

SPOTLIGHT

The evolving concept of cell identity in the single cell era

Samantha A. Morris^{1,2,3,*}

ABSTRACT

Fueled by recent advances in single cell biology, we are moving away from qualitative and undersampled assessments of cell identity, toward building quantitative, high-resolution cell atlases. However, it remains challenging to precisely define cell identity, leading to renewed debate surrounding this concept. Here, I present three pillars that I propose are central to the notion of cell identity: phenotype, lineage and state. I explore emerging technologies that are enabling the systematic and unbiased quantification of these properties, and outline how these efforts will enable the construction of a high-resolution, dynamic landscape of cell identity, potentially revealing its underlying molecular regulation to provide new opportunities for understanding and manipulating cell fate.

KEY WORDS: Cell atlases, Cell identity, Cell state, Single cell

Introduction

For centuries, biologists have sought to deconstruct the complexity of biological systems by breaking them down into their component parts – cells – and cataloging these individual units according to their identity. Establishing such a cellular taxonomy provides a universal scheme to standardize cell biology, yet the notion of cell identity, or cell type, remains poorly defined. Historically, cells have been classified by features such as morphology, location, ontogeny and interactions with other cell types. Over time, new assays were developed to measure the physiological function of cells, and these, accompanied by advances in molecular biology that enable the quantification of gene and protein expression, have allowed for more nuanced cell type classifications.

Fundamentally, though, no general method to accurately define cell identity currently exists. This represents a barrier to cataloging cell types in organisms where the full repertoire of cell identities remains unknown, such as in mouse (Han et al., 2018; Tabula Muris Consortium et al., 2018) and human (Regev et al., 2017). Thus, although several cell atlas construction endeavors are under way, these efforts have reignited debate around how cell identity can be effectively and accurately curated, revealing many differing viewpoints on this subject (Various authors, 2017; see also Xia and Yanai, 2019 in this issue). Here, I draw on new and established notions to synthesize a framework consisting of three pillars (Fig. 1) that I propose are central to the concept of cell identity: (1) phenotype (and function) – representing a central pillar for the definition of cell identity, this defines the broad range of physical, molecular and functional features that can be captured and analyzed to enable systematic and unbiased cell type categorization;

(2) lineage – to fully characterize cell identity, it is also valuable to understand the lineage relationships between different cell types and their genesis. Tracing the developmental origins of cell identity may allow a cellular taxonomy to be constructed, enabling similar cell types to be grouped together, potentially helping to characterize new cell species; (3) State – cell identity is stable; however, in response to diverse stimuli, the same cell type can exhibit a range of different phenotypes (states). Curating the cell states associated with a given cell type enables identity to be distinguished from state. Moreover, mapping the landscape of cell states lays the foundation for identifying when a cell travels out of normal physiological bounds into a pathological state. Together, a consideration of these three pillars can enable the construction of a high-resolution, dynamic cell identity landscape, potentially providing new opportunities for understanding and manipulating cell fate.

Phenotype and function: curating high-resolution snapshots of cellular features to characterize identity

Inferring cell identity from phenotype

The characterization of cell phenotype is central to defining cell identity and represents a longstanding focus of biologists. In the 1600s, aided by light microscopy, Robert Hooke initially described the cells that made up a sample of cork (Hooke, 1665). Two-hundred years later, the first histological stains using carmine, silver, and Hematoxylin and Eosin emerged, thus allowing relatively detailed cytological observations to be made (Pearse, 1984). It was around this time that Ramón y Cajal used Golgi's silver staining method to describe neurons, providing evidence that the nervous system isn't a continuum of fibers but is composed of individual units, neurons (Ramón y Cajal, 1888). Since these early discoveries, cell visualization using ever increasingly sophisticated microscopy and imaging techniques has remained central to cell type identification; probing key features such as cell shape, size, location and interactions with other cell types facilitates the classification of cells into discrete categories. With advances in molecular biology came the ability to stain cells for specific markers of identity (Coons et al., 1941). Eventually, distinct cell types could be labeled with fluorescent tags such as GFP (Chalfie et al., 1994), enabling the detailed investigation of cell phenotype within whole biological systems.

