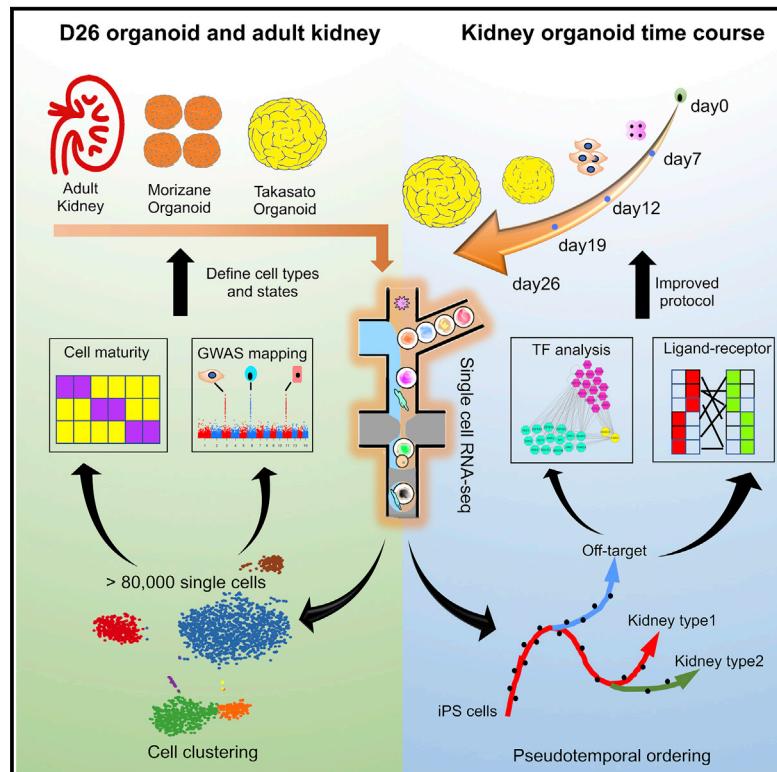


Comparative Analysis and Refinement of Human PSC-Derived Kidney Organoid Differentiation with Single-Cell Transcriptomics

Graphical Abstract



Authors

Haojia Wu, Kohei Uchimura,
Erinn L. Donnelly, Yuhei Kirita,
Samantha A. Morris,
Benjamin D. Humphreys

Correspondence

humphreysbd@wustl.edu

In Brief

A strong characterization of cell types, lineages, and differentiation states present in human PSC-derived kidney organoids is critical to improve differentiation protocols. Wu et al. use single-cell transcriptomics to reveal non-renal cell types, describe lineage-specific expression of regulatory genes, and report a broadly applicable strategy to reduce off-target cell populations.

Highlights

- Two human kidney organoid protocols were compared by single-cell transcriptomics
- Both protocols generated 10%–20% non-renal cells
- Kidney organoid cells are immature compared with fetal and adult human kidney
- Inhibiting BDNF-NTRK2 signaling reduces off-target cell types by 90%



Comparative Analysis and Refinement of Human PSC-Derived Kidney Organoid Differentiation with Single-Cell Transcriptomics

Haojia Wu,^{1,4} Kohei Uchimura,^{1,4} Erinn L. Donnelly,¹ Yuhei Kира,¹ Samantha A. Morris,^{2,3} and Benjamin D. Humphreys^{1,2,5,*}

¹Division of Nephrology, Department of Medicine, Washington University School of Medicine, St. Louis, MO, USA

²Department of Developmental Biology, Washington University School of Medicine, St. Louis, MO, USA

³Department of Genetics, Washington University School of Medicine, St. Louis, MO, USA

⁴These authors contributed equally

⁵Lead Contact

*Correspondence: humphreysbd@wustl.edu

<https://doi.org/10.1016/j.stem.2018.10.010>

SUMMARY

Kidney organoids derived from human pluripotent stem cells have great utility for investigating organogenesis and disease mechanisms and, potentially, as a replacement tissue source, but how closely organoids derived from current protocols replicate adult human kidney is undefined. We compared two directed differentiation protocols by single-cell transcriptomics of 83,130 cells from 65 organoids with single-cell transcriptomes of fetal and adult kidney cells. Both protocols generate a diverse range of kidney cells with differing ratios, but organoid-derived cell types are immature, and 10%–20% of cells are non-renal. Reconstructing lineage relationships by pseudotemporal ordering identified ligands, receptors, and transcription factor networks associated with fate decisions. Brain-derived neurotrophic factor (BDNF) and its cognate receptor NTRK2 were expressed in the neuronal lineage during organoid differentiation. Inhibiting this pathway improved organoid formation by reducing neurons by 90% without affecting kidney differentiation, highlighting the power of single-cell technologies to characterize and improve organoid differentiation.

INTRODUCTION

Chronic kidney disease affects 26–30 million adults in the United States, and 11% of individuals with stage 3 chronic kidney disease (CKD) will eventually progress to end-stage renal disease (ESRD), requiring dialysis or kidney transplantation (Coresh et al., 2007). In 2015, 18,805 kidney transplants were performed in the United States, but 83,978 patients were left waiting for a transplant because of a shortage of organs (United States Renal Data System, 2017). New treatments to slow progression of kidney disease are desperately needed, but progress has been slow, in part because the kidney is a complex organ but also because the relevance of rodent kidney

models to human kidney disease is debated (de Caestecker et al., 2015).

In this context, the emergence of methods to direct the differentiation of human pluripotent stem cells (PSCs) to kidney organoids has been received with great excitement (Lam et al., 2014; Morizane and Bonventre, 2017; Taguchi and Nishinakamura, 2017; Takasato et al., 2015; Xia et al., 2013). Over the last 4 years, several groups have published stepwise protocols, all based on kidney development during embryogenesis, resulting in generation of kidney tissue *in vitro* (Morizane and Bonventre, 2017; Taguchi and Nishinakamura, 2017; Takasato et al., 2016; Xia et al., 2014). These protocols modulate the activity of several signaling pathways, principally Wnt and Fgf, to generate renal progenitor populations that ultimately self-organize. Mature organoids contain up to hundreds of nephron structures, including glomeruli, properly segmented tubules, and interstitial cell types.

The ability to grow kidney organoids from patient-derived tissue offers unprecedented opportunities for the investigation of human kidney development, homeostasis, and disease. For example, kidney organoids have been used to successfully model and screen for modifiers of autosomal dominant polycystic kidney disease (Czerniecki et al., 2018; Freedman et al., 2015), acute kidney injury (Morizane et al., 2015), and vascularization of the glomerular tuft (Sharmin et al., 2016). A long-term goal is to generate transplantable kidneys grown in the laboratory, although many challenges remain. Bulk RNA sequencing has suggested that kidney organoids are most similar to first-trimester kidney (Takasato et al., 2015), and a recent marker analysis indicates that organoid nephrons are in the late capillary loop stage (Przepiorksi et al., 2018), so improving organoid maturation is one such challenge. No comprehensive analysis of exactly which cells kidney are generated by these protocols, their degree of maturation with respect to adults, and the extent to which off-target cells contaminate organoids has been undertaken to date. This information is a prerequisite for optimizing differentiation protocols to ultimately leverage kidney organoids for investigation of the most common adult kidney diseases, such as CKD, diabetic nephropathy, and acute kidney injury.

Here we used single-cell RNA sequencing (scRNA-seq) and single-nucleus RNA-seq (snRNA-seq) to generate comprehensive molecular maps describing kidney organoid cell diversity in two separate, commonly employed differentiation protocols and two separate pluripotent cell lines as well as in adult human



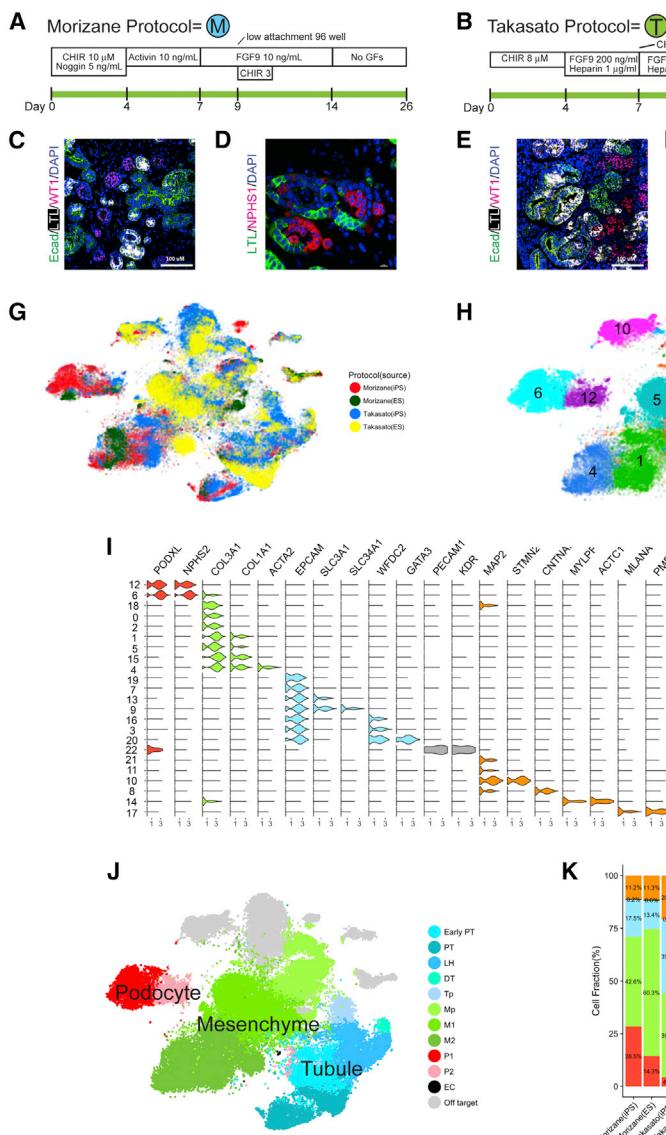


Figure 1. Comprehensive Single-Cell RNA Sequencing Demonstrates Development of a Spectrum of Cell Types in Kidney Organoids

(A and B) Diagram of Morizane (A) and Takasato (B) human iPSC directed differentiation protocols.

(C–F) Immunofluorescence analysis of day 26 organoids for proximal tubules (lotus tetragonolobus lectin [LTL]), distal tubules (ECAD), and podocytes (WT1 and NPHS1) from the Morizane protocol (C and D) and Takasato protocol (E and F). Scale bars, 50 μ m.

(G) tSNE projection of all day 26 organoid cells according to protocol (Morizane or Takasato) and cell line (induced PSCs [iPSCs] or hESCs).

(H) Unsupervised clustering of all organoid cells reveals 23 separate clusters.

(I) Violin plot showing cluster-specific expression of marker genes.

(J) Major kidney cell populations depicted after semi-supervised analysis.

(K) Proportions of kidney and off-target cell types according to protocol and cell source.

kidney. Our analysis reveals new insights, including the following: both protocols generate at least 12 separate kidney cell types; off-target non-renal cell types are present in all kidney organoids at similar ratios in human induced PSCs (iPSCs) versus human embryonic stem cells (hESCs); lineage relationships were revealed through pseudotemporal ordering during kidney organoid differentiation; kidney organoid cell types are immature when benchmarked against fetal and adult human single-cell datasets; and brain derived neurotrophic factor (BDNF) inhibition reduces off-target neuronal populations by 90% without altering kidney differentiation. These datasets provide a framework for evaluating and improving organoid differentiation protocols using single-cell transcriptomics.

RESULTS

scRNA-seq Defines Cell Diversity in Kidney Organoids

We used the hESC line H9 and the iPSC line BJFF.6; the latter was created from newborn male foreskin fibroblasts and reprog-

ammed with Sendai virus. We confirmed that the BJFF.6 line could efficiently generate kidney organoids using the protocol described by Takasato et al. (2015, 2016) and the protocol described by Morizane and Bonventre (2017) and Morizane et al. (2015) (Figures 1A and 1B; hereafter referred to as the Takasato or Morizane protocol, respectively). Each protocol generated nephron-like structures that closely resembled published reports (Figures 1C–F).

Using DropSeq, we isolated and sequenced mRNA from a total of 71,390 cells harvested from day 26 organoids. Organoids were generated using both hESCs and iPSCs. The Takasato protocol generated larger organoids, so we sequenced one or two each from separate batches. For the smaller Morizane protocol organoids, we combined 12 organoids each from separate batches. We detected \sim 1,930 unique transcripts from \sim 1,115 genes for each cell (Table S1). After correcting for batch effects by matching mutual nearest neighbors (Haghverdi et al., 2018), we reduced dimensionality by running a principal-component analysis (PCA) on the most highly variable genes, performed graph-based clustering on the significant principal components (PCs), and finally visualized distinct cell sub-groups using t-distributed stochastic neighbor embedding (tSNE). To examine protocol-dependent effects as well as differences between hESC- and iPSC-derived organoids, we projected cells according to protocol and cell source. This revealed co-clustering of cells predominantly based on the protocol used, with less difference attributable to hESC or iPSC source (Figure 1G).

Unsupervised clustering of the entire pooled dataset identified 23 transcriptionally distinct populations present in organoids generated from either the Morizane or the Takasato protocol

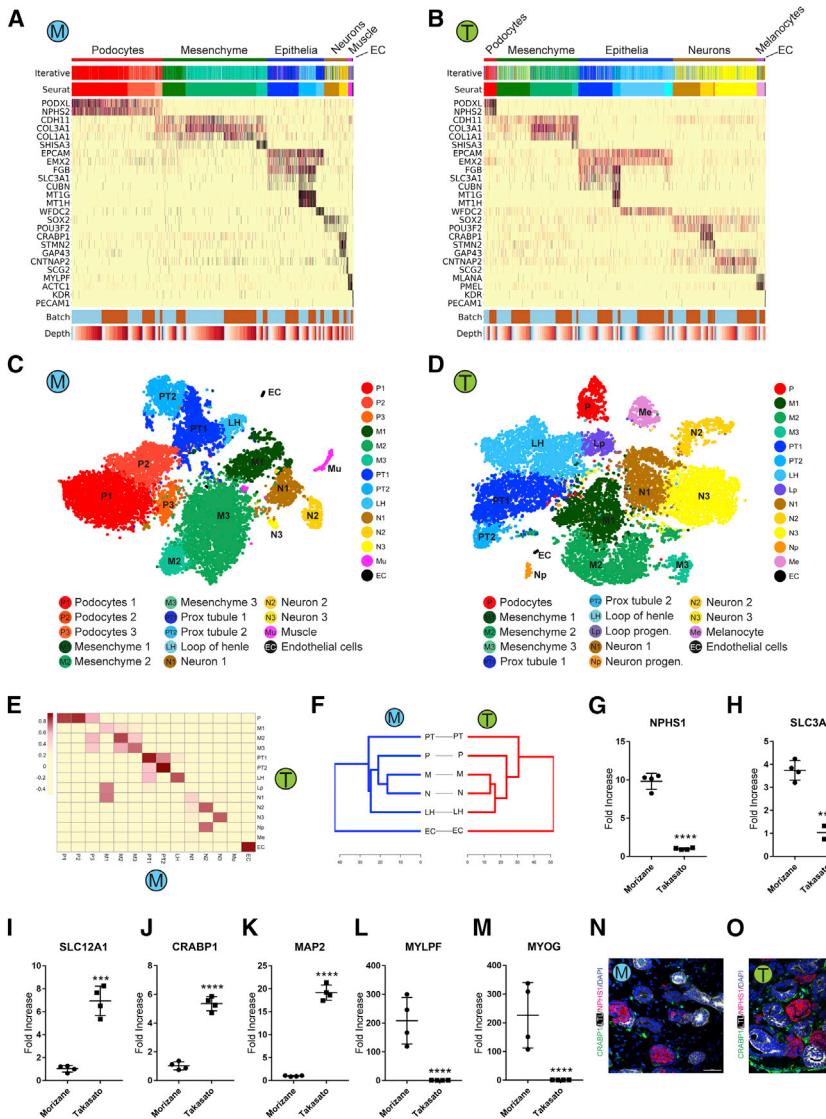


Figure 2. Comparison of Kidney Cell Types and Differentiation State in iPSC-Derived Kidney Organoids Generated with Both Protocols

(A and B) Heatmap of all cells clustered by recursive hierarchical clustering and Louvain-Jaccard clustering (Seurat), showing selected marker genes for every population of the Morizane protocol (A) and Takasato protocol (B). The bottom bars indicate the batch of origin (Batch) and number of unique molecular identifier (UMI) detected per cell (Depth).

(C and D) tSNE plot of cells based on the expression of highly variable genes for the day 26 organoids from the Morizane protocol (C) and Takasato protocol (D). The detected clusters are indicated by different colors.

(E) Heatmap indicating Pearson's correlations on the averaged profiles among common cell types for Morizane and Takasato organoids.

(F) Dendrogram showing relationships among the cell types in Morizane (left) and Takasato (right) organoids. The dendrogram was computed using hierarchical clustering with average linkage on the normalized expression value of the highly variable genes.

(G–M) qPCR comparing cell marker expression for podocytes (NPHS1) (G), proximal tubule (PT) (SLC3A1) (H), loop of henle (LOH) (SLC12A1) (I), neurons (CRABP1 and MAP2) (J and K), and muscle (MYLPF and MYOG) (L and M) between organoid protocols. *** $p < 0.001$ and **** $p < 0.0001$. Error bars indicate \pm SEM of fold change. (N and O) Immunofluorescence analysis of neural marker CRABP1 expression (green) in the Morizane (N) and Takasato (O) protocols. Cells were co-stained with PT (LTL, white) and podocyte (NPHS1, red) markers. Scale bars, 50 μ m.

and derived from either hESCs or iPSCs (Figure 1H). We annotated broad cluster classes by comparing unique transcript expression with existing RNA-seq datasets and the literature. In Figure 1I, violin plots show the expression of marker genes across these clusters. There were four broad classes of cell types in the pooled analysis: podocytes, mesenchyme, tubular epithelia, and off-target cells (Figure 1J). The fractions of these cell classes differed both according to protocol and according to cell source. For example, podocytes made up 28.5% of Morizane organoids derived from iPSCs but only 14.3% derived from hESCs. Off-target cell types, by contrast, were similar, at about 11% of both iPSC- or hESC-derived Morizane organoids, whereas they represented about 21% of both iPSC or hESC Takasato organoids (Figure 1K).

Variations in cell composition between hESC- and iPSC-derived organoids complicated efforts to reveal subtle distinctions between cells from the two protocols. We therefore analyzed organoids derived from iPSCs and hESCs separately to evaluate differences between the two protocols. In iPSC-derived organo-

ids, we analyzed 29,922 single-cell transcriptomes from two batches of the Morizane (15,951 cells) or Takasato (14,731 cells) organoids. To examine potential batch effects and to quantify variability among organoids, we projected cells from different batches of iPSC-derived organoids onto the same tSNE diagram, which showed that cells were intermixed regardless of batch (Figures S1A and S1B). Furthermore, cluster-based correlation analysis on both protocols revealed that the correlation for cells in the same cell cluster from different batches was always greater than the correlation for cells in the same batch from different cell clusters (Figures S1C and S1D). Proportions of cell clusters from different batches were also similar (Figures S1E and S1F). An alternative clustering approach, iterative hierarchical clustering (Baron et al., 2016), identified the same major organoid cell populations (Figures 2A and 2B).

Clustering of iPSC-derived organoids alone revealed a similar variability in cell frequency between Morizane and Takasato protocols, as observed in the global clustering analysis (Figures 2A and 2B). For example, Morizane organoids contained more podocytes, which were marked by the expression of PODXL and NPHS2 (Schwarz et al., 2001), whereas the Takasato protocol produced more tubular epithelial cells based on the

expression of EPCAM, SLC3A1, and WFDC2 (Figures 2A–2D; Litvinov et al., 1994). Despite this variability, both protocols generated very similar cell types because each pair of cell types was highly correlated (Figure 2E). Additionally, dendograms of analogous cell types from both cell types revealed very similar cell relationships (Figure 2F). We observed substantial numbers of non-renal cell types in both protocols. Morizane organoids contained three neuronal clusters and one muscle cluster. Organoids generated using the Takasato protocol contained four neuronal clusters and one cluster that we could not annotate but that expressed some melanocyte markers such as MLANA and PMEL (Kawakami et al., 1994). Similar findings were observed from a separate clustering analysis of hESC-derived organoids, except that muscle cells were common off-target cells present in both protocols (Figures S1G–S1M).

We confirmed differences in relative abundance of both renal and non-renal cell types by comparing marker gene expression for podocytes (NPHS1) and loop of Henle cells (SLC12A1) as well as muscle (MYLPF and MYOG) and neuronal (CRABP1 and MAP2) cells by qPCR (Figures 2G–2M). To localize neuronal cells, we performed immunostaining for CRABP1, a gene expressed in neuronal clusters from both protocols. CRABP1 protein expression localized to spindly cells present in the interstitium (Figures 2N and 2O). We could identify co-expression of MAP2, a microtubule-associated protein expressed exclusively in neurons (Dehmelt and Halpain, 2005), in many CRABP1+ interstitial cells (Figure S2A), further supporting a neuronal lineage. However, we could also detect coexpression of MEIS1, a marker of kidney stroma (Chang-Panesso et al., 2018), in some CRABP1+ cells (Figure S2B). Gene imputation analysis showed CRABP1 expression to be present in a small subset of mesenchymal cells from both protocols (Figures S2C and S2D). Expression of CRABP1 and MAP2 was low at earlier time points but rose substantially by day 26 (Figures S2E–S2I). Re-analysis of the bulk RNA-seq data in Takasato et al. (2015) confirmed the presence of many neuronal genes identified by our analysis (Figure S2J), suggesting that some degree of differentiation toward neural fates is a common outcome of current organoid differentiation protocols. Because CRABP1 was recently identified as a marker of stromal progenitors in embryonic day 14.5 (E14.5) kidney (Magella et al., 2018), we tested whether CRABP1 expression might mark a stromal progenitor cluster. The correlation between our neural clusters and E14.5 mouse stroma was very poor, however, and very few of the top 50 stromal progenitor genes were coexpressed in the CRABP1 cluster (Figures S2K and S2L). These analyses confirm the predominant neural identity of CRABP1-expressing cells in the kidney organoid.

Cell-Cycle Analysis

During kidney development, progenitor cell populations are characterized by rapid cell cycle progression that slows progressively with differentiation (Short et al., 2014). We analyzed cell cycle status as a proxy for the degree of differentiation (Kowalczyk et al., 2015). We scored all cells from both protocols based on cell cycle gene expression and assigned a cell cycle phase (G2M, S, or G1). The total fraction of cells in G2M, S, or G1 was similar between the protocols and cell source, at 39.1% and 43.9% in the Morizane and Takasato organoids derived from iPSC-derived organoids

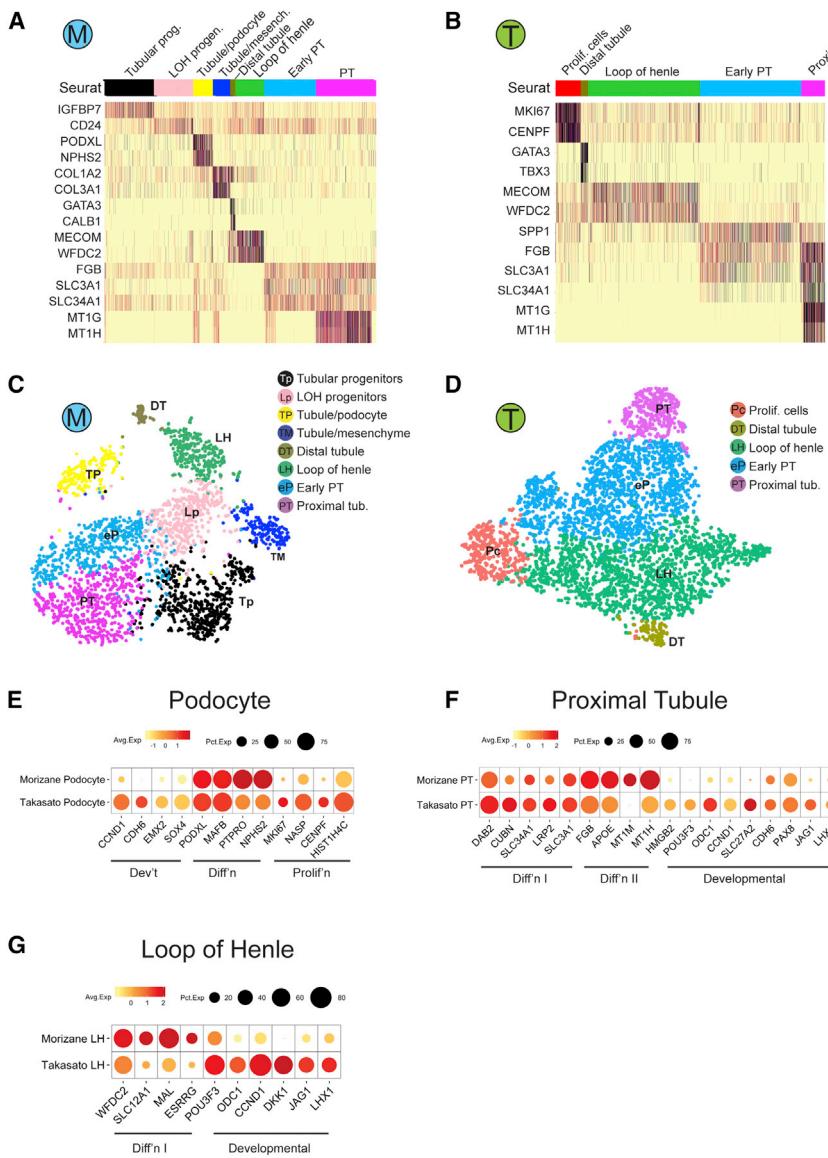
(Figures S3A and S3B) and 37.7% and 45.1% in hESC-derived organoids, respectively (Figures S3C and S3D). However, in the Morizane organoids, cells in G2M were limited to two cell clusters: a mesenchymal and neuronal cluster. By contrast, in Takasato organoids, G2M-phase cells were present as a subset of 6 separate clusters (two mesenchymal, two neuron-like, one unidentified, and an epithelial cluster). Cell cycle gene expression was not driving cluster identity. Both the Morizane M1 cluster and the Takasato Lp cluster expressed high levels of the cell cycle gene MKI67 but showed divergent expression of COL3A1 (M1) and POU3F3 (Lp) (Figure S3F).

We interpreted the broader proliferative distribution of the Takasato organoid to potentially reflect that the organoid had been harvested before it was fully differentiated. Indeed, cell cycle analysis of cells collected at different time points using the Takasato protocol revealed that the proportion of cycling cells decreased along the kidney organoid differentiation process (Figure S3E), suggesting that the degree of cell differentiation or cell type maturity might be negatively correlated with the proliferative capacity.

Kidney Organoid Cell Subsets

Re-clustering of tubular cells identified additional cell clusters in both protocols. We detected 8 and 5 tubular subtypes in Morizane and Takasato organoids, respectively (Figures 3A–3D). This includes a subpopulation that expressed the ureteric bud marker GATA3 (Labastie et al., 1995) in both protocols. Prior reports have suggested that kidney organoids contain derivatives of both major progenitor populations, the metanephric mesenchyme and the ureteric bud (Takasato et al., 2015). However, several lines of evidence suggest that the GATA3 cluster is actually a metanephric mesenchyme-derived distal tubule. First, the GATA3 cell cluster did not express mature collecting duct markers (e.g., AQP2 or AQP4), although this could also be explained by immaturity. Second, in addition to principal cells, GATA3 is also expressed in the distal convoluted tubule and connecting segment in both human and mouse kidney (Figures S4A and S4B). Third, the Morizane organoid GATA3 cluster also expressed calbindin, a marker of distal tubules (Bindels et al., 1991). We verified that, in post-natal day 1 (P1) mouse kidney scRNA-seq, calbindin mRNA is expressed in the distal tubule and the ureteric bud tip, whereas, in our adult kidney snRNA-seq data, calbindin mRNA was exclusively expressed in the distal convoluted tubule and the connecting segment but not in principal cells (Figures S4A and S4B). Finally, comparison of the GATA3 cluster with adult kidney cell types shows that it is equally or more similar to the distal tubule and connecting segment than to principal cells by Pearson's correlation (Figure 4C). These findings raise significant doubts that the ureteric bud and its derivatives are generated using either the Morizane or the Takasato protocol.

We next compared a panel of developmental and differentiation genes in podocyte, proximal tubule, and loop of Henle cell clusters across protocols. This revealed higher expression of the kidney developmental markers CDH6, EMX2, and SOX4 in Takasato podocytes (Brunskill et al., 2008). Morizane podocytes had somewhat higher expression of podocyte differentiation markers and lower expression of proliferation markers (Figure 3E). For the proximal tubule, both Morizane and Takasato



organoids had similar expression of differentiation markers, but the Takasato proximal tubule had higher expression of developmental markers. The Morizane proximal tubule had higher expression of genes that were difficult to interpret, including the metal-binding genes MT1M and MT1H (Figure 3F). The loop of Henle was more differentiated in Morizane organoids (Figure 3G).

Quantifying Organoid Kidney Cell Maturity

A critical question is the degree to which kidney organoid cell types resemble their native counterparts in molecular terms. We addressed this question in three ways. First, we compared organoid cell type gene signatures with a recent mouse P1 scRNA-seq dataset (Adam et al., 2017). Again, there was no clear ureteric bud or collecting duct population identified in the organoid datasets (Figures S4C and S4D). For both organoid protocols, the M1 mesenchymal clusters showed medium correlation to cap mesenchyme in addition to stroma. Notably, none

Figure 3. Human Kidney Organoids Contain Subclasses of Tubular Epithelial Cells

(A and B) Heatmap showing selected marker genes for every tubular subpopulation of the Morizane protocol (A) and Takasato protocol (B) generated from iPSCs. (C and D) tSNE plot of tubular subclusters in kidney organoids from the Morizane protocol (C) and Takasato protocol (D). The detected clusters are indicated by different colors.

(E–G) Dotplot comparing the expression of cell type signature and developmental or proliferating genes on podocytes (E), proximal tubules (F), and LOH (G) between the two protocols.

of the off-target clusters (neural, muscle, and melanocyte-like) correlated to cell types found in P1 kidney.

We also compared organoid cell clusters with a recently generated single-cell dataset generated from human week 16 kidney (Lindström et al., 2018b). This analysis revealed excellent correlation of organoid kidney cell types to fetal kidney cell types, either by Pearson's correlation or using a multiclass random forest classifier (Habib et al., 2017; Figures S4E–S4I). Notably, the off-target cell clusters did not map to any human fetal kidney cell types, with the sole exception of the Takasato N1 cluster, which showed some correlation to a fetal kidney cluster annotated as cycling. N1 also expresses a strong cell cycle gene signature, likely explaining the correlation (Figure S4H).

Finally, we compared kidney organoid cell types with their adult human counterparts. Attempts of scRNA-seq failed; however, we were successful in generating adult human kidney snRNA-seq data from a 62-year-old white male

with a serum creatinine level of 1.03 mg/dL using the 10X Chromium platform. We sequenced 4,524 nuclei to a similar depth (Table S1) as the organoid datasets and identified 12 distinct epithelial cell clusters, including podocytes, proximal tubule cells (S1–S3), loop of Henle cells (descending and ascending), distal tubule cells, connecting segment cells, principal cells, and intercalated cells (type A and type B) (Figures 4A and 4B). The absence of stromal or leukocyte populations most likely reflects either dissociation bias and/or a cell frequency below our limit of detection (Wu et al., 2018).

scRNA-seq measures transcripts from both cytoplasm and nucleus whereas single nucleus RNA-seq measures only nuclear transcripts. Nuclei contain only a fraction of total cell RNA, and although nuclear and cytoplasmic mRNAs correlate highly (Barthelson et al., 2007), some protein-coding mRNAs are retained in the nucleus (Bahar Halpern et al., 2015). Despite these differences, single-cell and snRNA-seq datasets predict cell types comparably and with high concordance (Habib et al., 2017;

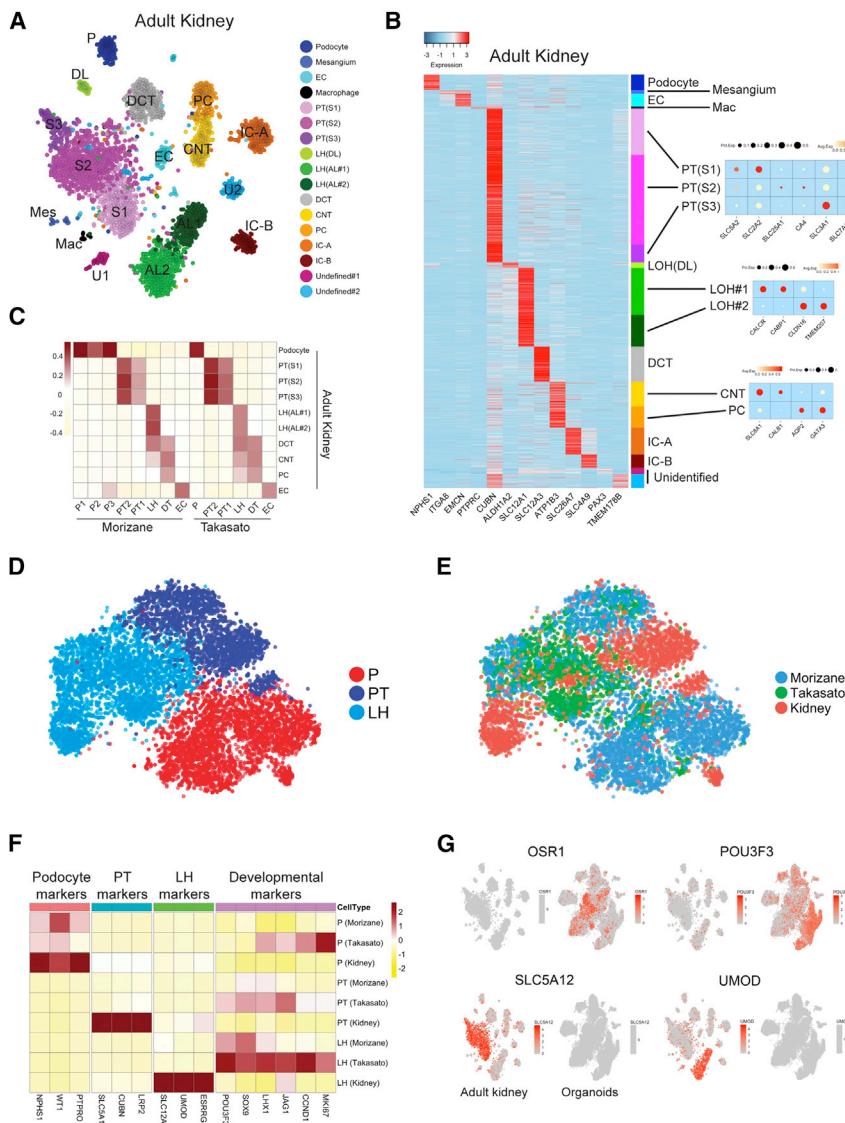


Figure 4. Organoid Cell Types Are Immature Compared with Benchmarked Adult Kidney Cell Types

(A) Unsupervised clustering of single-nucleus RNA sequencing (snRNA-seq) of adult human kidney identified 17 distinct cell types in human adult kidney. That includes 11 tubular cell types, podocytes, mesangium, endothelial cells, and macrophages.

(B) Heatmap showing uniquely expressed genes for each cluster.

(C) Pearson correlation analysis comparing the organoid cell types and their endogenous counterparts in human kidney. The color bar indicates the correlation score.

(D) Reclustering of podocytes, proximal tubule (S1 and S2), and loop of Henle cells derived from both organoids and adult kidney, analyzed using canonical component analysis.

(E) Cellular origins (Morizane, Takasato, or adult kidney) visualized in the tSNE reveal poor overlap between organoid-derived and adult-derived cells within each cluster.

(F) Comparison of the average expression of marker genes and developmental genes between organoid cell types and adult kidney cell types. Expression value was scaled by Z score.

(G) Expression of the developmental factors OSR1 and POU3F3 is strong in organoids but almost undetectable in adult kidney. Expression of the S1 marker SLC5A12 and loop of Henle marker UMOD is strong in adult kidney and undetectable in organoids.

Lake et al., 2017). Therefore, we next correlated all kidney organoid epithelial cell types with their corresponding endogenous counterparts from human adult kidney. We observed an expected correlation between corresponding cell types of organoid and human kidney (Figure 4C). Organoid loop of Henle correlated with adult loop of Henle but also distal tubule and collecting duct, likely reflecting their developmental immaturity.

Although organoid and adult kidney epithelial cell types correlated well, our prior analysis suggested that organoid-derived cells expressed developmental markers. To visualize overall similarities and differences in cellular transcriptomes from specific organoid and adult cell types, we performed unsupervised clustering of podocytes, proximal tubule, and loop of Henle clusters from both organoids and adult kidney after batch effects were corrected by canonical correlation analysis (Butler et al., 2018) and projected the data by tSNE (Figure 4D). As expected, the analysis revealed three separate clusters. When we then projected cellular origin onto the tSNE, however, there was relatively poor overlap between the organoid-derived versus adult kidney

cells within each cluster (Figure 4E). Further emphasizing the transcriptional differences between organoid-derived cells and adult kidney, expression of differentiation markers was much higher in adult cell types, whereas developmental marker expression was much higher in organoid cell types (Figure 4F). For some

of these developmental markers (OSR1 and POU3F3), expression was high in many organoid cell types in the pooled dataset but undetectable in adult kidney. In the same way, some differentiation markers (SLC5A12 and UMOD) were strongly expressed in adult clusters but undetectable in organoids (Figure 4G). These results indicate that organoid cell types are substantially immature compared with their adult counterparts.

We identified 123 receptors and 97 cognate ligands and mapped their expression to specific adult human kidney cells. Most of these mapped to a single predominant cell type (Figures S5A and S5B). This allowed for development of a simple connectome model for how mature kidney cells might intercommunicate during homeostasis (Figure S5C). Current protocols incubate organoids without any growth factors after 12–14 days, but the expression of so many ligands in mature kidney suggests a possible need to include soluble factors during the maintenance phase of organoid maturation.

Because transcription factors regulate cell state, we next tested the hypothesis that organoid cell immaturity might reflect

partial expression of the gene-regulatory network present in mature kidney cells. We identified 54 transcription factors present in adult human proximal tubules, over half of which have not been reported previously (Table S2). For example, high mobility group nucleosome-binding domain-containing protein 3 (HMGN3) is a thyroid hormone binding receptor that regulates gene expression and is strongly expressed in proximal tubules, and thyroid hormone is known to regulate renal fluid and electrolyte handling, suggesting that HMGN3 may mediate thyroid hormone actions in the proximal tubule (Michael et al., 1972). In human adult podocytes, we identified 38 transcription factors, 25 of which we believe have not been reported previously. We validated the expression of six of these transcription factors at the protein level (Figures S5D–S5I). Many of the proximal tubule transcription factors are expressed solely in proximal tubules, whereas podocyte transcription factors are more widely expressed across kidney cell types (Mendeley, <https://doi.org/10.1763/m4rfg9wb29.1>).

Both proximal tubule cells and podocytes derived from organoids expressed only a fraction of the transcription factors we identified in the adult cell types. For example, Takasato protocol proximal tubules expressed 11 of 54 adult proximal tubule transcription factors and Morizane proximal tubules only 9 of 54 (Table S2). Similarly, both Takasato and Morizane protocol podocytes expressed 7 of 38 adult podocyte transcription factors (Table S3). This result suggests that organoid cells, despite expressing some markers of differentiated cells, are fundamentally different from their terminally differentiated adult counterparts. Collectively, these results identify lineage-specific expression of genes that likely regulate cell specification, differentiation, and proliferation during kidney organoid maturation.

Disease-Related Genes Predicted by GWASs Are Expressed in Single Cell Types in Adult and Organoid Kidney

Human kidney organoids are already being used to model monogenic human kidney diseases. However, there are many more complex trait disease genes that have been identified by genome-wide association studies (GWASs). Recently, Park et al. (2018) reported that many human monogenic and complex trait genes are expressed predominantly in a single mouse kidney cell type. To gauge how useful kidney organoids might be for modeling disease-relevant genes, we next compared our ability to detect gene expression of GWAS hits in adult kidney versus organoids. We used established GWAS gene lists including 117 genes for chronic kidney diseases, 275 genes for hypertension, and 777 genes for plasma metabolite levels.

We could map expression of 207 of these genes to cell types in our adult kidney snRNA-seq dataset (Figures 5A–5C). Of these 207 mapped genes, we could only detect expression of 40 of them (19%) in the correct corresponding organoid cell types (Figures 5D–5F). In most cases, we confirmed that these GWAS genes are expressed in only a single kidney cell type (Figure 5). Unexpectedly, podocytes and mesangial cells expressed a substantial number of hypertension genes. These glomerular cell types are not widely believed to play important roles in regulating blood pressure. Consistent with their central role in secretion and reabsorption, proximal tubules had the highest number of genes associated with plasma metabolite levels.

As a complementary approach, we used PolyPoly, a regression-based polygenic model that allows prioritization of trait-relevant cell types by combining GWAS and single-cell expression datasets (Calderon et al., 2017). We focused on CKD and estimated glomerular filtration rate (eGFR) GWAS hits and asked whether trait-relevant cell types identified using the adult kidney dataset were similar to those identified using the organoid datasets. For CKD, PolyPoly identified principal cells, type A intercalated cells, podocytes, and proximal tubules as trait-relevant cell types. By contrast, only distal tubules (which might also represent principal cells) were identified using Morizane data and only podocytes using Takasato organoids. Similarly, for eGFR, trait-relevant cell types using adult kidney data included proximal tubules, podocytes, ascending loop of Henle cells, and principal cells. The Morizane dataset identified proximal tubules, and the Takasato dataset failed to identify any trait-relevant cell types (Figures S5J–S5O).

These results confirm and extend those of Park et al. (2018), which was performed in mouse and not human cells, but also suggest that kidney organoids are limited in their ability to predict trait-relevant cell types in comparison with adult kidney because many GWAS hits are not expressed in organoid cell types.

Lineage Reconstruction during Kidney Differentiation

To explore lineage relationships and the mechanisms of cell fate decisions during kidney organoid differentiation, we performed scRNA-seq at separate time points during differentiation using the Takasato protocol (days 0, 7, 12, 19, and 26). A total of 9,190 cells from all five time points were projected by tSNE, and on days 0, 7, and 12, each formed single distinct clusters (Figure 6A). Pluripotency gene expression (e.g., POU5F1 or OCT4) was completely downregulated by day 7, with upregulation of metanephric mesenchyme markers (SALL1, FGF18, and HOXB9; Figure 6B; Brunskill et al., 2008). The day 12 cluster most closely resembled the pretubular aggregate, with genes such as JAG1 and LHX1 strongly enriched at this time point. Multiple clusters corresponding to differentiating cell types were present on days 19 and 26, and most later clusters contained cells from both time points, reflecting asynchronous differentiation.

We compared our results with the bulk RNA-seq data from Takasato et al. (2015) by deconvolving cell frequency across time using a bulk sequence single-cell deconvolution analysis pipeline (Baron et al., 2016). This confirmed downregulation of the pretubular aggregate and posterior intermediate mesoderm and increasing fractions of differentiated cell populations with time (Figures S6A–S6H). Although differentiation marker expression generally increased with time in our dataset, and certain progenitor markers such as CITED1 decreased over time (Figure 6C), many genes marking developmental cell types persisted on day 26. Genes reflecting the renal vesicle (DKK1) and S-shaped body (JAG1, CCND1, CDH6, and LHX1) continued to be expressed, for example, suggesting an ongoing nephrogenic program (Figures 3E–3G; Brunskill et al., 2008). Future enhancements to kidney organoid differentiation protocols will need to push maturation of these developmental states toward fully mature kidney cell types.

To detect gene expression changes during organoid differentiation, we reconstructed kidney lineage relationships by

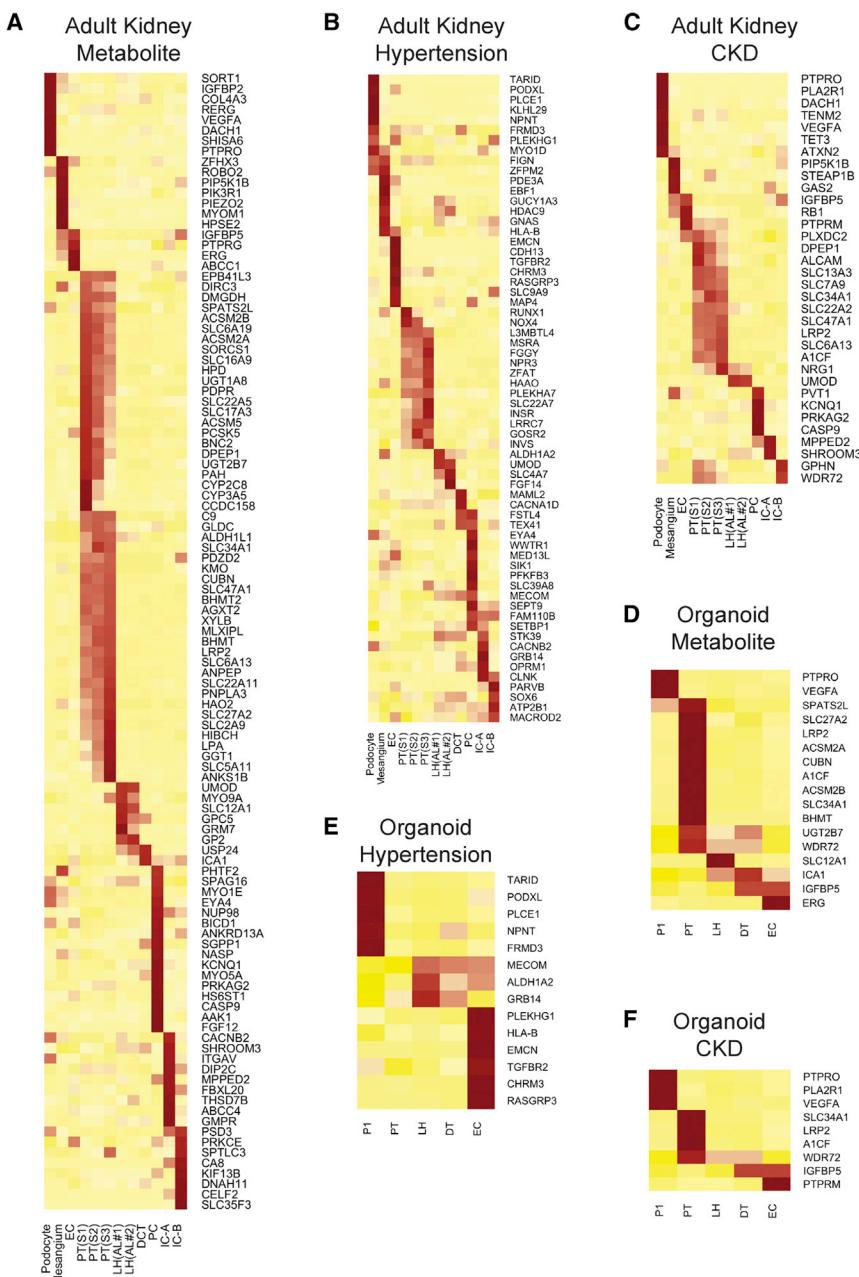


Figure 5. Cell-Specific Expression of Disease-Relevant Genes in Adult Kidney Compared with Organoids.

(A–C) Cell-specific expression of genes reported in plasma metabolite level-related GWAS (A), hypertension-related GWAS (B), and CKD-related GWAS (C) in adult kidney. Each gene reported in a kidney disease-related GWAS was assigned to the adult kidney cell type in which it was found to be differentially expressed (likelihood ratio test). A heatmap was used to visualize the Z score-normalized average gene expression of the candidate genes for each cell cluster.

(D–F) Disease-relevant genes identified in (A)–(C) for which cell-specific expression could also be detected in organoid cell types. (D) Plasma metabolite levels-related GWAS. (E) Hypertension-related GWAS. (F) CKD-related GWAS. Results from both protocols and both cell sources were pooled for the analysis.

transcription factors and signaling pathways whose modulation might improve kidney cell maturation and eliminate off-target cell types, we performed branched expression analysis modeling (BEAM). We identified a large number (56) of dynamically expressed transcription factors over the course of differentiation (Figure S6I). The analysis assigned expression of these genes to either of the main branches, but it could not resolve single-cell cluster expression. We therefore mapped expression of these transcription factors to all 12 major clusters (Figure S6J). Most genes were expressed in only one or a few cell types. We identified five genes (POU2F2, POU3F2, NHLH2, HES6, and LHX9) whose expression was limited to a neuronal cluster and confirmed that their expression corresponded to the neuronal branch by pseudotemporal ordering (Figure S6K). A subset of these transcription factors (TFs) has previously been implicated in neuronal

development. For example, loss of LHX9 prevents formation of the neocortex (Bulchand et al., 2001). Similarly, POU3F2 is required for survival of hypothalamic neural progenitors (Nakai et al., 1995). Thus, the induction of these TFs may be critical for one or more of the neuronal lineages present in kidney organoids. For finer mapping of gene expression changes during fate decisions, we performed BEAM on the podocyte and mesenchyme-neuron and on the mesenchyme and neuron branch (Mendeley, <https://doi.org/10.17632/m4rf9wb29.1>). This analysis included all differentially expressed genes as well as transcription factors alone.

To test whether longer organoid incubation times might improve the cell differentiation status, we grew organoids from the Takasato protocol out to 34 days and performed

performing pseudotemporal ordering using Monocle2 (Qiu et al., 2017). The resulting cell trajectories revealed one major branchpoint, separating loop of Henle and proximal tubule cell fates from podocyte, stromal, and neural cell fates (Figure 6E). A second branchpoint distinguished podocyte from stromal and neural fates. Cell fates were defined by projecting marker gene expression onto the pseudotime trajectories (Figure 6F).

Lineage-Specific Expression of Transcription Factors, Receptors, and Ligands during Organoid Differentiation

Although many steps in murine nephrogenesis are well characterized, the transcriptional pathways underlying human kidney development are less well characterized. To identify candidate

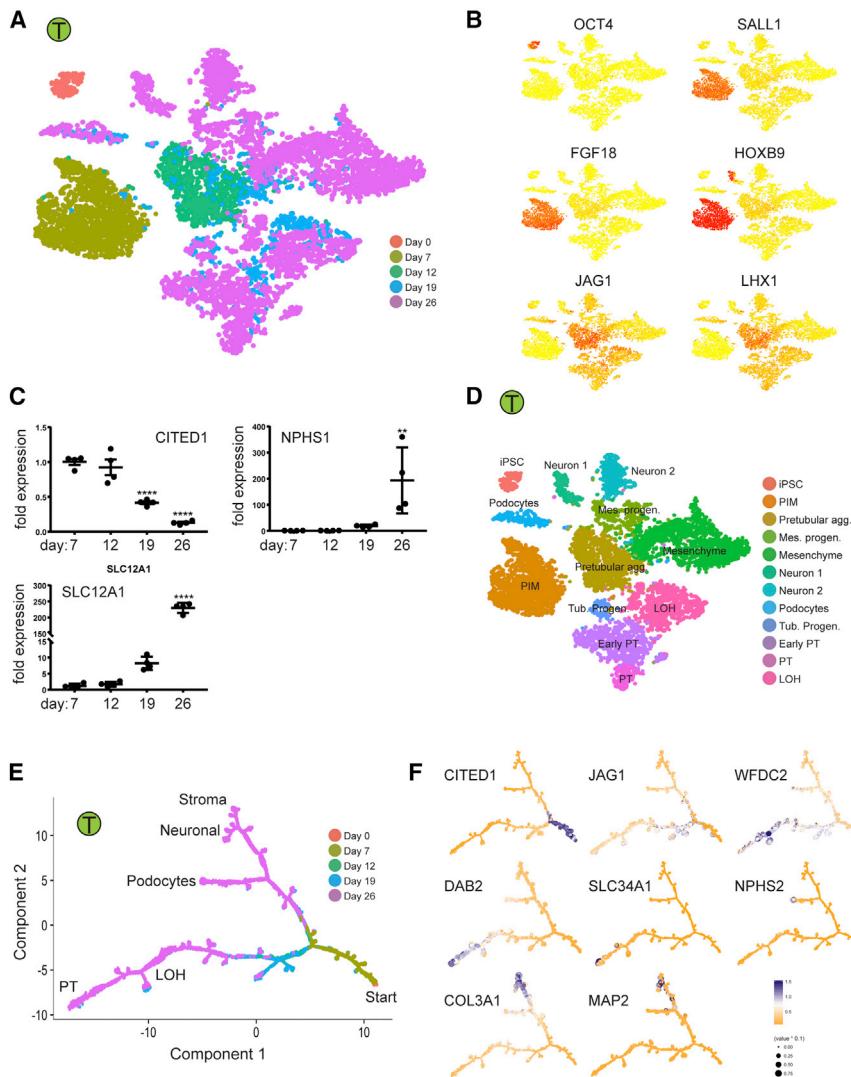


Figure 6. Time Course Analysis of Cells during Organoid Differentiation Reveals Lineage Relationships

(A and B) Projecting cells across time points to the tSNE. Cells were colored by the time point (A) or gene expression of stage-specific markers (B).

(C) Validation of the stage-specific marker by qPCR. ** $p < 0.01$ and *** $p < 0.001$ versus day 7. Error bars indicate \pm SEM of fold change.

(D) Annotation of cell clusters based on gene expression of cell-type-specific markers.

(E and F) Ordering of single-cell RNA sequencing (scRNA-seq) expression data according to the pseudotemporal position along the lineage (E) revealed a continuum of gene expression changes from iPSCs to differentiated cell types (F).

scRNA-seq on a total of 6,115 cells (Table S1). We compared the expression of differentiation markers across clusters and discovered that differentiation was generally worse, not better, at this later time point, with loss of endothelial cells, reduced expression of differentiation markers across most clusters, and the emergence of an off-target muscle cell cluster (Figure S7A). We next pooled the day 34 scRNA-seq results with our day 26 results, removed batch effects by matching nearest neighbors (Haghverdi et al., 2018), and reclustered. Although there was overlap with kidney cell clusters, the muscle cluster was unrelated to any day 26 cluster (Figures S7B and S7C). Furthermore, we could detect a separate new cluster, also specific only to day 34, that expressed a high percentage of mitochondrial genes, indicating that these were unhealthy cells (Figure S7D). Overall, there was a reduction in the fraction of mature kidney cell types (for example, mature stroma and proximal tubule) and a substantial increase in off-target cell types on day 34 compared with day 26 (Figure S7E).

Because organoid differentiation is accomplished by exposure of iPSCs to sequential combinations of extrinsic factors, we also searched for ligand and receptor pairs whose expression

changed in a lineage-specific fashion during organoid differentiation. We identified 19 receptors with 24 cognate ligands in this way and mapped their expression to the major organoid cell types (Figure 7A). NTRK2, which encodes neurotrophic tyrosine kinase receptor, type 2, was expressed exclusively in neural clusters N1 and N3. Its ligand, BDNF, was also strongly induced in the podocyte-neuron-stroma branch.

Inhibition of BDNF-NTRK2 Signaling Reduces Off-Target Cells

Because BDNF promotes neuron survival, growth, and differentiation (Huang and Reichardt, 2001), we reasoned that inhibition of signaling by its receptor might reduce off-target neuron populations in kidney organoids. To test this hypothesis,

we first selected a dose of the NTRK2 inhibitor K252a (Tapley et al., 1992) that did not alter the gross tubular morphology (Figure S7F). We administered 250 nM K252a beginning on day 12 of the Takasato protocol (Figure 7B). Preliminary qPCR data suggested a reduction in off-target marker expression (Figures S7G and S7H). We therefore performed scRNA-seq on K252a-treated organoids, which revealed a 90% reduction in neuronal cells, from 20%–22% to 2.1%. There was also a decrease in mesenchymal cells from 39.8% to 15.9%, consistent with pseudotemporal ordering that placed neurons and kidney mesenchyme in the same branch. We observed an increase in tubular cells from 35.2% to 70.4%, accompanied by an increase in podocytes from 4.2% to 11.5% (Figures 7C and 7D). All kidney lineages expressed the expected marker genes (Figure 7E). The reduction in neuronal cells was confirmed by immunofluorescence analysis of an independent batch (Figure 7F).

