hours (fast responses to environmental stimuli), hours to days (cell differentiation) and up to years (pathogenesis). Characterizing the series of regulatory events that underlie such dynamic processes can be surprisingly challenging because cells are rarely in perfect synchrony and their dynamics are non-deterministic (Fig. 1). Genome-scale studies of regulatory dynamics in bulk samples require either successful synchronization or the ability to isolate specific subpopulations of cells at distinct functional points in the process (Fig. 1a). However, the robust genomic profiling of a process such as the cell cycle still remains a challenge[77,78].

Single-cell genomics partially alleviates these limitations. Although most current genomic techniques remain incompatible with live-cell tracking, in principle, the sampling of cells in an asynchronous population enables the development of computational models for cellular dynamics at any timescale that is represented in the cell population under study[79]. Single-cell genomics therefore suggests, in principle, a universal, computationally driven approach for inferring dynamics in genome function and regulation.

A basic principle that underlies the inference of cellular dynamics from single-cell data is maximum parsimony. It suggests that of all the possible models of dynamics that connect the observed cellular states, the one involving minimal changes in transcription is preferred. The principle is effective for inferring cellular dynamics that proceed directionally and irreversibly through a succession of focal points by means of coordinated changes in gene expression. In such cases, it should be possible to infer the 'ordering' of cells (that is, the point in the idealized temporal process that the cell is at) and the molecular states associated

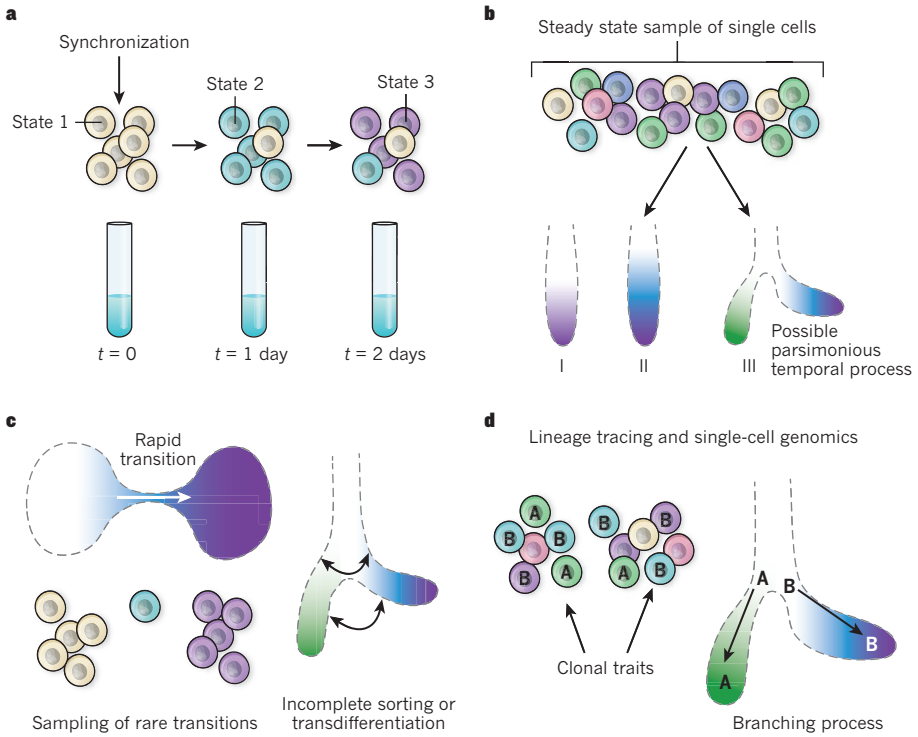

Figure 1 | The temporal axis. **a**, Bulk assays sample populations of cells that progressively lose synchrony (moving from state 1 to state 3), which limits the precise inference of temporal dynamics. **b**, Sampling a heterogenous mixture of single cells in various states can be used to infer temporal dynamics on the basis of the maximum parsimony principle, in which sampled cells are organized along a linear (I and II) or branched (III) process, such that differences in the molecular profiles of the sampled cells are captured by the inferred topology of the process. **c**, The maximum parsimony approach for inferring temporal dynamics from samples of single cells is challenged when transitions between frequent states are rare (left), or when cells undergo complex, non-hierarchical or non-deterministic dynamics (right). In either case, the maximum parsimony model for the data may become underdetermined. **d**, The addition of anchor points (represented here by A and B), such as known stem-cell states or states of differentiation, may help to distinguish between alternative dynamical models in single-cell data. Information from experiments on the clonal relationship in a sample of single cells can lead to the correct identification of a branching process.

with cells at prominent branching or differentiation points. Moreover, because every single cell that is sampled is positioned somewhere along the functional process, the 'residence time' in a particular interval of the process relates to the proportion of the cells sampled that were found in the interval[79]. If the number of cells sampled overall is substantial, even extremely transient (and hence rare) states can be positioned confidently in an inferred trajectory. Biological asynchrony therefore becomes an asset, and for short-term processes that recur continuously (for example, haematopoiesis), the entire dynamic process can be sampled effectively in a single experiment using very deep sampling of single cells[35,37,40,80–85] (reviewed in ref. 66).

A series of studies[35,37,40,80–85] has now used maximum parsimony to infer dynamics from the profiles of single cells while assuming linear[35,40,48,86], cyclic[45,87] or bifurcating[37,80–84] trajectories. For example, an early method called Wanderlust used single-cell multiplex protein measurements from mass cytometry data to build a trajectory of B-cell differentiation[35]. Related approaches successfully ordered cells from single-cell RNA-seq data sampled at multiple time points in myogenesis *in vitro*[40] or neurogenesis *in vivo*[48,86]. For the cell cycle, although it has proven challenging to compare bulk profiles of synchronized cell populations across systems[77,78], a cycle of single cells can be readily reconstructed that seems robustly conserved in both human and mouse cell lines and tissues[45,54,58,88,89]. The most recent methods have also made some encouraging progress towards reconstructing bifurcations[37,80–84]. Cellular dynamics have also been inferred from epigenomics data; this is of particular importance when major cell–cell variation is present, such as in studies of chromosomal conformation in cycling cells (ref. 90; P. Fraser & A.T., unpublished observation).

