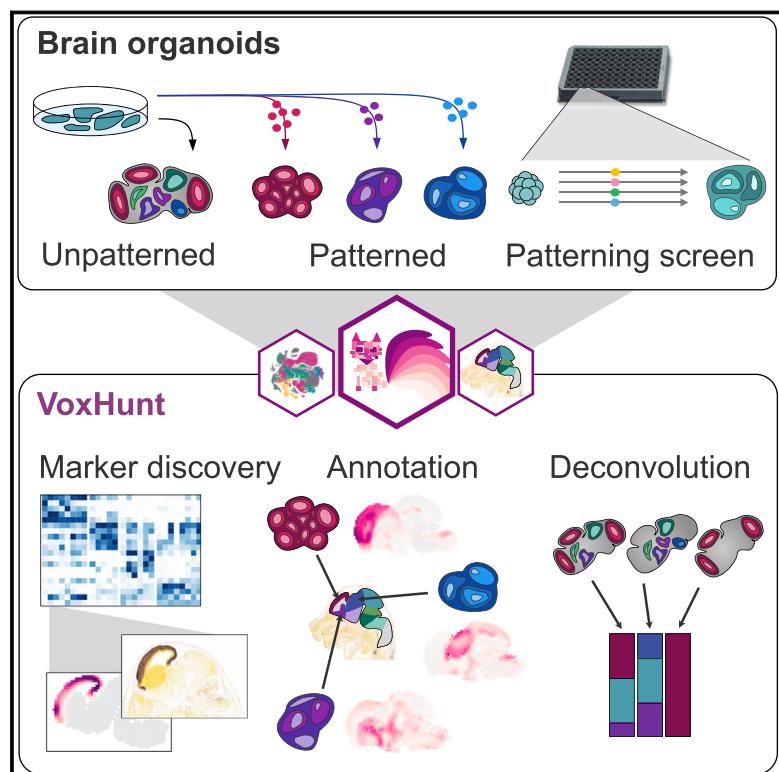


# Cell Stem Cell

## Resource

### Resolving organoid brain region identities by mapping single-cell genomic data to reference atlases

#### Graphical Abstract



#### Authors

Jonas Simon Fleck,  
Fátima Sanchís-Calleja, Zhisong He,  
Małgorzata Santel,  
Michael James Boyle, J. Gray Camp,  
Barbara Treutlein

#### Correspondence

grayson.camp@iob.ch (J.G.C.),  
barbara.treutlein@bsse.ethz.ch (B.T.)

#### In brief

Organoid tissues resembling parts of the human brain can be grown from stem cells. Treutlein, Camp, and colleagues develop a computational toolkit to explore the types of cells that are present in brain organoids by comparing gene features of individual organoid cells to primary brain atlases.

#### Highlights

- Computational toolkit to explore and visualize Allen Brain Atlas data
- Annotation of brain organoid single-cell genomic data via reference atlas queries
- Bulk transcriptome deconvolution through spatial brain map comparisons
- Multiplexed patterning screens to dissect the effect of morphogens



## Resource

# Resolving organoid brain region identities by mapping single-cell genomic data to reference atlases

Jonas Simon Fleck,<sup>1,5</sup> Fátima Sanchís-Calleja,<sup>1,5</sup> Zhisong He,<sup>1</sup> Małgorzata Santel,<sup>1</sup> Michael James Boyle,<sup>2</sup> J. Gray Camp,<sup>3,4,\*</sup> and Barbara Treutlein<sup>1,6,\*</sup>

<sup>1</sup>Department of Biosystems Science and Engineering, ETH Zürich, 4058 Basel, Switzerland

<sup>2</sup>Max Planck Institute for Evolutionary Anthropology, Leipzig, 04103 Sachsen, Germany

<sup>3</sup>Institute of Molecular and Clinical Ophthalmology Basel, 4031 Basel, Switzerland

<sup>4</sup>Department of Ophthalmology, University of Basel, 4031 Basel, Switzerland

<sup>5</sup>These authors contributed equally

<sup>6</sup>Lead contact

\*Correspondence: grayson.camp@iob.ch (J.G.C.), barbara.treutlein@bsse.ethz.ch (B.T.)

<https://doi.org/10.1016/j.stem.2021.02.015>

## SUMMARY

Self-organizing tissues resembling brain structures generated from human stem cells offer exciting possibilities to study human brain development, disease, and evolution. These 3D models are complex and can contain cells at various stages of differentiation from different brain regions. Single-cell genomic methods provide powerful approaches to explore cell composition, differentiation trajectories, and genetic perturbations in brain organoid systems. However, it remains a major challenge to understand the heterogeneity observed within and between individual organoids. Here, we develop a set of computational tools (VoxHunt) to assess brain organoid patterning, developmental state, and cell identity through comparisons to spatial and single-cell transcriptome reference datasets. We use VoxHunt to characterize and visualize cell compositions using single-cell and bulk genomic data from multiple organoid protocols modeling different brain structures. VoxHunt will be useful to assess organoid engineering protocols and to annotate cell fates that emerge in organoids during genetic and environmental perturbation experiments.

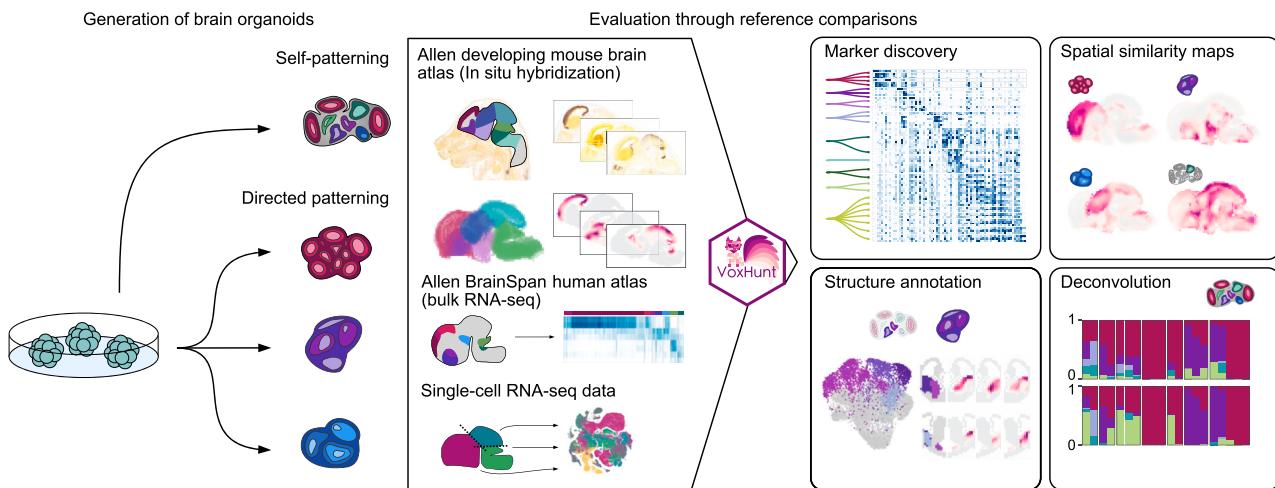
## INTRODUCTION

Spatiotemporal patterning events occurring during brain development lead to the formation of multiple forebrain, midbrain, and hindbrain structures composed of a diversity of cell types. Portions of this structural and cellular diversity can be recapitulated in stem-cell-derived neuronal tissues that self-organize in 3D culture *in vitro* (Lancaster and Knoblich, 2014). Over the past few years, the generation of human brain organoids and spheroids has enabled unprecedented analyses of human brain development and disease (reviewed in Arlotta, 2018; Giandomenico and Lancaster, 2017; Paşa, 2018), as well as evolution (Kanton et al., 2019; Mora-Bermúdez et al., 2016; Pollen et al., 2019). There are two general strategies for producing 3D human brain tissue cultures. In the first strategy, pluripotent stem cells are guided to form a layer of neuroepithelial stem cells that can differentiate into neurons. Culturing the neuroepithelium in permissive media can result in the generation of multiple brain structures within the same organoid. This self-patterning strategy (Kadoshima et al., 2013; Lancaster et al., 2013) offers the potential to understand how different parts of the brain self-organize and interact. However, there are often substantial differences in cell-type composition between individual organoids and between batches grown separately. The alternative

strategy is to use signaling molecules or other directive cues to control patterning of the neuroepithelium so that a defined structure forms, such as the cortex (Eiraku et al., 2008; Kadoshima et al., 2013; Paşa et al., 2015), ventral telencephalon (Bagley et al., 2017; Birey et al., 2017; Xiang et al., 2017), hypothalamus (Qian et al., 2016), and thalamus (Xiang et al., 2018). Multiple protocols have been published to generate modular brain structures and create higher-level interactions by fusing two or more brain structures to study neuronal migration or enable formation of inter-region neuronal connections (Bagley et al., 2017; Birey et al., 2017; Duan et al., 2019; Paşa, 2018). This technique might increase reproducibility of structure formation, but researchers have yet to define all of the signals that create each brain structure. Moreover, it is unclear whether certain parts of the brain can form in the absence of adjacent structures. Despite incredible progress in human brain organoid engineering, it is not yet possible to create a stereotypic organoid containing multiple brain structures.

Organoid protocol development requires robust strategies to assess organoid composition. Single-cell RNA sequencing (scRNA-seq) has emerged as a high-resolution approach to deconstruct organoid cell composition and reconstruct differentiation trajectories in these complex developing tissues (Camp et al., 2015; Kanton et al., 2019; Quadrato et al., 2017; Velasco





**Figure 1. Overview of the VoxHunt computational toolkit for exploring heterogeneity in the developing brain and brain organoid counterparts**

Brain organoids generated through various protocols or patterning regimens can give rise to cells resembling different brain structures. VoxHunt helps interpret transcriptomic or epigenomic data from these organoids by providing an interface to relevant reference datasets. VoxHunt implements functions to extract brain structure markers, compute similarity maps of organoid cells, assign regional identities, and deconvolute bulk RNA-seq data into structure proportions.

et al., 2019). Multiple labs have used scRNA-seq to assess heterogeneity in brain organoids, and direct comparison between engineered *in vitro* cortical structures and primary *in vivo* counterparts has revealed remarkable similarity in gene expression profiles (Camp et al., 2015; Pollen et al., 2019). However, it still remains a major challenge to understand and interpret gene expression and regulatory patterns from organoid single-cell genomic data. Here, we have developed a computational toolbox to assess brain organoid patterning, developmental state, and cell identity through systematic comparisons to annotated reference datasets (Figure 1). Primarily, we utilize *in situ* hybridization data from the Allen Developing Mouse Brain Atlas (ABA) as a well-annotated spatial gene expression map. Based on their transcriptome or accessible chromatin landscape, cells can be mapped to voxels (in this case, 3D positions rendered from spatial transcript patterns), which enables characterization and visualization of organoid regional composition. In addition, we provide an interface to perform comparisons with bulk RNA-sequencing (RNA-seq) data from human microdissected brain structures (BrainSpan; Thompson et al., 2014) as well as to single-cell RNA-seq data from the developing mouse brain (La Manno et al., 2020). Finally, we show that VoxHunt can help deconvolute bulk organoid transcriptome data as well as assess brain region composition in organoids grown under different conditions. Altogether, VoxHunt will be useful to assess organoid engineering protocols and to rapidly annotate cell fates that emerge in organoids during genetic and environmental perturbation experiments.

## RESULTS

### Spatiotemporal voxel maps of the developing mouse brain

To facilitate exploration of 3D gene expression patterns in the developing brain, we obtained *in situ* hybridization data of the mouse brain in eight different developmental stages (Figure 2A)

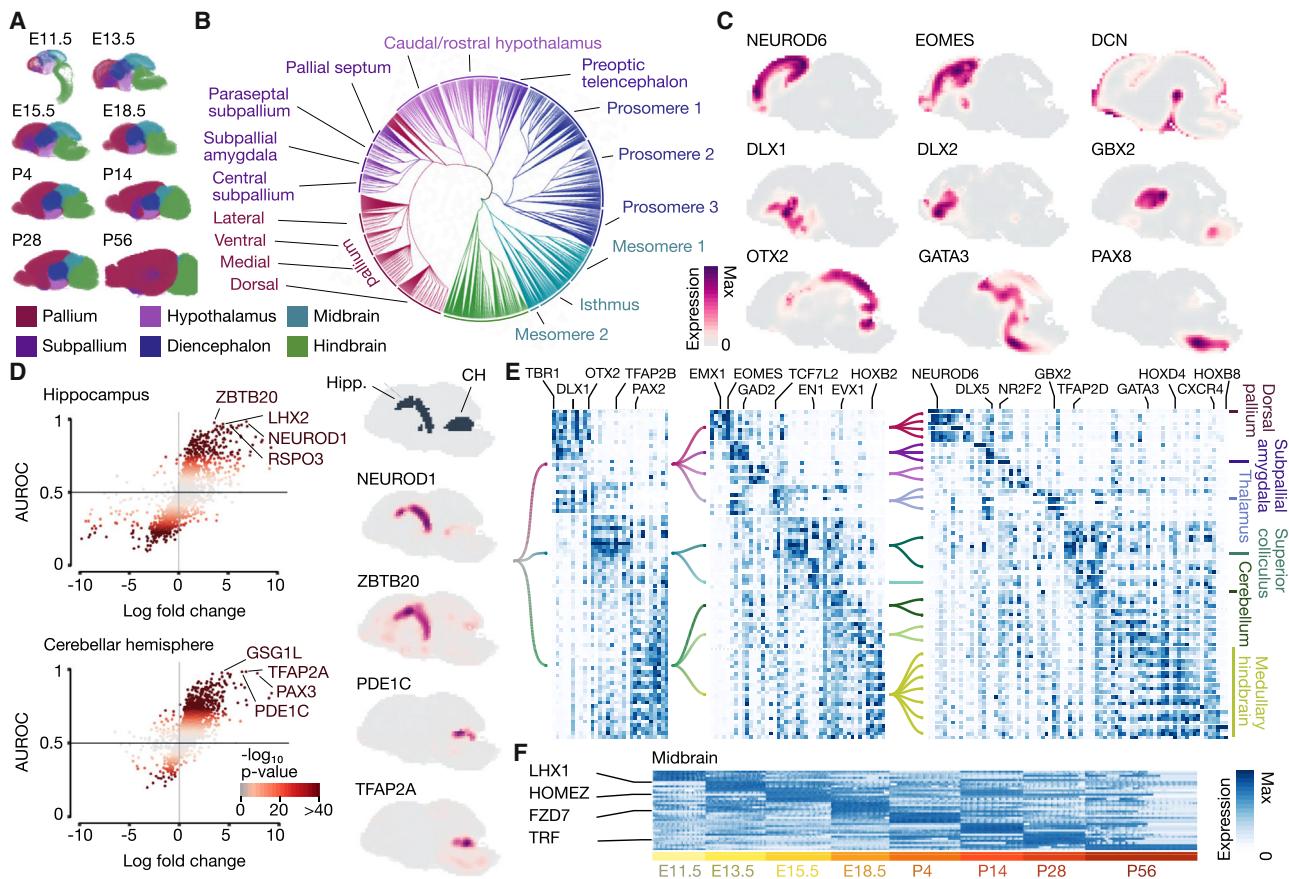
from the Allen Developing Mouse Brain Atlas (Thompson et al., 2014). This dataset is currently the most comprehensive spatially registered four-dimensional (4D) reference atlas of a mammalian brain. The dataset provides three-dimensional spatial expression patterns from different experiments that measure a total of 2,073 genes over time points ranging from mouse embryonic day 11.5 (E11.5) through postnatal day 56. Each expression quantification is registered to a single voxel grid for each time point (80–200  $\mu\text{m}$  resolution) (Figure S1A). Each voxel is further annotated with a brain structure derived from a hierarchical structure ontology. We preprocessed these data by summarizing *in situ* hybridization quantifications from different experiments measuring the same gene to obtain a single gene expression vector for each voxel at each developmental stage. To allow for comparison of structures with similar size throughout developmental stages, we further introduced additional layers of structure annotation by aggregating existing ontological levels, resulting in spatially and temporally resolved brain structures (Figures 2B and S1B).

