The Company of
Biologists

SPOTLIGHT

Human organomics: a fresh approach to understanding human development using single-cell transcriptomics

J. Gray Camp[1] and Barbara Treutlein[1,2,3,*]

ABSTRACT

Innovative methods designed to recapitulate human organogenesis from pluripotent stem cells provide a means to explore human developmental biology. New technologies to sequence and analyze single-cell transcriptomes can deconstruct these 'organoids' into constituent parts, and reconstruct lineage trajectories during cell differentiation. In this Spotlight article we summarize the different approaches to performing single-cell transcriptomics on organoids, and discuss the opportunities and challenges of applying these techniques to generate organ-level, mechanistic models of human development and disease. Together, these technologies will move past characterization to the prediction of human developmental and disease-related phenomena.

KEY WORDS: Organogenesis, Human development, Pluripotent stem cell, Organoid, Transcriptome, Single-cell transcriptomics

Introduction

Understanding how multiple different cell types come together to build an organ has been a long-standing fascination in developmental biology. Over the years, we have learned much with regard to the molecular events that instruct cell lineage, the specific growth factors that are required, and the morphological aspects that drive organ development. Most of this knowledge has been gained from studying non-human vertebrate organogenesis; however, the observation that differences exist between how organs are formed across a range of species has led us to question what it is that makes us uniquely human. The revelation that human pluripotent stem cells can self-organize into three-dimensional structures that contain multiple differentiated cell types organized to resemble primary human tissue has revitalized the field of human developmental biology (McCauley and Wells, 2017). In general, these structures are referred to as organoids, and protocols have been developed to generate gut, kidney, liver bud, multiple regions of the human brain, and other tissues (McCauley and Wells, 2017). Conventional strategies to analyze human organoid development often assess cell composition and differentiation using immunohistochemistry of a limited set of marker proteins, or cell tracking via a reporter gene. Because organoids are, by definition, composed of many different cell states and often show large organoid-to-organoid variability, high-throughput single-cell transcriptomics represents an exciting strategy to assess cell composition, lineage relationships, and gene networks in organoids. In this Spotlight article, we discuss how human

organomics, by which we mean the application of functional genomics to study human organ development at the single-cell level, can be applied to organoids in order to improve our understanding of human development. We focus on single-cell RNA sequencing (scRNA-seq), summarizing some of the commonly used methods and their advantages and limitations. We further discuss how we envisage improvements in single-cell transcriptomic methodology will enhance our comprehension of human developmental biology and disease. For a full glossary of technical terms and acronyms used throughout this article, please see Box 1.

Single-cell transcriptomics: one technology, many methods
Capturing RNA from single cells

Starting from a solid tissue, the general strategy of each scRNA-seq approach is to first dissociate the tissue into a single-cell suspension, then capture individual cells into isolated compartments, lyse the cells, prepare amplified cDNA from the RNA (usually mRNA) and at the end generate a multiplexed sequencing library, whereby all cDNA molecules from one individual cell contain the same unique sequence called a cell barcode. Existing single-cell transcriptomic methods differ in the way single cells are captured and compartmentalized, the way amplified cDNA is prepared and the way cell barcodes are introduced into the cDNA molecules. In this section, we summarise the most common methods for cell capture, for preparation of the cDNA and for cell barcoding. For greater detail on each capture and chemistry method, we refer the reader to two recent reviews (Kumar et al., 2017; Kolodziejczyk et al., 2015).

