Spatial profiling of the interplay between cell type- and vision-dependent transcriptomic programs in the visual cortex

Fangming Xie,[1,6,*] Saumya Jain,[1,5,6] Runzhe Xu,[1,6] Salwan Butrus,[2] Zhiqun Tan,[4] Xiangmin Xu,[4] Karthik Shekhar,[2,3,*] and S. Lawrence Zipursky[1,7,*]

[1]Department of Biological Chemistry, Howard Hughes Medical Institute, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA
[2]Department of Chemical and Biomolecular Engineering; Helen Wills Neuroscience Institute, California Institute for Quantitative Biosciences (QB3); Center for Computational Biology; Vision Sciences Graduate Program, University of California, Berkeley, Berkeley, CA 94720, USA
[3]Faculty Scientist, Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[4]Center for Neural Circuit Mapping (CNCM), Department of Anatomy and Neurobiology, Institute for Memory Impairments and Neurological Disorders (UCIMIND), University of California, Irvine, Irvine, CA 92697, USA
[5]School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA 30332, USA

[6]These authors contributed equally.
[7]Lead contact
*Correspondence: fmxie@ucla.edu (F.X.), kshekhar@berkeley.edu (K.S.), lzipursky@mednet.ucla.edu (S.L.Z.)

Abstract

How early sensory experience during “critical periods” of postnatal life affects the organization of the mammalian neocortex at the resolution of neuronal cell types is poorly understood. We previously reported that the functional and molecular profiles of layer 2/3 (L2/3) cell types in the primary visual cortex (V1) are vision-dependent (Tan et al., *Neuron*, **108**(4), 2020; Cheng et al., *Cell*, **185**(2), 2022). Here, we characterize the spatial organization of L2/3 cell types with and without visual experience. Spatial transcriptomic profiling based on 500 genes recapitulates the zonation of L2/3 cell types along the pial-ventricular axis in V1. By applying multi-tasking theory (Adler et al., *Cell Systems*, **8**, 2019), we suggest that the spatial zonation of L2/3 cell types is linked to the continuous nature of their gene expression profiles, which can be represented as a 2D manifold bounded by three archetypal cell types (“A”, “B”, and “C”). By comparing normally reared and dark reared L2/3 cells, we show that visual deprivation-induced transcriptomic changes comprise two independent gene programs. The first, induced specifically in the visual cortex, includes immediate-early genes and genes associated with metabolic processes. It manifests as a change in cell state that is orthogonal to cell type-specific gene expression programs. By contrast, the second program impacts L2/3 cell type identity, regulating a subset of cell type-specific genes and shifting the distribution of cells within the L2/3 manifold, with a depression of the B-type and C-type and a gain of the A-type. Through an integrated analysis of spatial transcriptomic measurements with single-nucleus RNA-seq data from our previous study, we describe how vision patterns L2/3 cortical cell types during the postnatal critical period.

Significance statement

Layer 2/3 (L2/3) glutamatergic neurons are important sites of experience-dependent plasticity and learning in the mammalian cortex. Their properties vary continuously with cortical depth and depend upon experience. Here, by applying spatial transcriptomics and different computational approaches in the mouse primary visual cortex, we show that vision regulates orthogonal gene expression programs underlying cell states and cell types. Visual deprivation not only induces an activity-dependent cell state, but also alters the composition of L2/3 cell types, which are appropriately described as a transcriptomic continuum. Our results provide insights into how experience shapes transcriptomes that may, in turn, sculpt brain wiring, function, and behavior.

Introduction

Early sensory experiences influence the development of neural circuitry throughout the mammalian brain (1–3). In the primary visual cortex (V1), visual experience is required to establish the circuitry for binocular vision (4, 5). Vision-dependent development of binocular circuits occurs in mice after eye-opening (postnatal day (P) 14). There are only a few binocular L2/3 neurons at eye opening and their number increases during the ensuing week. From P21 to P35, the tuning properties of these neurons improve markedly in a vision-dependent process. This developmental time window is referred to as the critical period (4, 6–8). Classically, the influence of early visual experience has not been examined at the level of the many cell types that form the building blocks of V1 circuitry. To address this gap, we previously performed single-nucleus RNA-seq (snRNA-seq) in normal- and dark-reared (NR, DR) mice during postnatal development, including the critical period (9). Using computational methods, we reconstructed the postnatal maturation of V1 cell types in NR mice and compared these profiles with DR mice. This enabled us to identify the cell types and gene expression programs impacted by vision. For most cell types, DR had little impact on molecular identity. However, DR had a pronounced effect on glutamatergic cell types in the supragranular layers (L) 2/3/4. In NR

mice, L2/3 glutamatergic neurons comprise three molecularly distinct types (A, B, and C). These L2/3 types occupy three partially overlapping sublayers along the pial-ventricular axis of L2/3 (upper, middle, and lower, respectively) based on the *in situ* expression patterns of three marker genes. DR altered the expression patterns of these marker genes, and a comparison of transcriptome-based clusters in DR did not correspond to NR types. Based on these observations, we concluded that visual deprivation during the critical period selectively disrupted L2/3 cell types. Furthermore, visual deprivation regulated large groups of genes, including determinants of cell adhesion and synapse formation. However, as clusters in DR could not be matched to NR types, we could not separate vision-dependent changes in cell state from vision-dependent changes in cell type, their organization, or both.

Here, we combined spatial transcriptomics and computational analyses to map the organization of L2/3 cell types in V1, and to assess the impact of visual deprivation on this organization. The gene expression changes due to visual deprivation comprise two independent transcriptomic programs. The first program is upregulated across multiple cell types throughout the visual areas and represents an activity-dependent cell state. The second program regulates a subset of cell type-specific genes within L2/3, and its impact on cell type composition can be interpreted using multi-tasking theory (10), a recently proposed framework for analyzing gene expression continua. Thus, visual deprivation affects both activity-dependent transcriptomic states and cell-type composition in L2/3.

