Sub-display channel: Science Forum

Title: The Human Cell Atlas

Abstract: The recent advent of methods for high-throughput single-cell molecular profiling has catalyzed a growing sense in the scientific community that the time is ripe to complete the 150-year-old effort to identify all cell types in the human body. The Human Cell Atlas Project is an international collaborative effort that aims to define all human cell types in terms of distinctive molecular profiles (such as gene expression profiles) and to connect this information with classical cellular descriptions (such as location and morphology). An open comprehensive reference map of the molecular state of cells in healthy human tissues would propel the systematic study of physiological states, developmental trajectories, regulatory circuitry and interactions of cells, and also provide a framework for understanding cellular dysregulation in human disease. Here we describe the idea, its potential utility, early proofs-of-concept, and some design considerations for the Human Cell Atlas, including a commitment to open data, code, and community.

Aviv Regev,[1,2,3,*] Sarah A. Teichmann,[4,5,6,*], Eric S. Lander[1,2,7,*], Ido Amit[8], Christophe Benoist[9], Ewan Birney[5], Bernd Bodenmiller[10], Peter Campbell[4,11], Piero Carninci[12], Menna Clatworthy[13], Hans Clevers[14], Bart Deplancke[15], Ian Dunham[5], James Eberwine[16], Roland Eils[17,18], Wolfgang Enard[19], Andrew Farmer[20], Lars Fugger[21], Berthold Göttgens[11,22], Nir Hacohen[1,23], Muzlifah Haniffa[24], Martin Hemberg[4], Seung Kim[25], Paul Klenerman[26,27], Arnold Kriegstein[28], Ed Lein[29], Sten Linnarsson[30], Emma Lundberg[31,32], Joakim Lundeberg[33], Partha Majumder[34], John C. Marioni[4,5,35], Miriam Merad[36], Musa Mhlanga[37], Martijn Nawijn[38], Mihai Netea[39], Garry Nolan[40], Dana Pe'er[41], Anthony Phillipakis[1], Chris P. Ponting[42], Steve Quake[43,44], Wolf Reik[4,45,46], Orit Rozenblatt-Rosen[1], Joshua Sanes[47], Rahul Satija[48,49], Ton N. Schumacher[50], Alex Shalek[1,51,52], Ehud Shapiro[53], Padmanee Sharma[54], Jay W. Shin[12], Oliver Stegle[5], Michael Stratton[4], Michael J.T. Stubbington[4], Fabian J. Theis[55,56], Matthias Uhlen[57,58], Alexander van Oudenaarden[59], Allon Wagner[60], Fiona Watt[61],

Jonathan Weissman[3,62,63,64], Barbara Wold[65], Ramnik Xavier[1,66,67,68], Nir Yosef[52,59], and the Human Cell Atlas Meeting Participants

1. Broad Institute of MIT and Harvard, Cambridge, USA
2. Department of Biology, Massachusetts Institute of Technology, Cambridge, USA
3. Howard Hughes Medical Institute, Chevy Chase, USA
4. Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, UK
5. EMBL-European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, UK
6. Cavendish Laboratory, Department of Physics, University of Cambridge, Cambridge, UK
7. Department of Systems Biology, Harvard Medical School, Boston, USA
8. Department of Immunology, Weizmann Institute of Science, Rehovot, Israel
9. Division of Immunology, Department of Microbiology and Immunobiology, Harvard Medical School, Boston, USA
10. Institute of Molecular Life Sciences, University of Zürich, Zürich, Switzerland
11. Department of Haematology, University of Cambridge, Cambridge, UK
12. Division of Genomic Technologies, RIKEN Center for Life Science Technologies, Yokohama, Japan
13. Molecular Immunity Unit, Department of Medicine, MRC Laboratory of Molecular Biology, University of Cambridge, Cambridge, UK
14. Hubrecht Institute, Princess Maxima Center for Pediatric Oncology and University Medical Center Utrecht, Utrecht, The Netherlands
15. Institute of Bioengineering, School of Life Sciences, Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland
16. Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, USA
17. Division of Theoretical Bioinformatics (B080), German Cancer Research Center (DKFZ), Heidelberg, Germany
18. Department for Bioinformatics and Functional Genomics, Institute for Pharmacy and Molecular Biotechnology (IPMB) and BioQuant, Heidelberg University, Heidelberg, Germany
19. Department of Biology II, Ludwig Maximilian University Munich, Martinsried, Germany

20. Takara Bio USA, Inc., Mountain View, USA
21. Oxford Centre for Neuroinflammation, Nuffield Department of Clinical Neurosciences, and MRC Human Immunology Unit, Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, University of Oxford, Oxford, UK
22. Wellcome Trust-MRC Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK
23. Massachusetts General Hospital Cancer Center, Boston, USA
24. Institute of Cellular Medicine, Newcastle University, Newcastle upon Tyne, UK
25. Departments of Developmental Biology and of Medicine, Stanford University School of Medicine, Stanford, USA
26. Peter Medawar Building for Pathogen Research and the Translational Gastroenterology Unit, Nuffield Department of Clinical Medicine, University of Oxford, Oxford, UK
27. Oxford NIHR Biomedical Research Centre, John Radcliffe Hospital, Oxford, UK
28. Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research, University of California San Francisco, San Francisco, USA
29. Allen Institute for Brain Science, Seattle, USA
30. Laboratory for Molecular Neurobiology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden
31 Science for Life Laboratory, School of Biotechnology, KTH Royal Institute of Technology, Stockholm, Sweden
32 Department of Genetics, Stanford University, Stanford, USA
33. Science for Life Laboratory, Department of Gene Technology, Royal Institute of Technology, Stockholm, Sweden
34. National Institute of Biomedical Genomics, Kalyani, India
35. Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, UK
36. Precision Immunology Institute , Icahn School of Medicine at Mount Sinai , New York, USA
37. Division of Chemical, Systems & Synthetic Biology, Institute for Infectious Disease & Molecular Medicine (IDM), Department of Integrative Biomedical Sciences, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa

38. Department of Pathology and Medical Biology, GRIAC Research Institute, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
39. Department of Internal Medicine and Radboud Center for Infectious Diseases, Radboud University Medical Center, Nijmegen, the Netherlands
40. Department of Microbiology and Immunology, Stanford University, Stanford, USA
41. Computational and Systems Biology Program, Sloan Kettering Institute, New York, USA
42. MRC Human Genetics Unit, MRC Institute of Genetics & Molecular Medicine, University of Edinburgh, Edinburgh, UK
43. Department of Applied Physics and Department of Bioengineering, Stanford University, Stanford, USA
44. Chan Zuckerberg Biohub, San Francisco, USA
45. Epigenetics Programme, The Babraham Institute, Cambridge, UK
46. Centre for Trophoblast Research, University of Cambridge, Cambridge, UK
47. Center for Brain Science and Department of Molecular and Cellular Biology, Harvard University, Cambridge, USA
48. Department of Biology, New York University, New York, USA
49. New York Genome Center, New York University, New York, USA
50. Division of Immunology, The Netherlands Cancer Institute, Amsterdam, The Netherlands
51. Institute for Medical Engineering & Science (IMES) and Department of Chemistry, Massachusetts Institute of Technology, Cambridge, USA
52. Ragon Institute of MGH, MIT and Harvard, Cambridge, USA
53. Department of Computer Science and Department of Biomolecular Sciences, Weizmann Institute of Science, Rehovot, Israel
54. Department of Genitourinary Medical Oncology, Department of Immunology, MD Anderson Cancer Center, University of Texas, Houston, USA
55. Institute of Computational Biology, German Research Center for Environmental Health, Helmholtz Center Munich, Neuherberg, Germany
56. Department of Mathematics, Technical University of Munich, Garching, Germany
57. Science for Life Laboratory and Department of Proteomics, KTH Royal Institute of Technology, Stockholm, Sweden

58. Novo Nordisk Foundation Center for Biosustainability, Danish Technical University (DTU), Lyngby, Denmark
59. Hubrecht Institute and University Medical Center Utrecht, Utrecht, The Netherlands
60. Department of Electrical Engineering and Computer Science and the Center for Computational Biology, University of California Berkeley, Berkeley, USA
61. Centre for Stem Cells and Regenerative Medicine, King's College London, London, UK
62. Department of Cellular & Molecular Pharmacology, University of California San Francisco, San Francisco, USA
63. California Institute for Quantitative Biomedical Research, University of California San Francisco, San Francisco, USA
64. Center for RNA Systems Biology, University of California San Francisco, San Francisco, USA
65. Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, USA
66. Center for Computational and Integrative Biology, Massachusetts General Hospital, Boston, USA
67. Gastrointestinal Unit and Center for the Study of Inflammatory Bowel Disease, Massachusetts General Hospital, Boston, USA
68. Center for Microbiome Informatics and Therapeutics, Massachusetts Institute of Technology, Cambridge, USA

* To whom correspondence should be addressed: aregev@broadinstitute.org (AR), st9@sanger.ac.uk (SAT), eric@broadinstitute.org (ESL)

<L1>**Introduction**</L1>

The cell is the fundamental unit of living organisms. Hooke reported the discovery of cells in plants in 1665 (Hooke, 1665) and named them for their resemblance to the cells inhabited by monks, but it took nearly two centuries for biologists to appreciate their central role in biology. Between 1838 and 1855, Schleiden, Schwann, Remak, Virchow and others crystalized an elegant Cell Theory (Harris, 2000), stating that all organisms are

composed of one or more cells; that cells are the basic unit of structure and function in life; and that all cells are derived from pre-existing cells (Mazzarello, 1999; **Figure 1**).

To study human biology, we must know our cells. Human physiology emerges from normal cellular functions and intercellular interactions. Human disease entails the disruption of these processes and may involve aberrant cell types and states, as seen in cancer. Genotypes give rise to organismal phenotypes through the intermediate of cells, because cells are the basic functional units, each regulating their own program of gene expression. Therefore, genetic variants that contribute to disease typically manifest their action through impact in a particular cell types: for example, genetic variants in the IL23R locus increase risk of autoimmune diseases by altering the function of dendritic cells and T-cells (Duerr et al., 2006), and DMD mutations cause muscular dystrophy through specific effects in skeletal muscle cells (Murray et al., 1982).

For more than 150 years, biologists have sought to characterize and classify cells into distinct types based on increasingly detailed descriptions of their properties, including their shape, their location and relationship to other cells within tissues, their biological function, and, more recently, their molecular components. At every step, efforts to catalog cells have been driven by advances in technology. Improvements in light microscopy were obviously critical. So too was the invention of synthetic dyes by chemists (Nagel, 1981), which biologists rapidly found stained cellular components in different ways (Stahnisch, 2015). In pioneering work beginning in 1887, Santiago Ramón y Cajal applied a remarkable staining process discovered by Camillo Golgi to show that the brain is composed of distinct neuronal cells, rather than a continuous syncytium, with stunningly diverse architectures found in specific anatomical regions (Ramón y Cajal, 1995); the pair shared the 1906 Nobel Prize in Physiology or Medicine for their work.

Starting in the 1930s, electron microscopy provided up to 5000-fold higher resolution, making it possible to discover and distinguish cells based on finer structural features. Immunohistochemistry, pioneered in the 1940s (Arthur, 2016) and accelerated by the advent of monoclonal antibodies (Kohler and Milstein, 1975) and Fluorescence-Activated Cell Sorting (FACS; Dittrich and Göhde, 1971; Fulwyler, 1965) in the 1970s, made it possible to detect the presence and levels of specific proteins. This revealed that morphologically indistinguishable cells can vary dramatically at the molecular level and led to exceptionally fine classification systems, for example, of hematopoietic cells, based

on cell-surface markers. In the 1980s, Fluorescence *in situ* Hybridization (FISH; Langer-Safer et al., 1982) enhanced the ability to characterize cells by detecting specific DNA loci and RNA transcripts. Along the way, studies showed that distinct molecular phenotypes typically signify distinct functionalities. Through these remarkable efforts, biologists have achieved an impressive understanding of specific systems, such as the hematopoietic and immune systems (Chao et al., 2008; Jojic et al., 2013; Kim and Lanier, 2013) or the neurons in the retina (Sanes and Masland, 2015).

Despite this progress, our knowledge of cell types remains incomplete. Moreover, current classifications are based on different criteria, such as morphology, molecules and function, which have not always been related to each other. In addition, molecular classification of cells has largely been ad hoc – based on markers discovered by accident or chosen for convenience – rather than systematic and comprehensive. Even less is known about cell states and their relationships during development: the full lineage tree of cells from the single-cell zygote to the adult is only known for the nematode *C. elegans*, which is transparent and has just ~1000 cells.

At a conceptual level, one challenge is that we lack a rigorous definition of what we mean by the intuitive terms 'cell type' and 'cell state'. Cell type often implies a notion of persistence (*e.g.*, being a hepatic stellate cell or a cerebellar Purkinje cell), while cell state often refers to more transient properties (*e.g.*, being in the G1 phase of the cell cycle or experiencing nutrient deprivation). But, the boundaries between these concepts can be blurred, because cells change over time in ways that are far from fully understood. Ultimately, data-driven approaches will likely refine our concepts.

The desirability of having much deeper knowledge about cells has been well recognized for a long time (Brenner, 2010; Eberwine et al., 1992; Shapiro, 2010; Van Gelder et al., 1990). However, only in the past few years has it begun to seem feasible to undertake the kind of systematic, high-resolution characterization of human cells necessary to create a systematic cell atlas.

The key has been the recent ability to apply genomic profiling approaches to single cells. By 'genomic approaches' we mean methods for large-scale profiling of the genome and its products, including DNA sequence, chromatin architecture, RNA transcripts, proteins, and metabolites (Lander, 1996). It has long been appreciated that such methods provide rich and comprehensive descriptions of biological processes. Historically,

however, they could only be applied to bulk tissue samples comprised of an ensemble of many cells, providing average genomic measures for a sample, but masking their differences across cells. The result is as unsatisfying as trying to understand New York, London or Mumbai based on the average properties of their inhabitants.

The first single-cell genomic characterization method to become feasible at large-scale is trancriptome analysis by single cell RNA-Seq (**Box 1;** Hashimshony et al., 2012; Jaitin et al., 2014; Picelli et al., 2013; Ramskold et al., 2012; Shalek et al., 2013). Initial efforts first used microarrays and then RNA-seq to profile RNA from small numbers of single cells, which were obtained either by manual picking from in situ fixed tissue, using flow-sorting or, later on, with microfluidic devices, adapted from devices developed initially for qPCR-based approaches (Crino et al., 1996; Dalerba et al., 2011; Marcus et al., 2006; Miyashiro et al., 1994; Zhong et al., 2008). Now, massively parallel assays can process tens and hundreds of thousands of single cells simultaneously to measure their transcriptional profiles at rapidly decreasing costs (Klein et al., 2015; Macosko et al., 2015; Shekhar et al., 2016) with increasing accuracy and sensitivity (Svensson et al., 2016; Ziegenhain et al., 2016). In some cases, it is even possible to register these sorted cells to their spatial positions in images (Vickovic et al., 2016). Single-cell RNA sequencing (scRNA-seq) is rapidly becoming widely disseminated.

Following this initial wave of technologies are many additional methods at various stages of development and high-throughput implementation. Techniques are being developed to assay: *in situ* gene expression in tissues at single-cell and even sub-cellular resolution (Chen et al., 2015b; Ke et al., 2013; Lee et al., 2014; Lubeck et al., 2014; Shah et al., 2016; Stahl et al., 2016); the distribution of scores of proteins at cellular or sub-cellular resolution (Angelo et al., 2014; Chen et al., 2015a; Giesen et al., 2014; Hama et al., 2011; Susaki et al., 2014; Yang et al., 2014); various aspects of chromatin state (Buenrostro et al., 2015; Cusanovich et al., 2015; Farlik et al., 2015; Guo et al., 2013; Lorthongpanich et al., 2013; Mooijman et al., 2016; Rotem et al., 2015a; Rotem et al., 2015b; Smallwood et al., 2014); and DNA mutations to allow precise reconstruction of cell lineages (Behjati et al., 2014; Biezuner et al., 2016; Shapiro et al., 2013; Taylor et al., 2003; Teixeira et al., 2013). Various groups are also developing single-cell multi-omic methods to simultaneously measure several types of molecular profiles in the same cell (Albayrak et al., 2016; Angermueller et

al., 2016; Behjati et al., 2014; Darmanis et al., 2016; Dey et al., 2015; Frei et al., 2016; Genshaft et al., 2016; Macaulay et al., 2015).

As a result, there is a growing sense in the scientific community that the time is now right for a project to complete the Human Cell Atlas that pioneering histologists began 150 years ago. Various discussions have taken place in a number of settings over the past two years, culminating in an international meeting in London in October 2016. In addition, several pilot efforts are already underway or in planning – for example, related to brain cells and immune cells. Prompted by such efforts, funding agencies, including the NIH, have sought information from the scientific community about the notion of creating cell or tissue atlases.