DISCUSSION

Human kidney and kidney organoids are composed of a wide array of cell types, all required for proper development and organ

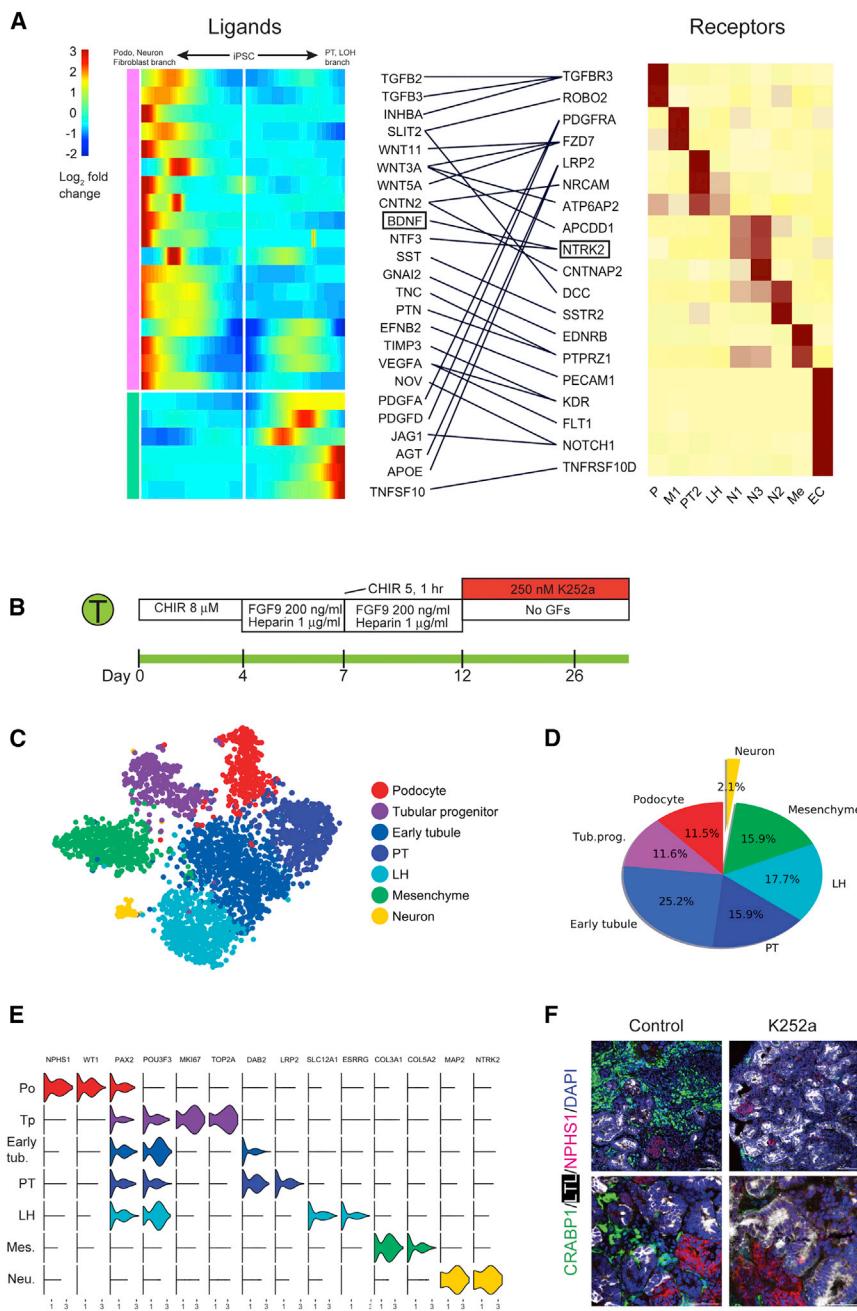


Figure 7. Reduction in Off-Target Cell Differentiation by Analysis of Cell-Specific Expression of Receptors and Ligands during Organoid Differentiation

(A) Heatmap showing the kinetics of branch-dependent ligand expression identified by BEAM (Monocle2) and corresponding cell-specific receptor expression in day 26 organoids from the Takasato protocol. The analysis identified that BDNF expression was induced in the podocyte-mesenchyme-neuron branch, and its receptor NTRK2 was exclusively expressed in neurons.

(B) Inhibition of the BDNF pathway using K252a (250 nM from days 12 to 26).

(C) tSNE of K252a-treated organoids showing a very small neuronal population.

(D) Off-target cells made up only 2.1% of the total cells in K252a-treated organoids.

(E) Violin plot showing marker gene expression across clusters in K252a-treated organoids.

(F) Verification of the strong reduction in neuronal cells by immunofluorescence staining of an independent organoid batch.

tion state. The information provided in this comprehensive dataset will guide future attempts to improve differentiation protocols.

These results will help guide protocol choice for investigators interested in modeling kidney function or disease. Based on increased expression of podocytes with the Morizane protocol, it is better suited for analysis of glomerular biology. Similarly, the Takasato protocol generated more tubular epithelium and is better suited for studying the tubulointerstitium. Unexpectedly, our analysis suggests the apparent absence of ureteric bud-derived cell types. The ureteric bud undergoes branching morphogenesis to form the collecting system and is required for the formation of an interconnected collecting duct (Costantini and Kopan, 2010). In a recent study, Taguchi and Nishinakamura (2017) induced mouse metanephric mesenchyme and ureteric bud progeni-

tors separately, and their recombination led to organoids with much more complex and interconnected collecting duct architecture than reported previously. Similar protocols for human PSC-derived kidney organoids have not yet been established. Our results suggest that establishing conditions that will support growth of the ureteric bud lineage in human kidney organoids is an immediate priority.

One measure of the usefulness of kidney organoids for modeling disease is the degree to which organoid cells express disease-relevant genes. We found that organoid cells expressed about 20% of trait-relevant genes defined by GWAS compared with their adult counterparts. Similarly, we could only detect about 20% of the transcription factors present in adult proximal

function. Recent studies have begun to characterize these cell types using scRNA-seq, revealing important new insights (Adam et al., 2017; Chen et al., 2017; Czerniecki et al., 2018; Der et al., 2017; Lindstrom et al., 2018a; Magella et al., 2018; Park et al., 2018). Fulfillment of the promise of human kidney organoids requires comprehensive characterization of their cell composition, comparison of differing protocols, and a better understanding of the degree to which they produce mature, differentiated kidney cell types. Using scRNA-seq, we established that current protocols generate a remarkable diversity of kidney cell types. We also provide the first direct comparison of separate differentiation protocols, revealing broadly similar outcomes but important differences in cell ratio and differentia-

tubule and podocytes compared with organoid equivalents. Consistent with these observations, a global comparison of organoid-derived cell types with human fetal and adult counterparts revealed that, although organoid-derived cells express some markers of terminal differentiation, they are all immature. Unexpectedly, longer organoid incubation did not improve differentiation but, rather, caused reduced expression of terminal markers and generated new off-target cells, suggesting kidney cell type dedifferentiation with time, consistent with a recent report (Przepiorski et al., 2018). These results indicate a need to identify conditions that will better support continued organoid maturation.

Off-target cell populations, primarily neurons, were present in kidney organoids generated from both protocols. By combining pseudotemporal ordering with lineage-specific expression of transcription factors, ligands, and receptors, our analysis provides a roadmap to understand lineage relationships and signaling during differentiation as well as a framework around which to test improvements to the differentiation protocol. The expression of BDNF and its receptor NTRK2 in neural clusters suggested a strategy to reduce these cell types by inhibiting BDNF-NTRK2 signaling. That this strategy reduced off-target cell types by 90% suggests that similar analyses could be applied broadly in the organoid field to reduce unwanted cell types.

We envision that analysis of signaling pathways and transcription factors expressed before and after branchpoints will suggest other potential strategies to regulate organoid cell fates. Future studies are necessary to determine how modulation of the gene expression patterns revealed here can be used to improve organoid maturation and, ultimately, better model physiological function.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - iPSC Culture
 - Kidney donor
- METHOD DETAILS
 - Kidney Organoid Differentiation
 - DropSeq scRNA-seq
 - Nuclei isolation and snRNA-seq of human kidney
 - Immunofluorescence
 - BDNF inhibitor (K252a) treatment
 - Real Time PCR Experiments
- QUANTIFICATION AND STATISTICAL ANALYSIS
 - Preprocessing of DropSeq data
 - Global clustering analysis
 - Clustering analysis on the organoid cells from different protocols or cell lines
 - sNuc-10X data processing and clustering analysis
 - Integrated analysis of multiple datasets
 - Cell cycle analysis
 - Correlation analysis of kidney organoid and mouse/human kidney

- Hierarchical clustering analysis on the cell types from kidney organoids
- Random Forest model to map cell types from human fetal kidney
- Cell type specific driving force analysis
- Deconvolution of bulk RNA-seq data
- GWAS analysis
- Kidney disease trait association with single cell-based cell types
- Pseudotemporal ordering
- Ligand-receptor interaction analysis
- Real Time PCR data quantification and statistical analysis

● DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and three tables and can be found with this article online at <https://doi.org/10.1016/j.stem.2018.10.010>.

ACKNOWLEDGMENTS

This work was supported by NIH/NIDDK grants DK103740 and DK107374, an established investigator award of the American Heart Association, and grant 173970 from the Chan Zuckerberg Initiative (all to B.D.H.). K.U. was supported by Japan Society for the Promotion of Science postdoctoral fellowships for research abroad.

AUTHOR CONTRIBUTIONS

H.W. designed and carried out experiments, analyzed results, and contributed to writing of the manuscript. K.U. designed and carried out experiments and analyzed results. E.L.D. and Y.K. contributed to experimental design, carried out experiments, and analyzed results. S.A.M. contributed to experimental design, analyzed results, and reviewed the manuscript. B.D.H. designed experiments, analyzed results, and wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 19, 2017

Revised: August 9, 2018

Accepted: October 6, 2018

Published: November 15, 2018

REFERENCES

- Adam, M., Potter, A.S., and Potter, S.S. (2017). Psychrophilic proteases dramatically reduce single-cell RNA-seq artifacts: a molecular atlas of kidney development. *Development* 144, 3625–3632.
- Bahar Halpern, K., Caspi, I., Lemze, D., Levy, M., Landen, S., Elinav, E., Ulitsky, I., and Itzkovitz, S. (2015). Nuclear Retention of mRNA in Mammalian Tissues. *Cell Rep.* 13, 2653–2662.
- Baron, M., Veres, A., Wolock, S.L., Faust, A.L., Gaujoux, R., Vetere, A., Ryu, J.H., Wagner, B.K., Shen-Orr, S.S., Klein, A.M., et al. (2016). A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. *Cell Syst.* 3, 346–360.e4.
- Barthelson, R.A., Lambert, G.M., Vanier, C., Lynch, R.M., and Galbraith, D.W. (2007). Comparison of the contributions of the nuclear and cytoplasmic compartments to global gene expression in human cells. *BMC Genomics* 8, 340.
- Bindels, R.J., Hartog, A., Timmermans, J.A., and van Os, C.H. (1991). Immunocytochemical localization of calbindin-D28k, calbindin-D9k and parvalbumin in rat kidney. *Contrib. Nephrol.* 91, 7–13.

- Brunskill, E.W., Aronow, B.J., Georgas, K., Rumballe, B., Valerius, M.T., Aronow, J., Kaimal, V., Jegga, A.G., Yu, J., Grimmond, S., et al. (2008). Atlas of gene expression in the developing kidney at microanatomic resolution. *Dev. Cell* 15, 781–791.
- Bulchand, S., Grove, E.A., Porter, F.D., and Tole, S. (2001). LIM-homeodomain gene Lhx2 regulates the formation of the cortical hem. *Mech. Dev.* 100, 165–175.
- Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* 36, 411–420.
- Calderon, D., Bhaskar, A., Knowles, D.A., Golan, D., Raj, T., Fu, A.Q., and Pritchard, J.K. (2017). Inferring Relevant Cell Types for Complex Traits by Using Single-Cell Gene Expression. *Am. J. Hum. Genet.* 101, 686–699.
- Chang-Panesso, M., Kadyrov, F.F., Machado, F.G., Kumar, A., and Humphreys, B.D. (2018). Meis1 is specifically upregulated in kidney myofibroblasts during aging and injury but is not required for kidney homeostasis or fibrotic response. *Am. J. Physiol. Renal Physiol.* 315, F275–F290.
- Chen, L., Lee, J.W., Chou, C.L., Nair, A.V., Battistone, M.A., Păunescu, T.G., Merkulova, M., Breton, S., Verlander, J.W., Wall, S.M., et al. (2017). Transcriptomes of major renal collecting duct cell types in mouse identified by single-cell RNA-seq. *Proc. Natl. Acad. Sci. USA* 114, E9989–E9998.
- Coresh, J., Selvin, E., Stevens, L.A., Manzi, J., Kusek, J.W., Eggers, P., Van Lente, F., and Levey, A.S. (2007). Prevalence of chronic kidney disease in the United States. *JAMA* 298, 2038–2047.
- Costantini, F., and Kopan, R. (2010). Patterning a complex organ: branching morphogenesis and nephron segmentation in kidney development. *Dev. Cell* 18, 698–712.
- Czerniecki, S.M., Cruz, N.M., Harder, J.L., Menon, R., Annis, J., Otto, E.A., Gulieva, R.E., Islas, L.V., Kim, Y.K., Tran, L.M., et al. (2018). High-Throughput Screening Enhances Kidney Organoid Differentiation from Human Pluripotent Stem Cells and Enables Automated Multidimensional Phenotyping. *Cell Stem Cell* 22, 929–940.e4.
- de Caestecker, M., Humphreys, B.D., Liu, K.D., Fissell, W.H., Cerdá, J., Nolin, T.D., Askenazi, D., Mour, G., Harrell, F.E., Jr., Pullen, N., et al.; ASN AKI Advisory Group (2015). Bridging Translation by Improving Preclinical Study Design in AKI. *J. Am. Soc. Nephrol.* 26, 2905–2916.
- Dehmelt, L., and Halpain, S. (2005). The MAP2/Tau family of microtubule-associated proteins. *Genome Biol.* 6, 204.
- Der, E., Ranabothu, S., Suryawanshi, H., Akat, K.M., Clancy, R., Morozov, P., Kustagi, M., Czuppa, M., Izmirly, P., Belmont, H.M., et al. (2017). Single cell RNA sequencing to dissect the molecular heterogeneity in lupus nephritis. *JCI Insight* 2, e93009.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21.
- Freedman, B.S., Brooks, C.R., Lam, A.Q., Fu, H., Morizane, R., Agrawal, V., Saad, A.F., Li, M.K., Hughes, M.R., Werff, R.V., et al. (2015). Modelling kidney disease with CRISPR-mutant kidney organoids derived from human pluripotent epiblast spheroids. *Nat. Commun.* 6, 8715.
- Guo, M., Wang, H., Potter, S.S., Whitsett, J.A., and Xu, Y. (2015). SINCERA: A Pipeline for Single-Cell RNA-Seq Profiling Analysis. *PLoS Comput. Biol.* 11, e1004575.
- Habib, N., Avraham-David, I., Basu, A., Burks, T., Shekhar, K., Hofree, M., Choudhury, S.R., Aguet, F., Gelfand, E., Ardlie, K., et al. (2017). Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* 14, 955–958.
- Haghverdi, L., Lun, A.T.L., Morgan, M.D., and Marioni, J.C. (2018). Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.* 36, 421–427.
- Hrvatin, S., Hochbaum, D.R., Nagy, M.A., Cicconetti, M., Robertson, K., Cheadle, L., Zilionis, R., Ratner, A., Borges-Monroy, R., Klein, A.M., et al. (2018). Single-cell analysis of experience-dependent transcriptomic states in the mouse visual cortex. *Nat. Neurosci.* 21, 120–129.
- Huang, E.J., and Reichardt, L.F. (2001). Neurotrophins: roles in neuronal development and function. *Annu. Rev. Neurosci.* 24, 677–736.
- Kawakami, Y., Eliyahu, S., Delgado, C.H., Robbins, P.F., Sakaguchi, K., Appella, E., Yannelli, J.R., Adema, G.J., Miki, T., and Rosenberg, S.A. (1994). Identification of a human melanoma antigen recognized by tumor-infiltrating lymphocytes associated with in vivo tumor rejection. *Proc. Natl. Acad. Sci. USA* 91, 6458–6462.
- Kowalczyk, M.S., Tiros, I., Heckl, D., Rao, T.N., Dixit, A., Haas, B.J., Schneider, R.K., Wagers, A.J., Ebert, B.L., and Regev, A. (2015). Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. *Genome Res.* 25, 1860–1872.
- Labastie, M.C., Catala, M., Gregoire, J.M., and Peault, B. (1995). The GATA-3 gene is expressed during human kidney embryogenesis. *Kidney Int.* 47, 1597–1603.
- Lake, B.B., Codeluppi, S., Yung, Y.C., Gao, D., Chun, J., Kharchenko, P.V., Linnarsson, S., and Zhang, K. (2017). A comparative strategy for single-nucleus and single-cell transcriptomes confirms accuracy in predicted cell-type expression from nuclear RNA. *Sci. Rep.* 7, 6031.
- Lam, A.Q., Freedman, B.S., Morizane, R., Lerou, P.H., Valerius, M.T., and Bonventre, J.V. (2014). Rapid and Efficient Differentiation of Human Pluripotent Stem Cells into Intermediate Mesoderm That Forms Tubules Expressing Kidney Proximal Tubular Markers. *J. Am. Soc. Nephrol.* 25, 1211–1225.
- Lèbre, S. (2009). Inferring dynamic genetic networks with low order independencies. *Stat. Appl. Genet. Mol. Biol.* 8, 9.
- Li, M., Li, Y., Weeks, O., Mijatovic, V., Teumer, A., Huffman, J.E., Tromp, G., Fuchsberger, C., Gorski, M., Lyttikäinen, L.P., et al.; CHARGE Glycemic-T2D Working Group; CHARGE Blood Pressure Working Group (2017). SOS2 and ACP1 Loci Identified through Large-Scale Exome Chip Analysis Regulate Kidney Development and Function. *J. Am. Soc. Nephrol.* 28, 981–994.
- Lindstrom, N.O., De Sena Brandine, G., Tran, T., Ransick, A., Suh, G., Guo, J., Kim, A.D., Parvez, R.K., Ruffins, S.W., Rutledge, E.A., et al. (2018a). Progressive Recruitment of Mesenchymal Progenitors Reveals a Time-Dependent Process of Cell Fate Acquisition in Mouse and Human Nephrogenesis. *Dev. Cell* 45, 651–660.e4.
- Lindström, N.O., Guo, J., Kim, A.D., Tran, T., Guo, Q., De Sena Brandine, G., Ransick, A., Parvez, R.K., Thornton, M.E., Baskin, L., et al. (2018b). Conserved and Divergent Features of Mesenchymal Progenitor Cell Types within the Cortical Nephrogenic Niche of the Human and Mouse Kidney. *J. Am. Soc. Nephrol.* 29, 806–824.
- Litvinov, S.V., Velders, M.P., Bakker, H.A., Fleuren, G.J., and Warnaar, S.O. (1994). Ep-CAM: a human epithelial antigen is a homophilic cell-cell adhesion molecule. *J. Cell Biol.* 125, 437–446.
- Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tiros, 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.
- Magella, B., Adam, M., Potter, A.S., Venkatasubramanian, M., Chetal, K., Hay, S.B., Salomonis, N., and Potter, S.S. (2018). Cross-platform single cell analysis of kidney development shows stromal cells express Gdnf. *Dev. Biol.* 434, 36–47.
- McMahon, A.P., Aronow, B.J., Davidson, D.R., Davies, J.A., Gaido, K.W., Grimmond, S., Lessard, J.L., Little, M.H., Potter, S.S., Wilder, E.L., and Zhang, P.; GUDMAP project (2008). GUDMAP: the genitourinary developmental molecular anatomy project. *J. Am. Soc. Nephrol.* 19, 667–671.
- Michael, U.F., Barenberg, R.L., Chavez, R., Vaamonde, C.A., and Papper, S. (1972). Renal handling of sodium and water in the hypothyroid rat. Clearance and micropuncture studies. *J. Clin. Invest.* 51, 1405–1412.
- Morizane, R., and Bonventre, J.V. (2017). Generation of nephron progenitor cells and kidney organoids from human pluripotent stem cells. *Nat. Protoc.* 12, 195–207.
- Morizane, R., Lam, A.Q., Freedman, B.S., Kishi, S., Valerius, M.T., and Bonventre, J.V. (2015). Nephron organoids derived from human pluripotent stem cells model kidney development and injury. *Nat. Biotechnol.* 33, 1193–1200.

- Nakai, S., Kawano, H., Yudate, T., Nishi, M., Kuno, J., Nagata, A., Jishage, K., Hamada, H., Fujii, H., Kawamura, K., et al. (1995). The POU domain transcription factor Brn-2 is required for the determination of specific neuronal lineages in the hypothalamus of the mouse. *Genes Dev.* 9, 3109–3121.
- Parekh, S., Ziegenhain, C., Vieth, B., Enard, W., and Hellmann, I. (2018). zUMIs - A fast and flexible pipeline to process RNA sequencing data with UMIs. *Gigascience* 7, giy059.
- Park, J., Shrestha, R., Qiu, C., Kondo, A., Huang, S., Werth, M., Li, M., Barasch, J., and Suszták, K. (2018). Single-cell transcriptomics of the mouse kidney reveals potential cellular targets of kidney disease. *Science* 360, 758–763.
- Pattaro, C., Teumer, A., Gorski, M., Chu, A.Y., Li, M., Mijatovic, V., Garnaas, M., Tin, A., Sorice, R., Li, Y., et al.; ICBP Consortium; AGEN Consortium; CARDIOGRAM; CHARGe-Heart Failure Group; ECHOGen Consortium (2016). Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. *Nat. Commun.* 7, 10023.
- Przepiorski, A., Sander, V., Tran, T., Hollywood, J.A., Sorrenson, B., Shih, J.H., Wolvetang, E.J., McMahon, A.P., Holm, T.M., and Davidson, A.J. (2018). A Simple Bioreactor-Based Method to Generate Kidney Organoids from Pluripotent Stem Cells. *Stem Cell Reports* 11, 470–484.
- Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H.A., and Trapnell, C. (2017). Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* 14, 979–982.
- Ramilowski, J.A., Goldberg, T., Harshbarger, J., Kloppmann, E., Lizio, M., Satagopam, V.P., Itoh, M., Kawaji, H., Carninci, P., Rost, B., and Forrest, A.R. (2015). A draft network of ligand-receptor-mediated multicellular signalling in human. *Nat. Commun.* 6, 7866.
- Schwarz, K., Simons, M., Reiser, J., Saleem, M.A., Faul, C., Kriz, W., Shaw, A.S., Holzman, L.B., and Mundel, P. (2001). Podocin, a raft-associated component of the glomerular slit diaphragm, interacts with CD2AP and nephrin. *J. Clin. Invest.* 108, 1621–1629.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504.
- Sharmin, S., Taguchi, A., Kaku, Y., Yoshimura, Y., Ohmori, T., Sakuma, T., Mukoyama, M., Yamamoto, T., Kurihara, H., and Nishinakamura, R. (2016). Human Induced Pluripotent Stem Cell-Derived Podocytes Mature into Vascularized Glomeruli upon Experimental Transplantation. *J. Am. Soc. Nephrol.* 27, 1778–1791.
- Shekhar, K., Lapan, S.W., Whitney, I.E., Tran, N.M., Macosko, E.Z., Kowalczyk, M., Adiconis, X., Levin, J.Z., Nemesh, J., Goldman, M., et al. (2016). Comprehensive Classification of Retinal Bipolar Neurons by Single-Cell Transcriptomics. *Cell* 166, 1308–1323.e30.
- Short, K.M., Combes, A.N., Lefevre, J., Ju, A.L., Georgas, K.M., Lamberton, T., Cairncross, O., Rumballe, B.A., McMahon, A.P., Hamilton, N.A., et al. (2014). Global quantification of tissue dynamics in the developing mouse kidney. *Dev. Cell* 29, 188–202.
- Taguchi, A., and Nishinakamura, R. (2017). Higher-Order Kidney Organogenesis from Pluripotent Stem Cells. *Cell Stem Cell* 21, 730–746.e6.
- Takasato, M., Er, P.X., Chiu, H.S., Maier, B., Baillie, G.J., Ferguson, C., Parton, R.G., Wolvetang, E.J., Roost, M.S., Chuva de Sousa Lopes, S.M., and Little, M.H. (2015). Kidney organoids from human iPS cells contain multiple lineages and model human nephrogenesis. *Nature* 526, 564–568.
- Takasato, M., Er, P.X., Chiu, H.S., and Little, M.H. (2016). Generation of kidney organoids from human pluripotent stem cells. *Nat. Protoc.* 11, 1681–1692.
- Tapley, P., Lamballe, F., and Barbacid, M. (1992). K252a is a selective inhibitor of the tyrosine protein kinase activity of the trk family of oncogenes and neurotrophin receptors. *Oncogene* 7, 371–381.
- Tirosh, I., Izar, B., Prakadan, S.M., Wadsworth, M.H., 2nd, Treacy, D., Trombetta, J.J., Rotem, A., Rodman, C., Lian, C., Murphy, G., et al. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* 352, 189–196.
- United States Renal Data System (2017). 2017 USRDS annual data report: Epidemiology of kidney disease in the United States. <https://wwwUSRDS.org/adr.aspx>.
- van Dijk, D., Sharma, R., Nainys, J., Yim, K., Kathail, P., Carr, A.J., Burdziak, C., Moon, K.R., Chaffer, C.L., Pattabiraman, D., et al. (2018). Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. *Cell* 174, 716–729.e7.
- Wu, H., Malone, A.F., Donnelly, E.L., Kiritu, Y., Uchimura, K., Ramakrishnan, S.M., Gaut, J.P., and Humphreys, B.D. (2018). Single-Cell Transcriptomics of a Human Kidney Allograft Biopsy Specimen Defines a Diverse Inflammatory Response. *J. Am. Soc. Nephrol.* 29, 2069–2080.
- Xia, Y., Nivet, E., Sancho-Martinez, I., Gallegos, T., Suzuki, K., Okamura, D., Wu, M.Z., Dubova, I., Esteban, C.R., Montserrat, N., et al. (2013). Directed differentiation of human pluripotent cells to ureteric bud kidney progenitor-like cells. *Nat. Cell Biol.* 15, 1507–1515.
- Xia, Y., Sancho-Martinez, I., Nivet, E., Rodriguez Esteban, C., Campistol, J.M., and Izpisua Belmonte, J.C. (2014). The generation of kidney organoids by differentiation of human pluripotent cells to ureteric bud progenitor-like cells. *Nat. Protoc.* 9, 2693–2704.

STAR★METHODS**KEY RESOURCES TABLE**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse anti-WT1	Santa Cruz	Cat#SC-7385; RRID: AB_628448
Rat anti-ECAD	Abcam	Cat#ab11512; RRID: AB_298118
Biotinylated LTL	Vector Labs	Cat#B-1325; RRID: AB_2336558
Sheep anti-NPHS1	R&D Systems	Cat#AF4269; RRID: AB_2154851
Rabbit anti-CRABP1	Cell Signaling	Cat#13163
Chicken anti-MAP2	Abcam	Cat#ab5392; RRID: AB_2138153
Mouse anti MEIS1	Active Motif	Cat#39795
Secondary antibodies included FITC-, Cy3, or Cy5-conjugated	Jackson ImmunoResearch	Cat#711-095-152, Cat#712-095-153, Cat#715-165-151, Cat#713-165-147, Cat#016-600-084, Cat#703-165-155
DAPI (4',6'-diamidino-2-phenylindole)	Thermo Fisher Scientific	Cat#D1306
Chemicals, Peptides, and Recombinant Proteins		
Barcoded Dropseq beads	ChemGenes	Cat#MACOSKO-2011-10
10mM Tris-HCl, pH 8.0	Teknova	Cat#T1173
UltraPure SSC	Invitrogen	Cat#15557036
Ficoll PM-400	GE Healthcare/Fisher Scientific	Cat#45-001-745
Sarkosyl	Sigma-Aldrich	Cat#L7414-50mL
Exonuclease I	New England Biolabs	Cat#M0293L
Perfluoro-1-octanol	Sigma-Aldrich	Cat#370533-25G
dNTP mix	Clontech	Cat#639125
Droplet Generation Oil	BioRad	Cat#186-4006
Tris-EDTA buffer, pH 8.0	Sigma-Aldrich	Cat#93283
1M DTT	Teknova	Cat#D9750
2M Tris pH 7.5	Sigma-Aldrich	Cat#T2944
Tween 20, RNase Free	Promega	Cat#H5152
NxGen RNase Inhibitor	Lucigen	Cat#30281-2
Nuclei EZ Lysis Buffer	Sigma-Aldrich	Cat#N-3408
RNasin Plus Ribonuclease Inhibitors	Promega	Cat#N2615
SUPERaseIN RNase Inhibitor	Thermo Fisher Scientific	Cat#AM2696
cComplete ULTRA Tablets, Mini, EDTA-free, EASYpack	Roche	Cat#05 892 791 001
CHIR	Tocris Bioscience	Cat#4423
APEL 2	STEMCELL Technologies	Cat#05275
Protein Free Hybridoma Medium II	GIBCO	Cat#12040077
Advanced RPMI 1640 medium	GIBCO	Cat#12633012
FGF9	R&D Systems	Cat#273-F9-025/CF
Heparin	Sigma-Aldrich	Cat#H4784-250MG
Trypsin-EDTA	Sigma-Aldrich	Cat#25200-114
Noggin	PeproTech	Cat#120-10C
GlutaMAX	Thermo Fisher Scientific	Cat#35050061
Activin	R&D Systems	Cat#338-AC-010
Accutase	STEMCELL Technologies	Cat#7920
TrypLE Select	Thermo Fisher Scientific	Cat#12563-029
ReLeSR	STEMCELL Technologies	Cat#05872
BDNF inhibitor (K252a)	Sigma-Aldrich	Cat#K1639

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Prolong Gold	Life Technologies	Cat#P36930
Essential 8 Medium	GIBCO	Cat#A2858501
Matrigel hESC-Qualified Matrix	Corning	Cat#354277
Antibiotic-Antimycotic Solution	Corning	Cat#30004CI
Critical Commercial Assays		
Maxima H Minus Reverse Transcriptase	ThermoFisher	Cat#EP0753
KAPA HiFi hotstart readymix	KAPA Biosystems	Cat#KK2602
Single Cell 3' Library and Gel Bead Kit V2	10x Genomics	Cat#120237
Chromium single cell chip kit V2	10x Genomics	Cat#120236
Nextera XT DNA Sample Preparation Kit	Illumina	Cat#FC-131-1096
Agilent High Sensitivity DNA Kit	Agilent	Cat#5067-4626
Agencourt AMPure XP - PCR Purification	Backman Coulter	Cat#A63881
Deposited Data		
Raw and analyzed data	This paper	GEO: GSE118184
Mendeley Data	This paper	https://doi.org/10.17632/m4rf9wb29.1
Experimental Models: Cell Lines		
Human iPSCs: BJFF.6 line	GEiC	N/A
Human ESCs: H9 line	GEiC	N/A
Software and Algorithms		
Drop-seq_tools v1.12	(Macosko et al., 2015)	http://mccarrolllab.org/dropseq
STAR v2.5.3a	(Dobin et al., 2013)	https://github.com/alexdobin/STAR
R 3.4.1	R project	https://www.r-project.org
Seurat v2.0	(Butler et al., 2018)	https://satijalab.org/seurat/
Monocle2	(Qiu et al., 2017)	https://github.com/cole-trapnell-lab/monocle-release
zUMIs	(Parekh et al., 2018)	https://github.com/sdparkh/zUMIs
Random Forest	CRAN-R	https://cran.r-project.org/web/packages/randomForest/index.html
SINCERA	(Guo et al., 2015)	https://github.com/xu-lab/SINCERA
BSeq-sc	(Baron et al., 2016)	https://github.com/shenorrlab/bseqsc
MNN	(Haghverdi et al., 2018)	https://github.com/MarioniLab/MNN2017/
rolypoly	(Calderon et al., 2017)	https://github.com/dcalderon/rolypoly
Cytoscape 3.6.1	(Shannon et al., 2003)	https://www.cytoscape.org/
Connectome	(Ramilowski et al., 2015)	https://github.com/Hypercubed/connectome
MAGIC	(van Dijk et al., 2018)	https://github.com/KrishnaswamyLab/MAGIC
Other		
Nuclei isolation protocol	This study	STAR Methods
Morizane organoid differentiation protocol	(Morizane et al., 2015)	N/A
Takasato organoid differentiation protocol	(Takasato et al., 2015)	N/A
Dropseq library preparation protocol	McCarroll's lab	http://mccarrolllab.org/dropseq/
10x library preparation protocol	10x genomics	https://support.10xgenomics.com/single-cell-gene-expression/library-prep/doc/user-guide-chromium-single-cell-3-reagent-kits-user-guide-v2-chemistry
Hemocytometer	INCYTO	Cat#DHC-F015
pluriStrainer 40 µm	pluriSelect	Cat#43-50040
pluriStrainer 20 µm	pluriSelect	Cat#43-50020
KONTES Dounce Tissue Grinders	Kimble Chase	Cat#KT885300-0002
Transwell polyester membrane cell culture inserts	Corning	Cat#3460
96 Well, Clear Round Bottom, Ultra Low	Corning	Cat# 7007
Syringe pumps	KD Scientific	Legato 100

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Magnetic mixing system	VP Scientific	Part #710D2
PDMS Microfluidic Device	FlowJEM	N/A
Mouse E14.5 kidney data	(Magella et al., 2018)	GEO: GSE104396
Mouse P1 kidney data	(Adam et al., 2017)	GEO: GSE94333
Human fetal kidney data	(Lindström et al., 2018b)	GEO: GSE102596
Organoid bulk RNA-seq data	(Takasato et al., 2015)	GEO: GSE70101

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Benjamin D. Humphreys (humphreysbd@wustl.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS**iPSC Culture**

All experiments utilized the BJFF6 human iPSC line reprogrammed by Sendai virus from human foreskin fibroblasts (Washington University Genome Engineering and iPSC Core). This line is confirmed to be karyotypically normal. BJFF6 cells were maintained in 6-well plates coated with matrigel (Corning) in Essential 8 medium (Thermo Fisher Scientific). iPSC cells were dissociated using ReLeSR (STEMCELL Technologies), confirmed to be mycoplasma free and maintained below passage 50.

Kidney donor

Institutional review board approval for research use of human tissue was obtained from Washington University. Renal cortex from a nephrectomy kidney was obtained and donor anonymity preserved. The donor was a 62 year-old white male with a serum creatinine of 1.03 mg/dL and BUN of 12 mg/dL.

METHOD DETAILS**Kidney Organoid Differentiation**

Kidney organoids were generated following either the protocol described by Takasato et al. (2016) or that of Morizane (Morizane and Bonventre, 2017) with minimal modifications. Briefly, for the Takasato approach, BJFF cells were treated with CHIR (8 uM, Tocris Bioscience) in basal medium - APEL 2 (STEMCELL Technologies) supplemented with 5% Protein Free Hybridoma Medium II (PFHMI, GIBCO) - for 4 days, followed by FGF9 (200 ng/mL, R&D Systems) and heparin (1 ug/mL, Sigma-Aldrich) for another 3 days. At day 7, cells were collected and dissociated into single cells using 0.25% Trypsin-EDTA (Thermo Fisher Scientific). Cells were spun down at 400 g for 3 min to form a pellet and transferred onto a trans-well membrane. Pellets were incubated with CHIR (5 uM) for 1 hour and then cultured with FGF9 (200 ng/mL) and heparin (1 ug/mL) for 5 days. For the next 13 days, organoids were cultured with basal medium changed every other day. For the Morizane approach, BJFF cells were treated with CHIR (10 uM) and Noggin (5 ng/mL, PeproTech) in basal medium - Advanced RPMI 1640 medium (GIBCO) supplemented with 1X L-GlutaMAX (Thermo Fisher Scientific) - for 4 days, followed by 3 days Activin (10 ng/mL, R&D Systems) and 2 days FGF9 (10 ng/mL) treatment. At day 9, the cells were dissociated with Accutase (StemCell technologies) and resuspended in the basic differentiation medium with CHIR (3 uM) and FGF9 (10 ng/mL), and placed in ultra-low attachment 96-well plates. Two days later the medium was changed to basal medium containing FGF9 (10 ng/mL) and cultured for 3 more days. After that, the organoids were cultured in basal medium with no additional factors until harvest at day 26. For ES (H9) cell line, we adjusted the concentration of CHIR to 5uM in Takasato protocol and to 8uM in Morizane protocol at the initial step.

DropSeq scRNA-seq

Organoids were dissociated using TrypLE Select (Thermo Fisher Scientific) at 37°C with shaking. After 5 min, cells were further dispersed by gentle pipetting and filtered through a 40μm cell strainer (pluriSelect). Single cell suspension was visually inspected under a microscope, counted by hemocytometer (INCYTO C-chip) and resuspended in PBS + 0.01% BSA. Single cells were coencapsulated in droplets with barcoded beads exactly as described (Macosko et al., 2015). In brief, cells were diluted to a concentration of 100 cells/μL, and co-encapsulated with barcoded beads (ChemGenes #Macosko201110), which were diluted to a concentration of 120 beads/μL. Droplets of about 1 nL in size were generated using microfluidic polydimethylsiloxane (PDMS) co-flow devices (FlowJEM Drop-seq chips). Droplets were collected in a 50-mL RNase-free Falcon tube for a total run time of about 15 min. Droplet emulsion was aliquoted into 1 mL each of cells and beads and were broken promptly by perfluorooctanol, following which barcoded beads with captured transcriptomes were washed and spun down at 4°C. Hybridized

RNA was reverse transcribed and exonuclease-treated using commercial kits (See [Key Resources Table](#)). The beads from one run were then equally distributed into individual PCR tubes with populations of 8,000 beads/tube (~400 cells). PCR tubes were separately amplified for 4+9 PCR cycles, and the PCR products were purified by the addition of 0.6x Agencourt AMPure XP beads (Beckman Coulter #A63881). The quality of the amplified cDNA was evaluated by Bioanalyzer (Agilent 2100) on a High Sensitivity DNA chip. Only cDNA with average insertion size > 1200 bp were used for downstream library preparation and sequencing. cDNA from an estimated 5,000 cells were prepared and fragmented by Nextera XT (Illumina) using 600 pg of cDNA input. cDNA library was amplified (12 cycles) using custom primers as described ([Macosko et al., 2015](#)). Amplified libraries were purified with a 0.6x volume of AMPure XP beads and quality was measured by Bioanalyzer. Libraries with average length of ~500-700 bp were submitted to Genome Technology Access Center (GTAC) of Washington University in St. Louis and sequenced on HiSeq 2500 and NovaSeq 6000 (Illumina). We routinely tested our DropSeq setup by running species mixing experiments prior to running on actual sample to assure that the cell doublet rate was below 5%. Information about experimental replicates and count statistics is specified in [Table S1](#).

Nuclei isolation and snRNA-seq of human kidney

Nuclei were isolated with Nuclei EZ Lysis buffer (Sigma #NUC-101) supplemented with protease inhibitor (Roche #5892791001) and RNase inhibitor (Promega #N2615, Life Technologies #AM2696). Samples were cut into < 2 mm pieces and homogenized using a Dounce homogenizer (Kimble Chase #885302-0002) in 2ml of ice-cold Nuclei EZ Lysis buffer and incubated on ice for 5 min with an additional 2ml of lysis buffer. The homogenate was filtered through a 40-μm cell strainer (pluriSelect #43-50040-51) and then centrifuged at 500 x for 5 min at 4°C. The pellet was resuspended and washed with 4 mL of the buffer and incubated on ice for 5 min. After another centrifugation, the pellet was resuspended with Nuclei Suspension Buffer (1x PBS, 0.07% BSA, 0.1% RNase inhibitor), filtered through a 20-μm cell strainer (pluriSelect 43-50020-50) and counted. RNA from single nucleus was encapsulated, barcoded and reversed transcribed on a 10x Chromium Single Cell Platform (10x Genomics). The library was sequenced in HiSeq2500.

Immunofluorescence

Organoids were fixed in 4% paraformaldehyde (Electron Microscopy Services), cryoprotected in 30% sucrose solution overnight and embedded in optimum cutting temperature (OCT) compound (Tissue Tek). Organoids were cryosectioned at 7 μm thickness and mounted on Superfrost slides (Thermo Fisher Scientific). Sections were washed with PBS (3 times, 5 minutes each), then blocked with 10% normal goat serum (Vector Labs), permeabilized with 0.2% Triton X-100 in PBS and then stained with primary antibody specific for mouse anti-WT1 (1:200, Santa Cruz Biotechnology, #SC-7385), rat anti-ECAD (1:200, Abcam, #ab11512), biotinylated LTL (1:200, Vector Labs, #B-1325), sheep anti-NPHS1 (1:200, R&D Systems, #AF4269) and rabbit anti-CRABP1 (1:200, Cell Signaling, #13163), chicken anti-MAP2 (1:200, Abcam, #ab5392) and mouse anti MEIS1 (1:200, Active Motif, #39795). Secondary antibodies included FITC-, Cy3, or Cy5-conjugated (Jackson ImmunoResearch). Then, sections were stained with DAPI (4',6'-diamidino-2-phenylindole) and mounted in Prolong Gold (Life Technologies). Images were obtained by confocal microscopy (Nikon C2+ Eclipse; Nikon, Melville, NY).

BDNF inhibitor (K252a) treatment

We treated the iPSC derived kidney organoids differentiated from Takasato protocol with different doses of a BDNF inhibitor, K252a (Sigma-Aldrich #K1639). A dose of 250 nM was selected because the organoid size and tubular morphology were not altered at this concentration. Starting from day 12, we supplemented K252a to the basal medium, and differentiated the organoid to day 26 in the presence of K252a. The medium containing K252a was replaced every 2 days. At day 26, two different batches of organoids were harvested for DropSeq, immunofluorescence and qPCR.

Real Time PCR Experiments

RNA from whole organoids was extracted using the RNeasy Mini Kit (QIAGEN) and 600 ng of total RNA was reverse transcribed with iScript (BioRad). Quantitative polymerase chain reactions were carried out with iQ-SYBR Green supermix (BioRad) and the BioRad CFX96 Real Time System with the C1000 Touch Thermal Cycler. Cycling conditions were 95°C for 3 minutes then 40 cycles of 95°C for 15 s and 60°C for 1 minute, followed by one cycle of 95°C for 10 s. Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was used as a housekeeping gene. Data was analyzed using the $2^{-\Delta\Delta Ct}$ method. The following primers were used: GAPDH: Fw 5'-GACAGTCAGCCGCATCTTCT -3'; Rv 5'-GCGCCAATACGACCAAATC -3'; Cited1: Fw 5'-CCTCACCTGCGAAGGAGGA -3'; Rv 5'-GGAGAGCCTATTGGAGATCCC -3'; NPHS1: Fw 5'-CTGCTGAAAACCTGACGGT -3'; Rv 5'-GACCTGGCACTCATACTCCG -3'; SLC3A1: Fw 5'-CAGGAGCCCCGACTTCAAGG -3'; Rv 5'-GAGGGCAATGATGGCTATGGT -3'

SLC12A1: Fw 5'-AGTGCCCAGTAATACCAATCGC -3'; Rv 5'-GCCTAAAGCTGATTCTGAGTCTT -3'; CRABP1: Fw 5'-GCAGCAGCGAGAATTTCGAC -3'; Rv 5'-CGTGGTGGATGTCTGATGTAGA -3'; MAP2: Fw 5'-CTCAGCACCGCTAACAGAGG -3'; Rv 5'-CATTGGCGCTTCGGACAAG -3'; MYLPF: Fw 5'-GAAGGACAGTAGAGGGCGGAA -3'; Rv 5'-TCTGGTCGATCACAGTGAAGG -3'; MYOG: Fw 5'-GGGGAAAACCTACCTGCCTGTC -3'; Rv 5'-AGGCGCTCGATGTACTGGAT -3'; MLANA: Fw 5'-GCTCACTCATCTATGGTTACCC -3'; Rv 5'-GACTCCCAGGATCACTGTCAG -3'; PMEL: Fw 5'-AGGTGCCTTCTCCGTGAG Rv 5'-AGCTTCA GCCAGATAGCCACT -3'

QUANTIFICATION AND STATISTICAL ANALYSIS

Preprocessing of DropSeq data

Paired-end sequencing reads were processed as previously described using the Drop-Seq Tools v1.12 software available in McCarroll's lab (<http://mccarrolllab.org/dropseq>). Briefly, each cDNA read (read2) was tagged with the cell barcode (the first 12 bases in read 1) and unique molecular identifier (UMI, the next 8 bases in read 1), trimmed of sequencing adaptors and poly-A sequences, and aligned to the human (GRCh38) or a concatenation of the mouse and human (for the species-mixing experiment) reference genome assembly using STAR v 2.5.3a (Dobin et al., 2013). Cell barcodes were corrected for possible bead synthesis errors using the DetectBeadSynthesisErrors program, and then collapsed to core barcodes if they were within an edit distance of 1 as previously described (Macosko et al., 2015). Digital gene expression (DGE) matrix was compiled by counting the number of unique UMIs for a given gene (as row) in each cell (as column).

Global clustering analysis

To assess the difference in cell composition across differentiation protocols and cell lines, we performed global clustering analysis on ten datasets from Dropseq sequencing of the day 26 organoids. First, we combined the UMI count matrices from different protocols and cell lines into one gene-cell matrix using the *merge* function in R. We then removed the low-quality cells with less than 500 or more than 4000 detected genes, or if their mitochondrial gene content was > 20%. Genes were filtered out that were detected in less than 10 cells. This filtering step resulted in 24,574 genes X 71,390 cells sampled from four batches Takasato iPS organoids (26,890 cells), three batches Morizane iPS organoids (18,072 cells), two batches Takasato ES organoids (19,380 cells) and one batch Morizane ES organoids (7,048 cells). The gene expression was then natural log transformed and normalized for scaling the sequencing depth to a total of 1e4 molecules per cell. Batch effect was corrected by matching mutual nearest neighbors using a recently published pipeline (incorporated in scran R package) (Haghverdi et al., 2018). Dimensionality reduction and clustering were performed on the batch effect corrected expression value using Seurat R tool kit.

A total of 23 clusters were classified from this analysis, consisting of cells from four broad "classes" defined as mesenchyme, tubule, podocyte and off-target cells based on their marker genes expression. We then performed a post hoc merging step, where transcriptionally indistinguishable clusters are merged back together. First, the distance between each pair of clusters within the broad cell class was computed based on the averaged expression value of the highly variable genes. Second, hierarchical clustering analysis was performed on the distance matrix using the *hclust* function from R. We then merged transcriptionally similar clusters that were placed adjacent on the hierarchical tree. With this approach, we identified five transcriptionally distinct subtypes in the tubule class, and 3 distinct subtypes in the mesenchyme class. Finally, we assessed the cell compositions by calculating the number of cells in each cluster or broad class, subdivided by the differentiation protocol and cell line.

Clustering analysis on the organoid cells from different protocols or cell lines

Seurat was used for quality control, dimensionality reduction and cell clustering for the Dropseq datasets generated by each protocol or cell line. In brief, raw DGE matrices from different batches for each protocol on each cell line were combined and loaded into the Seurat. For normalization, the DGE matrix was scaled by total UMI counts, multiplied by 10,000 and transformed to log space. Only genes found to be expressing in > 10 cells were retained. Additional filtering was set on the number of detected genes and mitochondrial gene content to remove the low-quality cells or cell doublets. We note that, depending on the sequencing depth and the variations in mitochondrial gene content from dataset to dataset, the cutoffs may need to be set on a case-by-case basis. Before clustering, variants arising from batch effects, library size and percentage of mitochondrial genes were regressed out by specifying the *vars.to.regress* argument in Seurat function *ScaleData*. The highly variable genes were identified using the function *FindVariableGenes*. The expression level of highly variable genes in the cells was scaled and centered along each gene, and was conducted to principal component analysis. We then assessed the number of PCs to be included in downstream analysis by (1) plotting the cumulative standard deviations accounted for each PC using the function *PCElbowPlot* in Seurat to identify the 'knee' point at a PC number after which successive PCs explain diminishing degrees of variance, and (2) by exploring primary sources of heterogeneity in the datasets using the *PCHeatmap* function in Seurat. Based on these two methods, we selected first top significant PCs for two-dimensional t-distributed stochastic neighbor embedding (tSNE), implemented by the Seurat software with the default parameters. We used *FindCluster* in Seurat to identify cell clusters for each protocol. Alternatively, A hierarchical clustering method devised by Baron et al. (2016) was recruited to validate the clusters identified by Seurat. To identify the marker genes, differential expression analysis was performed by the function *FindAllMarkers* in Seurat with likelihood-ratio test. Differentially expressed genes that were expressed at least in 25% cells within the cluster and with a fold change more than 0.25 (log scale) were considered to be marker genes. Gene expression of selected markers across clusters were visualized using a Python plotting library Matplotlib.

Sub-clustering tubular cells in Morizane protocol (cluster PT1, PT2, and LH) and Takasato protocol (cluster PT1, PT2, LH, and Lp) were first combined, then sub-clustered using the same approach described above. We identified 12 significant PCs in 3,056 tubular cells from Takasato protocol and 10 PCs in 4,933 tubular cells from Morizane protocol, which were further assigned into 9 and 5 sub-clusters, respectively. To further compare the cell types generated from the two protocols, we extracted the expression matrix for PT2 (a more mature PT cluster), LH and podocyte from each protocol, and used the combined matrices as input to Seurat. After regressing out potential protocol effect and library size difference, we performed differential gene analysis on the PT cells between

protocols with likelihood-ratio test. Developmental genes were selected based on the anchor/marker genes listed in GUDMAP (McMahon et al., 2008) and visualized by *DotPlot* function in Seurat.

For the time course Dropseq data, the datasets were preprocessed before being placed in the Seurat package. The matrices from all time points were merged into one single matrix with the *merge* function in R. To retain the gene expression variants across time course, all genes from each time point were kept in the merged matrix by assigning zeros to the genes for those cells who have missing expression value. 20 significant PCs calculated from 1,345 high variable genes were selected for tSNE and clustering analysis. We identified 14 clusters including key cell types corresponding to the important developmental states. We selected genes that represent each development stages and visualized them in tSNE map. Initially, we used the normalized gene expression value but found that the high dropout events in single cell data obscure the dynamic gene expression changes across time course. We therefore applied a gene imputation approach named MAGIC (van Dijk et al., 2018) to computationally fill in the missing value for all genes in our time course dataset. In brief, the merged time-course count matrix was loaded into the MAGIC pipeline written in Python (<https://github.com/KrishnaswamyLab/MAGIC>). Gene expression value in all cells were normalized, dimensionally reduced and transformed by the internal algorithms in MAGIC with the parameters: *n_pca_components* = 20, *random_pca* = True, *t* = 6, *k* = 30, *ka* = 10, *epsilon* = 1, *rescale_percent* = 99. Developmental genes and marker gene expression across time points after MAGIC imputation were visualized in the tSNE map constructed by Seurat and on the pseudotemporal trajectory tree constructed by Monocle (see methods below).

sNuc-10X data processing and clustering analysis

We used a newly developed pipeline, zUMIs (Parekh et al., 2018), to process the single nucleus sequencing data from human adult kidney. In brief, we first filtered out the low-quality barcodes or UMs based on sequence with the internal read filtering algorithm built in zUMIs. We then used zUMIs to map the filtered reads to human reference genome (GRCh38) using STAR 2.5.3a (two-pass mapping mode). Next, zUMIs quantified the reads that were uniquely mapped to exonic, intronic or intergenic region of the genome and inferred the true barcodes that mark cells by fitting a k-dimensional multivariate normal distribution with mclust package. Finally, a UMI count table utilizing both exonic and intronic reads were generated for downstream analysis. The whole data processing was executed by running the script on a HPC cluster with 96 × 2.3GHz computing cores (http://brc.wustl.edu/?page_id=12) with the following example script: *bash zUMIs-master.sh -f R1.fastq.gz -r R2.fastq.gz -c 1-16 -m 17-26 -l 98 -n Human_kidney -g GRCh38_ref_genome -a GRCh38_ref_genes.gtf -p 30*. Clustering analysis was performed on Seurat with a similar approach used for analyzing the organoid Dropseq datasets.

Integrated analysis of multiple datasets

To compare the cell types derived from different organoid protocols/cell lines and those from adult human kidney, we performed comparative analysis on multiple datasets by utilizing a recently developed computational strategy for integrated analysis (implemented in Seurat v2.0) (Butler et al., 2018). We first selected the union of the top 3,000 genes with the highest dispersion from all datasets for a canonical correlation analysis (CCA) to identify common sources of variation across the datasets. Then CCA was performed based on the normalized expression value of the highly-dispersed genes. Next, we selected the top dimensions of the CCA by examining a saturation in the relationship between the number of principle components and the percentage of the variance explained using the *MetageneBicorPlot* function in Seurat. We obtained a new dimensional reduction matrix by aligning the CCA subspaces with the top dimensions computed above. With the new dimensional reduction matrix, we performed clustering analysis on the cells or nuclei from different datasets by setting an optimal clustering parameters. We visualized the cells by their original identity or by their cluster identity classified by this integrated analysis. Differential gene analysis was performed on the cells or nuclei from different datasets but grouped in the same cluster after the alignment analysis. Differential genes were visualized using the *FeatureHeatmap* or *DotPlot* function in Seurat. We applied this computational strategy to compare the matched cell types (i.e., podocyte, PT and LH) from organoids and adult human kidney.

Cell cycle analysis

We assigned a cell cycle score on each cell according to its gene expression of G2/M and S phase markers (Tirosh et al., 2016). Based on this scoring system, we classified each cell in either G2M, S or G1 phase using the *CellCycleScoring* function in Seurat. The cells at different cell cycle classifications were visualized in the tSNE map, and the expression of cell cycle genes were plotted out using *FeaturePlot* function in Seurat.

Correlation analysis of kidney organoid and mouse/human kidney

To assess the similarity between kidney organoid cell types and embryonic kidney, we re-analyzed three previously published datasets from mouse E14.5 kidney (GEO: GSE104396) (Magella et al., 2018), P1 kidney (GEO: GSE94333) (Adam et al., 2017) and human fetal kidney (GEO: GSE102596) (Lindström et al., 2018b). We used the Seurat clustering parameters described by the authors and reproduced the same cell types from the datasets. We calculated the Pearson correlation based on the expression patterns of highly variable genes between cell populations within the mouse embryonic kidney dataset against the organoid cell types and the DT cell type identified from subclustering analysis. we performed the same analysis to compare organoid cell types to the human adult kidney cell types identified from the snRNA-seq dataset (data from this manuscript). Correlation matrix were visualized by R package *pheatmap*. Color keys (and dot sizes) represent the range of the coefficients of determination (r^2) in each analysis.

Hierarchical clustering analysis on the cell types from kidney organoids

To reveal the relationship among the matched cell types derived from Morizane and Takasato protocols, we extracted the expression profiles for P1, M3, PT2, LH and N2 subtypes from the Morizane organoid datasets and expression profiles for P, M3, PT2, LH, N2, and EC from the Takasato organoid datasets. Then we normalized the gene expression and computed the distance between each pair of cell types based on their average gene expression. Hierarchical clustering was performed on the distance matrix using *hclust* function in R and was visualized by dendrogram. We used the dendextend R package to compare the dendrogram from the Morizane organoid dataset and the Takasato organoid dataset.

Random Forest model to map cell types from human fetal kidney

To determine the congruence between cell types obtained from our organoid datasets and those in a recent fetal kidney dataset (Lindström et al., 2018b), we trained a multiclass random forest classifier (Habib et al., 2017; Shekhar et al., 2016) on the fetal kidney cell clusters and used it to map the organoid cell type. First, we composed a ‘training set’ by sampling 60% of the cells from 5 fetal kidney clusters representing mesenchyme, LH, PT, podocyte, endothelium and a cycling cell population (defined as mesenchyme progenitor). We next trained a random forest using 1,000 trees on the training set using the R package randomForest. We then used the remaining 40% of the cells from each cluster from the human fetal kidney dataset to validate the performance of the trained classifier. We used this model to assign a class label (one of the 5 human fetal kidney cell types) to each cell from kidney organoids. Finally, we quantified the number of cells that were mapped to each class label and visualized the data using ggplot2 package.

Cell type specific driving force analysis

To identify the key regulators that control the cell states, we performed cell type specific driving force analysis using the SINCERA pipeline (Guo et al., 2015). This approach consists of three main steps. First, the candidate transcription factors (TFs) and their regulatory target genes (TGs) were extracted from the DEG list identified in each cluster. Second, cell type specific transcription regulatory network (TRN) was constructed by establishing the interaction between TF-TF and TF-TG (TG-TF and TF auto-regulations were not considered) as previously described (Lèbre, 2009). Finally, the key TFs were selected based on their network node importance. This was accomplished by collecting the value of six node importance metrics including Degree Centrality (DC), Closeness Centrality (CC), Betweenness Centrality (BC), Disruptive Fragmentation Centrality (DFC), Disruptive Connection Centrality (DCC) and Disruptive Distance Centrality (DDC). TFs were ordered by taking the average of the node importance from these six matrices.