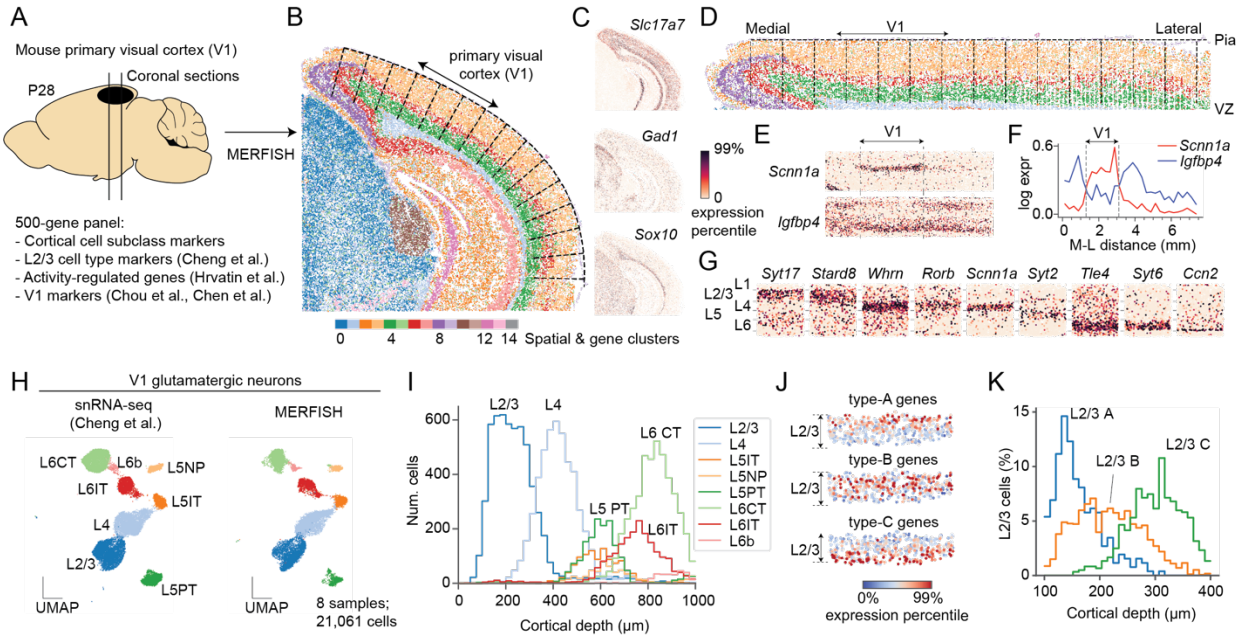


Figure 1. MERFISH recapitulates the spatial organization of L2/3 glutamatergic neurons in V1. (A) MERFISH was performed on mouse coronal brain sections using a panel of 500 genes. The panel was designed to resolve cell types and cell states in V1 based on published studies (9, 11–13). **(B)** An overview of the MERFISH data, showing a thin (10 μm) coronal section including the neocortex, parts of the hippocampus, and midbrain structures. Individual cells are colored by their cluster membership based on both gene expression similarity and spatial proximity (see **Figure S1A**). **(C)** *In situ* gene expression patterns of *Slc17a7,* a glutamatergic neuron marker; *Gad1,* a GABAergic neuron marker; and *Sox10*, an oligodendrocyte marker. Colors represent gene expression levels going from 0 to 99 percentile. **(D)** Cortical cells with defined locations along the medial-lateral (M-L) and pial-ventricular (P-V) axis of the cortex. Tissue was straightened *in silico* to facilitate downstream analysis. Cells are colored as in panel B. **(E)** Two genes, *Scnn1a* and *Igfbp4* with area-specific *in situ* signatures. **(F)** Expression levels of genes in

panel E along the M-L axis. The location of V1 is highlighted. **(G)** Examples of genes with layer-specific signatures in V1. **(H)** Subclasses of V1 glutamatergic neurons are represented in UMAP embeddings obtained from integrating snRNA-seq (left panel; (9)) and MERFISH (right panel; this study) using Harmony (14). **(I)** Spatial distributions of glutamatergic subclasses along the cortical depth (P-V) axis. **(J)** Aggregated expression in V1 L2/3 for gene groups defining types A (n=64 genes), B (n=35 genes) and C (n=71 genes). Type A, B, and C genes are distributed in upper, middle, and lower L2/3, respectively. Colors represent gene expression levels going from 0 to 99 percentile. **(K)** Spatial distributions of types A, B, and C along the cortical depth spanning L2/3. MERFISH cell labels were transferred from previously published snRNA-seq data (9).

Results

Spatial transcriptomic analysis of V1 reveals the continuous sublayered arrangement of L2/3 glutamatergic cell types.

We first sought to explore the spatial arrangement of L2/3 cell types. To do this, we performed multiplexed error-robust fluorescence *in situ* hybridization (MERFISH (15)) on brain coronal sections containing V1 obtained from normally-reared (NR) mice at postnatal day (P) 28 (**Figure 1A**). MERFISH is an imaging-based approach that allows simultaneous mapping of hundreds of genes in individual cells within an intact tissue section. Using computational methods, cells can be classified into types based on combinatorial gene expression patterns, and the spatial arrangement of each cell type can be visualized. For these experiments, we selected a panel of 500 genes that included cortical area- and layer-specific markers, marker genes for all major cortical cell populations (9, 13), and activity-regulated genes (11) (**Tables S1, S2**). Notably, given our specific goal of studying cell type-specific changes within L2/3, we included 170 genes selected from the 286 L2/3 cell-type-identity genes we previously reported (9) (see **Table S1**).

We performed a preliminary clustering of the cells from the entire section, defining clusters using a combination of molecular similarity and spatial proximity (**Methods**). This analysis identified 15 cell clusters, which highlighted different high-level anatomical structures of the mouse brain, including the neocortex with its characteristic laminar structure, parts of the hippocampus, and midbrain structures (**Figures 1B, S1A**). These associations were supported by *in situ* expression patterns of well-known genes. For example, *Slc17a7* (*Vglut1*), a marker for glutamatergic neurons, is prominently expressed in the cortex and parts of the hippocampus, two regions where glutamatergic neurons comprise 80~90% of all neurons (16–18) (**Figure 1C**). In contrast, *Gad1*, a marker for GABAergic neurons, is expressed at low levels in the cortex but high levels in the midbrain, where inhibitory neurons represent ~50% of neuronal cells (17, 18). *Sox10*, a marker for oligodendrocytes, is enriched in the white matter beneath the cortex.

We used the pial surface to define the tangential (medial-lateral) and vertical (pial-ventricular) coordinates of each cell and demarcated areas of the neocortex (**Figures 1B-G, S1B**). We localized the primary visual cortex (V1) based on the enrichment of *Scnn1a* and the depletion of *Igfbp4* along the tangential axis. Both markers, described in recent reports (12, 13), allowed us to identify V1 as a ~2mm wide x 1mm deep region (**Figures 1E,F**).

We used an integrative approach to classify each V1 cell in MERFISH into one of three classes (excitatory neurons, inhibitory neurons, and non-neuronal cells) and each neuron into one of 12 subclasses, as in Cheng et al. (9) (**Figures 1H, S1C-E**). The relative frequencies of the subclasses tightly corresponded between MERFISH and snRNA-seq (Spearman correlation = 0.95), and the different subclasses of glutamatergic neurons were spatially localized within expected layers along the pial-ventricular axis (**Figures 1I, S1C-F**).

Previously, we found that L2/3 cells can be clustered into three types (L2/3A, L2/3B, and L2/3C). Individual marker genes for these types – *Cdh13* for L2/3A, *Trpc6* for L2/3B, and *Chrm2* for L2/3C - were expressed in upper, middle, and lower regions of L2/3 along the pial-ventricular axis (9). However, the MERFISH data allowed us to validate the three-layered zonation of L2/3 types using a much larger gene panel (**Figures 1J, S2**). Combinatorial gene signatures based on 170 type-identity genes enabled classification of MERFISH L2/3 cells as L2/3A, L2/3B, or L2/3C. As expected, soma of the three cell types localized to the upper, middle, and lower sublayers within L2/3 (**Figure 1K**). Type A and C identity genes were expressed in a more restricted spatial pattern, whereas type B identity genes were expressed more broadly.