### Brain structure marker identification

We performed differential expression analysis between annotated brain structures and used the single-feature area under the receiver operating characteristic curve (AUROC) as a metric for selecting structure-specific genes with spatially confined expression (Figures 2C, 2D, and S1C). In this way, one can perform large-scale feature selection based on various criteria such as structure or stage specificity (Figures 2E, 2F, and S1D). These brain-structure-enriched gene expression profiles could serve as a resource of marker genes to validate cell-fate engineering and organoid protocols.

### VoxHunt accurately maps reference populations

We first validated that the mouse *in situ* hybridization (ISH) reference atlas is a suitable resource for cell annotation of human single-cell transcriptome data. We used VoxHunt to map single-cell



**Figure 2. Feature extraction from voxel maps of developing mouse brain *in situ* hybridization data**

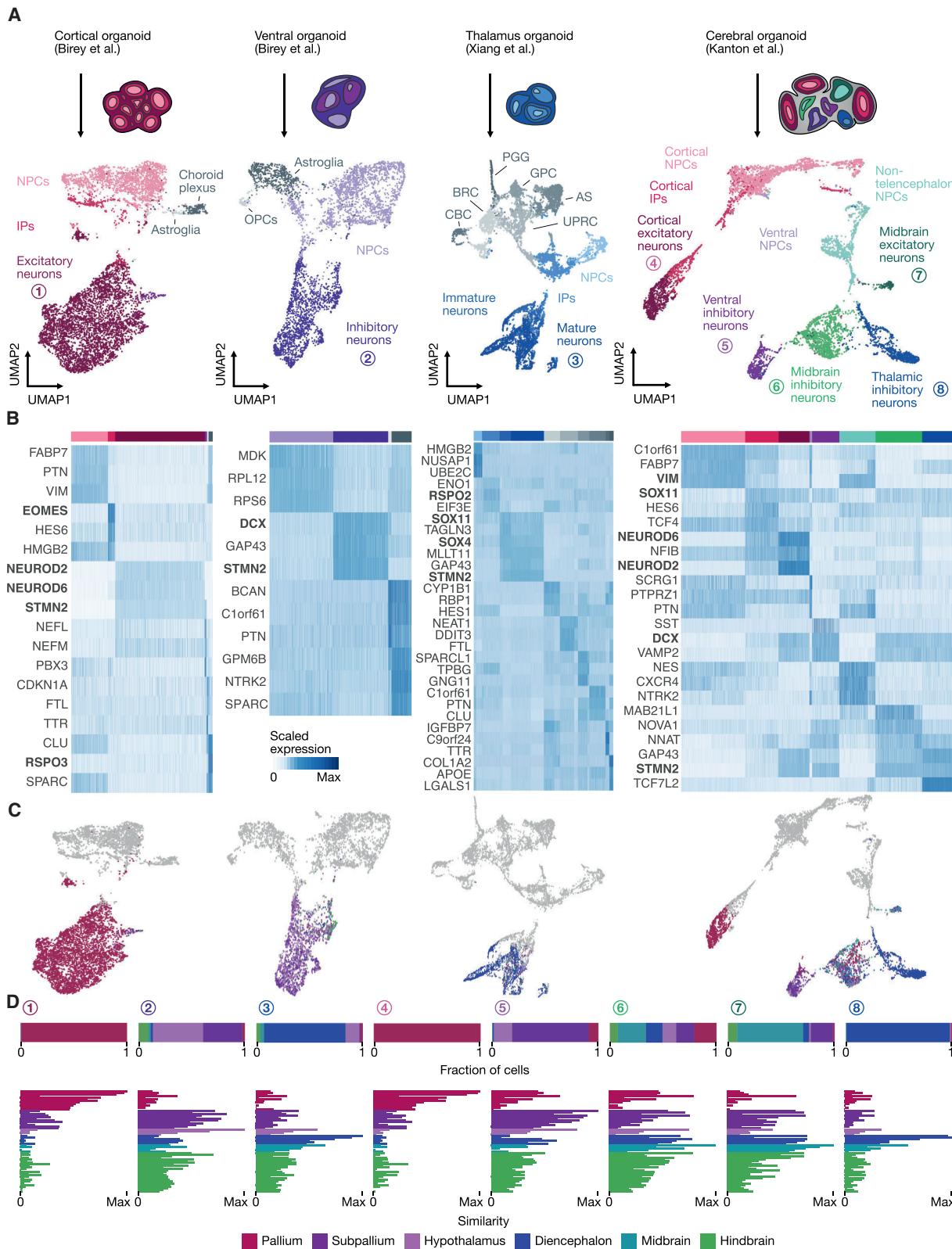
- (A) Schematics show mouse brain development over time, with high-level structure annotations color-coded.
- (B) Dendrogram showing the hierarchically organized structure ontology from the Allen Developing Mouse Brain Atlas.
- (C) Spatial expression patterns of selected markers as maximum intensity projections across five central sagittal sections (8–12) in the E15.5 mouse brain.
- (D) Scatterplot showing feature selection for two brain structures, hippocampus (Hipp.) and cerebellar hemisphere (CH), at E15.5, together with the expression profile for the top two marker genes per structure.
- (E and F) Heatmap showing marker gene expression across brain structures at different levels of structure resolution at stage E15.5 (E) and across developmental stages for the midbrain (F).

transcriptome data (split-pool barcoding) from the developing mouse brain (Rosenberg et al., 2018), which resulted in mouse-to-mouse mapping of annotated cell populations to voxelized structures (Figure S2A). Using this dataset, we assessed the accuracy (fraction of cells with correctly assigned structures) and contrast (similarity to second best structure/similarity to assigned structure) of the spatial mapping under different conditions for feature selection (e.g., differentially expressed genes between structures versus most variable features) and tested the stability of structure assignments through downsampling of unique molecular identifier (UMI) in the single-cell RNA-seq data (Figures S2B–S2D). We found that accuracy was proportional to the number of features used, and this result was consistent across brain structures. However, for certain structures (e.g., pallium), contrast decreased at low or high feature inclusion. Accuracy has some dependency on UMI count, with certain structures (e.g., rostral midbrain) being more sensitive to this aspect of the scRNA-seq data. Nonetheless, even with the relatively sparse split-pool barcoding data, we could identify

optimal parameters that facilitated mapping with high accuracy and contrast. We then compared bulk RNA-seq data from microdissected human brain structures from the BrainSpan Developing Human Brain Atlas (Thompson et al., 2014; Kanton et al., 2019) (Figures S2B and S2C), as well as four scRNA-seq datasets from primary human brain tissues (Fan et al., 2018; Nowakowski et al., 2017; Polioudakis et al., 2019; Zhong et al., 2018) (Figure S3). We found that each of the datasets exhibited consistently clear similarity patterns, supporting the use of mouse 4D4D gene expression data to match homologous brain regional data and predict properties of human cell types.

#### Organoid region composition annotation

We next tested if VoxHunt could be used for unsupervised annotation of cell types generated through different organoid protocols. For this, we obtained published single-cell transcriptome data from three brain-structure-specific organoid protocols (thalamus, cortex, and ventral telencephalon) (Birey et al., 2017; Xiang et al., 2018) as well as one cerebral (unpatterned)



**Figure 3. Single-cell transcriptomes deconstruct region composition in brain organoids**

(A) UMAP projections of published single-cell transcriptome data from different organoid protocols, including cortical (Birey et al., 2017), ventral (Birey et al., 2017), thalamus (Xiang et al., 2018), and cerebral (Kanton et al., 2019). Cells are colored based on the annotations in the original publications.

(legend continued on next page)

organoid protocol (Kanton et al., 2019). We used Uniform Manifold Approximation and Projection (UMAP) (Becht et al., 2018) embeddings to project cells onto a two-dimensional space for each protocol and annotated the clusters as described in the original studies (Figure 3A). All four datasets exhibit a diversity of cell populations, including heterogeneous progenitor and neuronal cells. For neuronal cells from structure-specific organoids, successful patterning is supported by the expression of canonical marker genes (Figure 3B). For cerebral organoids, multiple brain structures can form in any given organoid, and marker gene expression analysis suggests the emergence of neuronal cell types from the telencephalon (pallium and subpallium), diencephalon, midbrain, and hindbrain, supporting the annotations reported in the original publication (Figure 3B). We assessed if annotations from each dataset could be confirmed by correlating single-cell transcriptomes from each neuronal cluster with *in situ* expression patterns in the mammalian brain at high-level annotations that are consistent across time for telencephalon (pallium and subpallium), diencephalon, midbrain, and hindbrain (Figures 3C and 3D). Cells annotated as cortical neurons derived from both cortical or cerebral organoid protocols showed far higher correlation to pallium than to any other structure. Thalamic neurons showed a similarly clear pattern, with the majority (>90%) of cells having the highest correlation to voxels within the diencephalon. For clusters of inhibitory neurons annotated as ventral telencephalon, we observed differences between the two organoid protocols. While cells derived from cerebral organoids had consistently higher correlation with subpallial structures, a large fraction of cells (48.3%) from ventrally patterned organoids were most highly correlated with the hypothalamus. Cell clusters annotated as midbrain neurons derived from cerebral organoid protocols showed a noisier correlation pattern and overall lower correlation to *in situ* maps than any of the other clusters. While average correlation within each cluster was highest to the midbrain, correlation of single cells did not show clear agreement, especially for cluster 6, annotated as “midbrain excitatory neurons.”

#### Spatial projection of organoid cell transcriptomes

To select a suitable developmental stage for voxel-resolved visualization of spatial similarity patterns, we computed the correlation of each cell in a cluster to its matching structure from all fetal developmental stages (E11.5–E18.5) (Figures 4A and S4A). From the previously decomposed clusters, the majority showed the highest median correlation to E13.5, with the exception of thalamic neuron clusters, which were most highly correlated with structures at E11.5 (Figure 4B). For cortical neurons, we further found that their most similar developmental stage was proportional to the age of their organoid of origin (Figures S4B and S2C). For each organoid cluster, we visualized the correlation to each voxel in either a sagittal view (Figures 4C and S4A)

or two-dimensional coronal (Figure 4D) sections. To further aid the assessment of cell-type compositions of individual organoids, each cell from each organoid can be projected to spatial locations within the brain based on maximum similarity to loci within the voxel maps (Figure 4E). These analyses reveal the spatial location within the brain tissue where each cluster or cell has maximal correlation, which can be used to assess the specificity of the regional neuronal population and help annotate previously unknown cell populations. Similarity patterns can also be explored in three-dimensional registrations of the voxel maps (Videos S1, S2, S3, S4, and S5).

#### Complementary reference datasets as high-dimensional search spaces

To test whether the structure annotations obtained from mapping to ISH data could be further enhanced through other reference datasets, we compared organoid-derived neurons to time course bulk RNA-seq data from the human brain (BrainSpan) (Thompson et al., 2014) (Figure S4D) and a single-cell transcriptomic atlas of the developing mouse brain (La Manno et al., 2020) (Figure S4E). Similarities of neuronal clusters to BrainSpan samples largely agreed with the comparison to ISH data. However, BrainSpan covers a limited number of brain structures and lacks midbrain samples, which makes it unsuitable for annotating neurons from certain protocols. The single-cell atlas of the developing mouse brain provides coarse (forebrain, midbrain, and hindbrain; Figure S4E) annotations of cell types arising during early mammalian brain development and provides functional annotations that cannot be derived from the ISH data alone. We found that organoid-derived cortical neurons showed the highest similarity to excitatory neurons, while ventral neurons mapped to inhibitory (GABAergic) neurons, which is in line with expectations (Figure S4F). Comprehensive whole-brain scRNA-seq datasets from human and mouse will be valuable references for functionally annotating neuronal and non-neuronal cell types.

#### Assessing region specificity across organoid protocols

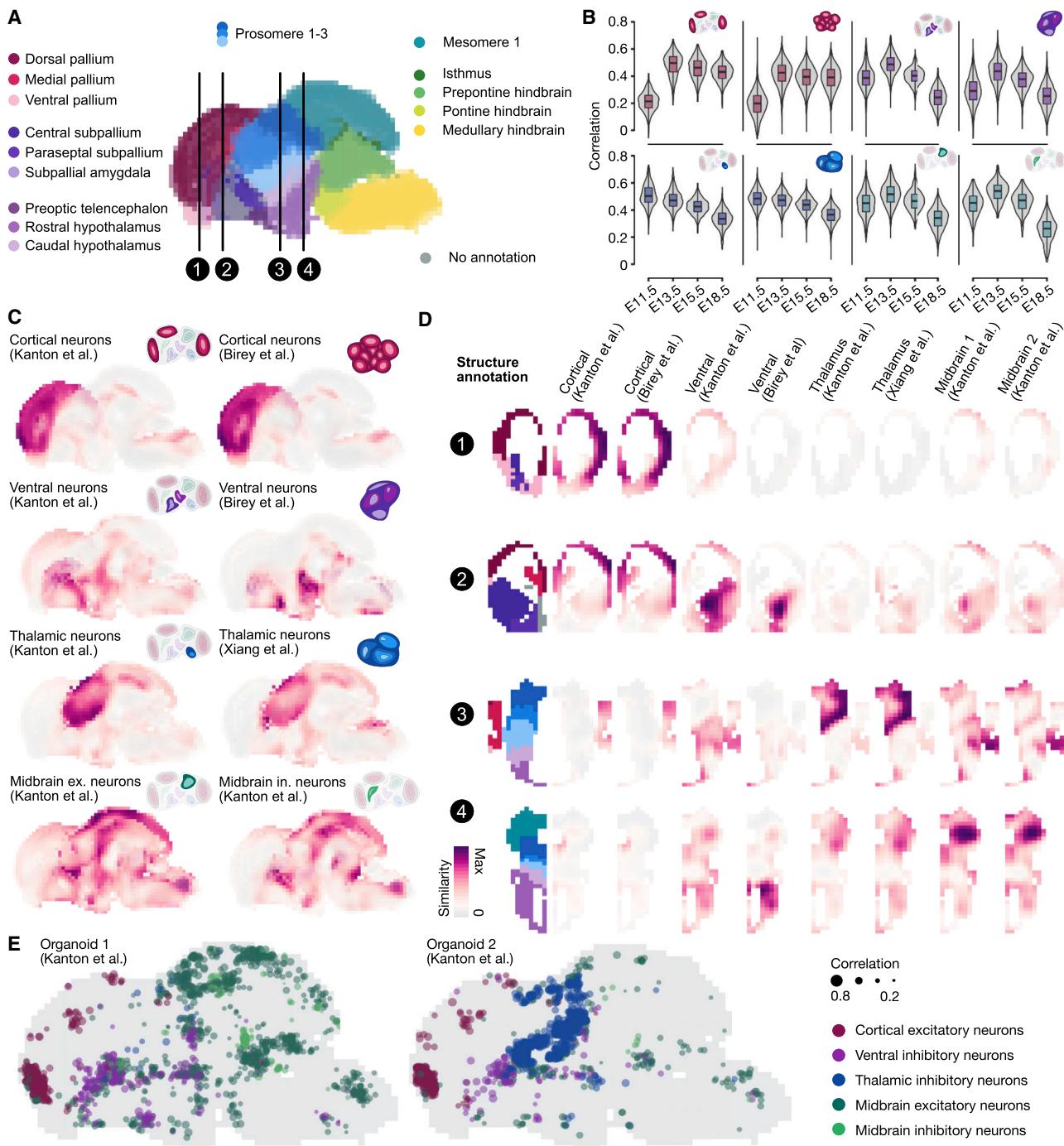
We used VoxHunt to assess cell identities generated in organoids from different protocols. We analyzed scRNA-seq data from nine different organoid protocols (Bhaduri et al., 2020; Birey et al., 2017; Giandomenico et al., 2019; Kanton et al., 2019; Pellegrini et al., 2020; Pollen et al., 2019; Velasco et al., 2019; Xiang et al., 2018) that have been reported to generate diverse neuronal populations. We found that spatial similarity maps and assigned region identities by VoxHunt corresponded well with the original annotations (Figure S5). We found that choroid plexus (ChP) cells from central nervous system barrier-forming organoids (Pellegrini et al., 2020) showed a distinct correlation to the roof plate of the invaginated telencephalic vesicle, a ChP precursor. In some cases, VoxHunt enabled annotation of neuronal cell populations with a previously uncertain identity.