Imaging-based phenotypic assessment, along with other established techniques, such as flow cytometry, provides high resolution in terms of capturing information on an individual cell basis. Furthermore, these analyses can be deployed in intact cells and organisms, enabling cell function to be probed. However, the information yielded by these assays is comparatively low dimensional, i.e. relatively few phenotypic features are captured from many cells. In addition, the selection of these features tends to be driven by prior knowledge of the biological system under study, limiting and potentially biasing assessment of cell identity. In contrast, methods supporting genome-wide analysis of RNA and protein abundance support the collection of broader and more objective measurements. Indeed, increasing the number of molecular features used to define cell types has enabled more systematic and unbiased

¹Department of Developmental Biology, Washington University School of Medicine in St Louis, St Louis, MO 63110, USA. ²Department of Genetics, Washington University School of Medicine in St Louis, St Louis, MO 63110, USA. ³Center of Regenerative Medicine, Washington University School of Medicine in St Louis, St Louis, MO 63110, USA.

*Author for correspondence (s.morris@wustl.edu)

 S.A.M., 0000-0001-8561-4340

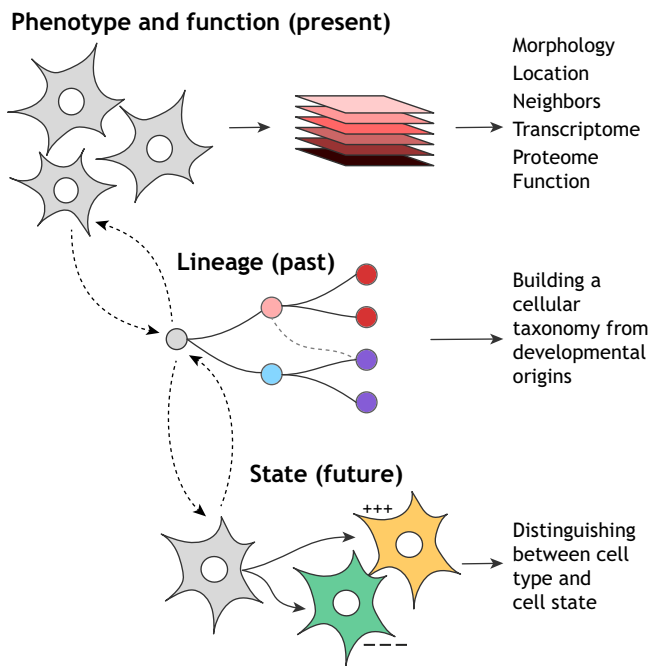


Fig. 1. The three pillars of cell identity. Phenotype and function: the main pillar concerns measurements of a cell in the present that enable systematic and unbiased cell type categorization. Lineage: this represents the past state of the cell and allows a cellular taxonomy to be constructed, enabling similar cell types to be grouped together to fully characterize cell identity. However, cell identity cannot accurately be defined by lineage alone. State represents the landscape of different future phenotypes within which a given cell type can exist in response to diverse stimuli. Together, these three pillars enable the construction of a high-resolution dynamic landscape of cell identity.

assessments of cell identity, based on gene expression alone (Cahan et al., 2014; Roost et al., 2015). Nevertheless, these approaches have relied on bulk analysis of mixed cell populations, blending signals from different sub-populations and altogether masking rare cell species, limiting the precision of cell type identification.

Recently developed single cell technologies have served to bridge the gap between detailed studies of individual cells and bulk studies of cell populations. These methods enable the capture of many thousands of features, without the requirement for experimental cell enrichment, thus generating a rigorous and unbiased picture of the range of cell phenotypes that exists within any given tissue. Of the current suite of technologies, which include genetic, epigenetic and proteomic profiling (Stuart and Satija, 2019), single cell RNA-sequencing (scRNA-seq), has seen rapid and wide adoption since its recent emergence (Tang et al., 2009). Although early iterations requiring cell separation in wells were relatively low-throughput and expensive, more recently developed microfluidic-based technologies have brought huge gains in cell capture rate (Klein et al., 2015; Macosko et al., 2015). Presently, pool-and-split cell labeling strategies are yielding even greater cell capture rates and further reductions in cost (Cao et al., 2017; Rosenberg et al., 2018).

scRNA-seq delivers relatively high-dimensional datasets, consisting of thousands of measurements across thousands of individual cells. Computational tools based on dimensionality reduction seek to reduce this complexity, clustering cells based on transcriptional similarity and enabling their visualization within two-dimensional space (Becht et al., 2018; Satija et al., 2015). It is important to note here that cluster-specific gene expression is used to infer cell type, representing an initial prediction of identity that must be orthogonally validated. One key limitation of scRNA-seq is that it

requires tissue disruption and cell destruction, resulting in loss of spatial information that is valuable for cell type identification. Maintaining this spatial information has been a recent focus of new single cell techniques (reviewed by Mayr et al., 2019 in this issue). For example, multiplexed *in situ* hybridization and sequencing technologies have enabled the measurement of gene expression at subcellular spatial resolution within intact tissues (Chen et al., 2015; Lee et al., 2014). Although these approaches initially required the upfront selection of genes for analysis, information on the expression of thousands of transcripts (Eng et al., 2019) and even genome-wide gene expression can be now captured (Rodrigues et al., 2019). Overall, these technologies are particularly promising, offering high-resolution visualization of many cellular features *in situ*, thereby allowing powerful predictions of cell identity to be made, based on phenotype.