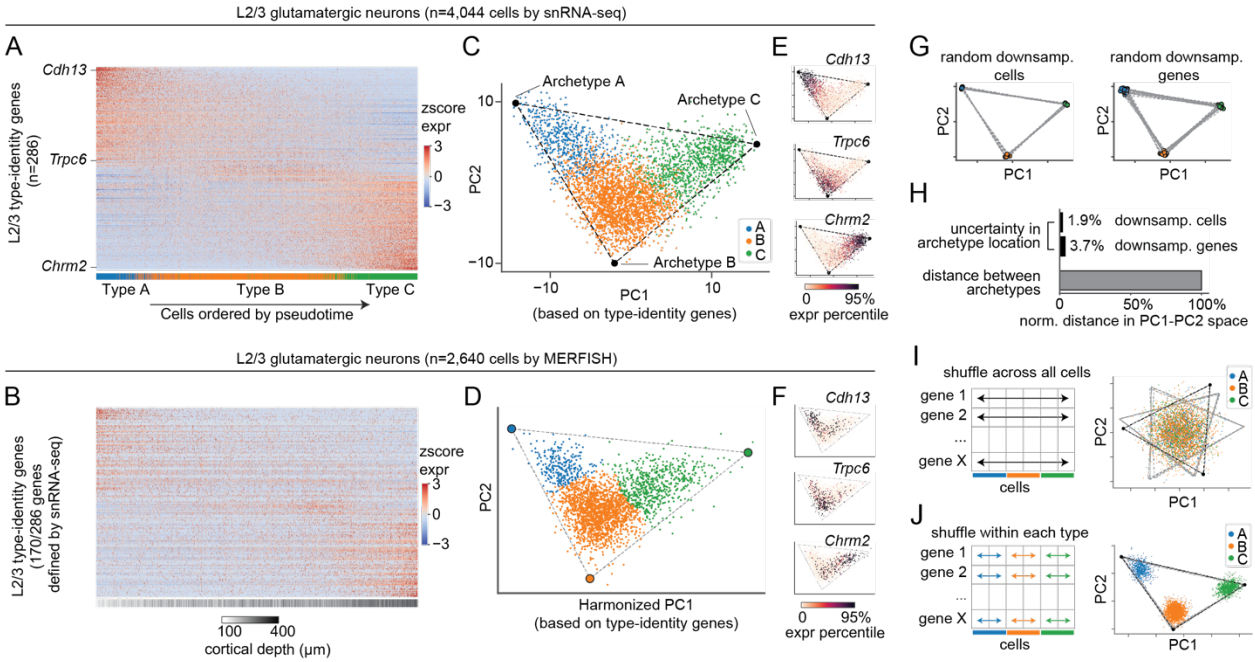


Figure 2. L2/3 glutamatergic neurons form a continuous manifold bounded by a triangle, whose vertices represent archetypes. (A-B) Expression heatmaps of type-identity genes (**Table S1**) across L2/3 glutamatergic neurons for snRNA-seq (A) and MERFISH (B). Expression was quantified as z-scored, log- and size-normalized counts. Cells (x-axis) were ranked by diffusion pseudotime (DPT) and colored by cell-type identity in (A) and cortical depth in (B) (annotation bar, bottom). **(C-D)** L2/3 cells represented in PC1 and PC2 space based on the same type-identity genes in (A-B). Cells are colored by type. The bounding triangle was inferred using computational procedures described in Adler et al. (10) (see Methods). **(E-F)** Expression levels of *Cdh13*, *Trpc6,* and *Chrm2* in PC1/PC2 space for scRNA-seq (E) and MERFISH (F). **(G-H)** Stability of the bounding triangle over randomized trials involving random down-sampling of the original dataset to 80% of cells (G, left panel) or 80% of genes (G, right panel). Comparison % variation in archetype coordinates during randomizations relative to the typical inter-archetype distances in these trials. Distances are measured by Euclidean distance in PC1-PC2 space. (H). **(I-J)** L2/3 cells in PC1/PC2 space after shuffling each gene independently across all cells (I), and after shuffling each gene independently across cells within each type (J), respectively. The gray lines represent triangular fits using 80% of cells, randomly down-sampled ten times independently. Thus, these cells form a continuum of types rather than three noisy cell types (see Figure S4). Panels G-J used snRNA-seq data only.

Multi-tasking theory relates the spatial zonation of L2/3 glutamatergic neuronal types to their continuous transcriptomic variation

The gene expression patterns that underlie the transcriptomic continuum are most evident when we analyzed the data using diffusion pseudotime (DPT) (19), an approach based on diffusion maps (20). For snRNA-seq, DPT orders cells according to their type identity along the x-axis and their corresponding type-identity genes along the y-axis (**Figure 2A**). The genes are expressed at different levels in a graded fashion, with no clear boundaries between the domains of each cell type. DPT produces a similar result for the MERFISH dataset, with cells now ordered based on their location along the cortical depth **Figure 2B**), further supporting the link between continuous transcriptomic variation and spatial zonation.

Strikingly, such a link is predicted by a recently proposed framework known as multi-tasking theory (10). This theory posits that *"if gene expression lies on a D-dimensional manifold (e.g., D=3 for a tetrahedron, D=2 for a triangle, or D=1 for a line), the cells should show at least a D-dimensional spatial zonation [in tissue]"* (10). Applying this framework to our data, we find that L2/3 neurons lie on a V-shaped manifold along the first two principal components (PCs), which is bounded by a triangle (**Figure 2C; Methods**). The structure of the manifold and the coordinates of the bounding triangle are similar for both the snRNA-seq and the MERFISH datasets (**Figures 2D, S3**). According to multi-tasking theory, vertices of the triangle represent "archetypal" cell types, with the gene expression profile of each archetype specialized for a particular function (see below). Indeed, the three vertices map to the L2/3 A, B and C types identified by clustering, and cells inside the triangle are mixtures of the archetypal gene expression profiles (**Figure 2E,F**). Finally, consistent with the theory, the organization along the manifold from A to B to C is mirrored in the spatial zonation along the pial-ventricular axis (**Figure 1J,K**).

We performed additional analyses to assess the robustness of the V-shaped manifold and the bounding triangle. Repeating the analyses with subsets of cells or genes resulted in only minor perturbations of the bounding triangle (**Figure 2G,H**). Repeating the analyses with shuffled versions of the original gene expression matrix further substantiated the continuous nature of the transcriptomic variation. First, we shuffled gene expression values independently for each gene across L2/3 cells, which preserves individual gene expression distributions while disrupting their correlations. This resulted in all cells collapsing towards the center and a total disruption of the triangular structure ($p<0.001$; t-ratio test) (**Figure 2I, S4**). We then shuffled gene expression values independently for each gene within each cell type (**Figure 2J**). This procedure distinguishes between two scenarios: a genuine continuum vs. discrete types seemingly continuous due to noise in the data, which could arise from the intrinsic stochasticity of gene expression or sampling noise in single-cell sequencing (21). When applied to clusters that span a continuum, the shuffling procedure splits continua into discrete clusters as long as the level of noise in the data are low to intermediate (**Figures S5A,C,E**). In contrast, when the clusters are already discrete it has no effect regardless of noise level (**Figures S5B,D,F**). Applying this shuffling procedure to our data splits L2/3 cells into three clusters (**Figure 2J**), supporting that transcriptomic continuum is genuine and not an artifact of noise.