Deconvolution of bulk RNA-seq data

To examine the possible use of our single cell data to infer the cell type compositions (deconvolution) from the bulk RNA profiling, we retrieved the previously published RNA-Seq dataset from Takasato et al. (GEO: GSE70101) (Takasato et al., 2015). This dataset contains gene expression profiles of time-course organoids collected at similar time-points (day 7, day 10, day 18 and day 25) as those collected for our time-course single cell study. We then applied a single cell deconvolution algorithm, BSeq-sc (Baron et al., 2016), to estimate the proportion of iPS cells, PIM, pretubular aggregate, PT, LH, podocyte and neurons in each time point from bulk RNA-seq datasets. The identified marker genes for each cell type from our time-course single cell data were used as an input to estimate the cell proportion in each time point of the bulk-seq data according to the tutorial from BSeq-sc package (<https://shenorlab.github.io/bseqsc/vignettes/bseq-sc.html>).

GWAS analysis

We downloaded GWAS gene lists associated with chronic kidney disease, hypertension and metabolite from the GWAS site (<https://www.ebi.ac.uk/gwas/>). Each associated gene that was identified as significant in a GWAS (reported gene) was mapped to the cell type marker gene list obtained from Seurat *FindAllMarker* function. To plot the GWAS genes, we normalized the gene expression value for each cell by z-scores and generated a new gene expression matrix with mean z-scores for each GWAS gene by averaging the z-score value from all individual cells in the same cluster (i.e., a mean z-scores matrix with GWAS genes as rows and cell type ID as columns). *Heatmap.2* function in gplots R package was used to create all heatmap graphs in this analysis.

Kidney disease trait association with single cell-based cell types

We used RoLyPoly, a polygenic method that identifies trait-involved cell types by analyzing the enrichment of GWAS signal in cell type specific gene expression (Calderon et al., 2017), to associate the kidney disease trait with gene expression profile from our single cell data. Since RoLyPoly requires four indispensable components as data input (GWAS summary statistics, expression data, an expression data annotation file, and linkage disequilibrium (LD) information), we prepared each of them as follows. First, we obtained the summary meta-analysis data from the CKDGen consortium (<http://ckdgen.imbi.uni-freiburg.de>) for the associations between genotype and CKD or eGFR on the basis of serum creatinine (eGFRcrea) based on Li et al.’s study (Li et al., 2017; Pattaro et al., 2016). To prepare cell type expression data, we computed the average expression on each cell type from Takasato’s iPS organoid, Morizane’s iPS organoid, and adult human kidney. To link gene expression with the location of GWAS variants, we defined a block as a 10kb window centered around each gene’s transcription start site (TSS) as recommended by rolypoly. The TSS for all known hg19 genes was downloaded from UCSC Genome Browser (<https://genome.ucsc.edu>). Finally, the linkage disequilibrium (LD) information was provided by rolypoly based on the calculation using PLINK for 1000 g phase 3 genomes filtered for values of $R^2 > 0.2$.

(<https://cran.r-project.org/web/packages/rolypoly/vignettes/intro.html>). We input all these data into the main rolypoly function call using the default parameters. Significant association between cell type and CKD or eGFR was determined by p value.

Pseudotemporal ordering

We used Monocle2 (Qiu et al., 2017) (default settings) to draw a minimal spanning tree connecting the 9,190 cells collected from time course organoids into multiple lineages. As input into Monocle2, we selected the ordering genes using a semi-supervised approach as described in the Monocle2 tutorial (http://cole-trapnell-lab.github.io/monocle-release/docs_mobile/). Specifically, we first defined the genes that mark the kidney organoid differentiation process and terminal cell lineage. The criteria for the gene selection include: 1) they are putatively reported as markers for the kidney developmental state or terminally kidney cell fate. 2) they are specific marker genes differentially expressed in the clusters identified by Seurat as noted above. We used them to select the ordering genes that co-vary with these markers using the Monocle function *markerDiffTable*. With this approach, we generated a final gene set containing 1,604 genes. We then reduced the data space to two dimensions with '*DDRTree*' method and ordered the cells using the *orderCells* function in Monocle. Individual cells were color-coded based on the time points where they were collected to validate the accuracy of the cell ordering. We further colored the cells with the cell type identity classified by Seurat as noted above to assure that the ordering is meaningful. We plotted the gene expression on the Monocle tree using MAGIC imputed value as described above.

Ligand-receptor interaction analysis

To study the ligand-receptor interaction at single cell level, we used a human ligand–receptor list comprising 2,557 ligand–receptor pairs curated by Database of Ligand–Receptor Partners (DLRP), IUPHAR and Human Plasma Membrane Receptome (HPMR) (Hrvatin et al., 2018; Ramilowski et al., 2015). We selected the receptors that were only differentially expressed in each cell type from the day 26 organoid (Takasato protocol), and the ligands that were induced in either branch on the cell trajectory based on the results computed by the branched expression analysis modeling (BEAM) algorithm in Monocle. To determine the ligand-receptor pairs to plot on the heatmap, we required that (i) The ligands are branch-dependent with significant score q-val < 0.01 based on the score table from BEAM analysis; (ii) The receptors are uniquely expressed in each cell type (q-val < 0.05 and logFC > 0.6); (iii) Each receptor should have at least one corresponding ligand to pair with. We used *plot_genes_branched_heatmap* function from the Monocle2 package to plot the ligands and heatmap.2 function from gplots package to show the receptors expressed in each cell type. We used the same list of ligand-receptor pairs to identify cell-type specific ligands and receptors in human adult kidney and employed a network approach to visualize the cell type communication based on connections of ligand-receptor pairs (Ramilowski et al., 2015).

Real Time PCR data quantification and statistical analysis

Data were presented as mean ± SEM. ANOVA with post hoc Bonferroni correction was used for multiple group comparison. Student t test was used to compare 2 different groups. Graph-Pad Prism software, version 6.0c (GraphPad Software Inc., San Diego, CA) and SPSS version 22 were used for statistical analysis. P value < 0.05 was considered as statistical significant difference.

Experimental Design

The number of replicate organoids is included in Table S1. Experimental groups were neither randomized nor blinded. No sample-size estimates were made and the inclusion criteria for analyzing adult human kidney was having normal kidney function and no known kidney disease.

DATA AND SOFTWARE AVAILABILITY

The accession number for the RNA sequencing data reported in this paper is NCBI GEO: GSE118184. Additional figures can be assessed at Mendeley Data: <https://doi.org/10.17632/m4rf9wb29.1>



Single-cell transcriptomics reveals multiple neuronal cell types in human midbrain-specific organoids

Lisa M. Smits¹ · Stefano Magni¹ · Kaoru Kinugawa² · Kamil Grzyb¹ · Joachim Luginbühl³ · Sonia Sabate-Soler¹ · Silvia Bolognin¹ · Jay W. Shin³ · Eiichiro Mori² · Alexander Skupin^{1,4} · Jens C. Schwamborn¹

Received: 23 January 2020 / Accepted: 22 June 2020 / Published online: 31 July 2020
© The Author(s) 2020

Abstract

Human stem cell-derived organoids have great potential for modelling physiological and pathological processes. They recapitulate in vitro the organization and function of a respective organ or part of an organ. Human midbrain organoids (hMOs) have been described to contain midbrain-specific dopaminergic neurons that release the neurotransmitter dopamine. However, the human midbrain contains also additional neuronal cell types, which are functionally interacting with each other. Here, we analysed hMOs at high-resolution by means of single-cell RNA sequencing (scRNA-seq), imaging and electrophysiology to unravel cell heterogeneity. Our findings demonstrate that hMOs show essential neuronal functional properties as spontaneous electrophysiological activity of different neuronal subtypes, including dopaminergic, GABAergic, glutamatergic and serotonergic neurons. Recapitulating these *in vivo* features makes hMOs an excellent tool for in vitro disease phenotyping and drug discovery.

Keywords Neural stem cells · Midbrain organoids · Neuronal subtypes · Single-cell RNA sequencing · Electrophysiological activity

Introduction

Current in vitro approaches to model physiology and pathology of human neurons are mainly based on pure cultures of neurons grown under 2D conditions. It has been shown that

Stefano Magni, Kaoru Kinugawa and Kamil Grzyb contributed equally to this work.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00441-020-03249-y>) contains supplementary material, which is available to authorized users.

Jens C. Schwamborn
jens.schwamborn@uni.lu

Kaoru Kinugawa
kinugawa_ko@naramed-u.ac.jp

¹ Luxembourg Centre for Systems Biomedicine (LCSB), Developmental and Cellular Biology, University of Luxembourg, Belvaux, Luxembourg

² Department of Future Basic Medicine, Nara Medical University, Kashihara, Nara, Japan

³ Division of Genomic Technologies, RIKEN Center for Life Science Technologies, Yokohama, Kanagawa, Japan

⁴ University California San Diego, La Jolla, CA, USA

the differentiation potential of human induced pluripotent stem cells (iPSCs) provides a unique source of different neural cell types (Takahashi and Yamanaka 2006). Until now, many protocols for generating iPSC-derived neural cultures have been described. The resulting cell culture monolayers have been proven as useful tools to study disease mechanisms and to identify potential neuroprotective compounds (Nguyen et al. 2011; Cooper et al. 2012; Sánchez-Danés et al. 2012; Reinhardt et al. 2013b; Ryan et al. 2013). However, these culture conditions do not recapitulate several characteristics, which are relevant to the human brain, like cyto-architecture or complex cell-cell interactions. This may result in inaccurate modelling of the human brain (patho-)physiology with the consequence that candidate compounds might prove efficacy in 2D in vitro studies but are ineffective in clinical trials or vice versa (Abe-Fukasawa et al. 2018). The recent establishment of new 3D neuronal cell culture models has contributed to mimic key aspects of human brain development (Lancaster et al. 2013; Tieng et al. 2014; Muguruma et al. 2015; Jo et al. 2016; Qian et al. 2016; Monzel et al. 2017). Studies using human cerebral brain organoids have shown the acquisition of neuronal maturity and network activity (Quadrato et al. 2017; Matsui et al. 2018). Their complex, multicellular architecture enables the study of neuronal diseases and has already led to novel insights on, e.g. Zika virus-induced microcephaly

(Ming et al. 2016; Qian et al. 2017). Besides this unique in vitro disease modelling potential, human brain organoids provide a platform for advanced drug screening (Kelava and Lancaster 2016; Di Lullo and Kriegstein 2017). In this study, we focused on a detailed characterization of the different neuronal subtypes in human midbrain-specific organoids (hMOs). With single-cell transcriptome analysis, we examined the presence of different neuronal subtypes, and subsequently studied the effect of chemical compounds on the electrophysiological activity of the neuronal network. Our findings demonstrate that hMOs contain, beside dopaminergic neurons, other neuronal subtypes including GABAergic, glutamatergic and serotonergic neurons. hMOs showed essential neuronal functional properties during the course of differentiation, like synapse formation and spontaneous electrophysiological activity. These features indicate that hMOs recapitulate specific characteristics of functional human midbrain tissue, thus making them a valuable resource for in vitro disease modelling and drug discovery.

Material and methods

Data availability

The data that support the findings of this study, including the original single-cell RNA Sequencing data, are publicly available at this doi: [www.doi.org/10.17881/lcsb.20190326.01](https://doi.org/10.17881/lcsb.20190326.01).

Furthermore, a previous version of this manuscript is available as pre-print under: <https://doi.org/10.1101/589598>.

Pluripotent stem cell culture

hiPSC lines were provided by Bill Skarnes, Wellcome Trust Sanger Institute (iPSC Bill), Alstem (iPS15, derived from human peripheral blood mononuclear cells, episomal reprogrammed) or previously described in Reinhard et alia (Reinhardt et al. 2013b). The cells were cultured on Matrigel-coated (Corning, hESC-qualified matrix) plates, maintained in Essential 8 medium (Thermo Fisher Scientific) and cultured with and split 1:6 to 1:8 every 4 to 5 days using Accutase (Sigma). Ten μ M ROCK inhibitor (Y-27632, Abcam) was added to the media for 24 h following splitting.

Derivation of midbrain floorplate neural progenitor cells

The derivation and maintenance of midbrain floorplate neural progenitor cells (mfNPCs) has been described previously (Smits et al. 2019).

In brief, embryoid bodies (EBs) were formed with 2000 iPSCs each, using AggreWell 400 (Stemcell Technologies).

The cells were cultured in Knockout DMEM (Invitrogen) with 20% Knockout Serum Replacement (Invitrogen), 100- μ M beta-mercaptoethanol (Gibco), 1% nonessential amino acids (NEAA, Invitrogen), 1% penicillin/streptomycin/glutamine (Invitrogen), freshly supplemented with 10- μ M SB-431542 (SB, Ascent Scientific), 250-nM LDN-193189 (LDN, Sigma), 3- μ M CHIR99021 (CHIR, Axon Medchem), 0.5- μ M SAG (Merck) and 5- μ M ROCK inhibitor (Sigma). After 24 h, EBs were transferred to a non-treated tissue culture plate (Corning). On day two, medium was replaced with N2B27 medium consists of DMEM-F12 (Invitrogen)/Neurobasal (Invitrogen) 50:50 with 1:200 N2 supplement (Invitrogen), 1:100 B27 supplement lacking vitamin A (Invitrogen) with 1% penicillin/streptomycin/glutamine, supplemented with 10- μ M SB, 250-nM LDN, 3- μ M CHIR and 0.5- μ M SAG. On day four and six, medium was exchanged with the same but including 200- μ M ascorbic acid (AA, Sigma). On day eight, EBs with neuroepithelial outgrowth were triturated into smaller pieces and diluted in a 1:10 ratio. For following passages, 1× TrypLE Select Enzyme (Gibco)/0.5-mM EDTA (Invitrogen) in 1× PBS was used and 10,000 to 20,000 cells per 96-well ultra-low attachment plate (round bottom, Corning) were seeded. The cells were always kept under 3D culture conditions and from passage 1 on cultured in N2B27 medium freshly supplemented with 2.5- μ M SB, 100-nM LDN, 3- μ M CHIR, 200- μ M AA and 0.5- μ M SAG. After every cell split, the ultra-low attachment plate was centrifuged for 3 min at 200×g to assure the aggregation of single cells at the bottom of the well. Additionally, a 5- μ M ROCK inhibitor was added. The cells were split every 7 to 14 days and the medium was changed every third day. After four to five passages, mfNPCs were used as a starting population for hMOs.

Generation of midbrain-specific organoids

To start the generation of hMOs, 3000 cells per well were seeded to an ultra-low attachment 96-well round bottom plate, centrifuged for 3 min at 200×g and kept under maintenance conditions for 7 days. LDN and SB were withdrawn of mfNPC expansion medium and after three additional days, the concentration of CHIR was reduced to 0.7 μ M. On day nine of differentiation, medium was changed to neuronal maturation N2B27 medium including 10-ng/ml BDNF (Peprotech), 10-ng/ml GDNF (Peprotech), 200- μ M AA (Sigma), 500- μ M dbcAMP (Sigma), 1-ng/ml TGF- β 3 (Peprotech), 2.5-ng/ml ActivinA (Life Technologies) and 10- μ M DAPT (Cayman). The organoids were kept under static culture conditions with media changes every third day for 35 or 70 days. Detailed information about the generation of hMOs has been published recently (Smits et al. 2019).

Immunofluorescence

hMOs were fixed with 4% PFA overnight at 4 °C and washed 3× with PBS for 15 min. After treatment, they were embedded in 3–4% low melting point agarose in PBS. The solid agarose block was sectioned with a vibratome (Leica VT1000s) into 50 or 70-μm sections. The sections were blocked on a shaker with 0.5% Triton X-100, 0.1% sodium azide, 0.1% sodium citrate, 2% BSA and 5% normal goat or donkey serum in PBS for 90 min at RT. Primary antibodies were diluted in the same solution but with only 0.1% Triton X-100 and were applied for 48 h at 4 °C.

After incubation with the primary antibodies (Supplementary Table 2), sections were washed 3× with PBS and subsequently blocked for 30 min at RT on a shaker. Then sections were incubated with the secondary antibodies in 0.05% Tween-20 in PBS for 2 h at RT and washed with 0.05% Tween-20 in PBS and Milli-Q water before they were mounted in Fluoromount-G mounting medium (Southern Biotech).

STAINperfect Immunostaining Kit (ImmunoSmol) was used according to manufacturer's protocol to detect dopamine, serotonin, GABA and L-glutamine. Nuclei were counterstained with Hoechst 33342 (Invitrogen).

For qualitative analysis, three randomly selected fields per organoid section were acquired with a confocal laser scanning microscope (Zeiss LSM 710) and images were further processed with OMERO Software. Three-dimensional surface reconstructions of confocal z-stacks were created using Imaris software (Bitplane).

Quantitative image analysis

Immunofluorescence 3D images of hMOs were analysed in Matlab (Version 2017b, Mathworks). The in-house developed image analysis algorithms automate the segmentation of nuclei, astrocytes and neurons with structure-specific feature extraction. The image preprocessing for the segmentation of nuclei was computed by convolving the raw Hoechst channel with a Gaussian filter. By selecting a pixel threshold to identify apoptotic cells, a pyknotic nuclei mask was identified and subtracted from the nuclei mask.

For the segmentation of neurons, a median filter was applied to the raw TUJ1 channels. The expression levels were expressed in two ways as follows: (i) positive pixel of the marker, normalized by the pixel count of Hoechst; (ii) cells positive for a marker expressed as a percentage of the total number of cells. In this latter case, the nuclei were segmented and a watershed function was applied. Considering the high cell density of the specimens, steps to ensure high quality in the segmentation process were implemented and structures with a size higher than 10,000 pixels were removed (this indicated incorrect segmentation, e.g. clumps). In the nuclei

successfully segmented as a single element, a perinuclear zone was identified. In case the marker of interest was positive in at least 1% of the perinuclear area, the corresponding cell was considered as positive.

Single-cell RNA sequencing using droplet-sequencing (Drop-Seq)

Single-cell RNA sequencing (scRNA-seq) data were generated using the Droplet-Sequencing (Drop-Seq) technique (Macosko et al. 2015) as described previously (Walter 2019). In this work, we performed scRNA-seq of hMOs derived from hiPSC line H4 (see Supplementary Table 1). For each time point, 35 days and 70 days after dopaminergic differentiation, we pooled and analysed 30 hMOs each.

Pre-processing of the digital expression matrices from scRNA-seq

The result of the Drop-Seq scRNA-seq pipeline and subsequent bioinformatics processing is a digital expression matrix (DEM) representing the number of mRNA molecules captured per gene per droplet. Here, we obtained two DEMs, one corresponding to 35-day hMOs and the other to 70-day hMOs. After quality cut based on knee plots, we retained for each sample 500 cells with the highest number of total transcripts measured and performed normalization of the DEM separately. Finally, the two DEMs were merged for the comparison analysis of the two time points based on 24,976 expressed genes in 1000 cells. The data was analysed by our customized Python analysis pipeline (Python version 3.6.0, with anaconda version 4.3.1) including dimensionality reduction by t-distributed stochastic neighbourhood embedding (t-SNE) (van der Maaten and Hinton 2008) and differential gene expression analysis.

Analysis of differentially expressed genes from scRNA-seq data

To determine which and how many genes were differentially expressed between 35-day and 70-day hMOs, we applied a one-way ANOVA test, a one-way ANOVA test on ranks (Kruskal-Wallis test), and a Mutual Information based test. The minimum *p* value obtained for each gene across these three tests was retained and statistical significance was set to *p* < 0.01 after Bonferroni correction for multiple hypothesis testing of differentially expressed genes (DEGs).

Cumulative gene expressions from scRNA-seq data

From literature, we extracted cell type-specific gene lists (Supplementary Table 3) for stem cells, neurons and neuronal subtypes (dopaminergic, glutamatergic, GABAergic and

serotonergic neurons) (Reinhardt et al. 2013a; La Manno et al. 2016; Cho et al. 2017). Note that not all genes listed therein have been measured in our dataset; these were highlighted in Supplementary Table 3.

For each list, we defined a score, which we refer to as cumulative gene expression, computed as the sum of the expression of the corresponding genes from normalized DEM for each cell. Since the expression levels were measured at single cell level, we can consider the cells' distributions across the cumulative genes expression scores (Fig. 2a). These histograms exhibit the cumulative gene expression scores normalized to their maxima on the horizontal axis. Thus, on the horizontal axis, a value of 1 corresponds to the maximal cumulative gene expression for one list of genes, while 0 corresponds to no expression of any genes from that list. The vertical axis exhibits the number of cells falling into the corresponding bin of the histogram. In each subpanel, the distributions for day 35 and for day 70 are shown. Population differences were assessed by Z-test of the means with Bonferroni correction.

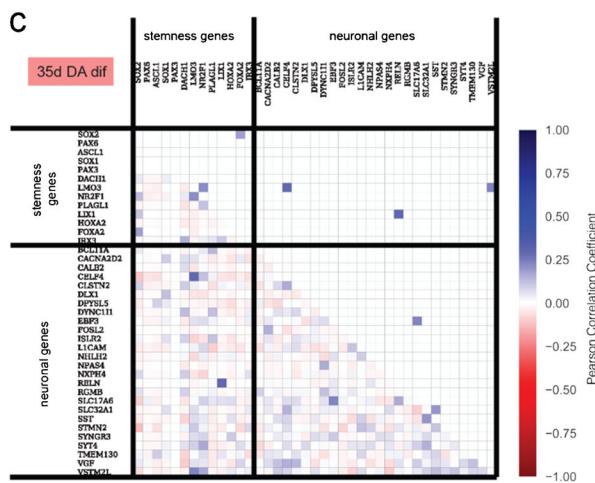
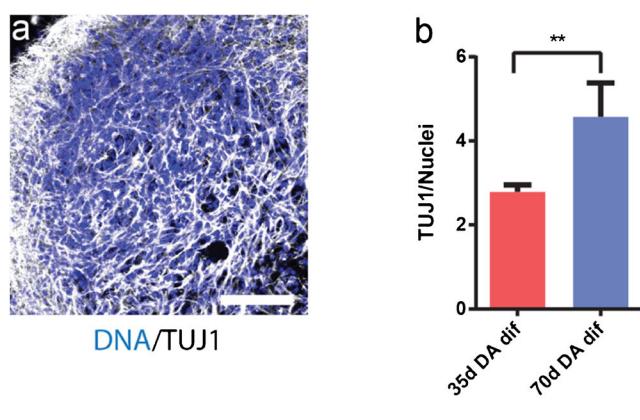


Fig. 1 Identification of neuronal population in midbrain-specific organoids. (a) Immunohistological staining of TUJ1 expressing neurons in 35-day organoid sections (50- μ m thickness, scale bar 100 μ m). (b) The ratio of TUJ1 positive pixels normalized against Hoechst (35 days $n = 59$, 70 days $n = 48$). (b') Quantification of Hoechst positive pixel (35 days

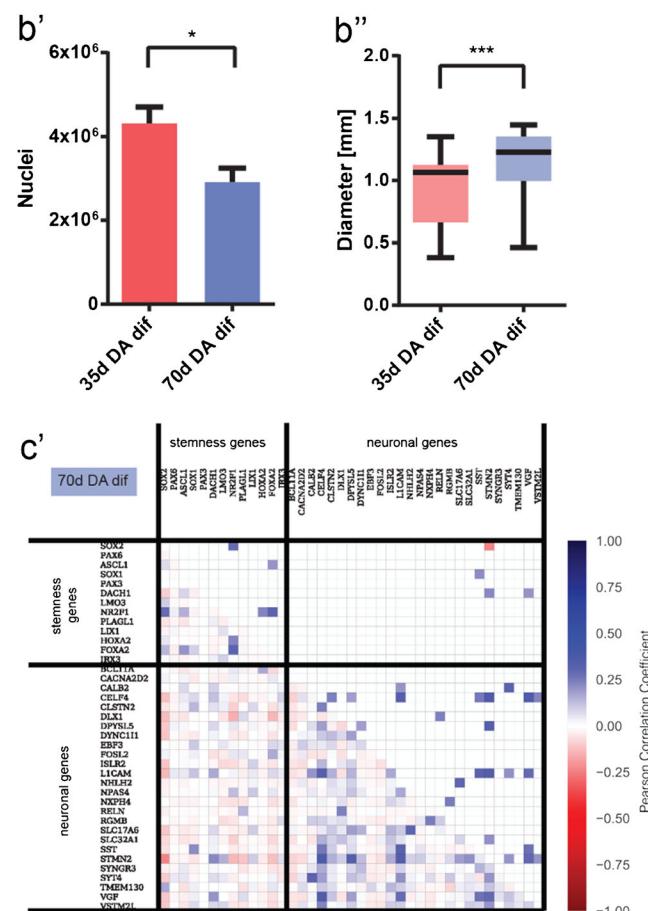
Gene-gene correlations from scRNA-seq data

From the scRNA-seq data, we also computed gene-gene Pearson correlation coefficients for stemness-specific and neuron-specific genes. Analysis was performed independently for the two samples (35-day DA dif and 70-day DA dif) resulting in two correlation matrices (Fig. 1c).

In the lower triangular matrix, all correlation values are shown, whereas the upper triangular matrix contains only statistically significant correlations (p value < 0.05 after Bonferroni correction). For visual clarity, diagonal elements and undetected genes were excluded.

Fold changes of gene expression from scRNA-seq data

For individual genes, we considered the normalized gene expression across cell populations. For each selected gene, we compared its expression within the 35-day cells with the one within the 70-day cells by computing the logarithmic fold change (log2FC). We performed this analysis for the genes



$n = 22$, 70 days $n = 29$. (b'') Average size of four different organoid lines. Whiskers present minimum and maximum (35 days $n = 21$, 70 days $n = 44$). Data presented as mean \pm SEM. (c) Gene-gene correlation matrices, for genes at day 35 (c), and day 70 (c'')

specific of neuronal subtypes including glutamatergic neurons, GABAergic neurons and dopaminergic neurons (Fig. 2c–d), where negative values indicate that a gene is less expressed at day 35 than at day 70 and positive numbers the opposite. *p* values are based on Z-test with Bonferroni correction and significance levels correspond to * = *p* value < 0.05, ** = *p* value < 0.01, *** = *p* value < 0.001, and **** = *p* value < 0.0001. Error bars represent SEM based on the individual sample average and error propagation.

scRNA-seq data analysis for UMAP plot, dot plot and violin plot

ScRNA-seq data were generated using the Droplet-Sequencing (Drop-Seq) technique (1). After bioinformatics

processing, we obtained two digital expression matrices (DEM), corresponding to day 35 and day 70 after differentiation into human midbrain organoids (hMOs).

In an alternative analysis approach, which is independent and complementary to the scRNA-Seq analysis described above, further data processing was performed using the Seurat v.3.0.0 R package (Satija et al. 2015). Cells with more than 4000 or less than 500 detected genes, as well as those with mitochondrial transcripts proportion higher than 7.5% were excluded. We collected a total of 1295 cells (505 cells at day 35 and 790 cells at day 70). The datasets were log normalized and scaled to 10,000 transcripts per cells. The top 2000 highly variable genes for day 35 and day 70 were determined using the variance-stabilizing transformation method. The datasets from day 35 and day 70 were integrated

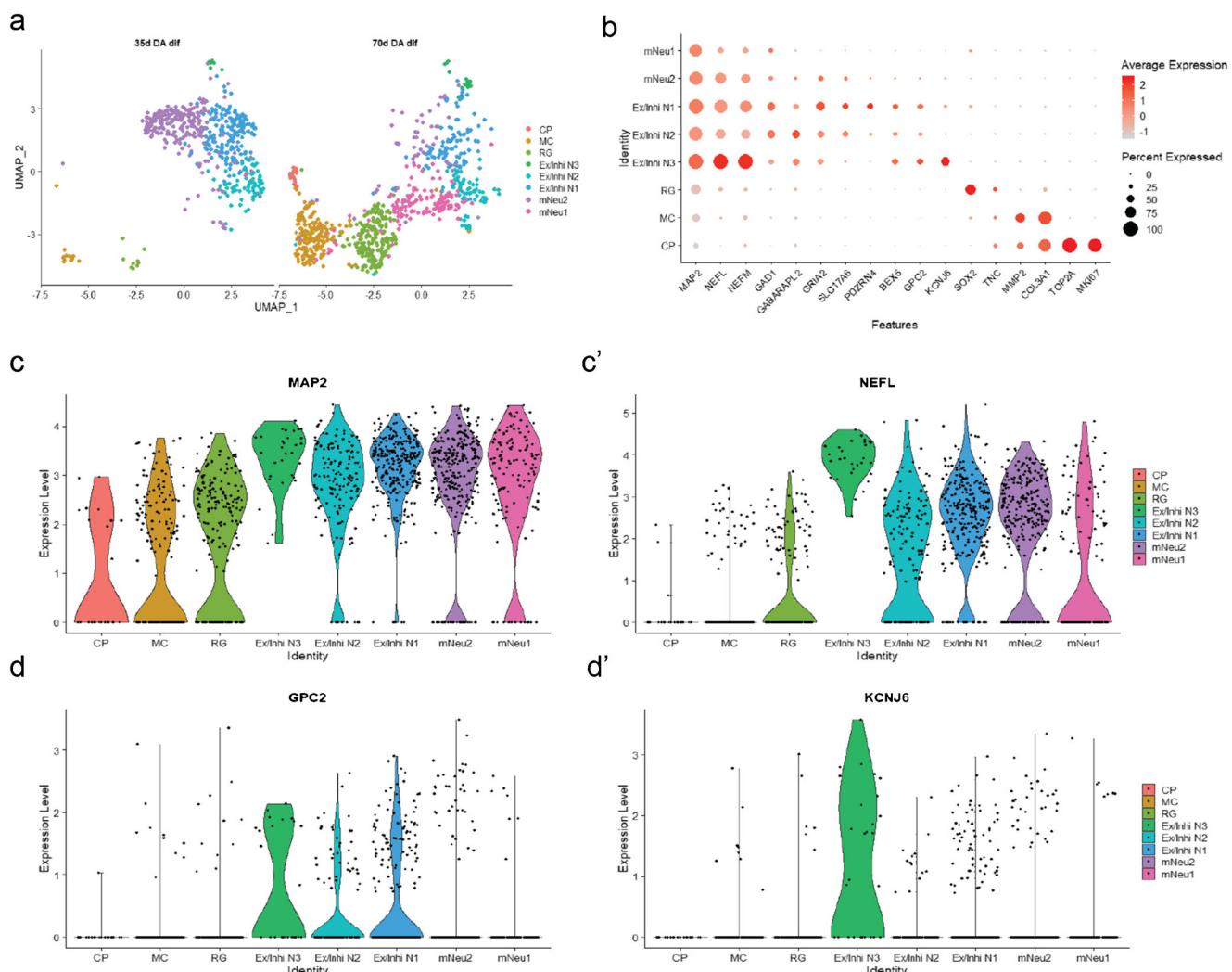


Fig. 2 Single-cell RNA sequencing analysis of midbrain-specific organoids. (a) Uniform manifold approximation and projection (UMAP) plot shows that the total 1295 cells identified 8 cell populations. Each dot corresponds to a single cell. Cell populations are coloured and annotated based on their expressing genes. CP, cycling progenitors; MC, mesenchymal cells; RG, radial glia cells; ExInhi N, excitatory/inhibitory

neurons; mNeu, mature neurons. UMAP plots shows difference in gene expressions between day 35 and day 70. (b) Dot plot shows the expression of each cell type-specific genes. (c–d) Violin plots show the distribution of expression of each marker gene. The mNeu clusters expressed MAP2 (c) and NEFL (c'), the ExInhi N clusters expressed GPC2 (d) and KCNJ6 (d')

using canonical correlation analysis (CCA) in the Seurat package (Stuart et al. 2019). The datasets were integrated based on the top 30 dimensions from CCA using the Seurat method by identifying anchors and integrating the datasets. The resulting integrated data were scaled and principal component analysis (PCA) was performed. Clustering was performed based on the top 30 principal components (PCs), using the shared nearest neighbour (SNN) modularity optimization with a resolution of 0.8. Cluster identities were assigned based on cluster gene markers as determined by the “FindAllMarkers” function in Seurat and gene expression of known marker genes.

TEM morphology

Sixty-three-day-old hMO specimens were immersion-fixed in a solution of 2% PFA and 2.5% glutaraldehyde in 0.1-M sodium cacodylate buffer (pH 7.4, Electron Microscopy Sciences, Hatfield, PA) for 3 h, rinsed several times in cacodylate buffer and further post-fixed in 2% glutaraldehyde in 0.1-M sodium cacodylate buffer for 2 h at room temperature on a gentle rotator; fixative was allowed to infiltrate an additional 48 h at 4 °C. Specimens were rinsed several times in cacodylate buffer, post-fixed in 1.0% osmium tetroxide for 1 h at room temperature and rinsed several times in cacodylate buffer. Samples were then dehydrated through a graded series of ethanol to 100% and dehydrated briefly in 100% propylene oxide. Tissue was then allowed to pre-infiltrate 2 h in a 2:1 mix of propylene oxide and Eponate resin (Ted Pella, Redding, CA), then transferred into a 1:1 mix of propylene oxide and Eponate resin and allowed to infiltrate overnight on a gentle rotator. The following day, specimens were transferred into a 2:1 mix of Eponate resin and propylene oxide for a minimum of 2 h, allowed to infiltrate in fresh 100% Eponate resin for several hours, and embedded in fresh 100% Eponate in flat moulds; polymerization occurred within 24–48 h at 60 °C. Thin (70 nm) sections were cut using a Leica EM UC7 ultramicrotome, collected onto formvar-coated grids, stained with uranyl acetate and Reynold’s lead citrate and examined in a JEOL JEM 1011 transmission electron microscope at 80 kV. Images were collected using an AMT digital imaging system with proprietary image capture software (Advanced Microscopy Techniques, Danvers, MA).

Microelectrode array

The Maestro microelectrode array (MEA, Axion BioSystems) platform was used to record spontaneous activity of the hMOs. A 48-well MEA plate containing a 16-electrode array per well was precoated with 0.1-mg/ml poly-D-lysine hydrobromide (Sigma-Aldrich). Sixty to seventy days old organoids of two different passages were briefly treated for 5 min with 1× TrypLE Select Enzyme, resuspend in 10 µg/ml laminin (Sigma-Aldrich) and placed as a droplet onto the

array. After 1 h incubation, neuronal maturation media was added and cells were cultured for 1–2 weeks. Spontaneous activity was recorded at a sampling rate of 12.5 kHz for 5 min at 37 °C over several days. Axion Integrated Studio (AxIS 2.1) was used to assay creation and analysis. A Butterworth band pass filter with 200–3000 Hz cutoff frequency and a threshold of 6× SD were set to minimize both false positives and missed detections. The spike raster plots were analysed using the Neural Metric Tool (Axion BioSystems). Electrodes with an average of ≥ 5 spikes/min were defined as active, for the pharmacological treatment 24 electrodes were analysed. The organoids were consecutively treated with Gabazine, D-AP-5, NBQX (Cayman Chemical, end concentration: 50 mM each), and Quinpirole (Sigma Aldrich, end concentration: 5 µM). To block all neuronal activity and thus verify spontaneous spiking activity of the cells, tetrodotoxin (TTX, Cayman Chemical, 1 µM) was applied at the end. The spike count files generated from the recordings were used to calculate the number of spikes/active electrode/min. Further details regarding the MEA system were previously described (Bardy et al. 2015).

Statistical analyses

If not stated otherwise, experiments were performed with three independently generated organoid cultures from three different cell lines ($n = 9$). Gaussian distribution was evaluated by performing D’Agostino and Pearson omnibus normality test. In case the data were normally distributed, Grubbs’ test was performed to detect significant outliers. Unpaired *t* test with Welch’s correction or nonparametric Kolmogorov-Smirnov test was performed to evaluate statistical significance. Data are presented as mean ± SEM. The statistical analyses of scRNA-seq data are described in the corresponding sections.

Results

Characterization of the neuronal differentiation dynamics in midbrain-specific organoids

Previously, we demonstrated that human iPSC-derived midbrain floor plate neural progenitor cells (mfNPCs) can give rise to 3D human organoids that contain high amounts of dopaminergic neurons (Smits et al. 2019). To have a better insight into the dynamics of the neuronal differentiation, we evaluated TUJ1 staining, as a marker for neuronal differentiation, at two time points during the differentiation of hMOs (Fig. 1 a and b). An in-house developed image analysis algorithm was used to segment Hoechst-positive nuclei and TUJ1-positive neurons to create specific nuclear and neuronal

masks. These masks contain all positive pixel counts for Hoechst and TUJ1, respectively.

The TUJ1 signal normalized to the Hoechst signal significantly increased after 70 days compared with 35 days, demonstrating a progressive differentiation into post-mitotic neurons. Whereas, the nuclear marker signal was significantly decreased at 70 days compared with 35 days, which might indicate selection in the cell population, as reported by Suzanne and Steller (2013) (Fig. 1 b and b'). Along with these findings, we observed that the size of the organoids significantly increased during the course of the differentiation. This suggests that the increased TUJ1 volume and organoid size are due to the increased tissue complexity (e.g. neuronal arborisation) within the hMO (Fig. 1b").

To further characterize the neuronal differentiation dynamics at the gene expression level, we performed scRNA-seq on samples from the two time points mentioned above. The experiments were conducted using the Drop-Seq technique (Macosko et al. 2015), and the standard bioinformatics processing of the data resulted in two sample-specific digital expression matrices (DEM), which were further normalized and merged (see Methods section).

To investigate how the differentiation of precursor cells into neurons evolves over time, we computed the gene-gene correlation for the genes of the neuron-specific list and of the stemness-specific list, altogether. Comparing these two lists, we found that at 35 days there are low values of correlation between genes exclusively specific for neurons or stem cells and also between neuron- and stemness-specific genes (Fig. 1c). Very few of the correlation values are significantly different from zero and were substituted by zeros in the upper triangular matrix (Fig. 1c). While correlations between stemness genes and neuron-stemness correlations at day 70 remain similar to day 35, correlations between neuron-specific genes increased considerably at day 70. This significant increase of neuron-specific gene correlations indicates a higher commitment of the cells towards the neuronal fate at day 70 compared with day 35 and supports the finding of a progressive maturation of post-mitotic neurons (Fig. 1 c and c').

To visualize the so-obtained high-dimensional single-cell data, we performed dimensionality reduction of the DEM by uniform manifold approximation and projection (UMAP) (van der Maaten and Hinton 2008), where each dot corresponds to a cell (Fig. 2a). After processing, quality control and filtering, we analysed a total of 1295 cells, 505 cells at day 35 and 790 cells at day 70 (Fig. S1 a and b). To identify distinct cell populations based on shared and unique patterns of gene expression, we performed dimensionality reduction and unsupervised cell clustering. All 1295 cells from the two time points were analysed together and plotted onto two uniform manifold approximation and projection (UMAP) plots (Fig. 2a). We identified eight distinct cell populations

expressing known markers of major cell types (Fig. 2b). The cell populations comprised five neuronal clusters and three non-neuronal clusters. The five neuron clusters were divided into two mature neuron cluster (mNeu1 and mNeu2) and three excitatory/inhibitory neuron clusters (Ex/Inhi N1, Ex/Inhi N2, Ex/Inhi N3). The mNeu clusters expressed MAP2, NEFL and NEFM (Fig. 2 b, c and c'). The Ex/Inhi N clusters expressed BEX5, GPC2, KCNJ6, PDZRN4, GRIA2, SLC17A6, GAD1 and GABARAPL2 (Fig. 2 b, d and d'). Non-neuronal clusters were divided into radial glia cells (RG), mesenchymal cells (MC) and cycling progenitors (CP) subtypes (Fig. 2 b, S2 a and b). Cell populations distinctly changed between day 35 and day 70 (Fig. 2a). Interestingly the day 70 cell populations showed more non-neuronal clusters.

Midbrain-specific organoids consist of different neuronal subtypes

From previous studies, we know that hMOs are rich in dopaminergic neurons (Jo et al. 2016; Qian et al. 2016; Monzel et al. 2017; Smits et al. 2019; Kim et al. 2019). We wanted to further explore which other neuronal subtypes develop besides midbrain dopaminergic neurons within the hMOs.

Therefore, we investigated the expression of genes typical for dopaminergic, glutamatergic, GABAergic and serotonergic neurons by analysing the scRNA-seq data. We plotted the distributions of cells across the cumulative gene expression scores, which were obtained from the lists of genes specific of a neuronal subtype (Fig. 3a–a''). While the cell distribution over cumulative expression score for GABAergic neurons was very similar between the samples at 35 days and 70 days (Fig. 3a''), we detected statistically significant differences between the distributions of cells over scores for the other three types of neurons. The expression of the selected genes for the glutamatergic and dopaminergic neurons was increased at day 35 compared with day 70, which is consistent with the observations for the neuron specific score (Fig. 2a).

In order to additionally highlight the presence of dopaminergic, glutamatergic, GABAergic and serotonergic neurons, we made use of the UMAP plots and highlighted the expression of marker genes for dopaminergic, glutamatergic, GABAergic and serotonergic neurons (Fig. 4 and S3). In agreement with the fact that the fraction of non-neuronal cells is higher in the day 70 organoids (Fig. 1), the counts for the various neuronal subtypes tend to be higher at day 35.

To further verify the presence of the addressed neuronal subtypes, we conducted an immunohistochemistry staining for the respective neurotransmitters. This allowed us to robustly detect dopaminergic, glutamatergic and GABAergic neurons as well as even a few serotonergic neurons within hMOs (Fig. 5a–d).

Finally, among the non-neuronal cells, we surprisingly found a cluster of mesenchymal cells (MC in Fig. 1d). A

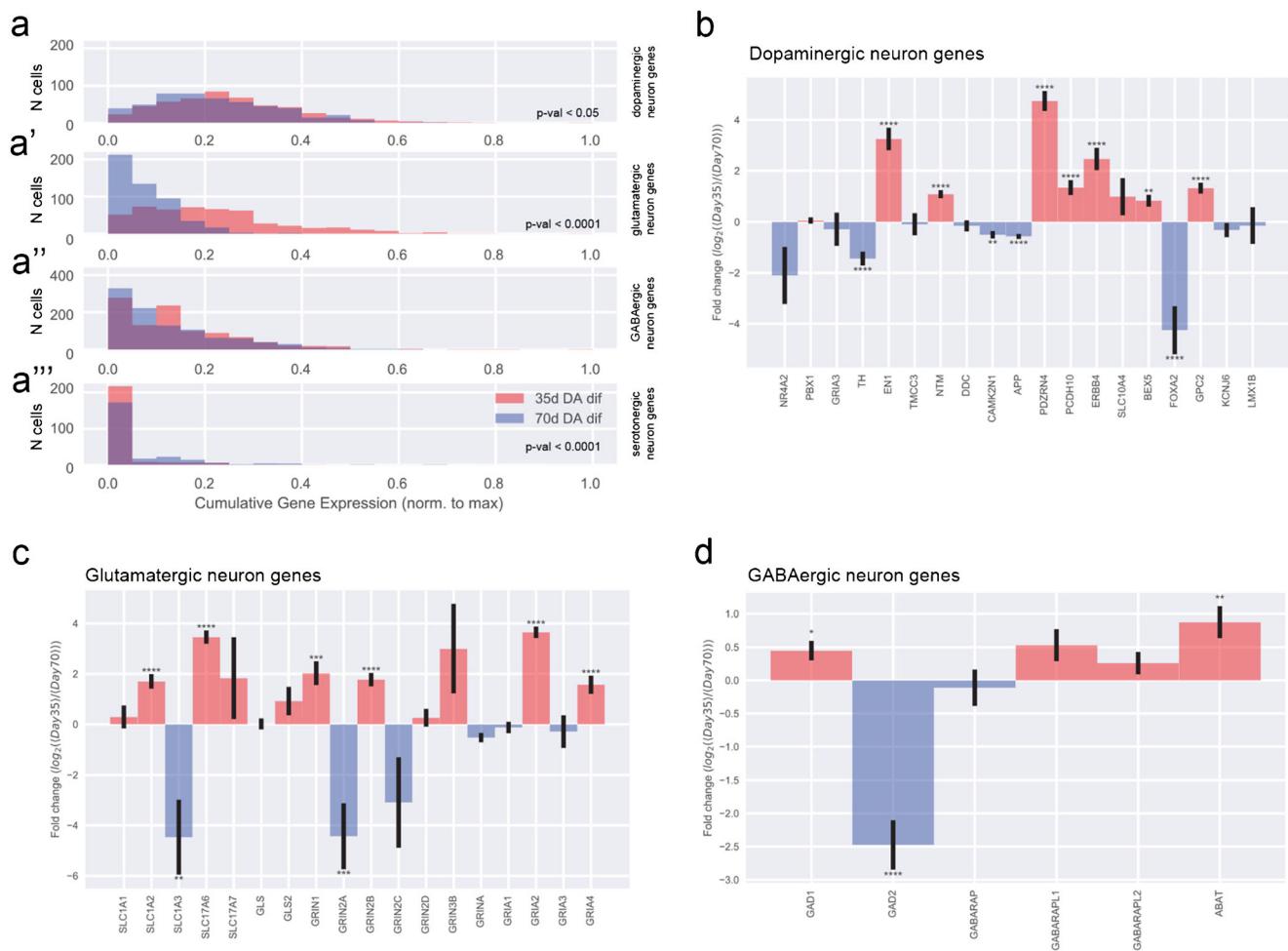


Fig. 3 Neuronal subtypes in midbrain-specific organoids. **(a)** Distributions (histograms) of cells across the cumulative gene expression scores, obtained from the lists of genes specific for the main neuronal subtypes present in the organoids, namely, dopaminergic (**a**), glutamatergic (**a'**), GABAergic (**a''**), and serotonergic neurons (**a'''**). **(b-d)** Log₂

fold-changes between day 35 and day 70 in gene expression for individual genes corresponding to the lists of genes typical of the neuronal subtypes: **(b)** dopaminergic neurons, **(c)** glutamatergic neurons, and **(d)** GABAergic neurons

similarly surprising finding has been reported recently for a single-cell RNA sequencing analysis of dopaminergic neurons transplanted in the rodent brain, where fibroblast-like cells were detected (Tiklova et al. 2019). These cells are reported to be positive for the marker COL1A1. Hence, we also stained hMOs with an anti-COL1A1 antibody and indeed we were able to confirm the presence of this cell type (Fig. 5e).

Midbrain-specific organoids express synaptic proteins

After identifying the presence of neurons and even specific neuronal subtypes on transcriptome expression levels by means of neurotransmitter staining and scRNA-seq, we investigated the actual interaction among the neuronal cells within the hMOs. We previously showed that hMOs synthesize and release the neurotransmitter dopamine (Smits et al. 2019). This already suggests the establishment of a

functional neuronal network. The basic requirement for neuronal network formation is the development of synapses. Hence, we evaluated the presence of synaptic connections using the presynaptic marker SYNAPTOPHYSIN and the postsynaptic marker PSD95 in organoid sections after 35 days and 70 days of culture (Fig. 6 a and b). Both proteins were detectable in a puncta-like organization, which is expected for synapses. With a subsequent 3D surface reconstruction, we observed that the signals for SYNAPTOPHYSIN and PSD95 were localized in close proximity, forming pre- and postsynaptic puncta (Fig. 3c). To further investigate whether actual functional synaptic connections were formed in the hMOs, we used a transmission electron microscopy (TEM) approach (Fig. 3d). EM micrographs show excitatory synapses characterized by electron dense post-synaptic density proteins (full arrow) and pre-synaptic synapse (asterisks) loaded with synaptic vesicles.

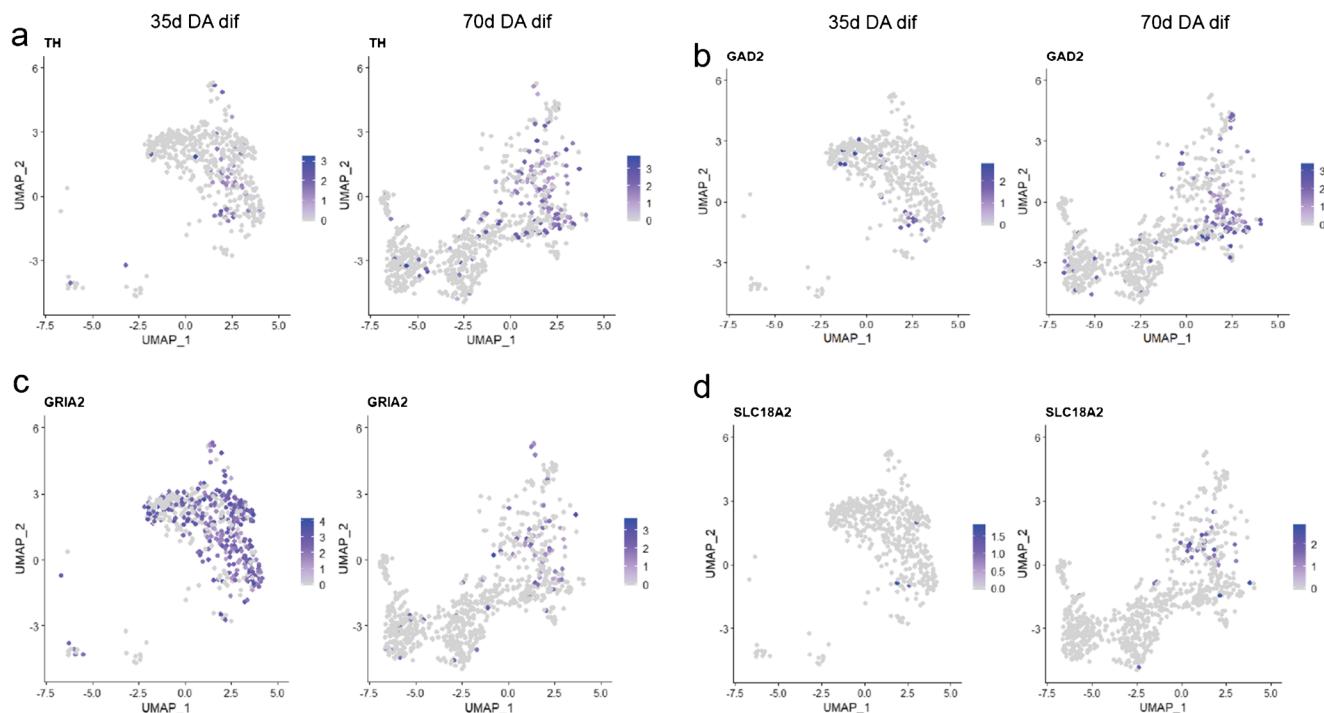


Fig. 4 UMAP plot analysis of neuronal subtypes. UMAP plots show the gene expressions at day 35 at day 70 for markers of dopaminergic (**a**), GABAergic (**b**), glutamatergic (**c**), and serotonergic (**d**) neurons. Each dot is coloured according to the expression level

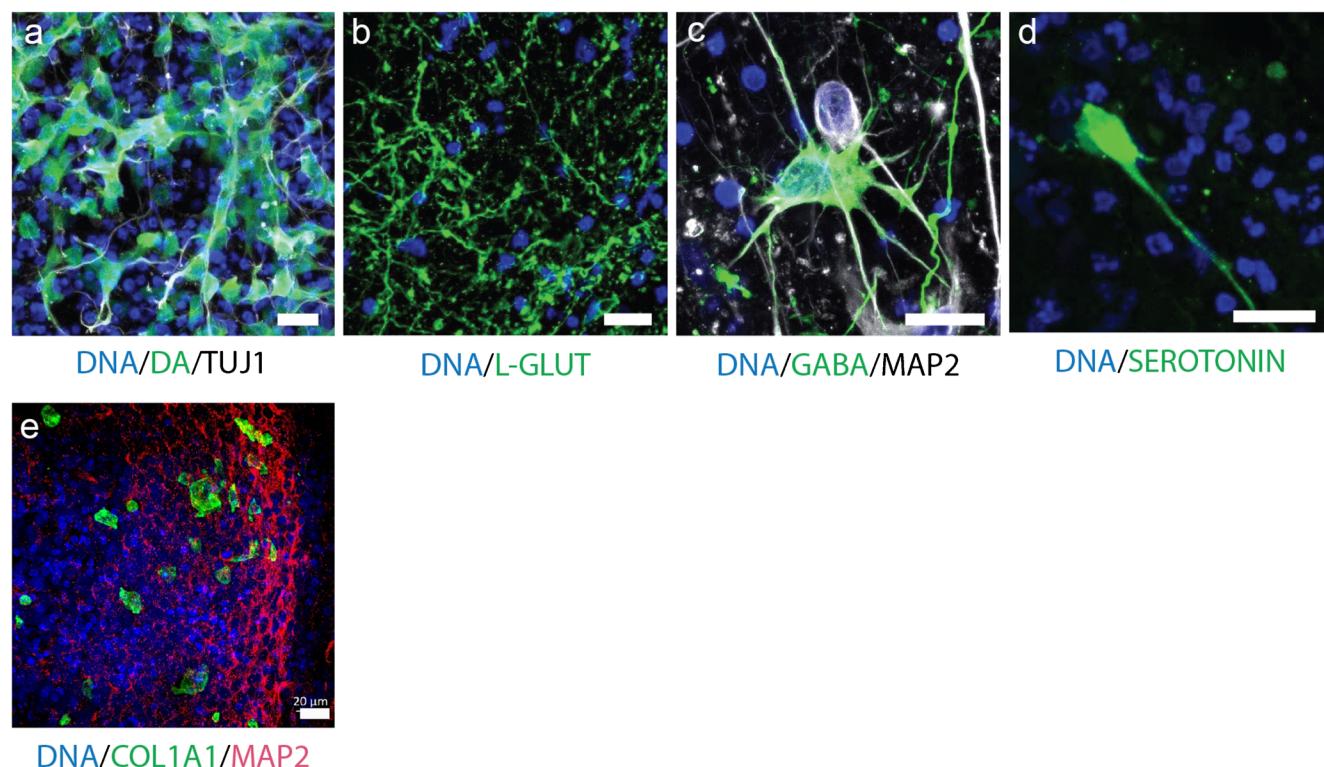


Fig. 5 Immunofluorescence staining analysis of neuronal subtypes. (**a–d**) Immunohistological staining of organoid sections (50-μm thickness). Detection of the neurotransmitters dopamine (**a**), L-glutamine (**b**),

GABA (**c**), and serotonin (**d**). Scale bar is 20 μm. (**e**) Immunohistological staining of organoid sections (70-μm thickness) for the detection of the fibroblast marker COL1A1. Scale bar is 20 μm

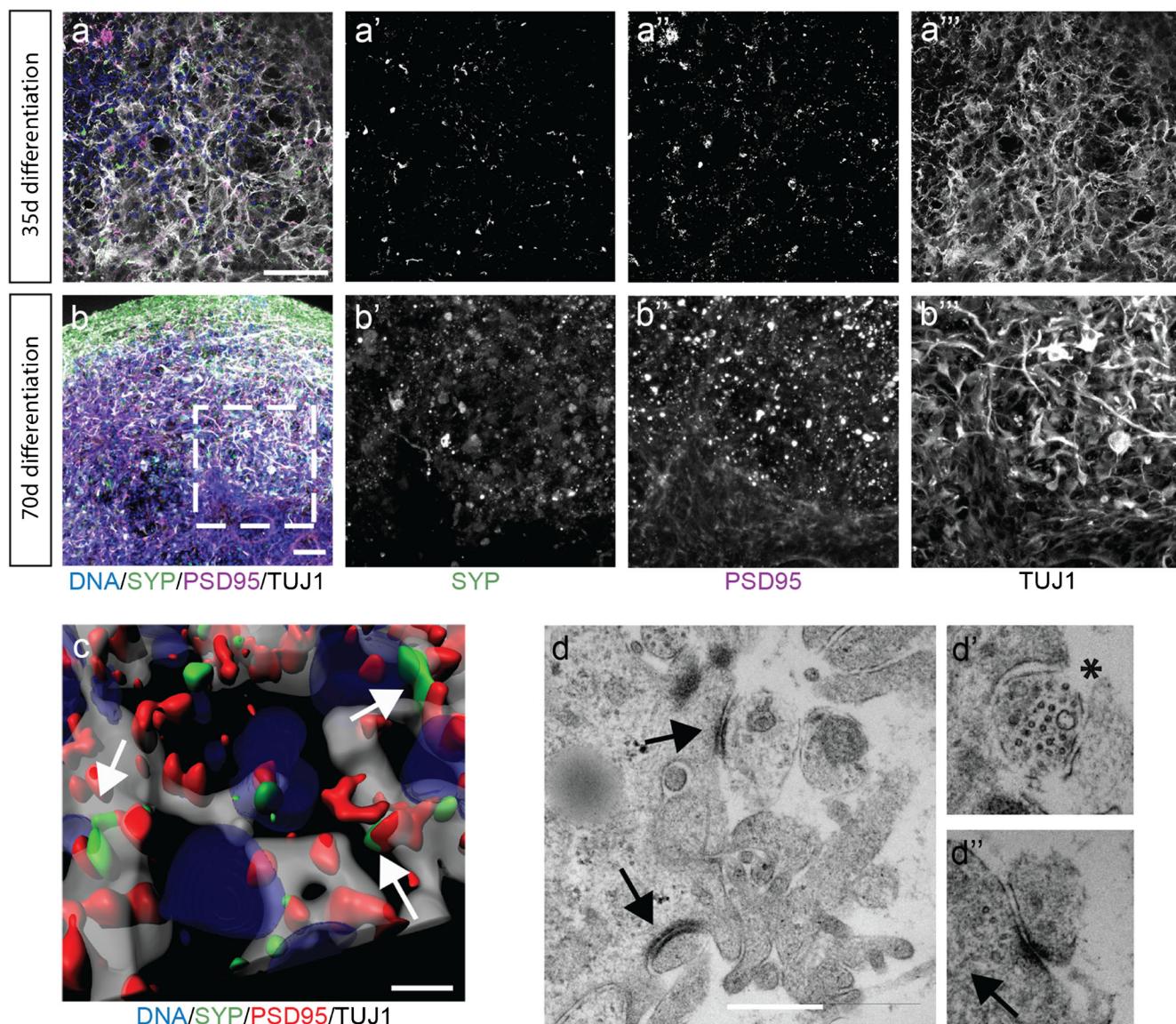


Fig. 6 Midbrain-specific organoids express synaptic proteins. **(a–b)** Immunostaining of pre- and postsynaptic markers at day 35 (**a**) and day 70 (**b**). Dashed lines indicate the region of magnification. Scale bar is 50 μ m. **(c)** 3D surface reconstructions of confocal z-stacks of an organoid

at day 70 of differentiation. Scale bar is 10 μ m. **(d)** Representative electron micrographs of synaptic terminals from 63-day organoids. Scale bar is 500 nm

Midbrain-specific organoids develop GABAergic, glutamatergic and dopaminergic electrophysiological activity

Non-invasive multielectrode array (MEA) measurements can give insights into physiological properties, like the generation of spontaneous neuronal activity of in vitro cultured, self-organized networks (Luhmann et al. 2016). As the assessment of neuronal activity is important to evaluate the functional maturation, we tested the spontaneous electrophysiological activity of hMOs by MEA measurements (Odawara et al. 2016). We measured extracellular field potentials, which are generated by action potentials. At days 50–60 of differentiation, hMOs were seeded in 48-well tissue culture MEA plates

on a grid of 16 electrodes (Fig. 7 a and b). After 10–20 days of culturing, we recorded spontaneous activity, on several electrodes, over several days, in the form of mono- and biphasic spikes (Fig. 7 a' and a''). To investigate which neuronal subtypes were functionally active in the hMOs, we applied specific drugs following a previously reported experimental design (Illes et al. 2014). We recorded spiking patterns from 24 active electrodes: in Fig. 7 c and d representative recordings of one electrode are displayed. After treating the organoids with gabazine, a GABA_A receptor antagonist, we detected an increase of spontaneous spiking (22.5% increase, Fig. 7d'). Following the gabazine-induced disinhibition, we applied the AMPA/Kainate-receptor antagonist NBQX and the NMDA-receptor antagonist D-AP-5. The inhibition of the

excitatory neurons resulted in a 28.1% decrease of spontaneous activity (Fig. 7d'). After the inhibition of GABAergic and glutamatergic neurons in the hMOs, we added the D2/D3 receptor agonist quinpirole (Fig. 7 c' and d''), which resulted in a 47.8% decrease of neuronal activity. Confirming the findings displayed in Fig. 2, we conclude from these experiments that hMOs contain functional GABAergic, glutamatergic and dopaminergic neurons.