Cell function: a ground truth of cell identity

Ultimately, cell identity is best defined by function. One powerful method for investigating cell function involves the physical elimination of cells, followed by observation of any physiological or behavioral impact on the organism. For example, laser ablation of a specific subset of *C. elegans* neurons revealed their role in locomotion (Chalfie et al., 1985). Alternatively, where a cell type is exclusively marked by expression of a specific gene, genetic ablation is possible, as illustrated by the targeted expression of a toxin gene to selectively kill pancreatic acinar cells (Palmiter et al., 1987). Although elegant in approach, ablation experiments are limited if cells cannot be physically accessed or are not marked by exclusive gene expression. For example, in the context of assessing cell function in humans, ablation experiments are clearly not feasible. Under these more limited circumstances, cells can be isolated and their function tested *in vitro* or in xenograft models. These approaches are being facilitated by single cell technologies that can identify new cell surface marker combinations at the proteomic level, for a given transcriptional state, enabling new cell species to be captured by flow cytometry and functionally assessed (Peterson et al., 2017; Stoekius et al., 2017). However, assigning cell function to a previously undescribed cell type would require an intractable array of assays to be deployed. Moreover, isolated cells often quickly lose their phenotype and function if culture conditions are not optimized, as illustrated by the dedifferentiation of *ex vivo* cultured hepatocytes (Elaut et al., 2006). Therefore, how do we begin to explore the function of novel cell types?

Where it is impractical to validate cell identity based on functional assays, will it be possible to predict cell function? Gene ontology serves as one commonly implemented method to predict cell function and behavior based on gene expression patterns (Ashburner et al., 2000). However, this approach often returns vague annotations, as gene expression does not directly translate to cell function. Considering that proteins are key effectors of cell function, measurement of protein abundance may be a more accurate predictor. Indeed, machine learning approaches have been deployed to infer cellular function based on tissue-specific protein function (Zitnik and Leskovec, 2017). To improve these predictions, quantifying protein localization in addition to protein abundance, e.g. via spatial proteomics, will undoubtedly prove beneficial. In this context, the recent construction of high-resolution cell atlases of protein expression, based on immunostaining of 12,003 proteins across 56 human cell lines, is extremely valuable (Thul et al., 2017). Also promising are machine learning algorithms that can be used to predict protein expression and localization in cells, based on light microscopy images alone (Christiansen et al., 2018). Indeed, the

broader application of machine learning to weave a more comprehensive picture of cell phenotype may serve well to infer cell function and classify cell identity (Smith et al., 2018). However, where possible, these predictive approaches must ultimately be supported by experimental evidence.

Lineage: new tracking technologies reveal cellular origins

So far, I have discussed some of the key tools that can be used to measure cellular phenotype and function, and how they serve to define cell identity. Consider a situation where the composite of these measurements reveals a previously undescribed novel cell type. It may be possible to assign some function to this new cell species using predictive tools, ideally confirmed by experimental assays. Even so, to fully understand cell identity is to place it within the context of all other cell types, in a cellular taxonomy. Constructing such a classification of cell identities from a snapshot of the adult organism in homeostasis is challenging, especially at present where datasets are sparse and we are still working to best integrate them in a meaningful way (Stuart and Satija, 2019). Instead, understanding the origins of a cell's identity, its developmental lineage, is a powerful and simple way to position a cell within a much more complex hierarchy. At a minimum, the new cell species can then be connected to its nearest relatives to provide further clues as to its role in the organism. Can, then, developmental origins alone provide sufficient information to define cell identity?

Lineage tracing, the identification of all progeny stemming from an individual cell, originates from Whitman's light microscopy studies of cell cleavage and eventual cell fate in invertebrate embryos (reviewed by Kretzschmar and Watt, 2012). Following on from these early studies, *C. elegans* has proven to be a particularly powerful model for lineage tracing, given its amenability to imaging, its relative small number of somatic cells and its invariant cell lineage. Indeed, a complete lineage tree for every cell in the *C. elegans* embryo has been constructed via non-invasive live imaging, documenting how cells decrease in potential and increase in specialization as development progresses (Sulston et al., 1983).

