possible to link regulatory networks with clustering and trajectory inference, which will in turn increase the power of causal inference. Follow-up validation of candidate regulatory elements using genetic deletions or CRISPR interference (CRISPRi) would be crucial (Box 1).

Finally, integration with spatial methods will enable identification of context-specific, functionally relevant relationships and how these shape cellular phenotypes. Novel computational methods offer an opportunity to exploit the full potential of single-cell multimodal omics sequencing techniques and will deepen our understanding of cellular identity and responses in both health and disease. ❐

Mirjana Efremova[1] and Sarah A. Teichmann[1,2]*
[1]*Wellcome Sanger Institute, Wellcome Genome Campus, Cambridge, UK.* [2]*Theory of Condensed Matter Group, Cavendish Laboratory, University of Cambridge, Cambridge, UK.*
**e-mail: st9@sanger.ac.uk*

Published online: 6 January 2020
https://doi.org/10.1038/s41592-019-0692-4

References

1. Stuart, T. et al. *Cell* **177**, 1888–1902.e21 (2019).
2. Welch, J. D. et al. *Cell* **177**, 1873–1887.e17 (2019).
3. Lopez, R. et al. A joint model of unpaired data from scRNA-seq and spatial transcriptomics for imputing missing gene expression measurements. Preprint at *arXiv* https://arxiv.org/abs/1905.02269 (2019).
4. Kester, L. & van Oudenaarden, A. *Cell Stem Cell* **23**, 166–179 (2018).
5. Ludwig, L. S. et al. *Cell* **176**, 1325–1339.e22 (2019).
6. Xu, J. et al. *eLife* **8**, e45105 (2019).
7. McCarthy, D. J. et al. Cardelino: integrating whole exomes and single-cell transcriptomes to reveal phenotypic impact of somatic variants. Preprint at *bioRxiv* https://doi.org/10.1101/413047 (2018).
8. Satpathy, A. T. et al. *Nat. Med.* **24**, 580–590 (2018).
9. Cuomo, A. S. E. et al. Single-cell RNA-sequencing of differentiating iPS cells reveals dynamic genetic effects on gene expression. Preprint at *bioRxiv* https://doi.org/10.1101/630996 (2018).
10. Aibar, S. et al. *Nat. Methods* **14**, 1083–1086 (2017).
11. Hainer, S. J., Bošković, A., McCannell, K. N., Rando, O. J. & Fazzio, T. G. *Cell* **177**, 1319–1329.e11 (2019).
12. Welch, J. D., Hartemink, A. J. & Prins, J. F. *Genome Biol.* **18**, 138(2017).
13. Burdziak, C., Azizi, E., Prabhakaran, S. & Pe'er, D. A nonparametric multi-view model for estimating cell type-specific gene regulatory networks. Preprint at *arXiv* https://arxiv.org/abs/1902.08138 (2019).
14. Henriksson, J. *Single Cell Methods.: Methods. Mol. Biol.* **1979**, 395–406 (2019).
15. Krishnaswamy, S. et al. *Science* **346**, 1250689 (2014).
16. Qin, X. et al. Single-cell signalling analysis of heterocellular organoids. Preprint at *bioRxiv* https://doi.org/10.1101/659896 (2019).
17. Stoeckius, M. et al. *Nat. Methods* **14**, 865–868 (2017).
18. Peterson, V. M. et al. *Nat. Biotechnol.* **35**, 936–939 (2017).
19. Gayoso, A. et al. A joint model of RNA expression and surface protein abundance in single cells. Preprint at *bioRxiv* https://doi.org/10.1101/791947 (2019).
20. Markowetz, F., Kostka, D., Troyanskaya, O. G. & Spang, R. *Bioinformatics* **23**, i305–i312 (2007).
21. Pirkl, M. & Beerenwinkel, N. *Bioinformatics* **34**, i964–i971 (2018).
22. Mayr, U., Serra, D. & Liberali, P. *Development* **146**, dev176727 (2019).
23. Halpern, K. B. et al. *Nature* **542**, 352–356 (2017).
24. Karaiskos, N. et al. The *Drosophila embryo* at single cell transcriptome resolution. *Science* **358**, 194–199 (2017).
25. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. *Nat. Biotechnol.* **33**, 495–502 (2015).
26. Achim, K. et al. *Nat. Biotechnol.* **33**, 503–509 (2015).
27. Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB v2.0: inferring cell-cell communication from combined expression of multi-subunit receptor-ligand complexes. Preprint at *bioRxiv* https://doi.org/10.1101/680926 (2019).
28. Colomé-Tatché, M. & Theis, F. J. *Curr. Opin. Syst. Biol.* **7**, 54–59 (2018).
29. Packer, J. & Trapnell, C. *Trends Genet.* **34**, 653–665 (2018).
30. Argelaguet, R. et al. MOFA: a probabilistic framework for comprehensive integration of structured single-cell data. Preprint at *bioRxiv* https://doi.org/10.1101/837104 (2019).