Discussion

The in vitro human brain organoid technology has become a valuable tool allowing advances in the field of basic research as well as in translational applications (Fatehullah et al. 2016). Organoids specifically modelling the human midbrain hold great promise for studying human development and for modelling Parkinson's disease (PD) (Jo et al. 2016; Monzel et al. 2017; Kim et al. 2019; Smits et al. 2019). In contrast to 2D monolayer cultures, hMOs can recapitulate complex interactions of midbrain dopaminergic neurons with other cell types of the central nervous system (CNS) in a 3D environment. However, human midbrain organoid research has so far focused mainly on dopaminergic neurons. In a detailed study of Borroto-Escuela et al. (2018), it has been described that released dopamine can diffuse into synaptic regions of glutamate and GABA synapses and directly affect other striatal cell types possessing dopamine receptors. Furthermore, *substantia nigra* dopaminergic neurons are directly controlled by GABAergic input (Tepper and Lee 2007). Evidences from these studies suggest that the presence of other neuronal subtypes is important to model multifactorial disease like PD. In our study, we have demonstrated that the derivation of hMOs leads to functional neuronal networks, containing different neuronal subtypes of the human midbrain. Single-cell transcriptomic data from hMOs demonstrated that there is an increased expression of neuronal-specific genes in 35 days compared with 70 days old hMOs. On the other hand, the gene-gene correlations between only neuron-specific genes increased considerably at day 70, suggesting an increased commitment of cells towards the neuronal cell fate during the course of the organoid development. This further supports the finding of a progressive maturation of post-mitotic neurons (Fig. 1c). The identification of these neuron-specific genes revealed that the genes upregulated at the earlier time point are relevant in the process of neurogenesis and neuronal migration and differentiation (EBF3 (Garcia-Dominguez et al. 2003), L1CAM (Patzke et al. 2016)). Whereas the upregulated genes at the later time point have been for instance implicated in subpopulations like GABAergic neurons (DLX1, CALB2 (Al-Jaberi et al. 2015)). This indicates a higher commitment of the

cells towards their intended fate and a progressive maturation of the post-mitotic neurons within the hMOs. Furthermore, single-cell analysis of the hMOs also proved the presence of specific neuronal subtypes, like dopaminergic, glutamatergic, GABAergic and serotonergic neurons. Supporting the findings of currently published midbrain-specific organoid models (Jo et al. 2016; Qian et al. 2016; Monzel et al. 2017; Smits et al. 2019), we detected a significant upregulation of tyrosine hydroxylase (TH) within the cell population of 70 days old hMOs compared with 35 days old hMOs.

In the here presented data, we see a strong underrepresentation of neurons among recovered cells in the scRNA-seq data (see, e.g. data for 70 days of differentiation in Fig. 2a). However, this has been seen by others before and is explained by difficulties in the mechanic dissociation of complex 3D neuronal tissues into single cells. Particularly, neurons with their long and branched processes have the tendency to be lost in this process. Hence, while scRNA-seq is an excellent tool for the qualitative analysis of cell types, particularly for neural cultures, it might not be the ideal method for cell type quantification. In this context, the identification of mesenchymal cells was surprising. However, a previous study identified similar cells in human dopaminergic neuron grafts in the rodent brain (Tiklova et al. 2019). Therefore, these cells have been described as vascular leptomeningeal cells, a cell type that includes barrier forming fibroblasts. These data are consistent with our findings.

The activity of neurons and their different receptors can be analysed by the specific response to chemical compounds. It has been shown that quinpirole, a specific D2/D3 receptor agonist, suppresses the firing in hMOs (Jo et al. 2016; Monzel et al. 2017). In addition to the previously reported analyses in hMOs, we blocked inhibitory and excitatory communication, to further isolate and attribute the recorded signals to neuronal subtypes. Gabazine induces a disinhibition of GABAergic neurons, whereas NMDA-receptor and AMPA/Kainate-receptor antagonists inhibit glutamatergic excitatory communication (Illes et al. 2014). Together with the characteristic hallmarks of synapse formation (Fig. 6a–d) and the previous findings of dopamine release (Smits et al. 2019), these data confirm the presence of functional dopamine receptors in dopaminergic neurons as well as functional GABAergic and glutamatergic neurons within hMOs. As neurons do not exist in isolation in the CNS but form functional networks with other neurons and non-neuronal cells, it is important to expand our research of neurodegenerative diseases using 3D models that are able to recapitulate cell autonomous as well as non-cell autonomous aspects. Utilizing 3D cell culture models that comprise a variety of neuronal subtypes could lead to new insights into the selective vulnerabilities, which are observed in neurodegeneration. Indeed, evidence

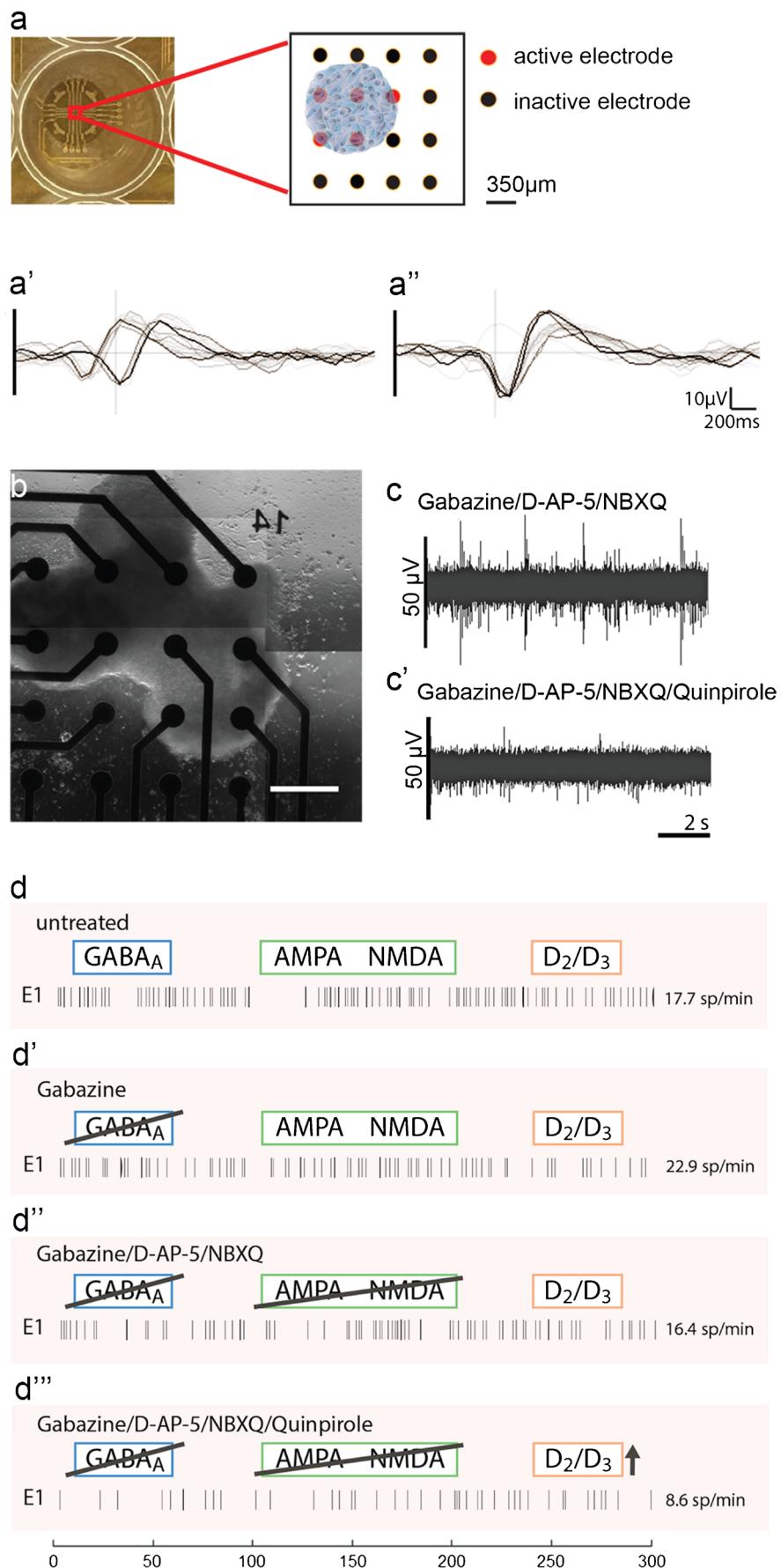


Fig. 7 Electrophysiological activity in midbrain-specific organoids. (a) Representative scheme of positioned midbrain organoid on a 16-electrode array in a 48-well tissue culture plate. Examples of mono- and biphasic spikes detected by individual electrodes of a multielectrode array (MEA) system (a', a''). (b) Representative image of midbrain organoid positioned on a 16-electrode array in a 48-well tissue culture plate. Scale bar is 350 μ m. (c–d) Evaluation of the spontaneous activity by addressing inhibitory (blue) and excitatory (green) neurotransmitter receptors using multielectrode array (MEA) system. (c, c') Representative raw data traces show the effect of Quinpirole in absence of inhibitory and excitatory synaptic communication. (d) Representative spike raster plots demonstrate effects of applied compounds

suggests that specific regulation of the excitability of dopaminergic neurons by other neuronal subtypes in the midbrain might explain their selective vulnerability in PD (Korotkova et al. 2004). This underlines the importance and the enormous potential for future disease modelling of the here described hMO model, as it contains functionally connected heterogeneous neuronal cell populations.

Acknowledgements We would like to thank Dr. Sebastian Illes (University of Gothenburg, Sweden) for his help with the experimental MEA design and Diane Capen (Massachusetts General Hospital, Boston, USA) for her EM work. We thank Dr. Jared Sterneckert (Technical University of Dresden, Germany) and Dr. Bill Skarnes (The Wellcome Trust Sanger Institute, Cambridge, UK) for human iPSC lines. Furthermore, we would like to thank Yohan Jarosz from the LCSB Bioinformatics Core Group for his support in data management. This project was supported by the LCSB pluripotent stem cell core facility.

Authors' contributions LMS designed and performed cell culture and imaging experiments, prepared the figures and wrote the original draft. KG performed the scRNA-seq experiments and related bioinformatics approaches. SSS did the fibroblast analysis. SM, KK, JL and JWS performed the computational analysis of the single-cell RNA-Seq data, edited the manuscript and contributed to the figs. AS, JWS and EM supervised the design and implementation of the single-cell experiments and associated computational data analysis. SB initiated the project, supervised it and edited the manuscript. JCS conceived and supervised the project, designed the experiments and edited the manuscript.

Funding information The JCS lab is supported by the Fonds National de la Recherche (FNR) (CORE, C13/BM/5791363 and Proof-of-Concept program PoC15/11180855 & PoC16/11559169). This is an EU Joint Programme-Neurodegenerative Disease Research (JPND) project (INTER/JPND/14/02; INTER/JPND/15/11092422). Further support comes from the SysMedPD project which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 668738. LMS is supported by fellowships from the FNR (AFR, Aides à la Formation-Recherche). Electron microscopy was performed in the Microscopy Core of the Center for Systems Biology/Program in Membrane Biology, which is partially supported by an Inflammatory Bowel Disease Grant DK043351 and a Boston Area Diabetes and Endocrinology Research Center (BADERC) Award DK057521. SM is supported by the University of Luxembourg and the National Research Fund through the CriTiCS DTU FNR PRIDE/10907093/CRITICS. KG and AS were supported by the Luxembourg National Research Fund (FNR) through the C14/BM/7975668/CaSCAD project and AS additionally by the National Biomedical Computation Resource (NBCR) through the NIH P41 GM103426 grant from the National Institutes of Health.

Compliance with ethical standards

Conflict of interest JCS is co-founder of the biotech company OrganoTherapeutics SARL.

Ethical approval Written informed consent was obtained from all individuals who donated samples to this study and the here conducted work was approved by the responsible ethics commissions. The cell lines used in this study are summarized in Supplementary Table 1.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abe-Fukasawa N, Otsuka K, Aihara A, Itasaki N, Nishino T (2018) Novel 3D liquid cell culture method for anchorage-independent cell growth, cell imaging and automated drug screening. *Sci Rep* 8(1):3627. <https://doi.org/10.1038/s41598-018-21950-5>
- Al-Jaberi N, Lindsay S, Sarma S, Bayatti N, Clowry GJ (2015) The early fetal development of human neocortical GABAergic interneurons. *Cereb Cortex* 25(3):631–645. <https://doi.org/10.1093/cercor/bht254>
- Bardy C, van den Hurk M, Eames T, Marchand C, Hernandez RV, Kellogg M et al (2015) Neuronal medium that supports basic synaptic functions and activity of human neurons in vitro. *Proc Natl Acad Sci*:201504393. <https://doi.org/10.1073/pnas.1504393112>
- Borroto-Escuela DO, Perez De La Mora M, Manger P, Narvaez M, Beggiato S, Crespo-Ramirez M et al (2018) Brain dopamine transmission in health and Parkinson's disease: modulation of synaptic transmission and plasticity through volume transmission and dopamine heteroreceptors. *Front Synaptic Neurosci* 10:20. <https://doi.org/10.3389/fnsyn.2018.00020>
- Cho GS, Lee DI, Tampakakis E, Murphy S, Andersen P, Uosaki H et al (2017) Neonatal transplantation confers maturation of PSC-derived cardiomyocytes conducive to modeling cardiomyopathy. *Cell Rep* 18(2):571–582. <https://doi.org/10.1016/j.celrep.2016.12.040>
- Cooper O, Seo H, Andrabí S, Guardia-Laguarda C, Graziotto J, Sundberg M et al (2012) Pharmacological rescue of mitochondrial deficits in iPSC-derived neural cells from patients with familial Parkinson's disease. *Sci Transl Med* 4(141):141ra190. <https://doi.org/10.1126/scitranslmed.3003985>
- Di Lullo E, Kriegstein AR (2017) The use of brain organoids to investigate neural development and disease. *Nat Rev Neurosci* 18(10):573–584. <https://doi.org/10.1038/nrn.2017.107>
- Fatehullah A, Tan SH, Barker N (2016) Organoids as an *in vitro* model of human development and disease. *Nat Publ Group* 18:246–254. <https://doi.org/10.1038/ncb3312>
- Garcia-Dominguez M, Poquet C, Garel S, Charnay P (2003) Ebf gene function is required for coupling neuronal differentiation and cell cycle exit. *Development* 130(24):6013–6025. <https://doi.org/10.1242/dev.00840>

- Illes S, Jakab M, Beyer F, Gelfert R, Couillard-Despres S, Schnitzler A et al (2014) Intrinsically active and pacemaker neurons in pluripotent stem cell-derived neuronal populations. *Stem Cell Reports* 2(3): 323–336. <https://doi.org/10.1016/j.stemcr.2014.01.006>
- Jo J, Xiao Y, Sun AX, Cukuroglu E, Tran HD, Goke J et al (2016) Midbrain-like organoids from human pluripotent stem cells contain functional dopaminergic and neuromelanin-producing neurons. *Cell Stem Cell* 19(2):248–257. <https://doi.org/10.1016/j.stem.2016.07.005>
- Kelava I, Lancaster MA (2016) Dishing out mini-brains : current progress and future prospects in brain organoid research. *Dev Biol* 420:199–209. <https://doi.org/10.1016/j.ydbio.2016.06.037>
- Kim H, Park HJ, Choi H, Chang Y, Park H, Shin J et al (2019) Modeling G2019S-LRRK2 sporadic Parkinson's disease in 3D midbrain organoids. *Stem Cell Reports* 12(3):518–531. <https://doi.org/10.1016/j.stemcr.2019.01.020>
- Korotkova TM, Ponomarenko AA, Brown RE, Haas HL (2004) Functional diversity of ventral midbrain dopamine and GABAergic neurons. *Mol Neurobiol* 29(3):243–259. <https://doi.org/10.1385/MN:29:3:243>
- La Manno G, Gyllborg D, Codeluppi S, Nishimura K, Salto C, Zeisel A et al (2016) Molecular diversity of midbrain development in mouse, human, and stem cells. *Cell* 167(2):566–580.e519. <https://doi.org/10.1016/j.cell.2016.09.027>
- Lancaster MA, Renner M, Martin CA, Wenzel D, Bicknell LS, Hurles ME et al (2013) Cerebral organoids model human brain development and microcephaly. *Nature* 501(7467):373–379. <https://doi.org/10.1038/nature12517>
- Luhmann HJ, Sinning A, Yang JW, Reyes-Puerta V, Stutzen MC, Kirischuk S et al (2016) Spontaneous neuronal activity in developing neocortical networks: from single cells to large-scale interactions. *Front Neural Circuits* 10:40. <https://doi.org/10.3389/fncir.2016.00040>
- Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M et al (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161(5):1202–1214. <https://doi.org/10.1016/j.cell.2015.05.002>
- Matsui TK, Matsubayashi M, Sakaguchi YM, Hayashi RK, Zheng C, Sugie K et al (2018) Six-month cultured cerebral organoids from human ES cells contain matured neural cells. *Neurosci Lett* 670:75–82. <https://doi.org/10.1016/j.neulet.2018.01.040>
- Ming GL, Tang H, Song H (2016) Advances in Zika virus research: stem cell models, challenges, and opportunities. *Cell Stem Cell* 19(6): 690–702. <https://doi.org/10.1016/j.stem.2016.11.014>
- Monzel AS, Smits LM, Hemmer K, Hachi S, Moreno EL, van Wuellen T et al (2017) Derivation of human midbrain-specific organoids from neuroepithelial stem cells. *Stem Cell Reports* 8(5):1144–1154. <https://doi.org/10.1016/j.stemcr.2017.03.010>
- Muguruma K, Nishiyama A, Kawakami H, Hashimoto K, Sasai Y (2015) Self-organization of polarized cerebellar tissue in 3D culture of human pluripotent stem cells. *Cell Rep* 10:537–550. <https://doi.org/10.1016/j.celrep.2014.12.051>
- Nguyen HN, Byers B, Cord B, Shcheglovitov A, Byrne J, Gujar P et al (2011) LRRK2 mutant iPSC-derived DA neurons demonstrate increased susceptibility to oxidative stress. *Cell Stem Cell* 8:267–280. <https://doi.org/10.1016/j.stem.2011.01.013>
- Odawara A, Katoh H, Matsuda N, Suzuki I (2016) Physiological maturation and drug responses of human induced pluripotent stem cell-derived cortical neuronal networks in long-term culture. *Sci Rep* 6: 26181. <https://doi.org/10.1038/srep26181>
- Patzke C, Acuna C, Giam LR, Wernig M, Sudhof TC (2016) Conditional deletion of L1CAM in human neurons impairs both axonal and dendritic arborization and action potential generation. *J Exp Med* 213(4):499–515. <https://doi.org/10.1084/jem.20150951>
- Qian X, Nguyen HN, Song MM, Hadiono C, Ogden SC, Hammack C et al (2016) Brain-region-specific organoids using mini-bioreactors for modeling ZIKV exposure. *Cell* 165(5):1238–1254. <https://doi.org/10.1016/j.cell.2016.04.032>
- Qian X, Nguyen HN, Jacob F, Song H, Ming GL (2017) Using brain organoids to understand Zika virus-induced microcephaly. *Development* 144(6):952–957. <https://doi.org/10.1242/dev.140707>
- Quadrato G, Nguyen T, Macosko EZ, Sherwood JL, Min Yang S, Berger DR et al (2017) Cell diversity and network dynamics in photosensitive human brain organoids. *Nature* 545(7652):48–53. <https://doi.org/10.1038/nature22047>
- Reinhardt P, Glatza M, Hemmer K, Tsytysura Y, Thiel CS, Hoing S et al (2013a) Derivation and expansion using only small molecules of human neural progenitors for neurodegenerative disease modeling. *PLoS One* 8(3):e59252. <https://doi.org/10.1371/journal.pone.0059252>
- Reinhardt P, Schmid B, Burbulla LF, Schondorf DC, Wagner L, Glatza M et al (2013b) Genetic correction of a LRRK2 mutation in human iPSCs links parkinsonian neurodegeneration to ERK-dependent changes in gene expression. *Cell Stem Cell* 12(3):354–367. <https://doi.org/10.1016/j.stem.2013.01.008>
- Ryan SD, Dolatabadi N, Chan SF, Zhang X, Akhtar MW, Parker J et al (2013) Isogenic human iPSC Parkinson's model shows nitrosative stress-induced dysfunction in MEF2-PGC1 α transcription. *Cell* 155:1351–1364. <https://doi.org/10.1016/j.cell.2013.11.009>
- Sánchez-Danés A, Richaud-Patin Y, Carballo-Carabalí I, Jiménez-Delgado S, Caig C, Mora S et al (2012) Disease-specific phenotypes in dopamine neurons from human iPS-based models of genetic and sporadic Parkinson's disease. *EMBO Mol Med* 4:380–395. <https://doi.org/10.1002/emmm.201200215>
- Satija R, Farrell JA, Gennert D, Schier AF, Regev A (2015) Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* 33(5):495–502. <https://doi.org/10.1038/nbt.3192>
- Smits LM, Reinhardt L, Reinhardt P, Glatza M, Monzel AS, Stanslawsky N et al (2019) Modeling Parkinson's disease in midbrain-like organoids. *npj Parkinson's Dis* in press. <https://doi.org/10.1038/s41531-019-0078-4>
- Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, Hao Y, Stoeckius M, Smibert P, Satija R (2019) Comprehensive integration of single-cell data. *Cell* 177(7):1888–1902. <https://doi.org/10.1016/j.cell.2019.05.031>
- Suzanne M, Steller H (2013) Shaping organisms with apoptosis. *Cell Death Differ* 20(5):669–675. <https://doi.org/10.1038/cdd.2013.11>
- Takahashi K, Yamanaka S (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126(4):663–676. <https://doi.org/10.1016/j.cell.2006.07.024>
- Tepper JM, Lee CR (2007) GABAergic control of substantia nigra dopaminergic neurons. *Prog Brain Res* 160:189–208. [https://doi.org/10.1016/S0079-6123\(06\)60011-3](https://doi.org/10.1016/S0079-6123(06)60011-3)
- Tieng V, Stoppini L, Villy S, Fathi M, Dubois-Dauphin M, Krause KH (2014) Engineering of midbrain organoids containing long-lived dopaminergic neurons. *Stem Cells Dev* 23(13):1535–1547. <https://doi.org/10.1089/scd.2013.0442>
- Tiklová K, Nolbrant S, Fiorenzano A, Björklund ÅK, Sharma V, Heuer A, Gillberg L, Hoban DB, Cardoso T, Adler AF, Birtele M, Lundén-Miguel H, Volakakis N, Kirkeby A, Perlmann T, Parmar M (2019) Single cell gene expression analysis reveals human stem cell-derived graft composition in a cell therapy model of Parkinson's disease. *bioRxiv*:720870. <https://doi.org/10.1101/720870>
- van der Maarten L, Hinton G (2008) Visualizing data using t-SNE. In: *Journal of Machine Learning Research*.
- Walter J (2019) Neural stem cells of Parkinson's disease patients exhibit aberrant mitochondrial morphology and functionality. *Stem Cell Reports* in press



NEUROPSYCHOPHARMACOLOGY REVIEWS

PsychENCODE and beyond: transcriptomics and epigenomics of brain development and organoids

Alexandre Jourdon¹, Soraya Scuderi¹, Davide Capauto¹, Alexej Abyzov² and Flora M. Vaccarino^{1,3,4}

Crucial decisions involving cell fate and connectivity that shape the distinctive development of the human brain occur in the embryonic and fetal stages—stages that are difficult to access and investigate in humans. The last decade has seen an impressive increase in resources—from atlases and databases to biological models—that is progressively lifting the curtain on this critical period. In this review, we describe the current state of genomic, transcriptomic, and epigenomic datasets charting the development of normal human brain with a particular focus on recent single-cell technologies. We discuss the emergence of brain organoids generated from pluripotent stem cells as a model to compensate for the limited availability of fetal tissue. Indeed, comparisons of neural lineages, transcriptional dynamics, and noncoding element activity between fetal brain and organoids have helped identify gene regulatory networks functioning at early stages of brain development. Altogether, we argue that large multi-omics investigations have pushed brain development into the “big data” era, and that current and future transversal approaches needed to leverage both fetal brain and organoid resources promise to answer major questions of brain biology and psychiatry.

Neuropsychopharmacology (2021) 46:70–85; <https://doi.org/10.1038/s41386-020-0763-3>

INTRODUCTION: HUMAN-SPECIFIC FEATURES OF BRAIN DEVELOPMENT

Understanding the development of the human central nervous system is fundamental for unraveling its complex functions, evolutionary innovations, and pathophysiology of neuropsychiatric disorders. The central nervous system originates from the neural tube, which forms within the ectoderm in the dorsal part of the embryo. In early development, succeeding orthogonal waves of diffusible molecules, called morphogens, control the specification and differentiation of its broad domains (i.e., telencephalon, diencephalon, mesencephalon, and spinal cord), their subsequent subdivisions into regions (for instance, the distinction of hippocampus and neocortex in the dorsal telencephalon or the lateral, medial, and caudal ganglionic eminences in the basal telencephalon), and further refined suborganizations (such as the neuronal layers of the neocortex) [1]. This tissue organization is accompanied on the cellular scale by several transitions, starting from a transition of symmetrically dividing neuroepithelial cells into neurogenic radial glia (RG), the stem cells of the developing brain. RG generate over time and space sequential waves of neuronal progenitors, neurons, astrocytes, and oligodendrocytes. In parallel, cells migrate across different regions and form reciprocal interconnections through axonogenesis, finally resulting in neurons and glia developing network electrical activities and completing the process of brain development.

Compared with other mammals, the human brain has more neurons, larger neuronal diversity and pronounced morphological differences—with a complex pattern of gyri and sulci—and a greater brain lateralization compared with nonhuman primates. These features are the result of an increased duration of

neurogenesis, an increased number and diversity of progenitors, and an increased complexity of cell fate programs. This divergence is supported by a longer gestational period and a postnatal maturation finishing only in the third decade of life (reviewed in [2]). The cerebral cortex is the region that has undergone the most remarkable number of changes at the anatomical, cellular, and molecular levels over the course of evolution. The human cerebral cortex has a six-layered laminar architecture, which is relatively conserved in mammals, but with a surface area 1000 times larger than that of a mouse [3]. Cortical areas, with their inherent neuronal architecture and connections, are much more diversified in humans, with many novel areas associated with cognitive functions [4, 5]. The cortex is an anatomical and functional map of our interactions with the external world, and, together with its interconnected regions, forms a biological entity responsible for higher human cognitive abilities and associated psychopathology.

The emerging discipline of functional genomics is crucial to understand the establishment of this cellular and functional architecture (see Box 1 for a glossary of important terms and technologies). Gene regulatory networks, composed by cascades of transcription factors (TFs) and epigenetic modifications, are central to the correct differentiation of every neural cell as the brain develops [6]. TFs bind to genomic noncoding regions called enhancers and the 3D chromatin architecture enables their interactions with gene promoters and other regulatory proteins, even over long distances [7]. These transcriptional machineries control the expression of their target genes in space and time (Fig. 1b). Enhancer activity is regulated in part through epigenetic modification of nearby histones (i.e., acetylation, methylation, and hydroxylation). Altogether, enhancer activities are used as an

¹Child Study Center, Yale University, New Haven, CT 06520, USA; ²Department of Health Sciences Research, Center for Individualized Medicine, Mayo Clinic, Rochester, MN 55905, USA; ³Department of Neuroscience, Yale University, New Haven, CT 06520, USA and ⁴Kavli Institute for Neuroscience, Yale University, New Haven, CT 06520, USA

Correspondence: Flora M. Vaccarino (flora.vaccarino@yale.edu)

These authors contributed equally: Alexandre Jourdon, Soraya Scuderi

Received: 13 May 2020 Revised: 24 June 2020 Accepted: 25 June 2020
Published online: 13 July 2020

Box 1 Glossary

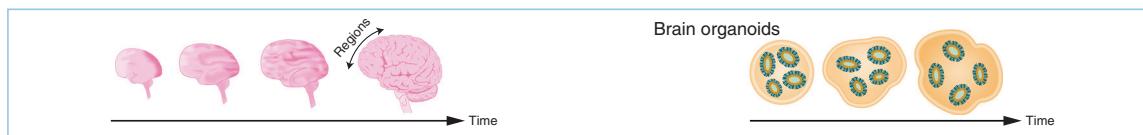
Transcription factors (TFs)	Proteins that bind specific target DNA sequences, located within enhancers or promoters, contributing to the regulation of RNA transcription of specific gene(s).
Promoter	DNA regulatory region typically present upstream of the transcription start site (TSS) of a gene. RNA polymerase recruitment at those loci triggers the transcription of the gene(s).
Enhancer	DNA element that modulates the transcription by binding TFs and bringing them in physical interaction with the cognate promoter(s) of gene(s). In contrast to promoters, enhancers act independently of the distance (up to hundreds of kilobases or even megabases) and orientation to the respective target gene(s). They are characterized by accessible chromatin devoid of nucleosomes and flanked by nucleosomes with specific histone modifications (H3K27ac and/or H3K4me1).
ChIP-seq	Chromatin immunoprecipitation followed by sequencing is a method for finding DNA–protein interactions by combining immunoprecipitation and DNA-sequencing. Can be applied to TF to validate their binding sites or to covalent histone modifications (e.g., H3K27 acetylation, H3K4 methylation) to identify putative regulatory regions—a process referred to as chromatin segmentation—and their differential activity.
RNA-seq	RNA-sequencing is a next-generation sequencing method that provides an overall unbiased quantification of RNA content, which can be refined by ribosomal RNA depletion or mRNA enrichment before cDNA synthesis, library preparation, and sequencing. The relative abundance of transcripts obtained constitute the input to evaluate differential gene expression between samples or conditions.
scRNA-seq	Single-cell RNA-sequencing; a technology that allows the gene expression analysis at single-cell resolution. Isolating individual cells, this technique highlights transcriptional differences between cells of the same biological sample, otherwise obscured using only the bulk RNA-seq analysis.
Hi-C	Hi-C is based on the chromosome conformation capture (3C) technology and it is used to detect, in a genome-wide manner, the chromatin interactions inside the nucleus. Chromatin is crosslinked with formaldehyde; the spatially close genomic fragments are ligated generating chimeric DNA fragments which are captured and identified by deep sequencing.
TAD	Topologically associate domains (TADs) are 3D chromosome structures, whose boundaries are relatively conserved while their internal conformation can change depending upon epigenetic marks and nucleosome structures. Within TADs, chromatin looping brings together regulatory element in close proximity (i.e., enhancers with promoters).
MPRA	Massively parallel reporter assay; method that provides a quantitative measurement of the activity of thousand potential DNA regulatory sequences in parallel. A library of candidate sequences cloned upstream of a promoter/reporter construct are transfected into a cellular model system and identified by deep sequencing.
STARR-seq	Self-transcribing active regulatory region sequencing; like MPRA, this technology assesses the activity of thousand potential regulatory regions, in a cellular model. Candidate sequences are cloned downstream of a promoter/reporter construct. Potential regulatory regions are self-transcribed and identified by RNA-seq to measure candidate regions' activity.
CRISPR/Cas9	Clustered regularly interspaced short palindromic repeats. Cas9 is an endonuclease that can be directed to the target region by a synthetic guide RNA (sgRNA) complementary to the target causing a DNA strand break. A mutated version of Cas9 without endonuclease activity (dead or dCas9) if fused with activator or repressor domains can induce chromatin changes leading to inactivation (CRISPRi) or activation (CRISPRa) of the target regions.
GWAS	Genome-wide association study. It is an observational study of a set of genetic variants (genome-wide) in different individuals to see if any variant is associated with a specific trait. Variants are detected by microarray or whole-genome sequencing technology.
eQTL	Expression quantitative trait loci. Regions of DNA in which genetic variation is associated with variability in the expression of one or more genes.
DNase-seq	A method based on DNasel hypersensitivity for identifying accessible regions of the genome.
ATAC-seq	Assay for transposase accessible chromatin; a method for identifying accessible regions of the genome, based on transposase activity.
WGCNA	Weighted gene co-expression network analysis identifies modules of genes which exhibit correlated patterns of gene expression across samples and often represent similar cell types. Trends in the network can be summarized by eigengenes or hub genes that are central to the network structure.

integrative mechanism for the control of gene expression and are central to the establishment and maintenance of cell identity, affecting the fate of multiple classes of neural cells [8, 9]. As such, they are controlled on a more refined level than the expression of TFs [10, 11]. This fundamental role highlights the importance of identifying TF-enhancer-gene relationships as well as their epigenetic regulation to understand how external information is integrated by the cell. Enhancers, along with other regulatory regions, therefore constitute the main “data” necessary to understand the control of brain development. Finally, mutations within those regions, such as single-nucleotide polymorphisms (SNPs), can affect transcription by, for instance, modifying TF binding affinity or chromatin conformation. Therefore, mutations in noncoding regions can have a more subtle, precise, and context-dependent impact than when affecting coding regions,

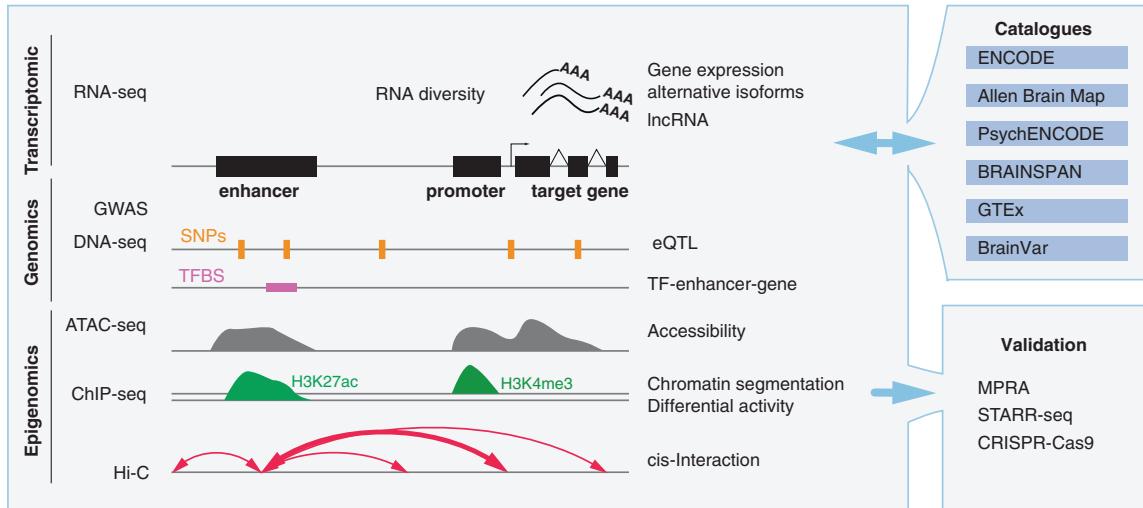
making them an important subject to understand complex genetic diseases, including many neuropsychiatric diseases [12, 13].

Studying functional genomics of the developing human brain is challenging due to the limited availability of fetal tissue for science. The generation of embryonic stem cell (ESC) and human induced pluripotent stem cell (iPSC) lines have recently offered an alternative. Using those cell lines, several labs have developed 3D *in vitro* models of brain development generally referred to as brain organoids [14–16]. By mimicking the morphology of the embryonic nervous system—especially the apico-basal polarity of the RG in the ventricular zone (VZ) and the generation of the outer VZ typical of the human brain — organoids have the inherent capability of generating the multiple cellular lineages of the brain while reproducing *in vivo* cell-to-cell communications and

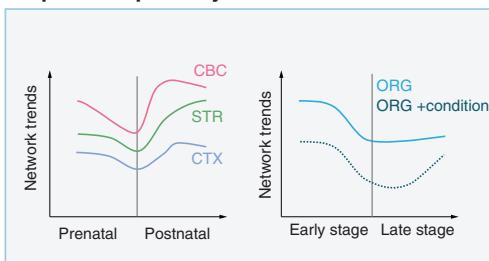
a Early human brain development modeled by Organoids



b Multi Omics integration analysis



c Spatiotemporal dynamics



d Single-cell Omics

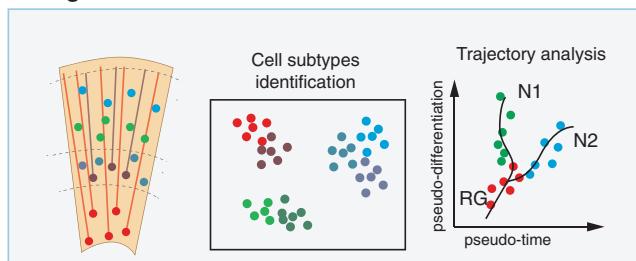


Fig. 1 Integrative approaches to study human brain development. **a** Illustration of two complementary models, postmortem human brain tissue and iPSC-derived brain organoids. **b** Potential of multi-omics approaches to connect/integrate genomic, transcriptomic, and epigenomic information generated or available through datasets of *in vivo/in vitro* studies. Enhancer activity (identified from ATAC-seq, ChIP-seq) influences transcription (RNA-seq) of gene/s through DNA looping (identified through Hi-C). Similarly, eQTL connects risk variants (discovered via DNA-seq, SNPs, GWAS) to gene expression. These variants may disrupt transcription factor binding sites (TFBS) within enhancers. Human-specific gene regulatory mechanisms can be validated in live cell systems, i.e., iPSC-derived organoids, or human progenitor cell lines, by high throughput methods (e.g., MPRA, STARR-seq). **c** Studying gene-enhancer dynamics over time both *in vivo* and in organoids (ORG) can reveal biological insight into the formation and evolution of the human brain (e.g., the hourglass shape of interregional diversity over time between cortex (CTX), striatum (STR), cerebellum (CRB)). **d** Single-cell Omics can define those dynamics at the cellular levels, for instance revealing how gene regulation evolves between radial glia (RG) and neuronal derivatives (N1, N2).

organization [17]. The organoid model presents multiple advantages, from the ability of studying individual variations using iPSCs from different human genetic backgrounds to the ability of conducting longitudinal analyses in the same genomic context. Brain organoids can be easily perturbed to model the effects of environmental and genetic factors on neural development in a controlled setting. While holding invaluable promises, the brain organoid field is still emerging and many developments remain necessary to translate its discoveries into the clinical world (see below, "Exploiting scRNA-seq to study organoids cellular composition and gene networks" section).

In this review, we will describe how the application of functional genomics to the human brain has led to the development of invaluable resources and datasets. We will then focus on recent studies that leveraged those resources to characterize the unique features of the developing human brain, and how the boundaries have been pushed by the emergence of single-cell technologies. The applications of the brain organoid model are presented and

we discuss how they compare and can be integrated with their *in vivo* counterpart, illustrating this by our recent work and the plethora of single-cell studies recently published (Table 1). Overall, we present a blueprint of integrative studies and techniques at the crossroad of developmental biology, stem cell biology, neuroscience, genomics, and system biology to understand brain development and its disorders (Fig. 1).

OMICS AND BIG DATASETS

Genomics and epigenomics

The development of cost-efficient sequencing in the last 10 years has drastically accelerated the reproducible generation of high throughput nucleic acid sequence data (hereafter referred generally as Omics). These include genomic variant discovery, like SNPs, by whole-exome or whole-genome sequencing (WGS). Transcriptome analysis by RNA-sequencing (RNA-seq) profiles global expression patterns of distinct RNA species, including

Table 1. List of scRNA-seq and scATAC-seq studies in human brain organoids.

Authors	Reference	Main focus of study	Type of organoid protocol	Single-cell technology	Number of iPSC line	Stages	Biological samples	Total cells	Main analytical toolkits; visualization, clustering, batch effect, trajectory, gene network	Main cluster annotations	Brain tissue reference (human scRNA-seq/scRNA-seq)	Main findings and observations related to the scRNA-seq of human organoids
									P	D	E	M
Camp et al.	[85]	Organoid model validation	Undirected	Fluidigm C1	1 iPSC, 1 ESC	X		508	Seurat, BackSPIN, t-SNE, Monocle	early/ventral NPC, dorsal NPC, dorsal neur., ventral neur., mesenchmal cells, RSPO + cells	226 cells neocortex (12–13 pcw)	
Quadrato et al.	[145]	Organoid model validation	Undirected	Drop-Seq	1 iPSC, 1 ESC	X X	31	#####	Seurat, t-SNE	NEC, div., Forebrain (RG, IPC, Callosal and Corticofugal neur., IN), Ast., dopa neur., retina, mesodermal cells	Pollen et al. [93].	First study to demonstrate a similar single-cell expression profile in organoid and cortex. Yet, organoids presented less basal progenitor than apical progenitors compared with brain.
Birey et al.	[134]	Regional specification and interaction	Directed: cortical and subpallial +	BD Resolve/ Smart-Seq2	1 iPSC (?) + DLX-lineage	X	4	#####	Seurat, t-SNE	hCS: RG, IPC, Glut., Neur., Choroid, hSS: GABA neur., ventral prog., Ast., OPC		Organoid mimick brain neurogenesis with coherent cell subtypes progression. Presence of mesodermal cells. Biofactors batch influences cell composition.
Sloan et al.	[123]	Cell-subtype specification	Directed: cortical spheroid	Smart-Seq2	2 iPSC	X X	8	710	PAGODA, t-SNE, Monocle	Neur., div. glial, intermediate glial, mature Ast.	Pollen et al. [93].	Specified forebrain subregions. Assembloids reproduce regional interactions, with migration of interneurons expressing subpallial and cortical IN markers (e.g., DLXs).
Xiang et al.	[135]	Regional specification and interaction	Directed: cortical and MGE	10X Chromium	1 ESC, 1 iPSC	X X	8	#####	Seurat, t-SNE, random-forest	NPC, IPC, cortical neur., IN, Non-committed neur., Subplate, Oligo., Ast., Glia, OPC, Ependymocyte	Darmantis et al. [77].	Described the maturation of astrocytes over ~20 months in cortical spheroids, with transcriptional and functional similarities to primary tissue.
Madhavan et al.	[125]	Cell-subtype specification	Directed: oligocortical	10X Chromium	1 ESC	X	1?	#####	CellRanger, t-SNE	Glia and neuronal clusters	Nowakowski et al. [86].	MGE organoids generate Nkx2-1+ IN lineage and oligodendrocytes. Cortical organoids generate interneurons (Nkx2-1 negative). Ability to promote oligodendrocyte lineage while keeping overall cortical fate in spheroids.
Amiri et al.	[70]	Organoid model validation	Directed: dorsal forebrain	10X Chromium	1 iPSC	X X	3	#####	Seurat, t-SNE, Monocle	early RG, ORG, tRG, MGE-RG, divRG, IPC, early EN, late EN, Glycossis, Choroid, U	Nowakowski et al. [86]; Liu et al. [104].	Organoids model cortical development equivalent to 5 and 16 pcw brain tissue, reproducing radial glial and early neuronal diversity. Mapping of enhancer elements and target genes active over organoid development.
Yoon et al.	[144]	Reproducibility	Directed: cortical	BD Rhapsody system	3 iPSC	X	3	#####	BD Rhapsody, t-SNE	RG, IPC, ventral prog., Glut. Neur., GABA choroid, Ast., OPC	Previous data	Reliability of cell composition over multiple iPSC lines, culture conditions, and differentiation.

Table 1. continued

Authors	Reference	Main focus of study	Type of organoid protocol	Single-cell technology	Number of FSC line	Stages	Biological samples	Total cells	Main analytical toolkits: clustering, batch effect, trajectory, gene network	Main cluster annotations	Brain tissue reference (human scRNA-seq/scRNA-seq)	Main findings and observations related to the scRNA-seq of human organoids	P D E M L				
													P	D	E	M	L
Marton et al.	[124]	Cell-subtype specification	Directed: oligodendrocyte	Smart-seq2	2 iPSC	X	2	295	k-mean, t-SNE, Monocle	Neurons, Astrocytes, Microglia, Endothelial, OPC, div, prog, myelinating Oligo.	Darnamis et al. [123]	Induced oligodendrocytes, myelinated neurons with transcriptional similarity with <i>in vivo</i> counterparts.					
Pollen et al.	[151]	Evolution	Directed: telencephalic	Fluidigm C1	10 PSC	X X	32	1223	NN-Louvain, CCA, t-SNE, WGNA	RG, div (G1/S/G2/M), Reelin+, EN, Maturing, EN, IN, Glycans, hindbrain, Choroid, retina, mesenchymal cells, outliers	Darnamis et al. [123]; Sloan et al. [177]	Gene networks are preserved in organoids. Some lines generate off-target cells (mesenchymal, hindbrain, retina).					
Field et al.	[150]	Evolution and lncRNA	Directed: cortical	10X Chromium	1 ESC	X X X	4	#####	k-mean, t-SNE	RG, Neuroepithelium, Cajal-Retzius cells	Darnamis et al. [123]	Noted a variability in composition, presence of glycolysis related gene modules and less interneurons than in primary samples.					
Xiang et al.	[116]	Regional specification and interaction	Directed: Thalamic	10X Chromium	1 ESC	X X	2	#####	Seurat, CCA, Destiny?	NPC, IPC, immature neur., glial prod., Ast., BMP-response, Cilium, proteoglycan, UPR	Darnamis et al. [123]; Thalamus (Allen Brain Atlas)	Described the cell type-specific presence of evolutionary conserved transiently-expressed lncRNA in human organoids.					
Giandomenico et al.	[120]	Regional interaction	Undirected	10X Chromium	2 ESC	X	6	#####	Seurat, t-SNE, Monocle	VZ/SVZ RG, div, prog, immature upper layer neur., mature neur., IN	Camp et al. [85]	Air-liquid interface promotes survival and axon outgrowth with intracortical and subcortical identities.					
Velasco et al.	[143]	Reproducibility	Directed: Dorsal forebrain	10X Chromium	4 iPSC, 1 ESC	X X X	21	#####	Seurat, CCA, t-SNE, Monocle	RG, ORG, div, prog., IPC, Cajal-Retzius cells, immat./mat. projection neur. (callosal, corticofugal), div, IN	Nowakowski et al. [86]; Habib et al. [179], Fan et al. [89]	Dorsally patterned forebrain organoids generate similar cell type composition with low intra- and inter-line variabilities, similarly to brain tissue datasets.					
Trujillo et al.	[119]	Organoid model validation	Directed: cortical spheroid	10X Chromium	Not explicit (?)	X X X	4	#####	Seurat, t-SNE, Monocle	Prog., div, Prog., IPC, Glut. neur., GABA neur., Glia, U	Prog., div, Prog., IPC, Glut. neur., GABA neur., Glia, U	Functional GABAergic interneurons develop after 6 months in cortical organoids. Glial cells become more abundant than progenitors after 3 months.					

Table 1. continued

Authors	Reference	Main focus of study	Type of organoid protocol	Single-cell technology	Number of PSC line	Stages	Biological samples	Total cells	Main analytical toolkits: visualization, clustering, batch effect, trajectory gene network	Main cluster annotations	Brain tissue reference (human scRNA-seq)	Main findings and observations related to the scRNA-seq of human organoids
			P	D	E	M	L					
Cakir et al.	[127]	Cell-subtype specification	Directed; cortical and vascular cells	10X Chromium	1 ESC	X	2	#####	Seurat, CCA, t-SNE, Monocle	RG, IPC, cortical neur., IN, Ast, glial prog., endothelial cells and progenitors, cilium, BMP-response, proteoglycan, EMT, UPR	Zhong et al. [88]	Induction of proto-vasculogenesis in organoids resulted in acceleration of neuronal maturation. Endothelial-like cells expressed vasoconstrictors, peroxisomes, collagen and cell adhesion markers.
Kanton et al.	[152]	Evolution	Undirected	10X Chromium	6 iPSC, 1 ESC	X X X X	36	#####	Seurat, CCA t-SNE, RS, SPRING Diffusion map, RNA velocity	Stem cells, NPC forebrain (dorsal and ventral telencephalon, diencephalon), midbrain, hindbrain and retina, EN, IN, Ast	Nowakowski et al. [86]; Pollen et al. [151]; Miller et al. [6]; BrainSpan; Camp et al. [85]; Mora-Bermudez et al. [149]	Excitatory and inhibitory lineages are distinguished after 2 months, and astrocytes at 4 months. iPSC lines varied in cell composition but lineage-specific signatures remained highly correlated. Human organoids present a delayed maturation compare with chimpanzee and macaque organoids.
Bhaduri et al.	[87]	Organoid model validation	2 directed + 1 undirected	10X Chromium	3 iPSC, 1 ESC	X X X X	36	#####	MACS2, Cicero, t-SNE, Diffusion map	Pluripotency, Neuroectoderm, Neuroepithelium, NPC, Neuron	Human VISTA enhancer	Differentially accessible peaks with chimpanzee with cell specific pattern are enriched for single-nucleotide evolutionary change.
Qian et al.	[122]	Organoid model validation	Directed; cortical (sliced)	SPLiT-seq (nuclei)	1 iPSC	X	1	#####	Seurat, t-SNE	RG, div, IPC, Neurons Layer I, upper layer, deep layer	Zhong et al. [88]; Zhang et al. [180]	Presence of layer specific neocortical neurons and mature signatures.

The information aggregated in this table focuses solely on human pluripotent stem cell (PSC) i.e., ESC- or iPSC-derived brain organoid scRNA-seq data published as of April 2020. The table highlights the historic progression of scRNA-seq usage and reflects the evolution of the technology (from low number of cells at high coverage to high number at lower coverage) and methods. All study metrics (col. F-M) were reported from the published "Method" section, main text, figures, supplementary information, or tables to the best of our comprehension, as no standard for scRNA-seq report exists. Annotations of the main cell types (col. O) have been harmonized when judged equivalent. A difference in type of clusters depends on the protocol used (col D). For instance, undirected protocol generates multiple regional identities (forebrain, midbrain, hindbrain) and mesenchymal fates and regional-directed protocols focus the annotation on relevant cell types. Difficulties and inconsistencies in annotation can partially be explained by the fact that accurate unsupervised clustering (col N) is difficult and its level of resolution is highly dependent on the question central to the study. This highlights the importance of validating annotations though some form of correlation analysis with scRNA-seq or RNA-seq from human fetal brain references (col P).

organoids (>6 months). Abbreviations for methods (col N): *NN-Louvain* Nearest-neighbor graph combined with Louvain clustering method (also used in Seurat), *CCA* canonical correlation analysis for batch effect, *WGCA* weighted correlation network analysis, *RSS* reference similarity spectrum.

Abbreviations for cluster annotations (col O): NPC neural precursor cells, NEC neuroepithelial cell, *neur.* neurons, *div.* dividing/cycling cells, *RG* radial glia, *IPC* intermediate precursor cell, *IN* interneuron, *Ast* astrocyte, *dopa* dopaminergic, *prog*, progenitor, *Glut*, glutamatergic, *OPC* oligodendrocyte precursor cell, *Choroid* Choroid Plexus cells, *Oligo*, oligodendrocyte, *U* unknown/undetermined/unlabeled, *EN* excitatory neurons, *UPR* unfolded protein response related, *VZ* ventricular zone, *SVZ* subventricular zone, *(im)mat.* (immature, *EMT*) epithelial-mesenchymal transition cells, *int.* intermediate.

messenger RNA (mRNA), microRNA or long noncoding RNA (lncRNA) and their eventual posttranscriptional modification, like alternative splicing or RNA editing. Finally, ChIP-seq directed at histone modifications and Hi-C or ATAC-seq that inform about chromatin structure, can be used to evaluate epigenomic changes (Fig. 1b) (Glossary in Box 1).

Large-scale Omics datasets are key to establish exhaustive repertoires of cellular and molecular features of normal and diseased-affected brain development, estimate the validity and fidelity of in vitro models and feed in silico drug discovery. Several public-private research initiatives, national or international (e.g., NIH, EU, Simons Foundation, Allen Institute for Brain Science), have funded collaborative efforts to catalog and analyze genomic, transcriptomic and epigenomic data of cells and tissue in human, nonhuman primates, and model organisms. Data have been uploaded in open repositories such as Gene expression Omnibus or in user-friendly databases to allow further analysis by different groups and enable cross comparisons.

It is now known that many neurodevelopmental and neuropsychiatric disorders, such as autism spectrum disorders (ASD), bipolar disorder, or schizophrenia, are genetically heterogeneous and do not always follow a Mendelian inheritance [13, 18, 19]. Instead, such polygenic disorders arise from multiple causative variants, both common and rare, and complex environmental factors. Genome and exome sequencing studies in patient cohorts and normal individuals have identified potential causative genetic variations of all sizes, from SNPs to large structural variation in the DNA [13, 20, 21], however the major challenge is to interpret their functional impact. The importance of sharing genetic and phenotypic data has fostered global networks since 2015, like Matchmaker Exchange, Decipher and GeneMatcher [22, 23], in which patients with similar genetic variants and/or phenotypes were matched. Other essential resources have been developed and implemented, including databases of diseases-associated genes, like OMIM and ORPHANET [24], or clinical interpretation of variation, like ClinVar or ClinGen [25, 26], initiating a move towards precision medicine and research. Population-wide cataloging of genetic variation, like gnomAD, has allowed the exclusion of variants that were too common at the population level to be plausible causes of diseases [27].