What is the functional significance of this continuum? According to multi-tasking theory, such transcriptomic continua arise as an optimal solution for division of labor among cells in tissues where the a cell's functional identity is related to its spatial position. While the theory does not specify these functions, examining the genes that define the archetypes may provide some clues. By performing gene ontology (GO) analyses, we found that archetype-defining genes are enriched for biological processes related to neuronal development and functions (**Figure S6**).

Programs involved in neural wiring, including “axon guidance” and “cell-cell adhesion,” show up in all three types (A, B, and C). For example, among known cell-recognition molecules related to neural wiring, type A expresses *Cdh13, Cntn5, Epha6, Sema6a* and *Robo1*; type B expresses *Epha3* and *Sema4a;* and type C expresses *Cdh12, Cntn2* and *Robo3* (**Table S3**). The differential expression of these genes may be related to the distinct projection patterns of A, B and C types to the higher visual areas (9, 22). Other programs, such as “sensory perception of light stimulus” and “detection of mechanical stimulus involved in sensory perception,” are specific to types B and C, respectively. Although one can only speculate on the functional meaning at this stage (see **Discussion**), the manifold representation of L2/3 neurons proves useful in understanding global and cell type-specific changes due to visual deprivation, which we analyze in the next section.

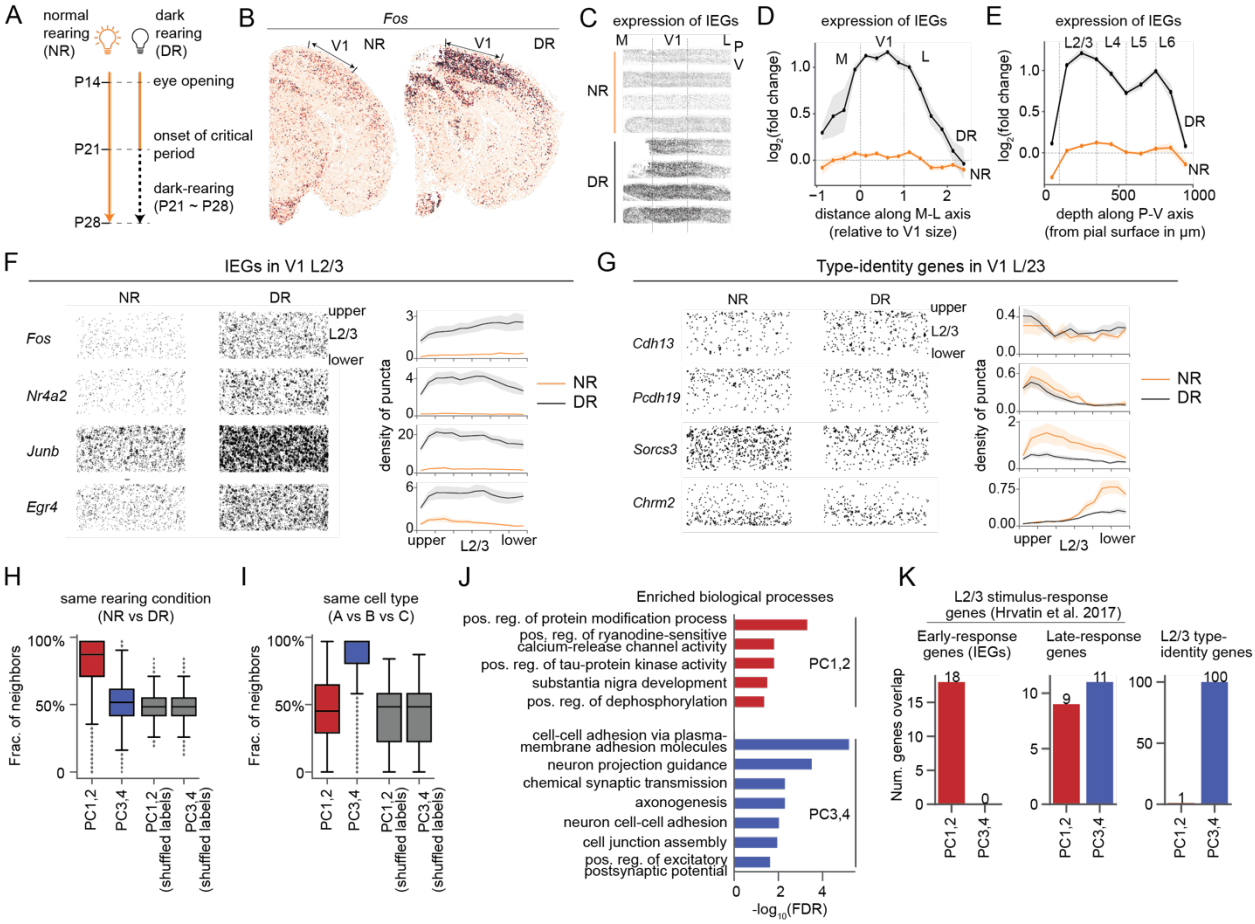


Figure 3. Gene programs driving transcriptomic heterogeneity in normal- and dark-reared L2/3 neurons. (A) Protocol for dark-rearing (9). **(B)** *In situ* expression patterns of *Fos* in NR and DR coronal sections. Locations of V1 are highlighted. **(C)** Aggregated *in situ* expression of IEGs (11) in NR and DR in V1 and its flanking cortical regions. Rows in each condition correspond to biological replicates. **(D-E)** Aggregated expression of IEGs along the M-L axis (D) and the P-V axis (E). Fold change is quantified relative to the mean expression level over the entire V1 region for all NR samples. **(F,G)** *In situ* expression of representative IEGs (F) and type-identity genes (G). Individual dots are transcript molecules detected by MERFISH. Line plots show the distribution of puncta along the P-V axis from upper to lower L2/3. The density of puncta were defined as the number of puncta per 100 μm^2 area and were averaged across samples. Error bars represent standard error of the mean. **(H-I)** Distribution of L2/3 cells' neighbor identities in terms of the the fraction of k nearest neighbors (k=30) that are of the same rearing condition (NR vs DR; panel H) and of the same type (A vs B vs C; panel I). PCs are computed from n=6,360 HVGs (see Methods and Figure S9). **(J)** Gene ontology (GO) analysis shows biological processes enriched in

genes driving PC1-PC2 (in red) and PC3-PC4 (in blue). Raw results from the enrichment analysis were filtered to remove redundant terms (see Methods). A full list of the top 10 significant GO terms with FDR < 0.05 is shown in **Figure S9E**. Top 100 genes for each PC ranked by the absolute value of PC loadings are used. **(K)** Overlap between PC-driving genes and previously defined gene groups. Activity-regulated genes, including early response genes (immediate early genes - IEGs) and late-response genes in L2/3 glutamatergic neurons are from Ref (11). Type-identity genes from Ref (9) are listed in Table S1.