Acknowledgements

We thank E. Dann and M. Sarkin Jain for careful and critical reading of the manuscript. We are grateful to J. Eliasova for help with the illustrations.

Author contributions

M.E and S.A.T wrote the manuscript.

Competing interests

The authors declare no competing interests.

Single-cell biology: beyond the sum of its parts

The field of single-cell RNA sequencing (scRNA-seq) has been paired with genomics, epigenomics, spatial omics, proteomics and imaging to achieve multimodal measurements of individual cellular phenotypes and genotypes. In its purest form, single-cell multimodal omics involves the simultaneous detection of multiple traits in the same cell. More broadly, multimodal omics also encompasses comparative pairing and computational integration of measurements made across multiple distinct cells to reconstruct phenotypes. Here I highlight some of the biological insights gained from multimodal studies and discuss the challenges and opportunities in this emerging field.

Alexander F. Schier

"Single-cell sequencing" was the Method of the Year in 2013. Since then, the impact of scRNA-seq has only grown, largely because throughput has increased dramatically through droplet or combinatorial indexing approaches, and application has been democratized through the development of off-the-shelf equipment, reagents and analysis tools. A primary contribution of these technologies has been the identification of previously uncharacterized cell types and cell states in heterogeneous samples. For example, in the well-studied human lung, an ionocyte-like cell type was discovered that expresses cystic fibrosis transmembrane conductance regulator and might underlie cystic fibrosis[1,2]. Even in the extensively analyzed early zebrafish embryo, a novel cell state was discovered that has hallmarks of apoptosis and cellular stress but whose origin and fate are mysterious[3]. Progress in scRNA-seq has been particularly remarkable in the nervous and immune systems, with the definition of dozens or even hundreds of transcriptionally distinct cell types.

As exemplified by this year's Method of the Year, "Single-cell multimodal omics," the field of single-cell genomics is moving from scRNA-seq snapshots to multimodal measurements of cellular phenotypes and genotypes. Here I discuss recent examples of biological insights gained by single-cell multimodal omics and speculate what the future might hold. For more detailed coverage of the different subfields, the reader is referred to excellent recent reviews[4–10].

Single-cell trajectories

scRNA-seq snapshots are powerful indicators of cellular diversity, but they tell us little about the history or biography of a cell. What was a cell's previous molecular trajectory to arrive at the current state? What was the division pattern that defines

its lineage relationship with other cells? Not only are these questions important for developmental biologists, but they are also defining features of diseases such as cancer.

Computational methods have helped to reconstruct the molecular trajectories of cells[11,12]. These maps are multimodal in the sense that they provide both snapshots and the potential transcriptional history of a cell. Current trajectories reach from simple differentiation paths to complex trees, networks or state manifolds. For example, transcriptional trajectories of vertebrate embryogenesis have been reconstructed from the pluripotent state to organ formation[13–17]. These studies revealed unexpected cellular plasticity at decision branchpoints, identified convergent paths to create specific cell types and phenotyped developmental mutants at whole-transcriptome scale.