There are multiple platforms to perform cell capture, including valve-based microfluidics (Fluidigm C1 or home-built chips), droplet-based microfluidics [Drop-seq, inDrop, commercial brands Chromium (10x Genomics) and BDTM Resolve System (BD Genomics)], sorting single cells into wells of a multi-well plate (Smart-seq1/2, MARS-seq, CEL-seq2, SCRB-seq) or randomly dispensing cells into wells on microwell plates (commercial brand WaferGen), and each platform has its own advantages and drawbacks. Once individual cells are isolated, cDNA is usually generated by reverse transcription priming off the poly-A tail of mRNA. The resulting cDNA can then be amplified, either exponentially in a PCR reaction (Drop-seq, Smart-seq1/2, STRT-seq, SCRB-seq) or quasi-linearly through *in vitro* transcription (CEL-seq1/2, inDrop). The PCR-based method Smart-seq2 is based on template-switching reverse transcriptases that ensure the generation of full-length cDNA and sequencing across the entire transcript (Picelli et al., 2013). Most other methods only sequence the 3′ or 5′ end of transcripts, which allows for attachment of unique molecular identifiers that enable individual molecules to be counted (STRT-seq, Drop-seq, inDrop, SCRB-seq, CEL-seq, MARS-seq) rather than estimating mRNA abundance by normalizing against mapped reads. In all cases, a major challenge in cell capture is to generate a suspension such that the cells are mostly viable singlets.

[1]Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany. [2]Max Planck Institute of Molecular Cell Biology and Genetics, 01307 Dresden, Germany. [3]Technical University Munich, Department of Biosciences, 85354 Freising, Germany.

*Author for correspondence (barbara_treutlein@eva.mpg.de)

B.T., 0000-0002-3299-5597

DEVELOPMENT

Numerous variations in enzymes, dissociation times, the method of titration, and so on should be tried in order to optimize the dissociation step. Additional consideration must be given to the possible enrichment or exclusion of certain cell types in some dissociation protocols, as well as the fact that the very nature of the protocol might alter the transcriptome of a cell. In the brain, for example, the differential fragility of cell types upon tissue dissociation to single cells means that the collection and profiling can be biased towards the more robust cells, such as astrocytes and neural progenitors. This is also a problem when using scRNA-seq to determine the identity of the cells present, as fragile cells might express markers of cell death, or cells can become activated during tissue disruption. Finally, some cell types, for example pancreatic cells, might contain enzymes that impact cDNA synthesis from lysed single cells. Methods to sequence transcriptomes from single nuclei provide an attractive strategy to combat the problem of differential fragility and perturbation of gene expression during the RNA capture stage, as nuclei are relatively robust and can be easily purified by FACS (Habib et al., 2016).

With all of these methods available, which should biologists choose to study organoid development at the transcriptional level? Three key variables to consider are the number of cells to be sampled, the sequencing depth (that is, the number of reads or transcripts per cell), and the coverage across the transcriptome. To analyze allele-specific expression or alternative splicing, one should choose a method that gives whole-gene coverage, for example Smart-seq2. This increased coverage across the transciptome comes at a cost, however, because many sequencing reads are needed for the same gene, which limits cell throughput. Nonetheless, this method has a very high sensitivity and accuracy relative to other methods. High cell throughput is crucial in order to reconstruct organoid development for a single organoid and compare across organoids, induced pluripotent stem cell lines, conditions, or individuals. From a sequencing perspective, cell throughput can be enhanced by sequencing only the end of the transcript (Jaitin et al., 2014). This will of course be dependent on the types of cells and degree of heterogeneity within an organoid. For plate-sorting,

manual labor can be reduced by employing liquid-handling robotics. The droplet-microfluidics approaches are striking improvements over previous technologies, dramatically increasing cell throughput through decreasing cost-per-cell and manual labor (Klein et al., 2015; Macosko et al., 2015). There are other high-throughput approaches on the horizon based on combinatorial indexing of cellular and nuclei RNA that are likely to compete with droplet-based methods (Rosenberg et al., 2017 preprint; Cao et al., 2017 preprint). One should note that high-throughput approaches often detect fewer genes per cell than high-coverage strategies, and might therefore miss any heterogeneity that is defined by only a few genes or by genes expressed at a low level. To circumvent this limitation, it might be possible to aggregate gene expression across all cells in a cluster and generate mean transcriptomes to reach the sensitivity of high-coverage technologies.