Transcriptomic heterogeneity in vision-deprived L2/3 neurons involves two distinct gene programs

To characterize the changes induced by visual deprivation, we obtained MERFISH data from mice dark-reared (DR) between P21-P28, following our published experimental protocol (**Figure 3A; Methods**) (9). Comparisons of DR and NR sections collected at P28 showed an upregulation of 13 genes (FC>2, FDR < 0.05, **Figure S7A**), a small subset of vision-dependent genes detected by snRNA-seq (see below). All 13 genes were canonical immediate early genes (IEGs) (11), and their upregulation was localized to the visual areas (**Figure 3B**). Along the tangential axis, IEG levels were significantly higher in V1 than in its flanking regions (**Figures 3C,D, S7B**). Consistent with prior snRNA-seq data, we observed that the up-regulation of IEGs was broadly shared by all neuronal subsets and, therefore, visible throughout the cortical depth (**Figures 3C,E, S7C**). Within L2/3, canonical IEGs, such as *Fos*, *Nr4a2, Junb and Egr4*, were up-regulated in DR similarly in all sublayers (**Figures 3F, S8A,C**).

We previously reported that DR selectively disrupts the transcriptomic signatures of L2/3 cell types in single-nucleus (sn) RNA-seq data from V1 (9). Our results were based on unsupervised clustering of snRNA-seq profiles from NR and DR mice. Most cell types correspond 1:1 between NR and DR with the notable exception of L2/3 neurons. In the MERFISH data, while type-identity genes by and large maintained their sublayered expression, they also exhibited some vision-dependent alterations in DR. L2/3A-specific genes were relatively unchanged, but L2/3B- and L2/3C-specific genes were downregulated (**Figures 3G, S8B-D**). Taken together, **Figures 3A-G** suggest that vision impacts L2/3 transcriptomes in two distinct modes, one comprising broadly expressed genes (e.g. IEGs), and the other impacting genes associated with cell type identity (**Figure S8E,F**).

These results motivated a re-analysis of our published snRNA-seq dataset (9). We combined NR and DR datasets at P28 and applied principal component analysis (PCA) using 6,360 highly variable genes (HVGs; see Methods). We focused on the top four PCs, which exhibited a clear spectral gap from other components (**Figure S9A**). PC1 and PC2 separated cells by rearing condition (NR vs. DR mice), while NR and DR cells intermixed in the PC3-PC4 space (**Figures 3H, S9B-C**). Conversely, PC3-PC4, but not PC1-PC2, separated NR cells by type identity (**Figures 3I, S9B-C**), recapitulating the triangular manifold in **Figure 2B**, which was estimated using 286 type-identity genes. Thus, transcriptomic variation in L2/3 glutamatergic neurons is driven by two gene programs: the first, encoded in PC1-PC2, captures vision-dependent changes, while the second, encoded in PC3-PC4, captures cell type identity.

As PCs are derived from orthogonal directions in transcriptomic space, we hypothesized that the gene programs encoded in PC1-PC2 vs. PC3-PC4 represent distinct biological processes. Gene ontology analysis shows that genes driving PC1 and PC2 were enriched in the regulation of protein modification and signaling pathways (**Figures 3J, S9E**), likely reflecting shifts in cell state due to experience-dependent activity. By contrast, PC3 and PC4 were enriched for genes associated with cell-cell adhesion, axonogenesis, neuron projection guidance, and chemical synaptic transmission (**Figures 3J, S9E, Table S4**), which are associated with neuronal cell-

type identity (23–26). We also compared the top-loading genes within each PC with a list of known activity-regulated genes in L2/3 glutamatergic neurons from Hrvatin et al. (11). This list includes 42 “early-response” genes, among which many are canonical IEGs, that were conserved across cell types in the original study, and 37 “late-response” genes, which were largely cell type-specific (**Table S5**). PC1-PC2 genes strongly overlapped with IEGs (odds ratio = 25, $p<10^{-10}$; Fisher’s exact test; significant enrichment) and barely overlapped with type-identity genes (odds ratio = 0.1, p=0.002; Fisher’s exact test; significant depletion) (**Figures 3K**). By contrast, PC3-PC4 genes did not contain IEGs and significantly overlapped with type-identity genes (odds ratio = 40, $p<10^{-10}$). Late-response genes featured in equal numbers across PC1-PC2 and PC3-PC4.

These results highlight the consistency between the MERFISH and the snRNA-seq datasets and present the following overall picture. The transcriptomic variance in NR and DR layer 2/3 cells can be decomposed into two independent programs: (1) PC1-PC2, associated with experience-regulated cell states, and (2) PC3-PC4, associated with cell type identity. However, we find that the contribution of these programs to the overall variance within a dataset differs by rearing condition. For NR, the projected variance along PC3-PC4 is 2.9-fold higher than that along PC1-PC2. By contrast, for DR, the projected variance along PC1-PC2 is 2.8-fold higher than along PC3-PC4. This disparity, however, drives the lack of correspondence in the HVG-based clustering of NR and DR L2/3 neurons reported in Cheng et al (9). Consistent with this, a “focused” clustering of DR L2/3 neurons using the 286 type-identity genes recovers cell clusters corresponding 1:1 with types A, B and C in NR (**Figure S10**). Overall, this suggests that despite being masked by experience-dependent changes, cell identity programs persist in DR (**Figure S9D**), albeit with major alterations (**Figure 3G**). We now analyze these alterations.