Despite the impressive progress in reconstructing developmental trajectories, it is important to note that these trajectories are only *models* of cell type specification, as it is not yet feasible to measure the same cell at the whole-transcriptome level more than once. For example, the trajectories often portray differentiation as a continuous path, but it is unclear whether a single cell's gene-expression path is indeed continuous or whether cells jump from state to state. We also do not know which representation most accurately reflects the phenotypic landscape during development[13–18]. Does the Waddingtonian view of valleys and hills hold true, or do we need to consider other landscapes? Is "landscapes" even the proper analogy for multidimensional phenotypic complexity? Addressing these questions requires the multiplex in vivo measuring of dozens of transcripts over time and at single-cell resolution—a Holy Grail technology that is not yet available.

Another limitation of current trajectory reconstructions is the necessary focus on transcription. However, most developmental decisions are initiated by the post-transcriptional regulation of signaling pathways, such as the phosphorylation or proteolysis of transcriptional regulators. Transcriptomes can therefore only capture cell fate decisions after they have been initiated. Hence, many reconstructed bifurcation points are temporally delayed with respect to the first step of cell type diversification.

Single-cell lineages

Molecular trajectories do not necessarily reflect the lineage relationships of cells. The recent mapping of single-cell transcriptomes onto the *Caenorhabditis elegans* lineage tree offers important lessons for the reconstruction and multimodal comparison of developmental trajectories and lineages[19]. Notably, there is limited correlation between transcriptome and lineage as cells become highly specialized. Moreover, there are many paths cells can take to converge on the same terminal transcriptomic state. These results indicate that the relationship between lineage and molecular trajectory can be complex, and that it would be naive to assume that an identical cascade of transcriptional regulators specifies a particular cell type independent of its lineage history. Indeed, a recent study identified a transcription factor, UNC-130, that acts as a lineage selector for one specific sublineage to generate a particular glial cell type[20]. The same cell type is also generated by a different sublineage that develops independently of UNC-130. Hence, different lineages and different trajectories can result in the generation of the same terminal cell type.

To track the lineage relationships between cells at global scales, genomic barcoding methods have been developed[4,5]. Akin to the phylogenetic reconstruction of evolutionary relationships based on mutations, these clonal or lineage-tracing methods introduce changes in the genome as cells divide. Inheritance of shared genomic changes can then reveal the ancestral relationships between cells. Using naturally accumulating genomic changes, lineage trees can even be constructed from human samples, particularly in tumors with high mutation rates and in the immune system by analyzing the T cell receptor repertoire[21–23].

Genomic lineage tracing, in combination with scRNA-seq or other modalities, has provided interesting insights into the dynamics of cell specification. For example, it identified striking differences between steady-state and regenerative hematopoiesis[24], revealed fine-grained clonal relationships in the hypothalamus[25], showed that only a handful of pre-gastrula cells give rise to the large majority of differentiated cells of a given organ[4] and identified a new lineage of myeloid-like cells[5]. In the future, these technologies have the potential to address at global scales the role of lineage history in determining the final cellular phenotype. Studies in *Drosophila melanogaster* have shown remarkable correlation between lineage and cell type, transmitter phenotype, morphology and connectivity of neurons[26,27]. In vertebrates, however, the fundamental rules are often unknown, and more stochastic and regulative processes might be at play, rather than determinate lineages[28,29]. Such scenarios raise the question of what kind of lineage motifs are used to generate organs of the correct size and cell type composition, and could lead to a field of developmental statistics that compares the distribution of lineage and trajectory motifs between individuals or species.

Despite the progress in genomic lineage tracing, it is important to note that the current barcode-derived lineage trees capture only a small part of the ancestral relationships of all cells—they are trees that lack most branches and leaves, because current barcode diversity does not uniquely define each cell, most cells are not recovered upon dissociation or sequencing, and barcode editing is temporally limited. It is also not clear when and in what progenitor type a particular edit was introduced, and instances of apoptosis are not captured.

Clonal resampling combined with cumulative barcoding can address some of these issues. In one example, a population of fibroblast cells was genetically barcoded during reprogramming to endoderm. Single-cell barcodes and transcriptomes were sampled at different stages of reprogramming, and cells were then mapped onto the RNA trajectory manifold and clonal coupling was analyzed[30]. This analysis identified dead-end paths of reprogramming and revealed a putative methyltransferase associated with the successful reprogramming trajectory. Resampling analyses can reveal the potential paths that cells can take during differentiation and reprogramming. However, this approach is not easily applicable in vivo. Therefore, most in vivo systems will require barcode editing that is targeted to specific cells at specific times, and will combine genomic lineage tracing with imaging approaches.