(B) Heatmap showing the expression of top marker genes for each cluster. Canonical marker genes are highlighted in bold.

(C) UMAP projections of neuronal cells from each dataset colored by structure annotation of the voxel with maximum correlation.

(D) Top: fraction of cells from each cluster annotated in (A) that have maximum correlation to each brain structure based on comparison to *in situ* hybridization voxel maps. Bottom: bar plots showing the mean similarity of each cluster to different brain structures. Each bar represents one brain structure at custom annotation level 4.

IP, immediate progenitor; NPC, neural progenitor cell; OPC, oligodendrocyte progenitor cells; CBC, cilia-bearing cells; BRC, BMP-related cells; PGG, proteoglycan-expressing glia; GPC, glial progenitor cell; AS, astrocytes; UPRC, unfolded protein response-related cells.



**Figure 4. Comparing brain organoid cell populations to spatial brain maps**

(A) Structure annotations of the E13.5 mouse brain (sagittal view), which was used for organoid cluster annotation. Locations 1–4 are used for coronal sections shown in subsequent panels.

(B) Violin plots showing the organoid cluster correlation distributions to voxel maps of the previously assigned structure (pallium, subpallium, diencephalon, and mesencephalon) across different developmental time points.

(C) Sagittal projections colored by scaled similarity scores of different organoid clusters from each dataset to voxel maps of the E13.5 mouse brain.

(D) Coronal sections visualizing regional annotation and scaled similarity scores to different brain structures from different clusters and protocols.

(E) Sagittal view of a maximum-correlation projection of cells from two individual cerebral organoids to 8 central sections (5–12) of the E13.5 mouse brain.

For example, we observed that neurons in ChP organoids mapped distinctly to diencephalon structures and appear distinct from neurons generated through a telencephalic organoid protocol (Figure S5I). We also observed substantial brain region composition heterogeneity between individual cerebral organoids consistent with previous reports (Kanton et al., 2019), especially when comparing organoids from different human induced pluripotent stem cell (iPSC) lines (Figure S5J). In comparison, the majority of cells from the dorsal forebrain organoid model match pallial structures across organoids and iPSC lines. We note that we were able to annotate a previously unknown cluster present in 1-month-old cerebral organoids (Camp et al., 2015) (Figures S6A and S6B). This population was marked by R-spondin 2/3 (RSPO2/3) and WNT2B, which localize to the invaginated telencephalic roof plate (Kamata et al., 2004) and transient structures annotated as pallium at the boundary of the developing telencephalon and diencephalon (Figures S6C–S6E).

#### Annotating organoid cell identity at substructure resolution

We were curious whether we could go beyond the structure ontology provided by the ABA to reveal a higher-resolved annotation of organoid cell types. For this, we selected two high-level structures, the dorsal pallium and the subpallium, and performed unbiased clustering on voxel transcriptomes (Figures 5A and 5B). We aggregated these clusters based on the expression of canonical marker genes to reveal neural progenitor cell (NPC), immediate progenitor (IP), and neuronal layers in the dorsal pallium and the lateral, medial, and caudal ganglionic eminences (LGE, MGE, and CGE) in the subpallium. These custom annotations matched the correlation patterns of single-cell transcriptomic data derived from the annotated structures in mice (Loo et al., 2019; Mayer et al., 2018) as well as the expected anatomic position and morphology (Figures 5C and 5D). We found that organoid cells from the cortical trajectory clearly mapped onto the annotated pallial zones (Figure 5E). Moreover, ventral neuron clusters from cerebral organoids could be clearly annotated as LGE-, MGE-, or CGE-like interneurons based on their similarities to the substructures (Figure 5F). We further used these newly annotated subpallial structures to resolve and compare the heterogeneity of inhibitory neurons derived from three organoid protocols (Birey et al., 2017; Kanton et al., 2019; Xiang et al., 2017). We found that all protocols gave rise to diverse neuronal populations, some of which matched distinct subpallial structures (Figures S6G–S6M), while others were more similar to the hypothalamus and did not express any canonical ganglionic eminence (GE) marker genes (Figures S5G and S5H). We note that these results may be confounded by maturation differences, and further analysis would benefit from availability of time-course data from human samples. Nevertheless, these data demonstrate that comparison to spatially resolved reference atlases can elucidate previously unappreciated and biologically meaningful heterogeneity within brain organoids.

#### Spatial mapping of chromatin accessibility data

Next, we show that similar analyses can be performed for single-cell assay for transposase-accessible chromatin using sequencing (scATAC-seq) data. In this case, we obtained paired

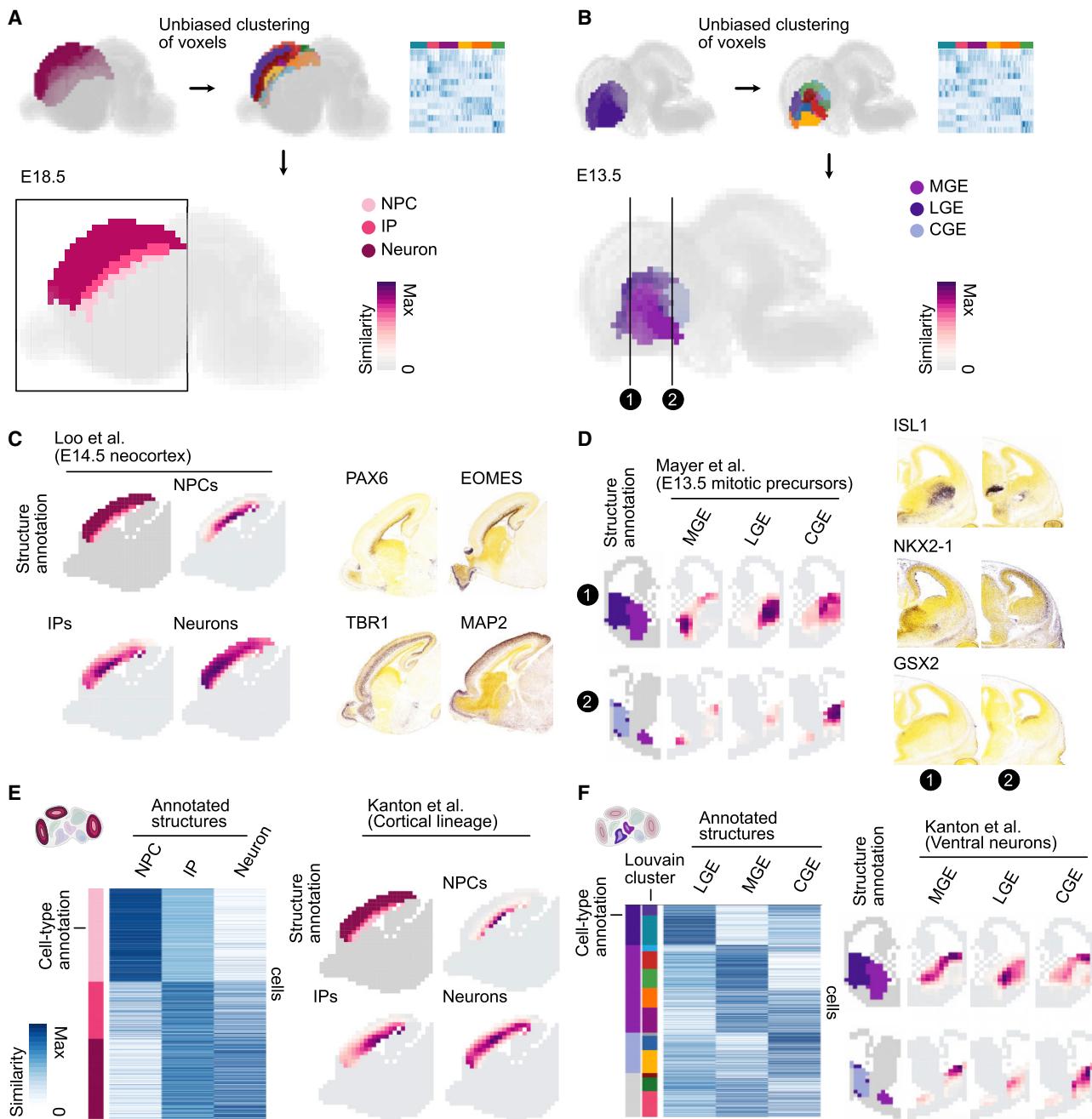
scATAC-seq and scRNA-seq data performed on the same cell suspension generated from microdissected regions from 2- to 4-month cerebral organoids (Kanton et al., 2019) (Figure 6A). We determined accessibility scores at the transcription start site (TSS) of each gene measured in the *in situ* hybridization atlas (Figure 6B) and then calculated TSS access and expression correlation across each structure annotated in the voxel map (Figures 6C and 6D). Coloration of the 3D voxel map based on correlation revealed that both transcriptomes and chromatin accessibility scores of neuronal cells from the microdissected region have the highest correlation with the developing cortex (Figure 6E). This demonstrates that VoxHunt can annotate cells across multiple single-cell-omics modalities.

#### Reference-based deconvolution of organoid bulk transcriptomes

There are ongoing efforts to steer brain organoid development toward distinct regional identities by supplying cues in defined culture conditions. Whole-organoid bulk transcriptome measurements provide a high-information-content metric of organoid states at relative low cost and high throughput. Therefore, we established a framework within VoxHunt to assess organoid brain structure representation using bulk transcriptomics. The pseudo-bulk data can be accurately deconvoluted into proportions of brain regions that are represented within heterogeneous cerebral organoids, and this is further validated with data from directed patterning into cortex or thalamus (Figures S7A and S7B). We note that this approach is particularly powerful at organoid stages where neurons are abundant, as neurons harbor strong structure-specific multi-gene signatures.

#### Assessing organoid patterning using multiplexed transcriptomics and reference mapping

Finally, we highlight how reference-mapping organoid transcriptomes can be used to assess the outcome of culture condition manipulations. We performed a proof-of-principle experiment in which we incubated developing organoids with several different patterning molecules including R-spondin 2 (RSPO2), R-spondin 3 (RSPO3), sonic hedgehog (SHH), an activator of the Wnt pathway (CHIR99021), and two small molecules for dual SMAD inhibition (SB431542 and dorsomorphin) (Figures 7A and S7C). We generated organoids in a multi-well plate and performed bulk RNA-seq of multiple individual organoids after transient treatment with two concentrations of each of these molecules, as well as control untreated organoids, resulting in 61 total bulk transcriptome measurements. Principal-component analysis (PCA) showed transcriptome heterogeneity induced by morphogen treatment that were largely consistent across organoids for a given condition (Figure 7B). Next, we correlated each bulk transcriptome to each voxel within the ABA developing mouse brain reference (E11.5) and found variation between conditions that was consistent across organoids for a given condition (Figure 7C). Interestingly, we found that RSPO2 and RSPO3 conditions had similar and reproducible patterns with enriched mapping to a particular location within the developing brain. As noted above, RSPO2/3 expression localizes to the invaginated telencephalic roof plate (Kamata et al., 2004) and transient structures annotated as pallium at the



**Figure 5. Substructure resolution of transcriptomic maps**

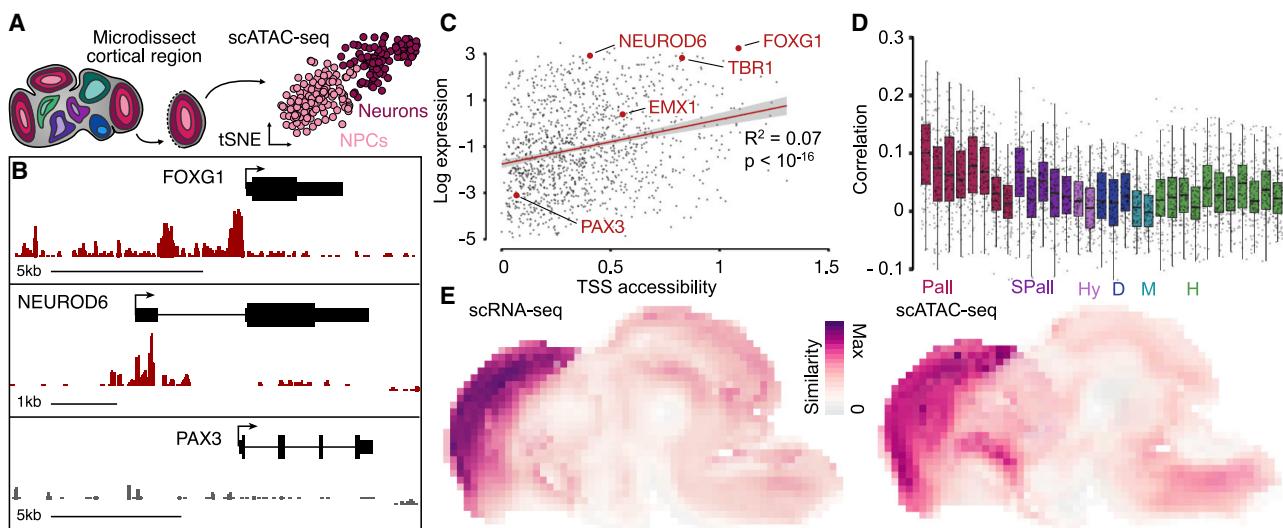
(A and B) Schematic of substructure annotation. Unbiased clusters of voxels were annotated based on the expression of canonical marker genes to identify fine structures beyond existing annotations in the dorsal (A) and ventral (B) telencephalon.

(C) Left: sagittal sections visualizing custom substructure annotations and scaled similarity scores from E14.5 mouse neocortex NPC, IP, and neuron primary scRNA-seq data to voxel maps of the E18.5 mouse brain. Right: ISH images of marker gene expression in shown sections.

(D) Left: coronal sections visualizing custom substructure annotations and scaled similarity scores from E13.5 mouse mitotic precursors from LGE, MGE, and CGE primary scRNA-seq data (Mayer et al., 2018) to voxel maps of the E13.5 mouse brain. Right: E13.5 mouse ISH images of marker gene expression in shown sections.