However, *de novo* inference of chains of events or the branching structure of a dynamic process is computationally challenging, even for deeply sampled processes, and it may quickly become underdetermined by the data, especially if little or no prior knowledge exists to determine the anchor points. In particular, the principle of maximum parsimony may not always be appropriate. Analogous to difficulties in phylogenetic reconstruction under a model that considers the lateral transfer of genes[91], the possibility of transdifferentiation, plasticity or incomplete cell-lineage sorting calls for the integration of new computational and experimental approaches with the current parsimony-based reconstruction algorithms. Moreover, any given cell undergoes multiple dynamic processes simultaneously (for example, responding to a nutrient and undergoing cell division at a particular point in its differentiation), and methods are needed to delineate these processes and to distinguish programs that involve incomplete sorting and transdifferentiation from those of commitment and differentiation. It will be important to peg the inferred dynamical models to other independent measures of the process such as morphological[92] or genetic features. For example, trajectory analysis in B cells identified an early putative population that comprised only 0.007% of all cells analysed. The temporal position of this new population could be confirmed by examining the status of the immunoglobulin heavy chain locus. Indeed, in T cells and B cells, T-cell antigen receptor and B-cell receptor immunoglobulin sequences provide a unique tool because their genetic status can be determined directly from single-cell RNA-seq data[53,54,93].

The inference of cellular dynamics will be greatly enhanced by advances in assays that measure both DNA and RNA in the same cell. Cell-lineage maps, which hierarchically relate sampled cells to common progenitor cells, can be inferred from genetic information — through randomized genome editing in model organisms[94] or the natural accrual of mutations at each DNA replication event[95,96]. Coupled with knowledge of the functional identity of a cell, gained through single-cell RNA-seq, for example, it should be possible to further derive cell-fate maps to determine which earlier cell types give rise to later ones, which is a cornerstone of developmental biology[97].

Understanding the temporal dynamics of single cells is important for shedding light on diseases in humans, especially cancer, in which cellular histories must often be inferred rather than measured directly. Tumourigenesis is highly dynamic and involves both genetic and epigenetic changes in the context of a heterogeneous environment[98]. However, using current methods, only one or a few samples can be obtained from a patient's tumour. Simultaneous measurements of both the genetic and functional state of each malignant cell in a tumour biopsy, as well as of the cells in the tumour microenvironment, would provide an extraordinary opportunity to track the evolution of a tumour, to understand metastasis and to monitor and predict the response to therapy. Emerging studies have shown the promise of this and related approaches in melanoma[54], glioblastoma[38,99], breast cancer[23], oligodendroglioma[100] and leukaemia[101].

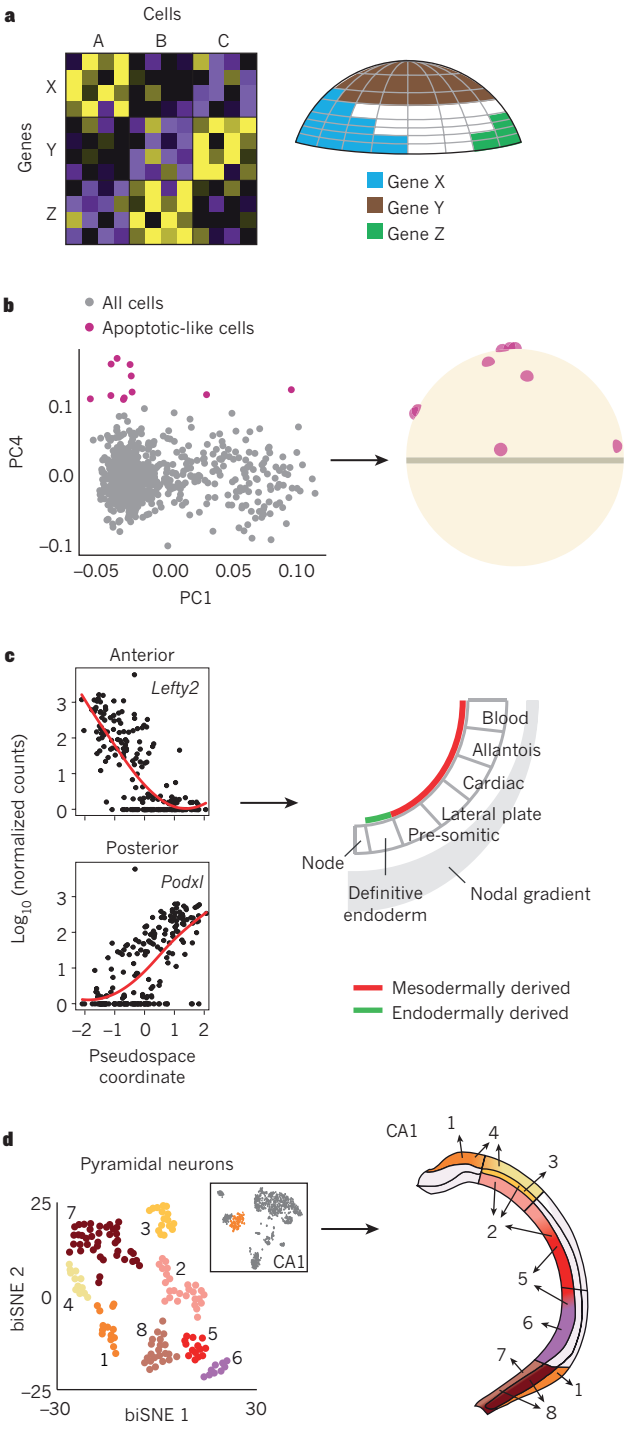

Figure 2 | The spatial axis. **a**, Spatial mapping uses single-cell gene-expression profiles (left) and a reference map of the spatial expression patterns of a small number of landmark genes (right) as input. The expression profile of the landmark genes in a cell is used to determine its spatial position. (Yellow indicates induced expression, purple shows repressed expression and black indicates no change in expression.) **b–d**, Various examples of successful spatial mapping: in the early fish embryo (**b**), several cells with a distinct apoptotic-like profile (principal component (PC) analysis plot; left, purple dots) were mapped to a salt-and-pepper pattern (right) (adapted from ref. 56); in the early mesoderm (**c**), patterns of gene expression in single cells (left) can define an anterior–posterior inferred 'pseudospace' (right) (adapted from ref. 104); and in the hippocampus of the brain (**d**), pyramidal neuron cell clusters from the region CA1 first grouped by proximity in a dimensionality reduced space (biSNE plot, left) can be mapped to positions along the lateral–medial and anterior–posterior axes (adapted from ref. 48). biSNE, biclustering on stochastic neighbor embedding.