With sequencing technologies, new methods to diagram the relationships between lineal ancestry and prospective cell fates have emerged. These stemmed from DNA-based barcoding approaches, where cells are labeled with random heritable DNA sequences (Lu et al., 2011), later progressing to transcribed barcodes that allow clonal relationships and cell identity to be read in parallel (Yao et al., 2017). These early approaches enable clonal analysis, i.e. all descendants of an ancestrally marked founder cell can be identified via inheritance of their integrated barcodes. However, lineage relationships between clonal descendants cannot be mapped using these techniques. New single cell tracking approaches are emerging to fill this gap (reviewed by McKenna and Gagnon, 2019 in this issue). For example, sequential rounds of labeling with transcribed barcodes has enabled the construction of lineage trees (Bidy et al., 2018). In an alternative approach, CRISPR/Cas9-based genome editing has been leveraged to introduce mutable genetic labels into individual cells (Alemany et al., 2018; Raj et al., 2018; Spanjaard et al., 2018). Yet another method, transposon-based TracerSeq (Wagner et al., 2018), exploits the Tol2 transposase to randomly integrate unique heritable labels into individual cell genomes; asynchronous insertion over successive cell divisions then permits lineage tree reconstruction. When applied to zebrafish development, TracerSeq revealed evidence of convergent differentiation, where clonally distinct embryonic fields give rise to similar cell types (Wagner et al., 2018). In contrast, some clonally related cells diverged toward distant identities, supporting the case for divergent

differentiation. Thus, lineage analyses do not always produce an expected tree structure, i.e. cells from diverse embryonic origins can converge on a similar identity. This is not surprising given classic *C. elegans* lineage tracing studies showing that similar neuron types can be generated by distinct lineages (Sulston et al., 1983). More recently, the same phenomenon has been suggested in mouse development, where myocytes are produced by two convergent trajectories, and neurons by several trajectories (Cao et al., 2019). However, it is important to note that in this study of mouse development, trajectories were inferred via computational methods and are not based on ground truth data. Nonetheless, taken together, we must bear these examples of convergent differentiation in mind when considering the utility of lineage alone in defining cell identity.

Another limitation of relying on lineage to facilitate cell type identification is its deployment in the context of human development. How, in the absence of ground truth data that can be used to map lineage relationships, can we infer a meaningful and accurate cell developmental hierarchy? Representing a relatively simple experimental strategy, retrospective lineage tracing exploits naturally occurring genetic variation to trace clonally related cells (Ludwig et al., 2019), but this is limited in scale and cannot produce detailed lineage trees. As an alternative, computational approaches enable temporal reconstruction of scRNA-seq data (Saelens et al., 2019; Tritschler et al., 2019 in this issue). However, the resulting trajectories are inferred, relying on sufficient capture and sampling of intermediate cell states. This can be problematic, particularly for tracing the origins of human cell identities. Here, *in vitro* models of mammalian development (Huch and Koo, 2015) could offer valuable insights into human development. Another possibility is to leverage non-human primate models, performing cross-species comparisons to infer lineage (Boroviak et al., 2018).

Altogether, considering the restricted opportunity for ground truth lineage tracing in humans, and the above evidence of convergent differentiation, defining cell identity based on lineage alone may not provide accurate cell type classification. However, combining lineage with phenotypic and anatomical features could be powerful, especially given that spatial transcriptomics is now poised to enable the generation of fate maps by supplementing lineage trees with positional information.

State: same identity, different guise

In the previous sections, I explored how high-resolution snapshots of cell phenotype and function, together with lineage, can serve to define cell identity. A third and essential facet of cell identity is 'state', which can be described as the range of cellular phenotypes arising from the interaction of a defined cell type with its environment. T cells serve as a well-characterized example: these cells exist in different activation states, which arise in response to different stimuli, yet they maintain their T-cell identity (Zemmour et al., 2018). Indeed, cell identity is generally stable, maintained by the autoregulation of identity-specifying transcription factors (Holmberg and Perlmann, 2012). In this respect, cell identity can be thought of as 'hard-wired', although it is reprogrammable under defined conditions (as exemplified by Takahashi and Yamanaka, 2006). On the contrary, cell state can be thought of as 'soft-wired', where a given cell type can exist in a range of subtly different states, raising the issue of how cell identity and state can be distinguished for previously uncharacterized cell types. For example, how can we be confident that a novel transcriptional signature represents a new cell type rather than a known cell type in an unrecognized state? As the cell transcriptome adjusts rapidly in response to changes in environmental conditions, reliance on