(E) Left: heatmap showing similarities of cortical cells from cerebral organoids (Kanton et al., 2019) to annotated substructures. Right: coronal sections visualizing custom annotations and scaled similarity scores from organoid cell states to voxel maps of the E13.5 mouse brain. LGE, MGE, and CGE, lateral, medial, and caudal ganglionic eminence; NPC, neural progenitor cell; IP, immediate progenitor.

(F) Left: heatmap showing similarities of ventral neurons from cerebral organoids (Kanton et al., 2019) to annotated substructures and their resulting annotations. Right: coronal sections visualizing custom annotations and scaled similarity scores from organoid ventral neuron populations to voxel maps of the E13.5 mouse brain.



**Figure 6. Projecting chromatin accessibility profiles to spatial brain maps**

(A) Single-cell RNA-seq (Camp et al., 2015; Kanton et al., 2019) and single-cell ATAC-seq (Kanton et al., 2019) was performed on single cortical regions microdissected from cerebral organoids between 2 and 4 months of age.

(B) Signal tracks showing accessibility (normalized read count) at the transcription start site (TSS) of the forebrain marker FOXG1, the cortex neuron marker NEUROD6, and the hindbrain marker PAX3.

(C) Expression of all measured genes in the E13.5 pallium (cortex) as a function of TSS accessibility in cortical organoid cells. The fit of a linear model is indicated in red, the gray area shows the standard error, and the indicated p-value was obtained through an F-test.

(D) Boxplots showing Pearson correlation distributions of organoid TSS accessibility (log-normalized peak count) to gene expression in different brain structures.

(E) Sagittal projections of the E13.5 mouse brain colored by scaled similarity scores of organoid TSS accessibility (scATAC-seq) or expression (scRNA-seq) with voxel map expression of the downstream gene.

Pall, pallium; SPall, subpallium; Hy, hypothalamus; D, diencephalon; M, midbrain; H, hindbrain.

boundary of the developing telencephalon and diencephalon (Figures S6D and S6E). The RSPO2/3-induced signature in organoids was distinct from control organoids, as assessed through PCA and differential expression analysis (Figures 7B, 7D, and 7E). The locations with highest mapping of the RSPO2/3-treated organoids were voxels adjacent to where both of these genes are highly and specifically expressed in the developing brain (Figures 7F and 7H). In addition, many genes that are differentially expressed between control and RSPO2/3-treated organoids are also expressed in voxels adjacent to the RSPO2/3-positive location, including the ChP markers TTR, CLU, and HTR2C (Figures 7F–7H). It is unclear how RSPO2/3 might induce ChP formation, and known R-spondin receptors such as leucine-rich-repeat-containing G-protein-coupled receptors (LGRs) are not probed for in the ABA. Nonetheless, these data provide a potential hint into how local morphogen expression within the developing brain might impact the development of cell phenotypes within neighboring structures. Altogether, this experiment provides an experimental and computational strategy for modulating and assessing developing brain organoids in high-throughput through transcriptome sequencing and reference atlas comparison.

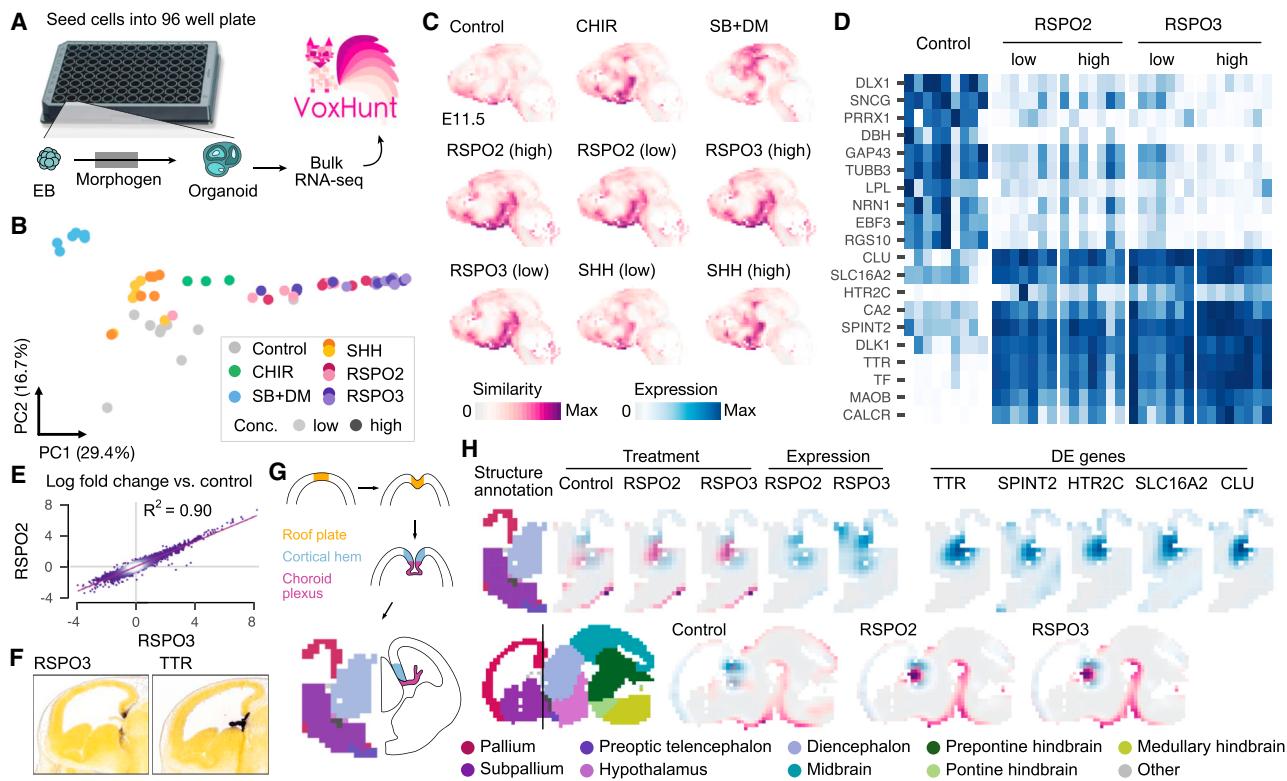
## DISCUSSION

Here, we have developed a set of computational tools that enable researchers to quickly and flexibly explore the vast amount of data within the ABA and other brain reference

resources. We provide multiple examples for how these tools can be implemented to interpret single-cell genomic data from 3D brain tissue cultures through systematic comparison to these references. There are still many uncertainties about how to effectively generate brain structures from humans and other species *in vitro*. VoxHunt (<https://github.com/quadbiolab/voxhunt>) will be a powerful approach to assess organoid engineering protocols, annotate cell fates that emerge in organoids during genetic and environmental perturbation experiments, and develop data-driven hypotheses about the mechanisms underlying mammalian brain patterning that can be tested *in vitro*.

## Limitations of study

One major limitation with the current VoxHunt implementation is that the most comprehensive brain transcriptome atlases have been generated using mouse brain tissue. The human brain has evolved features that are different from other mammals and primates, and there are certain scenarios where comparison to non-human tissues may be inadequate. In addition, single-cell sequencing has illuminated a great diversity of molecularly distinct neuronal and non-neuronal states in the brain, which is partially recapitulated in engineered human brain tissues. However, there is not yet a comprehensive spatial reference with single-cell resolution across the entire brain to use a query space for organoid-reference comparisons. High-resolution 4D reference brain maps will serve as valuable resources for unbiased and high-dimensional search spaces, and VoxHunt will be an



**Figure 7. Assessing the effects of morphogens on brain organoids using multiplexed bulk transcriptomics**

- (A) Schematic of the experimental setup. Single developing organoids were exposed to different morphogens in a 96-well plate. Bulk RNA-seq libraries were prepared from each organoid and multiplexed for sequencing.
- (B) First two principal components of organoid transcriptomes. Each point represents one individual organoid and is colored by condition.
- (C) Spatial similarity maps of each condition to the E11.5 developing mouse brain ISH atlas.
- (D) Heatmap showing the expression of differentially expressed genes between control and R-spondin conditions.
- (E) Correlation of transcriptomic effect sizes (log fold change) for RSPO2 and RSPO3 conditions compared to control samples.
- (F) *In situ* hybridization images of RSPO3 and the choroid plexus marker TTR in coronal sections of E13.5 mouse brain.
- (G) Schematic of roof plate invagination and early stages of choroid plexus development.
- (H) Top left: coronal sections visualizing regional annotation and scaled similarity scores overlaid with RSPO3 expression for control and R-spondin conditions. Top right: coronal sections visualizing expression of differentially expressed genes between control and R-spondin conditions. Bottom: sagittal sections visualizing the regional annotation and scaled similarity scores overlaid with RSPO3 expression for control and R-spondin conditions.

important tool to explore and interpret bulk and single-cell sequencing data from brain organoid systems.

## STAR METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - hESC experiments approval
  - hESC line culture and maintenance
  - Brain organoid generation and incubation with patterning molecules

## ● METHOD DETAILS

- Collection of patterned organoids and RNA purification
- Smartseq2 library preparation
- Sequencing of SS2 libraries
- Acquisition and preprocessing of Allen Brain Atlas *in situ* hybridization data
- Visualization of the structure annotation ontology
- Acquisition of organoid single-cell gene expression and chromatin accessibility data
- Acquisition of primary bulk and single-cell gene expression data

## ● QUANTIFICATION AND STATISTICAL ANALYSIS

- Preprocessing of single-cell gene expression data
- Differential expression analysis of ISH and single-cell data
- Correlation of single-cell transcriptomes to brain structures
- Calculation of similarity to developmental stages

- Evaluating the effect of feature selection and down-sampling
- Spatial correlation maps of organoid cells transcriptomes
- Comparison of organoid cells transcriptomes to other reference datasets
- Substructure resolution of transcriptomic maps
- Substructure-resolved annotation of ventral neurons across organoid protocols
- Spatial correlation maps of organoid cells chromatin accessibility profiles
- Preprocessing and analysis of bulk RNA-seq data
- Deconvolution of organoid pseudo-bulk RNA-seq data

#### SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.stem.2021.02.015>.

#### ACKNOWLEDGMENTS

We thank the Treutlein and Camp labs for helpful discussions and Sophie Janzen and Ryoko Okamoto for experimental support. We also want to thank the Kirkeby lab for kindly sharing the Nkx2.1<sup>GFP/w</sup> cell line. This project was made possible in part by the Chan Zuckerberg Initiative DAF (grant CZF2017-173814 to J.G.C. and B.T.), a Silicon Valley Community Foundation advised fund, the European Research Council (Anthropoid-803441 to J.G.C. and Organomics-758877 and Braintime-874606 to B.T.), the Swiss National Science Foundation (project grant 310030\_84795 to J.G.C. and project grant 310030\_192604 to B.T.), and the National Center of Competence in Research, Molecular Systems Engineering. J.S.F. was supported by a Ph.D. fellowship of the Boehringer Ingelheim Fonds.

#### AUTHOR CONTRIBUTIONS

J.S.F. developed the computational tool and performed computational analyses with support from Z.H. and M.J.B. F.S.C. designed and executed the patterning experiment with assistance from M.S. J.S.F., F.S.-C., B.T., and J.G.C. designed the study and wrote the manuscript.

#### DECLARATION OF INTERESTS

B.T. is a member of the *Cell Stem Cell* advisory board.

Received: January 6, 2020

Revised: October 23, 2020

Accepted: February 8, 2021

Published: March 11, 2021; corrected online: March 25, 2021

#### REFERENCES

- Arlotta, P. (2018). Organoids required! A new path to understanding human brain development and disease. *Nat. Methods* **15**, 27–29.
- Baglama, J., Reichel, L., and Lewis, B.W. (2019). irlba: Fast Truncated Singular Value Decomposition and Principal Components Analysis for Large Dense and Sparse Matrices. R package version 2.3.3. <https://CRAN.R-project.org/package=irlba>.
- Bagley, J.A., Reumann, D., Bian, S., Lévi-Strauss, J., and Knoblich, J.A. (2017). Fused cerebral organoids model interactions between brain regions. *Nat. Methods* **14**, 743–751.
- Becht, E., McInnes, L., Healy, J., Dutertre, C.-A., Kwok, I.W.H., Ng, L.G., Girhoux, F., and Newell, E.W. (2018). Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* **37**, 38–44.
- Bhaduri, A., Andrews, M.G., Mancia Leon, W., Jung, D., Shin, D., Allen, D., Jung, D., Schmunk, G., Haeussler, M., Salma, J., et al. (2020). Cell stress in cortical organoids impairs molecular subtype specification. *Nature* **578**, 142–148.
- Birey, F., Andersen, J., Makinson, C.D., Islam, S., Wei, W., Huber, N., Fan, H.C., Metzler, K.R.C., Panagiotakos, G., Thom, N., et al. (2017). Assembly of functionally integrated human forebrain spheroids. *Nature* **545**, 54–59.
- Blondel, V.D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *J. Stat. Mech.* **2008**, P10008.
- Bult, C.J., Blake, J.A., Smith, C.L., Kadin, J.A., and Richardson, J.E. (2019). Mouse Genome Database (MGD). *Nucleic. Acids Res.* **47**, D801–D806.
- Camp, J.G., Badsha, F., Florio, M., Kanton, S., Gerber, T., Wilsch-Bräuninger, M., Lewitus, E., Sykes, A., Hevers, W., Lancaster, M., et al. (2015). Human cerebral organoids recapitulate gene expression programs of fetal neocortex development. *Proc. Natl. Acad. Sci. USA* **112**, 15672–15677.
- 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.
- Duan, B., Zhou, C., Zhu, C., Yu, Y., Li, G., Zhang, S., Zhang, C., Ye, X., Ma, H., Qu, S., et al. (2019). Model-based understanding of single-cell CRISPR screening. *Nat. Commun.* **10**, 2233.
- Eiraku, M., Watanabe, K., Matsuo-Takahashi, M., Kawada, M., Yonemura, S., Matsumura, M., Wataya, T., Nishiyama, A., Muguruma, K., and Sasai, Y. (2008). Self-organized formation of polarized cortical tissues from ESCs and its active manipulation by extrinsic signals. *Cell Stem Cell* **3**, 519–532.
- Fan, X., Dong, J., Zhong, S., Wei, Y., Wu, Q., Yan, L., Yong, J., Sun, L., Wang, X., Zhao, Y., et al. (2018). Spatial transcriptomic survey of human embryonic cerebral cortex by single-cell RNA-seq analysis. *Cell Res.* **28**, 730–745.
- Giandomenico, S.L., and Lancaster, M.A. (2017). Probing human brain evolution and development in organoids. *Curr. Opin. Cell Biol.* **44**, 36–43.
- Giandomenico, S.L., Mierau, S.B., Gibbons, G.M., Wenger, L.M.D., Masullo, L., Sit, T., Sutcliffe, M., Boulanger, J., Tripodi, M., Derivery, E., et al. (2019). Cerebral organoids at the air-liquid interface generate diverse nerve tracts with functional output. *Nat. Neurosci.* **22**, 669–679.
- Glur, C. (2020). data.tree: General Purpose Hierarchical Data Structure. (R package version 1.0.0). <https://CRAN.R-project.org/package=data.tree>.
- Goulburn, A.L., Alden, D., Davis, R.P., Micallef, S.J., Ng, E.S., Yu, Q.C., Lim, S.M., Soh, C.L., Elliott, D.A., Hatzistavrou, T., et al. (2011). A targeted Nkx2.1 human embryonic stem cell reporter line enables identification of human basal forebrain derivatives. *Stem Cells* **29**, 462–473, <https://doi.org/10.1002/stem.587>.
- Kadoshima, T., Sakaguchi, H., Nakano, T., Soen, M., Ando, S., Eiraku, M., and Sasai, Y. (2013). 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.
- Kamata, T., Katsube, K., Michikawa, M., Yamada, M., Takada, S., and Mizusawa, H. (2004). R-spondin, a novel gene with thrombospondin type 1 domain, was expressed in the dorsal neural tube and affected in Wnt mutants. *Biochim. Biophys. Acta* **1676**, 51–62.
- Kanton, S., Boyle, M.J., He, Z., Santel, M., Weigert, A., Sanchis-Calleja, F., Guijarro, P., Sidow, L., Fleck, J.S., Han, D., et al. (2019). Organoid single-cell genomic atlas uncovers human-specific features of brain development. *Nature* **574**, 418–422.
- Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.R., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296.
- La Manno, G., Siletti, K., Furlan, A., Gyllborg, D., Vinsland, E., Langseth, C.M., Khven, I., Johnsson, A., Nilsson, M., Lönnberg, P., et al. (2020). Molecular architecture of the developing mouse brain. *bioRxiv*.
- Lancaster, M.A., and Knoblich, J.A. (2014). Generation of cerebral organoids from human pluripotent stem cells. *Nat. Protoc.* **9**, 2329–2340.
- Lancaster, M.A., Renner, M., Martin, C.-A., Wenzel, D., Bicknell, L.S., Hurles, M.E., Homfray, T., Penninger, J.M., Jackson, A.P., and Knoblich, J.A. (2013). Cerebral organoids model human brain development and microcephaly. *Nature* **501**, 373–379.

- Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12, 323.
- Loo, L., Simon, J.M., Xing, L., McCoy, E.S., Niehaus, J.K., Guo, J., Anton, E.S., and Zylka, M.J. (2019). Single-cell transcriptomic analysis of mouse neocortical development. *Nat. Commun.* 10, 134.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550.
- Lun, A.T.L., Riesenfeld, S., Andrews, T., Dao, T.P., Gomes, T., and Marioni, J.C.; participants in the 1st Human Cell Atlas Jamboree (2019). EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. *Genome Biol.* 20, 63.
- Mayer, C., Hafemeister, C., Bandler, R.C., Machold, R., Batista Brito, R., Jaglin, X., Allaway, K., Butler, A., Fishell, G., and Satija, R. (2018). Developmental diversification of cortical inhibitory interneurons. *Nature* 555, 457–462.
- Mora-Bermúdez, F., Badsha, F., Kanton, S., Camp, J.G., Vernot, B., Köhler, K., Voigt, B., Okita, K., Maricic, T., He, Z., et al. (2016). Differences and similarities between human and chimpanzee neural progenitors during cerebral cortex development. *eLife* 5, e18683.
- Nowakowski, T.J., Bhaduri, A., Pollen, A.A., Alvarado, B., Mostajo-Radji, M.A., Di Lullo, E., Haeussler, M., Sandoval-Espinosa, C., Liu, S.J., Velmeshev, D., et al. (2017). Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science* 358, 1318–1323.
- Pašca, S.P. (2018). The rise of three-dimensional human brain cultures. *Nature* 553, 437–445.
- Pašca, A.M., Sloan, S.A., Clarke, L.E., Tian, Y., Makinson, C.D., Huber, N., Kim, C.H., Park, J.-Y., O'Rourke, N.A., Nguyen, K.D., et al. (2015). Functional cortical neurons and astrocytes from human pluripotent stem cells in 3D culture. *Nat. Methods* 12, 671–678.
- Pellegrini, L., Bonfio, C., Chadwick, J., Begum, F., Skehel, M., and Lancaster, M.A. (2020). Human CNS barrier-forming organoids with cerebrospinal fluid production. *Science* 369, eaaz5626.
- Picelli, S., Faridani, O.R., Björklund, A.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* 9, 171–181.
- Polioudakis, D., de la Torre-Ubieta, L., Langerman, J., Elkins, A.G., Shi, X., Stein, J.L., Vuong, C.K., Nichterwitz, S., Gevorgian, M., Opland, C.K., et al. (2019). A Single-Cell Transcriptomic Atlas of Human Neocortical Development during Mid-gestation. *Neuron* 103, 785–801.e8.
- Pollen, A.A., Bhaduri, A., Andrews, M.G., Nowakowski, T.J., Meyerson, O.S., Mostajo-Radji, M.A., Di Lullo, E., Alvarado, B., Bedolli, M., Dougherty, M.L., et al. (2019). Establishing Cerebral Organoids as Models of Human-Specific Brain Evolution. *Cell* 176, 743–756.e17.
- Qian, X., Nguyen, H.N., Song, M.M., Hadiono, C., Ogden, S.C., Hammack, C., Yao, B., Hamersky, G.R., Jacob, F., Zhong, C., et al. (2016). Brain-Region-Specific Organoids Using Mini-bioreactors for Modeling ZIKV Exposure. *Cell* 165, 1238–1254.
- Quadrato, G., Nguyen, T., Macosko, E.Z., Sherwood, J.L., Min Yang, S., Berger, D.R., Maria, N., Scholvin, J., Goldman, M., Kinney, J.P., et al. (2017). Cell diversity and network dynamics in photosensitive human brain organoids. *Nature* 545, 48–53.
- Racle, J., de Jonge, K., Baumgaertner, P., Speiser, D.E., and Gfeller, D. (2017). Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. *eLife* 6, 6.
- Rosenberg, A.B., Roco, C.M., Muscat, R.A., Kuchina, A., Sample, P., Yao, Z., Graybuck, L.T., Peeler, D.J., Mukherjee, S., Chen, W., et al. (2018). Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* 360, 176–182.
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. *Cell* 177, 1888–1902.e1821.
- Thompson, C.L., Ng, L., Menon, V., Martinez, S., Lee, C.K., Glattfelder, K., Sunkin, S.M., Henry, A., Lau, C., Dang, C., et al. (2014). A high-resolution spatiotemporal atlas of gene expression of the developing mouse brain. *Neuron* 83, 309–323.
- Velasco, S., Kedaigle, A.J., Simmons, S.K., Nash, A., Rocha, M., Quadrato, G., Paulsen, B., Nguyen, L., Adiconis, X., Regev, A., et al. (2019). Individual brain organoids reproducibly form cell diversity of the human cerebral cortex. *Nature* 570, 523–527.
- Xiang, Y., Tanaka, Y., Patterson, B., Kang, Y.J., Govindaiah, G., Roselaar, N., Cakir, B., Kim, K.Y., Lombroso, A.P., Hwang, S.M., et al. (2017). Fusion of Regionally Specified hPSC-Derived Organoids Models Human Brain Development and Interneuron Migration. *Cell Stem Cell* 21, 383–398.e7.
- Xiang, Y., Tanaka, Y., Cakir, B., Patterson, B., Kim, K.Y., Sun, P., Kang, Y.J., Zhong, M., Liu, X., Patra, P., et al. (2018). hESC-Derived Thalamic Organoids Form Reciprocal Projections When Fused with Cortical Organoids. *Cell Stem Cell* 24, 487–497.e7.
- Yu, G. (2020a). tidytree: A Tidy Tool for Phylogenetic Tree Data Manipulation. (R package version 0.3.3.). <https://CRAN.R-project.org/package=tidytree>.
- Yu, G. (2020b). Using ggtree to Visualize Data on Tree-Like Structures. *Curr. Protoc. Bioinformatics* 69, e96, <https://doi.org/10.1002/cbpi.96>. <https://currentprotocols.onlinelibrary.wiley.com/doi/abs/10.1002/cbpi.96>.
- Yu, G., Wang, L.-G., and He, Q.-Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* 31, 2382–2383.
- Zhong, S., Zhang, S., Fan, X., Wu, Q., Yan, L., Dong, J., Zhang, H., Li, L., Sun, L., Pan, N., et al. (2018). A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. *Nature* 555, 524–528.

## STAR★METHODS

## KEY RESOURCES TABLE

| REAGENT or RESOURCE  | SOURCE                   | IDENTIFIER                   |
|--|--------------------------|------------------------------|
| <b>Chemicals, peptides, and recombinant proteins</b>           |                          |                              |
| Matrigel® hESC-Qualified Matrix, LDEV-free                     | Corning                  | Cat#354277                   |
| mTeSR™ Plus  | StemCell Technologies    | Cat#05825                    |
| TrypLE Express Enzyme  | Thermo Fisher Scientific | Cat#12605010                 |
| DMEM/F12, HEPES with L-glutamine and phenol red                | Thermo Fisher Scientific | Cat#11330032 or Cat#31330038 |
| N-2 supplement (100X)  | Thermo Fisher Scientific | Cat#17502048                 |
| GlutaMAX™ supplement   | Thermo Fisher Scientific | Cat#35050061                 |
| MEM Non-essential Amino Acid Solution (MEM-NEAA)               | Sigma-Aldrich            | Cat#M7145                    |
| Heparin  | Sigma-Aldrich            | Cat#H3149                    |
| Neurobasal™ Medium   | Thermo Fisher Scientific | Cat#21103049                 |
| B-27 Supplement (50X), minus vitamin A                         | Thermo Fisher Scientific | Cat#12587010                 |
| B-27 Supplement (50X), serum free                              | Thermo Fisher Scientific | Cat#17504044                 |
| Insulin solution (human)                                       | Sigma-Aldrich            | Cat#I9278                    |
| 2-Mercaptoethanol  | Sigma-Aldrich            | Cat# M6250                   |
| Penicillin-Streptomycin  | Thermo Fisher Scientific | Cat# 15140122                |
| Matrigel® Growth Factor Reduced (GFR) Basement Membrane Matrix | Corning                  | Cat#356231. Lot#9231005      |
| Recombinant Human Sonic Hedgehog/Shh (C24II) N terminus        | R&D Systems              | Cat#1845-SH-025              |
| Recombinant Human R-Spondin-2                                  | Peprotech                | Cat#120-43                   |
| Recombinant Human R-Spondin-3                                  | Peprotech                | Cat#120-44                   |
| SB431542   | StemCell Technologies    | Cat# #72232                  |
| Dorsomorphin   | Sigma-Aldrich            | Cat#P5499                    |
| CHIR99021  | StemCell Technologies    | Cat#72054                    |
| TRI Reagent™ Solution  | Thermo Fisher Scientific | Cat#AM9738                   |
| 1-bromo-3-chloropropane (BCP)                                  | Lucerna-Chem             | Cat# BP151                   |
| Betaine  | Sigma-Aldrich            | Cat# B0300-1VL               |
| RNase inhibitor  | Takara                   | Cat# 2313A                   |
| dNTPs (10 mM)  | Thermo Fisher Scientific | Cat# R0191                   |
| MgCl <sub>2</sub>  | Sigma-Aldrich            | Cat# M1028-100ML             |
| AMPure XP  | Beckman Coulter          | Cat# A63881                  |
| Ethanol ≥ 99.8%  | Honeywell                | Cat# 02860-2.5L              |
| Isopentane (2-methylbutane)                                    | Sigma-Aldrich            | Cat# M32631-1L               |
| EB solution  | QIAGEN                   | Cat# 19086                   |
| Nuclease-Free water  | Invitrogen               | Cat# AM993                   |
| <b>Critical commercial assays</b>                              |                          |                              |
| EZ-PCR Mycoplasma Detection kit                                | Biological Industries    | Cat# 20-700-20               |
| RNAlater-ICE Frozen Tissue Transition Solution                 | Thermo Fisher Scientific | Cat# AM7030                  |
| MagMAX-96 Total RNA Isolation Kit                              | Thermo Fisher Scientific | Cat# AM1830                  |
| Superscript II kit with DTT. (contains 5X First-Strand buffer) | Thermo Fisher Scientific | Cat# 18064071                |
| Kapa HiFi HotStart ReadyMix                                    | Roche                    | Cat# 07958927001             |
| Nextera XT DNA Library Preparation Kit (96 samples)            | Illumina                 | Cat# FC-131-1096             |

(Continued on next page)

**Continued**

| REAGENT or RESOURCE  | SOURCE                   | IDENTIFIER  |
|--|--------------------------|---|
| Nextera XT Index Kit v2 Set B (96 indexes)   | Illumina                 | Cat# FC-131-2002  |
| Qubit dsDNA high-sensitivity (HS) kit  | Invitrogen               | Cat# Q32851   |
| Qubit 2.0 fluorometer  | Invitrogen               | Cat# Q32866; RRID: SCR_020553   |
| Agilent high-sensitivity DNA kit   | Agilent Technologies     | Cat# 5067-4626  |
| Agilent 2100 Bioanalyzer   | Agilent Technologies     | Cat# G2938C; RRID: SCR_018043   |
| <b>Deposited data</b>  |                          |   |
| Developmental diversification of cortical inhibitory interneurons (transcriptome profiling)  | Mayer et al., 2018       | <a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE104158">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE104158</a>   |
| Human whole brain organoid single-cell genomic atlas   | Kanton et al., 2019      | <a href="https://www.ebi.ac.uk/arrayexpress/">https://www.ebi.ac.uk/arrayexpress/</a> . E-MTAB-7552 (cerebral sc-RNaseq). E-MTAB-8089. (cerebral scATAC-seq)                            |
| Single-cell profiling of the developing mouse brain  | Rosenberg et al., 2018   | <a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE110823">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE110823</a>   |
| Single-cell gene expression atlas from the developing mouse brain  | La Manno et al., 2020    | <a href="http://mousebrain.org/downloads.html">http://mousebrain.org/downloads.html</a>   |
| Single-cell gene expression data from the developing mouse neocortex   | Loo et al., 2019         | <a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE12335">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE12335</a>   |
| Single-cell RNA-seq developmental atlas of human prefrontal cortex   | Zhong et al., 2018       | <a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE104276">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE104276</a>   |
| Single cell RNA-seq analysis of human embryonic cortex   | Fan et al., 2018         | <a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE103723">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE103723</a>   |
| Single Cell Transcriptomic Analysis of Human Neocortical Development   | Polioudakis et al., 2019 | <a href="https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001836.v1.p1">https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001836.v1.p1</a> |
| Spatiotemporal gene expression profiling of the developing human cortex  | Nowakowski et al., 2017  | <a href="https://cells.ucsc.edu/?ds=cortex-dev">https://cells.ucsc.edu/?ds=cortex-dev</a>   |
| BrainSpan Developing Human Brain Atlas   | Thompson et al., 2014    | <a href="http://www.brainspan.org/static/download.html">http://www.brainspan.org/static/download.html</a> ; RRID: SCR_008083  |
| Allen Developing Mouse Brain Atlas   | Thompson et al., 2014    | <a href="http://help.brain-map.org/display/devmouse/API">http://help.brain-map.org/display/devmouse/API</a> ; RRID: SCR_002990  |
| Single cell profiling of cortical and ventral brain organoids  | Birey et al., 2017       | <a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE93811">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE93811</a>   |
| Single cell profiling of Medial Ganglionic Eminence and cortical organoids   | Xiang et al., 2017       | <a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE97882">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE97882</a>   |
| Single cell gene expression profiling of thalamic organoids  | Xiang et al., 2018       | <a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE122342">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE122342</a>   |
| Single cell gene expression profiling of cortical organoids  | Velasco et al., 2019     | <a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE129519">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE129519</a>   |
| VoxHunt 3D gene expression maps of the developing mouse brain  | This paper               | <a href="https://data.mendeley.com/datasets/g4xg38mwcn/">https://data.mendeley.com/datasets/g4xg38mwcn/</a>   |
| Bulk-RNA seq data from Patterning Screening  | This paper               | <a href="https://www.ebi.ac.uk/arrayexpress/">https://www.ebi.ac.uk/arrayexpress/</a> . Accession number E-MTAB-10037   |
| <b>Experimental models: cell lines</b>   |                          |   |
| Human: HES3 NKX2.1 <sup>GFP/w</sup>  | Goulburn et al., 2011    | <a href="http://doi.wiley.com/10.1002/stem.587">http://doi.wiley.com/10.1002/stem.587</a> ; RRID: N/A   |
| <b>Oligonucleotides</b>  |                          |   |
| TSO-LNA-blocked . (5'-AAGCAGTGTTATCAACG CAGAGTACATrGrG+G-3'), rG, riboGuanosine; +G, LNA-modified Guanosine. Dissolved in TE buffer, 100 µM. | Eurogentec               | N/A   |
| IS PCR primers. (5'-AAGCAGTGTTATCAACG CAGAGT-3') . Dissolved in TE buffer, 100 µM.   | IDT                      | N/A   |