Inferring the spatial axis

Although physiological processes take place in tissues and spatial organization is crucial for tissue function, most current single-cell genomics approaches dissociate tissue samples and cannot maintain a registry of the 3D organization of the component cells. A general solution is not yet fully realized; however, this is an area of intense study, with several solutions evolving[56,102–110]. One approach[48,56,102–104] relegates the spatial resolution problem to the computational model (Fig. 2). The transcriptome of the cell often carries the imprint of its location, and by combining the profiles of single cells with a reference map of a small number of marker genes, several studies have mapped single cells back to their spatial position[48,56,102–104]. For example, a study in early fish embryos[56] combined a reference map of legacy *in situ* expression data for a few dozen markers (each of which was measured independently and not at single-cell resolution) with single-cell RNA-seq data from the same developmental stage to assign cells into spatial 'bins' of up to 100 cells. This enabled gene-expression signatures that are strongly coupled to spatial gradients to be distinguished from those that involve spatially independent cell-type specification processes. A similar computational strategy[103] mapped cells distributed in punctate patterns in the brain of a worm. Other studies have shown a similar imprint of both discrete regions and morphogen gradients in the profiles of single cells from early mouse embryos[104] and the adult mouse hippocampus[48].

These examples suggest that experimental designs in which cruder information (for example, from microdissection) is combined with single-cell profiles can help to provide rich information about spatial genomics. However, the compatibility of single-cell genomics methods with such dissection must be improved to minimize batch bias and to increase consistency across tissue sections. More fundamentally, computational spatial reconstruction as implemented at present relies on the idea of canonical tissue organization in embryogenesis or organogenesis, such that multiple experiments can faithfully reproduce the same structure. We suggest that reproducible spatial structure can even be used for computational analysis in less-constrained developmental, differentiation or pathological scenarios. For example, histopathology suggests that there are higher-order features, at various length scales, that are preserved across samples, which form the basis of clinical pathology. To devise universal strategies for spatial analysis in single-cell genomics, new and flexible computational approaches that can be used at several scales (for example, the identification of spatial motifs) would have to be integrated with improved experimental techniques that interrogate spatial structures directly.

Techniques for the *in situ* genome-scale analysis of single cells in tissue sections are evolving rapidly to meet these challenges. Multiplexed RNA fluorescence *in situ* hybridization (FISH) techniques such as MERFISH[105] reliably measure the expression and spatial position of thousands of different transcripts concurrently. Other techniques have produced encouraging proof-of-concept data for *in situ* RNA sequencing in preserved tissue sections and cells[106,109]. It is also possible to determine the spatial expression of dozens of proteins in multiplex by coupling laser or ion-beam ablation of the tissue with mass-cytometry measurements of each ablated 'pixel'[107,108]. Recent studies have used such imaging mass cytometry (B. Bodenmiller, personal communication and ref. 107) to analyse breast-cancer tissue and to group cells into types by the neighbourhood that they inhabit rather than their intrinsic profiles. When using these types of methods, the concept of cell type may be generalized to include spatial context or it may even be redefined completely as a spatial feature on which gene-regulatory mechanisms and programs must be superimposed.

An important implication of the spatial structure of a tissue is the relative localization of specific cell types and the molecular composition of the physical contacts between them. Direct cellular contacts are of particular importance when modelling intercellular regulatory mechanisms. Although a complete 3D map of a tissue can, in theory, characterize all cell–cell junctions, approaches are also being developed to

measure these directly, for example, by assessing small, tightly coupled bone-marrow microniches of two cells (A. van Oudenaarden, personal communication). When restricting the spatial organization problem to the distribution of contacts between cell types, it may be possible to devise computational strategies for analysing correlations between the existence and proportions of specific cell types across many samples. The results can be used to generate hypotheses on the mutual dependency and cellular interactions in complex niches, as has already been shown in tumours[54].

The mechanism axis and modelling gene regulation

To move beyond phenomenology, new approaches are needed for modelling the regulatory mechanisms that underlie the observed repertoire of cellular behaviours. Despite intensive research, the systematic dissection of such mechanisms has remained a substantial challenge. Single-cell genomics provides fresh opportunities for combining observational, mechanistic and perturbational approaches to the inference and modelling of regulatory mechanisms both inside and between cells (Fig. 3). Given the natural cellular resolution of the data, studies in this domain are initially focusing on the intracellular level and are often restricted to transcriptional regulation, whereas future studies will probably focus on integration at the intercellular and tissue level.

The observational approach to inferring mechanisms of gene regulation was applied classically through the inference of correlations between molecular profiles of mRNA levels or epigenetic marks across a diverse set of bulk samples — for example, different tissue types, cell types or stimuli. Such analyses were confounded by the inherently mixed state of the profiled cellular ensembles, which could generate indirect correlations or hide other important regulatory interactions. By contrast, single-cell analysis has more power for covariation analysis at the intracellular level (Fig. 3a). The most basic approach takes a large number of single-cell profiles that capture various transcriptional states and generates candidate regulatory interactions by computing gene–gene correlations. These analyses therefore identify prospective regulators of cell types through the correlation between their pattern of expression and the profiles of genes that define the transcriptional signatures of the cell types. At higher resolution, single-cell analysis refines the definition of cell types and subtypes and enables the sensitive identification of correlation (or its absence) in small cellular niches, facilitating the progressive exclusion of more spurious gene–gene putative interactions. Such analyses have helped to predict the regulators that control cell types in immune cells[16], epithelial cells[44] and neurons[48,52,111,112].