The goal of this article is to engage the wider scientific community in this conversation. Although the timing is driven by technologies that have recently appeared or are expected to mature in the near-future, the project itself is fundamentally an intellectual endeavor. We therefore articulate the concept of a cell atlas and explore its potential utility for biology and medicine. We discuss how an atlas can lead to new understanding of histology, development, physiology, pathology, and intra- and inter-cellular regulation, and enhance our ability to predict the impact of perturbations on cells. It will also yield molecular tools with applications in both research and clinical practice.

As discussed below, a Human Cell Atlas Project would be a shared international effort involving diverse scientific communities. More details are available in the Human Cell Atlas White Paper (https://www.humancellatlas.org/files/HCA_WhitePaper_18Oct2017.pdf): the first version of this 'living document', which will updated on a regular basis, was released on October 18, 2017.

<L1>**What is the Human Cell Atlas, and what could we learn from it?**</L1>

At its most basic level, the Human Cell Atlas must include a comprehensive reference catalog of all human cells based on their stable properties and transient features, as well as their locations and abundances. Yet, an atlas is more than just a catalog: it is a *map* that aims to show the *relationships* among its elements. By doing so, it can sometimes reveal fundamental processes – akin to how the atlas of Earth suggested continental drift through the correspondence of coastlines.

To be useful, an atlas must also be an abstraction, comprehensively representing certain features, while ignoring others. The writer Jorge Luis Borges – a master at capturing the tension between grandeur and grandiosity – distilled this challenge in his one-paragraph story, "*On Exactitude in Science*", about an empire enamored with science of cartography (Box 2; Borges and Hurley, 2004). Over time, the cartographers' map of the realm grew more and more elaborate, and hence bigger, until – *expandio ad absurdum* – the map reached the size of the entire empire itself and became useless.

Moreover, an atlas must provide a system of coordinates on which one can represent and harmonize concepts at many levels (geopolitical borders, topography, roads, climate, restaurants, and even dynamic traffic patterns). Features can be viewed at any level of magnification, and high-dimensional information collapsed into simpler views.

So, a key question is how a Human Cell Atlas should abstract key features, provide coordinates, and show relationships. A natural solution would be to describe each human cell by a defined set of molecular markers. For example, one might describe each cell by the expression level of each of the ~20,000 human protein-coding genes: that is, each cell would be represented as a point in ~20,000-dimensional space. Of course, the set of markers could be expanded to include the expression levels of non-coding genes, the levels of the alternatively spliced forms of each transcript, the chromatin state of every promoter and enhancer, and the levels of each protein or each post-translationally modified form of each protein. The optimal amount and type of information to collect will emerge based on a balance of technological feasibility and the biological insight provided by each layer (Corces et al., 2016; Lorthongpanich et al., 2013; Paul et al., 2015). For specific applications, it will be useful to employ reduced representations. Solely for concreteness, we will largely refer below to the 20,000-dimensional space of gene expression, which can already be assayed at high-throughput.

The Atlas should have additional coordinates or annotations to represent histological and anatomical information (*e.g.*, a cell's location, morphology, or tissue context), temporal information (*e.g.*, the age of the individual or time since an exposure), and disease status. Such information is essential for harmonizing results based on molecular profiles with rich knowledge about cell biology, histology and function. How best to capture and represent this information requires serious attention.

In some respects, the Human Cell Atlas Project (whose fundamental unit is a cell) is analogous to the Human Genome Project (whose fundamental unit is a gene). Both are ambitious efforts to create 'Periodic Tables' for biology that comprehensively enumerate the two key 'atomic' units that underlie human life (cells and genes) and thereby provide a crucial foundation for biological research and medical application. As with the Human Genome Project, we will also need corresponding atlases for important model organisms, where conserved cell states can be identified and genetic manipulations and other approaches can be used to probe function and lineage. Yet, the Human Cell Atlas differs in important ways from the Human Genome Project: the nature of cell biology means that it will require a distinct experimental toolbox, and will involve making choices concerning molecular and cellular descriptors. Assessing the distance to completion will also be a challenge.

As a Borgesian thought experiment, we could conceive of an imaginary Ultimate Human Cell Atlas that represents: all markers in every cell in a person's body; every cell's spatial position (by adding three dimensions for the body axes); every cell at every moment of a person's lifetime (by adding another dimension for time relating the cells by a lineage); and the superimposition of such cell atlases from every human being, annotated according to differences in health, genotype, lifestyle and environmental exposure.

Of course, it is not possible to construct such an Ultimate Atlas. However, it is increasingly feasible to sample richly from the distribution of points to understand the key features and relationships among all human cells. We return below to the question of how the scientific community might go about creating a Human Cell Atlas. First, we consider the central scientific question: What could we hope to learn from such an atlas?

A Human Cell Atlas would have a profound impact on biology and medicine by bringing our understanding of anatomy, development, physiology, pathology, intracellular regulation, and intercellular communication to a new level of resolution. It would also provide invaluable markers, signatures and tools for basic research (facilitating detection, purification and genetic manipulation of every cell type) and clinical applications (including diagnosis, prognosis and monitoring response to therapy).

In the following sections, we outline reasonable expectations and describe some early examples. We recognize that these concepts will evolve based on emerging data. It is

clear that a Human Cell Atlas Project will require and will motivate the development of new technologies. It will also necessitate the creation of new mathematical frameworks and computational approaches that may have applications far beyond biology – perhaps analogous to how biological 'big data' in agriculture in the 1920s led to the creation, by R.A. Fisher and others, of key statistical methods, including the analysis of variance and experimental design (Parolini, 2015).

<L1>**Taxonomy: cell types**</L1>

The most fundamental level of analysis is the identification of cell types. In an atlas where cells are represented as points in a high-dimensional space, 'similar' cells should be 'close' in some appropriate sense, although not identical, owing to differences in physiological states (*e.g.*, cell-cycle stage), the inherent noise in molecular systems (Eldar and Elowitz, 2010; Kharchenko et al., 2014; Kim et al., 2015; Shalek et al., 2013), and measurement errors (Buettner et al., 2015; Kharchenko et al., 2014; Kim et al., 2015; Shalek et al., 2013; Shalek et al., 2014; Wagner et al., 2016). Thus, a cell 'type' might be defined as a region or a probability distribution (Kim and Eberwine, 2010; Sul et al., 2012) either in the full-dimensional space or in a projection onto a lower-dimensional space that reflects salient features.

While this notion is intuitively compelling, it is challenging to give a precise definition of a 'cell type'. Cell-type taxonomies are often represented as hierarchies based on morphological, physiological, and molecular differences (Sanes and Masland, 2015). Whereas higher distinctions are easily agreed upon, finer ones may be less obvious and may not obey a strict hierarchy, either because distinct types share features, or because some distinctions are graded and not discrete. Critically, it remains unclear whether distinctions based on morphological, molecular and physiological properties agree with each other. New computational methods will be required both to discover types and to better classify cells and, ultimately, to refine the concepts themselves (Grun and van Oudenaarden, 2015; Shapiro et al., 2013; Stegle et al., 2015; Tanay and Regev, 2017; Wagner et al., 2016). Unsupervised clustering algorithms for high-dimensional data provide an initial framework (Grun et al., 2015; Grun et al., 2016; Jaitin et al., 2014; Levine et al., 2015; Macosko et al., 2015; Shekhar et al., 2016; Vallejos et al., 2015), but substantial advances will be needed in order to select the 'right' features, the 'right' similarity metric, and the 'right'

level of granularity for the question at hand, control for distinct biological processes, handle technical noise, and connect novel clusters with legacy knowledge. Whereas cell types are initially defined based on regions in feature space, it will be important eventually to distill them into simpler molecular and morphological signatures that can be used to index cells in the atlas, aggregate and compare results from independent labs and different individuals, and create tools and reagents for validation and follow up studies.

For all the reasons above, we have not attempted to propose a precise definition of 'cell type'. Rather, the definition should evolve based on empirical observation.

Despite these challenges, recent studies in diverse organs – including immune, nervous, and epithelial tissues – support the prospects for comprehensive discovery of cell types, as well as harmonization of genomic, morphological, and functional classifications (**Figure 2A-C**). For example, analysis of immune cells from mouse spleen (Jaitin et al., 2014) and human blood (Horowitz et al., 2013) showed that well-established functional immune cell types and subtypes could be readily distinguished by unsupervised clustering of single-cell expression profiles. Similarly, single-cell expression profiles of epithelial cells from gut organoids (Grun et al., 2015) distinguished known cell subtypes, each with distinctive functional and histological characteristics, while also revealing a new subtype of enteroendocrine cells, which was subsequently validated experimentally.

The nervous system, where many cell types have not yet been characterized by any means, illustrates both the promise and the challenge. Whereas each of the 302 individual neurons in *C. elegans* can be distinctly defined by its lineage, position, connectivity, molecular profile and functions, the extent to which the ~ 10^{11} neurons in the human brain are distinctly defined by morphological, physiological, lineage, connectivity, and electrical-activity criteria, and have distinct molecular profiles, remains unknown. Cellular neuroanatomy is deeply rooted in the concept of cell types defined by their morphologies (a proxy for connectivity) and electrophysiological properties (Ascoli et al., 2008), and extensive efforts continue to classify the types in complicated structures like the retina and neocortex (Jiang et al., 2015; Markram et al., 2015; Sanes and Masland, 2015). Critically, it remains unclear whether distinctions based on morphological, connectional, and physiological properties agree with their molecular properties.

The mouse retina provides an ideal testing ground to test this correspondence because cell types follow highly stereotyped spatial patterns (Macosko et al., 2015; Sanes

and Masland, 2015). Analysis of 31,000 retinal bipolar cells (Shekhar et al., 2016) automatically re-discovered the 13 subtypes that had been defined over the past quarter-century based on morphology and lamination, while also revealing two new subtypes with distinct morphological and laminar characteristics. These subtypes included one with a 'bipolar' expression pattern and developmental history, but a unipolar morphology in the adult (Shekhar et al., 2016), which has distinct functional characteristics in the neural circuits of the retina (Della Santina et al., 2016). In this example, known morphological and other non-molecular classifications matched perfectly to molecular types, and new molecularly-defined cell types discovered in the single-cell transcriptomic analysis corresponded to unique new morphology and histology. In other complex brain regions such as the neocortex and hippocampus there are also a large number of transcriptionally defined types (Darmanis et al., 2015; Gokce et al., 2016; Habib et al., 2016a; Lake et al., 2016; Pollen et al., 2014; Tasic et al., 2016; Zeisel et al., 2015), but it has been more difficult to find consensus between data modalities, and the relationship between transcriptomic types and anatomical or morphological types is unclear. In this light, technologies that can directly measure multiple cellular phenotypes are essential. For example, electrophysiological measurements with patch clamping followed by scRNA-seq used in a recent study of a particular inhibitory cortical cell type showed that the transcriptome correlated strongly with the cell's physiological state (Cadwell et al., 2016; Foldy et al., 2016). Thus, the transcriptome appears to provide a proxy for other neuronal properties, but much more investigation is needed.

<L1>**Histology: Cell neighborhood and position**</L1>

Histology examines the spatial position of cells and molecules within tissues. Over the past century, we have learnt a great deal about cell types, markers, and tissue architecture, and this body of knowledge will need to be further refined and woven seamlessly into the Human Cell Atlas. With emerging highly multiplexed methods for *in situ* hybridization (Chen et al., 2015b; Shah et al., 2016) or protein staining (Angelo et al., 2014; Giesen et al., 2014), it should be possible to spatially map multiple cell types at once based on expression signatures to see how they relate to each other and to connect them with cell types defined by morphology or function. It should also be possible to extend

observations of continuous gradients for individual genes (such as morphogens) to multi-gene signatures.

Computational approaches could then allow iterative refinement of cellular characterization based on both a cell's molecular profile and information about its neighborhood; methods perfected in the analysis of networks could provide a helpful starting point (Blondel et al., 2008; Rosvall and Bergstrom, 2008). Conversely, expression data from a cell can help map its position in absolute coordinates or relative terms, as well as in the context of pathology, highlighting how disease tissue differs from typical healthy tissue. Combining molecular profiles with tissue architecture will require new computational methods, drawing perhaps on advances in machine vision (Xu et al., 2015; Zheng et al., 2015).

New methods for integrating single-cell genomics data into a spatial context have been developed recently. Single-cell analyses of tissues from early embryos (Satija et al., 2015; Scialdone et al., 2016) to adult (Achim et al., 2015) demonstrate how physical locations can be imprinted in transcriptional profiles (Durruthy-Durruthy et al., 2014) and can be used to infer tissue organization (**Figure 2D**). In the early zebrafish embryo, for example, a cell's expression profile specifies its location to within a small neighborhood of ~100 cells; the related expression patterns of individual genes in turn fall into only nine spatial archetypes (Satija et al., 2015). In the early mouse embryo, key spatial gradients can be recovered by a 'pseudospace' inferred from reduced dimensions of single cell profiles (Scialdone et al., 2016). In adult mouse hippocampus, cell profiles show clear clusters corresponding to discrete functional regions as well as gradients following dorsal/ventral and medial/lateral axes (Habib et al., 2016a). In the annelid brain, even finer punctate spatial patterns can be resolved (Achim et al., 2015).

<L1>**Development: transitions to differentiated cell types**</L1>

Cells arrive at their final differentiated cell types through partly asynchronous branching pathways of development (Ferrell, 2012), which are driven by and reflected in molecular changes, especially gene-expression patterns (see, for example, Chao et al., 2008; Jojic et al., 2013). It should therefore be possible to reconstruct development as trajectories in high-dimensional space, mirroring Waddington's landscape (Ferrell, 2012; Waddington, 1957) – just as it would be possible to infer the ski lifts and trails on a

mountain from snapshots of the positions of enough skiers. One can even infer sharp transitions, provided enough cells are observed. The required sampling density will depend on the number and complexity of paths and intersections, and sorting strategies can help to iteratively enrich for rare, transient populations. Notably, the relative proportions of cells observed at different points along the developmental paths can help convey critical information, both about the duration of each phase (Antebi et al., 2013; Kafri et al., 2013) and the balance of how progenitor cells are allocated among fates (Antebi et al., 2013; Lönnberg et al., 2017; Moris et al., 2016), especially when information about the rate of cell proliferation and/or death can be incorporated as inferred from the profiles.

In animal models, it should be possible to create true lineage trees by marking a common progenitor cell type. For example, one might use synthetic circuits that introduce a molecular barcode only in cells expressing an RNA pattern characteristic of the cell type in order to recognize its descendants (Gagliani et al., 2015; McKenna et al., 2016). In humans, immune cells naturally contain lineage barcodes through VDJ recombination in T and B cells and somatic hypermutation in B cells (Stubbington et al., 2016). More generally, it may be feasible to accomplish lineage tracing in human cells by taking advantage of the steady accumulation of DNA changes (such as somatic point mutations, or repeat expansions at microsatellite loci) at each cell division (Behjati et al., 2014; Biezuner et al., 2016; Martincorena et al., 2015; Reizel et al., 2012; Shlush et al., 2012) or as a molecular clock (Taylor et al., 2003; Teixeira et al., 2013).

Initial computational methods have already been developed for inferring dynamic trajectories from large numbers of single-cell profiles, although better algorithms are still needed. Critical challenges include accurately inferring branching structures, where two or more paths diverge from a single point; reconstructing 'fast' transitions, where only few cells can be captured; and accounting for the fact that a cell may be following multiple dynamic paths simultaneously – for example, differentiation, the cell cycle, and pathogen response (see below) – that may affect each other. The reconstruction algorithms themselves could incorporate insights from theoretical studies of dynamical systems, and learned models could be analyzed in light of such frameworks (Ferrell, 2012; May, 1976; Thom, 1989).

Recent studies provide proofs-of-principle for how simultaneous and orthogonal biological processes can be inferred from single-cell RNA-seq data (**Figure 3;** Angerer et al., 2016; Bendall et al., 2014; Chen et al., 2016b; Haghverdi et al., 2015; Haghverdi et al., 2016; Lönnberg et al., 2017; Marco et al., 2014; Moignard et al., 2015; Setty et al., 2016; Trapnell et al., 2014; Treutlein et al., 2016). Linear developmental trajectories have been reconstructed, for example, from single-cell protein expression during B-cell differentiation (Bendall et al., 2014), and from single-cell RNA expression during myogenesis *in vitro* (Trapnell et al., 2014), early hematopoiesis (Nestorowa et al., 2016), neurogenesis *in vivo* (Habib et al., 2016a; Shin et al., 2015), and reprogramming from fibroblasts to neurons (Treutlein et al., 2016). With a large enough number of cells, analysis of B-cell development was able to highlight a rare (0.007%) population corresponding to the earliest B-cell lymphocytes and confirm the identification by reference to rearrangements at the IgH locus. In direct reprogramming to neurons, scRNA-seq revealed unexpected trajectories (Treutlein et al., 2016). Bifurcated trajectories have also been reconstructed in the differentiation of embryonic stem cells, T helper cells, and hematopoietic cells (Chen et al., 2016b; Haghverdi et al., 2015; Haghverdi et al., 2016; Lönnberg et al., 2017; Marco et al., 2014; Moignard et al., 2015; Setty et al., 2016), and have helped address open questions about whether myeloid progenitor cells in bone marrow are already skewed towards distinct fates (Olsson et al., 2016; Paul et al., 2015) and when T helper cell commit to their fate (Lönnberg et al., 2017).