(Continued on next page)

**Continued**

| REAGENT or RESOURCE   | SOURCE                 | IDENTIFIER  |
|---|------------------------|---|
| <b>Oligo-dT30VN</b> . (5'-AAGCAGTGTTATCAAC<br>GCAGAGTACT30VN-3') . V, Adenosine/Cytosine/<br>Guanine; N, any base. Dissolved in TE<br>buffer, 100 µM. | IDT                    | N/A   |
| <b>Software and algorithms</b>  |                        |   |
| VoxHunt   | This paper             | <a href="https://github.com/quadbiolab/voxhunt">https://github.com/quadbiolab/voxhunt</a>   |
| bcl2fastq v2.20.422   | Illumina               | <a href="https://emea.support.illumina.com/downloads/bcl2fastq-conversion-software-v2-20.html">https://emea.support.illumina.com/downloads/bcl2fastq-conversion-software-v2-20.html</a> |
| STAR v2.5.1b  | Dobin et al., 2013     | <a href="https://github.com/alexdobin/STAR">https://github.com/alexdobin/STAR</a> ; RRID:<br>SCR_015899   |
| RSEM v1.3.1   | Li and Dewey, 2011     | <a href="https://github.com/deweylab/RSEM">https://github.com/deweylab/RSEM</a> ; RRID:<br>SCR_013027   |
| DESeq2 (R package)  | Love et al., 2014      | <a href="http://bioconductor.org/packages/release/bioc/html/DESeq2.html">http://bioconductor.org/packages/release/bioc/html/DESeq2.html</a> ; RRID: SCR_015687                          |
| Seurat 3.0 (R package)  | Stuart et al., 2019    | <a href="http://satijalab.org/seurat/articles/install.html">http://satijalab.org/seurat/articles/install.html</a> ;<br>RRID: SCR_016341   |
| EPIC (R package)  | Racle et al., 2017     | <a href="http://epic.gfellerlab.org">http://epic.gfellerlab.org</a>   |
| CHIPseeker (R package)  | Yu et al., 2015        | <a href="http://www.bioconductor.org/packages/release/bioc/html/ChIPseeker.html">http://www.bioconductor.org/packages/release/bioc/html/ChIPseeker.html</a>                             |
| Harmony (R package)   | Korsunsky et al., 2019 | <a href="https://github.com/immunogenomics/harmony">https://github.com/immunogenomics/harmony</a>   |
| DropletUtils (R package)  | Lun et al., 2019       | <a href="https://bioconductor.org/packages/release/bioc/html/DropletUtils.html">https://bioconductor.org/packages/release/bioc/html/DropletUtils.html</a>                               |
| presto (R package)  | Korsunsky et al., 2019 | <a href="https://github.com/immunogenomics/presto">https://github.com/immunogenomics/presto</a>   |
| irlba (R package)   | Baglama et al., 2019   | <a href="https://cran.r-project.org/web/packages/irlba/index.html">https://cran.r-project.org/web/packages/irlba/index.html</a>   |
| UMAP  | Becht et al., 2018     | <a href="https://github.com/lmcinnes/umap">https://github.com/lmcinnes/umap</a> ; RRID:<br>SCR_018217   |
| data.tree (R package)   | Glur, 2020             | <a href="https://github.com/gluc/data.tree">https://github.com/gluc/data.tree</a>   |
| tidytree (R package)  | Yu, 2020a              | <a href="https://github.com/YuLab-SMU/tidytree">https://github.com/YuLab-SMU/tidytree</a>   |
| ggtree (R package)  | Yu, 2020b              | <a href="https://bioconductor.org/packages/release/bioc/html/ggtree.html">https://bioconductor.org/packages/release/bioc/html/ggtree.html</a>   |
| ggraph (R package)  | CRAN                   | <a href="https://github.com/thomasp85/ggraph">https://github.com/thomasp85/ggraph</a>   |
| <b>Other</b>  |                        |   |
| Mouse Genome Informatics Database   | Bult et al., 2019      | <a href="http://www.informatics.jax.org/">http://www.informatics.jax.org/</a> ; RRID:<br>SCR_012953   |

**RESOURCE AVAILABILITY****Lead Contact**

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Barbara Treutlein ([barbara.treutlein@bsse.ethz.ch](mailto:barbara.treutlein@bsse.ethz.ch)).

**Materials availability**

HES3, NKX2.1<sup>GFP/w</sup> hESCs will be provided upon execution of a suitable Materials Transfer Agreement (MTA) with Ed Stanley and Andrew G. Elefanty (Murdoch Childrens Research Institute, Melbourne).

**Data and code availability**

VoxHunt was implemented as an R package and is available at <https://github.com/quadbiolab/voxhunt>. R markdown scripts reproducing the main analyses in this study are available at [https://github.com/quadbiolab/voxhunt\\_reproducibility](https://github.com/quadbiolab/voxhunt_reproducibility). All datasets used in this study are available for download in public repositories (see [key resources table](#) for individual accession numbers). The accession number for the bulk RNA-sequencing data reported in this paper is ArrayExpress: E-MTAB-10037.

**EXPERIMENTAL MODEL AND SUBJECT DETAILS****hESC experiments approval**

Stem cell experiments with NKX2.1<sup>GFP/w</sup> hESCs were approved by the Bundesamt für Gesundheit (Swiss Health Federal Office) with disposition number 606.0000-1/31 / 19.018224; and by the Ethikkommission Nordwest- und Zentralschweiz (Ethics Commission for Northern and Central Switzerland).

**hESC line culture and maintenance**

NKX2.1<sup>GFP/w</sup> hESCs were obtained from Agneta Kirkeby's research group at the University of Copenhagen, after arrangement of an MTA with Prof. Ed Stanley and Prof. Andrew G. Elefanty (Murdoch Childrens Research Institute, Melbourne). NKX2.1<sup>GFP/w</sup> hESCs were grown at 37°C, 5% CO<sub>2</sub> in feeder-free conditions, on 6-well tissue culture plates coated with hESC-Qualified Matrigel (Corning). hESCs were fed every other day with mTesR Plus (StemCell Technologies) and passaged when reaching around 80% confluence with EDTA for gentle dissociation. NKX2.1<sup>GFP/w</sup> hESCs were tested at passage 67 for copy number changes using the Agilent ISCA 8x60K v2 array and no abnormalities were detected. Cells were also karyotyped at passage 78 and showed apparently normal female karyotype in 20 cells examined. These analyses were performed by the Cell Guidance Systems Genetics Service (Cambridge, UK). Regular PCR-based Mycoplasma testing (Biological Industries) was performed to discard potential Mycoplasma infections.

**Brain organoid generation and incubation with patterning molecules**

NKX2.1<sup>GFP/w</sup> hESCs (passage 69) were trypsinized using TrypLE™ Express Enzyme (Thermo Fisher Scientific) to obtain single cell suspensions and 500 cells were plated in each well of ultra-low attachment 96-well plates (day 0). Cells in suspension were cultured for one day in mTesR Plus (StemCell Technologies) to allow for aggregation and Embryoid Body (EB) formation. On day 1, EBs were transitioned to Neural Induction Media (NIM, containing DMEM/F12, 1% N2 supplement (v/v), 1% GlutaMAX supplement (v/v), 1% MEM-NEAA (v/v) and 1 µg/mL Heparin). From day 3 to day 6, organoids were incubated in Neural differentiation media (Ndif-VA, one-half DMEM/F1, one-half Neurobasal, 0.5% N2 supplement (v/v), 1% B27 supplement without Vitamin A (v/v), 1:4000 insulin (Sigma Aldrich), 49.5 µM 2-Mercaptoethanol, 1% GlutaMAX (v/v) (Invitrogen), 0.5% MEM-non essential amino acids and 1% Penicillin/Streptomycin). Finally, organoids were transferred to Neural Maturation Media (Ndif+VA, same recipe as Ndif-VA but with B27 supplement containing Vitamin A) on day 6 and media was changed every 2-3 days until day 14. Matrigel was added to the corresponding media at increasing dilutions during organoid development. From day 3 to 6, brain organoids were supplemented with 1:50 Growth Factor Reduced (GFR) Matrigel (Corning); from day 6 to 8, with 1:100 GFR Matrigel; and from then on with 1:200 GFR Matrigel. Patterning factors were added in low and high concentrations from day 3 to 6: SHH (low, 50 ng/mL; high, 100 ng/mL), RSPO2 (low, 100 ng/mL; high, 1 µg/mL), RSPO3 (low, 100 ng/mL; high, 1 µg/mL), SB-431542 and Dorsomorphin (1 µM and 2.5 µM, respectively) and CHIR (low, 0.5 µM; high, 5 µM).

**METHOD DETAILS****Collection of patterned organoids and RNA purification**

On day 14, all culture media was removed and patterned brain organoids were snap-frozen in a bed of isopentane (Sigma-Aldrich) and dry ice. Immediately following freezing, organoids were transferred to -80°C and stored at this temperature until the day prior to RNA extraction. On that day, organoids were immersed in 200 µL of RNA Later ICE (Thermo Fisher) and defrosted overnight to protect RNA from degradation during the following steps. Organoids were transferred to TriReagent (Thermo Fisher) and immediately vortexed for 30 s to homogenize the tissue. After 5 minutes of incubation at room temperature, 0.1X volumes of 1-Bromo-3-chloropropane (BCP, Lucerna-Chem) were added, followed by a 10 s-vortexing step and 5 minutes of incubation at room temperature. Samples were centrifuged at 12,000 g for 10 minutes (4°C) and 100 µL of supernatant (aqueous phase) was collected to proceed with RNA purification using the spin procedure of the MagMax-96 total isolation kit (Thermo Fisher).

**Smartseq2 library preparation**

Purified bulk RNA samples from each individual organoid were then processed for Smartseq2 (SS2) library generation using the protocol published by Picelli and colleagues (Picelli et al., 2014) with slight modifications. As our starting material was extracted RNA and not single cells, Triton X-100 in lysis buffer was replaced by Nuclease-Free water. Therefore, lysis buffer composition was 1:10 RNase inhibitor (Takara) in Nuclease-free water (Invitrogen). 1 µL of extracted bulk RNA was incubated with 1 µL of lysis buffer, 1 µL of oligo-dT primer and 1 µL of dNTP mix (Thermo Fisher Scientific) for 3 minutes at 72°C.

Next, samples were reverse transcribed using the SuperScript II kit with DTT (Thermo Fisher Scientific). The RT mix included 100 U SuperScript II, 10 U RNase inhibitor (Takara), 1X First-Strand buffer, 5 mM DTT, 1 µM Template-Switch Oligo (TSO), 1 M Betaine (Sigma-Aldrich) and 6 mM MgCl<sub>2</sub> (Sigma-Aldrich) previously diluted in Nuclease-Free water (Invitrogen). Samples were incubated at 42°C for 90 min, followed by 10 cycles at 50°C for 2 min and 42°C for 2 min to prevent secondary RNA structures from stopping the polymerase and enable completion of RT. Finally, samples underwent an enzyme inactivation step at 70°C for 15 minutes.

Next, a PCR preamplification step (SS2 PCR) was performed. The PCR mix contained 10 µL of reverse-transcribed sample, 1X KAPA HiFi HotStart ReadyMix (Roche), 0.1 µM IS PCR primers (IDT) and Nuclease-Free water (Invitrogen) up to a total volume of

25 µL. PCR was performed using the following program: initial denaturation step at 98°C for 3 min, 15 cycles of 98°C for 20 s - 67°C for 15 s - 72°C for 6 min, and a final elongation step at 72°C for 5 min.

SS2 reactions were cleaned up using AMPure XP reagent (Beckman Coulter) at 1X ratio and subsequently quality-checked for fragment sizes and concentrations with Qubit 2.0 (Invitrogen) and Agilent 2100 Bioanalyzer (Agilent Technologies). All samples were then adjusted to concentrations around 200 pg/µL with EB solution (QIAGEN) prior to Nextera library preparation.

Fragmentation of cDNA was performed at 55°C for 5 min, using the Nextera XT DNA sample preparation kit (Illumina). Each fragmentation reaction was composed of 2.5 µL (1X) Tagment DNA (TD) buffer, 1.25 µL Amplicon Tagment Mix (ATM) and 250 pg (around 1.25 µL) of purified DNA from PCR. Immediately after, Tn5 transposase was stripped off the DNA by incubating each reaction with 1.25 µL of Neutralize Tagment (NT) buffer for 5 min at room temperature.

A final amplification of fragmented DNA (indexing PCR) was performed with index primers from Set B of Nextera XT Index Kit v2 (Illumina). 1.25 µL of each i5 and i7 adaptor and 3.75 µL of Nextera PCR Mastermix (NPM) were added to 1.25 µL of each sample prior to running the indexing PCR with Illumina's recommended settings: extension at 72°C for 3 min, denaturation at 95°C for 30 s, 12 cycles of 95°C for 10 s - 55°C for 30 s - 72°C for 30 s, and a final elongation step at 72°C for 5 min.

Immediately after indexing PCR, samples were pooled in equal ratios by adding 3 µL of each sample to a 1.5 mL tube. Pooled libraries were then cleaned up twice with AMPure XP (Beckman Coulter) at 0.9X ratio. Of note, a 0.9X ratio of AMPure XP was used instead of 1X ratio (as stated in [Picelli et al., 2014](#)) because it proved more efficient removing small carryover primers. A final quality control with Qubit 2.0 (Invitrogen) and Agilent 2100 Bioanalyzer (Agilent Technologies) showed cDNA concentration in the range of 10 nM.

### Sequencing of SS2 libraries

The pooled libraries were appropriately diluted and sequenced at the Genomics Facility Basel (GFB) using NextSeq, with paired-end configuration and 38 cycles. Sequencing data was demultiplexed by the GFB, using bcl2fastq version 2.20.422 with the following parameters: –ignore-missing-bcls –ignore-missing-controls –ignore-missing-positions –ignore-missing-filter –no-bgzf-compression –barcode-mismatches 1.