But even at the level of the single cell, correlation does not imply causation, and the predictive modelling of gene regulation must rely both on perturbations of the system and on the integration of mechanistic constraints into the inferred models. Because there is inherent variation between cells — even in a seemingly homogeneous population — owing to extrinsic noise[66,113] or asynchrony in response[17], each cell can be viewed as its own perturbation system. In this case, observational strategies such as correlation can theoretically, and indeed practically, support causal inference. For example, an early study used covariation analysis across 15 dendritic cells after stimulation with lipopolysaccharide[17] to recover a module of co-expressed antiviral genes and correctly associate the proteins signal transducer and activator of transcription (STAT) 2 and interferon regulatory factor 7 (IRF7) as its regulators. A similar approach detected IRF7 as a regulator of a common *in vivo* response to lipopolysaccharide in several dendritic-cell subtypes, each of which implement a distinct program of basal transcription[16].

In other systems, including those in the steady state, cells in a population may naturally span a spectrum of transcriptional states rather than a few discrete modes, corresponding to some phenotypic variation. For example, cell-cycle regulated gene modules can be detected[45,89] or modelled separately from programs that control cellular differentiation[46]. Single-cell profiling was also used to recover a spectrum of transcriptional states that correspond to the pathogenicity and autoimmunogenic potential of T-helper cells that express interleukin-17, identifying gene modules and putative regulators associated with this spectrum, which were validated in animal models[43]. The computational sophistication of the techniques used to infer regulatory models from single-cell data is bounded by the size of the data set and the reliability of the readout. Very large numbers of single cells were shown to support more quantitative modelling in mass cytometry data[36], which suggests that a similar approach may soon become applicable for single-cell genomics.

Temporal resolution can have an important role in inferring regulatory mechanisms from observational data. For dynamic processes that occur over a timescale longer than transcription, such as the stable induction of cell-type specific transcription factors, it is possible to use data sets ordered along temporal trajectories to considerably enhance the power of causal inference by identifying time lags between the activity profiles of early and late regulators that are associated with specific transitions. This approach has been applied to predict and validate a role for the interleukin-7/STAT5 pathway in early B-cell development *in vivo*[35], for various transcriptional regulators in myoblast differentiation *in vitro*[40] and for neurogenesis during embryonic development[86] and in the adult[48]. However, many regulatory processes occur over a much shorter timescale, or involve post-transcriptional and translational mechanisms that are not observed at the transcriptional level, which thereby restrict our ability to infer regulatory relationships from observations — even in time-resolved single-cell data sets.

The incorporation of epigenomics data, especially at the single-cell level, into models of gene regulation can increase the power of a causal inference framework by adding considerations that are not represented in observations of transcriptional states. Population and reference epigenomics may help to restrict the list of potential regulatory interactions by linking transcription factors with binding sites, or enhancer elements with target genes, facilitating the exclusion of putative regulatory relationships that are incompatible with the mechanistic constraints of epigenomics. Single-cell epigenomics, performed independently or simultaneously with RNA profiling, can define the regulatory landscape in a genome at an even better resolution, linking epigenetic activity at particular regulatory elements with activity at other elements or the RNA output of target genes. Such epigenetic activity can be identified by methods for assessing chromosome accessibility[26,27] or DNA hypomethylation[28,31]. The inference of gene regulation from single-cell epigenomics is still limited by the depth and breadth of the data, and computational approaches that enable the pooling of a large number of epigenomic profiles to compute robust correlations[25–27,31] must be further developed. Such approaches may eventually help to develop quantitative, mechanistic models of transcription, and to uncover the molecular basis of the intrinsic and extrinsic factors that drive variability and noise in transcription.

Ultimately, combining single-cell genomics with experimental perturbations of the system of interest provides the most direct avenue for causal inference. Analysis of classic knockout models, *ex vivo*[17,39] or *in vivo*[46], has enabled validation and refinement of the regulatory interactions between factors, for example, the CCAAT/enhancer-binding proteins C/EBPα and C/EBPε during myeloid cell differentiation[46]. Modern high-throughput perturbation methods, especially those based on clustered regularly interspaced short palindromic repeat (CRISPR) technology, can be combined with single-cell genomics to perform causal analysis at an unprecedented scale and resolution. This can be achieved by coupling existing CRISPR screens with a single-cell RNA-seq readout, providing an intimate dissection of the molecular response to perturbations using the high-content phenotype offered by an expression profile in lieu of the low-content readout of a single expression marker. Such a combination requires massive throughput from single-cell RNA-seq technology and strategies for reading the one or more applied perturbations from each of the analysed single cells. Perturbation screens can also be designed to specifically perturb regulatory systems and to test hypotheses derived by mechanistic models of single-cell gene regulation.