<L1>**Physiology and homeostasis: cycles, transient responses and plastic states**</L1>

In addition to development and differentiation, cells are constantly undergoing multiple dynamic processes of physiological change and homeostatic regulation (Yosef and Regev, 2011, 2016). These include *cyclical processes*, such as the cell cycle and circadian rhythms; *transient responses* to diverse factors, from nutrients and microbes to mechanical forces and tissue damage; and *plastic states* that can be stably maintained over longer time scales, but can change in response to new environmental cues. (The precise boundary between plastic states and cell types, it must be noted, remains to be clarified.) The molecular phenotype of a cell reflects a superposition of these various processes and their interactions (Wagner et al., 2016).

Studies of physiological processes from bulk tissue samples are hampered by asynchrony and heterogeneity among cells, which blur the signals of individual processes and states; investigators strive to create homogeneous cell populations through synchronization and purification. By contrast, single-cell analysis exploits asynchrony and heterogeneity, leveraging variation within a cell population to reveal underlying structures. The difference is analogous to two approaches in structural biology: X-ray crystallography, which requires molecules to be in a crystalline order, and cryo-electron microscopy, which depends on observing large numbers of molecules in randomly sampled poses.

From asynchronous observations of cyclical and transient processes, it should be possible to 'order' cells with respect to the process (as for development), with cell proportions reflecting residence time (*e.g.*, the length of a phase of the cell cycle). As was initially shown for single-cell measurement of a few features of the cell cycle (Kafri et al., 2013), analysis of many systems could yield a near-continuous model of the process, provided that a sufficient number of cells is sampled. This can occur either because all phases co-occur (*e.g.*, in asynchronously cycling cells) or because enough time points are sampled to span the full process. If very rapid and dramatic discontinuities exist, recovering them would likely require direct tracing, for example by genetic tracers or live analysis in cell cultures, organoids, or animal models.

Once the cells are ordered, one can derive gene-signatures that reflect each phase and use them to further sharpen and refine the model. With sufficient data, it should also be possible to tease apart interactions among processes occurring in parallel (such as the cell cycle, response to a pathogen, and differentiation). For plastic states, it may be possible to capture transient transitions between them, especially if they can be enriched by appropriate physiological cues. Finally, we will likely learn about the nature of stable states: while we often think of stable states as discrete attractor basins (Waddington, 1957), there may also be troughs that reflect a continuous spectrum of stable states (*e.g.*, the ratio of two processes may vary across cells, but are stable in each; Antebi et al., 2013; Gaublomme et al., 2015; Huang, 2012, 2013; Rebhahn et al., 2014; Zhou et al., 2012; Zhou et al., 2016). Some key aspects of processes may be difficult to uncover solely from observations of transitions among molecular states, and will likely require directed perturbations and detailed mechanistic studies.

Recent studies have shown that cyclical processes and transient responses – from the cell cycle (Buettner et al., 2015; Gut et al., 2015; Kafri et al., 2013; Kowalczyk et al., 2015; Macosko et al., 2015; Proserpio et al., 2016; Tirosh et al., 2016a) to the response of immune cells to pathogen components (Avraham et al., 2015; Shalek et al., 2013; Shalek et al., 2014) – can be traced in single-cell profiles. It is possible to order the cells temporally, define coordinately expressed genes with high precision, identify the time scale of distinct phases, and relate these findings to orthogonal measures (**Figure 4**). For example, in the cell cycle, analysis of single-cell profiles readily shows a robust, reproducible and evolutionarily conserved program that can be resolved in a near-continuous way across human and mouse cell lines (Macosko et al., 2015), primary immune cells (Buettner et al., 2015; Kowalczyk et al., 2015), and healthy and disease tissues (Patel et al., 2014; Tirosh et al., 2016a; Tirosh et al., 2016b). This approach has made it possible to determine the relative rates of proliferation of different cell subpopulations within a dataset (Buettner et al., 2015; Kolodziejczyk et al., 2015; Kowalczyk et al., 2015; Tsang et al., 2015), a feat difficult to accomplish using bulk synchronized populations along the cell cycle (Bar-Joseph et al., 2008; Lu et al., 2007). Notably, the cell cycle could also be reconstructed by similar approaches when applied to imaging data of very few molecular markers along with salient spatial features (Gut et al., 2015) or with morphology alone (Blasi et al., 2016; Eulenberg et al., 2017). Similar principles apply to transient responses. In the response of dendritic cells to pathogen components, single-cell profiling uncovered a small subset (<1%) of 'precocious' cells: these early-appearing cells express a distinctive module of genes, initiate production of interferon beta, and coordinate the subsequent response of other cells through paracrine signaling (Shalek et al., 2014).

<L1>**Disease: Cells and cellular ecosystems**</L1>

The Human Cell Atlas will be a critical reference for studying disease, which invariably involves disruption of normal cellular functions, interactions, proportions, or ecosystems. The power of single-cell analysis of disease is evident from decades of histopathological studies and FACS analysis. It will be substantially extended by the routine ability to characterize cells and tissues with rich molecular signatures, rather than focusing on a limited number of pre-defined markers or cell populations. It will also support the growing interest in understanding interactions between frankly abnormal

cells and all other cells in a tissue's ecosystem in promoting or suppressing disease processes (*e.g.*, between malignant cells and the tumor microenvironment).

Single-cell analysis of disease samples will also likely be critical to see the full range of normal cellular physiology, because disease either elicits key perturbs cellular circuitry in informative ways. A clear example is the immune system, where only in the presence of a 'challenge' is the full range of appropriate physiological behaviors and potential responses by a cell revealed.

Single-cell information across many patients will allow us to learn about how cell proportions and states vary and how this variation correlates with genome variants, disease course and treatment response. From initial studies of a limited number of patients, it should be possible to derive signatures of key cell types and states and use them to deconvolute cellular proportions in conventional bulk-tissue or blood samples (Levine et al., 2015; Tirosh et al., 2016a). Future studies may expand single-cell analysis to thousands of patients to directly investigate how genetic variation affects gene transcription and regulation.

The hematopoietic system will be an early and fruitful target. A study involving signatures of cell-signaling assays by single-cell mass cytometry of healthy hematopoietic cells led to more accurate classification of hematopoietic stem and progenitor cells (HSPCs) in Acute Myeloid Leukemia; a previous classification was error-prone, because the 'classical' cell-surface markers of healthy cells do not correctly identify the corresponding population in disease, whereas a richer signature allows accurate identification (Levine et al., 2015). Monitoring rare immune populations first discovered in a normal setting can help zero in on the relevant aberrations in disease. For example, the rare population associated with VDJ recombination first identified by trajectory analysis of B cell development (Bendall et al., 2014) is expanded in pediatric Acute Lymphoblastic Leukemia, and drastically more so in recurrence (Gary Nolan, unpublished results).

The greatest impact, at least in the short term, is likely to be in cancer. Early studies used single-cell qPCR to investigate the origin of radioresistance in cancer stem cells (Diehn et al., 2009) and to dissect the heterogeneity and distortions of cellular hierarchy in colon cancer (Dalerba et al., 2011). With the advent of high-throughput methods, single-cell genome analysis has been used to study the clonal structure and evolution of tumors in both breast cancer (Wang et al., 2014) and acute lymphoblastic

leukemia (Gawad et al., 2014), and to infer the order of earliest mutations that cause acute myeloid leukemia (Corces-Zimmerman et al., 2014; Jan et al., 2012).

In recent studies of melanoma (Tirosh et al., 2016a), glioblastoma (Patel et al., 2014), low-grade glioma (Tirosh et al., 2016b), and myeloproliferative neoplasms (Kiselev et al., 2016), single-cell RNA-seq of fresh tumors resected directly from patients readily distinguished among malignant, immune, stromal and endothelial cells. Among the malignant cells, it identified distinct cell states – such as cancer stem cells (Patel et al., 2014; Tirosh et al., 2016b), drug-resistant states (Tirosh et al., 2016a), proliferating and quiescent cells (Patel et al., 2014; Tirosh et al., 2016a; Tirosh et al., 2016b) – and related them to each other, showing, for example, that only stem-like cells proliferate in low-grade glioma (Tirosh et al., 2016b) and that individual sub-clones can be readily identified in one patient (Kiselev et al., 2016). Among the non-malignant cells, it found distinct functional states for T-cells, and revealed that, while activation and exhaustion programs are coupled, the exhausted state is also controlled by an independent regulatory program in both human tumors (Tirosh et al., 2016a) and a mouse model (Singer et al., 2016). To associate patterns observed in a few (5-20) patients with effects on clinical phenotypes, single-cell based signatures were used to deconvolute hundreds of bulk tumor profiles that had been collected with rich clinical information (Levine et al., 2015; Patel et al., 2014; Tirosh et al., 2016a).

<L1>**Molecular mechanisms: intracellular and inter-cellular circuits**</L1>

A Human Cell Atlas can also shed light on the molecular mechanisms that control cell type, differentiation, responses and states – within cells, between cells, as well as between cells and their tissue matrix.

For example, over the past several decades, biologists have sought to infer the circuitry underlying gene regulation by observing correlations between the expression of particular regulators and specific cellular phenotypes, drawing inferences about regulation, and testing their models through targeted genetic perturbations. Single-cell data provide a massive increase not only in the quantity of observations, but also in the range of perturbations. The number of cells profiled in a single-cell RNA-seq experiment can far exceed the number of profiles produced even by large consortia (such as ENCODE, FANTOM, TCGA, and GTEx). Moreover, each single cell is a perturbation system in which

664 the levels of regulatory molecules vary naturally – sometimes subtly, sometimes
665 dramatically – due to both stochastic and controlled phenomena within a single genetic
666 background, providing rich information from which to reconstruct cellular circuits
667 (Krishnaswamy et al., 2014; Sachs et al., 2005; Shalek et al., 2013; Stewart-Ornstein et al.,
668 2012).

669 Initial studies have shown that such analyses can uncover intracellular regulators
670 governing cell differentiation and response to stimuli. For example, co-variation of RNA
671 levels across a modest number of cells from a relatively 'pure' population of immune
672 dendritic cells responding to a pathogen component was sufficient to connect antiviral
673 transcription factors to their target genes, because of asynchrony in the responses (Shalek
674 et al., 2013). Similarly, co-variation analysis of a few hundred Th17 cells spanning a
675 continuum from less to more pathogenic states revealed regulators that control
676 pathogenicity, but not other features, such as cell differentiation (Gaublomme et al., 2015).
677 Co-variation identified a role for pregnenolone biosynthesis in the response of Th2 cells to
678 helminth infection (Mahata et al., 2014), and new regulators of pluripotency in mouse
679 embryonic stem cells (Kolodziejczyk et al., 2015). Computationally ordering cells along a
680 time-course of development provides another way to infer regulators – a strategy that has
681 been successful in, for example, differentiating B cells (Bendall et al., 2014), myoblasts
682 (Trapnell et al., 2014), neurons (Habib et al., 2016a; Shin et al., 2015), and T helper cells
683 (Lönnberg et al., 2017). Finally, when circuitry is already known, variation across single
684 cells can be used to infer exquisite – and functionally important – quantitative distinctions
685 about how signal is processed and propagated. An elegant example is a recent analysis of
686 signaling pathways downstream from the T cell receptor, where single-cell proteomics
687 data has shown how the same cellular circuitry processes signals differently in naïve and
688 antigen-exposed T cells (Krishnaswamy et al., 2014).

689 Beyond transcriptome analysis, single-cell multi-omic profiles (**Box 1**) will improve
690 the inference of cellular circuitry by connecting regulatory mechanisms and their targets
691 (Tanay and Regev, 2017). For example, simultaneous measurement of chromatin
692 accessibility and RNA levels may help identify which regulatory regions – and by inference
693 which *trans*–acting regulators – control the levels of which genes. Concomitant
694 measurement of DNA mutations and transcriptional profiles in cancer cells may allow

similar causal connections to be drawn, as has been recently shown for mutations in the *CIC* gene and the expression of its regulatory targets (Tirosh et al., 2016b).

Studies can be extended from naturally occurring variation among cells to engineered perturbations, by using pooled CRISPR libraries to manipulate genes and reading out both the perturbation and its effects on cellular phenotype in single cells – for example, by single-cell RNA-Seq (Adamson et al., 2016; Dixit et al., 2016; Jaitin et al., 2016).

A cell atlas can also help shed light on intercellular communication, based on correlated profiles across cell types and patients. For example, analysis of single-cell profiles from many small clusters of a few aggregated cells allowed the construction of a cell-cell interaction network in the bone marrow, uncovering specific interaction between megakaryocytes and neutrophils, as well as between plasma cells and neutrophil precursors (Alexander van Oudenaarden, unpublished results). Cell-cell interactomes have also been inferred from profiles of purified cell populations, based on the secreted and cell surface molecules that they express (Ramilowski et al., 2015).

In tumors from melanoma patients, gene-expression analysis (involving single-cell data obtained from some patients and bulk tumor data from many more patients, deconvoluted based on signatures learned from the single cells) found genes that are expressed in one cell type, but whose expression levels are correlated with the proportion of a different cell type that does not express them; this analysis revealed that high expression of the complement system in cancer-associated fibroblasts in the tumor microenvironment is correlated with increased infiltration of T cells (Tirosh et al., 2016a). Analysis of individual subcutaneous adipose stem cells revealed the existence of a novel cell population that negatively controls the differentiation of the resident stem cells into adipocytes, thus influencing adipose tissue growth and homeostasis (Bart Deplancke, unpublished results). In breast cancer tissues, spatial analysis of multiplex protein expression by imaging mass cytometry (Giesen et al., 2014) allowed classification of infiltrating immune cells and malignant cells based on the neighborhood of surrounding cells, highlighting new functional interactions (Bernd Bodenmiller, personal communication).

<L1>**A user's guide to the Human Cell Atlas: applications in research and medicine**</L1>

The Human Genome Project had a major impact on biomedicine by providing a comprehensive reference, a DNA sequence in which answers could be readily looked up and from which unique 'signatures' could be derived (*e.g.*, to recognize genes on microarrays or protein fragments in mass spectrometry). A Human Cell Atlas could provide similar benefits from basic research to clinically relevant applications.

Scientists will be able, for example, to look up precisely in which cell types a gene of interest is expressed and at which level. Today, it is surprisingly challenging to obtain definitive answers for most human genes beyond tissue- or organ-level resolution, although there have been pioneering efforts for the brain and immune system in mouse (Bakken et al., 2016; Hawrylycz et al., 2012; Kim and Lanier, 2013; Miller et al., 2014) and for protein expression in human (Thul et al., 2017; Uhlen et al., 2015). Yet, the question is of enormous importance to basic biologists studying development or comparing a model system to human biology, medical scientists examining the effect of a disease-causing mutation, and drug developers concerned about the potential toxicities of a small molecule or a CAR-T cell targeting a specific protein (Brudno and Kochenderfer, 2016).

Researchers will also be able to derive expression signatures that uniquely identify cell types. Such signatures provide a starting point for a vast range of experimental assays – from molecular markers for isolating, tagging, tracing or manipulating cells in animal models or human samples, to characterization of the effect of drugs on the physiological state of a tissue. Such descriptors of cellular identity will be widely used in clinical assays. For example, today's Complete Blood Count (CBC), a census of a limited number of blood components, may be supplemented by a 'CBC 2.0' that provides a high-resolution picture of the nucleated cells, including the number and activity states of each type in comparison with healthy reference samples. Analogous measures should be possible for other tissues as well. For example, gut biopsies from patients with ulcerative colitis or colon cancer could be analyzed for the type, response, state and location of each of the diverse epithelial, immune, stromal and neural cells that comprise them.

<L1>**Toward a Human Cell Atlas**</L1>

How might the biomedical community build a Human Cell Atlas? As with the Human Genome Project, a robust plan will best emerge from wide-ranging scientific discussions and careful planning involving biologists, technologists, pathologists,

physicians, surgeons, computational scientists, statisticians, and others. As noted above, various discussions have taken place for over two years about the idea of a comprehensive Human Cell Atlas, as well as about specific atlases for the brain and the immune system. Several pilot efforts are already underway. Moreover, over the past year discussions have been underway to create an initial plan for a Human Cell Atlas Project (which is articulated in the White Paper mentioned above). Among the key points for consideration are the following:

<L2>*Phasing of goals*</L2>

While the overall goal is to build a comprehensive atlas with diverse molecular measurements, spatial organization, and interpretation of cell types, histology, development, physiology and molecular mechanisms, it will be wise to set intermediate goals for 'draft' atlases at increasing resolution, comprehensiveness, and depth of interpretation. The value of a phased approach was illustrated by the Human Genome Project, which defined milestones along the way (genetic maps, physical maps, rough-draft sequence, finished sequence) that held the project accountable and provided immediate utility to the scientific community.