scRNA-seq-based technologies alone is likely insufficient to address these questions. In this respect, probing heritable, epigenetic signatures of cell identity (reviewed by Ludwig and Bintu, 2019 in this issue) may provide a more stable measure of cell type, permitting identity to be distinguished from state. For example, ATAC-seq (assay for transposase-accessible chromatin using sequencing) provides information on chromatin accessibility and can now be applied at single cell resolution (Cusanovich et al., 2018). Ideally, ‘multi-omic’ measurements will be collected from the same individual cell (Cao et al., 2018), revealing the different transcriptional states that are associated with the same epigenetic signatures. Ultimately, though, these technologies provide only a ‘snapshot’ of cell phenotype within a tissue, with connections between identity and state largely inferred, providing little objective measurement of the states associated with a given identity.

To provide a direct measure of cell state potential, single cell clonal or lineage mapping can be applied to map the emergence of different cell states from a given cell identity. To achieve this, the introduction of perturbations will be essential. For example, several recent methods have employed pooled CRISPR/Cas9 genome editing to introduce a large array of genetic perturbations into a population of cells, followed by measurement of the effects via scRNA-seq or scATAC-seq (Adamson et al., 2016; Dixit et al., 2016; Rubin et al., 2018 preprint). This approach could be modified to expose cells to a range of different environmental perturbations, e.g. exposure to different cytokines, tracking features of clonally related cells under different conditions and pushing given cell types into their full range of potential states. Altogether, this will provide ground truth data that reflect the different cell states that can arise from the same cell identity in response to different environmental cues. Using these approaches, we might also explore more extreme scenarios where cells are pushed over their boundaries, into different identities. In such cases, lineage may prove helpful to distinguish the line between a change in identity versus a dramatic change in state. Overall, for each cell identity, we can attach to it the probability that it will exist in a given state under defined conditions, potentially revealing the molecular regulation underlying hard-wired cell identity and soft-wired cell state.

Perspectives

Here, I have outlined three pillars of cell identity – phenotype (and function), lineage and state – each encompassing a unique and complementary set of measurements that together can serve to define cell identity in a systematic and unbiased manner. This approach will undoubtedly reveal new cell identities that can be placed within a larger cellular taxonomy, providing valuable clues to their physiological role. The full application of this framework in a human context may be somewhat limited at present, due to a reliance on *in vitro* culture systems that do not fully recapitulate *in vivo* counterparts. However, continued efforts to improve human tissue culture models will prove beneficial in this context. Altogether, these three pillars of cell identity will support the construction of high-resolution dynamic cell atlases, with the promise to reveal novel facets of the molecular regulation controlling cell identity, and providing new opportunities for understanding and manipulating cell fate. These endeavors raise some interesting questions: first, is there a minimal set of observations that will serve to universally define cell identity across all cell types and organisms? This leads to a second question: what information do we need to capture from cells to be able to predict their past and future from their present state? This is particularly exciting, as the construction of a probabilistic model of

cell identity could enable, for example, the future disease state of a cell to be predicted, providing new insight into disease progression and diagnosis. These questions are also relevant for the cell fate reprogramming field where, at a minimum, we will gain a high-resolution template to recapitulate the identity of major functional cell types. Once we have amassed a critical amount of information, will the landscape of cell identity be continuous or discrete? If cell identity can indeed exist as a continuum, this presents the opportunity to stabilize transient phenotypes and to create new cell identities, endowing known cell types with new functions. Through our continued efforts to define cell identity, we come closer to realizing these possibilities.