For complex diseases, numerous sequencing-based gene association studies have been done to link phenotype differences with variant allele frequencies. Genome-wide association studies (GWAS)—the first dating back to early 2000s—have been charting common genomic variants across the genome in individuals with or without the disease, initially using genome-wide SNPs arrays and more recently WGS. GWAS studies have demonstrated that most of such variants are found in noncoding regions of the genome, 60% associated to enhancer and super-enhancer, and so are more likely to be involved in gene regulation [28]. This approach has led to the recent identification of common risk variants for schizophrenia, albeit it required more than 36,000 subjects and 110,000 controls [29]. A GWAS effort focused upon neuropsychiatric disorders is the Psychiatric Genomics Consortium that covered 11 psychiatric disorders including attention-deficit/hyperactivity disorder, Alzheimer's disease, ASD, bipolar disorder, eating disorders, major depressive disorder, obsessive-compulsive disorder/Tourette syndrome, post-traumatic stress disorder, schizophrenia, substance use disorders, and all other anxiety disorders [30]. Of note, the NHGRI-EBI published a catalog of GWAS ([ebi.ac.uk/gwas](http://www.ncbi.nlm.nih.gov/gwas)). More focused consortia subsequently developed, for instance the Brain Somatic Mosaicism Network (BSMN), aimed at studying somatic mosaicism both in neurotypical postmortem human brains and in schizophrenia, ASD, bipolar disorder, Tourette syndrome, and epilepsy patient populations. The BSMN aims at cataloging the frequency and pattern of somatic mutations, which are not inherited but occur during the life. To overcome the challenges inherent to discovering somatic

mutations, which are present at low frequency in a subset of an individual's cells, members integrate a variety of complementary approaches which include clonal analyses, deep coverage DNA-sequencing, single-cell genomics, and cutting-edge bioinformatics, while the BSMN enables a cross-platform integrated analysis with other omic-datasets [31–33].

Several brain transcriptomic studies have been performed in both human and nonhuman samples offering the opportunity to probe the molecular basis of neuronal function, understanding its developmental process, and discovering conserved evolutionary mechanisms or diversity between species (reviewed in [34]). The analysis of transcripts has proved challenging, as RNA is more unstable than DNA, especially for postmortem samples. Yet, modification of transcriptional activity remains the core way to link genomic variants to alteration in gene expression through identification of expression quantitative trait loci (eQTL) and other analyses (Fig. 1). Pioneer transcriptome studies were performed in human, macaque, and mouse brain regions across lifespan by microarrays, highlighting time-dependent, layer-, region-, and species-specific features in cortical gene expression profiles [6, 35–41]. Larger transcriptomic datasets were soon generated using RNA-seq by several consortia (e.g., Allen brain Map, BRAINSPAN) (see ALLEN BRAIN ATLAS data portal: <https://portal.brain-map.org/>) from multiple cell lines, human brain regions at mid-fetal and adult stages, and from embryonic and adult mouse cortex. The GTEx project [42, 43] generated transcriptome datasets across tissues in many "normal", non-diseased tissues and each donor was genotyped for common SNPs, creating one of the biggest eQTL studies. Such studies have offered not only the characterization of variation in gene expression levels, but also its link to genetic variants and the basic process of gene regulation. With its extension in 2017, Enhancing GTEx proposes to integrate previous data with telomere length, DNA accessibility, histone modifications, DNA and RNA methylation, somatic mutation, allele-specific expression, and protein quantification across individuals [44].

Another important role of these multicentric studies was to investigate the exquisite gene regulatory mechanisms upstream to the transcriptome, through epigenomic studies. The Encyclopedia of DNA Elements (ENCODE) consortium offered the first functional annotation of regulatory elements in the genome, both coding and noncoding and systematically in human, mouse, fly and worm. To date, the Encode includes 10,868 projects and several bio-sample types. Through many assays (DNA binding, accessibility, methylation, transcription, RNA binding, replication timing, and 3D chromatin structure) the ENCODE performed the first segmentation of the human genome into different categories of functional elements [45]. These include active enhancers, which are typically enriched in H3K27ac-labeled histones; poised enhancers, which exhibit H3K4me1; promoters, which are associated with H3K4me3; and repressed chromatin, associated with H3K27me3.

Finally, our group has taken part in an integrative omics analysis initiative called the Psychiatric Encyclopedia of DNA Elements (PsychENCODE), a consortium focused on understanding gene regulatory mechanisms in the human brain [46]. This is in contrast to the ENCODE, that focused largely on human cell lines. The PsychENCODE consortium has generated a comprehensive online resource (<http://www.psychencode.org/>) of transcriptomic, epigenomic, and genomic data from postmortem developing and adult human brains, both normal and diseased (schizophrenia, ASD, and bipolar disorder), and human cellular model systems. Three main research areas were pursued: dissecting human brain development, studying disease transcriptomes and its regulation, and finally integrating bulk tissue and single-cell data with deep learning approaches to deconvolute the unique features of the human brain. In the following sections, we present some of the main findings of the consortium and related research.

Emergence of single-cell omics

The recent years have seen an explosion of technologies to study genome, transcriptome, and epigenome at the single-cell level [47]. This advancement was enabled by improvements in single-cell isolation and barcoding techniques, coupled to a reduction in sequencing costs. Many single-cell isolation methods now exist, each having different advantages and caveats [48–51]. Alternative methods that rely on combinatorial barcoding to identify single cells without requiring physical isolation have also been applied successfully to neural tissues [52].

Single-cell RNA-sequencing (scRNA-seq) generates transcriptomic signatures of hundreds to millions of single cells, revealing both cellular composition (Fig. 1d) and cell-type-specific gene networks in pluricellular structures such as brains or organoids (Table 1) [47]. The quality of scRNA-seq is highly dependent on correct cell isolation, to avoid doublets (two cells instead of one), and unbiased transcript capture and amplification from each cell, to avoid representation artefacts. Indeed, most techniques capture only a limited fraction of the cell's RNA content, leading to transcripts *dropout*. This unavoidable stochastic loss of transcripts requires the aggregated analysis of multiple cells to recover statistically significant information. Consequently, scRNA-seq output describes the state of *cell subpopulations* and not of single cells per se [47]. Although sequencing coverage is an important parameter, it has been shown that low coverage scRNA-seq (i.e., 50,000 reads per cell) is enough to identify and reconstitute cell diversity in the developing cortex [48]. While most single-cell platforms analyze only the three-prime end of mRNA, some library preparations allow to sequence the full-length mRNA transcript, improving sensitivity in isoforms detection [53].

Bioinformatic analyses of scRNA-seq have become highly complex and are still under active development (reviewed in [47, 54]). For best practice recommendations and a workflow in scRNA-seq data analysis, see [55]. While an extensive overview of current analytical steps is beyond the scope of this review, we wanted to highlight some commonly used tools. The Seurat package (developed in R by the Satija Lab) [56] has become the most commonly used in our field (referenced in 13 out of 19 scRNA organoid studies listed in Table 1) largely due to its centralized handling of the scRNA-seq analytical pipeline, including normalization, batch-effect correction, clustering, visualization, and multifeature integration [57]. Visualization is an important part of scRNA-seq interpretation and mainly relies on nonlinear dimensionality reduction (e.g., t-SNE or UMAP) to reduce the data to a 2D plot of single cells. In developmental datasets, such as fetal brain or organoids, cells are evolving along lineages (e.g., from RG to neurons). This hidden dimension can be revealed along a pseudo-time or pseudo-differentiation axis projection through trajectory analysis [58], appropriate visualization tools [59, 60] (see below, Fig. 1d and Table 1) and further verified by elegant methods such as RNA velocity analysis [61].

Improvements in scRNA-seq analytical tools are still required, especially to merge together the increasing number of datasets generated across multiple technologies and studies, including integration with multiple other cellular features. Indeed, similar to bulk methods presented in the previous section, single-cell studies are progressively becoming multimodal, capturing multiple information from the same sample or even from the same cell, including spatial, epigenomic, morphological, immunophenotype, DNA sequence or mutations, or even electrophysiological [62, 63]. Spatial transcriptomics is a recent development that allows capturing transcriptomic data from a given location in a tissue slice while retaining spatial information close to single-cell resolution [64]. Similarly, profiling single-cell epigenetic information, such as open chromatin state through scATAC-seq or DNA methylation, opens a new feature of classification of cell diversity [65]. Finally, the simultaneous collection of electrophysiological (e.g., patch-clamp or calcium imaging) and transcriptomic data

from the same neural cells constitutes an important innovation for neuroscience [66, 67].

GENOMICS TRAJECTORIES OF THE DEVELOPING BRAIN

Reconstituting human neural development from postmortem human tissues through transcriptomic, epigenomic, and integrative analyses: the PsychENCODE Consortium

The PsychENCODE project is aimed at defining a comprehensive map of functional regulatory genomic elements active in the human brain, differently than the ENCODE project, that mainly focused on peripheral and/or transformed cell lines [45]. The main success of PsychENCODE [46] has been the multi-omic approach that allowed a systematic characterization of noncoding elements, along with the transcriptome, in neurotypical developing and adult brains, in individuals with neuropsychiatric disorders and in human cellular models [12, 68–73]. Coupled with other notable recent studies in developmental genomics [74, 75] and single-cell studies (discussed below) this resource provides new insights into the biology of brain development and its diseases.

Among the main findings, it was observed that the overall transcriptomic signature of all brain regions undergoes a sharp transition phase between mid-fetal and late-fetal stages, suggesting that major changes occur around birth [68]. This temporal trend was accompanied by a transient drop in interregional variability. This suggests that cortical regions become more similar around birth and that adult region-specific signatures arise mainly after the late infancy stages. Part of these dynamic changes could be explained by a regional variability in cell type composition, including differences in progenitor populations during the prenatal stages and differences in mature cell types and functional diversification during later postnatal stages [68, 69]. Different levels of alternative splicing contributed to the overall transcriptional variability over time and space. As splicing dysregulation has been shown to be involved in neurodevelopmental diseases, including ASD, schizophrenia, and bipolar disorder [71], this highlights the importance of studying alternative splicing in early development.

Spatiotemporal variability was also described at the epigenomic level. Major changes in chromatin accessibility as assessed by ATAC-seq between the germinal zone and the cortical plate in fetal cortical samples reflected the transcriptomic changes that happen during neurogenesis [74]. The study also associated putative enhancers to their corresponding TF using binding sites enrichment analysis, confirming that the germinal zone accessible regions are enriched in binding sites for TF implicated in neural progenitor specification (i.e., PAX6, SOX2, ARX, EMX1/2, LHX2, etc.), although the study stopped short of comprehensively defining actual enhancers. This is important because many studies, including large-scale chromatin conformation analysis in mid-gestation brain samples, have revealed that most enhancer-promoter interactions within topologically associating domains (TAD) were long range and not with the adjacent genes [75].

Overall, there is a good concordance between DNA methylation, histone marks, and gene expression over brain development [68]. For instance, enhancers active during the fetal period were associated with genes linked to neural development functions, and became hypermethylated over the postnatal period, heralding the expected decrease in target gene activity. Genome-wide, chromatin accessibility correlated relatively well with gene expression, both at transcription start sites (TSS, $r = 0.417$) and at regulatory regions (putative enhancers), especially when the latter were defined using Hi-C chromatin interactions ($r = 0.456$) [74, 75].

All those studies constituted the bases of an integrative model for the discovery and interpretation of functional genomics of the adult human brain within the PsychEncode consortium [69]. This

included adult brain bulk transcriptome, chromatin, genotype, and Hi-C and single-cell datasets from major human brain regions and merged these datasets with others available through GTEx, ENCODE, and Roadmap Epigenomics (see Fig. 1). All the datasets were uniformly processed to create many fundamental resources, including a list of brain-expressed genes, co-expression modules, 79,000 brain-active enhancers, and their putative targets; more than 2.5 million eQTLs, including relationship with splice isoforms, cell fractions, and chromatin activity. The study generated a brain gene regulatory network where TFs, enhancers, and target genes are linked to each other, based on QTLs, element-activity correlation and Hi-C data. Disease genes were linked to GWAS variants for psychiatric disorders. The regulatory network was used as an input for a machine-learning model to predict psychiatric phenotypes, giving back a threefold increase in prediction compare with other models, highlighting the value of having both epigenome and transcriptome data. This integration remains to be extended to the developmental brain and functionally validated. Interestingly, the integrative model revealed that cell composition is the major contributor to the overall developmental trajectory signature of the human brain, a central result that was only possible to obtain by using single-cell resources.

A parallel longitudinal *in vitro* study of iPSC-derived organoids and fetal brains (described in more detail below) generated a dataset of roughly 96,000 enhancer elements active in early brain development and linked to genes by chromatin conformation analyses, and demonstrated a good correlation between enhancer activity and gene expression along neural differentiation (see below, Fig. 1 and [70]). Altogether, this integration between transcriptome and epigenomic studies demonstrated the validity of multi-omic approaches to deconvolute complex processes like neurodevelopment.

SINGLE-CELL TRANSCRIPTOMIC OF POSTMORTEM DEVELOPING BRAIN SAMPLES

Single-cell nucleus analysis of adult human cortex have been successfully applied to decipher human-specific diversity and organization (e.g., [76, 77]), such as transcriptomic signature of multiple cortical areas [78] and integrated analyses between transcriptomic and epigenomic signatures [79]. In addition, scRNA-seq has revealed cell-type-specific alterations in multiple human brain disorders, like ASD [80], glioblastoma [81], multiple sclerosis [82, 83], or Alzheimer's disease [84].

Single-nucleus analysis of fetal neocortex brain tissue from different post-conceptual weeks (PCW) and gestational weeks (GW) were described in a seminal series of publications by the Kriegstein group and others (GW 16 in Pollen et al. [48]; PCW12-13 in Camp et al. [85]; PCW 6-37, in Nowakowski et al. [86]; GW6-22 in Bhaduri et al. [87]). Other groups have refined and extended the single-cell spatiotemporal dynamics (GW 16-18, Darmanis et al. [77]; GW8-26 in Zhong et al. [88]; PCW 22-23 in Fan et al. [89]; 5-20 PCW in Li et al. [68]; GW17-18 in Polioudakis et al. [90]). Although the neocortex captures most of the attention, similar valuable datasets have been produced for other regions, such as the hippocampus (GW 16-27, [91]) or the embryonic ventral midbrain (PCW 6-11, La Manno et al. [92]). The quality, number of cells, and depth of analysis in those studies have followed the advances in single-cell technology. Many of these studies confirm decades of neurodevelopmental research: from the organization into ventricular, subventricular, subplate, and mantle zones, the order in which neurons of each cortical layer are specified over time, to the switch towards gliogenesis in the later stage. The full diversities of cell type transcriptomes are being characterized, covering not only RG, intermediate progenitor cells, inhibitory, and excitatory neurons, but also astrocytes and oligodendrocyte precursor cells, microglia, choroid plexus cells, mural cells, and endothelial cells [86].

In addition to cataloging the transcriptomic signatures of every single population, scRNA-seq brings the promise of refining or giving a new take on important unresolved questions of forebrain development biology in human brain, and specifically the most evolved regions such as the cerebral cortex.

Identifications of the outer- and truncated-RG subtypes (oRG, tRG) have received a particular attention for their hypothesized impact on human-specific brain size and morphology. Multiple studies have now established their molecular profiles, such as the oRG marker HOPX [93–96]. Interestingly, among the RG subpopulation, oRG presents an enrichment of the mTOR pathway, which has been implicated in pathological conditions related to dysplastic growth and defective cortical migration with co-morbid epilepsy such as focal cortical dysplasia type 2 and hemimegalencephaly [86, 97–99]. Another interesting aspect highlighted by these recent studies is that humans gained enhancers (referred to as HGE) which increase in activity over the course of human evolution, seem to preferentially target genes expressed in the oRG and regulate their growth during neurogenesis [74].

Another important question that has been debated in the mouse literature is the establishment of cortical patterning following either a premitotic model (specialized progenitors) or a postmitotic model (common progenitor but specialized neurons). By comparing visual and prefrontal cortex areas, Nowakowski et al. [86] noted that RG do not present regional transcriptional signatures but show evidence of a progressive divergence in gene networks during neurogenesis. They also noted that those two areas seem to mature at different speeds. This asynchronicity was also observed in another study analyzing 20 different areas, where the authors also observed different proportions of interneurons across the cortical regions [89]. This is in apparent contrast to a more recent study that found cortical areal signatures already present in RG progenitor cells [87].

The development of cortical layers and subpopulations of excitatory neurons seems to go through a phase where immature neurons expressed combination of genes, in particular TFs, that are known to be expressed by distinct cortical layers in adult neurons [68]. An example is the co-expression in embryonic and mid-fetal excitatory cells of BCL11B (CTIP2) and FEZF2 (both known markers of layers V/VI) with CUX2, an upper layer marker (layers II/IV). Another example is the co-expression of RELN and PCP4, specific for layer I and deep layers, respectively. This suggests that human neuronal cell types could be very malleable during early postmitotic differentiation and their molecular identities not completely resolved before the end of mid-fetal development.

The origin and establishment of interneurons diversity in the neocortex have received major attention and there is still a controversy regarding the capacity of dorsal cortical RG to generate interneurons [100–103]. It seems that similarly to excitatory neurons, different proportion of interneurons subtypes populate different cortical areas [89]. There is an early presence of GABAergic progenitors and SST and CALB2+ interneurons in the cortex [88, 89] and one study suggested that some SST-expressing cells could originate from the cortex around GW7 although without entering the cell cycle [88], but there is no evidence of interneurons progenitors later in the neocortex at GW17 [90]. Hence, generation of interneurons from the human neocortical neuroepithelium remains controversial.

Using GW13 to 23 samples, Liu et al. [104] cataloged the expression of lncRNAs in human neocortex development and applied it to detect 1400 lncRNA in single cells from four neocortex samples (19–23 GW) and in previously published samples (Pollen et al. [93]). They resolved cell-type-specific expression (e.g., higher expression of MEG3 or DLX6-AS lncRNA in interneurons compared with excitatory neurons), showed that some lncRNA that are barely detectable in bulk tissue are enriched only in certain cell types, and showed examples of a lncRNA regulating proliferation in RG.

The availability of scATAC-seq allows the mapping of regulatory elements in a cell-type-specific fashion. This is particularly important since regulatory elements exhibit far more cell-type specificity than genes [105, 106]. Open DNA (i.e., DNA accessible to transcriptional regulation) and the transcriptome can then be intersected across single cells [107] and used to infer gene regulatory network at the cell level [90]. This information can be integrated with cell-type enrichment of TFs and co-factors, and intersected with published epigenomic datasets [74]. Using disease variants datasets and scATAC-seq datasets, it is now possible to identify cell type enrichment not only in neurodevelopmental disorder risk genes, but also in SNPs within enhancers active in particular cell types [90, 108].

Finally, it is now possible to link transcriptomic information to functional heterogeneity, as Mayer et al. [67] demonstrated by coupling scRNA-seq with calcium imaging in dissociated cells from mid-gestational (PCW14–22) cortical plate, subplate, and germinal zone.

Altogether, multi-omics and single-cell information of developing brain samples constitute an invaluable resource. In addition to allowing to reconstitute the dynamics of brain development, they represent a necessary reference to validate and improve results obtained using human *in vitro* models, such as brain organoids.

BRAIN ORGANOID, AN IN VITRO MODEL TO VALIDATE FEATURES OF HUMAN BRAIN DEVELOPMENT

Although a diversity of *in vitro* models exists relying on human cells, we will focus on recent developments in the organoid field, presenting the incremental improvements of protocols, the multi-omics characterization and integration with brain data and the main conclusions from 5 years of single-cell study on many aspects of organoids biology (Table 1). Finally, we review innovative approaches for characterization of noncoding elements which, in our opinion, could leverage the power of organoids to answer long-standing questions of human genetics posed in the first part of this review.

Diversity of organoid protocols and applications

Organoid protocols can be separated in three major types—undirected, directed, and patterned—which reflects the extent to which molecular cues are used to guide neural differentiation. Each protocol has benefits and limitations to model different aspects of brain development. Undirected protocols rely on the capacity of neural progenitors to self-organize and yield multiple regional fates [16]. On the contrary, directed organoids take advantage of morphogen agonists or antagonists to mimic developmental cues, guiding cell fates during neural tube patterning *in vivo* [15, 109–112]. This encompasses canonical signaling pathways such as BMPs, TGFβ, Wnts, or SHH. Although at first many directed protocols focused on obtaining telencephalon and neocortex in particular [17, 110, 111], there now exist a full repertoire of protocols to generate regional organoids, including hippocampus [113], cerebellum [114], midbrain [115], thalamus [116], and others [117]. Finally, spatial patterning of organoids is a recent addition to the field where a local molecular signal, for instance SHH, allows long distance spatial organization inside the organoid, a process mimicking the morphogen gradients fundamental to establishing positional identity during development [118].

Maintenance of healthy organoids over a long period of time is crucial for the emergence of spontaneous neuronal activity and network oscillation patterns [119]. This time requirement and the aspiration to mimic other *in vivo* aspects has led to new inventive methodologies: moving from static towards dynamic culture systems (e.g., spinning bioreactors [112] or SpinΩ [117], growing organoids at the air-liquid interface [120] or even incorporating engineered microfilaments [121]). Recently, Qian et al. demonstrated that cultivating thick organoid sections—instead of

culturing whole organoids—improves nutrient access, decreases necrosis in the organoid core and results in an extended formation of most human neocortical layers [122]. In parallel, other groups have focused their efforts on obtaining all neural cell types, including astrocytes [123] or oligodendrocytes [124, 125]. Important for favoring neuronal maturation and energy exchanges, vascularization has been modeled by transplanting organoids in mouse brains [126] or by incorporating external mesodermal sources [127, 128]. Proper angiogenesis and reproducing vascular cues will certainly improve both longevity and fidelity of neural organoids since in mouse they strongly influence neurogenesis dynamics in a region-dependent manner [129]. Microglial cells are another mesodermal cell type that is a key player of neural development and are difficult to obtain in brain organoids. While it has been reported that an undirected protocol can yield some microglia [130], others proposed to rely on an external source [131–133]. Finally, to study interregional migration and connectivity over development, several studies work on a fusion of different regionally directed organoids, baptized *assembloids* [116, 134, 135]. Interestingly, human cells within organoids can form electrophysiologically active connections with mouse cells in xenografts, as nerve tracts have been modeled using co-culture with mouse spinal cord-muscle explants [120, 126]. Altogether, these studies show that the organoid research field is innovating new approaches at a rapid pace to explore what can be done using this model system.

In addition to studying normal development, the organoid model has also been successfully used to characterize the neurological impact of many conditions, including the Zika virus [117], drug exposure [136], syndromic mendelian gene mutations (e.g., in Rett syndrome [137] or Timothy syndrome [138]), and common idiopathic neuropsychiatric disorders such as ASD [110, 139] and schizophrenia [140, 141]. For this, organoids have the unprecedented advantage to be able to model and manipulate pathological mechanisms in a controlled human neural system, deciphering their impact on a large range of features in a dynamic manner (Fig. 1a, c), which is difficult to achieve in other relevant models such as postmortem human brain.

Multi-omics integration of organoids with the developing human brain

The correspondence between organoids and normal brain development has been difficult to investigate, in part due to scarcity of human brain samples, especially at the early stage of development, as well as genetic heterogeneity. The analysis of gene regulation in organoids has been pioneered by Amiri et al., who performed an integrative analysis of enhancer activity and gene expression in organoids derived from fetal skull fibroblasts of three individuals, and compared it to the isogenic fetal human brains [70]. Gene expression dynamics were assessed by RNA-seq during the transition from stem cell to neuronal progenitor cells and from progenitor to neurons, and reflected cell cycle exit, increase in neuronal differentiation, transcriptional regulation in cortical precursor cells and increase in synaptic transmission, cell adhesion, and axon guidance. Transcriptome comparison between organoids, isogenic brains and the PsychEncode developmental dataset showed that organoids mapped more closely to human cortex before 15 PCW.

The study also included cortical brain tissue from the same subjects (15–17 PCW), as internal reference tissue. In comparison with organoids, brain samples were enriched in more mature neuronal transcripts while depleted in transcript related to RG and cell division. Noncoding elements (enhancers, promoters, and repressed chromatin) were mapped by ChIP-seq for several histone posttranslational modifications. Comparison in enhancer number and activity between organoids and brains revealed 1.8 more enhancers in organoids, as 59% of the enhancers still active in organoids were already inactivated in the mid-fetal brain samples.

Overall, this study identified over 300,000 putative enhancers active in organoids and fetal brains during development [70]. Proximity and chromatin conformation analyses were used to link these putative enhancers with their target genes. About 30% (96,375) of the enhancers, among which 10% were novel, could be associated with protein coding genes. Among the ~96,000 gene-linked enhancers, 35% were shared with the isogenic human cortex. Based on correlations between enhancer activity and the expression level of their associated genes, enhancers were cataloged into potentially activating (A-reg) or repressing (R-reg) regulators of gene expression. This classification was reflected in A-reg and R-reg being significantly enriched in genes respectively upregulated or downregulated over time.

Gene expression and enhancer activities were then modeled into a weighted gene correlation network (WGCNA) encompassing 54 co-expressed gene modules, and 29 co-active enhancer modules with specific activity profiles and biological annotations consistent with the organoid's developmental trajectory [70]. Over 24% of the SFARI ASD-associated genes were differentially expressed in the organoids over time, and 80% were associated to enhancers active in both organoids and fetal brain. Genes associated with ASD by postmortem transcriptome analyses were significantly overrepresented in three gene modules related to synapse development and the regulation of cell proliferation. Similar enrichment was observed in the corresponding gene-associated enhancer modules, most of them showing an upregulated activity across development.

Organoids have also been shown to be a promising system to study the genetic mechanisms driving human brain evolution. Over 60% of the human gained enhancers [142], those set of enhancers that possess increased activity in early human brain development compared with rhesus macaque and mouse brains, were active in organoids, particularly at the earliest stages, and showed decreasing activity along differentiation. This evidence suggests that organoids can capture dynamic gene regulatory events, pointing them out as potentially involved in brain neurodevelopmental disorders.

The findings by Amiri et al. provides the most comprehensive integrative analysis of gene-enhancer interactions in human brain organoids so far, where enhancers were identified by a combination of peak calling (ChIP-seq) and chromatin segmentation, followed by identification of interacting gene-enhancer pairs achieved by cross-reference with Hi-C data of human fetal brain. One of the most notable results was the definition of a convergent gene and enhancer network defining global pattern of expression and activity along trajectories of neural cell differentiation and maturation. These enhancers harbored mutations found in ASD probands from external datasets, suggesting that organoids may provide a system to better understand the functional impact of disease-associated risk variants located in noncoding regions of the human genome and their potential to disrupt certain TF binding sites [70].

Exploiting scRNA-seq to study organoids cellular composition and gene networks

Brain organoids are a multi-cellular 3D model by definition. Therefore, the characterization of organoid models depends on identification of cell types and structures obtained, often by immunohistochemistry using antibodies against known markers of neural development and regional identities. The scRNA-seq technology was used early on to unbiasedly identify cellular diversity and lineage in organoid models [85] and has now become a standard for the field. This has led to a recent increase in the organoid single-cell transcriptomic data available. In the 19 studies that we have reviewed and listed in Table 1, the field has accumulated transcriptomic information over close to 800,000 cells, encompassing multiple relevant stages and types of human organoids.

Overall, organoid models seem to reproduce the mechanisms and temporal dynamics of neural system development, which can be observed both in cellular composition over time and gene expression networks (Table 1). Notably, the systematic presence of cell clusters with expression signatures typical of RG (marked by PAX6, NESTIN, SOX2, etc.), intermediate precursor cells (marked by EOMES/TBR2) and different subtypes of excitatory neurons (marked by TBR1, BCL11B, SATB2, CUX1, etc.) confirm the presence of cortical neurogenesis in most models. Interestingly, the ability to clearly distinguish between multiple subtypes of cortical neurons, including Cajal-Retzius cells/layer I neurons and lower/upper layer projection neurons, seems to be dependent on a sufficient period of culture [122]. As *in vivo*, astrocytes and oligodendrocytes require an extensive period of development to be generated (at least 8 months for "mature astrocytes") [123–125, 135]. Despite their subpallial origin, it is also surprising to note that a GABAergic lineage of interneuron is observed in cortically driven organoids [110, 119, 135, 143, 144], although they become more abundant when specifying ganglionic eminence fate. Finally, there seems to be other lineages emerging in many models, such as choroid plexus cells and ependymocytes of the VZ, although there is less clear consensus on their annotations (Table 1). Since many studies specifically use cortical organoids, a consensus on annotation strategies and markers of reference could be established in future work.

Despite this overall fidelity in reproducing brain cellular lineages and gene networks, Bhaduri et al. [87] recently exposed a divergence in the specification and maturation of different cell subtypes between organoids and brain samples (GW6-22). They link decreased neuronal maturation with an aberrant activation of oxidative and stress pathways *in vitro*, including glycolysis and endoplasmic reticulum stress, which could suggest improvements of organoid culture conditions (e.g., modifying glucose concentration or oxygen level).

Variability remains a major limitation of organoid models. Undirected brain organoids seem to present a higher intra- and inter-organoid regional variability, perhaps due to stochastic organization, with not only forebrain identity, but also regions such as retina, spinal cord, and others [16, 145]. Study of cortical development in those models often requires micro-dissection of cortex-like areas that can be visually identified from the rest of the organoid, although extracted regions can end up having non-cortical identity [85]. Dorsal forebrain patterned organoids seem to present less organoid to organoid and line to line variability [143, 144], although some variation, for instance in cortical areal identity pattern, have been reported [87]. The assessment of organoid cell composition variability, both between lines and between batches, is an important step relevant to statistical modeling and differential gene expression analysis. Notably, many published scRNA-seq studies are generated from 1 or 2 biologically distinct ESC or iPSC lines (14 out of 19 datasets in Table 1), and with rare intra-line variability estimation.

Integration efforts, especially across the diversity of existing protocols, should generate meaningful transversal conclusions. One limitation of the integration is the batch correction step when dealing with diverse datasets in terms of isolation technology, library preparation, and read depth level per cell, with data origin often driving the clustering [87]. Using canonical correlation analysis, Tanaka et al. recently integrated eight datasets from different studies containing both directed and undirected protocols and found similar cellular composition and gene expression per lineage [146]. With recent improvements in batch correction and integration methods [65, 147, 148], such transversal analysis should become more accessible, and lead to establishing a common single-cell transcriptomics organoid atlas defined through common marker genes and reference datasets. Such integration will lead to a clear transcriptomic definition of the cellular space of *in vitro* brain organoids, leading to a

protocol-independent definition of artefactual cell types and structuring the diverse lineage trajectories. Such integration is vitally important to determine inter- and intra-protocol variability in cell fate.

Finally, evolution has received a particular interest in single-cell organoid studies, owing to the capacity of generating iPSCs and brain organoids from multiple primate species, including chimpanzee, orangutan, and rhesus macaque [149–152]. It was demonstrated that compared with other apes, human organoids present a delayed maturation, with less mature neuronal signatures and astrocytes presence at equivalent stages, which agrees with the longer gestational period in humans [152]. Interestingly, transcriptomic divergence in the telencephalic lineage seems to consist mainly of gains in new gene expression in humans, with related functions spanning proliferation of RG, neuron migration, and neurite formation [152]. Using scATAC-seq in complement to scRNA-seq, Kanton et al. also showed that differentially accessible peaks between human and chimpanzee have a cell-type-specific pattern and are enriched in single-nucleotide evolutionary changes. Coherently, evolution seems to affect gene networks differently in different neural lineages, with major changes occurring in astrocytes [152].

Functional validation of noncoding elements in vitro

Identifying the physical location of putative gene regulatory elements does not represent definitive proof of their functional activity in regulating gene expression. This is true even if the degree of open chromatin, quantified by ATAC-seq or ChIP-seq signals, correlates well with gene expression. There is, therefore, a need to combine the biochemical annotation-based techniques, aimed at assessing both the accessibility of chromatin (DNaseI-seq and ATAC-seq) and its interactions (Hi-C and variant technologies such as capture Hi-C [153], PLAC-seq [154], and HiChIP [155]) with orthogonal assays more directed on demonstrating regulatory element activity and their effect on target genes.

In this regard, massively parallel reporter assay (MPRA) [156, 157] (Box 1) allows testing thousands of regulatory elements in a single experiment. A synthesized library of candidate sequences is cloned into a vector, wherein each candidate element is upstream of a basal promoter and a reporter gene. Each reporter gene is associated with a unique barcode providing a quantitative readout of the cognate candidate enhancer activity. Permitting the artificial introduction of SNPs [158], eQTL [159] or potential TF binding sites disrupting variants [160], MPRA is a powerful system for studying the effect of traits or disease-associated genomic variations on the functionality of regulatory regions. The shortcoming of this technique is the limited size of the tested fragments, that might not fully represent the entire enhancer's region. Furthermore, both the use of episomal and lentiviral (lentiMPRA [161, 162]) reporter vectors do not reflect the enhancers functionality in their endogenous genomic context. Finally, the MPRA is a good tool to measure enhancers' activity but fails to identify their endogenous target gene/s.

An alternative method for the enhancers' validation is STARR-seq (self-transcribing active regulatory region sequencing [163]) (Box 1). Differently than MPRA, the candidate regulatory sequences are placed downstream of the minimal promoter and reporter gene, and therefore they will be transcribed as "enhancer RNA" such that each enhancer sequence works as its own barcode. Like MPRA, STARR-seq allows to investigate the functional activity of regions of interest selected by other predictive analyses (CapSTARR-seq [164], CHIP STARR-seq [165]) or to test the functional effect of disease-associated SNPs but removes any epigenetic contextualization.

Using the CRISPR/Cas9 technology [166] (Box 1) is complementary to MPRA and STARR-seq as it allows to perturb the sequences of interest in their native context and can also reveal the cognate

gene(s) of enhancer regulation. Many studies use the active Cas9 and sgRNA libraries to "destroy" specific noncoding regions (up to hundreds of kilobases) in order to affect the expression of target genes [167–171]. An alternative approach alters the epigenomic landscape rather than the genomic sequence of the target region (CRISPR-epigenome editing). A catalytically inactive Cas9 (dCas9), fused with functional repressor or activator domains, triggers, respectively, repressive (CRISPR interference, or CRISPRi) or activating (CRISPR activation, or CRISPRa) chromatin modifications [172, 173]. Alongside CRISPR-mediated strategies based on detectable features (expression of the genetic reporter [171], drug resistance [169], or growth assays [174, 175]) recent works have combined CRISPRi-based enhancers perturbation with scRNA-seq in order to evaluate the variation of the transcriptome across the genome at single-cell resolution [176, 177]. However, all the CRISPR-based screens are not exempt from technical issues such as the potential presence of false positives and false negatives, or inefficiency of the Cas9-fused repressors/activators on certain enhancers [178].

The still unexplored combination of the above-mentioned methods (MPRA, STARR-seq, and CRISPR) with the powerful system of the cerebral organoids can open new avenues for a deeper understanding of the regulatory network involved in brain development and disorders.

FUTURE RESEARCH DIRECTIONS

There is an impressive amount of data available for the developmental neuroscience field, both from postmortem brain samples and in vitro models, which are largely derived from genome-scale sequencing efforts at both DNA and RNA levels. As demonstrated by some of the most recent studies presented here, we argue that the future lies in integration of these different levels of analyses. Impactful results can come from the intersections between transcriptomic information, epigenomic context, and genomic variation, which could be made even more compelling by integration of imaging and electrophysiological results. Targets relevant to the clinic can be obtained by intersecting those results with GWAS, WGS, exome and other databases to begin to understand disease pathophysiology during development of the brain. We tentatively summarize this multimodal approach in Fig. 1.

As all new technology, brain organoids come with obstacles and challenges. Although organoids resemble the embryonic to early fetal human brain at the molecular level, along with the capacity to generate most of the neural lineages found in humans, they are still less mature compared with adult neurons. Moreover, this system lacks the capacity to fully recapitulate features of human brain development like gyration, full distinct cortical neuronal layers, gliogenesis, and complex neuronal circuitry formation. Features that undoubtedly will be explored in future years are how to promote vascularization and proper morphogenetic patterning to acquire better fidelity to *in vivo* development and try to recapitulate later stages of fetal development. Meanwhile the human iPSC-derived brain organoids have opened new ways to analyze brain development for a specific individual and in a longitudinal fashion. They promise to help us gain better understanding of the field of functional genomics, including defining enhancer–gene relationships and other regulatory mechanisms that govern brain development (Fig. 1b).

Multimodal studies, incorporating genome-scale analyses with other biological features of both brain tissue and organoids, have become more complex to understand. Developmental neurobiology is undergoing a progressive transformation from traditional one gene-one function studies to integrative "big data" studies, and the shift towards single-cell analysis has exponentially increased the amount of information available. We predict that transversal and meta-analysis will reveal more than originally meet the eyes, and that computational biology and machine-learning

techniques will allow attaining a deeper understanding of brain development.

FUNDING AND DISCLOSURE

The authors declare no competing financial interests related to this article. FMV, DC, and SS are members of the PsychEncode consortium. This work was funded by NIH grants #R01 MH109648, #R56 MH114911, #U01 MH103365, and Simons Foundation grant# 632742.

ACKNOWLEDGEMENTS

We thank Jeremy Schreiner for proof editing this paper.

AUTHOR CONTRIBUTIONS

Paper conception/outline: FMV and AJ. Paper writing: FMV, AJ, SS, DC, and AA. Display item preparation (Fig. 1): AJ, SS, and FMV. (Table 1) curated by AJ. (Box 1) curated by AJ, SS, DC, FMV. All authors participated in paper editing and proofreading.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

1. Sansom SN, Livesey FJ. Gradients in the brain: the control of the development of form and function in the cerebral cortex. *Cold Spring Harb Perspect Biol*. 2009;1:a002519. <https://doi.org/10.1101/cshperspect.a002519>.
2. Molnar Z, Clowry GJ, Sestan N, Alzu'bi A, Bakken T, Hevner RF, et al. New insights into the development of the human cerebral cortex. *J Anat*. 2019;235:432–51. <https://doi.org/10.1111/joa.13055>.
3. Rakic P. A small step for the cell, a giant leap for mankind: a hypothesis of neocortical expansion during evolution. *Trends Neurosci*. 1995;18:383–8.
4. Shine JM, Breakspear M, Bell PT, Eghoetz Martens KA, Shine R, Koyejo O, et al. Human cognition involves the dynamic integration of neural activity and neuromodulatory systems. *Nat Neurosci*. 2019;22:289–96. <https://doi.org/10.1038/s41593-018-0312-0>.
5. Helfrich RF, Knight RT. Cognitive neurophysiology of the prefrontal cortex. *Handb Clin Neurol*. 2019;163:35–59. <https://doi.org/10.1016/B978-0-12-804281-6.00003-3>.
6. Miller JA, Ding SL, Sunkin SM, Smith KA, Ng L, Szafer A, et al. Transcriptional landscape of the prenatal human brain. *Nature*. 2014;508:199–206. <https://doi.org/10.1038/nature13185>.
7. Kim S, Shendure J. Mechanisms of interplay between transcription factors and the 3D genome. *Mol Cell*. 2019;76:306–19. <https://doi.org/10.1016/j.molcel.2019.08.010>.
8. Kim TK, Shiekhattar R. Architectural and functional commonalities between enhancers and promoters. *Cell*. 2015;162:948–59. <https://doi.org/10.1016/j.cell.2015.08.008>.
9. Long HK, Prescott SL, Wysocka J. Ever-changing landscapes: transcriptional enhancers in development and evolution. *Cell*. 2016;167:1170–87. <https://doi.org/10.1016/j.cell.2016.09.018>.
10. Visel A, Taher L, Gigris H, May D, Golonzka O, Hoch RV, et al. A high-resolution enhancer atlas of the developing telencephalon. *Cell*. 2013;152:895–908. <https://doi.org/10.1016/j.cell.2012.12.041>.
11. Dickel DE, Ypsilanti AR, Pla R, Zhu Y, Barozzi I, Mannion BJ, et al. Ultraconserved enhancers are required for normal development. *Cell*. 2018;172:491–9. <https://doi.org/10.1016/j.cell.2017.12.017>.
12. Rajarajan P, Borrman T, Liao W, Schrodé N, Flaherty E, Casino C, et al. Neuron-specific signatures in the chromosomal connectome associated with schizophrenia risk. *Science*. 2018;362. <https://doi.org/10.1126/science.aat4311>.
13. Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet*. 2019;51:431–44. <https://doi.org/10.1038/s41588-019-0344-8>.
14. Eiraku M, Watanabe K, Matsuo-Tasaki M, Kawada M, Yonemura S, Matsumura M, et al. Self-organized formation of polarized cortical tissues from ESCs and its active manipulation by extrinsic signals. *Cell Stem Cell*. 2008;3:519–32. <https://doi.org/10.1016/j.stem.2008.09.002>.
15. Mariani J, Simonini MV, Palejev D, Tomasini L, Coppola G, Szekely AM, et al. Modeling human cortical development in vitro using induced pluripotent stem cells. *Proc Natl Acad Sci USA*. 2012;109:12770–5. <https://doi.org/10.1073/pnas.1202944109>.
16. Lancaster MA, Renner M, Martin CA, Wenzel D, Bicknell LS, Hurles ME, et al. Cerebral organoids model human brain development and microcephaly. *Nature*. 2013;501:373–9. <https://doi.org/10.1038/nature12517>.
17. Bershteyn M, Nowakowski TJ, Pollen AA, Di Lullo E, Nene A, Wynshaw-Boris A, et al. Human iPSC-derived cerebral organoids model cellular features of lissencephaly and reveal prolonged mitosis of outer radial glia. *Cell Stem Cell*. 2017;20:435–49. <https://doi.org/10.1016/j.stem.2016.12.007>.
18. Gandal MJ, Leppa V, Won H, Parkash NN, Geschwind DH. The road to precision psychiatry: translating genetics into disease mechanisms. *Nat Neurosci*. 2016;19:1397–407. <https://doi.org/10.1038/nn.4409>.
19. Talkowski ME, Rosenfeld JA, Blumenthal I, Pillalamarri V, Chiang C, Heilbut A, et al. Sequencing chromosomal abnormalities reveals neurodevelopmental loci that confer risk across diagnostic boundaries. *Cell*. 2012;149:525–37. <https://doi.org/10.1016/j.cell.2012.03.028>.
20. Coe BP, Stessman HAF, Sulovari A, Geiseker MR, Bakken TE, Lake AM, et al. Neurodevelopmental disease genes implicated by de novo mutation and copy number variation morbidity. *Nat Genet*. 2019;51:106–16. <https://doi.org/10.1038/s41588-018-0288-4>.
21. Satterstrom FK, Kosmicki JA, Wang J, Breen MS, De Rubeis S, An JY, et al. Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. *Cell*. 2020;180:568–84. <https://doi.org/10.1016/j.cell.2019.12.036>. e23
22. Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D, et al. DECIPHER: database of chromosomal imbalance and phenotype in humans using Ensembl resources. *Am J Hum Genet*. 2009;84:524–33. <https://doi.org/10.1016/j.ajhg.2009.03.010>.
23. Philippakis AA, Azzariti DR, Beltran S, Brookes AJ, Brownstein CA, Brudno M, et al. The Matchmaker Exchange: a platform for rare disease gene discovery. *Hum Mutat*. 2015;36:915–21. <https://doi.org/10.1002/humu.22858>.
24. Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. OMIM.org: Online Mendelian Inheritance in Man (OMIM(R)), an online catalog of human genes and genetic disorders. *Nucleic Acids Res*. 2015;43:D789–98. <https://doi.org/10.1093/nar/gku1205>.
25. Rehm HL, Berg JS, Brooks LD, Bustamante CD, Evans JP, Landrum MJ, et al. ClinGen—the clinical genome resource. *N Engl J Med*. 2015;372:2235–42. <https://doi.org/10.1056/NEJMsr1406261>.
26. Landrum MJ, Chitipiralla S, Brown GR, Chen C, Gu B, Hart J, et al. ClinVar: improvements to accessing data. *Nucleic Acids Res*. 2020;48:D835–D44. <https://doi.org/10.1093/nar/gkz972>.
27. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536:285–91.
28. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA, et al. Super-enhancers in the control of cell identity and disease. *Cell*. 2013;155:934–47. <https://doi.org/10.1016/j.cell.2013.09.053>.
29. Schizophrenia Working Group of the Psychiatric Genomics C. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 2014;511:421–7. <https://doi.org/10.1038/nature13595>.
30. Psychiatric GCSC. A framework for interpreting genome-wide association studies of psychiatric disorders. *Mol Psychiatry*. 2009;14:10–7. <https://doi.org/10.1038/mp.2008.126>.
31. McConnell MJ, Moran JV, Abyzov A, Akbarian S, Bae T, Cortes-Ciriano I, et al. Intersection of diverse neuronal genomes and neuropsychiatric disease: The Brain Somatic Mosaic Network. *Science*. 2017;356. <https://doi.org/10.1126/science.aal1641>.
32. Abyzov A, Vaccarino FM, Urban AE, Sarangi V. Approaches and methods for variant analysis in the genome of a single cell. In: Moskalev A, editor. *Biomarkers of Human Aging and Longevity*: Springer, Cham; 2019. p. 203–28.
33. D'Gama AM, Walsh CA. Somatic mosaicism and neurodevelopmental disease. *Nat Neurosci*. 2018;21:1504–14. <https://doi.org/10.1038/s41593-018-0257-3>.
34. Keil JM, Qalieh A, Kwan KY. Brain transcriptome databases: a user's guide. *J Neurosci*. 2018;38:2399–412. <https://doi.org/10.1523/JNEUROSCI.1930-17.2018>.
35. Bakken TE, Miller JA, Ding SL, Sunkin SM, Smith KA, Ng L, et al. A comprehensive transcriptional map of primate brain development. *Nature*. 2016;535:367–75. <https://doi.org/10.1038/nature18637>.
36. Johnson MB, Kawasawa YI, Mason CE, Krsnik Z, Coppola G, Bogdanovic D, et al. Functional and evolutionary insights into human brain development through global transcriptome analysis. *Neuron*. 2009;62:494–509. <https://doi.org/10.1016/j.neuron.2009.03.027>.
37. Kang HJ, Kawasawa YI, Cheng F, Zhu Y, Xu X, Li M, et al. Spatio-temporal transcriptome of the human brain. *Nature*. 2011;478:483–9. <https://doi.org/10.1038/nature10523>.
38. Colantuoni C, Lipska BK, Ye T, Hyde TM, Tao R, Leek JT, et al. Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature*. 2011;478:519–23. <https://doi.org/10.1038/nature10524>.

39. Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, Shen EH, Ng L, Miller JA, et al. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature*. 2012;489:391–9. <https://doi.org/10.1038/nature11405>.
40. Ayoub AE, Oh S, Xie Y, Leng J, Cotney J, Dominguez MH, et al. Transcriptional programs in transient embryonic zones of the cerebral cortex defined by high-resolution mRNA sequencing. *Proc Natl Acad Sci USA*. 2011;108:14950–5. <https://doi.org/10.1073/pnas.1112213108>.
41. Belgard TG, Marques AC, Oliver PL, Aiba HO, Sirey TM, Hoerder-Suhedissen A, et al. A transcriptomic atlas of mouse neocortical layers. *Neuron*. 2011;71:605–16. <https://doi.org/10.1016/j.neuron.2011.06.039>.
42. Consortium GT. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*. 2015;348:648–60. <https://doi.org/10.1126/science.1262110>.
43. Consortium GT. Erratum: Genetic effects on gene expression across human tissues. *Nature*. 2018;553:530. <https://doi.org/10.1038/nature25160>.
44. e GP. Enhancing GTEx by bridging the gaps between genotype, gene expression, and disease. *Nat Genet*. 2017;49:1664–70. <https://doi.org/10.1038/ng.3969>.
45. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489:57–74. <https://doi.org/10.1038/nature11247>.
46. Psych EC, Akbarian S, Liu C, Knowles JA, Vaccarino FM, Farnham PJ, et al. The PsychENCODE project. *Nat Neurosci*. 2015;18:1707–12. <https://doi.org/10.1038/nn.4156>.
47. Lahmann D, Koster J, Szczurek E, McCarthy DJ, Hicks SC, Robinson MD, et al. Eleven grand challenges in single-cell data science. *Genome Biol*. 2020;21:31. <https://doi.org/10.1186/s13059-020-1926-6>.
48. Pollen AA, Nowakowski TJ, Shuga J, Wang X, Leyrat AA, Lui JH, et al. Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. *Nat Biotechnol*. 2014. <https://doi.org/10.1038/nbt.2967>
49. Fan HC, Fu GK, Fodor SP. Expression profiling. Combinatorial labeling of single cells for gene expression cytometry. *Science*. 2015;347:1258367. <https://doi.org/10.1126/science.1258367>.
50. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*. 2015;161:1202–14. <https://doi.org/10.1016/j.cell.2015.05.002>.
51. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell*. 2015;161:1187–201. <https://doi.org/10.1016/j.cell.2015.04.044>.
52. Rosenberg AB, Roco CM, Muscat RA, Kuchina A, Sample P, Yao Z, et al. Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science*. 2018;360:176–82. <https://doi.org/10.1126/science.aam8999>.
53. Picelli S, Bjorklund AK, Faridani OR, Sagasser S, Winberg G, Sandberg R. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat Methods*. 2013;10:1096–8. <https://doi.org/10.1038/nmeth.2639>.
54. Amezquita RA, Lun ATL, Becht E, Carey VJ, Carpp LN, Geistlinger L, et al. Orchestrating single-cell analysis with bioconductor. *Nat Methods*. 2020;17:137–45. <https://doi.org/10.1038/s41592-019-0654-x>.
55. Luecken MD, Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol Syst Biol*. 2019;15:e8746. <https://doi.org/10.1525/msb.20188746>.
56. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol*. 2015;33:495–502. <https://doi.org/10.1038/nbt.3192>.
57. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol*. 2019;20:296. <https://doi.org/10.1186/s13059-019-1874-1>.
58. Saelens W, Cannoodt R, Todorov H, Saeyns Y. A comparison of single-cell trajectory inference methods. *Nat Biotechnol*. 2019;37:547–54. <https://doi.org/10.1038/s41587-019-0071-9>.
59. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol*. 2018. <https://doi.org/10.1038/nbt.4314>.
60. Moon KR, van Dijk D, Wang Z, Gigante S, Burkhardt DB, Chen WS, et al. Visualizing structure and transitions in high-dimensional biological data. *Nat Biotechnol*. 2019;37:1482–92. <https://doi.org/10.1038/s41587-019-0336-3>.
61. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, et al. RNA velocity of single cells. *Nature*. 2018;560:494–8. <https://doi.org/10.1038/s41586-018-0414-6>.
62. van den Hurk M, Bardy C. Single-cell multimodal transcriptomics to study neuronal diversity in human stem cell-derived brain tissue and organoid models. *J Neurosci Methods*. 2019;325:108350. <https://doi.org/10.1016/j.jneumeth.2019.108350>.
63. Macaulay IC, Ponting CP, Voet T. Single-cell multiomics: multiple measurements from single cells. *Trends Genet*. 2017;33:155–68. <https://doi.org/10.1016/j.tig.2016.12.003>.
64. Rodrigues SG, Stickels RR, Goeva A, Martin CA, Murray E, Vanderburg CR, et al. Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science*. 2019;363:1463–7. <https://doi.org/10.1126/science.aaw1219>.
65. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3rd, et al. Comprehensive integration of single-cell data. *Cell*. 2019;177:1888–902.e1821. <https://doi.org/10.1016/j.cell.2019.05.031>.
66. Cadwell CR, Palasantza A, Jiang X, Berens P, Deng Q, Yilmaz M, et al. Electrophysiological, transcriptomic and morphologic profiling of single neurons using Patch-seq. *Nat Biotechnol*. 2016;34:199–203. <https://doi.org/10.1038/nbt.3445>.
67. Mayer S, Chen J, Velmeshov D, Mayer A, Eze UC, Bhaduri A, et al. Multimodal single-cell analysis reveals physiological maturation in the developing human neocortex. *Neuron*. 2019;102:143–58.e7. <https://doi.org/10.1016/j.neuron.2019.01.027>.
68. Li M, Santpere G, Imamura Kawasawa Y, Evgrafov OV, Gulden FO, Pochareddy S, et al. Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science*. 2018;362. <https://doi.org/10.1126/science.aat7615>.
69. Wang D, Liu S, Warrell J, Won H, Shi X, Navarro FCP, et al. Comprehensive functional genomic resource and integrative model for the human brain. *Science*. 2018;362. <https://doi.org/10.1126/science.aat8464>.
70. Amiri A, Coppola G, Scuderi S, Wu F, Roychowdhury T, Liu F, et al. Transcriptome and epigenome landscape of human cortical development modeled in organoids. *Science*. 2018;362. <https://doi.org/10.1126/science.aat6720>.
71. Gandal MJ, Zhang P, Hadjimichael E, Walker RL, Chen C, Liu S, et al. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science*. 2018;362. <https://doi.org/10.1126/science.aat8127>.
72. Rhie SK, Schreiner S, Witt H, Armoskus C, Lay FD, Camarena A, et al. Using 3D epigenomic maps of primary olfactory neuronal cells from living individuals to understand gene regulation. *Sci Adv* 2018;4:eaav8550. <https://doi.org/10.1126/sciadv.aav8550>.
73. An JY, Lin K, Zhu L, Werling DM, Dong S, Brand H, et al. Genome-wide de novo risk score implicates promoter variation in autism spectrum disorder. *Science*. 2018;362. <https://doi.org/10.1126/science.aat6576>.
74. de la Torre-Ubieta L, Stein JL, Won H, Opland CK, Liang D, Lu D, et al. The dynamic landscape of open chromatin during human cortical neurogenesis. *Cell*. 2018;172:289–304. <https://doi.org/10.1016/j.cell.2017.12.014>.
75. Won H, de la Torre-Ubieta L, Stein JL, Parikhshak NN, Huang J, Opland CK, et al. Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature*. 2016;538:523–7. <https://doi.org/10.1038/nature19847>.
76. Hodge RD, Bakken TE, Miller JA, Smith KA, Barkan ER, Graybuck LT, et al. Conserved cell types with divergent features in human versus mouse cortex. *Nature*. 2019. <https://doi.org/10.1038/s41586-019-1506-7>.
77. Darmanis S, Sloan SA, Zhang Y, Enge M, Caneda C, Shuer LM, et al. A survey of human brain transcriptome diversity at the single cell level. *Proc Natl Acad Sci USA*. 2015;112:7285–90. <https://doi.org/10.1073/pnas.1507125112>.
78. Lake BB, Ai R, Kaeser GE, Salathia NS, Yung YC, Liu R, et al. Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science*. 2016;352:1586–90. <https://doi.org/10.1126/science.aaf1204>.
79. Lake BB, Chen S, Sos BC, Fan J, Kaeser GE, Yung YC, et al. Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat Biotechnol*. 2018;36:70–80. <https://doi.org/10.1038/nbt.4038>.
80. Velmeshov D, Schirmer L, Jung D, Haeussler M, Perez Y, Mayer S, et al. Single-cell genomics identifies cell type-specific molecular changes in autism. *Science*. 2019;364:685–9. <https://doi.org/10.1126/science.aav8130>.
81. Bhaduri A, Di Lullo E, Jung D, Muller S, Crouch EE, Espinosa CS, et al. Outer radial glia–cancer stem cells contribute to heterogeneity of glioblastoma. *Cell Stem Cell*. 2020;26:48–63. <https://doi.org/10.1016/j.stem.2019.11.015>.
82. Jakel S, Aguirre E, Mendhana Falcao A, van Bruggen D, Lee KW, Knuesel I, et al. Altered human oligodendrocyte heterogeneity in multiple sclerosis. *Nature*. 2019;566:543–7. <https://doi.org/10.1038/s41586-019-0903-2>.
83. Schirmer L, Velmeshov D, Holmqvist S, Kaufmann M, Werneburg S, Jung D, et al. Neuronal vulnerability and multilineage diversity in multiple sclerosis. *Nature*. 2019;573:75–82. <https://doi.org/10.1038/s41586-019-1404-z>.
84. Mathys H, Davila-Velderrain J, Peng Z, Gao F, Mohammadi S, Young JZ, et al. Single-cell transcriptomic analysis of Alzheimer's disease. *Nature*. 2019;570:332–7. <https://doi.org/10.1038/s41586-019-1195-2>.
85. Camp JG, Badsha F, Florio M, Kanton S, Gerber T, Wilsch-Brauninger M, et al. Human cerebral organoids recapitulate gene expression programs of fetal neocortex development. *Proc Natl Acad Sci USA*. 2015;112:15672–7. <https://doi.org/10.1073/pnas.1520760112>.
86. Nowakowski TJ, Bhaduri A, Pollen AA, Alvarado B, Mostajo-Radji MA, Di Lullo E, et al. Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science*. 2017;358:1318–23. <https://doi.org/10.1126/science.aap8809>.
87. Bhaduri A, Andrews MG, Mancia Leon W, Jung D, Shin D, Allen D, et al. Cell stress in cortical organoids impairs molecular subtype specification. *Nature*. 2020;578:142–8. <https://doi.org/10.1038/s41586-020-1962-0>.