<L2>*Sampling strategies*</L2>

While an adult human has ~2 x 10^{13} nucleated cells, it is neither possible nor necessary to study them all to recover the fine distinctions among human cells. The key will be to combine sound statistical sampling, biological enrichment purification, and insights from studies of model organisms. It is likely beneficial to apply an adaptive, iterative approach with respect to both the number of cells and depth of profiles, as well as anatomical coverage and spatial resolution in the tissue, with initial sparse sampling driving decisions about further sampling. This adaptive approach, termed a 'Sky Dive', adjusts as resolution increases (and is further described in the HCA White Paper).

Such approaches can be facilitated by experimental techniques that allow fast and inexpensive 'banking' of partially processed samples, to which one can return for deeper analysis as methods mature. Advances in handling fixed or frozen tissues would further facilitate the process (**Box 1**). With respect to depth of profiling, recent studies suggest the utility of a mixed strategy: relatively low coverage of the transcriptome can identify many cell types reliably (Heimberg et al., 2016; Shekhar et al., 2016) and a smaller set of deep profiles can be help interpret the low-coverage data to further increase detection power.

As a result, the 'Sky Dive' begins with large-scale uniform sampling, follows with stratified sampling, and then employs specialized sampling at lower throughput.

<L2>*Breadth of profiles*</L2>

The atlas must combine two branches – a cellular branch, focused on the properties of individual cells, and a spatial branch, describing the histological organization of cells in the tissue. For the cellular branch, massively parallel transcriptome analysis of individual single cells or nuclei will likely be the workhorse for efforts in the first few years. However, other robust, high-throughput profiling methods are rapidly emerging, including techniques for studying chromatin, genome folding, and somatic mutations at single-cell resolution (**Box 1**). For the spatial branch, *in situ* analysis of the spatial patterns of RNA, proteins, and potentially epigenomics will be equally important. While some of these methods are already rapidly maturing, others will benefit from focused development efforts, as well as from cross-comparison among different techniques. Fortunately, most can be applied to preserved tissue specimens, allowing specimens collected now to be analyzed later, as methods mature.

<L2>*Biological scope*</L2>

It will be important to consider the balance among tissue samples from healthy individuals at various stages; small cohorts of individuals with diseases; and samples from model organisms, where key developmental stages are more accessible and manipulations more feasible. Well-chosen pilot projects could help refine strategies and galvanize communities of biological experts. Some communities and projects would be organized around organs (*e.g.*, liver, heart, brain), others around systems (*e.g.*, the immune system) or disease (*e.g.*, cancer), the latter distributed across many organs and tissues.

As outlined in the HCA White Paper, the first draft of the atlas might pursue roughly a dozen organs and systems, each from up to 100 individuals, collected across 3–4 geographical sites; each would be analyzed to obtain both cellular and spatial data, by means of uniform to stratified sampling. Tissue from post-mortem examination will play a key role, because it is the only way to obtain samples from a single individual across the entire body. These efforts will be complemented, where possible, by biopsy or resection material from healthy research participants, and by whole organs obtained from deceased transplant donors after transplantation organs have been harvested. In some cases, such as

the immune system, samples from individuals with a disease will be included to probe different functional states of a system.

The full atlas, will ultimately describe at least 10 billion cells, covering all tissues, organs, and systems. Specimens will come from both healthy research participants and small cohorts of patients with relevant diseases. The cells and tissues will be studied using a broad range of techniques, to obtain cellular and spatial information, from samples designed to represent the world's diversity. As with previous genomic projects, the Human Cell Atlas will be bounded in its resolution (with respect to the rarity of cell types/states and the spatial resolution), its coverage of disease and diversity (broadly representative but not obviating the need for direct genetic and clinical studies), and its functional assessment (to validate the existence of identified cells and facilitate – but not perform – detailed functional characterization).

<L2>*Model organisms*</L2>

The Human Genome Project and the broader scientific community benefitted from insights learned from genome projects conducted in parallel in model organisms. These projects empowered functional studies in model organisms, ushered a new era of comparative genomics, and provided important technical lessons. By analogy, we envision that key 'sister' atlases in model organisms will be developed in parallel and in coordination with the Human Cell Atlas. These projects should not delay progress on the human atlas (or *vice versa*), because current techniques are already directly applicable to biomedical research on human samples.

In some cases, model organism atlases can use techniques that are not possible in humans, such as engineering animals to facilitate lineage tracing. In many cases, the extensive validation and functional follow-up studies that can be performed in model organisms will help validate 'by proxy' conclusions drawn in the human atlas. Finally, comparing the atlases across organisms will provide invaluable lessons in evolution and function.

<L2>*Quality*</L2>

In creating a reference map to be used by thousands of investigators, it is critical to ensure that the results are of high quality and technically reproducible. This is especially

important in view of the inherent biological variation and expected measurement noise. Substantial investment will be needed in the development, comparison, and dissemination of rigorous protocols, standards, and benchmarks. Both individual groups and larger centers will likely have important roles in defining and ensuring high quality. It will also be important that the collected samples be accompanied by excellent clinical annotations, captured in consistent meta-data across the atlas.

Tissue processing poses special challenges, including the need for robust methods for dissociating samples into single cells so as to preserve all cell types, fixation for *in situ* methods, and freezing for transport. A related challenge is the difference in the amenability of specific cell types for different assays (T cells are very small and yield lower quality scRNA-seq; the fat content in adipocyte is challenging for many spatial methods; many neurons cannot currently be isolated with their axons and dendrites from adult tissue). Careful attention will also be needed to data generation and computational analysis, including validated standard operating procedures for experimental methods, best practices, computational pipelines, and benchmarking samples and data sets to ensure comparability.

<L2>*Global equity*</L2>

Geographical atlases of the Earth were largely developed to serve global power centers. The Human Cell Atlas should be designed to serve all people: it should span genders, ethnicities, environments, and the global burden of diseases – all of which are likely to affect the molecular profiles of cells and must be characterized to maximize the atlas's benefits. The project itself should encourage and support the participation of scientists, research centers and countries from around the globe, while recognizing the value of respecting and learning from diverse populations, cultures, mores, beliefs, and traditions.

<L2>*Open data*</L2>

The Human Genome Project made clear the power of open data that can be used by all and freely combined with other datasets. A Human Cell Atlas should similarly be an open endeavor, to the full extent permitted by participants' wishes and legal regulation. While the underlying sequence data contains many polymorphisms that make it 'identifiable', it should be possible to map the data onto 'standard models' of each gene to substantially mitigate this issue. To make the atlas useful, it will be critical to develop data

884 platforms that can provide efficient aggregation and storage, quality control, analytical
885 software, and user-friendly portals.
886 <L2>*Flexibility*</L2>
887 A Human Cell Atlas Project should be intellectually and technologically flexible.
888 The project should embrace the fact that its biological goals, experimental methods,
889 computational approaches, overall scale, and criteria for 'completion' will evolve rapidly as
890 insights and tools develop. For historical context, it is useful to remember that discussions
891 about a Human Genome Project began before the development of automated DNA
892 sequencing machines, the polymerase chain reaction, or large-insert DNA cloning, and the
893 project drove technological progress on many fronts. Moreover, the criteria for a 'finished'
894 genome sequence were only agreed upon during the last third of the project.
895 <L2>*Impact on the scientific community*</L2>
896 Large-scale efforts, such as a Human Cell Atlas, must be careful to appropriately
897 weigh the benefits to science and individual scientists with the potential costs. We
898 consider the key benefits to the broad scientific community to include: the core scientific
899 knowledge and discoveries that will result from having a reference map; the
900 empowerment of scientists working across any tissue or cell type to pursue their research
901 more precisely and effectively; the development, hardening and dissemination of
902 experimental techniques and computational methods in the context of big-data settings,
903 all of which will be openly shared; the inclusive and maximally open Human Cell Atlas
904 community, inviting participation by all individual labs and research centers; and the
905 coordination of efforts that would otherwise be unconnected, less extensive, and more
906 expensive.
907 At the same time, we must be aware of potential pitfalls, including: premature
908 restriction to specific technologies or approaches, which might limit innovation in a fast-
909 moving field; implicit restriction of participation, based on available resources; and
910 diversion of funding from other research directions. The unique organization and
911 community of the Human Cell Atlas Project will tackle these potential challenges by
912 committing to open membership, to the open and immediate data release with no
913 restrictions, and to open-source code for all computational approaches. We hope that the
914 new information and technology generated will more than repay the costs of the project

by increasing the speed and efficiency of biomedical research throughout the scientific community.

<L2>*Engagement with the non-scientific community*</L2>

The general public is a key stakeholder community for the Human Cell Atlas. Proper public engagement should involve many communities, including interested members of the public, citizen-scientists, schoolchildren, teachers and, where appropriate, research participants. Engagement will take diverse forms, including traditional media, social media, video and, importantly, direct sharing of the project's data. Across all channels, it will be important to articulate the goals, principles and motivations of the project. While explaining the intended benefits to the public with respect to advancing disease biology, drug discovery and diagnostics, it will be equally important to avoid 'hype': that is, we need to avoid making promises and raising expectations that are unrealistic in content or timing.

<L2>*Forward looking*</L2>

Any data produced today will be easier, faster, more accurate and cheaper to produce tomorrow. Any intermediate milestones achieved during the project will be supplanted by deeper, broader, more accurate, and more comprehensive successors within a few short years. However, as we define the goal of a Human Cell Atlas Project, we should view it not as a final product, but as a critical stepping-stone to a future when the study of human biology and medicine is increasing tractable.

<L2>**Conclusion**</L2>

The past quarter-century has shown again and again the value of the scientific community joining together in collaborative efforts to generate and make freely available systematic information resources to accelerate scientific and medical progress in tens of thousands of laboratories around the world. The Human Cell Atlas builds on this rich tradition, extending it to the fundamental unit of biological organization: the cell.

Many challenges will arise along the way, but we are confident that they can be met through scientific creativity and collaboration. It is time to begin.

Box 1: Key experimental methods for single-cell genomics

Over the past several years, powerful approaches have emerged that make it possible to measure molecular profiles and signatures at single-cell resolution. The field remains very active, with new methods being rapidly developed and existing ones improved.

Single-cell RNA-Seq (scRNA-seq) refers to a class of methods for profiling the transcriptome of individual cells. Some may take a census of mRNA species by focusing on 3'- or 5'-ends (Islam et al., 2014; Macosko et al., 2015), while others assess mRNA structure and splicing by collecting near-full-length sequence (Hashimshony et al., 2012; Ramskold et al., 2012). Strategies for single-cell isolation span manual cell picking, initially used in microarray studies (Eberwine et al., 1992; Van Gelder et al., 1990), FACS-based sorting into multi-well plates (Ramskold et al., 2012; Shalek et al., 2013), microfluidic devices (Shalek et al., 2014; Treutlein et al., 2014), and, most recently, droplet-based (Klein et al., 2015; Macosko et al., 2015) and microwell-based (Fan et al., 2015; Yuan and Sims, 2016) approaches. The droplet and microwell approaches, which are currently coupled to 3'-end counting, have the largest throughput, allowing rapid processing of tens of thousands of cells simultaneously in a single sample. scRNA-seq is typically applied to freshly dissociated tissue, but emerging protocols use fixed cells (Nichterwitz et al., 2016; Thomsen et al., 2016) or nuclei isolated from frozen or lightly fixed tissue (Habib et al., 2016b; Lake et al., 2016). Applications to fixed or frozen samples would simplify the process flow for scRNA-seq, as well as open the possibility of using archival material. Power analyses provides a framework for comparing the sensitivity and accuracy of these approaches (Svensson et al., 2016; Ziegenhain et al., 2016). Finally, there has been progress in scRNA-Seq with RNA isolated from live cells in their natural microenvironment using transcriptome in vivo analysis (Lovatt et al., 2014).

Mass cytometry (CyTOF) and related methods allow multiplexed measurement of proteins based on antibodies barcoded with heavy metals (Bendall et al., 2014; Levine et al., 2015). In contrast to comprehensive profiles, these methods involve pre-defined signatures and require an appropriate antibody for each target, but they can process many millions of cells for a very low cost per cell. They are applied to fixed cells. Recently, the approach has been extended to the measurement of RNA signatures through multiplex hybridization of nucleic-acid probes tagged with heavy metals (Frei et al., 2016).

Single-cell genome and epigenome sequencing characterizes the cellular genome. Genomic methods aim either to characterize the whole genome or capture specific pre-defined regions (Gao et al., 2016). Epigenomic methods may capture regions based on distinctive histone modifications (single-cell ChIP-Seq; Rotem et al., 2015a), accessibility (single-cell ATAC-Seq; Buenrostro et al., 2015; Cusanovich et al., 2015), or likewise characterize DNA methylation patterns (single-cell DNAme-Seq; Farlik et al., 2015; Guo et al., 2013; Mooijman et al., 2016; Smallwood et al., 2014) or 3D organization (single-cell Hi-C; Nagano et al., 2013; Ramani et al., 2016). Combinatorial barcoding strategies have been used to capture measures of accessibility and 3D organization in tens of thousands of single cells (Cusanovich et al., 2015; Ramani et al., 2016). Single cell epigenomics methods are usually applied to nuclei, and can thus use frozen or certain fixed samples. Some methods, such as single-cell DNA sequencing, are currently applied to relatively few cells, due to the size of the genome and the sequencing depth required. Other methods, such as single-cell analysis of chromatin organization (by either single-cell ATAC-Seq; Buenrostro et al., 2015; Cusanovich et al., 2015) or single-cell ChIP-Seq (Rotem et al., 2015a), currently yield rather sparse data, which presents analytic challenges and benefits from large numbers of profiled cells. Computational analyses have begun to address these issues by pooling of signal across cells and across genomic regions or loci (Buenrostro et al., 2015; Rotem et al., 2015a) and by imputation (Angermueller et al., 2016).

Single-cell multi-omics techniques aim to collect two or more types of data (transcriptomic, genomic, epigenomic, and proteomic) from the same single cell. Recent studies have simultaneously profiled the transcriptome together with either the genome (Angermueller et al., 2016; Dey et al., 2015; Macaulay et al., 2015), the epigenome (Angermueller et al., 2016), or protein signatures (Albayrak et al., 2016; Darmanis et al., 2016; Frei et al., 2016; Genshaft et al., 2016). Efforts to combine three and more approaches are underway (Cheow et al., 2016). Multi-omic methods could help fill in causal chains from genetic variation to regulatory mechanisms and phenotypic outcome in health and in disease, especially cancer.

Multiplex in situ analysis and other spatial techniques aim to detect a limited number of nucleic acids and/or proteins *in situ* in tissue samples – by hybridization (for RNA), antibody staining (for proteins), sequencing (for nucleic acids), or other tagging strategies. These *in situ* results can then be used to map massive amounts of single-cell genomic

information from dissociated cells onto the tissue samples providing important clues about spatial relationships and cell-cell communication. Some strategies for RNA detection, such as MERFISH (Chen et al., 2015b; Moffitt et al., 2016b) or Seq-FISH (Shah et al., 2016), combine multiplex hybridization with microscopy-based quantification to assess distributions at both the cellular and subcellular level; other early studies have performed *in situ* transcription (Tecott et al., 1988), followed by direct manual harvesting of cDNA from individual cells (Crino et al., 1996; Tecott et al., 1988). Some approaches for protein detection, such as Imaging Mass Cytometry (Giesen et al., 2014) and Mass Ion Bean Imaging (Angelo et al., 2014), involve staining a tissue specimen with antibodies, each labeled with a barcode of heavy metals, and rastering across the sample to measure the proteins in each 'pixel'. This technique permits the reconstruction of remarkably rich images. Finally, more recent studies have performed RNA-seq *in situ* in cells and in preserved tissue sections (Ke et al., 2013; Lee et al., 2014). Many *in situ* methods can benefit from tissue clearing and/or expansion to improve detection and spatial resolution (Chen et al., 2015a; Chen et al., 2016a; Moffitt et al., 2016a; Yang et al., 2014). The complexity and accuracy of these methods continues to improve with advances in sample handling, chemistry and imaging. Various methods are also used, for example, to measure transcriptomes *in situ* with barcoded arrays (Stahl et al., 2016).