Single cells in space

scRNA-seq experiments rest on the dissociation of tissues and result in the loss of spatial and morphological information. To reconstruct the spatial origin of newly characterized cells, initial approaches have used marker or reporter genes. In some cases, the computational pairing of scRNA-seq data with the spatial expression of landmark genes allows the genome-wide reconstruction of gene-expression domains[3,31]. More recently, barcoded arrays or multiplex in situ hybridization–imaging approaches have mapped the expression of dozens or even thousands of genes at (sub) cellular resolution[32,33].

In combination with other cellular phenotypes such as lineage[34], morphology or physiology, RNA localization studies have provided insights into cellular diversity and have raised the overarching question how a cell's transcriptome, location, morphology and physiology relate to each

other. Particularly in the nervous system, the relationships between different modalities are very complex. Although there is a general overlap of transcriptomic profiles with electrophysiological and morphological properties[8,9,35,36], there are also very clear examples in which morphological diversity is not reflected in the transcriptome. For example, mature *Drosophila* olfactory projection neurons can have indistinguishable transcriptomic profiles but different projection patterns[37]. Strikingly, these morphological changes are still guided by transcriptomic differences during the differentiation of these neurons but are no longer present upon maturation. Hence, correlations between single-cell modalities can be stage dependent.

The relationships between function and transcriptome are also complex and controversial. Two recent studies used scRNA-seq, multiplex in situ hybridization, axonal projection mapping and neuronal activity measurement to classify dozens of cell types and their locations in the mouse hypothalamus[38,39]. Surprisingly, specific neuronal populations exhibited behavior-specific activation in one[38] but not the other study[39]. The cause for these discrepancies is not known, but the studies point to complex relationships between the molecular identity and functional roles of individual neurons[8,9].

RNA meets DNA

Efforts to sequence both the genome and RNA from the same cell are still in their infancy, but targeted sequencing of genomic regions of interest in combination with scRNA-seq allows insights into genotype–phenotype connections. Fruitful applications involve the correlation of mutations and transcriptomes in CRISPR screens or during cancer heterogeneity. For example, distinct AML subtypes defined by scRNA-seq correspond to different genetic alterations[40]. These multimodal measurements will also inform attempts to target cancerous cells by gene-editing.

Transcription is driven by cis-regulatory DNA elements. To understand gene regulation, one would therefore want to define enhancer activity, chromatin state, chromosomal conformation and RNA expression. By miniaturizing approaches such as methylome sequencing, ATAC-seq and ChIP-seq, multimodal datasets have been generated that combine RNA with genomic features. For example, the correlation of transcriptome with chromatin accessibility data helps link cis-regulatory regions to genes[41]. In another recent example, the combination of methylome, nuclease-sensitivity and mRNA data revealed that during early mouse embryogenesis the exit from pluripotency coincides with the formation of a repressive chromatin landscape[42]. Mesoderm and endoderm progenitors display cell-type- and gene-specific increases in chromatin accessibility and decreases in enhancer DNA methylation. In contrast, the epigenome of ectodermal cells is already established before cell type specification, consistent with the classic view that ectodermal commitment is the default mode of germ layer formation. In another study, combined single-cell methylome and RNA sequencing suggested that aging muscle stem cells do not have a methylation clock but instead show uncoordinated methylation changes at promoters and increased transcriptional variability[43]. These studies highlight the potential of multimodal analyses to compare and contrast different phenotypic traits within single cells.

Challenges and opportunities

Single-cell multimodal omics approaches have become a powerful tool to phenotype cells, but many challenges remain. First, we need to capture the full complement of cellular traits. For example, one major void is the lack of comprehensive single-cell protein or ribosome profiling approaches, because current antibody-based methods are limited to the co-detection of a few dozen proteins[44–46]. We also need to be able to reconstruct and record the full history of cells. For example, it is currently impossible to faithfully record or reconstruct the molecular trajectories and lineages that result in the hundreds of cell types in the vertebrate brain. With such advances, the biographies of cells would no longer be mostly empty books.