References

- Adamson, B., Norman, T. M., Jost, M., Cho, M. Y., Nuñez, 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.e21. doi:10.1016/j.cell.2016.11.048
- Aleman, A., Florescu, M., Baron, C. S., Peterson-Maduro, J. and van Oudenaarden, A. (2018). Whole-organism clone tracing using single-cell sequencing. *Nature* **556**, 108-112. doi:10.1038/nature25969
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T. et al. (2000). Gene ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25-29. doi:10.1038/75556
- Becht, E., McInnes, L., Healy, J., Dutertre, C.-A., Kwok, I. W. H., Ng, L. G., Ginhoux, F. and Newell, E. W. (2018). Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* **37**, 38-44. doi:10.1038/nbt.4314
- Biddy, B. A., Kong, W., Kamimoto, K., Guo, C., Waye, S. E., Sun, T. and Morris, S. A. (2018). Single-cell mapping of lineage and identity in direct reprogramming. *Nature* **564**, 219-224. doi:10.1038/s41586-018-0744-4
- Boroviak, T., Stirparo, G. G., Dietmann, S., Hernando-Herraez, I., Mohammed, H., Reik, W., Smith, A., Sasaki, E., Nichols, J. and Bertone, P. (2018). Single cell transcriptome analysis of human, marmoset and mouse embryos reveals common and divergent features of preimplantation development. *Development* **145**, dev167833. doi:10.1242/dev.167833
- Cahan, P., Li, H., Morris, S. A., Lummertz da Rocha, E., Daley, G. Q. and Collins, J. J. (2014). CellNet: network biology applied to stem cell engineering. *Cell* **158**, 903-915. doi:10.1016/j.cell.2014.07.020
- Cao, J., Packer, J. S., Ramani, V., Cusanovich, D. A., Huynh, C., Daza, R., Qiu, X., Lee, C., Furlan, S. N., Steemers, F. J. et al. (2017). Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* **357**, 661-667. doi:10.1126/science.aam8940
- Cao, J., Cusanovich, D. A., Ramani, V., Aghamirzaie, D., Pliner, H. A., Hill, A. J., Daza, R. M., McFaline-Figueroa, J. L., Packer, J. S., Christiansen, L. et al. (2018). Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science* **361**, 1380-1385. doi:10.1126/science.aau0730
- Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D. M., Hill, A. J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F. J. et al. (2019). The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496-502. doi:10.1038/s41586-019-0969-x
- Chalfie, M., Sulston, J. E., White, J. G., Southgate, E., Thomson, J. N. and Brenner, S. (1985). The neural circuit for touch sensitivity in *Caenorhabditis elegans*. *J. Neurosci.* **5**, 956-964. doi:10.1523/JNEUROSCI.05-04-00956.1985
- Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W. and Prasher, D. C. (1994). Green fluorescent protein as a marker for gene expression. *Science* **263**, 802-805. doi:10.1126/science.8303295
- Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. and Zhuang, X. (2015). Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090-aaa6090. doi:10.1126/science.aaa6090
- Christiansen, E. M., Yang, S. J., Ando, D. M., Javaherian, A., Skibinski, G., Lipnick, S., Mount, E., O’Neil, A., Shah, K., Lee, A. K. et al. (2018). In silico labeling: predicting fluorescent labels in unlabeled images. *Cell* **173**, 792-803.e19. doi:10.1016/j.cell.2018.03.040
- Coons, A. H., Creech, H. J. and Jones, R. N. (1941). Immunological properties of an antibody containing a fluorescent group. *Exp. Biol. Med.* **47**, 200-202. doi:10.3181/00379727-47-13084P
- Cusanovich, D. A., Hill, A. J., Aghamirzaie, D., Daza, R. M., Pliner, H. A., Berletch, J. B., Filippova, G. N., Huang, X., Christiansen, L., DeWitt, W. S. et al. (2018). A single-cell atlas of *In Vivo* mammalian chromatin accessibility. *Cell* **174**, 1309-1324.e18. doi:10.1016/j.cell.2018.06.052
- Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C. P., Jerby-Aron, L., Marjanovic, N. D., Dionne, D., Burks, T., Raychowdhury, R. et al. (2016). Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* **167**, 1853-1866.e17. doi:10.1016/j.cell.2016.11.038