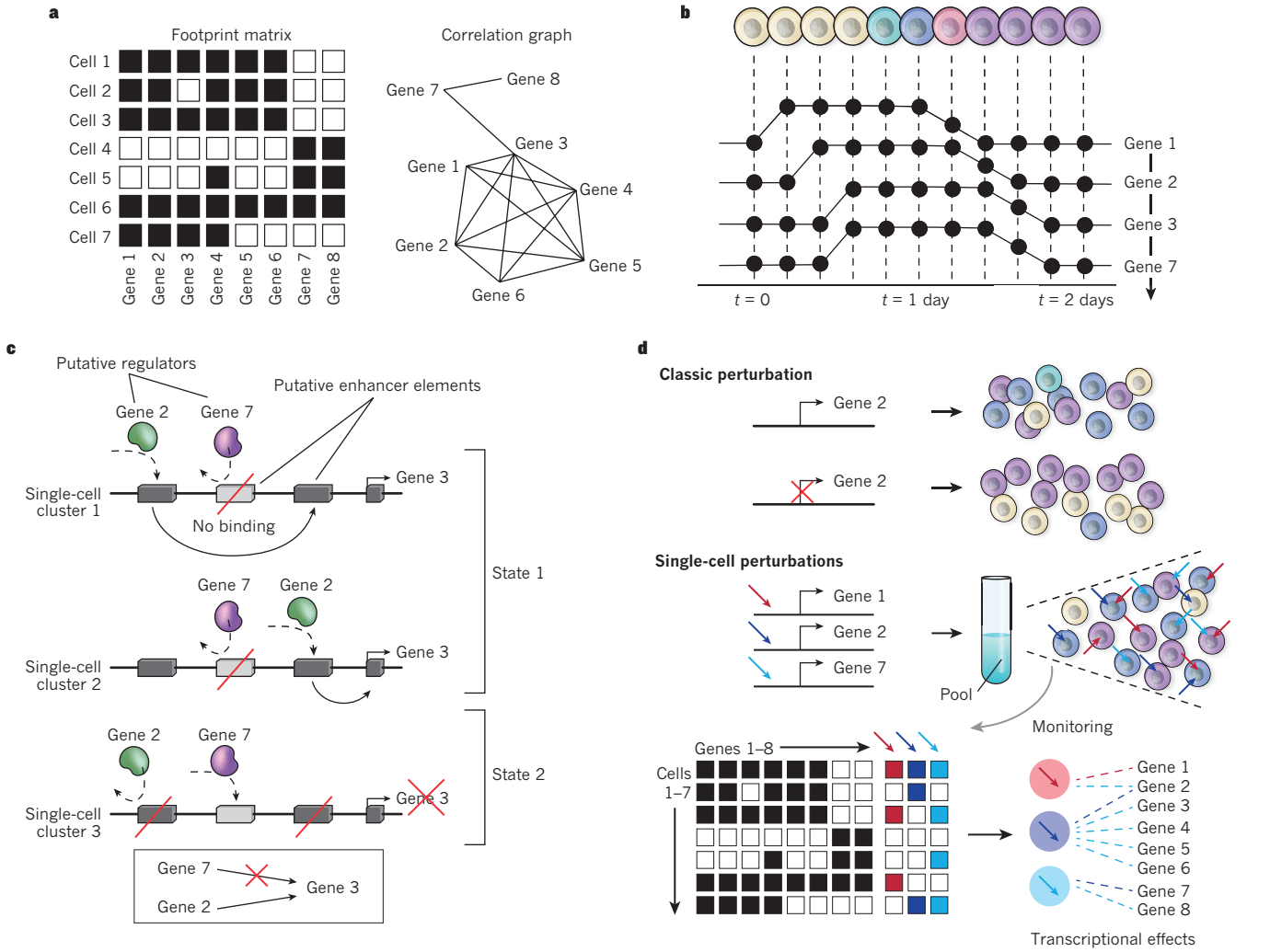

Figure 3 | The mechanism axis. **a**, Inference through covariation across single cells. Using expression profiles (left) for gene regulators and their targets (columns) across cells (rows), a graph of the correlation between genes (right) can be constructed to identify candidate regulators. With increasing numbers of cells, the correlation approach can help to exclude putative regulatory relationships if they are inconsistent with observed states. For example, gene 7 is unlikely to regulate genes 1, 2, 4, 5 or 6, but may regulate genes 3 and 8. **b**, Inference through temporally resolved single-cell data. Putative regulatory interactions are identified as time lags between the activity profiles of gene regulators and their potential targets. The temporal order of the onset of expression of four such genes is shown on the right. Here, the data suggest that gene 7 is unlikely to regulate gene 3 because there is no lag between the onset of expression of either gene. **c**, The refinement of regulatory models with epigenetic information. A regulatory region that surrounds gene 3 is depicted, including three putative enhancer elements that are targeted by two putative regulators, encoded by gene 2 (green) and gene 7 (purple). Pooled epigenomics data from single-cell clusters can be used to identify two states. In state 1, gene 3 is active and targeted by gene 2, and in state 2, gene 3 is inactive and targeted by gene 7. It can therefore be inferred that gene 2, but not gene 7, is activating gene 3. **d**, The causal inference of regulatory models by perturbation. Classic perturbation experiments that target specific genes followed by single-cell profiling help to determine causal relationships (top). Here, a perturbation that prevents the expression of gene 2 is shown to affect the composition of the sampled population of cells. Perturbation experiments performed on a pool of single cells (bottom), combinatorially for several genes, can be followed by the single-cell monitoring of both the perturbation and its effect on transcription, providing enhanced power for causal inference.

Outlook

Efforts towards the mapping and classification of cellular programs in humans and model organisms are becoming increasingly ambitious, aiming to provide a comprehensive atlas of the cell types and subtypes in organs and whole organisms. This opens up remarkable opportunities to move beyond descriptive studies of cell type and state and to develop mechanistic–predictive models of regulatory programs. Reference maps are an essential starting point for the inference and testing of predictive models. Mechanistic models, in turn, make it possible to dissect, annotate and contextualize large reference maps. Therefore, a comprehensive atlas of normal cellular states in humans or mice will pave the way for, but not substitute, the need for predictive and mechanistic models. This is because each disease can give rise to new variations on the reference state, and each genetic variant may modify it further, making a single-cell atlas only the basis for computing the effective states in any cell population of interest. In the context of an extensive reference atlas, the focus of gene-regulatory modelling need not be the complete *de novo* inference of states from, for example, sequences or epigenomes. Instead, it should focus on the ability to infer the effect of small perturbations, owing to the availability of measured states with very similar characteristics. Modelling how known cellular states are perturbed may therefore shed light on mechanisms of disease while providing a tractable path to predictive models of cellular function, despite the incredible computational complexity of the cellular process.