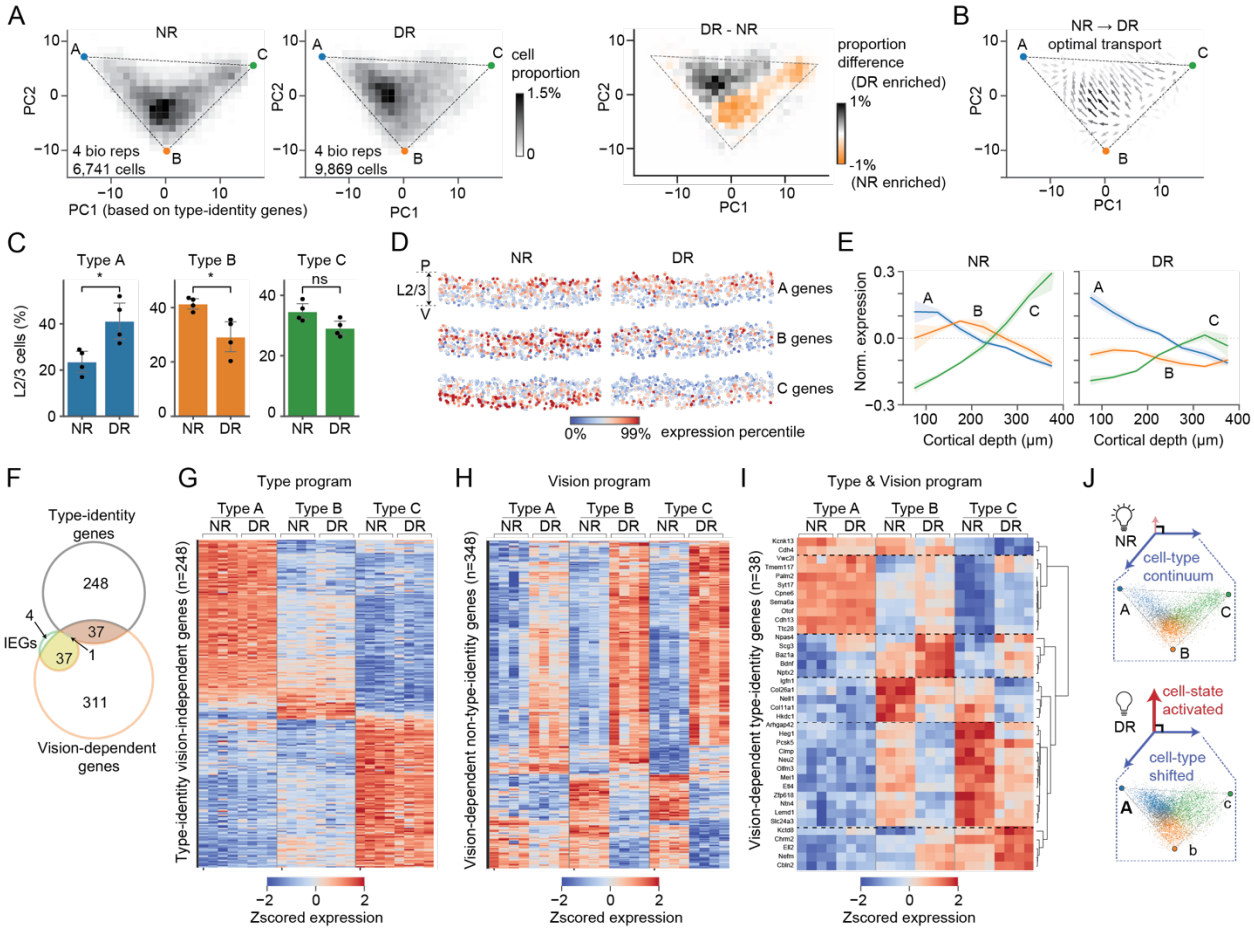


Figure 4. Visual deprivation redistributes cells within the L2/3 type triangle. (A) Distribution of L2/3 types within the L2/3 type-triangle. Cell density for NR cells (left), DR cells (middle), and the difference between NR and DR (right) are shown in PC1 and PC2 space derived from type-identity genes. (B) Optimal transport analysis (27, 28) showing the transport map connecting the NR to DR distribution. The arrows indicate the direction of local redistribution of cells in the reduced gene expression space. (C) The proportions of L2/3 cells assigned as types A, B and C for NR and DR samples. The type assignment is based on type A, B and C identity gene expression (see Methods). (D) *In situ* expression levels of A, B and C type-identity genes for NR and DR samples. (E) Expression of A, B and C identity genes along the P-V axis from upper to lower L2/3 for NR (left panel) and DR (right panel) samples. Mean expression levels were shown relative to the baseline levels, which are defined as the mean expression levels in NR samples averaged across L2/3. (F) Overlap between vision-dependent genes, type-identity genes and previously identified L2/3 IEGs (11). (G-I) Expression profiles of type-identity vision-*in*dependent genes (G), vision-dependent *non*-type-identity genes (H), and vision-dependent type-identity genes (I). Expression levels are z-score normalized across all samples independently for each gene. Gene numbers correspond to panel (F). (J) Diagram shows L2/3 transcriptomes comprise two orthogonal axes of variations – cell-state programs and cell-type programs. Visual deprivation activates cell-state programs and shifts cell-type programs away from archetypes B and C towards A.

Visual deprivation alters the composition of L2/3 cell types along the transcriptomic continuum

We used the 2D manifold representation (**Figure 2C,D**) to study cell type-specific changes induced by visual deprivation. We hypothesized that vision could regulate one or more of the following features: a) the locations of the archetypes, i.e., the bounding triangle; b) the

352 distribution of cells along the manifold, which correspond to changes in cell type composition
353 and spatial zonation; and c) the expression levels of one or more type-identity genes.
354
355 To evaluate whether archetype coordinates change, we recalculated PCs based on the 286
356 type-identity genes as in **Figure 2C**, but now using both NR and DR cells from the published
357 snRNA-seq data. Using this representation, we separately inferred archetype coordinates using
358 cells from each biological sample (n=4 replicates for each condition, including P38 data from ref
359 (9)). These calculations verified that L2/3 cells from NR and DR mice occupy similar triangular
360 regions, with no significant changes detected in archetype coordinates (3~10% changes relative
361 to the minimum distance between archetypes) (**Figure S11A,C**).
362
363 Individual cells, however, distributed differently between NR and DR, shifting away from the
364 region proximal to archetype B towards archetype A, and away from archetype C in DR (**Figure**
365 **4A; Figure S11B**). This reorganization between NR and DR was also predicted by two
366 alternative methods: optimal transport analysis (27, 28) (**Figure 4B; Methods**) and supervised
367 classification-based label transfer from NR to DR cells (**Figure 4C**). The compositional
368 difference between NR and DR at P28 or P38 was larger than the difference between P28NR
369 and P38NR, or between P28DR and P38DR (**Figure S11D**). This suggests that the effect of DR
370 on L2/3 transcriptomic identities is consistent over at least ten days. Furthermore, similar
371 changes in cell type composition were also recovered in the MERFISH data (**Figure S12**).
372 Additionally, MERFISH shows that the overall expression levels of type B and C genes
373 decrease in DR, especially in middle and lower L2/3 (**Figure 4D,E; Figure S13**). By contrast,
374 the expression of type A genes is relatively unchanged and consistently enriched in upper L2/3.
375
376 We performed differential gene expression analysis between NR and DR samples to identify
377 vision-dependent genes within each type. In total, we identified 386 unique vision-dependent
378 genes, of which 70 were regulated in type A, 226 in type B, and 304 in type C (fold change >2
379 and FDR <0.05). These genes included 90% (n=38/42) of previously identified L2/3 IEGs (11),
380 while the IEGs represent ~10% (n=38/386) of the vision-dependent genes (**Figure 4F**).
381 Moreover, most vision-dependent genes (90%; n=348/386) are not type-specific, and they were
382 up- or down-regulated in all or two of the three L2/3 types in DR (**Figure 4F-H**). Conversely,
383 most type-identity genes (87%; n=248/286) were vision-independent, consistently marking types
384 A, B, and C in NR and DR (**Figure 4F,G**). Of 286 type-specific genes 38 (13%) were vision-
385 dependent (**Figure 4F,I**). These *vision-dependent type-identity* genes fell into six groups based
386 on their patterns of regulation (**Figure 4I**). Overall, type A had smaller fold changes in DR
387 compared with types B and C in almost all genes. Unlike types B and C-specific genes, a group
388 of type A genes (Group 2; n=9) were up-regulated in types B and C, but did not change in type
389 A themselves. The changes in individual type-identity genes were consistent with the overall
390 trend that type A was less affected by DR than types B and C.
391
392 Together, these results exemplify the plasticity of the transcriptional programs defining A, B and
393 C cell types, resulting in a redistribution of cells in the 2D manifold associated with type identity
394 due to visual deprivation (**Figure 4J**). The nature of this redistribution is consistent across
395 biological replicates of snRNA-seq (**Figure S11**) and MERFISH (**Figure S12**). Finally, L2/3
396 glutamatergic neurons (also known as L2/3 intra-telencephalic neurons; L2/3 IT) are more
397 sensitive to vision than deeper-layer intra-telencephalic (L5/6 IT) neurons, consistent with prior
398 results (9) (**Figure S14**).