88. Zhong S, Zhang S, Fan X, Wu Q, Yan L, Dong J, et al. A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. *Nature*. 2018;555:524–8. <https://doi.org/10.1038/nature25980>.
89. Fan X, Dong J, Zhong S, Wei Y, Wu Q, Yan L, et al. Spatial transcriptomic survey of human embryonic cerebral cortex by single-cell RNA-seq analysis. *Cell Res*. 2018;28:730–45. <https://doi.org/10.1038/s41422-018-0053-3>.
90. Polioudakis D, de la Torre-Ubieta L, Langerman J, Elkins AG, Shi X, Stein JL, et al. A single-cell transcriptomic atlas of human neocortical development during mid-gestation. *Neuron*. 2019;103:785–801.e8. <https://doi.org/10.1016/j.neuron.2019.06.011>.
91. Zhong S, Ding W, Sun L, Lu Y, Dong H, Fan X, et al. Decoding the development of the human hippocampus. *Nature*. 2020;577:531–6. <https://doi.org/10.1038/s41586-019-1917-5>.
92. La Manno G, Gyllborg D, Codeluppi S, Nishimura K, Salto C, Zeisel A, et al. Molecular diversity of midbrain development in mouse, human, and stem cells. *Cell*. 2016;167:566–80.e19. <https://doi.org/10.1016/j.cell.2016.09.027>.
93. Pollen AA, Nowakowski TJ, Chen J, Retallack H, Sandoval-Espinosa C, Nicholas CR, et al. Molecular identity of human outer radial glia during cortical development. *Cell*. 2015;163:55–67. <https://doi.org/10.1016/j.cell.2015.09.004>.
94. Thomsen ER, Mich JK, Yao Z, Hodge RD, Doyle AM, Jang S, et al. Fixed single-cell transcriptomic characterization of human radial glial diversity. *Nat Methods*. 2016;13:87–93. <https://doi.org/10.1038/nmeth.3629>.
95. Johnson MB, Wang PP, Atabay KD, Murphy EA, Doan RN, Hecht JL, et al. Single-cell analysis reveals transcriptional heterogeneity of neural progenitors in human cortex. *Nat Neurosci*. 2015;18:637–46. <https://doi.org/10.1038/nn.3980>.
96. Nowakowski TJ, Pollen AA, Sandoval-Espinosa C, Kriegstein AR. Transformation of the radial glia scaffold demarcates two stages of human cerebral cortex development. *Neuron*. 2016;91:1219–27. <https://doi.org/10.1016/j.neuron.2016.09.005>.
97. Lee JH, Huynh M, Silhavy JL, Kim S, Dixon-Salazar T, Heiberg A, et al. De novo somatic mutations in components of the PI3K-AKT3-mTOR pathway cause hemimegalencephaly. *Nat Genet*. 2012;44:941–5. <https://doi.org/10.1038/ng.2329>.
98. Poduri A, Evrny GD, Cai X, Elhosary PC, Beroukhim R, Lehtinen MK, et al. Somatic activation of AKT3 causes hemispheric developmental brain malformations. *Neuron*. 2012;74:41–8. <https://doi.org/10.1016/j.neuron.2012.03.010>.
99. Marin-Valencia I, Guerrini R, Gleeson JG. Pathogenetic mechanisms of focal cortical dysplasia. *Epilepsia*. 2014;55:970–8. <https://doi.org/10.1111/epi.12650>.
100. Yu X, Zecevic N. Dorsal radial glial cells have the potential to generate cortical interneurons in human but not in mouse brain. *J Neurosci*. 2011;31:2413–20. <https://doi.org/10.1523/JNEUROSCI.5249-10.2011>.
101. Hansen DV, Lui JH, Flandin P, Yoshikawa K, Rubenstein JL, Alvarez-Buylla A, et al. Non-epithelial stem cells and cortical interneuron production in the human ganglionic eminences. *Nat Neurosci*. 2013;16:1576–87. <https://doi.org/10.1038/nn.3541>.
102. Letinic K, Zoncu R, Rakic P. Origin of GABAergic neurons in the human neocortex. *Nature*. 2002;417:645–9. <https://doi.org/10.1038/nature00779>.
103. Ma T, Wang C, Wang L, Zhou X, Tian M, Zhang Q, et al. Subcortical origins of human and monkey neocortical interneurons. *Nat Neurosci*. 2013;16:1588–97. <https://doi.org/10.1038/nn.3536>.
104. Liu SJ, Nowakowski TJ, Pollen AA, Lui JH, Horlbeck MA, Attenello FJ, et al. Single-cell analysis of long non-coding RNAs in the developing human neocortex. *Genome Biol*. 2016;17:67. <https://doi.org/10.1186/s13059-016-0932-1>.
105. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, et al. The accessible chromatin landscape of the human genome. *Nature*. 2012;489:75–82. <https://doi.org/10.1038/nature11232>.
106. Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015;518:317–30. <https://doi.org/10.1038/nature14248>.
107. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, et al. Comprehensive integration of single cell data. *Cell*. 2019;177:1888–1902.e1821.
108. Nott A, Holtman IR, Coufal NG, Schlaichetzki JCM, Yu M, Hu R, et al. Brain cell type-specific enhancer-promoter interactome maps and disease-risk association. *Science*. 2019;366:1134–9. <https://doi.org/10.1126/science.aaoy0793>.
109. Kadoshima T, Sakaguchi H, Nakano T, Soen M, Ando S, Eiraku M, et al. Self-organization of axial polarity, inside-out layer pattern, and species-specific progenitor dynamics in human ES cell-derived neocortex. *Proc Natl Acad Sci USA*. 2013;110:20284–9. <https://doi.org/10.1073/pnas.1315710110>.
110. Mariani J, Coppola G, Zhang P, Abzov A, Provini L, Tomasini L, et al. FOXG1-dependent dysregulation of GABA/glutamate neuron differentiation in autism spectrum disorders. *Cell*. 2015;162:375–90. <https://doi.org/10.1016/j.cell.2015.06.034>.
111. Pasca AM, Sloan SA, Clarke LE, Tian Y, Makinson CD, Huber N, et al. Functional cortical neurons and astrocytes from human pluripotent stem cells in 3D culture. *Nat Methods*. 2015. <https://doi.org/10.1038/nmeth.3415>.
112. Rigamonti A, Repetti GG, Sun C, Price FD, Reny DC, Rapino F, et al. Large-scale production of mature neurons from human pluripotent stem cells in a three-dimensional suspension culture system. *Stem cell Rep*. 2016;6:993–1008. <https://doi.org/10.1016/j.stemcr.2016.05.010>.
113. Sakaguchi H, Kadoshima T, Soen M, Narii N, Ishida Y, Ohgushi M, et al. Generation of functional hippocampal neurons from self-organizing human embryonic stem cell-derived dorsomedial telencephalic tissue. *Nat Commun*. 2015;6:8896. <https://doi.org/10.1038/ncomms9896>.
114. Muguruma K, Nishiyama A, Kawakami H, Hashimoto K, Sasai Y. Self-organization of polarized cerebellar tissue in 3D culture of human pluripotent stem cells. *Cell Rep*. 2015;10:537–50. <https://doi.org/10.1016/j.celrep.2014.12.051>.
115. Monzel AS, Smits LM, Hemmer K, Hachi S, Moreno EL, van Wullen T, et al. Derivation of human midbrain-specific organoids from neuroepithelial stem cells. *Stem cell Rep*. 2017;8:1144–54. <https://doi.org/10.1016/j.stemcr.2017.03.010>.
116. Xiang Y, Tanaka Y, Cakir B, Patterson B, Kim KY, Sun P, et al. hESC-derived thalamic organoids form reciprocal projections when fused with cortical organoids. *Cell Stem Cell*. 2019;24:487–97. <https://doi.org/10.1016/j.stem.2018.12.015>. e7
117. Qian X, Nguyen HN, Song MM, Hadjono C, Ogden SC, Hammack C, et al. Brain-region-specific organoids using mini-bioreactors for modeling ZIKV exposure. *Cell*. 2016;165:1238–54. <https://doi.org/10.1016/j.cell.2016.04.032>.
118. Cederquist GY, Asciolla JJ, Tchieu J, Walsh RM, Cornacchia D, Resh MD, et al. Specification of positional identity in forebrain organoids. *Nat Biotechnol*. 2019;37:436–44. <https://doi.org/10.1038/s41587-019-0085-3>.
119. Trujillo CA, Gao R, Negraes PD, Gu J, Buchanan J, Preissl S, et al. Complex oscillatory waves emerging from cortical organoids model early human brain network development. *Cell Stem Cell*. 2019;25:558–69. <https://doi.org/10.1016/j.stem.2019.08.002>.
120. Giandomenico SL, Mierau SB, Gibbons GM, Wenger LMD, Masullo L, Sit T, et al. Cerebral organoids at the air-liquid interface generate diverse nerve tracts with functional output. *Nat Neurosci*. 2019;22:669–79. <https://doi.org/10.1038/s41593-019-0350-2>.
121. Lancaster MA, Corsini NS, Wolfinger S, Gustafson EH, Phillips AW, Burkard TR, et al. Guided self-organization and cortical plate formation in human brain organoids. *Nat Biotechnol*. 2017;35:659–66. <https://doi.org/10.1038/nbt.3906>.
122. Qian X, Su Y, Adam CD, Deutschmann AU, Pather SR, Goldberg EM, et al. Sliced human cortical organoids for modeling distinct cortical layer formation. *Cell Stem Cell*. 2020. <https://doi.org/10.1016/j.stem.2020.02.002>.
123. Sloan SA, Darmanis S, Huber N, Khan TA, Birey F, Caneda C, et al. Human astrocyte maturation captured in 3D cerebral cortical spheroids derived from pluripotent stem cells. *Neuron*. 2017;95:779–90. <https://doi.org/10.1016/j.neuron.2017.07.035>.
124. Marton RM, Miura Y, Sloan SA, Li Q, Revah O, Levy RJ, et al. Differentiation and maturation of oligodendrocytes in human three-dimensional neural cultures. *Nat Neurosci*. 2019;22:484–91. <https://doi.org/10.1038/s41593-018-0316-9>.
125. Madhavan M, Nevin ZS, Shick HE, Garrison E, Clarkson-Paredes C, Karl M, et al. Induction of myelinating oligodendrocytes in human cortical spheroids. *Nat Methods*. 2018;15:700–6. <https://doi.org/10.1038/s41592-018-0081-4>.
126. Mansour AA, Goncalves JT, Bloyd CW, Li H, Fernandes S, Quang D, et al. An in vivo model of functional and vascularized human brain organoids. *Nat Biotechnol*. 2018;36:432–41. <https://doi.org/10.1038/nbt.4127>.
127. Cakir B, Xiang Y, Tanaka Y, Kural MH, Parent M, Kang YJ, et al. Engineering of human brain organoids with a functional vascular-like system. *Nat Methods*. 2019;16:1169–75. <https://doi.org/10.1038/s41592-019-0586-5>.
128. Worsdorfer P, Dalda N, Kern A, Kruger S, Wagner N, Kwok CK, et al. Generation of complex human organoid models including vascular networks by incorporation of mesodermal progenitor cells. *Sci Rep*. 2019;9:15663. <https://doi.org/10.1038/s41598-019-52204-7>.
129. Paredes I, Himmels P, Ruiz de Almodovar C. Neurovascular communication during CNS development. *Dev Cell*. 2018;45:10–32. <https://doi.org/10.1016/j.devcel.2018.01.023>.
130. Ormel PR, Vieira de Sa R, van Bodegraven EJ, Karst H, Harschnitz O, Sneeboer MAM, et al. Microglia innately develop within cerebral organoids. *Nat Commun*. 2018;9:4167. <https://doi.org/10.1038/s41467-018-06684-2>.
131. Song L, Yuan X, Jones Z, Vied C, Miao Y, Marzano M, et al. Functionalization of brain region-specific spheroids with isogenic microglia-like cells. *Sci Rep*. 2019;9:11055. <https://doi.org/10.1038/s41598-019-47444-6>.
132. Abud EM, Ramirez RN, Martinez ES, Healy LM, Nguyen CHH, Newman SA, et al. iPSC-derived human microglia-like cells to study neurological diseases. *Neuron*. 2017;94:278–93. <https://doi.org/10.1016/j.neuron.2017.03.042>.
133. Abreu CM, Gama L, Krasemann S, Chesnut M, Odwin-Dacosta S, Hogberg HT, et al. Microglia increase inflammatory responses in iPSC-derived human Brain-Spheres. *Front Microbiol*. 2018;9:2766. <https://doi.org/10.3389/fmicb.2018.02766>.
134. Birey F, Andersen J, Makinson CD, Islam S, Wei W, Huber N, et al. Assembly of functionally integrated human forebrain spheroids. *Nature*. 2017;545:54–9. <https://doi.org/10.1038/nature22330>.
135. Xiang Y, Tanaka Y, Patterson B, Kang YJ, Govindaiah G, Roselaar N, et al. Fusion of regionally specified hPSC-derived organoids models human brain development and interneuron migration. *Cell Stem Cell*. 2017;21:383–98. <https://doi.org/10.1016/j.stem.2017.07.007>.

136. Dang J, Tiwari SK, Agrawal K, Hui H, Qin Y, Rana TM. Glial cell diversity and methamphetamine-induced neuroinflammation in human cerebral organoids. *Mol Psychiatry*. 2020. <https://doi.org/10.1038/s41380-020-0676-x>
137. Marchetto MC, Carromeu C, Acab A, Yu D, Yeo GW, Mu Y, et al. A model for neural development and treatment of Rett syndrome using human induced pluripotent stem cells. *Cell*. 2010;143:527–39. <https://doi.org/10.1016/j.cell.2010.10.016>.
138. Pasca SP, Portmann T, Voineagu I, Yazawa M, Shcheglovitov A, Pasca AM, et al. Using iPSC-derived neurons to uncover cellular phenotypes associated with Timothy syndrome. *Nat Med*. 2011;17:1657–62. <https://doi.org/10.1038/nm.2576>.
139. Marchetto MC, Belinson H, Tian Y, Freitas BC, Fu C, Vadodaria K, et al. Altered proliferation and networks in neural cells derived from idiopathic autistic individuals. *Mol Psychiatry*. 2017;22:820–35. <https://doi.org/10.1038/mp.2016.95>.
140. Brennan KJ, Simone A, Jou J, Gelboin-Burkhart C, Tran N, Sangar S, et al. Modelling schizophrenia using human induced pluripotent stem cells. *Nature*. 2011;473:221–5. <https://doi.org/10.1038/nature09915>.
141. Fromer M, Roussos P, Sieberts SK, Johnson JS, Kavanagh DH, Perumal TM, et al. Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat Neurosci*. 2016;19:1442–53. <https://doi.org/10.1038/nn.4399>.
142. Reilly SK, Yin J, Ayoub AE, Emera D, Leng J, Cotney J, et al. Evolutionary genomics evolutionary changes in promoter and enhancer activity during human corticogenesis. *Science*. 2015;347:1155–9. <https://doi.org/10.1126/science.1260943>.
143. Velasco S, Kedaigle AJ, Simmons SK, Nash A, Rocha M, Quadrato G, et al. Individual brain organoids reproducibly form cell diversity of the human cerebral cortex. *Nature*. 2019;570:523–7. <https://doi.org/10.1038/s41586-019-1289-x>.
144. Yoon SJ, Elahi LS, Pasca AM, Marton RM, Gordon A, Revah O, et al. Reliability of human cortical organoid generation. *Nat Methods*. 2019;16:75–8. <https://doi.org/10.1038/s41592-018-0255-0>.
145. Quadrato G, Nguyen T, Macosko EZ, Sherwood JL, Min Yang S, Berger DR, et al. Cell diversity and network dynamics in photosensitive human brain organoids. *Nature*. 2017;545:48–53. <https://doi.org/10.1038/nature22047>.
146. Tanaka Y, Cakir B, Xiang Y, Sullivan GJ, Park IH. Synthetic analyses of single-cell transcriptomes from multiple brain organoids and fetal brain. *Cell Rep*. 2020;30:1682–9. <https://doi.org/10.1016/j.celrep.2020.01.038>.
147. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods*. 2019;16:1289–96. <https://doi.org/10.1038/s41592-019-0619-0>.
148. Haghverdi L, Lun ATL, Morgan MD, Marioni JC. Batch effects in single-cell RNA-seqencing data are corrected by matching mutual nearest neighbors. *Nat Biotechnol*. 2018;36:421–7. <https://doi.org/10.1038/nbt.4091>.
149. Mora-Bermudez F, Badsha F, Kanton S, Camp JG, Vernot B, Kohler K, et al. Differences and similarities between human and chimpanzee neural progenitors during cerebral cortex development. *eLife*. 2016;5. <https://doi.org/10.7554/eLife.18683>.
150. Field AR, Jacobs FMJ, Fiddes IT, Phillips APR, Reyes-Ortiz AM, LaMontagne E, et al. Structurally conserved primate lncRNAs are transiently expressed during human cortical differentiation and influence cell-type-specific genes. *Stem Cell Rep*. 2019;12:245–57. <https://doi.org/10.1016/j.stemcr.2018.12.006>.
151. Pollen AA, Bhaduri A, Andrews MG, Nowakowski TJ, Meyerson OS, Mostajo-Radji MA, et al. Establishing cerebral organoids as models of human-specific brain evolution. *Cell*. 2019;176:743–56.e17. <https://doi.org/10.1016/j.cell.2019.01.017>.
152. Kanton S, Boyle MJ, He Z, Santel M, Weigert A, Sanchis-Calleja F, et al. Organoid single-cell genomic atlas uncovers human-specific features of brain development. *Nature*. 2019;574:418–22. <https://doi.org/10.1038/s41586-019-1654-9>.
153. Dryden NH, Broome LR, Dudbridge F, Johnson N, Orr N, Schoenfelder S, et al. Unbiased analysis of potential targets of breast cancer susceptibility loci by Capture Hi-C. *Genome Res*. 2014;24:1854–68. <https://doi.org/10.1101/gr.175034.114>.
154. Fang R, Yu M, Li G, Chee S, Liu T, Schmitt AD, et al. Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq. *Cell Res*. 2016;26:1345–8. <https://doi.org/10.1038/cr.2016.137>.
155. Mumbach MR, Rubin AJ, Flynn RA, Dai C, Khavari PA, Greenleaf WJ, et al. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat Methods*. 2016;13:919–22. <https://doi.org/10.1038/nmeth.3999>.
156. Patwardhan RP, Lee C, Litvin O, Young DL, Pe'er D, Shendure J. High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. *Nat Biotechnol*. 2009;27:1173–5. <https://doi.org/10.1038/nbt.1589>.
157. Melnikov A, Murugan A, Zhang X, Tesileanu T, Wang L, Rogov P, et al. Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat Biotechnol*. 2012;30:271–7. <https://doi.org/10.1038/nbt.2137>.
158. Ulirsch JC, Nandakumar SK, Wang L, Giani FC, Zhang X, Rogov P, et al. Systematic functional dissection of common genetic variation affecting red blood cell traits. *Cell*. 2016;165:1530–45. <https://doi.org/10.1016/j.cell.2016.04.048>.
159. Tewhey R, Kotliar D, Park DS, Liu B, Winnicki S, Reilly SK, et al. Direct identification of hundreds of expression-modulating variants using a multiplexed reporter assay. *Cell*. 2016;165:1519–29. <https://doi.org/10.1016/j.cell.2016.04.027>.
160. Grossman SR, Zhang X, Wang L, Engreitz J, Melnikov A, Rogov P, et al. Systematic dissection of genomic features determining transcription factor binding and enhancer function. *Proc Natl Acad Sci USA*. 2017;114:E1291–E300. <https://doi.org/10.1073/pnas.1621150114>.
161. Inoue F, Kircher M, Martin B, Cooper GM, Witten DM, McManus MT, et al. A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity. *Genome Res*. 2017;27:38–52. <https://doi.org/10.1101/gr.212092.116>.
162. Inoue F, Kreimer A, Ashuach T, Ahituv N, Yosef N. Identification and massively parallel characterization of regulatory elements driving neural induction. *Cell Stem Cell*. 2019;25:713–27. <https://doi.org/10.1016/j.stem.2019.09.010>.
163. Arnold CD, Gerlach D, Stelzer C, Boryn LM, Rath M, Stark A. Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science*. 2013;339:1074–7. <https://doi.org/10.1126/science.1232542>.
164. Vanhille L, Griffon A, Maqbool MA, Zacarias-Cabeza J, Dao LT, Fernandez N, et al. High-throughput and quantitative assessment of enhancer activity in mammals by CapStarr-seq. *Nat Commun*. 2015;6:6905. <https://doi.org/10.1038/ncomms7905>.
165. Vockley CM, D'ippolito AM, McDowell IC, Majoros WH, Safi A, Song L, et al. Direct GR binding sites potentiate clusters of TF binding across the human Genome. *Cell*. 2016;166:1269–81. <https://doi.org/10.1016/j.cell.2016.07.049>.
166. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. *Science*. 2013;339:819–23. <https://doi.org/10.1126/science.1231143>.
167. Canver MC, Smith EC, Sher F, Pinello L, Sanjana NE, Shalem O, et al. BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature*. 2015;527:192–7. <https://doi.org/10.1038/nature15521>.
168. Gasperini M, Findlay GM, McKenna A, Milbank JH, Lee C, Zhang MD, et al. CRISPR/Cas9-mediated scanning for regulatory elements required for HPRT1 expression via thousands of large, programmed genomic deletions. *Am J Hum Genet*. 2017;101:192–205. <https://doi.org/10.1016/j.ajhg.2017.06.010>.
169. Sanjana NE, Wright J, Zheng K, Shalem O, Fontanillas P, Joung J, et al. High-resolution interrogation of functional elements in the noncoding genome. *Science*. 2016;353:1545–9. <https://doi.org/10.1126/science.aaf7613>.
170. Rajagopal N, Srinivasan S, Kooshesh K, Guo Y, Edwards MD, Banerjee B, et al. High-throughput mapping of regulatory DNA. *Nat Biotechnol*. 2016;34:167–74. <https://doi.org/10.1038/nbt.3468>.
171. Diao Y, Li B, Meng Z, Jung I, Lee AY, Dixon J, et al. A new class of temporarily phenotypic enhancers identified by CRISPR/Cas9-mediated genetic screening. *Genome Res*. 2016;26:397–405. <https://doi.org/10.1101/gr.197152.115>.
172. Hilton IB, D'ippolito AM, Vockley CM, Thakore PI, Crawford GE, Reddy TE, et al. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol*. 2015;33:510–7. <https://doi.org/10.1038/nbt.3199>.
173. Thakore PI, D'ippolito AM, Song L, Safi A, Shivakumar NK, Kabadi AM, et al. Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat Methods*. 2015;12:1143–9. <https://doi.org/10.1038/nmeth.3630>.
174. Fulco CP, Munschauer M, Anyoha R, Munson G, Grossman SR, Perez EM, et al. Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science*. 2016;354:769–73. <https://doi.org/10.1126/science.aag2445>.
175. Korkmaz G, Lopes R, Ugalde AP, Nevedomskaya E, Han R, Myacheva K, et al. Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. *Nat Biotechnol*. 2016;34:192–8. <https://doi.org/10.1038/nbt.3450>.
176. Xie S, Duan J, Li B, Zhou P, Hon GC. Multiplexed engineering and analysis of combinatorial enhancer activity in single cells. *Mol Cell*. 2017;66:285–99. <https://doi.org/10.1016/j.molcel.2017.03.007>.
177. Gasperini M, Hill AJ, McFaline-Figueira JL, Martin B, Kim S, Zhang MD, et al. A genome-wide framework for mapping gene regulation via cellular genetic screens. *Cell*. 2019;176:377–90. <https://doi.org/10.1016/j.cell.2018.11.029>.
178. Klein JC, Chen W, Gasperini M, Shendure J. Identifying novel enhancer elements with CRISPR-based screens. *ACS Chem Biol*. 2018;13:326–32. <https://doi.org/10.1021/acscchembio.7b00778>.
179. Habib N, Avraham-David I, Basu A, Burks T, Shekhar K, Hofree M, et al. Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat Methods*. 2017;14:955–8. <https://doi.org/10.1038/nmeth.4407>.
180. Zhang Y, Sloan SA, Clarke LE, Caneda C, Plaza CA, Blumenthal PD, et al. Purification and characterization of progenitor and mature human astrocytes reveals transcriptional and functional differences with mouse. *Neuron*. 2016;89:37–53. <https://doi.org/10.1016/j.neuron.2015.11.013>.

ARTICLE



<https://doi.org/10.1038/s41467-021-27464-5>

OPEN

Single-cell transcriptomics captures features of human midbrain development and dopamine neuron diversity in brain organoids

Alessandro Fiorenzano  ¹✉, Edoardo Sozzi  ¹, Marcella Birtele  ¹, Janko Kajtez  ¹, Jessica Giacomoni  ¹, Fredrik Nilsson  ¹, Andreas Bruzelius  ², Yogita Sharma¹, Yu Zhang  ¹, Bengt Mattsson¹, Jenny Emnéus³, Daniella Rylander Ottosson  ², Petter Storm  ¹ & Malin Parmar  ¹

Three-dimensional brain organoids have emerged as a valuable model system for studies of human brain development and pathology. Here we establish a midbrain organoid culture system to study the developmental trajectory from pluripotent stem cells to mature dopamine neurons. Using single cell RNA sequencing, we identify the presence of three molecularly distinct subtypes of human dopamine neurons with high similarity to those in developing and adult human midbrain. However, despite significant advancements in the field, the use of brain organoids can be limited by issues of reproducibility and incomplete maturation which was also observed in this study. We therefore designed bioengineered ventral midbrain organoids supported by recombinant spider-silk microfibers functionalized with full-length human laminin. We show that silk organoids reproduce key molecular aspects of dopamine neurogenesis and reduce inter-organoid variability in terms of cell type composition and dopamine neuron formation.

¹ Developmental and Regenerative Neurobiology, Wallenberg Neuroscience Center, and Lund Stem Cell Center, Department of Experimental Medical Science, Lund University, Lund, Sweden. ² Regenerative Neurophysiology, Wallenberg Neuroscience Center, Lund Stem Cell Center, Department of Experimental Medical Science, Lund University, Lund, Sweden. ³ Department of Biotechnology and Biomedicine (DTU Bioengineering), Technical University of Denmark, Lyngby, Denmark. ✉email: alessandro.fiorenzano@med.lu.se

The ability to control cell-fate specification and drive differentiation of human pluripotent stem cells (hPSCs) into regionally specified neuronal subtypes has opened up new avenues of research into human-specific brain development, disease modeling, and cell-based therapies. hPSC differentiation is most routinely carried out in two-dimensional (2D) cultures, which poorly recapitulate the cellular composition and structural complexity of human brain. Although models of human neural development that better recreate the intricacy of different regions of the developing neural tube are being developed^{1–3}, it remains challenging to obtain mature cells in these systems. Three-dimensional (3D) human-brain organoids have rapidly become a widely adopted system to study the development and function/dysfunction of neuronal cells, as it provides a more physiologically relevant cellular context and allows for long-term maintenance of functionally mature neurons^{4–6}. Recent advancements in single-cell sequencing technologies have increased the scope for dissecting organoid cultures to define cell-type composition and provide an opportunity to study brain development, cell diversity, and gene regulation of otherwise inaccessible human cells. To date, most studies of this sort have been conducted in cerebral organoids. For example, single-cell transcriptomics were used to map cell types and developmental states within organoids⁷, to explore cortical development⁸, and to define human-specific gene-regulatory changes⁹. However, the full potential of organoid models is still restricted by issues of reproducibility and variability in terms of morphology, cellular makeup, and activity^{10–13}.

In this study, we patterned hPSC-derived brain organoids into a ventral midbrain (VM) identity using a protocol that results in the formation of midbrain dopamine (DA) progenitors and of functionally mature DA neurons after transplantation in xeno-graft models of Parkinson's disease (PD)¹⁴. DA neurons in these organoids exhibited mature electrophysiological properties, neuromelanin production, and the ability to release DA, confirming the long-term maintenance of functionally mature human DA neurons in 3D culture as previously reported in both hPSC-derived^{15–18} and neural progenitor-derived^{19–21} midbrain organoids. A time-course transcriptional analysis of human VM development and DA neuron differentiation at single-cell level revealed four populations of cells with high transcriptional similarity to VM floor-plate cells, followed by the stepwise emergence of neurons, vascular leptomeningeal cells (VLMCs), astrocytes, and oligodendrocytes. Importantly, we found that mature neurons and glia formed within the organoids displayed high similarity to the corresponding cell types in adult human midbrain²². The large number of cells within the DA neuron cluster (14,606 cells) enabled us to perform a detailed molecular dissection of mature human DA neurons not previously possible. We identified several molecularly distinct human DA neuron subgroups, similarly to recent observations in developing and adult mouse brain^{23–25}.

These brain organoids showed a lower level of variability compared with self-organized 3D structures^{4,7}, and a similar level of variability as other protocols based on extrinsic patterning factors^{15,26}. To further reduce variability in the organoids, we used recombinant spider silk microfibers functionalized with full-length human laminin. These bioengineered silk-VM organoids reproduced key molecular aspects of DA neurogenesis but with more precise patterning and less organoid-to-organoid as well as intra-organoid variation. Furthermore, detailed functional and molecular analyses showed that the silk scaffolds reduced necrosis and supported neuronal maturation in all regions of the organoid. In sum, silk-fiber scaffolding is an experimentally straightforward method that does not require specialized equipment or technical expertise and can thus be widely implemented in organoid culturing.

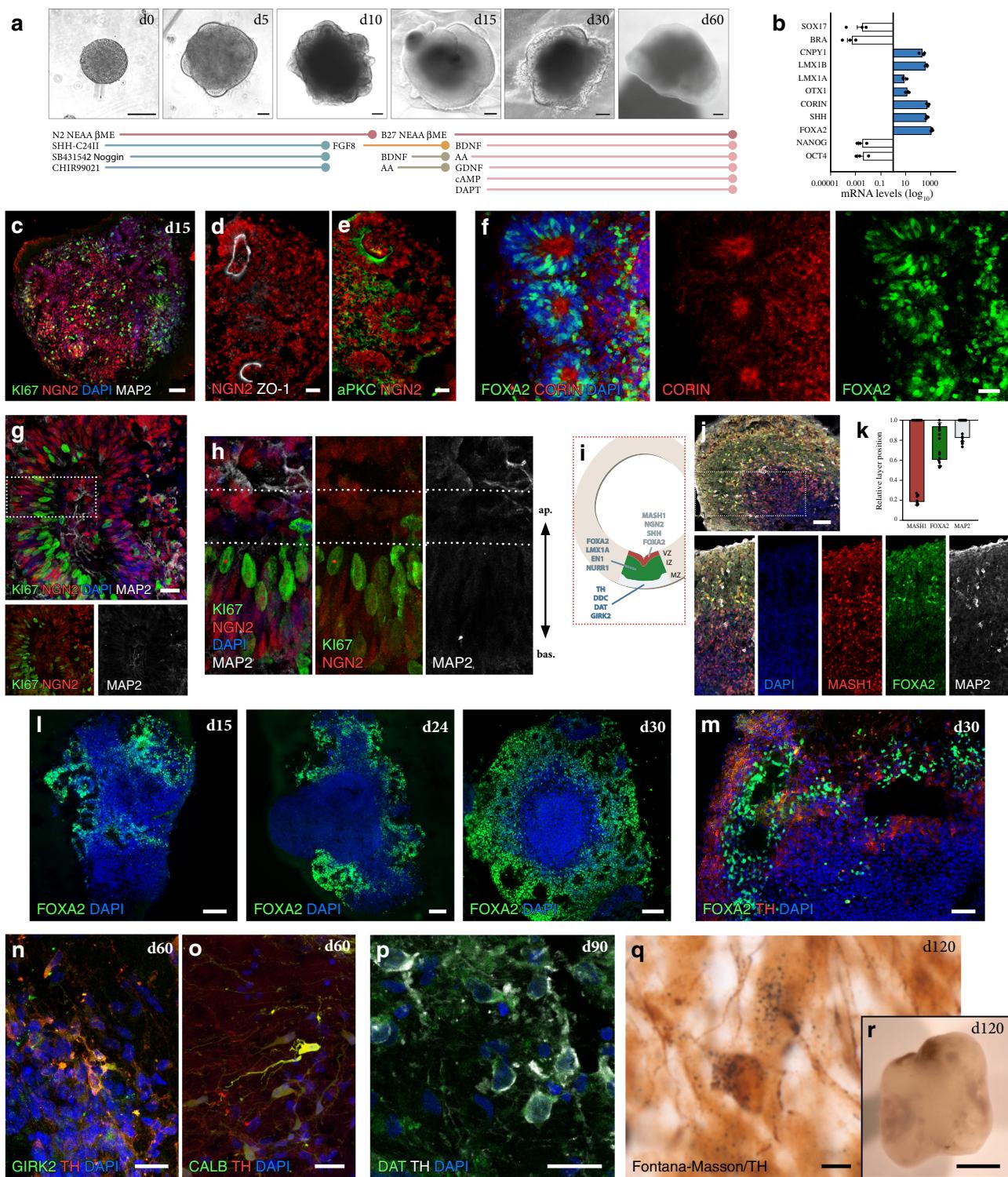
Results

Key features of human VM development can be recapitulated in 3D culture. We established VM organoids by adapting a commonly used protocol for the generation of forebrain organoids²⁷ with the addition of dual-SMAD inhibition combined with exposure to SHH, GSK3i, and FGF8 in the same temporal sequence used in 2D culture to promote neurogenic conversion of VM FP progenitors toward a DA neuron identity (Fig. 1a)^{28,29}. We confirmed that in 3D cultures, this patterning regime also resulted in efficient neuralization and ventral midbrain differentiation as evaluated by expression of *CNPY1*, *LMX1a*, *LMX1b*, *OTX1*, *CORIN*, *SHH*, and *FOXA2* (Fig. 1b). The expression of midbrain genes was accompanied by complete downregulation of pluripotency-associated genes *OCT4* and *NANOG* (Fig. 1b). Expression of early neural marker *NGN2* (Fig. 1c–e) as well as tight-junction protein ZO1 (Fig. 1d) and atypical protein kinase aPKC (Fig. 1e) was detected by immunocytochemistry at early stage of organoid formation. Although the architectural arrangements morphologically resembled the intermediate rosette stage typical of anterior neuroectoderm³⁰, they were PAX6-negative (Supplementary Fig. 1a–c) and expressed both *CORIN* and *FOXA2* (Fig. 1f), in accordance with their VM identity. Successful 3D VM differentiation was reproduced using two additional cell lines, H9 and HS1001 (Supplementary Fig. 1d–k), demonstrating the robustness of the protocol.

Staining for the cell cycle marker Ki67 showed that cell proliferation was largely confined to the basal region, while postmitotic neurons detected by MAP2 were located in the apical regions (Fig. 1g, h). We further assessed the presence of different developmental zones as defined by expression of *MASH1*, *FOXA2*, and *MAP2* (Fig. 1i–k) and found a layered-restricted organization (Fig. 1j). This was confirmed by quantification of cells in different anatomical locations of the organoids, showing that *FOXA2*- and *MAP2*-expressing cells were concentrated to the outer layer (Fig. 1k). By the second week, *FOXA2*-positive cells appeared in the organoids (Fig. 1l), and from day 30, TH-expressing neurons were detected (Fig. 1m). To directly compare our VM organoids with other hPSC-derived midbrain-patterned organoids generated using previously published protocols, we recreated midbrain-like organoids (MLO) according to Jo et al.¹⁵ and dorsomorphin A82-01 midbrain organoids (DA-MO) according to Kwak et al.¹⁷ as well as forebrain organoids (FBO) as reported in Lancaster et al.²⁷. Quantifications revealed a similar number of cells expressing the key VM markers *FOXA2*, *LMX1*, and *TH* in all three midbrain patterned organoids (Supplementary Fig. 1l–p), while *PAX6* was only expressed in FBO (Supplementary Fig. 1a, p).

After 60 days, expression of G-protein-regulated inwardly rectifying potassium channel 2 (GIRK2) (Fig. 1n and Supplementary Fig. 1q), calcium-binding protein 1 (CALB1) (Fig. 1o and Supplementary Fig. 1r), and DA transporter (DAT) (Fig. 1p) indicated that mature DA neuron subtypes had emerged within the organoids. Quantifications at day 60 confirmed comparable differentiation into mature DA neurons as from previously reported protocols (Supplementary Fig. 1q–w). Specification toward a mature and authentic A9-like DA population, which consists of pigmented neurons located in SNC in primate VM—and which is prevalently lost in PD—was corroborated by combined Fontana–Masson staining/TH immunohistochemistry revealing the presence of intracellular and extracellular neuromelanin, visible as dark granular pigmentation after long-term culture (Fig. 1q, r).

scRNASeq reveals cellular composition and developmental trajectory in VM organoids. We next performed a 10X genomics droplet-based single-cell time-course transcriptomic analysis of



human VM organoid development (Fig. 2a) by profiling a total of 91,034 single cells isolated from organoids at days 15, 30, 60, 90, and 120 of VM organoid differentiation with several replicates per time point (day 15, $n = 2$; day 30, $n = 4$; day 60, $n = 5$; day 90, $n = 2$; day 120, $n = 6$). After integration using the Harmony approach³¹, uniform manifold approximation and projection (UMAP) and graph-based clustering visualized eight different clusters (Fig. 2b), all assigned a neuroectodermal identity (Fig. 2b–d and Supplementary Fig. 2c) with no remaining pluripotent cells (Supplementary Fig. 3a), mesodermal (*T*), or endodermal derivatives (*SOX17*) (Supplementary Fig. 3b). All

identified clusters showed distinct separations with high average silhouette widths (Supplementary Fig. 2b).

The most highly differentially expressed genes were used together with canonical markers to manually annotate clusters with their respective cellular identities. The yellow cluster in the UMAP space was made up of cells expressing neural markers (*HES1*, *NES*, and *SOX2*) (Fig. 2d and Supplementary Fig. 3c) and cells with a highly proliferative signature (*TOP2A*, *CCNB2*, and *CENPF1*) (Fig. 2e, f), which we named FP-0. These cells display similar characteristics to cycling FP progenitors located in VM at early stages of embryonic development^{32,33}. FP-0 population was

Fig. 1 Dopamine neurogenesis in VM organoids. **a** Representative bright-field images of ventral midbrain (VM) organoid differentiation at different time points (upper) and schematic overview of the experimental design (lower). Scale bars, 200 μ m. **b** qRT-PCR of selected markers at day 15 of VM organoid differentiation. The results are given as fold change over undifferentiated hPSCs. Data represent mean \pm SEM of 3 independent organoids. **c–e** Immunocytochemistry of (**c**) NGN2/Ki67/MAP2, (**d**) NGN2/ZO-1, and (**e**) NGN2/aPKC at day 15 during VM organoid differentiation. Scale bars, 50 μ m. **f** Immunohistochemistry of FOXA2/CORIN and **g** NGN2/Ki67/MAP2 in VM organoid at day 15. Scale bars, 100 μ m (**f**) and 50 μ m (**g**). **h** High-power magnification of the inset in **g**. **i** Schematic representation of developing DA neurons in vivo showing genes expressed at different stages of development (MZ, mantle zone; IZ, intermediate zone; VZ, ventricular zone). **j** Immunohistochemistry of MASH1, FOXA2, and MAP2 showing layer-specific organization and **k** relative quantification of fluorescence staining in VM organoid at day 24. Data represent mean \pm SEM of 7 independent VM organoids. Scale bar, 100 μ m. **l** Immunohistochemistry of FOXA2 across a time course (day 15–30). Scale bars, 50 μ m. **m** Cryosection of VM organoid at month 1 showing TH/FOXA2 double staining. Scale bars, 100 μ m. **n, o** Immunohistochemistry of TH stained with GIRK2/CALB at day 60 and **p** with DAT at day 90. Scale bars, 50 μ m (**n, o**) and 20 μ m (**p**). **q** Fontana Masson/TH double-stained cryosection from long-term cultured VM organoid (month 4). Scale bars, 50 μ m. **r** Representative bright-field image of VM organoid at month 4. Scale bars, 1 mm. Nuclei were stained with DAPI in **g, h, j, l–p**. Source data are provided as a Source Data file.

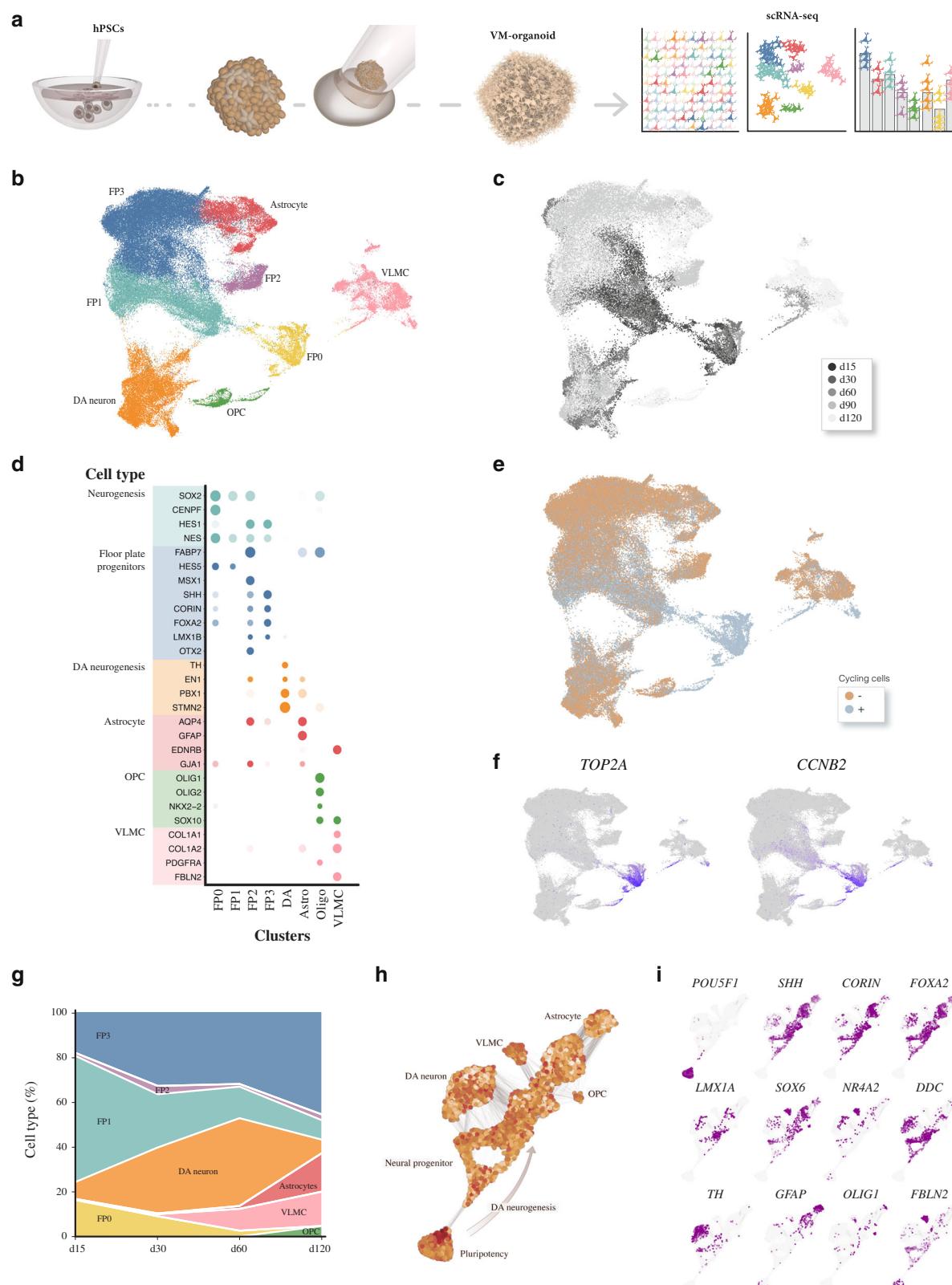
the prominent cluster at the early time points (Fig. 2c and Supplementary Fig. 4a, b), and based on pseudo-time-inference reconstruction analysis (Supplementary Fig. 4c) gave rise to all the other identified cell clusters (see later section). UMAP also visualized another large FP-like progenitor population that was further subdivided into three different clusters, referred to as FP-1 (light green), FP-2 (purple), and FP-3 (blue) (Fig. 2b). The concomitant expression of *HES5* and *SOX2* with *VIM* and *FABP7* (also known as *BLBP1*) (Fig. 2d and Supplementary Fig. 3c, d) indicated that FP-1 shares many features of radial glial cells and is also enriched in cell cycle genes (Fig. 2e). FP-2 was instead enriched in cells expressing *SHH* and *CORIN*, and also contained early DA progenitor markers, including *FOXA2*, *LMX1A*, *MSX1*, and *OTX2* (Fig. 2d and Supplementary Fig. 3e). The FP-3 cluster mainly differed from the FP-2 cluster in the additional expression of neuronal precursor markers *STMN2* and *SYT1*, and DA progenitor markers *EN1* and *DDC* (Fig. 2d and Supplementary Fig. 2c and Supplementary Fig. 3f, g), as well as its exit from cell cycle (Fig. 2e, f). Furthermore, scRNAseq analysis revealed the absence of forebrain (*FOXG1*) and only very few scattered hindbrain (*HOXA2*) cells, indicating efficient VM patterning (Supplementary Fig. 3d). In agreement, the neuronal cluster (orange in Fig. 2b) defined by expression of *DCX*, *SYT1*, and *STMN2*, primarily expressed genes associated with DA-fate identity (*PBX1*, *NR4A2/NURR1*, *EN1*, *TH*, and *DDC*) (Fig. 2d and Supplementary Fig. 2c and Supplementary Fig. 3f, g), with only few cells showing GABAergic (*VGAT*) and glutamatergic (*VGLUT1*) features (Supplementary Fig. 3k). VM organoids also contained a newly discovered class of perivascular-like cells termed vascular leptomeningeal cells (VLMCs) expressing *PDGFRa*, *COL1A1*, *COL1A2*, and *LUM*, astrocytes (*GFAP*, *AQP4*, and *EDNRB*) and a small cluster of oligodendrocyte progenitors (*OLIG1/2*, *PDGFRa*, and *SOX10*) (Fig. 2b, d and Supplementary Fig. 3h–j and l–n). Consistent with efficient VM patterning, the single-cell dataset generated here correlated well with published bulk and single-cell sequencing of midbrain 3D cultures derived from either hPSCs or neural progenitors (Supplementary Fig. 5a)^{15,20,21} and, as expected, displayed a lower correlation to FBO (Supplementary Fig. 5a). We also performed scRNAseq of one-month-old MLOs generated using the protocol described by Jo et al. (29,112 cells analyzed from two replicates), which showed the presence of the same cell types at the same timepoints (Supplementary Fig. 5b).

Analysis of organoid composition over time revealed that the different cell types appeared in a temporal pattern: the yellow cluster was the first emerging progenitor population, while FP-2 and FP-3 appeared slightly later (Fig. 2g and Supplementary Fig. 4a, b). Unlike FP-0 and FP-1, FP-3 increased proportionally in frequency over time and FP-3 was the only population still present in high numbers at day 120 (Fig. 2g and Supplementary

Fig. 4a, b). The DA cluster started to emerge from day 15 and was present in high numbers at all subsequent time points analyzed, although its relative proportion varied due to the increased presence of other cell types at later timepoints (Fig. 2g and Supplementary Fig. 4a, b). VLMCs, astrocytes, and oligodendrocyte progenitors appeared in a sequential manner (Fig. 2g and Supplementary Fig. 4a, b). When organizing cells—from pluripotent to mature differentiated states—according to transcriptional similarity along a temporal axis, force-directed *k*-nearest-neighbor graph-based pseudotime trajectory defined distinct branches segregating from the FP progenitor cells (Fig. 2h, i). At month 1, a first branch gave rise to DA progenitors, which by month 2 had started to mature into postmitotic DA neurons (Fig. 2h, i). A second branch trajectory toward vascular stromal progenitors able to differentiate into VLMC progenitors after month 1, but into more mature cell types only from month 3 (Fig. 2h, i). These findings were corroborated by Slingshot analysis (Supplementary Fig. 4c), where lineages are identified by treating clusters of cells as nodes in a graph and drawing a minimum spanning tree between nodes³⁴.

Molecular diversity in human DA neuron cluster. We next investigated whether the mature cell types generated in VM-patterned organoids were transcriptionally similar to those present in a recently published snRNA-seq dataset from adult human midbrain containing ~6000 midbrain nuclei derived from five adult individuals²². This study identified distinct cell types in the adult midbrain: astrocytes, oligodendrocytes, oligodendrocyte progenitors (OPCs), microglia, endothelial cells, and neurons²². Merged clustering with our VM organoid dataset showed that the cell types present in organoids and human midbrain were transcriptionally similar, with the exception of microglia, which were not present in the organoids (Fig. 3a). Interestingly, all cell types in VM-patterned organoids displayed much higher transcriptional similarity to the corresponding cell types in human midbrain than to those in the cortex from the same dataset (~10,700 cortical cells from the middle frontal gyrus) (Fig. 3b).

Histological analysis (Fig. 1) indicated that DA neurons mature over time in the organoids, and that markers enriched in the two subtypes A9 and A10 neurons were present in long-term cultures (Fig. 1n–r). Several recent studies based on scRNAseq describe a greater-than-expected molecular DA neuron diversity, and at least five different molecular subtypes are reported in adult mouse VM^{23–25,35}. However, similar datasets for mature human DA neurons do not yet exist. To determine if distinct molecular subtypes of human DA neurons appear in the organoids, we isolated the DA compartment (14,606 cells, all time points) and reran the integration (Harmony) and clustering. A SPRING plot visualized two major populations (Fig. 3c), one mostly present at



early time points and one present at late stages of VM organoid differentiation, which we termed DA^{Early} and DA^{Late}, respectively (Fig. 3c, d). Both DA^{Early} and DA^{Late} expressed TH (Fig. 3e). The expression of embryonic/early neural markers (NES, SOX2, and RFX4) in DA^{Early} confirmed their relatively immature neuronal state, whereas the expression of mature, postmitotic DA markers (NURR1/NR4A2, VMAT-2/SLC18A2) and voltage-

gated potassium- and sodium-channel subfamily members (KCNC2, SCN2A) defined the more mature DA population (Fig. 3e).

The resulting network map (Fig. 3f) indicated transcriptional diversity within this cluster, prompting us to perform a higher-resolution analysis in order to detect the potential existence of human DA neuron molecular subtypes. The refined graph-based

Fig. 2 Single-cell transcriptomics identifying VM organoid cell types. **a** Schematic overview of the experimental design. hPSCs differentiated into regionalized VM organoids for up to 120 days, were analyzed at single-cell resolution. **b** 2D scatterplot of uniform manifold approximation and projection (UMAP) embeddings showing clustering of 91,034 analyzed cells from VM organoids at days 15, 30, 60, 90, and 120. Cell-type assignments are indicated. **c** UMAP plot of cells color-coded by organoid of origin. **d** Dot plot showing expression levels of indicated genes in each cluster. Indicated genes are established markers for neural progenitors, floor-plate progenitors, DA neurons, astrocytes, oligodendrocyte progenitors (OPCs), and vascular leptomeningeal cells. **e** UMAP plots showing cell cycle classification of analyzed cells (Seurat CellCycleScoring predictions). Cycling cells shown with gray dots. **f** Expression levels of indicated cell cycle genes visualized in the UMAP plots. **g** Proportion of each cell type along the temporal axis during VM organoid differentiation (day 15–120). **h** Developmental trajectory from pluripotency to terminally differentiated stages in VM organoid reconstructed using SPRING in VM organoid. Pseudocells are color-graded by total count. **i** SPRING plot colored (purple) by marker gene expression in emerging cellular clusters.

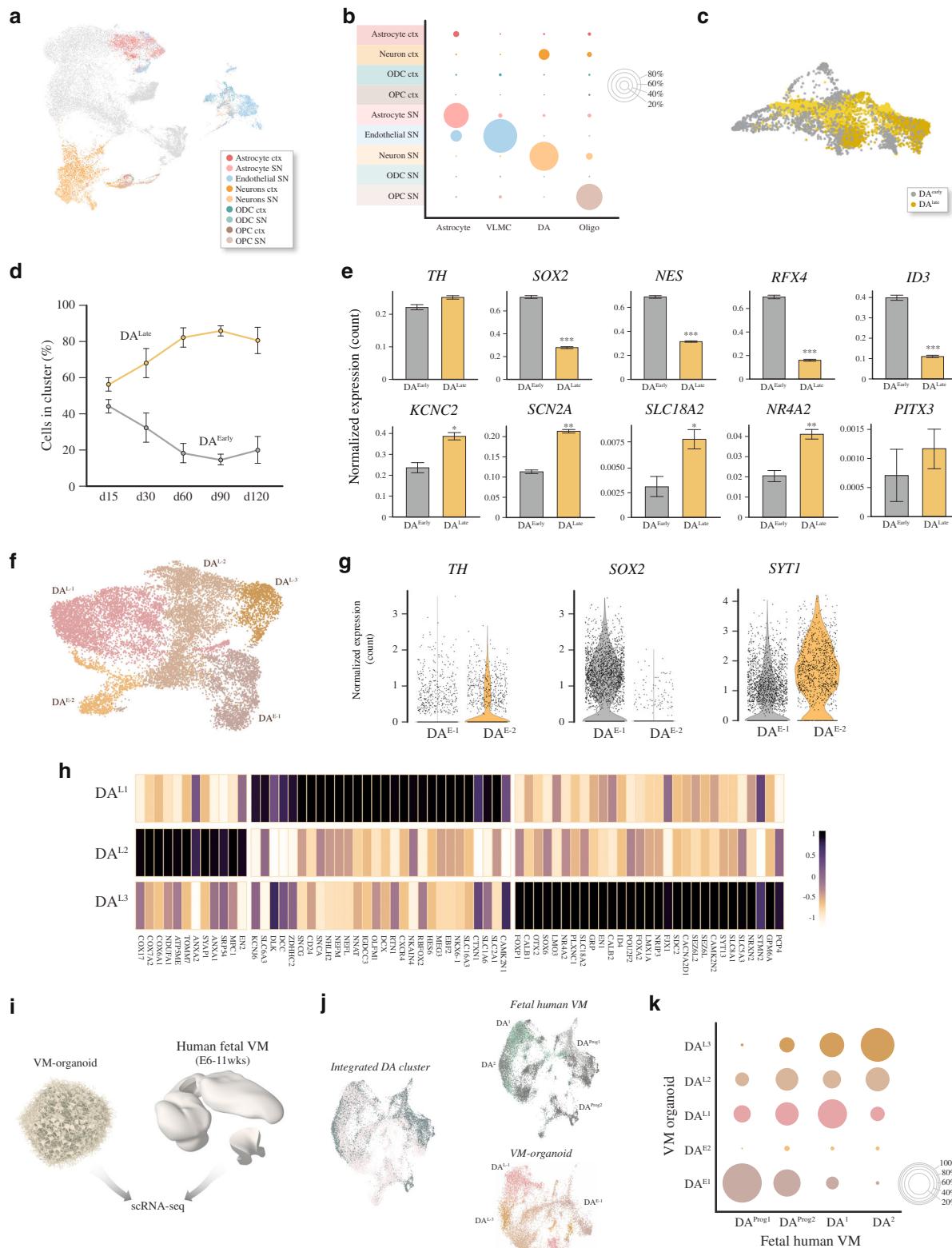
clustering segregated the early DA neurons into two distinct clusters (Fig. 3f): one with low TH (DA^{E-1}) and one with high TH (DA^{E-2}) (Fig. 3g). DA^{E-1} also showed increased expression of neural markers (SOX2), and reduced level of SYT1 compared with DA^{E-2} (Fig. 3g). The late DA neurons segregated into three distinct molecular identities (named DA^{L-1} , DA^{L-2} , and DA^{L-3}) (Fig. 3f). Within the DA^{L-1} subcluster we found concomitant expression of *DLK1*, *KCNJ6* (also known as *GIRK2*), *SLC6A3* (*DAT*), and *DCC*. Interestingly, *SNCG* and *SNCA* (encoding members of the synuclein family of proteins), glycoprotein CD24, and transcription factors *ZDHHC2* and *NHLH2*, which were all observed in this subcluster, were also found enriched in SNC from mouse bulk and scRNAseq datasets³⁶ (Fig. 3h). Synapse-associated protein 1 (*SYAP1*) transcription factors and engrailed homeobox 2 (EN2) were significantly expressed in DA^{L-2} . This subcluster also expressed *ANXA1*, encoding for a calcium-dependent phospholipid-binding protein, recently found associated with SNC at different developmental stages^{25,37,38}, and a large set of genes coding for components of respiratory electron-transport complexes (*COX17*, *NDUF*, and *ATP5ME*) as well as the brain mitochondrial receptor (*MPC1*) (Fig. 3h). The DA^{L-3} subcluster was molecularly defined by expression of *OTX2* and *CALB1*, markers of A10 DA neurons, while *DLK1* also appeared enriched in both DA^{L-3} and DA^{L-2} (Fig. 3h)^{23,39}. Importantly, the DA^{L-3} cluster was also enriched in genes previously identified at single-cell level during mouse development up until adulthood, such as *POU2F2* and *ID4*, as candidate regulators that may drive A10 subtype diversification (Fig. 3h). A set of genes associated with neuropsychiatric conditions, including Alzheimer's disease (*CLU*, *P4HA1*), schizophrenia (*CNIH2*, *DKK3*), and autism-spectrum disorders (*SEZ6L*, *SDC2*), was found particularly upregulated in this cluster (Fig. 3h)⁴⁰.