Cell lineage determination. Because mammals are not transparent and have many billions of cells, it is not currently possible to directly observe the fate of cells by microscopy. Various alternative approaches have been developed (Kretzschmar and Watt, 2012). In mice, cells can be genetically marked with different colors (Barker et al., 2007) or DNA barcodes (Lu et al., 2011; Naik et al., 2013; Perie and Duffy, 2016), and their offspring traced during development. Recent work has used iterative CRISPR-based genome editing to generate random genetic scars in the fetal genome and use them to reconstruct lineages in the adult animal (McKenna et al., 2016). In humans, where such methods cannot be applied, human cell lineages can be monitored experimentally *in vitro*, or by transplantation of human cells to immunosuppressed mice (Morton and Houghton, 2007; O'Brien et al., 2007; Richmond and Su, 2008), or can be inferred from *in vivo* samples by measuring the DNA differences between individual sampled cells, arising from random mutations during cell division, and using the genetic distances to construct cellular phylogenies, or lineages (Behjati et al., 2014; Shapiro et al., 2013).

Box 2: On Exactitude in Science. Jorge Luis Borges (1946)

". . . In that Empire, the Art of Cartography attained such Perfection that the map of a single Province occupied the entirety of a City, and the map of the Empire, the entirety of a Province. In time, those Unconscionable Maps no longer satisfied, and the Cartographers Guilds struck a Map of the Empire whose size was that of the Empire, and which coincided point for point with it. The following Generations, who were not so fond of the Study of Cartography as their Forebears had been, saw that that vast map was Useless, and not without some Pitilessness was it, that they delivered it up to the Inclemencies of Sun and Winters. In the Deserts of the West, still today, there are Tattered Ruins of that Map, inhabited by Animals and Beggars; in all the Land there is no other Relic of the Disciplines of Geography."

Purportedly from Suárez Miranda, Travels of Prudent Men, Book Four, Ch. XLV, Lérida, 1658.

Figure legends

Figure 1: A hierarchical view of human anatomy. A graphical depiction of the anatomical hierarchy from organs (such as the gut), to tissues (such as the epithelium in the crypt in the small intestine), to their constituent cells (such as epithelial, immune, stromal and neural cells).

Figure 2: Anatomy: cell types and tissue structure. The first three plots show single cells (dots) embedded in low-dimensional space based on similarities between their RNA-expression profiles (A, C) or protein-expression profiles (B), using either t-stochastic neighborhood embedding (A,B) or circular projection (C) for dimensionality reduction and embedding. (A) Bi-polar neurons from the mouse retina; reprinted from Shekhar et al., 2016 with permission from Elsevier. (B) Human bone marrow immune cells; reprinted from Levine et al., 2015 with permission from Elsevier. (C) Immune cells from the mouse spleen; reprinted from Jaitin et al., 2014 with permission from AAAS. (D) Histology. Projection of single-cell data onto tissue structures: image shows the mapping of individual cells onto locations in the marine annelid brain, based on the correspondence (color bar) between their single-cell expression profiles and independent FISH assays for a set of landmark transcripts; adapted from Achim et al., 2015 with permission from Macmillan Publishers Limited.

Figure 3: Developmental trajectories. Each plot shows single cells (dots; colored by trajectory assignment, sampled time point, or developmental stage) embedded in low-dimensional space based on their RNA (A-C) or protein (D) profiles, using different methods for dimensionality reduction and embedding: Gaussian process patent variable model (A); t-stochastic neighborhood embedding (B, D); diffusion maps (C). Computational methods then identify trajectories of pseudo-temporal progression in each case. (A) Myoblast differentiation *in vitro;* reprinted from Lönnberg et al., 2017 with permission from AAAS. (B) Neurogenesis in the mouse brain dentate gyrus; reprinted from Habib et al., 2016a with permission from AAAS. (C) Embryonic stem cell differentiation *in vitro;* adapted from Haghverdi et al., 2016 with permission from

Macmillan Publishers Limited. (D) Early hematopoiesis; adapted from Setty et al., 2016 with permission from Macmillan Publishers Limited.

Figure 4. Physiology. Each plot shows single cells (dots) embedded in low-dimensional space on the basis of their RNA profile, based on predefined gene signatures (A) or PCA (B, C), highlighting distinct dynamic processes. (A) The cell cycle in mouse hematopoietic stem and progenitor cells; adapted under terms of CC BY 4.0 from Scialdone et al., 2015. (B) Response to lipopolysaccharide (LPS) in mouse immune dendritic cells; adapted from Shalek et al., 2014 with permission from Macmillan Publishers Limited. (C) Variation in the extent of pathogenicity in mouse Th17 cells; reprinted from Gaublomme et al., 2015 with permission from Elsevier.

References

Achim, K., Pettit, J.B., Saraiva, L.R., Gavriouchkina, D., Larsson, T., Arendt, D., and Marioni, J.C. (2015). High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. Nature Biotechnology *33*, 503-509.

Adamson, B., Norman, T.M., Jost, M., Cho, M.Y., Nunez, J.K., Chen, Y., Villalta, J.E., Gilbert, L.A., Horlbeck, M.A., Hein, M.Y., *et al.* (2016). A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. Cell *167*, 1867-1882 e1821.

Albayrak, C., Jordi, C.A., Zechner, C., Lin, J., Bichsel, C.A., Khammash, M., and Tay, S. (2016). Digital Quantification of Proteins and mRNA in Single Mammalian Cells. Molecular Cell *61*, 914-924.

Angelo, M., Bendall, S.C., Finck, R., Hale, M.B., Hitzman, C., Borowsky, A.D., Levenson, R.M., Lowe, J.B., Liu, S.D., Zhao, S., *et al.* (2014). Multiplexed ion beam imaging of human breast tumors. Nature Medicine *20*, 436-442.

Angerer, P., Haghverdi, L., Buttner, M., Theis, F.J., Marr, C., and Buettner, F. (2016). destiny: diffusion maps for large-scale single-cell data in R. Bioinformatics *32*, 1241-1243.

Angermueller, C., Clark, S.J., Lee, H.J., Macaulay, I.C., Teng, M.J., Hu, T.X., Krueger, F., Smallwood, S.A., Ponting, C.P., Voet, T., *et al.* (2016). Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. Nature Methods *13*, 229-232.

Antebi, Y.E., Reich-Zeliger, S., Hart, Y., Mayo, A., Eizenberg, I., Rimer, J., Putheti, P., Pe'er, D., and Friedman, N. (2013). Mapping differentiation under mixed culture conditions reveals a tunable continuum of T cell fates. PLoS Biology *11*, e1001616.

Arthur, G. (2016). Albert Coons: harnessing the power of the antibody. Lancet Respir Med *4*, 181-182.

Ascoli, G.A., Alonso-Nanclares, L., Anderson, S.A., Barrionuevo, G., Benavides-Piccione, R., Burkhalter, A., Buzsaki, G., Cauli, B., Defelipe, J., *et al.* (2008). Petilla terminology: nomenclature of features of GABAergic interneurons of the cerebral cortex. Nature Reviews Neuroscience *9*, 557-568.

Avraham, R., Haseley, N., Brown, D., Penaranda, C., Jijon, H.B., Trombetta, J.J., Satija, R., Shalek, A.K., Xavier, R.J., Regev, A., *et al.* (2015). Pathogen Cell-to-Cell Variability Drives Heterogeneity in Host Immune Responses. Cell *162*, 1309-1321.

Bakken, T.E., Miller, J.A., Ding, S.L., Sunkin, S.M., Smith, K.A., Ng, L., Szafer, A., Dalley, R.A., Royall, J.J., Lemon, T., *et al.* (2016). A comprehensive transcriptional map of primate brain development. Nature *535*, 367-375.

Bar-Joseph, Z., Siegfried, Z., Brandeis, M., Brors, B., Lu, Y., Eils, R., Dynlacht, B.D., and Simon, I. (2008). Genome-wide transcriptional analysis of the human cell cycle identifies genes differentially regulated in normal and cancer cells. PNAS *105*, 955-960.

Barker, N., van Es, J.H., Kuipers, J., Kujala, P., van den Born, M., Cozijnsen, M., Haegebarth, A., Korving, J., Begthel, H., Peters, P.J., *et al.* (2007). Identification of stem cells in small intestine and colon by marker gene Lgr5. Nature *449*, 1003-1007.

Behjati, S., Huch, M., van Boxtel, R., Karthaus, W., Wedge, D.C., Tamuri, A.U., Martincorena, I., Petljak, M., Alexandrov, L.B., Gundem, G., *et al.* (2014). Genome sequencing of normal cells reveals developmental lineages and mutational processes. Nature *513*, 422-425.

Bendall, S.C., Davis, K.L., Amir el, A.D., Tadmor, M.D., Simonds, E.F., Chen, T.J., Shenfeld, D.K., Nolan, G.P., and Pe'er, D. (2014). Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. Cell *157*, 714-725.

Biezuner, T., Spiro, A., Raz, O., Amir, S., Milo, L., Adar, R., Chapal-Ilani, N., Berman, V., Fried, Y., Ainbinder, E., *et al.* (2016). A generic, cost-effective, and scalable cell lineage analysis platform. Genome Research *26*, 1588-1599.

Blasi, T., Hennig, H., Summers, H.D., Theis, F.J., Cerveira, J., Patterson, J.O., Davies, D., Filby, A., Carpenter, A.E., and Rees, P. (2016). Label-free cell cycle analysis for high-throughput imaging flow cytometry. Nat Commun *7*, 10256.

Blondel, V.D., Guillaume, J.L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. J Stat Mech-Theory E. 2008:P10008

Borges, J.L., and Hurley, A. (2004). *A Universal History of Iniquity* New York: Penguin.

Brenner, S. (2010). Sequences and consequences. Philosophical Transactions of the Royal Society of London Series B, Biological Sciences *365*, 207-212.

Brudno, J.N., and Kochenderfer, J.N. (2016). Toxicities of chimeric antigen receptor T cells: recognition and management. Blood *127*, 3321-3330.

Buenrostro, J.D., Wu, B., Litzenburger, U.M., Ruff, D., Gonzales, M.L., Snyder, M.P., Chang, H.Y., and Greenleaf, W.J. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. Nature *523*, 486-490.

Buettner, F., Natarajan, K.N., Casale, F.P., Proserpio, V., Scialdone, A., Theis, F.J., Teichmann, S.A., Marioni, J.C., and Stegle, O. (2015). Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. Nature Biotechnology *33*, 155-160.

Cadwell, C.R., Palasantza, A., Jiang, X., Berens, P., Deng, Q., Yilmaz, M., Reimer, J., Shen, S., Bethge, M., Tolias, K.F., *et al.* (2016). Electrophysiological, transcriptomic and morphologic profiling of single neurons using Patch-seq. Nature Biotechnology *34*, 199-203.

Chao, M.P., Seita, J., and Weissman, I.L. (2008). Establishment of a normal hematopoietic and leukemia stem cell hierarchy. Cold Spring Harbor Symposia on Quantitative Biology *73*, 439-449.

Chen, F., Tillberg, P.W., and Boyden, E.S. (2015a). Expansion microscopy. Science *347*, 543-548.

Chen, F., Wassie, A.T., Cote, A.J., Sinha, A., Alon, S., Asano, S., Daugharthy, E.R., Chang, J.B., Marblestone, A., Church, G.M., *et al.* (2016a). Nanoscale imaging of RNA with expansion microscopy. Nature Methods *13*, 679-684.

Chen, J., Schlitzer, A., Chakarov, S., Ginhoux, F., and Poidinger, M. (2016b). Mpath maps multi-branching single-cell trajectories revealing progenitor cell progression during development. Nat Commun *7*, 11988.

Chen, K.H., Boettiger, A.N., Moffitt, J.R., Wang, S., and Zhuang, X. (2015b). RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. Science *348*, aaa6090.

Cheow, L.F., Courtois, E.T., Tan, Y., Viswanathan, R., Xing, Q., Tan, R.Z., Tan, D.S., Robson, P., Loh, Y.H., Quake, S.R., *et al.* (2016). Single-cell multimodal profiling reveals cellular epigenetic heterogeneity. Nature Methods *13*, 833-836.

Corces, M.R., Buenrostro, J.D., Wu, B., Greenside, P.G., Chan, S.M., Koenig, J.L., Snyder, M.P., Pritchard, J.K., Kundaje, A., Greenleaf, W.J., *et al.* (2016). Lineage-specific and single cell chromatin accessibility charts human hematopoiesis and leukemia evolution. Nature Genetics. 48, 1193-1203

Corces-Zimmerman, M.R., Hong, W.J., Weissman, I.L., Medeiros, B.C., and Majeti, R. (2014). Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission. PNAS *111*, 2548-2553.

1197 Crino, P.B., Trojanowski, J.Q., Dichter, M.A., and Eberwine, J. (1996). Embryonic neuronal
1198 markers in tuberous sclerosis: single-cell molecular pathology. PNAS *93*, 14152-14157.
1199 Cusanovich, D.A., Daza, R., Adey, A., Pliner, H.A., Christiansen, L., Gunderson, K.L.,
1200 Steemers, F.J., Trapnell, C., and Shendure, J. (2015). Multiplex single cell profiling of
1201 chromatin accessibility by combinatorial cellular indexing. Science *348*, 910-914.
1202 Dalerba, P., Kalisky, T., Sahoo, D., Rajendran, P.S., Rothenberg, M.E., Leyrat, A.A., Sim, S.,
1203 Okamoto, J., Johnston, D.M., Qian, D.L., *et al.* (2011). Single-cell dissection of
1204 transcriptional heterogeneity in human colon tumors. Nature Biotechnology *29*, 1120-1127.
1205 Darmanis, S., Gallant, C.J., Marinescu, V.D., Niklasson, M., Segerman, A., Flamourakis, G.,
1206 Fredriksson, S., Assarsson, E., Lundberg, M., Nelander, S., *et al.* (2016). Simultaneous
1207 Multiplexed Measurement of RNA and Proteins in Single Cells. Cell Reports *14*, 380-389.
1208 Darmanis, S., Sloan, S.A., Zhang, Y., Enge, M., Caneda, C., Shuer, L.M., Gephart, M.G.H.,
1209 Barres, B.A., and Quake, S.R. (2015). A survey of human brain transcriptome diversity at
1210 the single cell level. PNAS *112*, 7285-7290.
1211 Della Santina, L., Kuo, S.P., Yoshimatsu, T., Okawa, H., Suzuki, S.C., Hoon, M.,
1212 Tsuboyama, K., Rieke, F., and Wong, R.O.L. (2016). Glutamatergic Monopolar
1213 Interneurons Provide a Novel Pathway of Excitation in the Mouse Retina. Current Biology
1214 *26*, 2070-2077.
1215 Dey, S.S., Kester, L., Spanjaard, B., Bienko, M., and van Oudenaarden, A. (2015). Integrated
1216 genome and transcriptome sequencing of the same cell. Nature Biotechnology *33*, 285-289.
1217 Diehn, M., Cho, R.W., Lobo, N.A., Kalisky, T., Dorie, M.J., Kulp, A.N., Qian, D.L., Lam, J.S.,
1218 Ailles, L.E., Wong, M.Z., *et al.* (2009). Association of reactive oxygen species levels and
1219 radioresistance in cancer stem cells. Nature *458*, 780-783.
1220 Dittrich, W.M., and Göhde, W.H. (1971). Flow-through Chamber for Photometers to
1221 Measure and Count Particles in a Dispersion Medium. US Patent.
1222 https://www.google.com/patents/US3761187
1223 Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C.P., Jerby-Arnon, L., Marjanovic, N.D.,
1224 Dionne, D., Burks, T., Raychowdhury, R., *et al.* (2016). Perturb-Seq: Dissecting Molecular
1225 Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. Cell *167*, 1853-
1226 1866 e1817.

Duerr, R.H., Taylor, K.D., Brant, S.R., Rioux, J.D., Silverberg, M.S., Daly, M.J., Steinhart, A.H., Abraham, C., Regueiro, M., Griffiths, A., *et al.* (2006). A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. Science *314*, 1461-1463.

Durruthy-Durruthy, R., Gottlieb, A., Hartman, B.H., Waldhaus, J., Laske, R.D., Altman, R., and Heller, S. (2014). Reconstruction of the Mouse Otocyst and Early Neuroblast Lineage at Single-Cell Resolution. Cell *157*, 964-978.

Eberwine, J., Yeh, H., Miyashiro, K., Cao, Y., Nair, S., Finnell, R., Zettel, M., and Coleman, P. (1992). Analysis of gene expression in single live neurons. PNAS *89*, 3010-3014.

Eldar, A., and Elowitz, M.B. (2010). Functional roles for noise in genetic circuits. Nature *467*, 167-173.

Eulenberg, P., Kohler, N., Blasi, T., Filby, A., Carpenter, A.E., Rees, P., Theis, F.J., and Wolf, F.A. (2017). Reconstructing cell cycle and disease progression using deep learning. Nat Commun *8*, 463.

Fan, H.C., Fu, G.K., and Fodor, S.P. (2015). Combinatorial labeling of single cells for gene expression cytometry. Science *347*, 1258367.