Defining the cellular composition of organoids

The power of single-cell transcriptomics lies in its ability to identify molecularly distinct cell types within a complex tissue without any prior purification of cell types. This depends to a large extent on the computational approaches that are used to identify different cell populations. Principal component analysis (PCA) is widely used to identify genes that vary across the sampled single cells. Cell relationships can then be visualized in two-dimensional space based on the expression of these genes, for example using tSNE or force-directed graphs, and clustered using a variety of different algorithms, such as BackSPIN, K-means and others (Kumar et al., 2017; Kolodziejczyk et al., 2015). A major challenge in deconstructing cell composition is to understand the source of heterogeneity in the dataset. Technical noise, most prominently 'dropout' due to inefficient capturing of mRNA and batch effects, can confound the analysis and obscure true biological noise. There are strategies to combat technical noise and remove confounding variables such as cell cycle state (Stegle et al., 2015). In addition, it is often unclear what constitutes a cell type and how to define discrete cell types versus continuous cell states. As a result, there is no definitive resolution on where to draw the line between different cell states. It is essential to iteratively explore the data using multiple approaches to fully understand the sources of heterogeneity within a dataset. Also, in order to identify rare cell types, many thousands of cells must be analyzed, or rare cell types need to be enriched using FACS, density gradient centrifugation or other approaches. As noted above, these methods of dissociation might enrich or exclude certain cell types from the suspension; therefore, it is generally unclear whether scRNA-seq can be trusted to quantify cell type abundance.

Reconstructing lineage relationships within an organoid

Organoids contain immature and mature cells on the same differentiation lineage. scRNA-seq provides a snapshot of the transcriptome states present in a tissue at any given moment. In a recent study of human fetal and organoid cortex, scRNA-seq detected the presence of intermediate cells that could be aligned to reconstruct a lineage path from progenitor to neuron (Fig. 1) (Camp et al., 2015). Therefore, a great power of this technology is to infer lineage trajectories through the presence of intermediate states. This has led to the generation of various computational models that try to align cells in a pseudotemporal order, and the expression of genes can be monitored as a function of pseudotime (Trapnell et al., 2014; Setty et al., 2016; Haghverdi et al., 2016). In addition, it is also possible to detect lineage bifurcations where a common progenitor gives rise to two or more differentiated cell types.
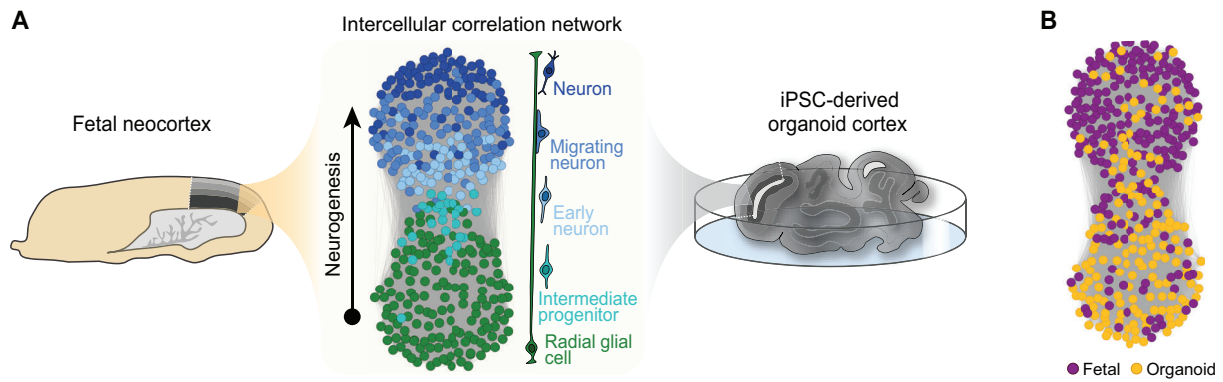
DEVELOPMENT

Fig. 1. Human organomics: reconstructing organoid development using scRNA-seq. (A) scRNA-seq comparisons between fetal human tissue and organoids bring insight into how well organoids recapitulate human development. As an example, scRNA-seq was used to analyze human fetal neocortex and cerebral organoids (Camp et al., 2015). An intercellular correlation network between fetal and organoid cells can reconstruct cell lineage relationships between radial glial progenitors, intermediates, and neurons. The network is colored based on the cell type. (B) Fetal and organoid cells are highly correlated and intermix in the network. The network is colored based on whether the cell is derived from the organoid or fetal cortex. iPSC, induced pluripotent stem cell.