By way of a metaphor, the existence of a 'periodic table' of cellular states may be sufficient for understanding the 'physics' of individual cells, but not the 'chemistry' through which they combine in tissues. Therefore, as single-cell genomics is revolutionizing our ability to map cellular states and to infer mechanisms for intracellular gene regulation (that is, the 'elements' of the periodic table), another grand challenge is

how to integrate individual cellular states into models of functioning tissues (that is, the chemistry). Cells are the building blocks of tissues, and the emerging techniques that we have discussed may enable their mapping in 3D or in their immediate neighbourhoods, or the tracking of their lineage trees. Methods to measure and model key intercellular molecular markers such as metabolites, signalling molecules and components of the extracellular matrix will be crucial for assembling single cells into cohesive models of tissues. Similar to models of cellular state, descriptive tissue models are only a starting point for understanding and predicting the higher-order organization and function of tissues. Although the complexity of the processes performed by communities of interacting cells, as well as the molecular decisions that they make, may seem daunting, it is probable that an intermediate molecular–anatomical level of 'tissue modules' with distinct functionalities exists[1]. We therefore envision a framework that combines information on spatial proximity and molecular communication between cells with the functional impact of such interactions on cellular states to identify and study recurrent multicell modules in tissues. Such modules may consist, as proposed previously[1], of cells with complementary functions, including the core specialized cell of the tissue (for example, adipocyte cells in fat, myocytes in muscle or epithelial cells in the gut mucosa) and accessory cells that provide important support functions[1]. When such modules can be characterized and studied experimentally, single-cell genomics may lead us towards a real revolution in our fundamental understanding of biology. ■

Received 19 September; accepted 14 November 2016.

1. Okabe, Y. & Medzhitov, R. Tissue biology perspective on macrophages. *Nature Immunol.* **17,** 9–17 (2016).
2. Mayr, E. *The Growth of Biological Thought: Diversity, Evolution, and Inheritance* (Belknap, 1982).
3. Szathmary, E. & Maynard-Smith, J. *The Major Transitions in Evolution* (Oxford University Press, 1995).
4. Gould, S. J. *Ontogeny and Phylogeny* (Belknap, 1977).
5. Richardson, L. *et al.* EMAGE mouse embryo spatial gene expression database: 2014 update. *Nucleic Acids Res.* **42,** D835–D844 (2014).
6. Oh, S. W. *et al.* A mesoscale connectome of the mouse brain. *Nature* **508,** 207–214 (2014).
7. Miller, J. A. *et al.* Transcriptional landscape of the prenatal human brain. *Nature* **508,** 199–206 (2014).
8. Hawrylycz, M. J. *et al.* An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489,** 391–399 (2012).
9. Bakken, T. E. *et al.* A comprehensive transcriptional map of primate brain development. *Nature* **535,** 367–375 (2016).
10. Jojic, V. *et al.* Identification of transcriptional regulators in the mouse immune system. *Nature Immunol.* **14,** 633–643 (2013).
11. Sanes, J. R. & Masland, R. H. The types of retinal ganglion cells: current status and implications for neuronal classification. *Annu. Rev. Neurosci.* **38,** 221–246 (2015).
12. Chao, M. P., Seita, J. & Weissman, I. L. Establishment of a normal hematopoietic and leukemia stem cell hierarchy. *Cold Spring Harb. Symp. Quant. Biol.* **73,** 439–449 (2008).
13. Ramsköld, D. *et al.* Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nature Biotechnol.* **30,** 777–782 (2012).
Refs 13–17 are initial reports on the development and scaling of single-cell RNA-seq.
14. Picelli, S. *et al.* Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nature Methods* **10,** 1096–1098 (2013).
15. Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep.* **2,** 666–673 (2012).
16. 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).
17. Shalek, A. K. *et al.* Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* **498,** 236–240 (2013).
18. Zong, C., Lu, S., Chapman, A. R. & Xie, X. S. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science* **338,** 1622–1626 (2012).
Refs 18–23 introduce and develop single-cell genome sequencing, which has applications in the inference of cancer evolution.
19. Xu, X. *et al.* Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. *Cell* **148,** 886–895 (2012).
20. Navin, N. *et al.* Tumour evolution inferred by single-cell sequencing. *Nature* **472,** 90–94 (2011).
21. Leung, M. L., Wang, Y., Waters, J. & Navin, N. E. SNES: single nucleus exome sequencing. *Genome Biol.* **16,** 55 (2015).
22. Hou, Y. *et al.* Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm. *Cell* **148,** 873–885 (2012).
23. Gao, R. *et al.* Punctuated copy number evolution and clonal stasis in triple-negative breast cancer. *Nature Genet.* **48,** 1119–1130 (2016).
24. Rotem, A. *et al.* High-throughput single-cell labeling (Hi-SCL) for RNA-Seq using drop-based microfluidics. *PLoS ONE* **10,** e0116328 (2015).
25. Rotem, A. *et al.* Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nature Biotechnol.* **33,** 1165–1172 (2015).
This paper develops the technique single-cell chromatin immunoprecipitation followed by sequencing (ChIP-seq) as well as the concept of single-cell pooling for the analysis of sparse epigenomic data.
26. Buenrostro, J. D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523,** 486–490 (2015).
Refs 26 and 27 develop two methods of single-cell assay for transposase-accessible chromatin using sequencing (ATAC-seq) and discuss their application to the identification of cell subtypes and to pooling information across loci.
27. Cusanovich, D. A. *et al.* Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348,** 910–914 (2015).
28. Smallwood, S. A. *et al.* Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nature Methods* **11,** 817–820 (2014).
Refs 28, 30 and 31 develop variants of single-cell DNA methylation profiling, which are applied to sparse single-cell epigenomic analysis.
29. Mooijman, D., Dey, S. S., Boisset, J. C., Crosetto, N. & van Oudenaarden, A. Single-cell 5hmC sequencing reveals chromosome-wide cell-to-cell variability and enables lineage reconstruction. *Nature Biotechnol.* **34,** 852–856 (2016).
This paper uses restriction enzymes to profile hydroxymethylation and perform very short-term lineage reconstruction.
30. 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).
31. Farlik, M. *et al.* Single-cell DNA methylome sequencing and bioinformatic inference of epigenomic cell-state dynamics. *Cell Rep.* **10,** 1386–1397 (2015).
32. Kind, J. *et al.* Genome-wide maps of nuclear lamina interactions in single human cells. *Cell* **163,** 134–147 (2015).
33. Nagano, T. *et al.* Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502,** 59–64 (2013).
A proof-of-concept report on single-cell Hi-C (a high-throughput derivative of chromosome conformation capture), which demonstrates the variability of T-cell chromosomal architectures.
34. 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).
35. Bendall, S. C. *et al.* Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. *Cell* **157,** 714–725 (2014).
36. Krishnaswamy, S. *et al.* Conditional density-based analysis of T cell signaling in single-cell data. *Science* **346,** 1250689 (2014).
37. Marco, E. *et al.* Bifurcation analysis of single-cell gene expression data reveals epigenetic landscape. *Proc. Natl Acad. Sci. USA* **111,** E5643–E5650 (2014).
38. Patel, A. P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344,** 1396–1401 (2014).
39. Shalek, A. K. *et al.* Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature* **510,** 363–369 (2014).
40. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature Biotechnol.* **32,** 381–386 (2014).
41. Arvey, A. *et al.* Genetic and epigenetic variation in the lineage specification of regulatory T cells. *eLife* **4,** e07571 (2015).
42. Avraham, R. *et al.* Pathogen cell-to-cell variability drives heterogeneity in host immune responses. *Cell* **162,** 1309–1321 (2015).
43. Gaublomme, J. T. *et al.* Single-cell genomics unveils critical regulators of Th17 cell pathogenicity. *Cell* **163,** 1400–1412 (2015).
44. Grün, D. *et al.* Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature* **525,** 251–255 (2015).
45. Kowalczyk, M. S. *et al.* Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. *Genome Res.* **25,** 1860–1872 (2015).
46. Paul, F. *et al.* Transcriptional heterogeneity and lineage commitment in myeloid progenitors. *Cell* **163,** 1663–1677 (2015).
47. Corces, M. R. *et al.* Lineage-specific and single cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nature Genet.* **48,** 1193–1203 (2016).
48. Habib, N. *et al.* Div-Seq: single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons. *Science* **353,** 925–928 (2016).
This paper adapts RNA-seq to the nuclei of single cells, with applications in rare cells and other challenging samples.
49. Matcovitch-Natan, O. *et al.* Microglia development follows a stepwise program to regulate brain homeostasis. *Science* **353,** aad8670 (2016).
50. Olsson, A. *et al.* Single-cell analysis of mixed-lineage states leading to a binary cell fate choice. *Nature* **537,** 698–702 (2016).
51. Proserpio, V. *et al.* Single-cell analysis of $CD4^+$ T-cell differentiation reveals three major cell states and progressive acceleration of proliferation. *Genome Biol.* **17,** 103 (2016).
52. Shekhar, K. *et al.* Comprehensive classification of retinal bipolar neurons by single-cell transcriptomics. *Cell* **166,** 1308–1323 (2016).
53. Stubbington, M. J. *et al.* T cell fate and clonality inference from single-cell transcriptomes. *Nature Methods* **13,** 329–332 (2016).

54. Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352,** 189–196 (2016).
55. Kumar, R. M. *et al.* Deconstructing transcriptional heterogeneity in pluripotent stem cells. *Nature* **516,** 56–61 (2014).
56. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nature Biotechnol.* **33,** 495–502 (2015).
57. Treutlein, B. *et al.* Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* **509,** 371–375 (2014).
58. Macosko, E. Z. *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161,** 1202–1214 (2015).
Refs 58 and 59 enhance throughputs in single-cell RNA-seq using droplet technology.
59. Klein, A. M. *et al.* Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161,** 1187–1201 (2015).
60. Hashimshony, T. *et al.* CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol.* **17,** 77 (2016).
61. Nichterwitz, S. *et al.* Laser capture microscopy coupled with Smart-seq2 for precise spatial transcriptomic profiling. *Nature Commun.* **7,** 12139 (2016).
62. Thomsen, E. R. *et al.* Fixed single-cell transcriptomic characterization of human radial glial diversity. *Nature Methods* **13,** 87–93 (2016).
63. Lake, B. B. *et al.* Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* **352,** 1586–1590 (2016).
64. Milo, R., Jorgensen, P., Moran, U., Weber, G. & Springer, M. BioNumbers — the database of key numbers in molecular and cell biology. *Nucleic Acids Res.* **38,** D750–D753 (2010).
65. Han, F. & Lillard, S. J. *In-situ* sampling and separation of RNA from individual mammalian cells. *Anal. Chem.* **72,** 4073–4079 (2000).
66. Wagner, A., Regev, A. & Yosef, N. Revealing the vectors of cellular identity with single-cell genomics. *Nature Biotechnol.* **34,** 1145–1160 (2016).
67. Islam, S. *et al.* Quantitative single-cell RNA-seq with unique molecular identifiers. *Nature Methods* **11,** 163–166 (2014).
68. Heimberg, G., Bhatnagar, R., El-Samad, H. & Thomson, M. Low dimensionality in gene expression data enables the accurate extraction of transcriptional programs from shallow sequencing. *Cell Syst.* **2,** 239–250 (2016).
69. Lavin, Y. *et al.* Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. *Cell* **159,** 1312–1326 (2014).
70. Macaulay, I. C. *et al.* G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. *Nature Methods* **12,** 519–522 (2015).
Refs 70–76 introduce techniques for the simultaneous profiling of RNA and DNA, RNA and DNA methylation, or RNA and proteins.
71. Dey, S. S., Kester, L., Spanjaard, B., Bienko, M. & van Oudenaarden, A. Integrated genome and transcriptome sequencing of the same cell. *Nature Biotechnol.* **33,** 285–289 (2015).
72. Angermueller, C. *et al.* Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nature Methods* **13,** 229–232 (2016).
73. Frei, A. P. *et al.* Highly multiplexed simultaneous detection of RNAs and proteins in single cells. *Nature Methods* **13,** 269–275 (2016).
74. Darmanis, S. *et al.* Simultaneous multiplexed measurement of RNA and proteins in single cells. *Cell Rep.* **14,** 380–389 (2016).
75. Albayrak, C. *et al.* Digital quantification of proteins and mRNA in single mammalian cells. *Mol. Cell* **61,** 914–924 (2016).
76. Genshaft, A. S. *et al.* Multiplexed, targeted profiling of single-cell proteomes and transcriptomes in a single reaction. *Genome Biol.* **17,** 188 (2016).
77. Lu, Y. *et al.* Combined analysis reveals a core set of cycling genes. *Genome Biol.* **8,** R146 (2007).
78. Bar-Joseph, Z. *et al.* Genome-wide transcriptional analysis of the human cell cycle identifies genes differentially regulated in normal and cancer cells. *Proc. Natl Acad. Sci. USA* **105,** 955–960 (2008).
79. Kafri, R. *et al.* Dynamics extracted from fixed cells reveal feedback linking cell growth to cell cycle. *Nature* **494,** 480–483 (2013).
80. Setty, M. *et al.* Wishbone identifies bifurcating developmental trajectories from single-cell data. *Nature Biotechnol.* **34,** 637–645 (2016).
Refs 80–87 introduce techniques for inferring dynamics from the steady state sampling of single-cell profiles.
81. Moignard, V. *et al.* Decoding the regulatory network of early blood development from single-cell gene expression measurements. *Nature Biotechnol.* **33,** 269–276 (2015).
82. Haghverdi, L., Buttner, M., Wolf, F. A., Buettner, F. & Theis, F. J. Diffusion pseudotime robustly reconstructs lineage branching. *Nature Methods* **13,** 845–848 (2016).
83. Haghverdi, L., Buettner, F. & Theis, F. J. Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics* **31,** 2989–2998 (2015).
84. Chen, J., Schlitzer, A., Chakarov, S., Ginhoux, F. & Poidinger, M. Mpath maps multi-branching single-cell trajectories revealing progenitor cell progression during development. *Nature Commun.* **7,** 11988 (2016).