### Acquisition and preprocessing of Allen Brain Atlas *in situ* hybridization data

We downloaded 3D expression grid data from all *in situ* hybridization (ISH) experiments corresponding to the eight developmental time points (E11.5, E13.5, E15.5, E18.5, P4, P14, P28, P56) using the application programming interface (API) provided by the Allen Brain Atlas (ABA) Developing Mouse Brain Data Portal (<http://help.brain-map.org/display/devmouse/API>). In order to obtain *in situ* intensity profiles (all genes) for each voxel, the data for each time point was summarized into a voxel by gene matrix. If multiple experiments were available for the same gene and time point, we calculated the arithmetic mean across experiments. To be able to perform comparisons to human transcriptomics data, we converted the gene names (official gene symbols) to the names of the human homolog, using homology annotations from the Mouse Genome Informatics database (<http://informatics.jax.org>). Genes without a human homolog were excluded from the dataset. The structure ontology was obtained by downloading the structure graph with ID 17 through the ABA web API (<http://www.brain-map.org/api/index.html>). The resulting json file was converted into a data frame. For differential expression testing and visualization, we introduced four additional levels of annotation by summarizing existing annotations. In these custom annotations, pallium (dorsal forebrain), subpallium (ventral forebrain), hypothalamus, diencephalon, midbrain, hindbrain and spinal cord are on the same level on the ontology tree and each splits further into finer structures. For the analysis shown here as well as for the VoxHunt package, we removed all voxels annotated as spinal cord, as this structure was only included in the earliest developmental stages. To remove background signal from *in situ* hybridization intensities we applied a minimum intensity threshold of 1. The final expression matrices including metadata and custom annotations have been deposited to Mendeley Data: <https://doi.org/10.17632/g4xg38mwcn.1>.

### Visualization of the structure annotation ontology

To visualize the structure ontology as a tree, we used the data.tree R package ([Glur, 2020](#)) to convert the structure ontology data frame into a phylo object. We used the tidytree R package ([Yu, 2020a](#)) to manipulate the tree and add custom annotations to the nodes. Finally, the tree was plotted with the ggtree R package ([Yu, 2020b](#)) using a circular layout and edges were colored according to our custom annotation level 3.

### Acquisition of organoid single-cell gene expression and chromatin accessibility data

With the goal of covering a range of different organoid protocols, we collected published single cell gene expression data from dorsal forebrain ([Birey et al., 2017](#); [Velasco et al., 2019](#)), ventral forebrain ([Birey et al., 2017](#)) thalamus ([Xiang et al., 2018](#)), and cerebral ([Kanton et al., 2019](#)) (unpatterned) organoids. Single cell gene expression and/or chromatin accessibility data for these organoids was obtained through the following accessions: GEO: GSE93811 for Birey et al. (cortical and ventral), GEO: GSE97882 for Xiang et al. (ventral), GEO: GSE122342 for Xiang et al. (thalamus), ArrayExpress: E-MTAB-7552 for Kanton et al. (cerebral scRNA-seq) and ArrayExpress: E-MTAB-8089 for Kanton et al. (cerebral scATAC-seq).

### Acquisition of primary bulk and single-cell gene expression data

We obtained a single-cell gene expression atlas from the developing mouse brain (La Manno et al., 2020) from the Mouse Brain Atlas website (<http://mousebrain.org/downloads.html>). Average cluster transcriptomes and annotations were obtained from the file dev\_all.agg.loom and tSNE embedding coordinates for and annotations for individual cells were obtained from the file dev\_all.loom. In addition, we obtained single-cell gene expression data (split-pool barcoding) from the developing mouse brain (Rosenberg et al., 2018) through the GEO accession GSE110823. From the dataset, we extracted one representative neuronal cluster for each evaluated structure: “OB mitral/Tufted Ms4a15” for olfactory bulb, “Medium Spiny Neurons” for striatum, “CTX PyrL6a” for cortex, “MTt Glut” for rostral midbrain, “THAL Glut” for thalamus, “Purkinje Late” for cerebellum, “MD Int Rxfp2” for medulla and “HIPP Granule Nrp2” for hippocampus. Single-cell gene expression data from the mouse ganglionic eminences (Mayer et al., 2018) was obtained through the GEO accession GSE104158 and we extracted ‘mitotic precursors’ annotated as CGE, LGE and MGE for our analysis. Single-cell gene expression data from the developing mouse neocortex (Loo et al., 2019) was obtained from the GEO accession GSE123335. Bulk RNA-seq data from microdissected regions of the human brain was obtained from BrainSpan (<http://www.brainspan.org/static/download.html>). We obtained four datasets with single-cell transcriptomic data of the developing human brain through the following GEO accessions: GSE104276 for Zhong et al. (Zhong et al., 2018), GSE10372 for Fan et al. (Fan et al., 2018), the dbGaP accession phs001836.v1.p1 for Polioudakis et al. (Polioudakis et al., 2019) and through the USCS Cell Browser portal (<https://cells.ucsc.edu/?ds=cortex-dev>) for Nowakowski et al. (Nowakowski et al., 2017).

### QUANTIFICATION AND STATISTICAL ANALYSIS

#### Preprocessing of single-cell gene expression data

Transcript count matrices were preprocessed separately for each dataset with consistent parameters using the R package Seurat 3.0 (Stuart et al., 2019). Cells were filtered based on unique molecular identifier (UMI) counts ( $> 500, < 50000$ ), the number of detected genes ( $> 200, < 6000$ ) and the fraction of mitochondrial genes ( $< 0.1$ ). For all datasets, we annotated cells based on the cell type labels from the original paper and filtered out unannotated cells. Counts for each cell were normalized by the total number of counts for that cell, multiplied by a scaling factor of 10000 and subsequently natural-log transformed (NormalizeData()). Log normalized expression values were used as an input to the function FindVariableFeatures() to find the 2000 most variable genes based on the ‘vst’ selection method. Expression levels of these genes were used to perform Principal Component Analysis (PCA) with the function prcomp\_irlba() from the irlba R package (Baglama et al., 2019) and the top 15 principal components (PCs) were used to produce a two-dimensional embedding with Uniform Manifold Approximation and Projection (UMAP) (Becht et al., 2018).

#### Differential expression analysis of ISH and single-cell data

To perform differential expression analysis (DE) on both ISH and single-cell expression data, we calculated the Wilcoxon rank sum test p values with a Benjamini-Hochberg multiple testing correction (q-value), AUROC values and log fold changes using the wilcox-auc() function from the R-package presto (Korsunsky et al., 2019). For the ISH data, we first performed DE on log2-transformed intensity values with a pseudocount of  $1 \times 10^{-10}$  between custom annotation levels filtered for statistical significance ( $q < 10^{-5}$ ) and sorted by AUROC value to find highly specific marker genes for each structure (structure markers). To find time point-specific genes, we further performed DE between developmental stages for each structure on custom annotation level 2 separately (stage markers). Additionally, we searched for genes whose expression in a structure is upregulated at any fetal developmental stage and is maintained in subsequent stages. For fetal stages E13.5, E15.5 and E18.5 we performed DE between the current stage and all, and older stages versus all younger stages (e.g., for genes upregulated at E15.5: E11.5 and E13.5 versus E15.5 and E18.5) for each structure on custom annotation level 2 separately (fetal development markers). For single-cell data, we performed DE on log normalized expression levels between annotated cell types for each dataset separately. Here, we determined unbiased cluster markers by taking 3 statistically significant ( $q < 10^{-5}$ ) genes with highest log fold change for each cluster.

#### Correlation of single-cell transcriptomes to brain structures

To determine best matching brain structures for each neuronal cluster in the organoid datasets, we calculated the correlation of single cell transcriptomes with gene expression of structures in the fetal E13.5 mouse brain. For each cell in each cluster we calculated the Pearson correlation coefficient between the log normalized transcript count for each cell and the expression vector of each voxel in the E13.5 mouse brain based on 189 structure marker genes (top 10 for each structure at level 3). Correlation coefficients for each cell were then summarized for each structure by calculating the arithmetic mean. From the mean correlation coefficient of each cell with each structure we then defined the similarity to the structure as

$$s_{ij} = \begin{cases} r_{ij} & \text{if } r_{ij} > 0 \\ 0 & \text{otherwise} \end{cases}$$

We assigned each cell to the structure with maximum similarity and displayed the mean similarity to each structure for all clusters.

### Calculation of similarity to developmental stages

To assess the relationship between the age of the neuronal population and the mouse developmental stage, we calculated for each cell in each neuronal population the highest correlating voxel in the best matching structure of custom annotation level 2 for this population. We selected genes that were informative with regard to the developmental stage for each structure separately by taking the union of the above mentioned fetal development markers (top 50 AUROC for each stage and structure) and stage markers specific to E11.5 (top 50 AUROC for each structure) and used their expression values as an input for computing the above described similarity based on Pearson correlation. In total, this resulted in 183 genes for the pallium, 181 genes for the subpallium, 179 genes for the midbrain and 181 genes for the diencephalon.

### Evaluating the effect of feature selection and downsampling

To test how size of the feature set and data quality affect the fidelity of the mapping, we used VoxHunt to map single-cell transcriptome data from the P2 mouse brain (Rosenberg et al., 2018) to voxel maps of the P4 mouse brain and evaluated its performance under different conditions. First, we extracted various feature sets from the P4 mouse brain voxel maps. In the case of DE genes, we adjusted the size of the feature set by taking the 1, 5, 10, 20, 30, 50 or 200 most specific genes for each structure at custom annotation level 3. In the case of variable features, we selected the 10, 50, 100, 200, 500 or 1000 most variable genes. For each feature set as well as for the full set of 1757 genes we computed the similarity of each cell to each structure as described above. For each feature set and brain structure, we evaluated accuracy as the fraction of correctly assigned structures and contrast as the ratio between the similarity to the second best structure and the assigned structure. To evaluate the effect of data quality on the mapping, we further used the function `downsampleMatrix()` from the R package `DropletUtils` (Lun et al., 2019) to downsample the UMI counts of the dataset to as low as 10%. For each downsample fraction and feature set, we computed similarity maps, assigned cells to their most similar structure and computed accuracy as described above.

### Spatial correlation maps of organoid cells transcriptomes

To compute spatial correlation patterns of single-cell transcriptomes to ISH maps, we first intersected the genes measured in the ABA with all genes detected in the dataset. From these genes, we then picked the 10-20 most specific genes for each structure on custom annotation level 3 based on AUROC values to receive 200-300 genes for each dataset (exact feature sets can be found in Table S1). Based on the expression values of the resulting gene set, we calculated the Pearson correlation coefficient between each cell and each voxel. Correlation coefficients for each cell cluster were summarized by computing the arithmetic mean for each voxel and subsequently transformed to a similarity metric as described above. Scaled correlation values were displayed as similarity maps across the brain in a sagittal view and on coronal sections. To summarize the similarities to discrete brain structures, we computed the arithmetic mean over the correlations to all voxels in this structure. For estimation of brain structure identities, each cell or cell cluster was assigned the structure of the most highly correlated voxel. For direct cell-to-voxel projection, we further assigned each cell to the voxel with maximum correlation considering the 8 central sections (5-12) in the E13.5 mouse brain.

### Comparison of organoid cells transcriptomes to other reference datasets

To compare organoid single-cell transcriptomes to RNA-seq data from human microdissected brain structures, we selected the top 50 differentially expressed genes for each structure group in the BrainSpan dataset. Based on these genes, we computed the similarity between each cell and each sample in BrainSpan. To perform comparisons to the single-cell transcriptomic atlas of the developing mouse brain, we used the top 20 spatial structure markers extracted from ISH data to compute the similarity of each cell to the mean transcriptome of each cluster in the reference data. We then visualized these similarities using the tSNE embedding of individual cells.

### Substructure resolution of transcriptomic maps

To annotate brain structures that were not directly included in the ABA structure ontology, we first performed unbiased clustering on voxelized ISH intensities of the higher-level structure (subpallium for ganglionic eminences (GEs) and dorsal pallium for pallial zones). For this, we performed PCA on voxel ISH intensities with all detected genes and used the first 20 PCs as an input for the louvain clustering algorithm (Blondel et al., 2008) (resolution of 2 for subpallium and 6 for dorsal pallium). Next, we annotated the resulting clusters based on canonical marker genes. We validated these annotations through computing spatial similarity maps of mouse scRNA-seq data from populations derived from the respective structures (Mayer et al., 2018) as described above. For the GEs, we used the newly annotated substructures to annotate cerebral organoid cell populations (Kanton et al., 2019) (ventral neurons from the H9 line). For this, we performed PCA on log-normalized transcript counts of the 2000 most variable genes and used the first 20 PCs to perform louvain clustering (Blondel et al., 2008) at a resolution of 1. The clusters were annotated based on their similarity to the annotated substructures. If the similarity patterns were not sufficiently clear, they were annotated as ‘other’. For pallial zones, we used the cell type annotations from the original publication (Kanton et al., 2019) (cortical lineage from the 409b2 line) and computed their similarities to the annotated substructures. In each case, we used the union of top 10 structure markers on custom annotation level 3 as described above and the top 5 makers (based on high AUROC) of each annotated substructure to compute the Pearson correlation (218 genes for GEs and 188 genes for the pallial zones).

**Substructure-resolved annotation of ventral neurons across organoid protocols**

To assess and compare the brain structure identity of inhibitory neuron populations from ventral and cerebral organoid protocols, we extracted neuronal clusters from each dataset and performed PCA on the joined dataset using the scaled expression of all transcription factor coding genes. We used Harmony ([Korsunsky et al., 2019](#)) with default parameters to integrate all datasets based on the first 10 principal components. In the integrated space, we performed louvain clustering with a resolution of 2 and computed the similarities of these clusters to GE substructures as described above. We treated the similarity of all cells to all voxels as background similarity and computed a z-score for the similarity of each cell to each GE substructure by subtracting the mean of the background similarity and dividing by its standard deviation. Based on this z-score, we determined the correlation threshold corresponding to a p value of 0.05. All clusters with a median similarity higher than this threshold were annotated with the maximum correlating structure, while all others were annotated as ‘other’. One cluster showed comparably high similarity above the significance threshold to all three substructures and was annotated as ‘unspecified GE’. To find specific genes for each annotated neuronal population, we performed a Wilcoxon rank sum test and selected genes based on statistical significance ( $q < 10^{-5}$ ) and fold change (top 10).

**Spatial correlation maps of organoid cells chromatin accessibility profiles**

To compute correlation coefficients between transcriptome and chromatin accessibility profiles, the latter was summarized for each gene by counting all peaks in its TSS region ( $TSS \pm 3kb$ ) using the R package CHIPseeker ([Yu et al., 2015](#)). TSS peak counts were further normalized by the total number of counts for that cell, multiplied by a scaling factor of 10,000 and subsequently natural-log transformed. We further selected the 10 most specific markers for each structure at custom annotation level 3 to obtain a total of 182 genes and used them to compute the Pearson correlation-based similarity between voxel expression values and log normalized gene accessibility scores for each cell as described above.

**Preprocessing and analysis of bulk RNA-seq data**

Sequencing reads from each sample were aligned to the human genome (hg38-3.0.0) using STAR v2.5.1b ([Dobin et al., 2013](#)). Transcript and gene counts were obtained from BAM files using RSEM v1.3.1 ([Li and Dewey, 2011](#)). Both tools were run with standard parameters. Gene counts were normalized to counts per million (CPM) by dividing by the total number of reads per sample and multiplying by a scaling factor of 1,000,000. CPM values were subsequently natural-log transformed (log CPM). Principal component analysis was performed based on scaled log CPM values of the 2000 most highly variable genes. Differential expression analysis data was performed with DESeq2 ([Love et al., 2014](#)) using the Seurat function FindMarkers(), followed by a Benjamini-Hochberg multiple testing correction (q-value). To find genes which drive the differences in spatial similarity patterns between condition and control, we selected genes with high statistical significance ( $q < 10^{-5}$ ) that were measured in the ABA. Out of the resulting gene set, we selected genes with the highest fold change for each direction separately. To compute spatial similarity maps, we used log CPM values of the top 20 structure markers on custom annotation level 3.