- Elaut, G., Henkens, T., Papeleu, P., Snykers, S., Vinken, M., Vanhaecke, T. and Rogiers, V. (2006). Molecular mechanisms underlying the dedifferentiation process of isolated hepatocytes and their cultures. *Curr. Drug Metab.* **7**, 629-660. doi:10.2174/138920006778017759
- Eng, C.-H. L., Lawson, M., Zhu, G., Dries, R., Kouloua, N., Takei, Y., Yun, J., Cronin, C., Karp, C., Yuan, G.-C. et al. (2019). Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* **568**, 235-239. doi:10.1038/s41586-019-1049-y
- Han, X., Wang, R., Zhou, Y., Fei, L., Sun, H., Lai, S., Saadatpour, A., Zhou, Z., Chen, H., Ye, F. et al. (2018). Mapping the mouse cell atlas by microwell-seq. *Cell* **172**, 1091-1107.e17. doi:10.1016/j.cell.2018.02.001
- Holmberg, J. and Perlmann, T. (2012). Maintaining differentiated cellular identity. *Nat. Rev. Genet.* **13**, 429-439. doi:10.1038/nrg3209
- Hooke, R. (1665). *Micrographia, or, Some Physiological Descriptions of Minute Bodies made by Magnifying Glasses: with Observations and Inquiries Thereupon* (ed. R. Hooke). London: Printed by Jo. Martyn and Ja. Allestry, printers to the Royal Society.
- Huch, M. and Koo, B.-K. (2015). Modeling mouse and human development using organoid cultures. *Development* **142**, 3113-3125. doi:10.1242/dev.118570
- Klein, A. M. M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D. A. A. and Kirschner, M. W. W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187-1201. doi:10.1016/j.cell.2015.04.044
- Kretzschmar, K. and Watt, F. M. (2012). Lineage tracing. *Cell* **148**, 33-45. doi:10.1016/j.cell.2012.01.002
- Lee, J. H., Daugharthy, E. R., Scheiman, J., Kalhor, R., Yang, J. L., Ferrante, T. C., Terry, R., Jeanty, S. S. F., Li, C., Amamoto, R. et al. (2014). Highly multiplexed subcellular RNA sequencing in situ. *Science* **343**, 1360-1363. doi:10.1126/science.1250212
- 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. *Nat. Biotechnol.* **29**, 928-933. doi:10.1038/nbt.1977
- Ludwig, C. H. and Bintu, L. (2019). Mapping chromatin modifications at the single cell level. *Development* **146**, dev170217. doi:10.1242/dev.170217
- Ludwig, L. S., Lareau, C. A., Ulirsch, J. C., Christian, E., Muus, C., Li, L. H., Pelka, K., Ge, W., Oren, Y., Brack, A. et al. (2019). Lineage tracing in humans enabled by mitochondrial mutations and single-cell genomics. *Cell* **176**, 1325-1339.e22. doi:10.1016/j.cell.2019.01.022
- 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. doi:10.1016/j.cell.2015.05.002
- Mayr, U., Serra, D. and Liberali, P. (2019). Exploring single cells in space and time during tissue development, homeostasis and regeneration. *Development* **146**, dev176727. doi:10.1242/dev.176727
- McKenna, A. and Gagnon, J. A. (2019). Recording development with single cell dynamic lineage tracing. *Development* **146**, dev169730. doi:10.1242/dev.169730
- Palmiter, R. D., Behringer, R. R., Quaife, C. J., Maxwell, F., Maxwell, I. H. and Brinster, R. L. (1987). Cell lineage ablation in transgenic mice by cell-specific expression of a toxin gene. *Cell* **50**, 435-443. doi:10.1016/0092-8674(87)90497-1
- Pearse, A. G. E. (1984). History of staining (third edition). Edited by G. Clark and F. H. Kasten. Williams and Wilkins, Baltimore, 1983. No. of pages: 304. Price: £22.25. ISBN: 0683 017 055. *J. Pathol.* **143**, 139-139. doi:10.1002/path.1711430209
- Peterson, V. M., Zhang, K. X., Kumar, N., Wong, J., Li, L., Wilson, D. C., Moore, R., McClanahan, T. K., Sadekova, S. and Klappenbach, J. A. (2017). Multiplexed quantification of proteins and transcripts in single cells. *Nat. Biotechnol.* **35**, 936-939. doi:10.1038/nbt.3973
- Raj, B., Wagner, D. E., McKenna, A., Pandey, S., Klein, A. M., Shendure, J., Gagnon, J. A. and Schier, A. F. (2018). Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. *Nat. Biotechnol.* **36**, 442-450. doi:10.1038/nbt.4103
- Ramón y Cajal, S. (1888). *Estructura de los Centros Nerviosos de las Aves*. Madrid, Spain: Jiménez y Molina.
- Regev, A., Teichmann, S. A., Lander, E. S., Amit, I., Benoist, C., Birney, E., Bodenmiller, B., Campbell, P., Carninci, P., Ciatworthy, M. et al. (2017). The human cell atlas. *Elife* **6**, e27041. doi:10.7554/eLife.27041
- Rodrigues, S. G., Stickels, R. R., Goeva, A., Martin, C. A., Murray, E., Vanderburg, C. R., Welch, J., Chen, L. M., Chen, F. and Macosko, E. Z. (2019). Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463-1467. doi:10.1126/science.aaw1219