Farlik, M., Sheffield, N.C., Nuzzo, A., Datlinger, P., Schonegger, A., Klughammer, J., and Bock, C. (2015). Single-cell DNA methylome sequencing and bioinformatic inference of epigenomic cell-state dynamics. Cell Reports *10*, 1386-1397.

Ferrell, J.E., Jr. (2012). Bistability, bifurcations, and Waddington's epigenetic landscape. Current Biology *22*, R458-466.

Foldy, C., Darmanis, S., Aoto, J., Malenka, R.C., Quake, S.R., and Sudhof, T.C. (2016). Single-cell RNAseq reveals cell adhesion molecule profiles in electrophysiologically defined neurons. PNAS *113*, E5222-E5231.

Frei, A.P., Bava, F.A., Zunder, E.R., Hsieh, E.W., Chen, S.Y., Nolan, G.P., and Gherardini, P.F. (2016). Highly multiplexed simultaneous detection of RNAs and proteins in single cells. Nature Methods *13*, 269-275.

Fulwyler, M.J. (1965). Electronic separation of biological cells by volume. Science *150*, 910-911.

Gagliani, N., Amezcua Vesely, M.C., Iseppon, A., Brockmann, L., Xu, H., Palm, N.W., de Zoete, M.R., Licona-Limon, P., Paiva, R.S., Ching, T., *et al.* (2015). Th17 cells transdifferentiate into regulatory T cells during resolution of inflammation. Nature *523*, 221-225.

1259 Gao, R., Davis, A., McDonald, T.O., Sei, E., Shi, X., Wang, Y., Tsai, P.C., Casasent, A.,
1260 Waters, J., Zhang, H., *et al.* (2016). Punctuated copy number evolution and clonal stasis in
1261 triple-negative breast cancer. Nature Genetics 48, 1119-1130
1262 Gaublomme, J.T., Yosef, N., Lee, Y., Gertner, R.S., Yang, L.V., Wu, C., Pandolfi, P.P., Mak,
1263 T., Satija, R., Shalek, A.K., *et al.* (2015). Single-Cell Genomics Unveils Critical Regulators of
1264 Th17 Cell Pathogenicity. Cell *163*, 1400-1412.
1265 Gawad, C., Koh, W., and Quake, S.R. (2014). Dissecting the clonal origins of childhood
1266 acute lymphoblastic leukemia by single-cell genomics. PNAS *111*, 17947-17952.
1267 Genshaft, A.S., Li, S., Gallant, C.J., Darmanis, S., Prakadan, S.M., Ziegler, C.G., Lundberg,
1268 M., Fredriksson, S., Hong, J., Regev, A., *et al.* (2016). Multiplexed, targeted profiling of
1269 single-cell proteomes and transcriptomes in a single reaction. Genome Biology *17*, 188.
1270 Giesen, C., Wang, H.A., Schapiro, D., Zivanovic, N., Jacobs, A., Hattendorf, B., Schuffler,
1271 P.J., Grolimund, D., Buhmann, J.M., Brandt, S., *et al.* (2014). Highly multiplexed imaging of
1272 tumor tissues with subcellular resolution by mass cytometry. Nature Methods *11*, 417-422.
1273 Gokce, O., Stanley, G.M., Treutlein, B., Neff, N.F., Camp, J.G., Malenka, R.C., Rothwell,
1274 P.E., Fuccillo, M.V., Sudhof, T.C., and Quake, S.R. (2016). Cellular Taxonomy of the Mouse
1275 Striatum as Revealed by Single-Cell RNA-Seq. Cell Reports *16*, 1126-1137.
1276 Grun, D., Lyubimova, A., Kester, L., Wiebrands, K., Basak, O., Sasaki, N., Clevers, H., and
1277 van Oudenaarden, A. (2015). Single-cell messenger RNA sequencing reveals rare intestinal
1278 cell types. Nature *525*, 251-255.
1279 Grun, D., Muraro, M.J., Boisset, J.C., Wiebrands, K., Lyubimova, A., Dharmadhikari, G.,
1280 van den Born, M., van Es, J., Jansen, E., Clevers, H., *et al.* (2016). De Novo Prediction of
1281 Stem Cell Identity using Single-Cell Transcriptome Data. Cell Stem Cell *19*, 266-277.
1282 Grun, D., and van Oudenaarden, A. (2015). Design and Analysis of Single-Cell Sequencing
1283 Experiments. Cell *163*, 799-810.
1284 Guo, H., Zhu, P., Wu, X., Li, X., Wen, L., and Tang, F. (2013). Single-cell methylome
1285 landscapes of mouse embryonic stem cells and early embryos analyzed using reduced
1286 representation bisulfite sequencing. Genome Research *23*, 2126-2135.
1287 Gut, G., Tadmor, M.D., Pe'er, D., Pelkmans, L., and Liberali, P. (2015). Trajectories of cell-
1288 cycle progression from fixed cell populations. Nature methods *12*, 951-954.

Habib, N., Li, Y., Heidenreich, M., Swiech, L., Avraham-Davidi, I., Trombetta, J.J., Hession, C., Zhang, F., and Regev, A. (2016a). Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons. Science *353*, 925-928.

Habib, N., Li, Y., Heidenreich, M., Swiech, L., Trombetta, J.J., Zhang, F., and Regev, A. (2016b). Div-Seq: A single nucleus RNA-Seq method reveals dynamics of rare adult newborn neurons in the CNS. bioRxiv. https://doi.org/10.1101/045989

Haghverdi, L., Buettner, F., and Theis, F.J. (2015). Diffusion maps for high-dimensional single-cell analysis of differentiation data. Bioinformatics *31*, 2989-2998.

Haghverdi, L., Buttner, M., Wolf, F.A., Buettner, F., and Theis, F.J. (2016). Diffusion pseudotime robustly reconstructs lineage branching. Nature Methods. 13, 845-845

Hama, H., Kurokawa, H., Kawano, H., Ando, R., Shimogori, T., Noda, H., Fukami, K., Sakaue-Sawano, A., and Miyawaki, A. (2011). Scale: a chemical approach for fluorescence imaging and reconstruction of transparent mouse brain. Nature Neuroscience *14*, 1481-1488.

Harris, H. (2000). The Birth of the Cell Yale University Press

Hashimshony, T., Wagner, F., Sher, N., and Yanai, I. (2012). CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. Cell Reports *2*, 666-673.

Hawrylycz, M.J., Lein, E.S., Guillozet-Bongaarts, A.L., Shen, E.H., Ng, L., Miller, J.A., van de Lagemaat, L.N., Smith, K.A., Ebbert, A., Riley, Z.L., *et al.* (2012). An anatomically comprehensive atlas of the adult human brain transcriptome. Nature *489*, 391-399.

Heimberg, G., Bhatnagar, R., El-Samad, H., and Thomson, M. (2016). Low Dimensionality in Gene Expression Data Enables the Accurate Extraction of Transcriptional Programs from Shallow Sequencing. Cell Syst *2*, 239-250.

Hooke, R. (1665). Micrographia. London: Royal Society

Horowitz, A., Strauss-Albee, D.M., Leipold, M., Kubo, J., Nemat-Gorgani, N., Dogan, O.C., Dekker, C.L., Mackey, S., Maecker, H., Swan, G.E., *et al.* (2013). Genetic and environmental determinants of human NK cell diversity revealed by mass cytometry. Science Translational Medicine *5*, 208ra145.

Huang, S. (2012). The molecular and mathematical basis of Waddington's epigenetic landscape: A framework for post-Darwinian biology? BioEssays *34*, 149-157.

Huang, S. (2013). Hybrid T-helper cells: stabilizing the moderate center in a polarized system. PLoS Biology *11*, e1001632.

Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., Lonnerberg, P., and Linnarsson, S. (2014). Quantitative single-cell RNA-seq with unique molecular identifiers. Nature Methods *11*, 163-166.

Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Elefant, N., Paul, F., Zaretsky, I., Mildner, A., Cohen, N., Jung, S., Tanay, A., *et al.* (2014). Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. Science *343*, 776-779.

Jaitin, D.A., Weiner, A., Yofe, I., Lara-Astiaso, D., Keren-Shaul, H., David, E., Salame, T.M., Tanay, A., van Oudenaarden, A., and Amit, I. (2016). Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. Cell *167*, 1883-1896 e1815.

Jan, M., Snyder, T.M., Corces-Zimmerman, M.R., Vyas, P., Weissman, I.L., Quake, S.R., and Majeti, R. (2012). Clonal Evolution of Preleukemic Hematopoietic Stem Cells Precedes Human Acute Myeloid Leukemia. Science translational medicine *4*, 149ra118

Jiang, X., Shen, S., Cadwell, C.R., Berens, P., Sinz, F., Ecker, A.S., Patel, S., and Tolias, A.S. (2015). Principles of connectivity among morphologically defined cell types in adult neocortex. Science *350*, aac9462.

Jojic, V., Shay, T., Sylvia, K., Zuk, O., Sun, X., Kang, J., Regev, A., Koller, D., Immunological Genome Project, C., Best, A.J., *et al.* (2013). Identification of transcriptional regulators in the mouse immune system. Nature Immunology *14*, 633-643.

Kafri, R., Levy, J., Ginzberg, M.B., Oh, S., Lahav, G., and Kirschner, M.W. (2013). Dynamics extracted from fixed cells reveal feedback linking cell growth to cell cycle. Nature *494*, 480-483.

Ke, R., Mignardi, M., Pacureanu, A., Svedlund, J., Botling, J., Wahlby, C., and Nilsson, M. (2013). In situ sequencing for RNA analysis in preserved tissue and cells. Nature Methods *10*, 857-860.

Kharchenko, P.V., Silberstein, L., and Scadden, D.T. (2014). Bayesian approach to single-cell differential expression analysis. Nature Methods *11*, 740-742.

Kim, C.C., and Lanier, L.L. (2013). Beyond the transcriptome: completion of act one of the Immunological Genome Project. Current Opinion in Immunology *25*, 593-597.

Kim, J., and Eberwine, J. (2010). RNA: state memory and mediator of cellular phenotype. Trends in Cell Biology *20*, 311-318.

Kim, J.K., Kolodziejczyk, A.A., Ilicic, T., Teichmann, S.A., and Marioni, J.C. (2015). Characterizing noise structure in single-cell RNA-seq distinguishes genuine from technical stochastic allelic expression. Nat Commun *6*, 8687.

Kiselev, V.Y., Kirschner, K., Schaub, M.T., Andrews, T., Yiu, A., Chandra, T., Natarajan, K.N., Reik, W., Barahona, M., Green, A.R., *et al.* (2016). SC3: consensus clustering of single-cell RNA-Seq data. Nature Methods 14, 483-486

Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell *161*, 1187-1201.

Kohler, G., and Milstein, C. (1975). Continuous cultures of fused cells secreting antibody of predefined specificity. Nature *256*, 495-497.

Kolodziejczyk, A.A., Kim, J.K., Tsang, J.C., Ilicic, T., Henriksson, J., Natarajan, K.N., Tuck, A.C., Gao, X., Buhler, M., Liu, P., *et al.* (2015). Single Cell RNA-Sequencing of Pluripotent States Unlocks Modular Transcriptional Variation. Cell Stem Cell *17*, 471-485.

Kowalczyk, M.S., Tirosh, I., Heckl, D., Rao, T.N., Dixit, A., Haas, B.J., Schneider, R.K., Wagers, A.J., Ebert, B.L., and Regev, A. (2015). Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. Genome Research *25*, 1860-1872.

Kretzschmar, K., and Watt, F.M. (2012). Lineage tracing. Cell *148*, 33-45.

Krishnaswamy, S., Spitzer, M.H., Mingueneau, M., Bendall, S.C., Litvin, O., Stone, E., Pe'er, D., and Nolan, G.P. (2014). Conditional density-based analysis of T cell signaling in single-cell data. Science *346*, 1250689.

Lake, B.B., Ai, R., Kaeser, G.E., Salathia, N.S., Yung, Y.C., Liu, R., Wildberg, A., Gao, D., Fung, H.L., Chen, S., *et al.* (2016). Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. Science *352*, 1586-1590.

Lander, E.S. (1996). The new genomics: global views of biology. Science *274*, 536-539.

Langer-Safer, P.R., Levine, M., and Ward, D.C. (1982). Immunological method for mapping genes on Drosophila polytene chromosomes. PNAS *79*, 4381-4385.

Lee, J.H., Daugharthy, E.R., Scheiman, J., Kalhor, R., Yang, J.L., Ferrante, T.C., Terry, R., Jeanty, S.S., Li, C., Amamoto, R., *et al.* (2014). Highly multiplexed subcellular RNA sequencing in situ. Science *343*, 1360-1363.

Levine, J.H., Simonds, E.F., Bendall, S.C., Davis, K.L., Amir el, A.D., Tadmor, M.D., Litvin, O., Fienberg, H.G., Jager, A., Zunder, E.R., *et al.* (2015). Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. Cell *162*, 184-197.

Lönnberg, T., Svensson, V., James, K.R., Fernandez-Ruiz, D., Sebina, I., Montandon, R., Soon, M.S.F., Fogg, L.G., Stubbington, M.J.T., Otzen Bagger, F., *et al.* (2017). Temporal mixture modelling of single-cell RNA-seq data resolves a CD4+ T cell fate bifurcation. bioRxiv. https://doi.org/10.1101/074971

Lorthongpanich, C., Cheow, L.F., Balu, S., Quake, S.R., Knowles, B.B., Burkholder, W.F., Solter, D., and Messerschmidt, D.M. (2013). Single-Cell DNA-Methylation Analysis Reveals Epigenetic Chimerism in Preimplantation Embryos. Science *341*, 1110-1112.

Lovatt, D., Ruble, B.K., Lee, J., Dueck, H., Kim, T.K., Fisher, S., Francis, C., Spaethling, J.M., Wolf, J.A., Grady, M.S., *et al.* (2014). Transcriptome in vivo analysis (TIVA) of spatially defined single cells in live tissue. Nature Methods *11*, 190-196.

Lu, R., Neff, N.F., Quake, S.R., and Weissman, I.L. (2011). Tracking single hematopoietic stem cells in vivo using high-throughput sequencing in conjunction with viral genetic barcoding. Nature Biotechnology *29*, 928-933.

Lu, Y., Mahony, S., Benos, P.V., Rosenfeld, R., Simon, I., Breeden, L.L., and Bar-Joseph, Z. (2007). Combined analysis reveals a core set of cycling genes. Genome Biology *8*, R146.

Lubeck, E., Coskun, A.F., Zhiyentayev, T., Ahmad, M., and Cai, L. (2014). Single-cell in situ RNA profiling by sequential hybridization. Nature Methods *11*, 360-361.

Macaulay, I.C., Haerty, W., Kumar, P., Li, Y.I., Hu, T.X., Teng, M.J., Goolam, M., Saurat, N., Coupland, P., Shirley, L.M., *et al.* (2015). G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. Nature Methods *12*, 519-522.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., *et al.* (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell *161*, 1202-1214.

Mahata, B., Zhang, X., Kolodziejczyk, A.A., Proserpio, V., Haim-Vilmovsky, L., Taylor, A.E., Hebenstreit, D., Dingler, F.A., Moignard, V., Gottgens, B., *et al.* (2014). Single-cell RNA sequencing reveals T helper cells synthesizing steroids de novo to contribute to immune homeostasis. Cell Reports *7*, 1130-1142.

Marco, E., Karp, R.L., Guo, G., Robson, P., Hart, A.H., Trippa, L., and Yuan, G.C. (2014). Bifurcation analysis of single-cell gene expression data reveals epigenetic landscape. PNAS *111*, E5643-5650.

Marcus, J.S., Anderson, W.F., and Quake, S.R. (2006). Microfluidic single-cell mRNA isolation and analysis. Analytical Chemistry *78*, 3084-3089.

Markram, H., Muller, E., Ramaswamy, S., Reimann, M.W., Abdellah, M., Sanchez, C.A., Ailamaki, A., Alonso-Nanclares, L., Antille, N., Arsever, S.*, et al.* (2015). Reconstruction and Simulation of Neocortical Microcircuitry. Cell *163*, 456-492.

Martincorena, I., Roshan, A., Gerstung, M., Ellis, P., Van Loo, P., McLaren, S., Wedge, D.C., Fullam, A., Alexandrov, L.B., Tubio, J.M.*, et al.* (2015). High burden and pervasive positive selection of somatic mutations in normal human skin. Science *348*, 880-886.

May, R.M. (1976). Simple mathematical models with very complicated dynamics. Nature *261*, 459-467.

Mazzarello, P. (1999). A unifying concept: the history of cell theory. Nat Cell Biol *1*, E13-15.