To assess to what extent the molecular identity of DA neurons in the organoids corresponds to that of authentic human DA neurons, we compared our single cell data with two scRNA-seq datasets of fetal VM DA populations, one previously reported by La Manno et al.³⁵ (Supplementary Fig. 5c, d) and the other obtained from 6 to 11-week post-conception human embryos and cultures thereof (Birtele et al., bioRxiv doi.org/10.1101/2020.10.01.322495) (Fig. 3i). In this latter fetal dataset, containing more mature neurons, 18,848 human fetal DA cells formed four molecularly distinct DA populations, two of which consisted of DA progenitors (gray) and the remaining two more mature populations (green) (Fig. 3j and Supplementary Fig. 5e). To confidently define organoid DA populations, we integrated fetal and hPSC-derived data, and normalized and clustered the gene expression matrix, identifying commonalities visualized via UMAP (Fig. 3k). We found a high similarity between developing and mature DA cell populations in hPSC-derived subclusters and their fetal DA neuronal counterparts (Fig. 3k), demonstrating that DA neurons in VM organoids have a similar molecular identity to authentic midbrain DA neurons sourced from human fetal brain. A detailed comparison between the four clusters of fetal VM-

derived DA neurons with the five clusters of DA neurons detected in VM organoids revealed that DA progenitors from fetal brain showed high transcriptional similarity to DA^{E-1} in the organoids, unlike DA^{E-2} (Fig. 3k). Moreover, $DA^{L1,2,3}$ all showed high similarity to the mature fetal VM-sourced DA neurons (Fig. 3k).

Molecular and functional heterogeneity in VM organoids. scRNAseq followed by clustering of sample-to-sample correlations (Pearson) (Fig. 4a) and principal component analysis (PCA) (Supplementary Fig. 6a) revealed that organoids analyzed at the same developmental stage (days 30, 60, and 120) contained the same cell types (Fig. 4b), confirming the reproducibility of this protocol. However, relative frequency analysis quantifying changes in cell-type composition revealed high variability in the proportion of cell types within each cluster from organoid to organoid (Fig. 4b) even though the VM organoids were generated from a small number of hPSCs (2500 cells) following an optimized protocol reported to reduce organoid-to-organoid heterogeneity and increase long-term viability of 3D structures^{7,41}. In addition to the variation between individual organoids, intra-organoid heterogeneity was observed in serial confocal TH-stained sections (Fig. 4c), showing that VM organoids exhibited a poorly developed core with sparse TH⁺ neurons, suggesting that nonsynchronous differentiation and maturation takes place. To test this, we performed whole-cell patch-clamp recordings to assess functional maturation of neurons within organoids using a recently reported method based on embedding in low-melting-point agarose⁴², allowing recordings in both the interior and exterior regions. We found that the cells at the surface of the organoids (Fig. 4d) exhibited more hyperpolarized resting-membrane potentials (Supplementary Fig. 6b, c) and rapidly inactivating inward sodium (Na⁺) and outward delayed-rectifier potassium (K⁺) currents (Fig. 4e) indicative of a mature neuronal state. In line with these findings, cells in the external part ($n = 16$) displayed the ability to fire induced action potentials (APs) upon current injections, indicating a neuronal function (Fig. 4f). These cells also showed spontaneous firing at resting-membrane potential (Fig. 4g) as well as a rebound depolarization (Fig. 4h) typical for DA neuron phenotype. In contrast, when recording from cells located in the inner region (Fig. 4i), no inward Na⁺ and outward K⁺ voltage-dependent currents or abortive APs were observed ($n = 20$ cells) (Fig. 4j–l). This distribution of functional cells located toward the edge of the organoids was confirmed using two additional protocols for PSC-derived mid-brain organoids (Supplementary Fig. 6d–i).

Generation and characterization of silk-bioengineered VM organoids. In an attempt to create more homogeneous VM organoids and further reduce organoid-to-organoid variability, we evaluated a biomaterial made of recombinant spider-silk protein^{43,44} that can self-assemble into a biocompatible cell scaffold. Silk scaffolds in the form of a network of microfiber



solution were obtained by placing a 20 μ l droplet at the bottom center of a hydrophobic well and then introducing air bubbles by repeatedly pipetting air into the droplet (Fig. 5a and Supplementary Fig. 7a). Via self-assembly of the silk protein, a thin film was formed around each air bubble, producing a temporary foam (Fig. 5b and Supplementary Fig. 7a). hPSCs were then dispersed throughout the 3D silk scaffold to obtain the controlled cell

distribution and adherence of cells within the network along the entire length of the microfibers (Fig. 5b and Supplementary Fig. 7b). Subsequently, during incubation in culture media, the foam collapses as the film around the air bubbles bursts, leaving a network of microfibers (Fig. 5c, d). Silk fibers were used either alone as an inert scaffold or functionalized with Lam-111, previously shown to promote DA patterning and support DA

Fig. 3 Single-cell transcriptomics mapping DA diversity in VM organoids. **a** UMAP cluster-integration analysis combining published scRNA-seq datasets of adult human midbrain¹⁸ and the hPSC-derived VM organoids with **b**, relative overlapping quantification. **c** SPRING network plot showing the distribution of single cells in 2 dopamine (DA) clusters (DA^{Early}, gray, and DA^{Late}, yellow) within the VM organoids. **d** Percent distribution of DA^{Early} and DA^{Late} clusters across a time course during VM organoid differentiation (d15 n = 2; d30 n = 5; d60 n = 5; d90 n = 2; d120 n = 6). Data represent mean ± SD per 10X run. **e** Bar plot of normalized expression for DA^{Early} and DA^{Late} clusters of immature and mature neuronal marker genes (d15 n = 2; d30 n = 5; d60 n = 5; d90 n = 2; d120 n = 6). Data represent mean ± SEM, two-tailed Wilcoxon Rank Sum test, KCNC2 p = 0.0045; SCN2A p = 0.0002; SLC18A2 p = 0.0269; NR4A2 p = 0.0013, ***p < 0.0001. **f** UMAP plot showing DA subclusters after reintegration and clustering (DA^{E-1}, DA^{E-2}, DA^{L-1}, DA^{L-2}, and DA^{L-3}). **g** Violin plots showing differential expression levels of indicated genes in each DA^{Early} subclusters. **h** Heatmap showing differentially expressed genes and manually selected markers in 3 DA^{Late} neuron subclusters (DA^{L-1}, DA^{L-2}, and DA^{L-3}). **i** Schematic overview of experimental design where scRNAseq data from dissected human fetal VM (6–11-week embryos) and 3D primary cultures thereof (1 month) Birtele et al., bioRxiv doi.org/10.1101/2020.10.01.322495. **j** Overlapping and individual UMAP plots showing DA subcluster-integration analysis from scRNA-seq dataset of human fetal VM and hPSC-derived VM organoid. **k** Relative overlapping quantification of DA organoid subtypes vs human fetal DA neuron dataset after prediction of DA neuronal subtypes using fetal data as reference (Seurat).

differentiation in 2D cultures^{14,28}. With time, the cells gradually occupied the surface and inner space of the scaffold (Fig. 5e, f), and at day 10, the resulting 3D structures were mechanically detached from the bottom of the plate (Fig. 5g, h and Supplementary Fig. 7c, d). We named the resulting bioengineered organoids *silk-VM organoids* (Fig. 5i, j and Supplementary Fig. 7g). Unlike VM-patterned organoids grown without a scaffold, *silk-VM* organoids was less round and more variable in shape (Fig. 5g, h and Supplementary Fig. 7d), yet displayed a less distinct boundary between outer and inner regions (Fig. 5g, i and Supplementary Fig. 7e, f). Cell-viability assay indicated that the self-arrangement of cells along silk fibers to enhance organoid generation did not affect their viability (Supplementary Fig. 7c). Immunocytochemistry after one month confirmed a robust expression of the VM FP progenitor cell markers ZO-1, SOX2, FOXA2, and LMX1A (Fig. 5k–m and Supplementary Fig. 7h–l, n, o), indicating a similar developmental progression to that observed in organoids without scaffolding. The establishment of midbrain DA neuronal fate was confirmed by FOXA2 (Fig. 5n, o), TH, and MAP2 expression (Fig. 5p, q and Supplementary Fig. 7m, p–v). Quantifications at day 50 revealed a similar patterning and higher number of TH-expressing neurons in silk organoids (Fig. 5q and Supplementary Fig. 7t–v), followed by expression of GIRK2 and CALB1, markers of A9 and A10 DA neuron subtypes, and DAT at month 3 (Fig. 5r–u).

Silk scaffolding reduces interorganoid variability in cell-type composition and DA neuron formation. We used scRNAseq to compare 1-month-old VM organoids grown without a scaffold, here defined as “conventional” organoids (12,830 cells analyzed), silk-VM organoids (16,740 cells analyzed), and silk-VM organoids functionalized with Lam-111 (silk-lam VM organoids) (15,520 cells analyzed) from three independent biological replicates and separate 10X runs. UMAP embedding and graph-based clustering resulted in six major clusters (Fig. 6a). To annotate the clusters, we exploited the cell types identified in conventionally generated organoids and projected these labels onto the new data using Seurat’s v3 label transfer⁴⁵. Frequency analysis quantifying the number of cells in each cluster showed a lower variability in both silk and silk-lam VM organoids than in organoids grown without a scaffold (Fig. 6b). UMAP plots (Fig. 6a) and chord diagram (Fig. 6c) visualizing cell-type interrelationships, revealed a larger DA cluster when the silk fibers were functionalized with Lam-111 (Fig. 6d, e and Supplementary Fig. 8a).

We next examined gene expression profiles of DA neurons and their progenitors in multiple independent batches of organoids generated with and without silk at month 2. Early and late DA markers were highly varied in inter- and intrabatches of conventionally generated VM brain organoids, as shown by RT-PCR, whereas self-arrangement into silk scaffolds alone or

functionalized with Lam-111 significantly limited batch variability (Fig. 6f). To more precisely quantify the reproducibility of the silk methodology at protein level, we used a CRISPR/Cas9-mediated gene-edited transgenic hPSC line where CRE is knocked into the first exon of the *TH* gene⁴⁶. When transduced with a flexed GFP lentiviral vector, this line serves as live reporter cell line where GFP is expressed specifically in DA neurons⁴⁶. Also using this reporter system, the TH neurons appeared much more heterogeneously distributed in conventional vs silk organoids (Fig. 6g, h). Flow-based quantification in individual organoids established from this *TH*-Cre knock-in line revealed that *silk-lam* VM organoids displayed a higher percentage and more homogeneous formation of DA neurons (Fig. 6i and Supplementary Fig. 8b).

We further used scRNAseq to analyze three independent silk-lam VM organoid batches after four months. UMAP analysis of 18,375 cells visualized the same eight distinct major clusters (FP0–3, DA neurons, astrocytes, oligodendrocytes, and VLMCs) found in conventional organoids (Fig. 6j). However, frequency analysis quantifying the number of cells in each cluster revealed a lower variability in silk-lam VM organoids than in organoids grown without a scaffold also at this time point (Fig. 6k, l). To test this statistically, we utilized intraclass correlation (ICC), a correlation metric testing the proportions of each cell type in each 10X run where an ICC near 1.0 indicates high agreement. The six batches produced using the standard protocol had an ICC of 0.51 (95% CI: 0.214–0.837) compared with 0.98 (0.96–0.99) for silk-lam VM organoids. Importantly, a high proportion of DA neurons was maintained long term in *silk-lam* organoids, indicating a precise and reproducible patterning (Fig. 6l). The FP2 population was also detected at this time point, suggesting that its absence at the early stage (Fig. 6a) reflects a slightly different developmental timing in silk organoids (Supplementary Fig. 8c). In addition, when analyzing the same number of cells and time points in both culture systems, DA neurons in silk-lam VM organoids showed a higher expression of TH and postmitotic DA neuron markers (Fig. 6m), indicating that a greater degree of maturation had been reached.

Silk scaffolding reduces intra-organoid variability. Immunolabeling-enabled imaging of solvent-cleared organs (iDISCO) (Fig. 7a, b) reconstructed a more complex and highly intricate DA region throughout the entire silk-engineered organoids (Supplementary Video 1), showing more extensive DA circuitry and more efficient generation of VM regions than in conventional 3D cultures (Fig. 7c–e). Quantification of TH, GIRK2, CALB, and DDC in the core vs edge of organoids confirmed that the distribution of DA cells was much more homogeneous in *silk-lam* VM organoids than in our conventional VM organoids (Fig. 7f), as well as in previously described midbrain-patterned organoids (Supplementary Fig. 9a–d)^{15,17}.

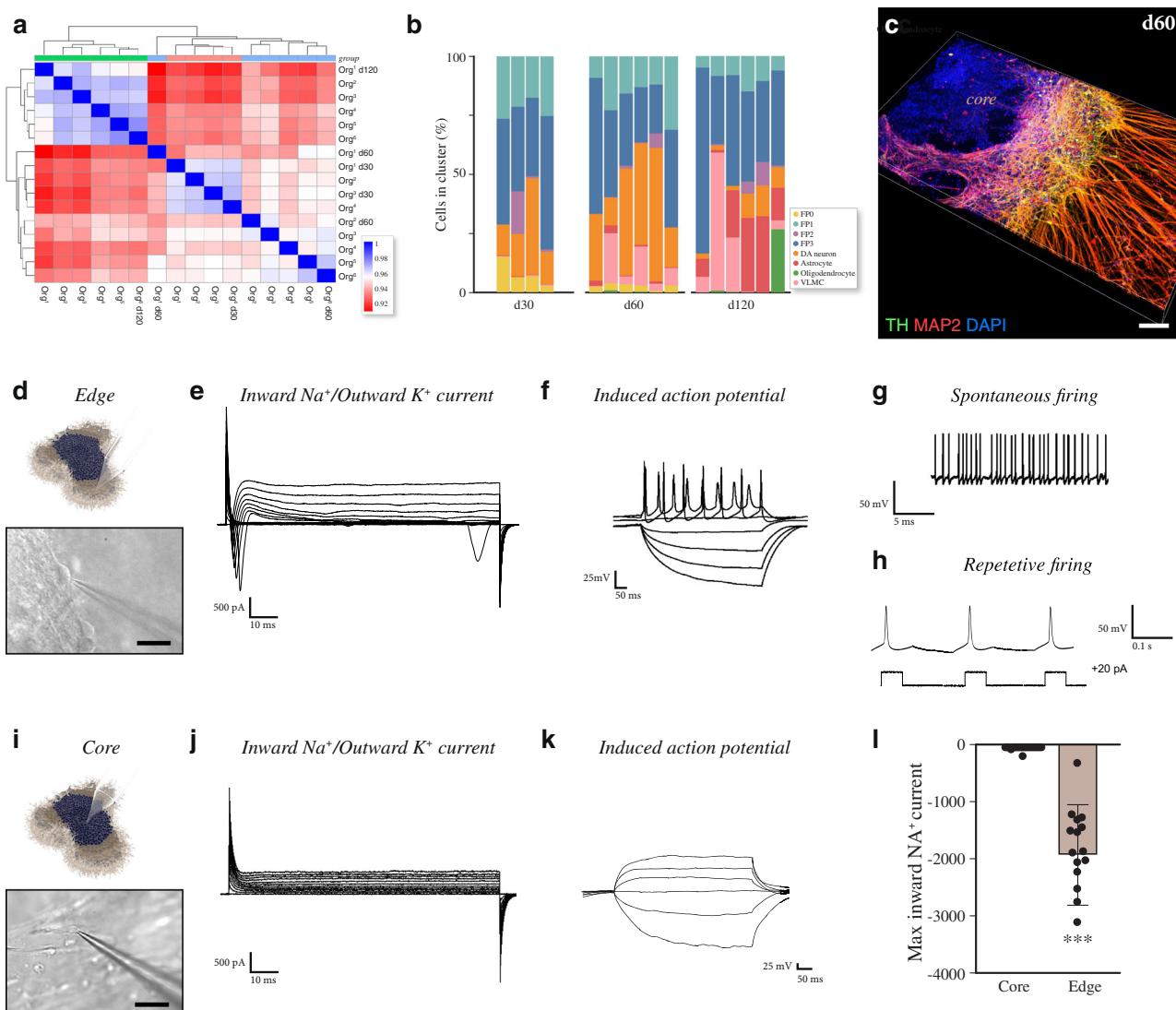
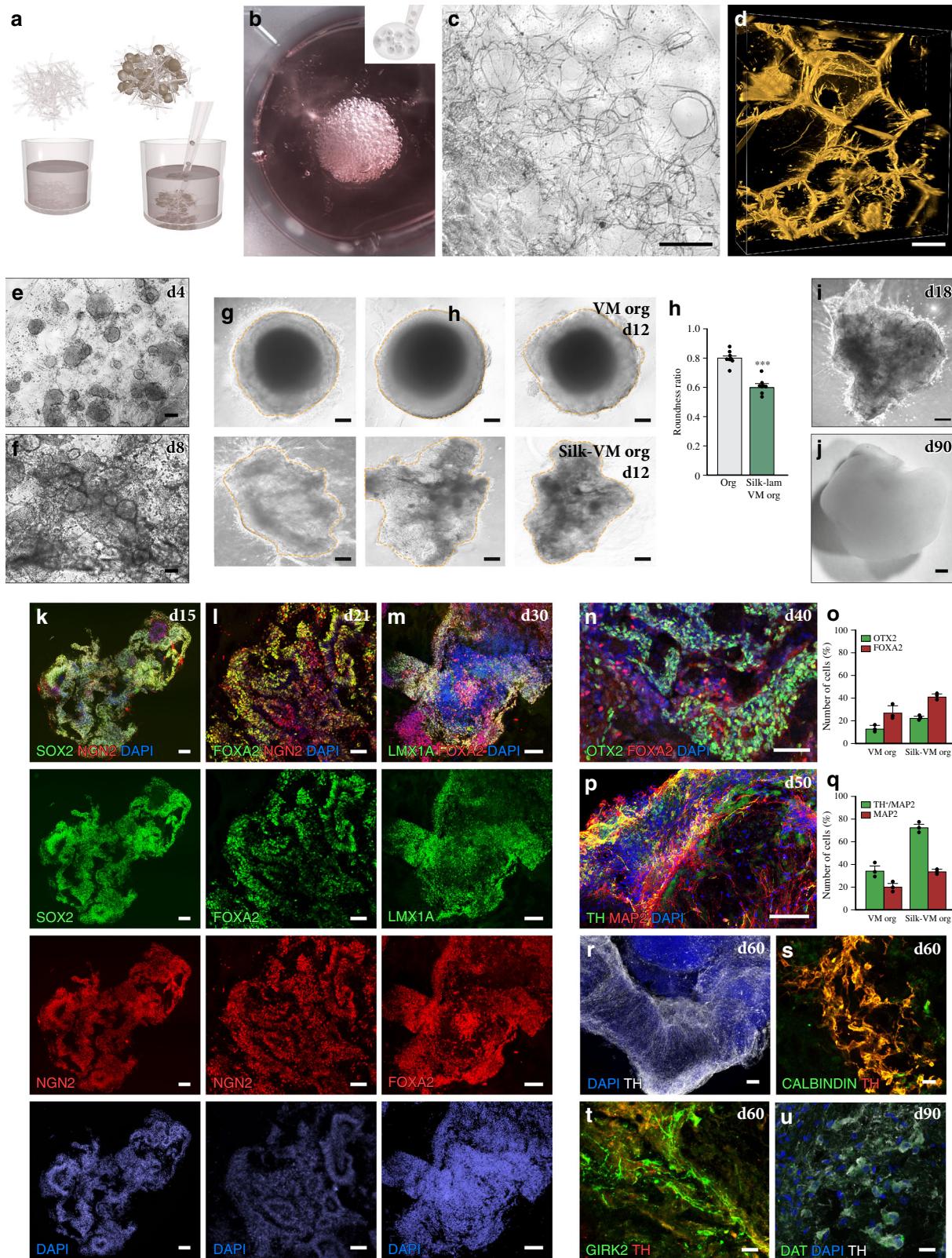


Fig. 4 Molecular and functional heterogeneity in human VM organoids. **a** Clustering of sample to sample correlations (Pearson) of organoids and different timepoints using Euclidean distance on normalized and log-transformed read counts. **b** Percentage of cells belonging to each cell cluster from individual organoids at days 30, 60, and 120. Intraclass correlation coefficient (correlation metric that considers group structure in the data) decreased from 0.717 at day 30 to 0.682 at day 60 and to 0.548 at day 120. **c** 3D reconstruction of an image stack from an 80 μm -thick optical section of TH and MAP2 immunohistochemistry at day 60. Scale bars, 100 μm . **d** Representative image of external functional recordings using whole-cell patch-clamp technique. Scale bars, 100 μm . **e** Representative trace from external patching showing inward sodium- and outward potassium-rectifying current traces of VM organoid at day 90 triggered by stepwise depolarization. **f** Patch-clamp recordings of external VM organoid cells depicting current-induced action potentials (APs) at day 90 (-85 pA to $+165 \text{ pA}$ with 20 pA steps). **g** External spontaneous firing at resting-membrane potential indicative of mature DA neuronal physiology in VM organoids at day 90. **h** Example trace of rebound depolarization after brief membrane depolarization (20 pA) indicative of DA phenotype in externally located cells. **i** Representative image of internal functional recordings using whole-cell patch-clamp technique. Scale bars, 100 μm . **j** Representative inward sodium- and outward potassium-rectifying current traces of internally located cells at day 90 triggered by stepwise depolarization. **k** Patch-clamp recordings of internal VM organoid cells depicting the absence of current-induced APs at day 90 (-85 pA to $+165 \text{ pA}$ with 20 pA steps). **l** Inward sodium current quantifications in externally ($n=20$) and internally ($n=16$) localized cells within VM organoids at day 90. Data represent mean \pm SD, unpaired two-tailed t -test, $p=0.0007$. Source data are provided as a Source Data file.

Studies on current organoid methods report that the increasing size of organoids can limit access to oxygen and nutrients in the inner layers, thereby affecting cell function and lifespan^{12,47,48}. We found that while conventional and silk organoids had a similar volume (Fig. 7c), the volume of the immature core was smaller in silk organoids (Fig. 7d, e). Even when larger silk organoids were generated by proportionally increasing the number of cells and silk fibers, the core volume in each organoid remained small independent of the size of the individual organoid (Fig. 7c–e). By analyzing 2D sequential imaging of DAPI-stained sections, we observed the presence of

porous microarchitectures in silk-engineered 3D structures (Fig. 7g) with an average cavity size of $3957 \pm 817 \mu\text{m}^2$, likely to promote an increase in oxygen and nutrients in the inner regions and thus reducing cell death. To test this hypothesis, we performed whole-organoid Western blot analysis of hypoxia-inducible factor-1 alpha (HIF-1 α), a key oxygen-labile protein. We found that silk scaffolding led to attenuation of hypoxic response pathway (Fig. 7h), which also persisted when silk-lam VM organoids were cultured for the first few hours under low oxygen tension (1%) in a gas-controlled chamber (Fig. 7i). Analysis based on scRNASeq showed that the global level of



stress-response signaling in DA neurons is lower in silk than in conventionally generated VM organoids (Fig. 7j), as is the expression of individual genes associated with metabolic dysfunctions including glycolysis, oxidative stress and DNA damage (Fig. 7k). In agreement with this finding, decreased interior cell death was also observed in *silk-lam* VM organoids (Fig. 7l–q).

Furthermore, the increased homogeneity, decreased cell death, and increased oxygen diffusion within the silk organoids resulted in mature and functional DA neurons in all regions of the organoid as assessed using whole-cell patch-clamp recordings from the outer and inner regions of the *silk-VM* organoids. This analysis revealed that in contrast to conventionally generated VM organoids, cells in the inner core of silk 3D culture also exhibited

Fig. 5 Generation and characterization of silk-VM organoids. **a** Schematic representation of silk-VM organoid generation. **b** Representative image of cells dispersed throughout silk foam. **c** Bright-field and **d** confocal images of 3D silk scaffold after reabsorption of foam. Scale bars, 50 μ m. **e**, **f** Bright-field images showing adherence and growth of cells along the length of silk microfibers at days 4 and 8. Scale bars, 200 μ m. **g** Representative bright-field images, and **h**, roundness measurement of VM organoids grown with and without scaffold at day 12. Scale bars, 100 μ m. Data represent mean \pm SEM of 8 biologically independent organoids, two-tailed Mann-Whitney test, *** p < 0.0001. **i** Representative bright-field images of a short-term and, **j**, long-term silk-VM organoid culture. Scale bar, 200 μ m. **k** Immunohistochemistry of SOX2/NGN2 from organoid at day 15 and **l–n** FOXA2 across a time course from day 21 to 40, stained with NGN2 (**l**), LMX1A (**m**), and OTX2 (**n**). Scale bars, 100 μ m (**l**, **n**) and 50 μ m (**k**, **m**). **o** Quantifications of OTX2⁺ and FOXA2⁺ cells in VM and silk-VM organoids. Data represent mean \pm SEM obtained from 3 independent organoids. **p** Immunohistochemistry of TH and MAP2 and **q**, quantifications of MAP2 and TH/MAP2 in VM and silk-VM organoids at day 50. Data represent mean \pm SEM obtained from 3 independent organoids per condition. Scale bars, 100 μ m. **r** Immunohistochemistry of TH and **s**, **t** with CALB1/GIRK2 at day 60. Scale bars, 20 μ m. **u** Immunohistochemistry of TH stained with DAT at day 90. Scale bars, 20 μ m. Nuclei were stained with DAPI in **j–l**, **m**, **o**, **q** and **t**. Source data are provided as a Source Data file.

inward Na⁺ and outward K⁺ currents, confirming expression of voltage-gated sodium and potassium channels and a mature neuronal phenotype ($n = 20$ cells, inner; $n = 19$ cells, outer) (Fig. 8a–l and Supplementary Fig. 9e, f). Moreover, cells in both the core and outer layers of silk-lam VM organoids revealed mature electrophysiological properties of DA neurons with the presence of induced APs as well as spontaneous firing and rebound depolarization (Fig. 8b–e, h–k and Supplementary Fig. 9e, f). In addition, calcium imaging of MAP2-GCaMP5-labeled neurons indicated active neuronal signaling, confirming that mature and functional DA networks were present in silk-VM organoids (Fig. 8m, n). Finally, we performed real-time chronoamperometric measurements of DA exocytosis using a carbon-coated fiber (200 μ m diameter) as a working electrode in a three-electrode electrochemical setup (Fig. 8o, p)⁴⁹. Although DA release confirmed the high maturation and functionality of DA neurons in conventionally generated organoids, a lower proportion of the recordings showed a release of DA than in silk-lam VM organoids (Fig. 8q). Thus, although the quality of individual DA neurons generated in 3D organoids is comparable in conventional and silk organoids, the silk-based tissue engineering technology is more robust and results in better DA patterning with less variation within and between organoids.

Discussion

We used VM organoids derived from pluripotent stem cells to perform a time course transcriptional analysis of human VM development and human DA neuron differentiation at single-cell level. This allowed us to (i) determine cellular composition, (ii) deduce the timing of cell-type appearance during organoid maturation, and (iii) trace developmental trajectories. We also more closely dissected the DA neuron cluster (14,606 cells), which led to the identification of three molecularly distinct subgroups of mature human DA neurons.

We used ventralizing and caudalizing factors known to direct a VM identity in 2D culture^{28,29} and to result in the formation of functional DA neurons after transplantation in preclinical rat models of PD^{50,51}. A subset of these factors was previously used to generate similar midbrain-patterned organoids derived from hPSCs^{15–17} and other studies report VM organoids derived from neural progenitors^{19–21}. Together, these reports show that it is possible to form and maintain DA neurons with mature functional properties for extended periods in 3D culture, and midbrain-patterned organoids have now started to be used for developmental studies, disease modeling, and high-throughput screening^{18,20}.

In this study, we performed single cell transcriptional profiling of over 120,000 cells at different time points and produced a comprehensive map of cell type composition in VM organoids. This type of large-scale transcriptional mapping at single cell level has previously been performed in cerebral organoids, identifying a large diversity of neuronal cell types expressing both markers of

multiple brain regions including forebrain, midbrain, hindbrain, and retina, as well as long-term excitatory/inhibitory neuronal identities^{9,52}. Our analysis revealed that VM organoids have a more restricted repertoire of cell types, and that the composition is analogous to that observed in DAergic transplants from similarly patterned hPSCs, as previously described⁵³. In addition, comparison of our single cell dataset with findings reported in adult human midbrain shows a high similarity between all mature cell types (neurons, astrocytes, and OPCs) within VM organoids and those present in human SNc, confirming the successful regionalization in vitro²² and supporting the use of midbrain organoids in developmental studies and in the therapeutic development.

The clear temporal appearance of different cell types, with FP-like cells being the dominant cell type at early time points followed by generation of DA neurons and subsequently development of astrocytes, OPCs, and VLMCs, enabled us to perform a detailed analysis of cell lineage using force-directed k -nearest-neighbor graph-based pseudotime trajectory to predict the future cellular states in each cluster. The readout of this analysis points to the fact that cells with an FP-like transcriptional profile serve as a common progenitor for all mature cell types arising in VM organoids. Further, the expression in both DA neuron and VLMC clusters of NR4A2 and SOX6, known to exert a critical role in the specification and maturation of DA phenotype, suggests that lineage segregation is a relatively late event during DA neurogenesis.

DA neurons in adult VM are traditionally divided into two main subtypes based on anatomical landmarks: A9 neurons located in SNc with projections into the dorsolateral striatum, and A10 neurons located in the ventral tegmental area (VTA) with more widespread projections including nucleus accumbens, septum, and prefrontal cortex^{54,55}. Recent investigations at single cell level in mouse VM revealed a greater-than-expected molecular diversity in midbrain DA neurons^{23,24,35}, but similar studies of human DA neurons are lacking. Previously, such a detailed analysis has not been possible using existing datasets from human midbrain^{22,35} due to the low number of mature DA neurons captured. To date, no scRNAseq data set from midbrain organoids obtained from pluripotent stem cells exists, and only one scRNA-seq analysis into neural progenitor-derived VM organoids is available in the literature²¹. This dataset²¹ describes a similar cell type composition to the one we identified in hPSC-derived VM organoids but did not analyze enough DA neurons or perform the late-stage profiling needed to define molecularly distinct DA neuron subtypes. The emergence of functionally mature DA neurons in high numbers within our VM patterned organoids derived using a protocol known to result in authentic and functional DA neurons capable of mediating functional recovery and circuitry reconstruction after transplantation in rat models of PD^{28,50,51}, thus uniquely enabled a molecular analysis at single cell level in order to map human DA neuron diversity. By

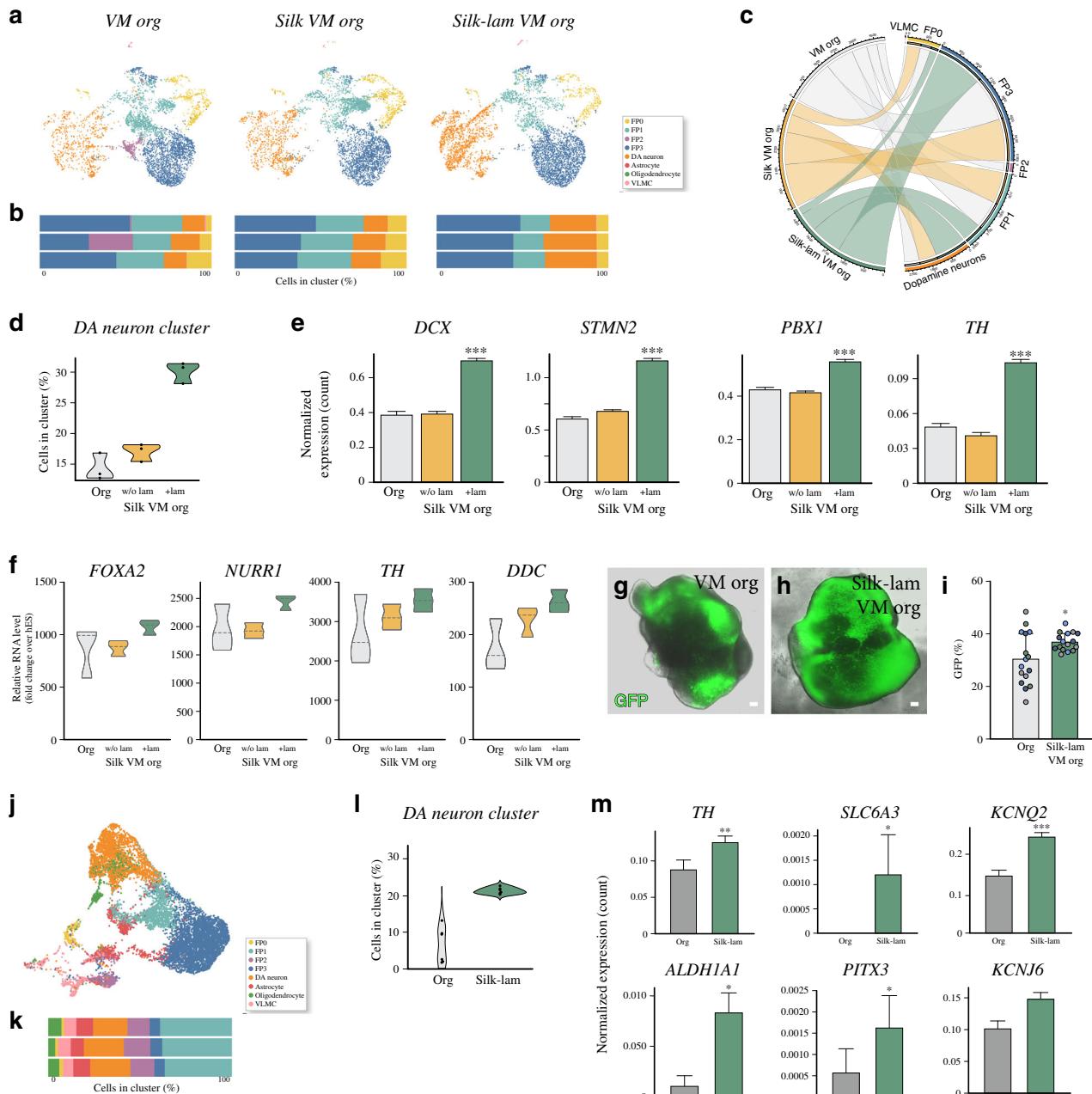
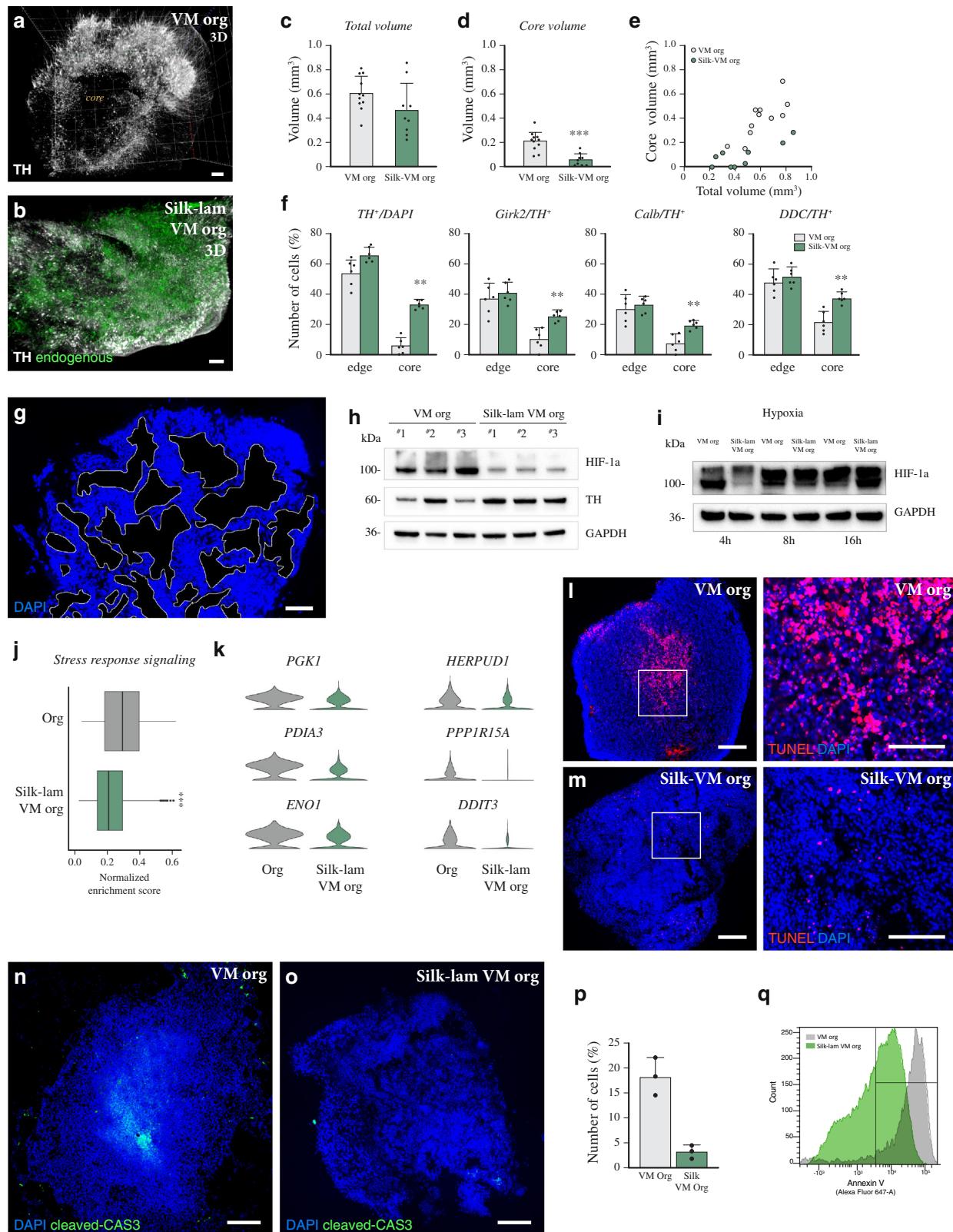


Fig. 6 Single-cell transcriptomics identifying silk-VM organoid cell composition. **a** UMAP plots showing cell clusters from conventional VM, silk-VM, and silk-lam VM organoids and **b**, percentage of cells belonging to each cell cluster from individual organoids at month 1. **c** Chord diagram visualizing cell-type interrelationships between conventional, silk-VM and silk-lam VM organoids. **d** Violin plot showing the percentage of cells belonging to DA neuron clusters from conventional, silk-VM, and silk-lam VM organoids at month 1 from three individual organoids per condition. **e** Expression of selected markers belonging to DA neuron cluster in conventional, silk-VM, and silk-lam VM organoids at 1 month. Data represent mean \pm SEM of 3 biologically independent organoids, two-tailed Wilcoxon Rank Sum test, *** p < 0.0001. **f** qRT-PCR analysis of early and late DA neuron markers in conventional and silk/silk-lam VM individual organoids at month 2. Data represent mean \pm SEM of 3 independent organoids per condition. **g, h** Representative images of GFP expression in conventional and silk-lam VM organoids differentiated from the CRISPR/Cas9-mediated gene-edited TH-Cre hPSC line. Scale bars, 100 μ m. **i** FACS-based quantification of GFP expression in conventional and silk-lam VM organoids differentiated from a CRISPR/Cas9-mediated gene-edited TH-Cre hPSC line in 4 biologically independent experiments shown as color-coded dots (green, light blue, blue and purple). Data represent mean \pm SD, two-tailed unpaired t test p = 0.0162. **j** UMAP plot showing cell clusters from silk-lam VM organoids and **k**, percentage of cells belonging to each cell cluster from individual organoids at month 4. **l** Violin plot of percentage of cells belonging to DA neuron cluster from conventional and silk-lam VM organoids at month 4. Data represent mean \pm SEM of 3 biologically independent organoids, two-tailed Wilcoxon Rank Sum test, TH p = 0.0045, SLC6A3 p = 0.010462, KCNQ2 p = 0.0007, ALDH1A1 p = 0.046. Source data are provided as a Source Data file.



performing an unbiased and comprehensive characterization of the mature DA cluster we identified three molecularly distinct subtypes (DA^{L-1}, DA^{L-2}, and DA^{L-3}). These separate DA clusters showed several points of similarity with previously published datasets from bulk and single cell mouse VM DA neuron populations³⁶. We also identified a set of genes in VM organoids which may define similar DA neuron subtypes in both mouse and

human. Our data suggest that DA^{L-1}, which mainly expresses *SLC6A3*, *KCNJ6*, and *SNCG*, corresponds to two DA groups (T-Dat^{high} and AT-Dat^{high}) reported to express high levels of DAT by Tiklova et al. and to another DA cluster (DA^{L-1}) identified in adult mouse brain by Poulin and colleagues. *ANXA1*, which was found enriched together with *SLC6A3* and *MCPI* in our DA^{L-2} cluster, was used as a marker to distinguish SNc from VTA and

Fig. 7 Silk fibers result in more homogeneous VM organoids. **a** iDISCO circuitry reconstruction obtained by mapping TH in conventional and **b** silk-lam VM organoids at day 60. Scale bar, 100 μ m. **c** iDISCO-based total volume quantification and **d**, **e** core quantification of conventionally and silk-lam-generated VM organoids. Data represent mean \pm SEM obtained from 11 and 9 independent conventionally and silk-lam-generated VM organoids respectively, two-tailed Mann-Whitney test, $p = 0.0002$. **f** Percentages of TH⁺, GIRK2⁺, CALB⁺, and DDC⁺ expressing cells located in the outer and inner layers in conventionally and silk-lam-generated VM organoids. Data represent mean \pm SEM obtained from 6 biologically independent organoids per condition, two-tailed Mann-Whitney test, $p = 0.002$. **g** Immunohistochemistry showing microporous dimension. Scale bars 100 μ m. **h** Representative Western blots of HIF-1 α protein and TH expression in conventional and silk-lam VM organoids in normoxia conditions (21% O₂). GAPDH was used as loading control. **i** Representative Western blots of HIF-1 α protein in conventional and silk-lam VM organoids across a time course of 4 h, 8 h and 16 h under hypoxia conditions (<1% O₂). GAPDH was used as loading control. **j** Gene Set Enrichment Analysis of Stress response signaling. Lower and upper hinges correspond to the first and third quartiles and the whisker extends from the hinge to the largest value no further than $|1.5 * IQR|$ from the hinge (where IQR is the interquartile range, or distance between the first and third quartiles); two-tailed Wilcoxon Rank Sum test, *** $p < 0.0001$. **k** Representative markers of metabolic stress of DA neurons in VM organoids grown with and without scaffold at 4 months. **l**, **m** TUNEL staining of VM organoids grown with and without scaffold at 6 months. Scale bars, 100 μ m. **n**, **o** Immunohistochemistry of cleaved caspase-3 and, **p** quantification of cleaved CAS3 over DAPI performed on conventional and silk-lam VM organoids at 6 months. Scale bars 100 μ m. Data represent mean \pm SEM of 6 biologically independent VM organoids per condition, two-tailed unpaired t-test, $p = 0.0028$. **q** FACS analysis for fluorescence intensity of Annexin-V staining in conventionally and silk-lam-generated VM organoids. Representative FACS plots of biological triplicates are shown. Nuclei were stained with DAPI in **g**, **j**. Source data are provided as a Source Data file.

to identify different SNC subpopulations during mouse VM development^{25,37,38}. D^{L-3}, characterized by CALB, OTX2, LMO3, and SOX6 expression, seems to resemble other clusters, DA^{2B} and mDA2, previously described in two mouse studies by Poulin et al. and La Manno et al., respectively.

Our datasets provide insights into early fundamental regulators involved in molecular mechanisms that may play an important role in driving segregation of different mature DA neuron subtypes. A greater understanding of how subidentity is established could lead to the design of more targeted and effective DA neuron differentiation strategies, with implications for stem-cell-based therapies and disease modeling in PD. However, current organoid studies are often hampered by problems of reproducibility within and between organoids, as well as incomplete maturation resulting from interior hypoxia and the emergence of an immature or necrotic inner core⁴¹. Such issues were also observed in our study and in previous reports of VM organoids^{15,19}. Inter-organoid variability can be partially reduced by more precise patterning²⁰, careful titration of the initial cell number⁷, starting with neural progenitor cells rather than pluripotent cells¹⁹, the adoption of scalable and automated culture systems^{20,16}, or the use of scaffolds to guide self-organization⁴¹.

Bioengineering efforts to generate more homogeneous organoids with viable and mature cells in all regions are actively being pursued. Recent findings describe two very different strategies, one based on the creation of a 2D scaffold using individual inert microfilaments to guide the self-organization of hPSCs into organoids with more reproducible neuroectoderm features⁴¹ and the other using biomaterials such as hydrogels, which support bioengineered 3D neural cultures by mimicking the native brain extracellular matrix^{56,57}. Exogenous vascularization in human organoids can also be achieved either by transplanting human organoids into a physiological environment in mouse brain⁵⁸ and ectopically expressing genes to induce a vascular-like structure⁵⁹, or by coculturing with endothelial cells⁶⁰. These approaches at least partly address issues of reproducibility and maturation, but with limited experimental control. Moreover, the *in vivo* model of vascularized human brain organoids limits their use in large-scale biomedical applications such as drug screening, and the coculture or genetic induction of vessels may affect the directed differentiation of hPSCs necessary to obtain organoids of a specific brain region.

In this study, we used recombinant spider-silk microfibers that provide an easily accessible *in vitro* platform to generate bioengineered VM organoids, which we termed *silk-VM* organoids. These spider-silk microfibers present several advantages over

existing methods using inert fibers. *First*, they are a recombinant chemically defined biomaterial that creates a strong, elastic, and biocompatible 3D scaffold able to guide the self-assembly of hPSCs into complex tissue-specific structures. *Second*, they can be easily functionalized with bioactive molecules that favor cell adhesion or control patterning. *Third*, they have the dual capacity to serve as both an anchored scaffold during early and a floating scaffold during late stages of organoid differentiation, allowing precise control of the ratio between cell density and number of fibers, thus decreasing variation between organoids. *Fourth*, silk scaffolds form porous microarchitectures, creating more favorable growth and differentiation conditions by allowing for the diffusion of oxygen and extrinsic patterning factors into the core, thereby reducing necrosis, reducing metabolic stress, and supporting neuronal maturation in all regions of the organoid.

Silk-VM organoids reproduce key molecular aspects of DA neurogenesis with a similar developmental progression pattern as conventionally generated organoids. The DAergic patterning in *silk-VM* organoids was further enhanced when functionalized with Lam-111. In addition, silk fibers sustain the homogeneous and functional generation of DA neurons throughout all compartments of the organoid in a highly efficient manner. Reduction of the necrotic core was previously achieved by mechanical cutting and generation of sliced human cortical organoid cultures to prevent hypoxia and cell death in the core¹², or by engineering smaller organoids^{16,20,59}. Unlike mechanically sliced organoids¹², *silk-VM* organoids are preserved in size and shape, thereby more likely to maintain functional properties as well as the intricacy of neuronal networks. The remarkable properties of silk scaffolds combined with their straightforward use requiring no specialized equipment thus provide an easily accessible *in vitro* methodological platform able to generate organoids in a reproducible and functional manner.

Methods

hPSC culture and 2D differentiation. Undifferentiated RC17 (Roslin Cells, cat. no. hPSCreg RCe021-A), H9 (WiCell, cat. no. hPSCreg WAe009-A), HS999 and HS1001 (kindly provided by the Karolinska Institute), and TH-Cre hPSC cells were maintained on 0.5 μ g/cm² Lam-521 (BioLamina, #LN-521)-coated plates in iPS Brew medium (Miltenyi, #130-104-368) and were differentiated into 2D VM-patterned progenitors using our GMP-grade protocol¹⁴. All procedures were in accordance with the European Union directive and approved by the local ethical committee at Lund University.

Human brain organoid culture. To start 3D VM organoid differentiation, RC17, H9, HS999, HS1001, and TH-Cre hPSC cells were detached from the culture dish with 0.5 mM Accutase (Thermo Fisher Scientific, #A1110501) to yield a single cell

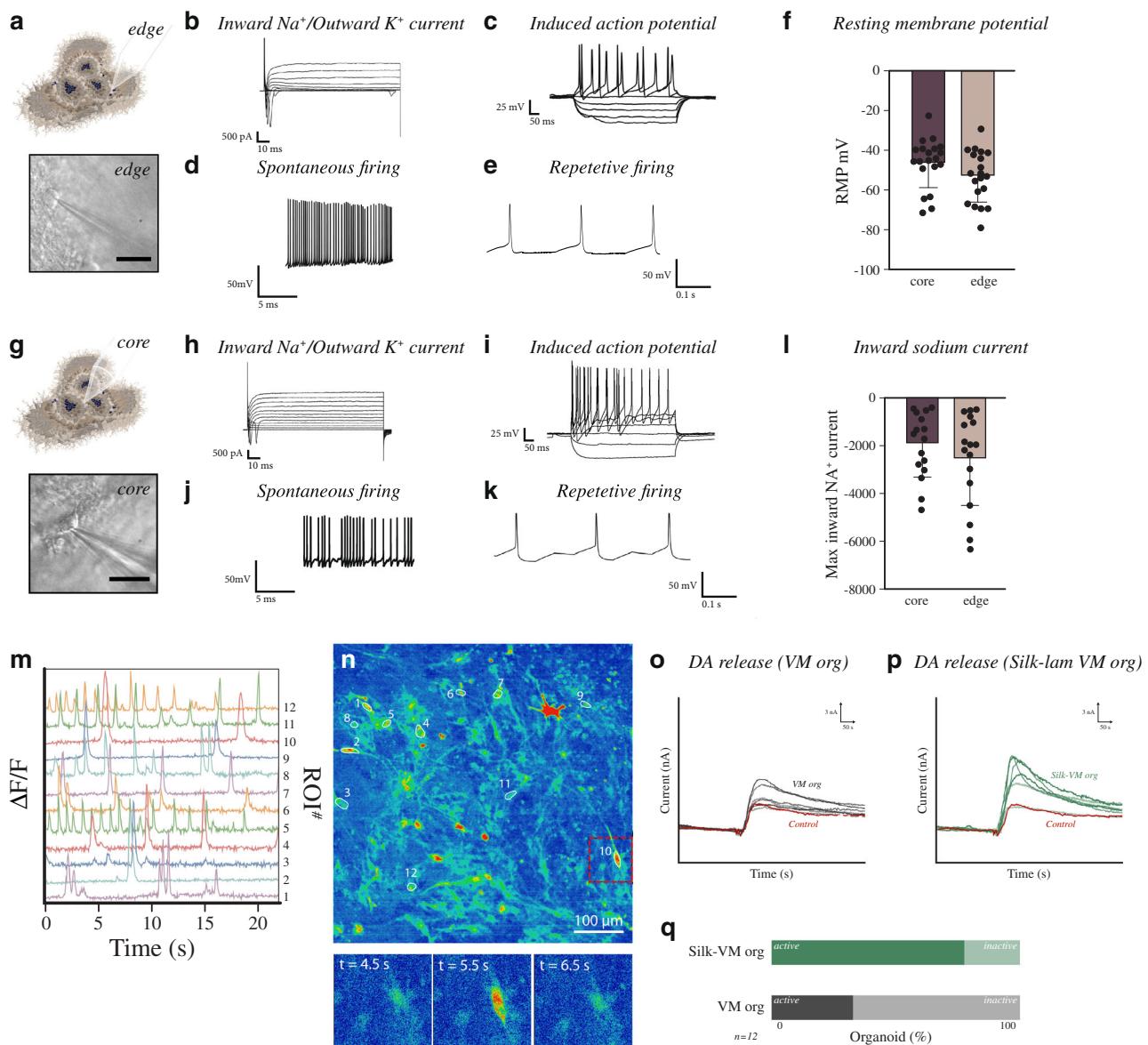


Fig. 8 Silk-VM organoids are functionally homogeneous. **a** Representative images of functional recordings from the external part using whole-cell patch-clamp technique. Scale bars, 100 μ m. **b** Representative inward sodium- and outward potassium-rectifying current trace of external VM organoid at day 90 triggered by stepwise depolarization. **c** Whole-cell patch-clamp recordings of external VM organoid cells depicting current-induced APs at day 90 (-85 pA to $+165$ pA with 20 pA steps). **d** Spontaneous firings at resting membrane potential indicative of mature DA neuronal physiology in silk-lam VM organoids in the external part at day 90. **e** Example trace of rebound depolarization after brief membrane depolarization (20 pA) indicative of DA phenotype in externally located cells. **f** Resting-membrane quantifications between externally ($n=20$) and internally localized cells ($n=20$) in VM organoids at day 90. Data represent mean \pm SD. **g** Representative images of functional recordings from the internal region of organoid using whole-cell patch-clamp technique. Scale bars, 100 μ m. **h** Representative internal inward sodium- and outward potassium-rectifying current trace of VM organoid at day 90 triggered by stepwise depolarization. **i**, **j** Whole-cell patch-clamp recordings in internal region of VM organoid cells depicting current-induced APs at day 90 (-85 pA to $+165$ pA with 20 pA steps). **k** Spontaneous firings at resting-membrane potential indicative of mature DA neuronal physiology in the internal region of silk-lam VM organoids at day 90. **l** Example trace of rebound depolarization after brief membrane depolarization (20 pA) indicative of DA phenotype in internally located cells. **m** Quantification of maximum inward sodium current recorded in internal ($n=16$ cells) and external ($n=17$ cells) regions. Data represent mean \pm SD. **n** Differential fluorescence-intensity profile of intracellular Ca^+ levels as a function of time in neurons expressing MAP2-GCaP5 at day 90. **o**, **p** Fluorescence image with marked regions of interest corresponding to recorded cells and three timeframes displaying the change in intracellular fluorescence intensity. Scale bar, 100 μ m. **q** Representative analysis of real-time DA release chronoamperometric measurements in conventional and silk-lam VM organoids and **q**, relative quantification ($n=12$).

suspension. Differentiation was initiated by plating 2500 single cells in each well of a 96-well U-bottom plate (Corning, #CLS7007) in iPS Brew with 10 μ M Y-27632 dihydrochloride (Miltenyi, #130-106-538), as previously described²⁷. After three days in culture, embryoid bodies were transferred to differentiation medium consisting of 1:1 DMEM/F12:Neurobasal medium (Thermo Fisher Scientific, #21331020 and #A1371201), 1:100 N2 supplement (Thermo Fisher Scientific, #A1370701), 10 μ M SB431542 (Miltenyi, #130-106-543), 150 ng/ml rhNoggin

(Miltenyi, #130-103-456), 400 ng/ml SHH-C24II (Miltenyi, #130-095-727), and 1.5 μ M CHIR99021 (Miltenyi, #130-106-539), and in the presence of 200 mM L-glutamine (Thermo Fisher Scientific, #25030081) and 10,000 U/mL penicillin-streptomycin (Thermo Fisher Scientific, #15140122). About 1% minimum essential medium nonessential amino acids (MEM-NEAA, Sigma-Aldrich, #M7145) and 0.1% 2-mercaptoethanol (Merck, #8057400005) were maintained for the entire differentiation period. On day 11, developing VM organoids were transferred to a

24-well plate containing 1:50 Neurobasal medium, B27 supplement without vitamin A (Thermo Fisher Scientific, #12587010), and 100 ng/mL FGF-8b (Miltenyi, #130-095-740). On day 14, 20 ng/mL BDNF (Miltenyi, #130-096-286) and 200 μ M L-Ascorbic acid (Sigma-Aldrich, #A4403-100MG) were added. At this point, VM organoids were embedded in 30 μ L droplets of Matrigel (BD Biosciences), as previously described²⁷. From day 16 onward, 0.5 mM db-cAMP (Sigma-Aldrich, #D0627-1G) and 1 μ M DAPT (R&D Systems, #2634) were added to the culture medium for terminal maturation for up to four months.

We define a batch of organoids as all organoids established at the same time and in the same dish.