**Deconvolution of organoid pseudo-bulk RNA-seq data**

Pseudo-bulk transcriptomes for each organoid were obtained from UMI counts by taking the arithmetic mean of each gene. From the resulting gene expression vector, log CPM values were obtained by dividing by the total number of reads per sample, multiplying by a scaling factor of 1,000,000 and natural-log transformation. For deconvolution based on the Developing Mouse Brain Atlas, ISH fluorescence intensities were analogously transformed to ‘pseudo-CPM’ by log-normalization of each voxel and subsequently summarized for each structure at custom annotation level 2 by calculating the arithmetic mean to obtain transcriptomic references for deconvolution. We used the mean transcriptomes of ‘pallium’, ‘subpallium’, ‘hypothalamus’, ‘diencephalon’, ‘midbrain’, ‘prepontine hindbrain’ and ‘pontine hindbrain’ as references for deconvolution. To perform deconvolution based on bulk RNA-seq data from BrainSpan, we summarized log-normalized counts of all samples aged 12 - 20 pcw to structure groups by calculating the arithmetic mean and used the resulting gene expression vectors as references for deconvolution. Deconvolution of pseudo-bulk organoid transcriptomes was done using EPIC ([Racle et al., 2017](#)) for both references separately. We constrained the input to the top 50 level 2 structure markers and used the top 20 level 2 structure markers as ‘sigGenes’. Additionally, we set ‘withOtherCells’ to False and used otherwise default parameters. The estimated structure proportions were further summarized to higher level structures ‘pallium’, ‘ventral telencephalon’ (subpallium and hypothalamus), ‘diencephalon’, and ‘non-telencephalon’ (midbrain and hindbrain).

Update

**Cell Stem Cell**

Volume 28, Issue 6, 3 June 2021, Page 1177–1180

DOI: <https://doi.org/10.1016/j.stem.2021.03.015>

**Correction****Resolving organoid brain region identities by mapping single-cell genomic data to reference atlases**

Jonas Simon Fleck, Fátima Sanchís-Calleja, Zhisong He, Małgorzata Santel, Michael James Boyle, J. Gray Camp,\* and Barbara Treutlein\*

\*Correspondence: [grayson.camp@iob.ch](mailto:grayson.camp@iob.ch) (J.G.C.), [barbara.treutlein@bsse.ethz.ch](mailto:barbara.treutlein@bsse.ethz.ch) (B.T.)

<https://doi.org/10.1016/j.stem.2021.03.015>

(Cell Stem Cell 28, 1148–1159.e1–e8; June 3, 2021)

In the version of the manuscript by Fleck et al. originally published online, the publishers mistakenly swapped the positions of Figure 5 and Figure 6, thus erroneously pairing the title and legend of Figure 5 with the actual image of Figure 6 and vice versa. The figures now appear in the correct positions in the manuscript online and below, and the publisher apologizes to the authors and the wider community for the oversight.



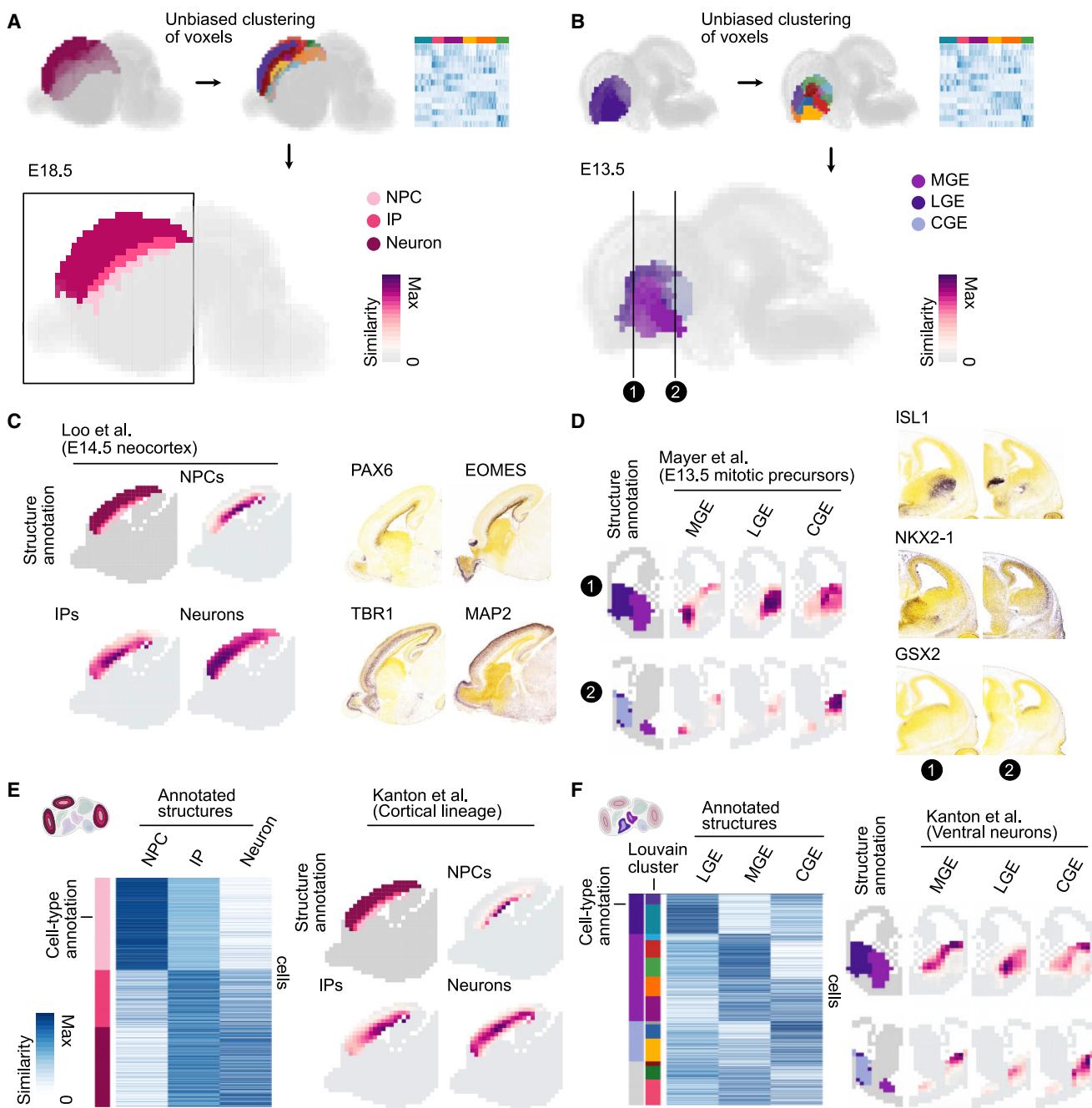


Figure 5. Substructure resolution of transcriptomic maps (corrected)

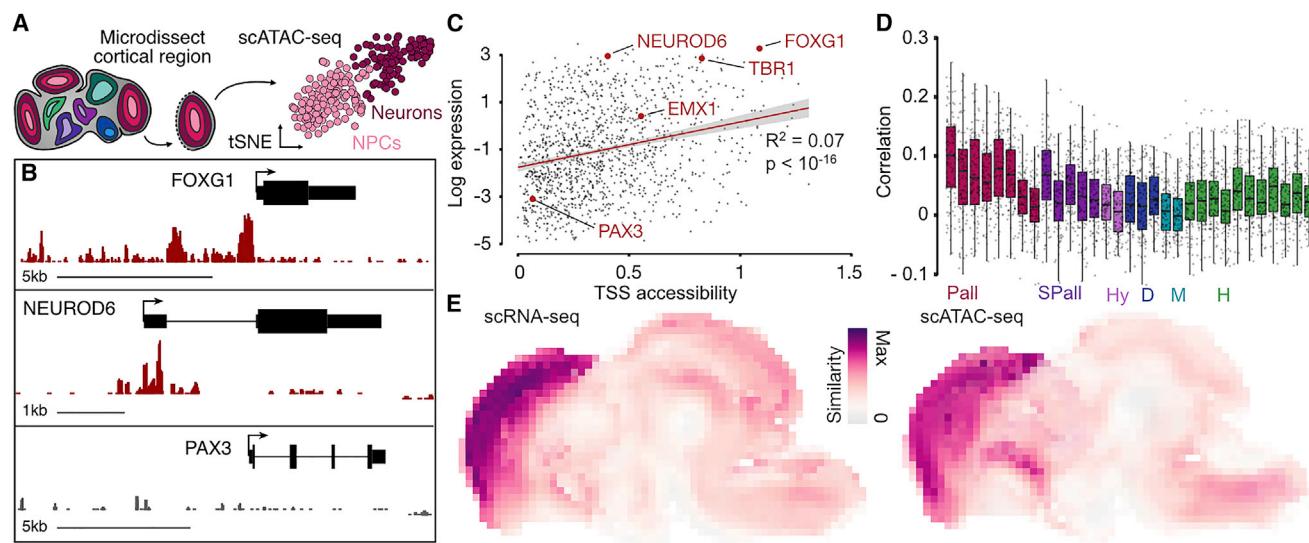


Figure 5. Substructure resolution of transcriptomic maps (original)

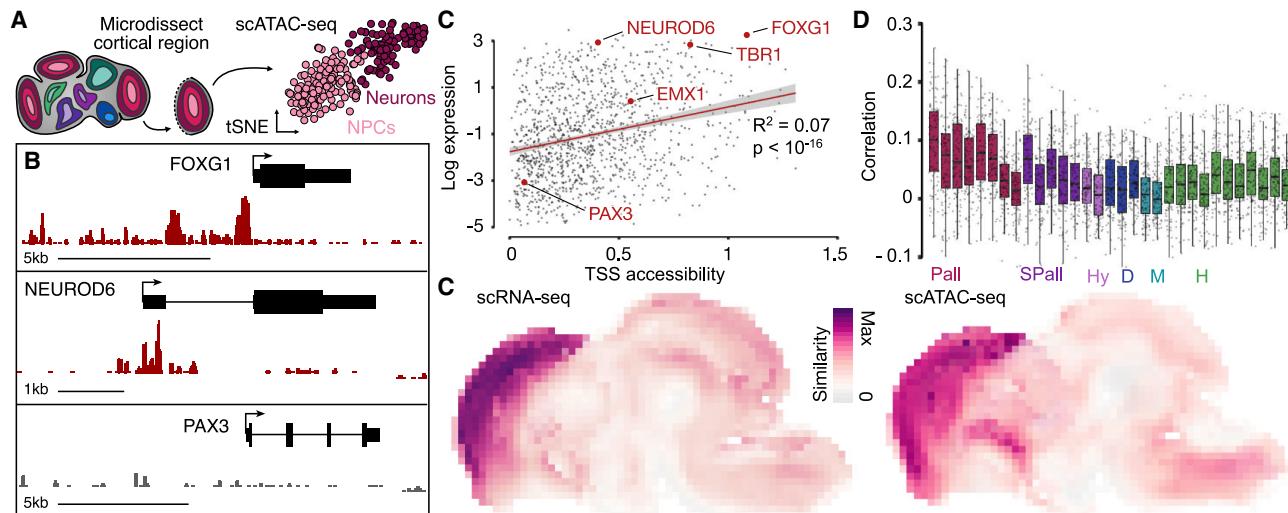


Figure 6. Projecting chromatin accessibility profiles to spatial brain maps (corrected)

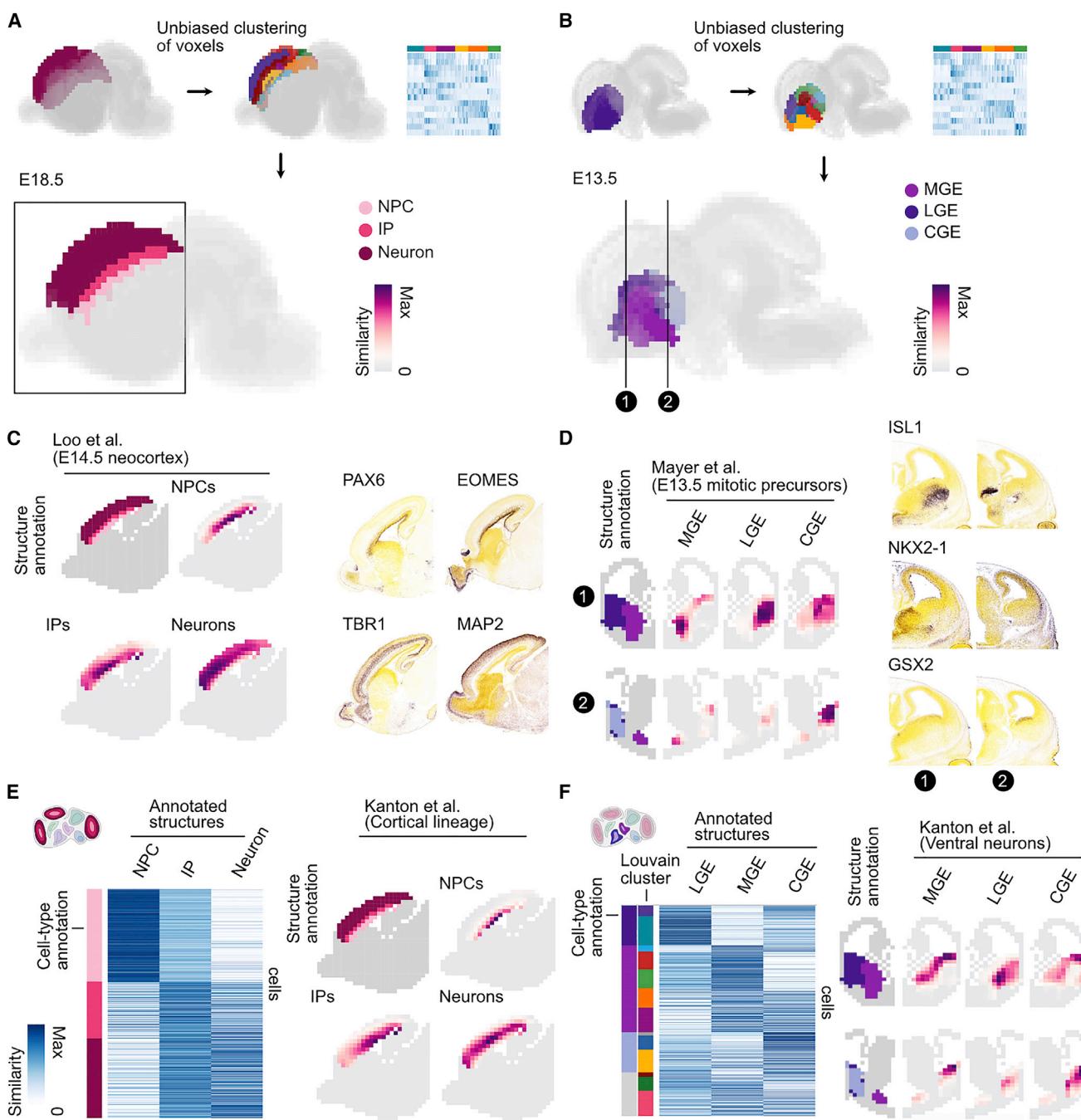


Figure 6. Projecting chromatin accessibility profiles to spatial brain maps (original)