85. Angerer, P. *et al.* destiny: diffusion maps for large-scale single-cell data in R. *Bioinformatics* **32,** 1241–1243 (2016).
86. Shin, J. *et al.* Single-cell RNA-Seq with waterfall reveals molecular cascades underlying adult neurogenesis. *Cell Stem Cell* **17,** 360–372 (2015).
87. Gut, G., Tadmor, M. D., Pe'er, D., Pelkmans, L. & Liberali, P. Trajectories of cell-cycle progression from fixed cell populations. *Nature Methods* **12,** 951–954 (2015).
88. Proserpio, V. *et al.* Single-cell analysis of $CD4^+$ T-cell differentiation reveals three major cell states and progressive acceleration of proliferation. *Genome Biol.* **17,** 103 (2016).
89. Buettner, F. *et al.* Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nature Biotechnol.* **33,** 155–160 (2015).
90. Naumova, N. *et al.* Organization of the mitotic chromosome. *Science* **342,** 948–953 (2013).
91. Koonin, E. V. Horizontal gene transfer: essentiality and evolvability in prokaryotes, and roles in evolutionary transitions. *F1000Res.* **5,** 1805 (2016).
92. Kellogg, R. A. & Tay, S. Noise facilitates transcriptional control under dynamic inputs. *Cell* **160,** 381–392 (2015).
93. 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 http://biorxiv.org/content/early/2016/08/31/072744 (2016).
94. McKenna, A. *et al.* Whole organism lineage tracing by combinatorial and cumulative genome editing. *Science* **353,** aaf7907 (2016).
95. Reizel, Y. *et al.* Cell lineage analysis of the mammalian female germline. *PLoS Genet.* **8,** e1002477 (2012).
96. Shlush, L. I. *et al.* Cell lineage analysis of acute leukemia relapse uncovers the role of replication-rate heterogeneity and microsatellite instability. *Blood* **120,** 603–612 (2012).
97. Sulston, J. E. & Horvitz, H. R. Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Dev. Biol.* **56,** 110–156 (1977).
98. Alizadeh, A. A. *et al.* Toward understanding and exploiting tumor heterogeneity. *Nature Med.* **21,** 846–853 (2015).
99. Wang, J. *et al.* Clonal evolution of glioblastoma under therapy. *Nature Genet.* **48,** 768–776 (2016).
100. Tirosh, I. *et al.* Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature* **539,** 309–313 (2016).
101. Li, S. *et al.* Distinct evolution and dynamics of epigenetic and genetic heterogeneity in acute myeloid leukemia. *Nature Med.* **22,** 792–799 (2016).
102. Durruthy-Durruthy, R. *et al.* Reconstruction of the mouse otocyst and early neuroblast lineage at single-cell resolution. *Cell* **157,** 964–978 (2014).
103. Achim, K. *et al.* High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nature Biotechnol.* **33,** 503–509 (2015).
104. Scialdone, A. *et al.* Resolving early mesoderm diversification through single-cell expression profiling. *Nature* **535,** 289–293 (2016).
105. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348,** aaa6090 (2015).
Refs 105 and 106 develop multiplexed RNA-FISH (ref. 105) or *in situ* sequencing (ref. 106) to enable the spatial mapping of a large number of different transcripts at single-cell resolution.
106. Lee, J. H. *et al.* Highly multiplexed subcellular RNA sequencing *in situ*. *Science* **343,** 1360–1363 (2014).
107. Giesen, C. *et al.* Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nature Methods* **11,** 417–422 (2014).
108. Angelo, M. *et al.* Multiplexed ion beam imaging of human breast tumors. *Nature Med.* **20,** 436–442 (2014).
109. Ke, R. *et al. In situ* sequencing for RNA analysis in preserved tissue and cells. *Nature Methods* **10,** 857–860 (2013).
110. Ståhl, P. L. *et al.* Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353,** 78–82 (2016).
111. Zeisel, A. *et al.* Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347,** 1138–1142 (2015).
112. Tasic, B. *et al.* Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nature Neurosci.* **19,** 335–346 (2016).
113. Stewart-Ornstein, J., Weissman, J. S. & El-Samad, H. Cellular noise regulons underlie fluctuations in *Saccharomyces cerevisiae*. *Mol. Cell* **45,** 483–493 (2012).

Acknowledgements We thank L. Gaffney for help with artwork. A.R. is a Howard Hughes Medical Institute investigator. A.T. is a Kimmel investigator and is supported by the Flight Attendant Medical Research Institute and the European Research Council.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: see go.nature.com/2jibe0f. Readers are welcome to comment on the online version of this paper at go.nature.com/2jibe0f. Correspondence should be addressed to A.T. (amos.tanay@weizmann.ac.il) or A.R. (aregev@broadinstitute.org).