Discussion

We previously showed that vision deprivation selectively impacts L2/3 glutamatergic neuronal types in V1 (9). Using unsupervised clustering of snRNA-seq data, we found that L2/3 neuronal clusters in DR mice "*poorly resembled the three types in NR animals, and the expression patterns of cell-type-specific marker genes were disrupted*". We interpreted these results as "*a global disruption of gene expression patterns throughout L2/3*", resulting in "*the loss of [L2/3] cell type identity in animals deprived of light*". The nature of the disruption in gene expression was, however, not clear due to our inability to relate cell types between the two conditions.

Here, combining spatial transcriptomics with new computational analyses enabled us to relate L2/3 cell types between NR and DR and thereby clarify the vision-dependent gene expression changes. Visual deprivation impacts two orthogonal gene programs, one associated with experience-induced cell states and the other overlapping with programs of cell type identity. The deprivation-induced state changes dominate the transcriptomic variation in DR, masking cell type distinctions in the unsupervised clustering analysis used previously.

Using multi-tasking theory, we showed that the transcriptomic variation among L2/3 cells can be represented as a continuous manifold in 2D bounded by a triangle whose vertices represent archetypes A, B, and C. These archetypes represent the "basis sets" used to construct the profiles of sampled L2/3 neurons. While preserving the archetypes themselves, dark-rearing redistributes cells within the manifold, shifting cells away from archetypes B and C. Thus, L2/3 cell types are not lost in animals deprived of light, but their composition is altered, and the underlying gene expression programs are masked by orthogonal activity-induced changes. These changes are consistently reproduced in both MERFISH and snRNA-seq datasets.

A key tenet of multi-tasking theory is to link transcriptomic continua to the spatial zonation of functionally distinct cell types. Previous observations indicate that neurons within different L2/3 sublayers (A, B, C) differ in their functional properties and projections to higher visual areas (9, 11, 29). Our results support the idea that vision is required to maintain the proper distribution of cells in the L2/3 cell-type continuum and that vision preferentially impacts types B and C more than type A. The unequal dependence on vision could be related to the spatial zonation, as previous studies have shown that deep, but not more superficial, L2/3 cells receive LGN inputs (30)**.** Consistent with this, our GO analyses suggest that types B and C but not type A genes are enriched in biological processes associated with sensory input (**Table S3**)**.**

Previous studies have shown that neural activity influences the transcriptomes of the developing neurons (9, 11, 13, 31, 32), with L2/3 glutamatergic neurons being the cortical cell types that are most sensitive to sensory experience (13, 32), which is in agreement with functional studies (33, 34). Several studies have highlighted the regulation of IEGs under sensory deprivation, although the direction of regulation may depend upon the experimental protocol (9, 11, 32). Our data, based on both snRNA-seq and MERFISH, suggests dark-rearing itself leads to a massive up-regulation of IEGs and the effect is specific to visual areas. This area-specificity suggests the upregulation of IEGs is unlikely to be an experimental artifact. In addition, known IEGs only represent about 10% of genes altered by dark-rearing (**Figure 4F**).

How the interplay between intrinsic and vision-dependent molecular programs establish the transcriptomic continuum of L2/3 glutamatergic neurons, and how this sculpts circuitry and visual perception are exciting and challenging questions. Emerging techniques correlating transcriptomics, neural activity, and/or connectivity at the single neuron level provide new tools to explore these mechanisms (35–38). We anticipate that these studies, alongside modeling

and computational analysis, will provide insights into the significance of the continuous nature of cell type variation for the proper function of V1. Multitasking theory (10, 39) suggests that transcriptomic continua aid the division of labor among cells in a tissue responsible for multiple functions. For continua bound by a polygon, cells close to each vertex (archetype) are specialized for a particular function, while cells in the middle are “generalists”, which can perform multiple functions at the cost of being sub-optimal for any single function. When gradients in gene expression space are also correlated with position, this can create spatial domains within a tissue, each suited for a unique complement of functions. Consistent with this idea, transcriptomic continua have recently been found to be a common trait in the mammalian brain, and they are often correlated with spatial and physiological continua (40–47). Indeed, the physiological and morphological features of L2/3 neurons in binocular V1 vary as a continuum along the pial-ventricular axis (48). Together, these findings suggest that vision controls the function of L2/3 glutamatergic neurons in V1, at least in part, by contributing to shaping the properties of this continuum. As L2/3 neuron types have selectively expanded in the primate cortex, continua of cell types may be particularly well suited to change by experience and may contribute to the enhanced cognitive capacities of primates.