Multimodal analyses also amplify the current struggle to define "cell type"[8,9]. Hierarchical definitions based on variably expressed genes provide some criteria to define the relationship between cells, but multimodal analysis raises the question how to weigh different traits. For example, are epithelial cells in the gut and skin primarily epithelial cells or primarily gut or skin cells? This issue is further amplified in cellular comparisons between disease states or through evolution[47–49]. Cell type definitions are likely to become context and modality dependent and will be guided by specific biological questions[50,51].

Although single-cell approaches have led to important discoveries in specific systems, detractors might argue that the main biological findings have 'only' resulted in the extension of established paradigms into more global rules. They might also argue that the large datasets themselves have not revealed any fundamental new principles. However, three recent examples suggest that multimodal analyses can reveal unexpected biology. First, large-scale in situ profiling of nascent transcripts revealed global transcript oscillations in mouse embryonic stem cells[52]. Oscillations follow a two-hour period and are not synchronized between cells, but it remains unclear how they are generated. Second, global analysis of apical versus basal localization of mRNAs in the gut epithelium revealed that more than 600 mRNAs were polarized[53]. Surprisingly, however, local protein abundance did not correlate with RNA localization. Instead, a significantly higher abundance of ribosomal proteins was found at the apical side, leading to higher local translation. Thus, the global subcellular analysis of proteins and RNAs revealed system properties that had not been recognized by smaller-scale approaches. Third, multimodal single-cell imaging reconstructed the emergence of a motor circuit in zebrafish embryos[54]. Combined tracking of neuron lineages, migration, marker gene expression and activity revealed that early-born motor neurons guide local circuit activity and that neighboring neurons from the same lineage share functional properties. Thus, multimodal single-cell imaging can discover how neuronal ensembles are assembled and activated. These three examples show the promise of multimodal studies in uncovering new biology and allow us to hope for many more discoveries achieved through single-cell multimodal omics. ❐

Alexander F. Schier[1,2,3]

[1]*Biozentrum, University of Basel, Basel, Switzerland.* [2]*Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA, USA.* [3]*Allen Discovery Center for Cell Lineage Tracing, University of Washington, Seattle, WA, USA.*
e-mail: alex.schier@unibas.ch

Published online: 6 January 2020
https://doi.org/10.1038/s41592-019-0693-3

References

1. Plasschaert, L. W. et al. *Nature* **560**, 377–381 (2018).
2. Montoro, D. T. et al. *Nature* **560**, 319–324 (2018).
3. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. *Nat. Biotechnol.* **33**, 495–502 (2015).
4. McKenna, A. & Gagnon, J. A. *Development* **146**, dev169730 (2019).
5. Baron, C. S. & van Oudenaarden, A. *Nat. Rev. Mol. Cell Biol.* **20**, 753–765 (2019).
6. Suvà, M. L. & Tirosh, I. *Mol. Cell* **75**, 7–12 (2019).
7. Nathan, A., Baglaenko, Y., Fonseka, C. Y., Beynor, J. I. & Raychaudhuri, S. *Curr. Opin. Immunol.* **61**, 17–25 (2019).
8. Zeng, H. & Sanes, J. R. *Nat. Rev. Neurosci.* **18**, 530–546 (2017).
9. Huang, Z. J. & Paul, A. *Nat. Rev. Neurosci.* **20**, 563–572 (2019).
10. Gottgens, B. & Kucinski, I. in *Stem Cells: From Biological Principles to Regenerative Medicine* (eds. Lo Celso, C., Red-Horse, K. & Watt, F. M.) (Cold Spring Harbor Laboratory Press, 2019); https://doi.org/10.17863/CAM.44977.
11. Trapnell, C. et al. *Nat. Biotechnol.* **32**, 381–386 (2014).
12. Bendall, S. C. et al. *Cell* **157**, 714–725 (2014).
13. Farrell, J. A. et al. *Science* **360**, eaar3131 (2018).