Challenges and opportunities in human organomics

Lineage-coupled transcriptomics

Current methods for understanding lineage decisions using scRNA-seq do not provide direct evidence of lineage relationships, as lineages are reconstructed through overlapping patterns of gene expression in cells at intermediate stages. There are multiple methodological advances that could enable the tracing of direct lineage relationships using single-cell transcriptomics in organoids. One strategy is to use viral libraries to infect cells, expressing in each cell a reporter with a unique barcode (Gerlach et al., 2013). Single-cell transcriptome measurements from these cells would include the lineage-defining barcode as well as the transcriptome. Another strategy is to induce targeted DNA mutations using CRISPR/Cas9 to create unique patterns of insertions and deletions, known as scars, which would serve as cell-specific markers that are transmitted to daughter cells (McKenna et al., 2016). Lineage and transcriptome information can be read either by sequencing DNA and RNA from the same cell, or by creating the mutations in an expressed gene, for example an integrated fluorescent reporter, and only sequencing the transcriptome (Junker et al., 2017 preprint).

Spatial transcriptomics

Current widely used scRNA-seq methods require dissociation of tissue into single-cell suspensions to compartmentalize cells for barcoding. However, this approach loses spatial resolution. Recent reviews have explained how multiple labs have demonstrated transcriptome-scale measurements of RNA abundance within tissues (Tanay and Regev, 2017; Crosetto et al., 2015). One approach for spatially resolved single-cell transcriptomics is to combine scRNA-seq data with reference maps generated by traditional *in situ* hybridization. Another approach uses sequential rounds of smFISH with sophisticated combinatorial fluorescent barcoding to quantify the expression of tens to hundreds of genes *in situ* (Shah et al., 2016). An additional strategy is to position histological sections on arrayed reverse transcription primers with unique positional barcodes, thus generating RNA-sequencing data with two-dimensional positional information maintained (Stahl et al., 2016). These approaches reveal the location of distinct cell types, and offer powerful advantages over dissociation-based methods. Because organoids are relatively heterogeneous and can lack a stereotyped developmental axis, localizing transcriptomes within tissue architecture will be very useful. The technical difficulties and

expertise required for spatial transcriptomics have hampered their wide adoption; however, this could change over the coming years.

Beyond cell atlases: mechanisms of human development and disease

A major goal in human single-cell transcriptomics is to generate a comprehensive and quantitative reference 'atlas' of every human cell type in both adult and fetal tissues. A human cell atlas will provide a reference for comparison with organoids to understand how cell composition and gene networks are recapitulated in these *in vitro* models. In addition, fetal human cell atlases could provide strategies to reverse engineer human tissues by identifying transcription factors specific to certain cell types, and bring insight into cell-cell communication through the identification of putative receptor-ligand pairs. Furthermore, comparisons between diseased human tissue and the atlas will help elucidate disease mechanisms at cellular resolution, and might even discover 'new' human cell types. However, surveys such as these describe phenomena, whereas human organoids offer the possibility of uncovering regulatory mechanisms through the exploration and perturbation of developmental processes within a controlled environment. As such, high-information content measurements and sophisticated computational approaches position scRNA-seq as a powerful instrument to infer developmental mechanisms, which can then be tested. Comparisons between organoids generated from healthy and diseased individuals using high-throughput, spatial, and lineage-coupled transcriptomics will be able to localize network aberrations and identify disregulated genes. Recently, high-throughput scRNA-seq has been coupled with CRISPR/Cas9 mutagenesis to explore network robustness (Jaitin et al., 2016; Dixit et al., 2016; Adamson et al., 2016). We believe that this combination of technologies will provide enormous insight into the regulation of cell differentiation, cell communication, tissue organization, and response to environmental variables during human development.