- Roost, M. S., van Iperen, L., Ariyurek, Y., Buermans, H. P., Arindrarto, W., Devalla, H. D., Passier, R., Mummery, C. L., Carlotti, F., de Koning, E. J. P. et al. (2015). KeyGenes, a tool to probe tissue differentiation using a human fetal transcriptional atlas. *Stem Cell Rep.* **4**, 1112-1124. doi:10.1016/j.stemcr.2015.05.002
- Rosenberg, A. B., Roco, C. M., Muscat, R. A., Kuchina, A., Sample, P., Yao, Z., Graybus, L. T., Peeler, D. J., Mukherjee, S., Chen, W. et al. (2018). Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* **360**, 176-182. doi:10.1126/science.aam8999
- Rubin, A. J., Parker, K. R., Satpathy, A. T., Qi, Y., Wu, B., Ong, A. J., Mumbach, M. R., Ji, A. L., Kim, D. S., Cho, S. W. et al. (2018). Coupled single-cell CRISPR screening and epigenomic profiling reveals causal gene regulatory networks. *Cell* **176**, 361-376. doi:10.1016/j.cell.2018.11.022
- Saelens, W., Cannoodt, R., Todorov, H. and Saeys, Y. (2019). A comparison of single-cell trajectory inference methods. *Nat. Biotechnol.* **37**. doi:10.1038/s41587-019-0071-9
- Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495-502. doi:10.1038/nbt.3192
- Smith, K., Piccinini, F., Balassa, T., Koos, K., Danka, T., Azizpour, H. and Horvath, P. (2018). Phenotypic image analysis software tools for exploring and understanding big image data from cell-based assays. *Cell Syst.* **6**, 636-653. doi:10.1016/j.cels.2018.06.001
- Spanjaard, B., Hu, B., Mitic, N., Olivares-Chauvet, P., Janjuha, S., Ninov, N. and Junker, J. P. (2018). Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars. *Nat. Biotechnol.* **36**, 469-473. doi:10.1038/nbt.4124
- Stoeckius, M., Hafemeister, C., Stephenson, W., Houck-Loomis, B., Chattopadhyay, P. K., Swerdlow, H., Satija, R. and Smibert, P. (2017). Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865-868. doi:10.1038/nmeth.4380
- Stuart, T. and Satija, R. (2019). Integrative single-cell analysis. *Nat. Rev. Genet.* **20**, 257-272. doi:10.1038/s41576-019-0093-7
- Sulston, J. E., Schierenberg, E., White, J. G. and Thomson, J. N. (1983). The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Dev. Biol.* **100**, 64-119. doi:10.1016/0012-1606(83)90201-4
- Tabula Muris Consortium, Overall coordination, Logistical coordination, Organ collection and processing, Library preparation and sequencing, Computational data analysis, Cell type annotation, Writing group, Supplemental text writing group and Principal investigators (2018). Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562**, 367-372. doi:10.1038/s41586-018-0590-4
- Takahashi, K. and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-676. doi:10.1016/j.cell.2006.07.024
- Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B. B., Siddiqui, A. et al. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377-382. doi:10.1038/nmeth.1315
- Thul, P. J., Åkesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Ait Blal, H., Alm, T., Asplund, A., Björk, L., Breckels, L. M. et al. (2017). A subcellular map of the human proteome. *Science* **356**, eaal3321. doi:10.1126/science.aal3321
- Tritschler, S., Büttner, M., Fischer, D. S., Lange, M., Bergen, V., Lickert, H. and Theis, F. J. (2019). Concepts and limitations for learning developmental trajectories from single cell genomics. *Development* **146**, dev170506. doi:10.1242/dev.170506
- Variou (2017). What is your conceptual definition of cell type in the context of a mature organism? *Cell Syst.* **4**, 255-259. doi:10.1016/j.cels.2017.03.006
- Wagner, D. E., Weinreb, C., Collins, Z. M., Briggs, J. A., Megason, S. G. and Klein, A. M. (2018). Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* **360**, 981-987. doi:10.1126/science.aar4362
- Xia, B. and Yanai, I. (2019). A periodic table of cell types. *Development* **146**, dev169854. doi:10.1242/dev.169854
- Yao, Z., Mich, J. K., Ku, S., Menon, V., Krostag, A.-R., Martinez, R. A., Furchtgott, L., Mulholland, H., Bort, S., Fuqua, M. A. et al. (2017). A single-cell roadmap of lineage bifurcation in human ESC models of embryonic brain development. *Cell Stem Cell* **20**, 120-134. doi:10.1016/j.stem.2016.09.011
- Zemmour, D., Zilionis, R., Kiner, E., Klein, A. M., Mathis, D. and Benoist, C. (2018). Single-cell gene expression reveals a landscape of regulatory T cell phenotypes shaped by the TCR. *Nat. Immunol.* **19**, 291-301. doi:10.1038/s41590-018-0051-0
- Zitnik, M. and Leskovec, J. (2017). Predicting multicellular function through multi-layer tissue networks. *Bioinformatics* **33**, i190-i198. doi:10.1093/bioinformatics/btx252