McKenna, A., Findlay, G.M., Gagnon, J.A., Horwitz, M.S., Schier, A.F., and Shendure, J. (2016). Whole organism lineage tracing by combinatorial and cumulative genome editing. Science. 353, aaf7907

Miller, J.A., Ding, S.L., Sunkin, S.M., Smith, K.A., Ng, L., Szafer, A., Ebbert, A., Riley, Z.L., Royall, J.J., Aiona, K.*, et al.* (2014). Transcriptional landscape of the prenatal human brain. Nature *508*, 199-206.

Miyashiro, K., Dichter, M., and Eberwine, J. (1994). On the nature and differential distribution of mRNAs in hippocampal neurites: implications for neuronal functioning. PNAS *91*, 10800-10804.

Moffitt, J.R., Hao, J., Bambah-Mukku, D., Lu, T., Dulac, C., and Zhuang, X. (2016a). High-performance multiplexed fluorescence in situ hybridization in culture and tissue with matrix imprinting and clearing. PNAS *113*, 14456-14461.

Moffitt, J.R., Hao, J., Wang, G., Chen, K.H., Babcock, H.P., and Zhuang, X. (2016b). High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. PNAS *113*, 11046-11051.

Moignard, V., Woodhouse, S., Haghverdi, L., Lilly, A.J., Tanaka, Y., Wilkinson, A.C., Buettner, F., Macaulay, I.C., Jawaid, W., Diamanti, E.*, et al.* (2015). Decoding the regulatory

network of early blood development from single-cell gene expression measurements. Nature Biotechnology *33*, 269-276.

Mooijman, D., Dey, S.S., Boisset, J.C., Crosetto, N., and van Oudenaarden, A. (2016). Single-cell 5hmC sequencing reveals chromosome-wide cell-to-cell variability and enables lineage reconstruction. Nature Biotechnology *34*, 852-856.

Moris, N., Pina, C., and Arias, A.M. (2016). Transition states and cell fate decisions in epigenetic landscapes. Nature Reviews Genetics *17*, 693-703.

Morton, C.L., and Houghton, P.J. (2007). Establishment of human tumor xenografts in immunodeficient mice. Nature Protocols *2*, 247-250.

Murray, J.M., Davies, K.E., Harper, P.S., Meredith, L., Mueller, C.R., and Williamson, R. (1982). Linkage relationship of a cloned DNA sequence on the short arm of the X chromosome to Duchenne muscular dystrophy. Nature *300*, 69-71.

Nagano, T., Lubling, Y., Stevens, T.J., Schoenfelder, S., Yaffe, E., Dean, W., Laue, E.D., Tanay, A., and Fraser, P. (2013). Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. Nature *502*, 59-64.

Nagel, M.C. (1981). Sir William Henry Perkin, pioneer in color. Journal of Chemical Education *58*, 305.

Naik, S.H., Perie, L., Swart, E., Gerlach, C., van Rooij, N., de Boer, R.J., and Schumacher, T.N. (2013). Diverse and heritable lineage imprinting of early haematopoietic progenitors. Nature *496*, 229-232.

Nestorowa, S., Hamey, F.K., Pijuan Sala, B., Diamanti, E., Shepherd, M., Laurenti, E., Wilson, N.K., Kent, D.G., and Gottgens, B. (2016). A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. Blood *128*, e20-31.

Nichterwitz, S., Chen, G., Aguila Benitez, J., Yilmaz, M., Storvall, H., Cao, M., Sandberg, R., Deng, Q., and Hedlund, E. (2016). Laser capture microscopy coupled with Smart-seq2 for precise spatial transcriptomic profiling. Nat Commun *7*, 12139.

O'Brien, C.A., Pollett, A., Gallinger, S., and Dick, J.E. (2007). A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. Nature *445*, 106-110.

Olsson, A., Venkatasubramanian, M., Chaudhri, V.K., Aronow, B.J., Salomonis, N., Singh, H., and Grimes, H.L. (2016). Single-cell analysis of mixed-lineage states leading to a binary cell fate choice. Nature. 537, 698-702

1476 Parolini, G. (2015). The emergence of modern statistics in agricultural science: analysis of
1477 variance, experimental design and the reshaping of research at Rothamsted Experimental
1478 Station, 1919-1933. Journal of the History of Biology *48*, 301-335.
1479 Patel, A.P., Tirosh, I., Trombetta, J.J., Shalek, A.K., Gillespie, S.M., Wakimoto, H., Cahill,
1480 D.P., Nahed, B.V., Curry, W.T., Martuza, R.L., *et al.* (2014). Single-cell RNA-seq highlights
1481 intratumoral heterogeneity in primary glioblastoma. Science *344*, 1396-1401.
1482 Paul, F., Arkin, Y., Giladi, A., Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Winter, D.,
1483 Lara-Astiaso, D., Gury, M., Weiner, A., *et al.* (2015). Transcriptional Heterogeneity and
1484 Lineage Commitment in Myeloid Progenitors. Cell *163*, 1663-1677.
1485 Perie, L., and Duffy, K.R. (2016). Retracing the in vivo haematopoietic tree using single-cell
1486 methods. FEBS Letters *590*, 4068-4083.
1487 Picelli, S., Bjorklund, A.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R.
1488 (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nature
1489 Methods *10*, 1096-1098.
1490 Pollen, A.A., Nowakowski, T.J., Shuga, J., Wang, X., Leyrat, A.A., Lui, J.H., Li, N.,
1491 Szpankowski, L., Fowler, B., Chen, P., *et al.* (2014). Low-coverage single-cell mRNA
1492 sequencing reveals cellular heterogeneity and activated signaling pathways in developing
1493 cerebral cortex. Nature Biotechnology *32*, 1053-1058.
1494 Proserpio, V., Piccolo, A., Haim-Vilmovsky, L., Kar, G., Lonnberg, T., Svensson, V.,
1495 Pramanik, J., Natarajan, K.N., Zhai, W., Zhang, X., *et al.* (2016). Single-cell analysis of CD4+
1496 T-cell differentiation reveals three major cell states and progressive acceleration of
1497 proliferation. Genome Biology *17*, 103.
1498 Ramani, V., Deng, X., Gunderson, K.L., Steemers, F.J., Disteche, C.M., Noble, W.S., Duan,
1499 Z., and Shendure, J. (2016). Massively multiplex single-cell Hi-C. Nature Methods 14, 263-
1500 266
1501 Ramilowski, J.A., Goldberg, T., Harshbarger, J., Kloppmann, E., Lizio, M., Satagopam, V.P.,
1502 Itoh, M., Kawaji, H., Carninci, P., Rost, B., *et al.* (2015). A draft network of ligand-receptor-
1503 mediated multicellular signalling in human. Nat Commun *6*, 7866.
1504 Ramón y Cajal, S. (1995). Histology of the Nervous System of Man and Vertebrates. Oxford
1505 University Press
1506 Ramskold, D., Luo, S., Wang, Y.C., Li, R., Deng, Q., Faridani, O.R., Daniels, G.A.,
1507 Khrebtukova, I., Loring, J.F., Laurent, L.C., *et al.* (2012). Full-length mRNA-Seq from

single-cell levels of RNA and individual circulating tumor cells. Nature Biotechnology *30*, 777-782.

Rebhahn, J.A., Deng, N., Sharma, G., Livingstone, A.M., Huang, S., and Mosmann, T.R. (2014). An animated landscape representation of CD4+ T-cell differentiation, variability, and plasticity: insights into the behavior of populations versus cells. European Journal of Immunology *44*, 2216-2229.

Reizel, Y., Itzkovitz, S., Adar, R., Elbaz, J., Jinich, A., Chapal-Ilani, N., Maruvka, Y.E., Nevo, N., Marx, Z., Horovitz, I., *et al.* (2012). Cell lineage analysis of the mammalian female germline. PLoS Genetics *8*, e1002477.

Richmond, A., and Su, Y. (2008). Mouse xenograft models vs GEM models for human cancer therapeutics. Disease Models & Mechanisms *1*, 78-82.

Rosvall, M., and Bergstrom, C.T. (2008). Maps of random walks on complex networks reveal community structure. PNAS *105*, 1118-1123.

Rotem, A., Ram, O., Shoresh, N., Sperling, R.A., Goren, A., Weitz, D.A., and Bernstein, B.E. (2015a). Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. Nature Biotechnology *33*, 1165-1172.

Rotem, A., Ram, O., Shoresh, N., Sperling, R.A., Schnall-Levin, M., Zhang, H., Basu, A., Bernstein, B.E., and Weitz, D.A. (2015b). High-Throughput Single-Cell Labeling (Hi-SCL) for RNA-Seq Using Drop-Based Microfluidics. PloS One *10*, e0116328.

Sachs, K., Perez, O., Pe'er, D., Lauffenburger, D.A., and Nolan, G.P. (2005). Causal protein-signaling networks derived from multiparameter single-cell data. Science *308*, 523-529.

Sanes, J.R., and Masland, R.H. (2015). The types of retinal ganglion cells: current status and implications for neuronal classification. Annual Review of Neuroscience *38*, 221-246.

Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nature Biotechnology *33*, 495-502.

Scialdone, A., Tanaka, Y., Jawaid, W., Moignard, V., Wilson, N.K., Macaulay, I.C., Marioni, J.C., and Gottgens, B. (2016). Resolving early mesoderm diversification through single-cell expression profiling. Nature *535*, 289-293.

Setty, M., Tadmor, M.D., Reich-Zeliger, S., Angel, O., Salame, T.M., Kathail, P., Choi, K., Bendall, S., Friedman, N., and Pe'er, D. (2016). Wishbone identifies bifurcating developmental trajectories from single-cell data. Nature Biotechnology *34*, 637-645.

Shah, S., Lubeck, E., Zhou, W., and Cai, L. (2016). In Situ Transcription Profiling of Single Cells Reveals Spatial Organization of Cells in the Mouse Hippocampus. Neuron *92*, 342-357.

Shalek, A.K., Satija, R., Adiconis, X., Gertner, R.S., Gaublomme, J.T., Raychowdhury, R., Schwartz, S., Yosef, N., Malboeuf, C., Lu, D., *et al.* (2013). Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. Nature *498*, 236-240.

Shalek, A.K., Satija, R., Shuga, J., Trombetta, J.J., Gennert, D., Lu, D., Chen, P., Gertner, R.S., Gaublomme, J.T., Yosef, N., *et al.* (2014). Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. Nature *510*, 363-369.

Shapiro, E. (2010). The Human Cell Lineage Flagship Initiative. http://www.lineage-flagship.eu/ (Accessed November 28, 2017)

Shapiro, E., Biezuner, T., and Linnarsson, S. (2013). Single-cell sequencing-based technologies will revolutionize whole-organism science. Nature Reviews Genetics *14*, 618-630.

Shekhar, K., Lapan, S.W., Whitney, I.E., Tran, N.M., Macosko, E.Z., Kowalczyk, M., Adiconis, X., Levin, J.Z., Nemesh, J., Goldman, M., *et al.* (2016). Comprehensive Classification of Retinal Bipolar Neurons by Single-Cell Transcriptomics. Cell *166*, 1308-1323 e1330.

Shin, J., Berg, D.A., Zhu, Y., Shin, J.Y., Song, J., Bonaguidi, M.A., Enikolopov, G., Nauen, D.W., Christian, K.M., Ming, G.L., *et al.* (2015). Single-Cell RNA-Seq with Waterfall Reveals Molecular Cascades underlying Adult Neurogenesis. Cell Stem Cell *17*, 360-372.

Shlush, L.I., Chapal-Ilani, N., Adar, R., Pery, N., Maruvka, Y., Spiro, A., Shouval, R., Rowe, J.M., Tzukerman, M., Bercovich, D., *et al.* (2012). Cell lineage analysis of acute leukemia relapse uncovers the role of replication-rate heterogeneity and microsatellite instability. Blood *120*, 603-612.

Singer, M., Wang, C., Cong, L., Marjanovic, N.D., Kowalczyk, M.S., Zhang, H., Nyman, J., Sakuishi, K., Kurtulus, S., Gennert, D., *et al.* (2016). A Distinct Gene Module for Dysfunction Uncoupled from Activation in Tumor-Infiltrating T Cells. Cell *166*, 1500-1511 e1509.

Smallwood, S.A., Lee, H.J., Angermueller, C., Krueger, F., Saadeh, H., Peat, J., Andrews, S.R., Stegle, O., Reik, W., and Kelsey, G. (2014). Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. Nature Methods *11*, 817-820.

1571 Stahl, P.L., Salmen, F., Vickovic, S., Lundmark, A., Navarro, J.F., Magnusson, J.,
1572 Giacomello, S., Asp, M., Westholm, J.O., Huss, M., *et al.* (2016). Visualization and analysis
1573 of gene expression in tissue sections by spatial transcriptomics. Science *353*, 78-82.
1574 Stahnisch, F.W. (2015). Joseph von Gerlach (1820–1896). Journal of Neurology *262*, 1397-
1575 1399.
1576 Stegle, O., Teichmann, S.A., and Marioni, J.C. (2015). Computational and analytical
1577 challenges in single-cell transcriptomics. Nature Reviews Genetics *16*, 133-145.
1578 Stewart-Ornstein, J., Weissman, J.S., and El-Samad, H. (2012). Cellular noise regulons
1579 underlie fluctuations in Saccharomyces cerevisiae. Molecular Cell *45*, 483-493.
1580 Stubbington, M.J., Lonnberg, T., Proserpio, V., Clare, S., Speak, A.O., Dougan, G., and
1581 Teichmann, S.A. (2016). T cell fate and clonality inference from single-cell transcriptomes.
1582 Nature Methods *13*, 329-332.
1583 Sul, J.Y., Kim, T.K., Lee, J.H., and Eberwine, J. (2012). Perspectives on cell reprogramming
1584 with RNA. Trends in Biotechnology *30*, 243-249.
1585 Susaki, E.A., Tainaka, K., Perrin, D., Kishino, F., Tawara, T., Watanabe, T.M., Yokoyama,
1586 C., Onoe, H., Eguchi, M., Yamaguchi, S., *et al.* (2014). Whole-Brain Imaging with Single-
1587 Cell Resolution Using Chemical Cocktails and Computational Analysis. Cell *157*, 726-739.
1588 Svensson, V., Natarajan, K.N., Ly, L.-H., Miragaia, R.J., Labalette, C., Macaulay, I.C., Cvejic,
1589 A., and Teichmann, S.A. (2016). Power Analysis of Single Cell RNA - Sequencing
1590 Experiments. Nature Methods 14, 381-387
1591 Tanay, A., and Regev, A. (2017). Scaling single-cell genomics from phenomenology to
1592 mechanism. Nature *541*, 331-338.
1593 Tasic, B., Menon, V., Nguyen, T.N., Kim, T.K., Jarsky, T., Yao, Z., Levi, B., Gray, L.T.,
1594 Sorensen, S.A., Dolbeare, T., *et al.* (2016). Adult mouse cortical cell taxonomy revealed by
1595 single cell transcriptomics. Nature Neuroscience *19*, 335-346.
1596 Taylor, R.W., Barron, M.J., Borthwick, G.M., Gospel, A., Chinnery, P.F., Samuels, D.C.,
1597 Taylor, G.A., Plusa, S.M., Needham, S.J., Greaves, L.C., *et al.* (2003). Mitochondrial DNA
1598 mutations in human colonic crypt stem cells. Journal of Clinical Investigation *112*, 1351-
1599 1360.
1600 Tecott, L.H., Barchas, J.D., and Eberwine, J.H. (1988). In situ transcription: specific
1601 synthesis of complementary DNA in fixed tissue sections. Science *240*, 1661-1664.

Teixeira, V.H., Nadarajan, P., Graham, T.A., Pipinikas, C.P., Brown, J.M., Falzon, M., Nye, E., Poulsom, R., Lawrence, D., Wright, N.A., *et al.* (2013). Stochastic homeostasis in human airway epithelium is achieved by neutral competition of basal cell progenitors. eLife *2*, e00966.

Thom, R. (1989). Structural Stability and Morphogenesis Addison Wesley Publishing Company.

Thomsen, E.R., Mich, J.K., Yao, Z., Hodge, R.D., Doyle, A.M., Jang, S., Shehata, S.I., Nelson, A.M., Shapovalova, N.V., Levi, B.P., *et al.* (2016). Fixed single-cell transcriptomic characterization of human radial glial diversity. Nature Methods *13*, 87-93.

Thul, P.J., Akesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Ait Blal, H., Alm, T., Asplund, A., Bjork, L., Breckels, L.M., *et al.* (2017). A subcellular map of the human proteome. Science *DOI:* 10.1126/science.aal3321

Tirosh, I., Izar, B., Prakadan, S.M., Wadsworth, M.H., 2nd, Treacy, D., Trombetta, J.J., Rotem, A., Rodman, C., Lian, C., Murphy, G., *et al.* (2016a). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science *352*, 189-196.