Two additional midbrain-patterned organoids were also generated according to previously published protocols. Midbrain-like organoids (MLOs) were grown in tissue growth induction medium containing midbrain-patterning factors, as described in Jo et al.¹⁵. Midbrain organoids (MOs) were cultured with the addition of 2 μ M dorsomorphin (Sigma), 2 μ M A83-01 (Peprotech), 3 μ M CHIR99021 (Miltenyi), and 1 μ M IWP2 (Biogems), as reported by Kwak et al.¹⁷. Forebrain organoids were also generated using a whole organoid differentiation protocol²⁷.

Human embryonic tissue source. Human fetal tissue from legally terminated embryos was collected in accordance with existing guidelines with approval of the Swedish National Board of Health and Welfare and informed consent from women seeking elective abortions. To determine the gestational age of the embryos, the crown-to-rump length was measured and the embryo was staged according to week post conception.

Hypoxic conditions. Both conventional and *silk*-VM organoids were cultured at 37 °C in 21% O₂ and 5% CO₂ in normoxia conditions. For hypoxia analysis, VM organoids were transferred to the hypoxia chamber (BINDER CB160), which was filled with 1% O₂ and 5% CO₂ and mixed with N₂, and collected after 4, 8, and 16 h.

Cell-viability analysis. Cell viability was measured using the colorimetric CyQuant cell proliferation assay (Invitrogen), following the manufacturer's instructions. Absorbance was analyzed at 480–520 nm, using 50,000 cells harvested from whole conventional and *silk*-VM organoids. Each organoid was analyzed in 6 replicates (i.e., 300 000 cells from each organoid) using Biochrom Asys Expert 96 Microplate Reader (Biochrom). Apoptosis was detected via flow cytometry after staining using the Alexa Fluor 647 Annexin V conjugate (BD Pharmingen, #A23204) and Click-iT plus Tunel assay (Invitrogen #C10617) according to the manufacturer's instructions.

Silk-scaffold preparation and hPSC integration. A 20 μ L droplet of Biosilk protein solution (BioLamina, #BS-0101) either alone or functionalized with Lam-111 (BioLamina, #LN111), was placed in the center of hydrophobic culture wells in a 24-well plate (Sarstedt, #833922500). Recombinant human laminin-111 was purchased from Biolamina (Sweden) and added to the silk protein solution at a final concentration of 10 μ g ml⁻¹. The FN-silk/LN-111 mixture was incubated 10 min at RT before usage. About 10 μ M Y-27632 dihydrochloride was added⁴⁴. Air bubbles were introduced into the droplet by pipetting up and down (10–15 strokes), creating a dense foam. Multiple foams are generated in series with uniform shape and size (Supplementary Fig. 7b). Foam can also assume different dimension by increasing/reducing the volume of silk droplet giving rise to different organoid size in culture. RC17, H9, and HS1001 hPSCs were detached from the culture dish with 0.5 mM Accutase (Thermo Fisher Scientific, #A1110501) and prepared as a concentrated single cell suspension (20,000 cells/ μ L in iPS Brew medium). A total of 80,000 cells from the cell suspension were added to the silk foam and dispersed throughout by 6–8 additional pipette strokes⁶¹. The interaction between cells and silk microfibers was stabilized at 37 °C in an incubator for 20 min. Prewarmed iPS Brew medium containing 10 μ M Y-27632 dihydrochloride was added to the foam-integrated cells.

Silk-bioengineered VM organoid generation and morphological characterization. After three days, differentiation medium consisting of 1:1 DMEM/F12:Neurobasal medium, 1:100 N2 supplement, 10 μ M SB431542, 150 ng/mL rhNoggin, 400 ng/mL SHH-C24II, and 1.5 μ M CHIR99021 was added from day 0 to 10, following the same protocol used for generating conventional VM organoids. At day 10, the resulting 3D structures were mechanically detached from the bottom of the plate with a spatula and transferred to a 6-well plate (Corning, #3471) and grown in suspension (free floating). *Silk*-VM organoids were embedded in 30 μ L droplets of Matrigel and cultured following the VM organoid-differentiation protocol described in the subsection “hPSC VM organoid differentiation”. Images were collected on phase-contrast inverted microscope (Olympus, #CKX31SF). Morphological classification (spherical/nonspherical) was performed in triplicate. Roundness measurements were based on bright-field images and calculated as the ratio between diameters of the largest inscribed and the smallest circumscribed circle of the organoid silhouettes (dotted line). Images were analyzed in ImageJ (NIH).

Library preparation, sequencing, and raw data processing. For 10x Genomics single-cell RNA sequencing, single-cell suspensions were loaded onto 10x Genomics Single Cell 3' Chips along with the mastermix as per the manufacturer's protocol (<https://support.10xgenomics.com/single-cell-gene-expression/index/doc/technical-note-chromium-single-cell-3-v3-reagent-workflow-and-software-updates>) for the Chromium Single Cell 3' Library to generate single cell gel beads in emulsion (GEMs, version 3 chemistry). The resulting libraries were sequenced on either a NextSeq500 or a NovaSeq 6000 with the following specifications Read1 28 cycles, Read2 98 cycles, and Index1 8 cycles using a 200-cycle kit. Raw base calls were demultiplexed and converted fastq files using cellranger mkfastq program (bcl2fastq 2.19/cellranger 3.0). Sequencing data were first preprocessed through the Cell Ranger pipeline (10x Genomics, Cellranger count v2) with default parameters (expect-cells set to the number of cells added to 10x system), aligned to GrCh38 (v3.1.0), and the resulting matrix files were used for subsequent bioinformatic analysis.

Bioinformatics analysis of sequencing data. Seurat (v 3.1 and R version 3.6.1) was applied to the scRNA data for downstream analysis of matrix files. Cells with at least 600 but no more than 12000 genes detected were kept for analysis. In addition, cells with more than 20% mitochondrial reads were excluded. After log-transformation, 4000 highly variable genes were identified using vst and z-transformed expression values followed by dimensionality reduction (PCA). To integrate data from different 10x runs, Harmony was applied using the R-package “Harmony” using individual 10X runs as grouping variable. Harmony converged after 9 iterations and corrected coordinates were used for downstream analysis. To identify clusters, Louvain clustering (resolution 0.4, Seurat) was applied to harmony embeddings. Differential expression analysis between clusters was carried out using the Wilcoxon rank sum test (Seurat) with genes with a FDR-adjusted p-value < 0.05 considered significant. Gene ontology overrepresentation analysis was performed using the enrichGO function in the clusterProfiler package (3.13) using MSigDB as the database. Gene Set Enrichment Analysis was performed using the R package escape (<https://www.nature.com/articles/s42003-020-01625-6>). For silhouette and tree analysis the cluster package (version 2.1) was used. Lineage inference and pseudotime reconstruction was performed using Slingshot (version 1.6.1). We first used the expression data from 4000 top variable genes to generate the minimum spanning tree of cells in a reduced-dimensionality space (Harmony-corrected UMAP embeddings). Global lineage structure was identified with a cluster-based minimum spanning tree and fitting simultaneous principal curves describing each lineage using “slingshot” function. Pseudotime analysis was also performed using a force-directed layout of k-nearest-neighbor graphs (SPRING, <http://pubmed.ncbi.nlm.nih.gov/29228172/>) on normalized expression counts using default settings.

qRT-PCR. Total RNAs were isolated using the RNeasy Micro Kit (QIA-GEN#74004) and reverse transcribed using random hexamer primers and Maxima Reverse Transcriptase (Thermo Fisher #K1642, Invitrogen). cDNA was prepared together with SYBR Green Master mix (Roche#04887352001) using the Bravo instrument (Agilent) and analyzed by quantitative PCR on a LightCycler 480 device using a 2-step protocol with a 60 °C annealing/elongation step. All quantitative RT-PCR (qRT-PCR) samples were run in technical triplicates and the results are given as fold change over undifferentiated hPSCs using each of the two housekeeping genes for normalization (ACTB and GAPDH). Details and list of primers are reported in Supplementary Table 1.

Organoid cryosectioning and immunofluorescence. Both conventional and *silk*-VM organoids were fixed in 4% paraformaldehyde for 5 h at 4 °C followed by washing in PBS three times for 10 min. Both conventional and *silk*-VM organoids were left to sink in 30% sucrose overnight. Sucrose solution was replaced with 1:1 OCT:30% sucrose mixture for 6 h and then transferred to a cryomold and filled with OCT. The embedded tissue was frozen on dry ice and either cryosectioned at 20 μ m or stored at -80 °C. For immunohistochemistry, sections were washed in PBS1X for 10 min and then blocked and permeabilized in 0.3% Triton X-100 and 5% normal donkey serum in PBS1X. After incubation with primary antibodies, the sections were incubated for 1 h with the appropriate secondary antibodies (Alexa Fluor 488, 594, and 647 used at 1:400, Molecular Probes) and then mounted on gelatin-coated slides and coverslipped with PVA-DABCO containing DAPI (1:1000). A list of primary antibodies is reported in Supplementary Table 2.

Quantification of fluorescence for developmental layers identified was performed using Image J software (NIH, v1.49). Measurements were performed by taking a radial line-intensity profile (80 μ m width) for each channel from the center of the organoid to the edge of the organoid, subtracting the background and recording the position where the positive signal begins and ends. Recordings were normalized to the length of the radial line and mean value with standard deviation plotted for each of three fluorescence channels.

Whole *silk*- and conventionally generated VM organoids were sequentially sectioned to obtain slices from edge to core. For quantifications, each section was quartered into 4 equal areas and then scanned using a confocal microscope under 20X magnification. The number of TH⁺, Calb⁺, Girk2⁺, and DDC⁺ cells was manually counted in each area using Image J software (NIH, v1.49), and summed

to give the total number of positive cells per slice. Counting was performed on every 3 sections and the final counts were corrected for the total number of sections per organoid.

Immunohistochemistry and neuromelanin staining. For diaminobenzidine (DAB) staining, the sections were incubated with secondary biotinylated horse antibodies diluted (1:200, Vector Laboratories) for 1 h at room temperature (RT), washed three times, and then incubated with avidin-biotin complex (ABC) for 1 h at RT for amplification. Peroxidase-driven precipitation of DAB was used to detect the primary antibody. In this step, the sections were incubated in 0.05% DAB for 1–2 min before addition of 0.01% H₂O₂ for a further 1–2 min. After development of DAB staining, the sections were placed in an ammoniacal silver solution to detect neuromelanin using a Masson Fontana Stain Kit (Atom Scientific, #RRSK12-100), according to the manufacturer's instructions. The sections were then mounted on gelatin-coated slides and dehydrated in an ascending series of alcohol concentrations cleared in xylene, and coverslipped with DPX mounting.

iDISCO. Both conventional and *silk*-VM organoids were fixed in 2% paraformaldehyde overnight at 4 °C and permeabilized in 0.2% Triton X-100 20% DMSO and then in 0.1% Triton X-100, 0.1% Tween20, 0.1% C₂₄H₃₉NaO₄, 0.1% NP40, and 20% DMSO overnight at 37 °C. After incubation with primary antibodies for 2 days at 37 °C, the organoids were incubated for 2 days with the appropriate secondary antibody, embedded in 1% agarose, and dehydrated in an ascending series of methanol concentrations and dichloromethane as previously described⁶². Samples were imaged in a chamber filled with DBE. The cleared brain organoids were imaged on an Ultra Microscope II (LaVision Biotec) equipped with an sCMOS camera (Andor Neo, model 5.5-CL3) and 12x objective lenses (LaVision LVMi-Fluor 4x/0.3 or 12x/0.53 MI Plan). We used two laser configurations with the following emission filters: 525/50 for endogenous background and AlexaFluor 488. Whole-organoid volumes were quantified based on endogenous background, and DA neuron number based on TH+ cell. Several stacks (mosaic acquisition) were taken with 10% overlap to cover the entire brain-organoid volume, including edge and core. Stacks were acquired with ImspectroPro64 (LaVision Biotec) using 3 μm z-steps to acquire the volume in 3D. These image stacks were stitched to visualize the brain organoid in 3D with Arivis Vision 4D 3.5.0 (Arivis AG). The high density of cells in the core volume, characterized by the absence of TH+ staining, was quantified based on 525/50 endogenous signal. Rendered movies were compiled in Final Cut Pro 10.4.3 (Apple Inc.).

Microscopy. Images were captured using an Epson Perfection V850 PRO flatbed scanner, a Leica DMI6000B widefield microscope, or a Leica TCS SP8 confocal laser-scanning microscope, or a Nikon inverted Ti2 microscope equipped with a CSU-W1 spinning-disk system. Image acquisition software was Leica LAS X and images were processed using Velocity 6.5.1 software (Quorum Technologies) and Adobe Photoshop. Any adjustments were applied equally across the entire image, and without the loss of any information.

Flow cytometry. TH-Cre hPSC-derived single-cell suspensions, differentiated either from conventional or *silk*-VM organoids, were obtained using the Papain Kit (Worthington, #LK003150). GFP⁺ cells were analyzed with BD FACSAria III (BD Biosciences) and all data plots were generated using FlowJo software.

Western blotting. VM-organoid lysates were prepared with ice-cold immunoprecipitation assay (RIPA) lysis buffer. Whole-cell lysates were then separated on gels (Invitrogen, NuPAGE 4–12%) and transferred to a PVDF membrane (Millipore, Immobilon-P Membrane, 0.45 μm). Detection was performed with ECL reagents (Amersham Biosciences). Details and list of antibodies are reported in Supplementary Table 2.

Organoid-slice preparation and electrophysiology. About 3% low-melting-point agarose (Promega, #V2111) was dissolved and melted in Neurobasal medium at 37 °C and placed in a cubic PDMS mold (1 cm³). At day 90, VM organoids were immersed and embedded in an agarose mold and left to solidify for 5–10 mins at RT. Sections of 400 μm thickness were prepared using a vibratome tissue slicer (Leica VT1000 S) at 0.1 mm/s speed and 1 mm vibration amplitude. Agarose blocks containing the VM-organoids were submerged in oxygenated artificial cerebrospinal fluid (ACSF) at 4 °C during the cutting procedure. VM organoid slices were equilibrated for 30 min in oxygenated ACSF at 37.5 °C prior to whole-cell patch clamp recordings⁴².

Whole-cell patch-clamp electrophysiological recordings were performed at day 90 of VM organoid differentiation. Both conventional and *silk*-VM organoids were transferred to a recording chamber containing Krebs solution gassed with 95% O₂ and 5% CO₂ at RT and exchanged every 20 min during recordings. The standard solution was composed of (in mM) 119 NaCl, 2.5 KCl, 1.3 MgSO₄, 2.5 CaCl₂, 25 glucose, and 26 NaHCO₃. For recordings, a Multiclamp 700B Microelectrode Amplifier (Molecular Devices) was used together with borosilicate glass pipettes (3–7 MΩ) filled with the following intracellular solution (in mM): 122.5 potassium gluconate, 12.5 KCl, 0.2 EGTA, 10 HEPES, 2 MgATP, 0.3 Na₃GTP, and

8 NaCl adjusted to pH 7.3 with KOH, as previously described²⁹. Data acquisition was performed with pCLAMP 10.2 software (Molecular Devices); current was filtered at 0.1 kHz and digitized at 2 kHz. Cells with neuronal morphology and round cell body were selected for recordings. Resting-membrane potentials were monitored immediately after breaking-in in current-clamp mode. Thereafter, cells were kept at a membrane potential of −60 mV to −80 mV, and 500 ms currents were injected from −85 pA to +165 pA with 20 pA increments to induce action potentials. For inward sodium and delayed rectifying potassium current measurements, cells were clamped at −70 mV and voltage-depolarizing steps were delivered for 100 ms at 10 mV increments. Spontaneous action potentials were recorded in current-clamp mode at resting membrane potentials.

Calcium imaging of MAP2-GCamP5-labeled neurons. Calcium imaging was performed at day 120 of VM cultures containing the MAP2-GCamP5 reporter. Imaging was performed on an inverted Ti2 microscope (Nikon) equipped with a CSU-W1 spinning-disc system (Yokogawa), a sCMOS camera (Teledyne Photometrics), and a 20 × objective. An environment control chamber was used to maintain the temperature at 37 °C and CO₂ level at 5% during imaging. Exposure time was set to 50 ms. Spontaneous activity was recorded from 3 different *silk*-VM organoids. Images were analyzed in ImageJ (NIH).

Chronoamperometry. Three-electrode setup with a pyrolyzed carbon fiber as the working electrode was used to detect dopamine. Electrode fabrication, electrochemical characterization, and setup assembly were performed as previously described⁴⁹. In short, to perform the measurements, an organoid was placed on the working electrode immersed in baseline buffer, a constant potential was applied, and the resulting current monitored. Dopamine release was chemically triggered through addition of stimulation buffer that elevated K⁺ concentration to 150 mM. Measured current was normalized to the baseline value and plotted with respect to time.

Statistics and reproducibility. Statistical analysis of qRT-PCR data and immunofluorescence-based quantifications was performed using two-tailed Student's *t*-test and *p*-values <0.05 were considered significant. For comparisons of electrophysiological properties, two-tailed Student's *t*-test and one-way analysis of variance (ANOVA) was used. Data were statistically analyzed with the GraphPad Prism 9 software and presented as mean ± SEM, except where stated otherwise. Statistical analysis of sequencing data was conducted using two-tailed Wilcoxon rank-sum test (Seurat v3.1) in R v3.6.1. Please refer to the bioinformatics analysis section above for more details. Immunohistochemical staining images are representative of 6–12 sections from at least 4 biologically independent organoids. Western blots are representative of 3 biological replicates.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The scRNA-seq data generated in this study have been deposited in the Gene Expression Omnibus under accession code “GSE168323”. A reporting summary for this article is available as Supplementary Information file. All other relevant data supporting the key findings of this study are available within the article and its Supplementary Information files or from the corresponding author upon reasonable request. Source data are provided with this paper.

Code availability

The code used for this paper is available on GitHub https://github.com/ParmarLab/scRNA-seq_silk_organoids with <https://doi.org/10.5281/zenodo.5603294>.

Received: 24 February 2021; Accepted: 9 November 2021;

Published online: 15 December 2021

References

1. Kelava, I. & Lancaster, M. A. Stem cell models of human brain development. *Cell Stem Cell* **18**, 736–748 (2016).
2. Rifes, P. et al. Publisher correction: modeling neural tube development by differentiation of human embryonic stem cells in a microfluidic WNT gradient. *Nat. Biotechnol.* **38**, 1357 (2020).
3. Demers, C. J. et al. Development-on-chip: in vitro neural tube patterning with a microfluidic device. *Development* **143**, 1884–1892 (2016).
4. Lancaster, M. A. et al. Cerebral organoids model human brain development and microcephaly. *Nature* **501**, 373–379 (2013).
5. Bagley, J. A. et al. Fused cerebral organoids model interactions between brain regions. *Nat. Methods* **14**, 743–751 (2017).

6. Fiorenzano, A. et al. Dopamine neuron diversity: recent advances and current challenges in human stem cell models and single cell sequencing. *Cells* **10**, 1366 (2021).
7. Quadrato, G. et al. Cell diversity and network dynamics in photosensitive human brain organoids. *Nature* **545**, 48–53 (2017).
8. Camp, J. G. et al. Human cerebral organoids recapitulate gene expression programs of fetal neocortex development. *Proc. Natl Acad. Sci. USA* **112**, 15672–15677 (2015).
9. Kanton, S. et al. Organoid single-cell genomic atlas uncovers human-specific features of brain development. *Nature* **574**, 418–422 (2019).
10. Qian, X., Song, H. & Ming, G. L. Brain organoids: advances, applications and challenges. *Development* **146**, dev166074 (2019).
11. Mansour, A. A., Schafer, S. T. & Gage, F. H. Cellular complexity in brain organoids: current progress and unsolved issues. *Semin Cell Dev. Biol.* **111**, 32–39 (2021).
12. Qian, X. et al. Sliced human cortical organoids for modeling distinct cortical layer formation. *Cell Stem Cell* **26**, 766–781 e9 (2020).
13. Bhaduri, A. et al. Cell stress in cortical organoids impairs molecular subtype specification. *Nature* **578**, 142–148 (2020).
14. Nolbrant, S. et al. Generation of high-purity human ventral midbrain dopaminergic progenitors for in vitro maturation and intracerebral transplantation. *Nat. Protoc.* **12**, 1962–1979 (2017).
15. Jo, J. et al. Midbrain-like organoids from human pluripotent stem cells contain functional dopaminergic and neuromelanin-producing neurons. *Cell Stem Cell* **19**, 248–257 (2016).
16. Qian, X. et al. Brain-region-specific organoids using mini-bioreactors for modeling ZIKV exposure. *Cell* **165**, 1238–1254 (2016).
17. Kwak, T. H. et al. Generation of homogeneous midbrain organoids with in vivo-like cellular composition facilitates neurotoxin-based Parkinson's disease modeling. *Stem Cells* **38**, 727–740 (2020).
18. Kim, H. et al. Modeling G2019S-LRRK2 sporadic Parkinson's disease in 3D midbrain organoids. *Stem Cell Rep.* **12**, 518–531 (2019).
19. Monzel, A. S. et al. Derivation of human midbrain-specific organoids from neuroepithelial stem cells. *Stem Cell Rep.* **8**, 1144–1154 (2017).
20. Renner, H. et al. A fully automated high-throughput workflow for 3D-based chemical screening in human midbrain organoids. *Elife* **2020**, 9.
21. Smits, L. M. et al. Single-cell transcriptomics reveals multiple neuronal cell types in human midbrain-specific organoids. *Cell Tissue Res.* **382**, 463–476 (2020).
22. Agarwal, D. et al. A single-cell atlas of the human substantia nigra reveals cell-specific pathways associated with neurological disorders. *Nat. Commun.* **11**, 4183 (2020).
23. Tiklova, K. et al. Single-cell RNA sequencing reveals midbrain dopamine neuron diversity emerging during mouse brain development. *Nat. Commun.* **10**, 1–12 (2019).
24. Poulin, J. F. et al. Defining midbrain dopaminergic neuron diversity by single-cell gene expression profiling. *Cell Rep.* **9**, 930–943 (2014).
25. Saunders, A. et al. Molecular diversity and specializations among the cells of the adult mouse brain. *Cell* **174**, 1015–1030 e16 (2018).
26. Birey, F. et al. Assembly of functionally integrated human forebrain spheroids. *Nature* **545**, 54–59 (2017).
27. Lancaster, M. A. & Knoblich, J. A. Generation of cerebral organoids from human pluripotent stem cells. *Nat. Protoc.* **9**, 2329–2340 (2014).
28. Kirkeby, A. et al. Predictive markers guide differentiation to improve graft outcome in clinical translation of hESC-based therapy for Parkinson's disease. *Cell Stem Cell* **20**, 135–148 (2017).
29. Nilsson, F. et al. Single-cell profiling of coding and noncoding genes in human dopamine neuron differentiation. *Cells* **10**, 137 (2021).
30. Elkabetz, Y. et al. Human ES cell-derived neural rosettes reveal a functionally distinct early neural stem cell stage. *Genes Dev.* **22**, 152–165 (2008).
31. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
32. Ono, Y. et al. Differences in neurogenic potential in floor plate cells along an anteroposterior location: midbrain dopaminergic neurons originate from mesencephalic floor plate cells. *Development* **134**, 3213–3225 (2007).
33. Bonilla, S. et al. Identification of midbrain floor plate radial glia-like cells as dopaminergic progenitors. *Glia* **56**, 809–820 (2008).
34. Street, K. et al. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 (2018).
35. La Manno, G. et al. Molecular diversity of midbrain development in mouse, human, and stem cells. *Cell* **167**, 566–580 e19 (2016).
36. Poulin, J. F. et al. Classification of midbrain dopamine neurons using single-cell gene expression profiling approaches. *Trends Neurosci.* **43**, 155–169 (2020).
37. Hook, P. W. et al. Single-Cell RNA-Seq of mouse dopaminergic neurons informs candidate gene selection for Sporadic Parkinson disease. *Am. J. Hum. Genet.* **102**, 427–446 (2018).
38. Kramer, D. J. et al. Combinatorial expression of Grp and Neurod6 defines dopamine neuron populations with distinct projection patterns and disease vulnerability. *eNeuro* **10**.1523/ENEURO.0152-18.2018 (2018).
39. Panman, L. et al. Sox6 and Otx2 control the specification of substantia nigra and ventral tegmental area dopamine neurons. *Cell Rep.* **8**, 1018–1025 (2014).
40. Braskie, M. N. et al. Common Alzheimer's disease risk variant within the CLU gene affects white matter microstructure in young adults. *J. Neurosci.* **31**, 6764–6770 (2011).
41. Lancaster, M. A. et al. Guided self-organization and cortical plate formation in human brain organoids. *Nat. Biotechnol.* **35**, 659–666 (2017).
42. Wiley, L. A. et al. A method for sectioning and immunohistochemical analysis of stem cell-derived 3-D organoids. *Curr. Protoc. Stem Cell Biol.* **37**, 1C 19 1–1C 19 11 (2016).
43. Widhe, M., Shalaly, N. D. & Hedhammar, M. A fibronectin mimetic motif improves integrin mediated cell bidding to recombinant spider silk matrices. *Biomaterials* **74**, 256–266 (2016).
44. Astrand, C. et al. Assembly of FN-silk with laminin-521 to integrate hPSCs into a three-dimensional culture for neural differentiation. *Biomater. Sci.* **8**, 2514–2525 (2020).
45. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 e21 (2019).
46. Fiorenzano, A. et al. Evaluation of TH-Cre knock-in cell lines for detection and specific targeting of stem cell-derived dopaminergic neurons. *Helyon* **7**, e06006 (2021).
47. Kadoshima, T. et al. Self-organization of axial polarity, inside-out layer pattern, and species-specific progenitor dynamics in human ES cell-derived neocortex. *Proc. Natl Acad. Sci. USA* **110**, 20284–20289 (2013).
48. Pasca, A. M. et al. Human 3D cellular model of hypoxic brain injury of prematurity. *Nat. Med.* **25**, 784–791 (2019).
49. Vasudevan, S. et al. Leaky optoelectrical fiber for optogenetic stimulation and electrochemical detection of dopamine exocytosis from human dopaminergic neurons. *Adv. Sci.* **6**, 1902011 (2019).
50. Grealish, S. et al. Human ESC-derived dopamine neurons show similar preclinical efficacy and potency to fetal neurons when grafted in a rat model of Parkinson's disease. *Cell Stem Cell* **15**, 653–665 (2014).
51. Hoban, D. B. et al. Impact of alpha-synuclein pathology on transplanted hESC-derived dopaminergic neurons in a humanized alpha-synuclein rat model of PD. *Proc. Natl Acad. Sci. USA* **117**, 15209–15220 (2020).
52. Velasco, S. et al. Individual brain organoids reproducibly form cell diversity of the human cerebral cortex. *Nature* **570**, 523–527 (2019).
53. Tiklova, K. et al. Single cell transcriptomics identifies stem cell-derived graft composition in a model of Parkinson's disease. (vol 11, 2434, 2020). *Nat. Commun.* **11**, 1–11 (2020).
54. Björklund, A. & Dunnett, S. B. Dopamine neuron systems in the brain: an update. *Trends Neurosci.* **30**, 194–202 (2007).
55. Brichta, L. & Greengard, P. Molecular determinants of selective dopaminergic vulnerability in Parkinson's disease: an update. *Front Neuroanat.* **8**, 152 (2014).
56. Kajtez, J. et al. 3D biomaterial models of human brain disease. *Neurochem. Int.* **147**, 105043 (2021).
57. Hofer, M. & Lutolf, M. P. Engineering organoids. *Nat. Rev. Mater.* **6**, 402–420 (2021).
58. Mansour, A. A. et al. An in vivo model of functional and vascularized human brain organoids. *Nat. Biotechnol.* **36**, 432–441 (2018).
59. Cakir, B. et al. Engineering of human brain organoids with a functional vascular-like system. *Nat. Methods* **16**, 1169–1175 (2019).
60. Shi, Y. et al. Vascularized human cortical organoids (vOrganoids) model cortical development in vivo. *PLoS Biol.* **18**, e3000705 (2020).
61. Johansson, U. et al. Assembly of functionalized silk together with cells to obtain proliferative 3D cultures integrated in a network of ECM-like microfibers. *Sci. Rep.* **9**, 6291 (2019).
62. Renier, N. et al. iDISCO: a simple, rapid method to immunolabel large tissue samples for volume imaging. *Cell* **159**, 896–910 (2014).

Acknowledgements

We thank BioLamina and Spiber® for providing Biosilk. We thank Dr Therese Kallur, Prof. My Hedhammar, Prof. Anders Björklund for helpful comments and discussions, and Ulla Jarl, Sol Da Rocha Baez, and Marie Persson Veijgården for technical assistance. The research leading to these results has received funding from the New York Stem Cell Foundation, European Research Council (ERC) under ERC Grant Agreement 771427, European Union-funded project NSC-Reconstruct (European Union, H2020, GA no 874758, 2020–23), Swedish Research Council (2016-00873), Swedish Parkinson Foundation (Parkinsonfonden), Swedish Brain Foundation, Strategic Research Area at Lund University Multipark, and Knut and Alice Wallenberg Stiftelse (KAW 2018-0040). Alessandro Fiorenzano is the recipient of a Thorsten and Elsa Segerfalk Foundation grant, which was used to support this study. Malin Parmar is a New York Stem Cell Foundation Robertson Investigator.

Author contributions

M.P. and A.F. conceived the project, designed experiments, and interpreted computational analysis, analyzed all results in the project, and wrote the paper. A.F. and E.S.

generated organoids, performed experiments and interpreted histological data with help from J.K., F.N., J.G., B.M., and Y.Z. P.S. and Y.S. analyzed and interpreted the single-cell RNA data. D.R.O., M.B., and A.B. performed and interpreted electrophysiological experiments. J.E. and J.K. analyzed and interpreted amperometry experiments. All authors read and approved the final paper.

Funding

Open access funding provided by Lund University.

Competing interests

M.P. is the owner of Parmar Cells AB and co-inventor of the following patents WO2016162747A2, WO2018206798A1, and WO2019016113A1. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-27464-5>.

Correspondence and requests for materials should be addressed to Alessandro Fiorenzano.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021, corrected publication 2022

High-throughput single-cell transcriptomics on organoids

Agnieszka Brazovskaja¹, Barbara Treutlein^{1,2,3} and J Gray Camp¹



Three-dimensional (3D) tissues grown in culture from human stem cells offer the incredible opportunity to analyze and manipulate human development, and to generate patient-specific models of disease. Methods to sequence DNA and RNA in single cells are being used to analyze these so-called ‘organoid’ systems in high-resolution. Single-cell transcriptomics has been used to quantitate the similarity of organoid cells to primary tissue counterparts in the brain, intestine, liver, and kidney, as well as identify cell-specific responses to environmental variables and disease conditions. The merging of these two technologies, single-cell genomics and organoids, will have profound impact on personalized medicine in the near future.

Addressees

¹ 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, 80333 Munich, Germany

Corresponding authors:

Treutlein, Barbara (barbara_treutlein@eva.mpg.de),
Camp, J Gray (gray_camp@eva.mpg.de)

Current Opinion in Biotechnology 2019, 55:167–171

This review comes from a themed issue on **Analytical biotechnology**

Edited by Saulius Klimasauskas and Linas Mazutis

For a complete overview see the [Issue](#) and the [Editorial](#)

Available online 29th November 2018

<https://doi.org/10.1016/j.copbio.2018.11.002>

0958-1669/© 2018 Published by Elsevier Ltd.

conventional 2D culture counterparts by creating complex microenvironments where multiple lineages structurally organize and communicate to maintain a balanced physiological status. However, it has been unclear how precisely these organoid systems recapitulate the cell state-specific gene expression landscapes of the tissues they intend to model. Single-cell genomic methods (reviewed in Ref. [2]) have brought new analytic approaches to characterize these organoid models with more resolution and less bias than previous immunohistological or bulk RNA-seq descriptions of organoid development.

Most single-cell RNA-seq methods require each cell to be physically captured in a small volume where cells can be lysed, and chemistry can be performed on the contents of each individual cell. Capture can be achieved by hand picking or flow cytometric sorting into multi-well plates [3], limited dilutions into wafers containing hundreds of microwells [4], or through valve [5] or droplet-based microfluidic approaches [6•,7•]. In addition, combinatorial barcoding strategies have been developed that enable barcoding of cellular RNAs without physically isolating the cells [8]. Each strategy has particular benefits or drawbacks depending on the application, however the approaches that enable high-throughput (thousands of cells per experiment) are best suited to efficiently sample the complex cellular diversity in organoids and to understand organoid to organoid variability. One limitation of the higher throughput approaches is that only the 3' or 5' end of the transcript is sequenced limiting inquiry into certain features of the transcriptome (e.g. alternative splicing).

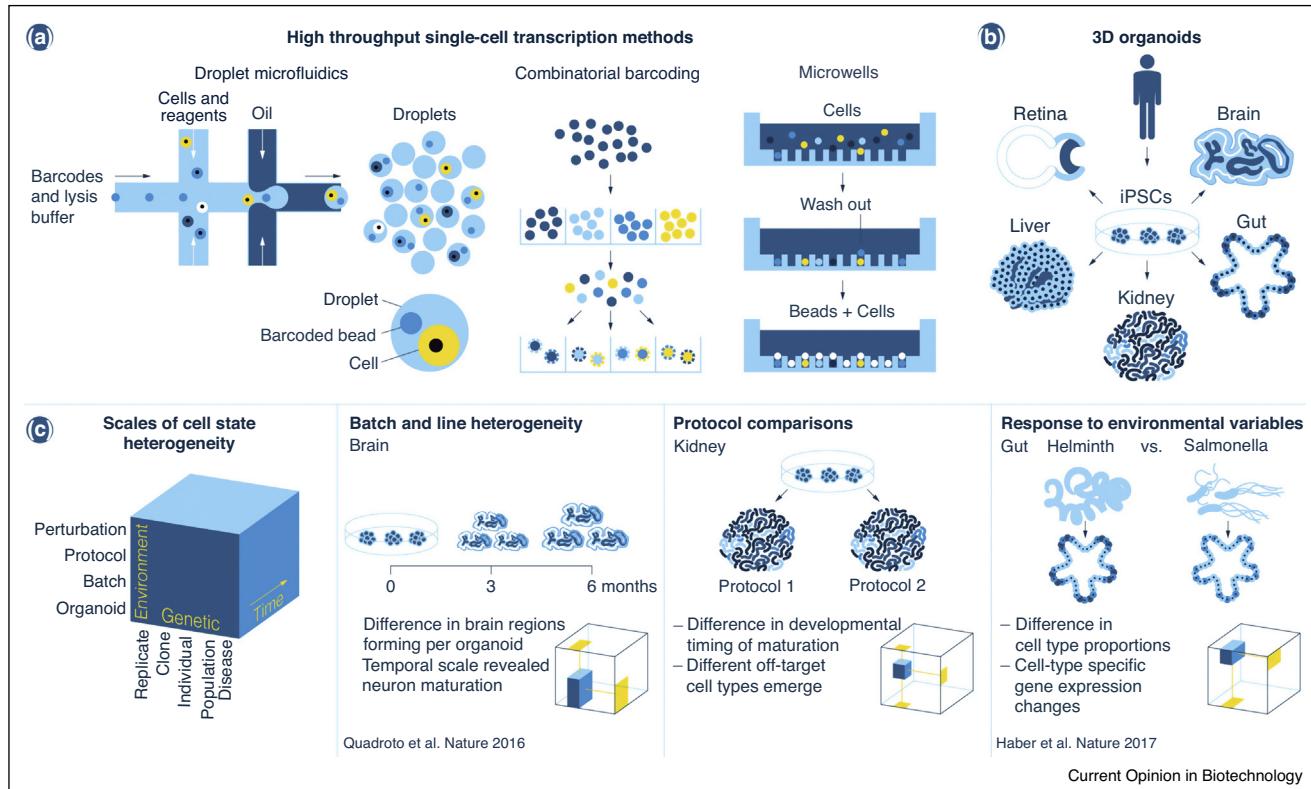
These quantitative SCG technologies are being used to study how each cell fate is regulated within complex multi-lineage human organoids (Figure 1). This data can be used to eventually generate organ-level computational models of human development, which may be used to predict disease mechanisms. Here we review recent advances at the intersection of single-cell genomics (SCG) and human tissue engineering, and highlight existing challenges in the organoid field where high-throughput SCG strategies can have an immediate impact. We concentrate on organoid modeling of the brain, intestine, liver, and kidney, where there has been the most recent progress.

Brain

There have been multiple methods published to generate 3D tissue cultures that resemble the developing human

Introduction

The human body is composed of an amazing diversity of cell fates, which transition through various states during organ development, disease, and regeneration. Methods to engineer diverse human tissues from stem cells are being developed at a rapid pace in order to study uniquely human organ development and disease in controlled culture environments. Depending on the tissue and disease, these culture models can be initiated from organ-specific adult stem cells or iPSCs generated from skin or blood [1]. These 3D tissues, commonly called ‘organoids’, offer the promise to more accurately model human development, physiology and disease than

Figure 1

Single-cell transcriptomics is enabling high-resolution analysis of cell heterogeneity in human organoids.

(a) New high-throughput single-cell transcriptomic methods based on droplet microfluidics, combinatorial barcoding, or microwell technologies have dramatically expanded the number of cells that can be analyzed per experiment, opening up new inroad into quantitative comparisons of cell states. **(b)-(c)** These high-throughput methods are enabling the analysis of human 3D organoids derived from induced pluripotent stem cells (b) across a range of environmental, genetic, and temporal scales (c). High-throughput methods will make it possible to quantitatively assess cell states between organoids, batches, protocols, and various perturbations (environmental scale); between replicates, clones, individuals from the same or different populations, and patients with a genetic disorder (genetic scale); and each of the genetic or environmental impacts on cell state can be assessed over a time course (time scale). The panels highlight that multiple publications have already started to address how various environmental or genetic conditions effect organoid cell states in brain, kidney and gut organoids.

brain. These approaches come in two distinct categories. First, iPSC-derived neuroectoderm can be allowed to self-organize into cerebral organoids that contain multiple interconnected brain regions. Second, iPSC aggregates can be patterned to generate distinct, independent brain regions. In both cases, progenitor cells exhibit very similar morphology and behaviors that have been observed in fetal tissue. Neurons mature, establish synapses, spontaneously fire action potentials, and may even respond to physiological stimuli. Camp, Badsha *et al.* was the first to directly compare cerebral organoids with the early fetal neocortex using scRNA-seq and found that cell composition, lineage relationships, and gene expression programs were largely recapitulated in the organoid cortical regions (Figure 1) [9]. A major limitation of this study was the scRNA-seq technology was low throughput, resulting less than 1000 cells. Quadratto *et al.* substantially advanced the characterization of multi-region cerebral organoids using

the Drop-seq method based on droplet microfluidics. It allowed the authors to sequence over 80 000 cells from 31 whole-brain organoids, which clustered into many distinct populations representing different brain developmental identities, including forebrain and retina [10••]. It also revealed, at the transcriptome level, that neurons progressively matured from three to six months in culture along multiple lineages (callosal projection neurons as well as Müller glial and bipolar cells). The authors also provided evidence that organoids from the same bioreactor contained more comparable cell types than between bioreactors, likely due to variation in organoid brain region composition [11].

Birey *et al.* analyzed human iPSC-derived dorsal and ventral forebrain spheroids before and after fusion [12]. The data showed that cells were remarkably well patterned. It was shown for the first time that cells which

migrated from the ventral to the dorsal region could functionally integrate into cortical networks. Migrated cells had increased complexity of dendrites branching, twice higher action potential generation rate than non-migrated or one region cells, expressed the presynaptic and postsynaptic proteins and formed functional synapses. In addition, the results of modeling a neurodevelopmental disorder on forebrain assembled spheroids from different patients' iPSC lines were compatible with the expected disease phenotype. In each of these cases, there was no robust characterization of how iPSC lines from different human individuals, or even different lines from the same individual compare in gene expression space. Such a quantitative characterization will be required in order to understand disease models. Furthermore, it is not clear if cortical organoids can accurately model the expansion of the outer radial glial populations that is observed in later stages of human neocortex development [13]. Indeed, the spatial heterogeneity of organoid cultures naturally requires spatial transcriptomics and there will be major advances in the coming years in this area of research. Finally, recent work has shown that organoids can generate complex network wiring, and it will be interesting to couple single-cell transcriptomics to activity-dependent electrical stimulations [14].

Gut

Gut organoids (stomach, small intestine, colon) can be established either from adult-derived intestinal epithelial stem cells [15,16] or through the differentiation of iPSCs through endoderm derived foregut and hindgut spheroids [17,18]. To date, scRNA-seq on mouse small intestine organoids has provided the most insight into the cellular composition of adult stem cell-derived organoids. Low-throughput methods were used to identify the distinct cardinal populations of the intestinal epithelium (stem cells, paneth cells, goblet cells, enteroendocrine cells, enterocytes) [19]. However, enteroendocrine cells needed to be enriched in order to identify enteroendocrine subpopulations due the use of low-throughput methods. Haber, Biton, Rogel *et al.* generated an atlas of mouse small intestine cell composition from 53 193 number cells. From this survey, they could identify each cell population as well as subtypes of enteroendocrine cells, such as early, middle and late precursors and mature cells [20••]. A new enteroendocrine cell taxonomy was created by comparing the expression of detected genes across two subtypes to canonical classification markers. Results revealed that previously defined markers, for example Sct, Cck, Gcg and Ghrl, were not subtype-specific but expressed across multiple cell types. Interestingly, there was heterogeneity in an enterochromaffin cell population, which split into two distinct subtypes. In addition, the authors compared cell composition and gene expression landscapes in organoids after exposure to different microbe populations. Extended to humans, this strategy will allow controlled experimentation of human

intestine to diverse dietary, microbial, or pharmaceutical manipulations. Currently, however, improved organoid culture methods are needed to enable human gut organoids to maintain a balance of progenitors and differentiated cells within a 3D structure, and there is currently no single-cell transcriptomics manuscript published that analyzed human intestinal organoids.

Liver

The dynamic developmental, structural, and cellular heterogeneity of the liver makes it challenging to recapitulate the growth pathways of this organ *in vitro* from pluripotency. The developing fetal liver initially serves as the reservoir for hematopoiesis before structural reorganization and maturation into the major metabolic organ of the body. Multiple different protocols have been developed to differentiate iPSCs in 2D monocultures to hepatic endoderm and then toward 'hepatocyte-like' cells [21]. However, these cells are not functioning as mature, metabolically complete hepatocytes and certain widely used protocols generate cells with only a modest similarity to human hepatocytes, which may even be off target cells with similarity to the intestinal epithelium [22]. Incorporating additional lineages thought to provide signals that specify hepatic fate acquisition, such as the transverse mesenchyme and nascent endothelium, into a 3D microenvironment has shown great promise in generating hepatic organoids [23]. We have analyzed this system and found that the hepatic cells within these *in vitro* and transplanted human organoids acquire a significant increase in similarity to fetal hepatocytes relative to the 2D counterparts [24]. However, in all cases we have analyzed thus far, there remains a major challenge to generate mature hepatocytes with fully metabolic functionality from human iPSCs.

It has been shown that 2D monocultures of primary adult human hepatocytes can only be maintained short-term due to dedifferentiation and cell death [25]. Recently, protocols were developed to isolate hepatic stem cells (HSCs) from adult tissues and culture HSCs in 3D matrix environments that support proliferation and differentiation of hepatic epithelium [26]. These methods are revealing the potential to maintain differentiated hepatic cells *in vitro*. Single-cell RNA-seq analyses on adult liver could in principle map the transcriptome states of the HSCs and mature hepatocytes *in vivo*, and be used to assess the accuracy and precision of the adult stem-cell derived organoids maintained *in vitro*. In any case, major current protocols lack the cellular diversity (e.g. kupffer cells, stellate cells, bile ducts, portal endothelium, etc.) and the structural organization of the human liver. Spatial maps of mouse liver hepatocyte transcripts confirmed that hepatocytes were ordered into a metabolic hierarchy that correlates with proximity to the portal vein and bile ducts [27]. In the future, full reconstructions of liver development, from fetus to adult, with structural and cell state resolution will enable reverse engineering and benchmarking of 3D liver organoid technologies.

Kidney

Multiple recent studies showed that 3D kidney organoids can be generated from pluripotent stem cells [28–30]. Bulk transcriptome analysis suggested that these organoid systems resemble approximately the first trimester of human kidney development. Recently, single-cell transcriptomics was used to dissect cell composition in human kidney organoids and to identify cell types that could be impacted by disease-related genes predicted from genome-wide association studies [30,31[•],32,33]. The mapping of disease-related genes was generally consistent with a scRNA-seq study of the mouse kidney, which highlights that discrete human disease phenotypes are due to mutations in genes that have a cell-type-specific gene expression pattern in the kidney [32]. In Wu *et al.*, the authors showed that both organoid systems contain very similar cardinal renal cell types (podocytes, mesenchyme, tubule cells), albeit at different proportions, and that the cells are relatively immature compared to fetal and adult renal cells. The authors also identified multiple populations of non-renal cells, and developed an inhibition strategy based on receptor expression to reduce the prevalence of these off-target cells. This manuscript is a great example of how single-cell transcriptomics can be used not only to assess the quality of the organoids, but also guide the engineering process.

Furthermore, establishment of automated high-throughput human organoid generation platforms that enable testing of culture conditions to enhance cell differentiation, predict chemical toxicity, and phenotype organoids promise rapid innovations in culture methods and assessment of disease phenotypes [34]. Czerniecki *et al.* used scRNA-seq to characterize kidney organoids that had been optimized by robotic manipulation of culture conditions in hundreds of mini-organoids in microwells. Image-based data of marker genes had suggested that the addition of vascular endothelial growth factor (VEGF) increased the abundance of endothelial cells in the organoids. However, single cell RNA-sequencing on 10535 cells from organoids treated with or without VEGF, revealed that there were very few mature endothelial cells present in the organoids, which had not been resolved by immunohistochemistry. The authors conclude that the VEGF treatment greatly increases the number of endothelial cell progenitors in the organoid cultures, but only a small minority of these cells reaches a mature endothelial cell differentiation state similar to that found *in vivo*. This data further underscores the importance of high-resolution descriptions of cell states that arise in organoid culture systems.

Future prospective

Personalized medicine is on the horizon, where an individual's genome can be integrated with personalized 3D tissue culture models to create avatars of disease particular to the patient. Depending on the tissue and disease, these culture models can be initiated from iPSCs,

organ-specific adult stem cells, or cancer tissue. Even though recent protocol enhancements enable longer-term growth and development of 3D organoids, the challenge remains to generate 3D models that recapitulate mature cellular and physiological phenotypes observed in postnatal human tissues. Single-cell sequencing methods are required to assess the accuracy, precision, and efficiency of protocol enhancements. Additionally, because of the cellular complexity and general batch heterogeneity of organoids, it is critical to sample as many cells as economically feasible in multiple organoids per experiment, ideally with spatial [35] and lineage resolution [36]. We expect that over the coming years there will be dedicated efforts to analyze organoids from hundreds of patients with particular diseases from iPSC and adult organoid biobanks. These efforts will require increases in sample throughput either through sample tagging [37], combinatorial barcoding [38[•]], or random composite measurements [39]. As the field progresses, robust computational strategies will be required to integrate the data and make biological sense of what is sure to be high-information content and extremely complex data.

Conflict of interest statement

Nothing declared.

Acknowledgement

This work was supported by the Max Planck Society.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Clevers H: **Modeling development and disease with organoids.** *Cell* 2016, **165**:1586–1597 <http://dx.doi.org/10.1016/j.cell.2016.05.082>.
 2. Picelli S: **Single-cell RNA-sequencing: the future of genome biology is now.** *RNA Biol* 2016;1–14 <http://dx.doi.org/10.1080/15476286.2016.1201618>.
 3. Picelli S *et al.*: **Smart-seq2 for sensitive full-length transcriptome profiling in single cells.** *Nat Methods* 2013, **10**:1096–1098 <http://dx.doi.org/10.1038/nmeth.2639>.
 4. Han X *et al.*: **Mapping the mouse cell atlas by microwell-seq.** *Cell* 2018, **172**:1091–1107 <http://dx.doi.org/10.1016/j.cell.2018.02.001> e1017.
 5. Treutlein B *et al.*: **Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq.** *Nature* 2014, **509**:371–375 <http://dx.doi.org/10.1038/nature13173>.
 6. Macosko EZ *et al.*: **Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets.** *Cell* 2015, **161**:1202–1214 <http://dx.doi.org/10.1016/j.cell.2015.05.002>.
- These two papers presented the first high-throughput single-cell RNA-seq data using droplet microfluidic technology, and the authors made their protocols widely accessible to scientists enabling the establishment of this specialized method in labs around the world. Subsequent commercializations of droplet-based methods have dramatically expanded the single-cell genomics field.
7. Klein AM *et al.*: **Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells.** *Cell* 2015, **161**:1187–1201 <http://dx.doi.org/10.1016/j.cell.2015.04.044>.

These two papers presented the first high-throughput single-cell RNA-seq data using droplet microfluidic technology, and the authors made their protocols widely accessible to scientists enabling the establishment of this specialized method in labs around the world. Subsequent commercializations of droplet-based methods have dramatically expanded the single-cell genomics field.

8. Rosenberg AB et al.: **Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding.** *Science* 2018, **360**:176–182 <http://dx.doi.org/10.1126/science.aam8999>.
9. Camp JG et al.: **Human cerebral organoids recapitulate gene expression programs of fetal neocortex development.** *Proc Natl Acad Sci U S A* 2015, **112**:15672–15677 <http://dx.doi.org/10.1073/pnas.1520760112>.
10. Quadrato G et al.: **Cell diversity and network dynamics in photosensitive human brain organoids.** *Nature* 2017, **545**:48–53 <http://dx.doi.org/10.1038/nature22047>.
- This was one of the first explorations of cell diversity in human iPSC-derived organoids using high-throughput single-cell transcriptomics (drop-seq protocol). The higher throughput method enabled comparison of cell composition across organoid batches and time points.
11. Renner M et al.: **Self-organized developmental patterning and differentiation in cerebral organoids.** *EMBO J* 2017, **36**:1316–1329 <http://dx.doi.org/10.1525/embj.201694700>.
12. Birey F et al.: **Assembly of functionally integrated human forebrain spheroids.** *Nature* 2017, **545**:54–59 <http://dx.doi.org/10.1038/nature22330>.
13. Pollen AA et al.: **Molecular identity of human outer radial glia during cortical development.** *Cell* 2015, **163**:55–67 <http://dx.doi.org/10.1016/j.cell.2015.09.004>.
14. Fuzik J et al.: **Integration of electrophysiological recordings with single-cell RNA-seq data identifies neuronal subtypes.** *Nat Biotechnol* 2016, **34**:175–183 <http://dx.doi.org/10.1038/nbt.3443>.
15. Sato T et al.: **Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche.** *Nature* 2009, **459**:262–265 <http://dx.doi.org/10.1038/nature07935>.
16. Stange DE et al.: **Differentiated Troy+ chief cells act as reserve stem cells to generate all lineages of the stomach epithelium.** *Cell* 2013, **155**:357–368 <http://dx.doi.org/10.1016/j.cell.2013.09.008>.
17. Spence JR et al.: **Directed differentiation of human pluripotent stem cells into intestinal tissue in vitro.** *Nature* 2011, **470**:105–109 <http://dx.doi.org/10.1038/nature09691>.
18. McCracken KW et al.: **Modelling human development and disease in pluripotent stem-cell-derived gastric organoids.** *Nature* 2014, **516**:400–404 <http://dx.doi.org/10.1038/nature13863>.
19. Grun D et al.: **Single-cell messenger RNA sequencing reveals rare intestinal cell types.** *Nature* 2015, **525**:251–255 <http://dx.doi.org/10.1038/nature14966>.
20. Haber AL et al.: **A single-cell survey of the small intestinal epithelium.** *Nature* 2017, **551**:333–339 <http://dx.doi.org/10.1038/nature24489>.
- This manuscript used scRNA-seq to first establish an atlas of cell diversity in the mouse small intestinal epithelium and 3D organoid cultures derived from adult stem cells, and then used scRNA-seq to analyze how each cell type within the intestinal organoid cultures respond to different intestinal microbes. Higher throughput methods enabled comprehensive comparisons of organoids in different environmental conditions.
21. Mallanna SK, Duncan SA: **Differentiation of hepatocytes from pluripotent stem cells.** *Curr Protoc Stem Cell Biol* 2013, **26** <http://dx.doi.org/10.1002/9780470151808.sc01gs26> Unit 1G 4.
22. Cahan P et al.: **CellNet: network biology applied to stem cell engineering.** *Cell* 2014, **158**:903–915 <http://dx.doi.org/10.1016/j.cell.2014.07.020>.
23. Takebe T et al.: **Vascularized and functional human liver from an iPSC-derived organ bud transplant.** *Nature* 2013, **499**:481–484 <http://dx.doi.org/10.1038/nature12271>.
24. Camp JG et al.: **Multilineage communication regulates human liver bud development from pluripotency.** *Nature* 2017, **546**:533–538 <http://dx.doi.org/10.1038/nature22796>.
25. Kegel V et al.: **Protocol for isolation of primary human hepatocytes and corresponding major populations of non-parenchymal liver cells.** *J Vis Exp* 2016:e53069 <http://dx.doi.org/10.3791/53069>.
26. Huch M et al.: **Long-term culture of genome-stable bipotent stem cells from adult human liver.** *Cell* 2015, **160**:299–312 <http://dx.doi.org/10.1016/j.cell.2014.11.050>.
27. Halpern KB et al.: **Single-cell spatial reconstruction reveals global division of labour in the mammalian liver.** *Nature* 2017, **542**:352–356 <http://dx.doi.org/10.1038/nature21065>.
28. Takasato M et al.: **Kidney organoids from human iPS cells contain multiple lineages and model human nephrogenesis.** *Nature* 2015, **526**:564–568 <http://dx.doi.org/10.1038/nature15695>.
29. Morizane R et al.: **Nephron organoids derived from human pluripotent stem cells model kidney development and injury.** *Nat Biotechnol* 2015, **33**:1193–1200 <http://dx.doi.org/10.1038/nbt.3392>.
30. Combes AN et al.: **High throughput single cell RNA-seq of developing mouse kidney and human kidney organoids reveals a roadmap for recreating the kidney.** *bioRxiv*. 2017 <http://dx.doi.org/10.1101/235499>.
31. Wu H et al.: **Comparative analysis of kidney organoid and adult human kidney single cell and single nucleus transcriptomes.** *bioRxiv*. 2017 <http://dx.doi.org/10.1101/232561>.
- This manuscript quantitatively compared cell composition, similarity to primary tissue, and off-target cell type emergence in two different human kidney organoid protocols derived from iPSCs.
32. Park J et al.: **Single-cell transcriptomics of the mouse kidney reveals potential cellular targets of kidney disease.** *Science* 2018, **360**:758–763 <http://dx.doi.org/10.1126/science.aar2131>.
33. Lindstrom NO et al.: **Progressive recruitment of mesenchymal progenitors reveals a time-dependent process of cell fate acquisition in mouse and human nephrogenesis.** *Dev Cell* 2018, **45**:651–660 <http://dx.doi.org/10.1016/j.devcel.2018.05.010> e654.
34. Czerniecki SM et al.: **High-throughput screening enhances kidney organoid differentiation from human pluripotent stem cells and enables automated multidimensional phenotyping.** *Cell Stem Cell* 2018, **22**:929–940 <http://dx.doi.org/10.1016/j.stem.2018.04.022> e924.
35. Lein E, Borm LE, Linnarsson S: **The promise of spatial transcriptomics for neuroscience in the era of molecular cell typing.** *Science* 2017, **358**:64–69 <http://dx.doi.org/10.1126/science.aan6827>.
36. Spanjaard B et al.: **Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars.** *Nat Biotechnol* 2018, **36**:469–473 <http://dx.doi.org/10.1038/nbt.4124>.
37. Stoeckius M et al.: **Cell “hashing” with barcoded antibodies enables multiplexing and doublet detection for single cell genomics.** *bioRxiv*. 2017 <http://dx.doi.org/10.1101/237693>.
38. Cao J et al.: **Comprehensive single-cell transcriptional profiling of a multicellular organism.** *Science* 2017, **357**:661–667 <http://dx.doi.org/10.1126/science.aam8940>.
- This paper described a combinatorial barcoding approach to generate single-cell transcriptomes without first compartmentalizing the cell into a well or droplet. This strategy could in principle be used to scale up the number of conditions (genetic, environmental, or temporal) that can be analyzed in organoids by lowering the cost and also because the method can be performed in 96 well plates, a convenient platform for some organoid culture protocols.
39. Cleary B, Cong L, Cheung A, Lander ES, Regev A: **Efficient generation of transcriptomic profiles by random composite measurements.** *Cell* 2017, **171**:1424–1436 <http://dx.doi.org/10.1016/j.cell.2017.10.023> e1418.