Conclusions

Human organoids are manipulable, genetically and otherwise, a feature once reserved for classical model systems such as yeast, worms, flies, fish and mice. As a technology, however, *in vitro* organogenesis is still in its infancy, and in many cases it is unclear exactly what cell types are present within organoids and whether each cell type can be created in a reproducible manner. scRNA-seq will help to address this uncertainty, providing a greater depth of

analysis of cell heterogeneity and reproducibility. As protocols continue to evolve, human organoids are likely to come even closer to recapitulating bona fide human organogenesis in a predictable and reproducible way, making organoids a highly relevant system for understanding human development. We feel that quantitative single-cell transcriptomic approaches will provide impressive resolution of cell composition, lineage relationships, and gene network function within developing organoids, and, together with other genomic approaches, will offer unprecedented insight into the mechanisms that underpin human organogenesis. Methods to analyze DNA, methylation, chromatin accessibility, non-messenger RNAs and proteins in single cells will further advance the field. The cost-per-cell of many single-cell approaches is rapidly reducing and new methods are emerging that are relatively simple to implement in the lab. Hence, we believe that these technologies applied to human organoids represent a new direction in human developmental biology, and will help pave the way towards a better appreciation of what makes us uniquely human.

Competing interests

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.e1821.

Camp, J. G., Badsha, F., Florio, M., Kanton, S., Gerber, T., Wilsch-Bräuninger, M., Lewitus, E., Sykes, A., Hevers, W., Lancaster, M. et al. (2015). Human cerebral organoids recapitulate gene expression programs of fetal neocortex development. *Proc. Natl. Acad. Sci. USA* **112**, 15672-15677.

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 by combinatorial indexing. *bioRxiv*. doi:10.1101/104844.

Crosetto, N., Bienko, M. and Van Oudenaarden, A. (2015). Spatially resolved transcriptomics and beyond. *Nat. Rev. Genet.* **16**, 57-66.

Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C. P., Jerby-Arnon, 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.e1817.

Gerlach, C., Rohr, J. C., Perie, L., Van Rooij, N., Van Heijst, J. W. J., Velds, A., Urbanus, J., Naik, S. H., Jacobs, H., Beltman, J. B. et al. (2013). Heterogeneous differentiation patterns of individual CD8+ T cells. *Science* **340**, 635-639.

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

Haghverdi, L., Buttner, M., Wolf, F. A., Buettner, F. and Theis, F. J. (2016). Diffusion pseudotime robustly reconstructs lineage branching. *Nat. Methods* **13**, 845-848.

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.

Junker, J. P., Spanjaard, B., Peterson-Maduro, J., Alemany, A., Hu, B., Florescu, M. and Van Oudenaarden, A. (2017). Massively parallel clonal analysis using CRISPR/Cas9 induced genetic scars. *bioRxiv*. doi:10.1101/056499.

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.

Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C. and Teichmann, S. A. (2015). The technology and biology of single-cell RNA sequencing. *Mol. Cell* **58**, 610-620.

Kumar, P., Tan, Y. and Cahan, P. (2017). Understanding development and stem cells using single cell-based analyses of gene expression. *Development* **144**, 17-32.

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.

McCauley, H. A. and Wells, J. M. (2017). Pluripotent stem cell-derived organoids: using principles of developmental biology to grow human tissues in a dish. *Development* **144**, 958-962.

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.

Picelli, S., Björklund, Å. K., Faridani, O. R., Sagasser, S., Winberg, G. and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10**, 1096-1098.

Rosenberg, A. B., Roco, C., Muscat, R. A., Kuchina, A., Mukherjee, S., Chen, W., Peeler, D. J., Yao, Z., Tasic, B., Sellers, D. L. et al. (2017). Scaling single cell transcriptomics through split pool barcoding. *bioRxiv*. doi:10.1101/105163.

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. *Nat. Biotechnol.* **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.

Stahl, P. L., Salmen, F., Vickovic, S., Lundmark, A., Navarro, J. F., Magnusson, J., Giacomello, S., Asp, M., Westholm, J. O., Huss, M. et al. (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78-82.

Stegle, O., Teichmann, S. A. and Marioni, J. C. (2015). Computational and analytical challenges in single-cell transcriptomics. *Nat. Rev. Genet.* **16**, 133-145.

Tanay, A. and Regev, A. (2017). Scaling single-cell genomics from phenomenology to mechanism. *Nature* **541**, 331-338.

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. *Nat. Biotechnol.* **32**, 381-386.

DEVELOPMENT