Tirosh, I., Venteicher, A.S., Hebert, C., Escalante, L.E., Patel, A.P., Yizhak, K., Fisher, J.M., Rodman, C., Mount, C., Filbin, M., *et al.* (2016b). Single-cell RNA-seq supports a developmental hierarchy in IDH-mutant oligodendroglioma. Nature *539, 309-313.*

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nature Biotechnology *32*, 381-386.

Treutlein, B., Brownfield, D.G., Wu, A.R., Neff, N.F., Mantalas, G.L., Espinoza, F.H., Desai, T.J., Krasnow, M.A., and Quake, S.R. (2014). Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. Nature *509*, 371-375.

Treutlein, B., Lee, Q.Y., Camp, J.G., Mall, M., Koh, W., Shariati, S.A.M., Sim, S., Neff, N.F., Skotheim, J.M., Wernig, M., *et al.* (2016). Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq. Nature *534*, 391-+.

Tsang, J.C., Yu, Y., Burke, S., Buettner, F., Wang, C., Kolodziejczyk, A.A., Teichmann, S.A., Lu, L., and Liu, P. (2015). Single-cell transcriptomic reconstruction reveals cell cycle and multi-lineage differentiation defects in Bcl11a-deficient hematopoietic stem cells. Genome Biology *16*, 178.

Uhlen, M., Fagerberg, L., Hallstrom, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Sjostedt, E., Asplund, A., *et al.* (2015). Proteomics. Tissue-based map of the human proteome. Science *347*, 1260419.

Vallejos, C.A., Marioni, J.C., and Richardson, S. (2015). BASiCS: Bayesian Analysis of Single-Cell Sequencing Data. PLoS Computational Biology *11*, e1004333.

Van Gelder, R.N., von Zastrow, M.E., Yool, A., Dement, W.C., Barchas, J.D., and Eberwine, J.H. (1990). Amplified RNA synthesized from limited quantities of heterogeneous cDNA. PNAS *87*, 1663-1667.

Vickovic, S., Stahl, P.L., Salmen, F., Giatrellis, S., Westholm, J.O., Mollbrink, A., Navarro, J.F., Custodio, J., Bienko, M., Sutton, L.A., *et al.* (2016). Massive and parallel expression profiling using microarrayed single-cell sequencing. Nat Commun *7*, 13182.

Waddington, C.H. (1957). The Strategy of the Genes London: Allen & Unwin

Wagner, A., Regev, A., and Yosef, N. (2016). Uncovering the vectors of cellular identity with single-cell genomics. Nature Biotechnology 34, 1145-1160.

Wang, Y., Waters, J., Leung, M.L., Unruh, A., Roh, W., Shi, X.Q., Chen, K., Scheet, P., Vattathil, S., Liang, H., *et al.* (2014). Clonal evolution in breast cancer revealed by single nucleus genome sequencing. Nature *512*, 155-160

Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhutdinov, R., Zemel, R.S., and Bengio, Y. (2015). Show, attend and tell: Neural image caption generation with visual attention. arXiv: https://arxiv.org/abs/1502.03044

Yang, B., Treweek, J.B., Kulkarni, R.P., Deverman, B.E., Chen, C.K., Lubeck, E., Shah, S., Cai, L., and Gradinaru, V. (2014). Single-Cell Phenotyping within Transparent Intact Tissue through Whole-Body Clearing. Cell *158*, 945-958.

Yosef, N., and Regev, A. (2011). Impulse control: temporal dynamics in gene transcription. Cell *144*, 886-896.

Yosef, N., and Regev, A. (2016). Writ large: Genomic dissection of the effect of cellular environment on immune response. Science *354*, 64-68.

Yuan, J., and Sims, P.A. (2016). An Automated Microwell Platform for Large-Scale Single Cell RNA-Seq. Scientific Reports *6*, 33883.

Zeisel, A., Munoz-Manchado, A.B., Codeluppi, S., Lonnerberg, P., La Manno, G., Jureus, A., Marques, S., Munguba, H., He, L., Betsholtz, C., *et al.* (2015). Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science *347*, 1138-1142.

Zheng, Y., Zemel, R.S., Zhang, Y.-J., and Larochelle, H. (2015). A Neural Autoregressive Approach to Attention-based Recognition. International Journal of Computer Vision *113*, 67-79.

Zhong, J.F., Chen, Y., Marcus, J.S., Scherer, A., Quake, S.R., Taylor, C.R., and Weiner, L.P. (2008). A microfluidic processor for gene expression profiling of single human embryonic stem cells. Lab Chip *8*, 68-74.

Zhou, J.X., Aliyu, M.D., Aurell, E., and Huang, S. (2012). Quasi-potential landscape in complex multi-stable systems. Journal of the Royal Society: Interface. 9, 3539-3553.

Zhou, J.X., Samal, A., d'Hérouël, A.F., Price, N.D., and Huang, S. (2016). Relative stability of network states in Boolean network models of gene regulation in development. BioSystems *142-143*, 15-24.

Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Smets, M., Leonhardt, H., Hellmann, I., and Enard, W. (2016). Comparative analysis of single-cell RNA sequencing methods. Molecular Cell 65, 631-643

Figure 1

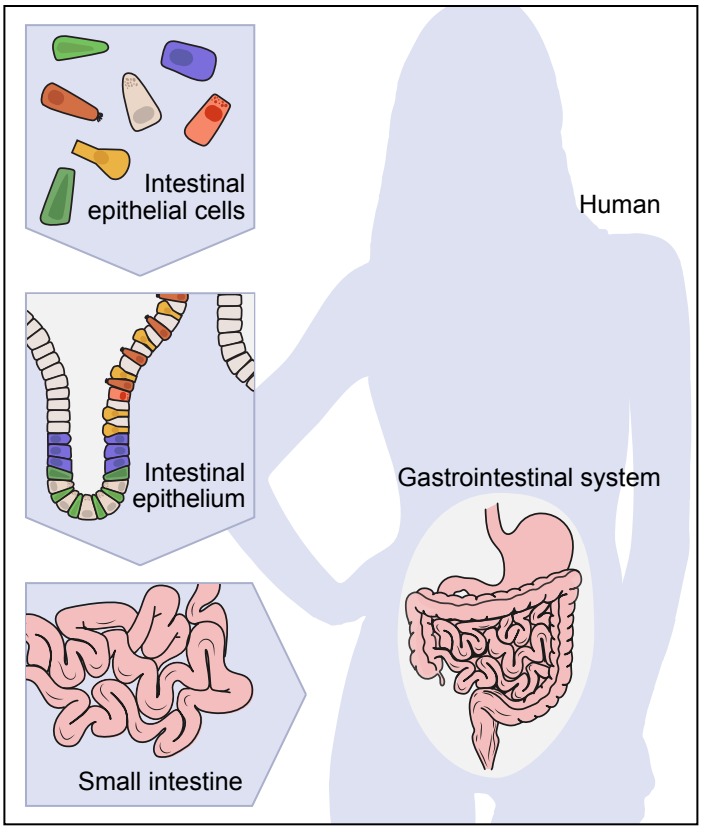

Intestinal epithelial cells
Intestinal epithelium
Small intestine
Human
Gastrointestinal system

Figure 2

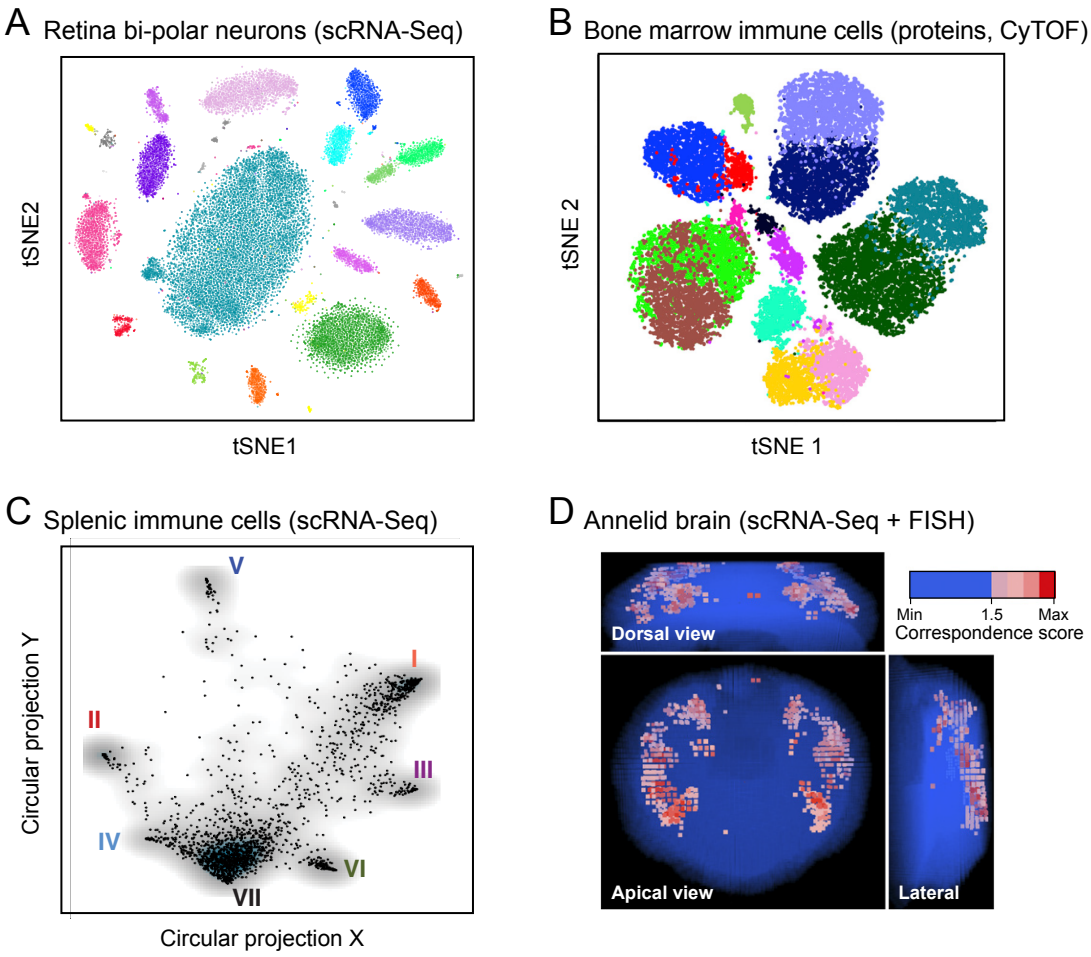

A Retina bi-polar neurons (scRNA-Seq)
tSNE2
tSNE1
B Bone marrow immune cells (proteins, CyTOF)
tSNE 2
tSNE 1
C Splenic immune cells (scRNA-Seq)
Circular projection Y
Circular projection X
V
I
II
III
IV
VI
VII
D Annelid brain (scRNA-Seq + FISH)
Dorsal view
Min
1.5
Max
Correspondence score
Apical view
Lateral

Figure 3

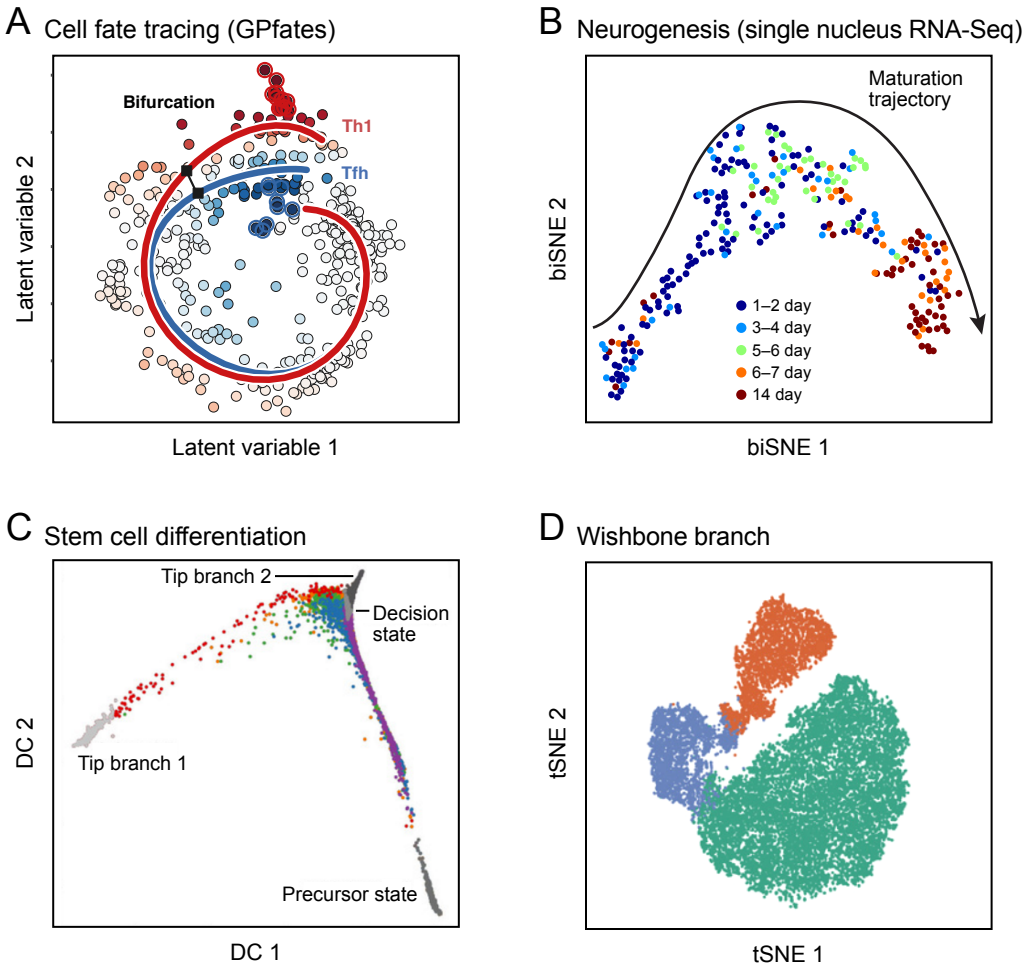
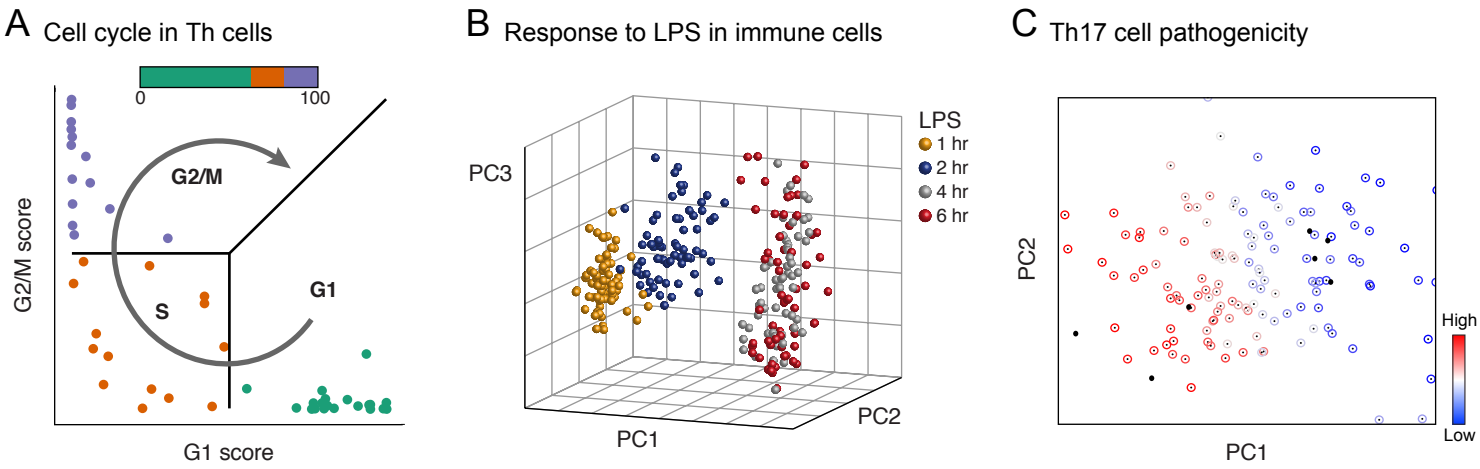

A Cell fate tracing (GPfates)
Bifurcation
Th1
Tfh
Latent variable 2
Latent variable 1
B Neurogenesis (single nucleus RNA-Seq)
Maturation trajectory
1–2 day
3–4 day
5–6 day
6–7 day
14 day
biSNE 2
biSNE 1
C Stem cell differentiation
Tip branch 2
Decision state
Tip branch 1
Precursor state
DC 2
DC 1
D Wishbone branch
tSNE 2
tSNE 1

Figure 4


A Cell cycle in Th cells
0
100
G2/M
S
G1
G2/M score
G1 score
B Response to LPS in immune cells
PC3
PC1
PC2
LPS
1 hr
2 hr
4 hr
6 hr
C Th17 cell pathogenicity
PC2
PC1
High
Low