14. Briggs, J. A. et al. *Science* **360**, eaar5780 (2018).
15. Wagner, D. E. et al. *Science* **360**, 981–987 (2018).
16. Cao, J. et al. *Nature* **566**, 496–502 (2019).
17. Pijuan-Sala, B. et al. *Nature* **566**, 490–495 (2019).
18. Soldatov, R. et al. *Science* **364**, eaas9536 (2019).
19. Packer, J. S. et al. *Science* **365**, eaax1971 (2019).
20. Mizeracka, K., Rogers, J. M., Shaham, S., Bulyk, M. L. & Heiman, M. G. Preprint at https://doi.org/10.1101/758508 (2019).
21. Stubbington, M. J. T. et al. *Nat. Methods* **13**, 329–332 (2016).
22. Lee-Six, H. et al. *Nature* **561**, 473–478 (2018).
23. Ludwig, L. S. et al. *Cell* **176**, 1325–1339.e22 (2019).
24. Sun, J. et al. *Nature* **514**, 322–327 (2014).
25. Raj, B. et al. *Nat. Biotechnol.* **36**, 442–450 (2018).
26. Lacin, H. et al. *eLife* **8**, 1113 (2019).
27. Doe, C. Q. *Annu. Rev. Cell Dev. Biol.* **33**, 219–240 (2017).
28. He, J. et al. *Neuron* **75**, 786–798 (2012).
29. Sánchez-Guardado, L. & Lois, C. *eLife* **8**, 766 (2019).
30. Biddy, B. A. et al. *Nature* **564**, 219–224 (2018).
31. Achim, K. et al. *Nat. Biotechnol.* **33**, 503–509 (2015).
32. Lein, E., Borm, L. E. & Linnarsson, S. *Science* **358**, 64–69 (2017).
33. Eng, C. L. et al. *Nature* **568**, 235–239 (2019).
34. Frieda, K. L. et al. *Nature* **541**, 107–111 (2017).
35. Muñoz-Manchado, A. B. et al. *Cell Reports* **24**, 2179–2190.e7 (2018).
36. Gouwens, N. W. et al. *Nat. Neurosci.* **22**, 1182–1195 (2019).
37. Li, H. et al. *Cell* **171**, 1206–1220.e22 (2017).
38. Moffitt, J. R. et al. *Science* **362**, eaau5324 (2018).
39. Kim, D.-W. et al. *Cell* **179**, 713–728.e17 (2019).
40. van Galen, P. et al. *Cell* **176**, 1265–1281.e24 (2019).
41. Cao, J. et al. *Science* **361**, 1380–1385 (2018).
42. Argelaguet, R. et al. *Nature* https://doi.org/10.1038/s41586-019-1825-8 (2019).
43. Hernando-Herraez, I. et al. *Nat. Commun.* **10**, 4361 (2019).
44. Saka, S. K. et al. *Nat. Biotechnol.* **37**, 1080–1090 (2019).
45. Goltsev, Y. et al. *Cell* **174**, 968–981.e15 (2018).
46. Gut, G., Herrmann, M. D. & Pelkmans, L. *Science* **361**, eaar7042 (2018).
47. Marioni, J. C. & Arendt, D. *Annu. Rev. Cell Dev. Biol.* **33**, 537–553 (2017).
48. Shafer, M. E. R. *Front. Cell Dev. Biol.* **7**, 175 (2019).
49. Camp, J. G., Platt, R. & Treutlein, B. *Science* **365**, 1401–1405 (2019).
50. Wee, C. L. et al. *Nat. Neurosci.* **22**, 1477–1492 (2019).
51. Lovett-Barron, M. et al. Preprint at *bioRxiv* https://doi.org/10.1101/745075 (2019).
52. Shah, S. et al. *Cell* **174**, 363–376.e16 (2018).
53. Moor, A. E. et al. *Science* **357**, 1299–1303 (2017).
54. Wan, Y. et al. *Cell* **179**, 355–372.e23 (2019).

Acknowledgements

I am grateful to J. Farrell, S. Mango and B. Raj for comments on the manuscript and to the US National Institutes of Health, the McKnight Endowment Fund for Neuroscience, the Allen Discovery Center for Cell Lineage Tracing, Harvard University and the University of Basel for support.

Competing interests

The author declares no competing interests.