Supplementary Figures

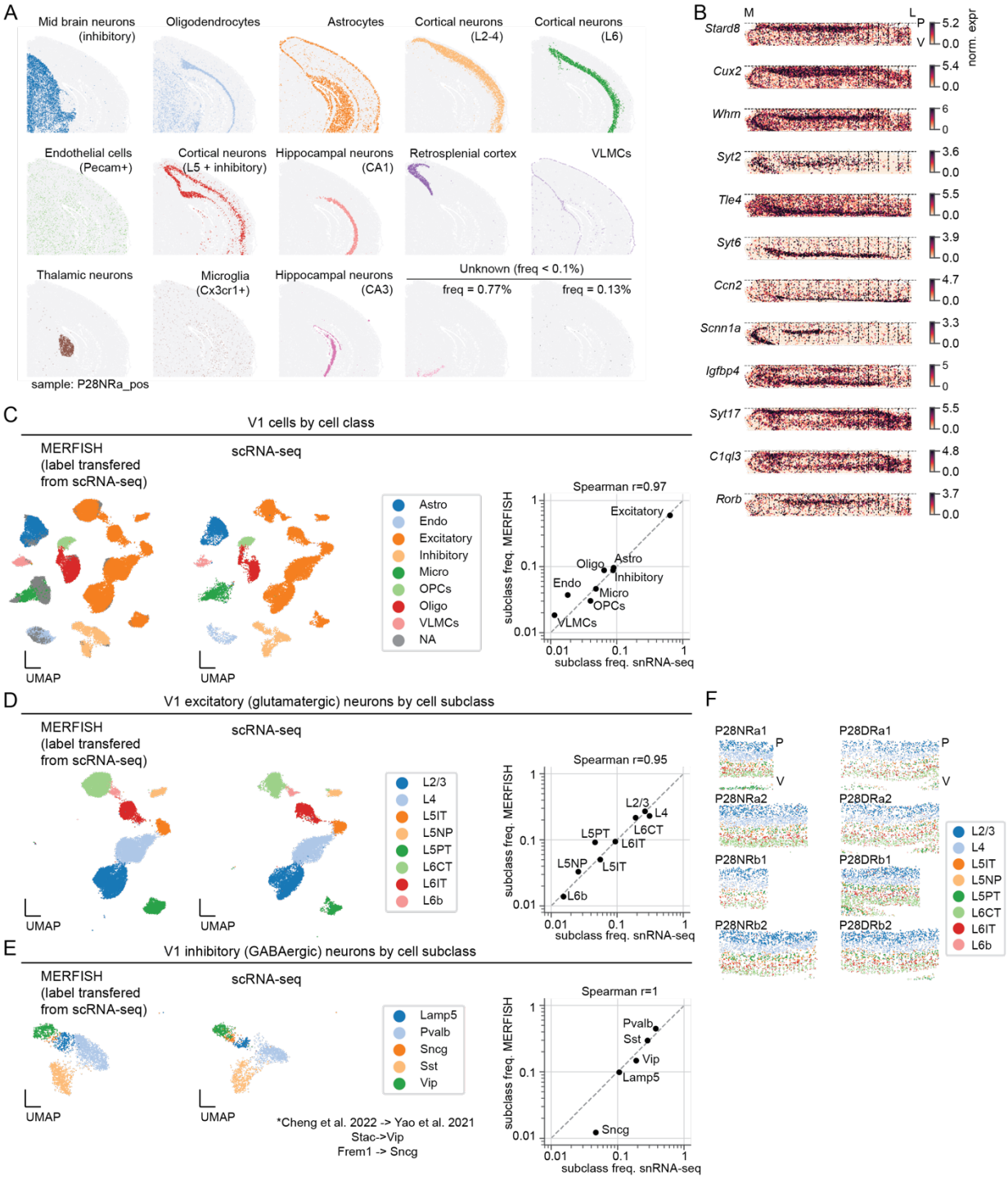
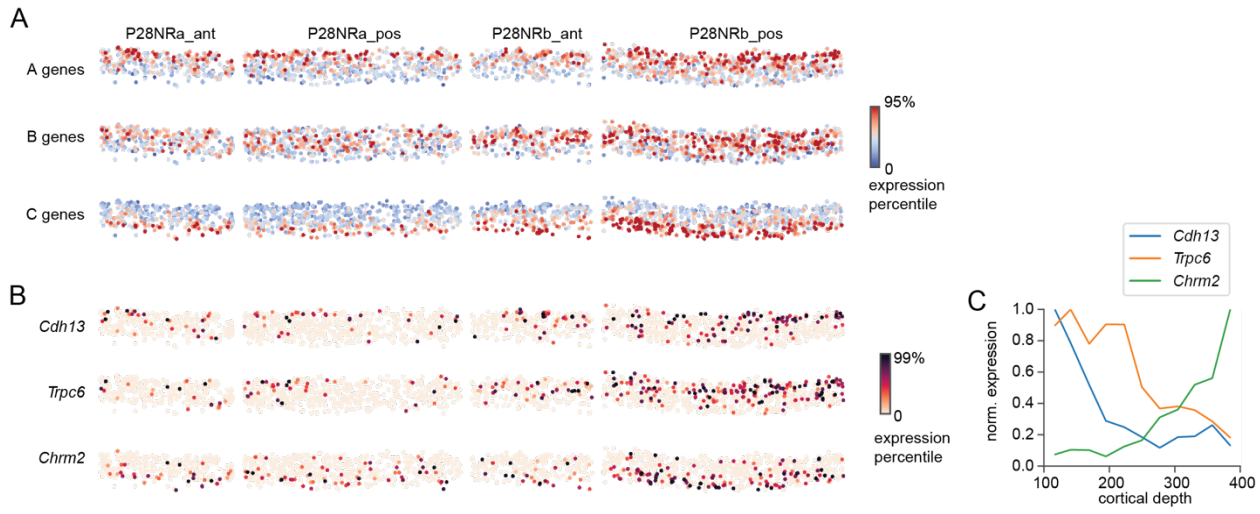


Figure S1. Related to Figure 1. MERFISH identifies major anatomical structures of the brain and cell subclasses in V1.

(A) Spatial distributions of individual cell clusters identified based on gene expressions and spatial proximity (Related to **Figure 1B**). (B) *In situ* gene expression patterns across the

472 straightened cortex. (C) V1 cell classes represented in UMAP obtained from integrating
473 MERFISH (left panel; this study) and snRNA-seq (middle panel (9)) using Harmony (14). Right
474 panel compares cell type proportions between snRNA-seq and MERFISH. 19.6% of MERFISH
475 cells could not be classified (marked as NA and colored gray) due to their low quality and were
476 filtered from downstream analyses. These cells are likely low-quality glia as they are co-
477 clustered with non-neuronal cells. (D-E) Subclasses of V1 glutamatergic neurons (D) and
478 GABAergic neurons (E) represented in UMAP embeddings obtained from integrating MERFISH
479 (left panel; this study) and snRNA-seq (middle panel (9)) using Harmony (14). Right panels
480 compare cell type proportions between snRNA-seq and MERFISH. (F) Spatial distributions of
481 V1 glutamatergic subclasses.



482

483 **Figure S2. Related to Figure 1. *In situ* distributions of L2/3 type-identity genes.**

484 (A) *In situ* expression patterns of types A, B and C identity genes. (B) *In situ* expression patterns
485 of *Cdh13*, *Trpc6* and *Chrm2*. *Cdh13* marks type A, *Trpc6* marks type B, and *Chrm2* marks type
486 C. (C) Line plots showing the distribution of marker gene expressions along the cortical depths
487 spanning L2/3.

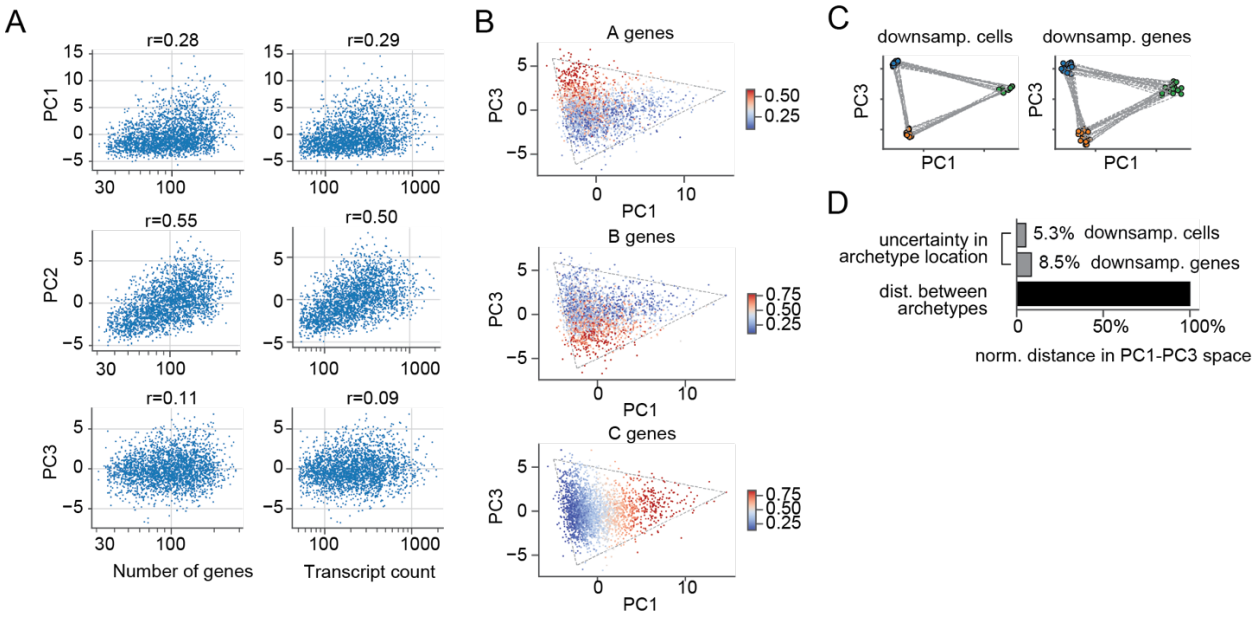
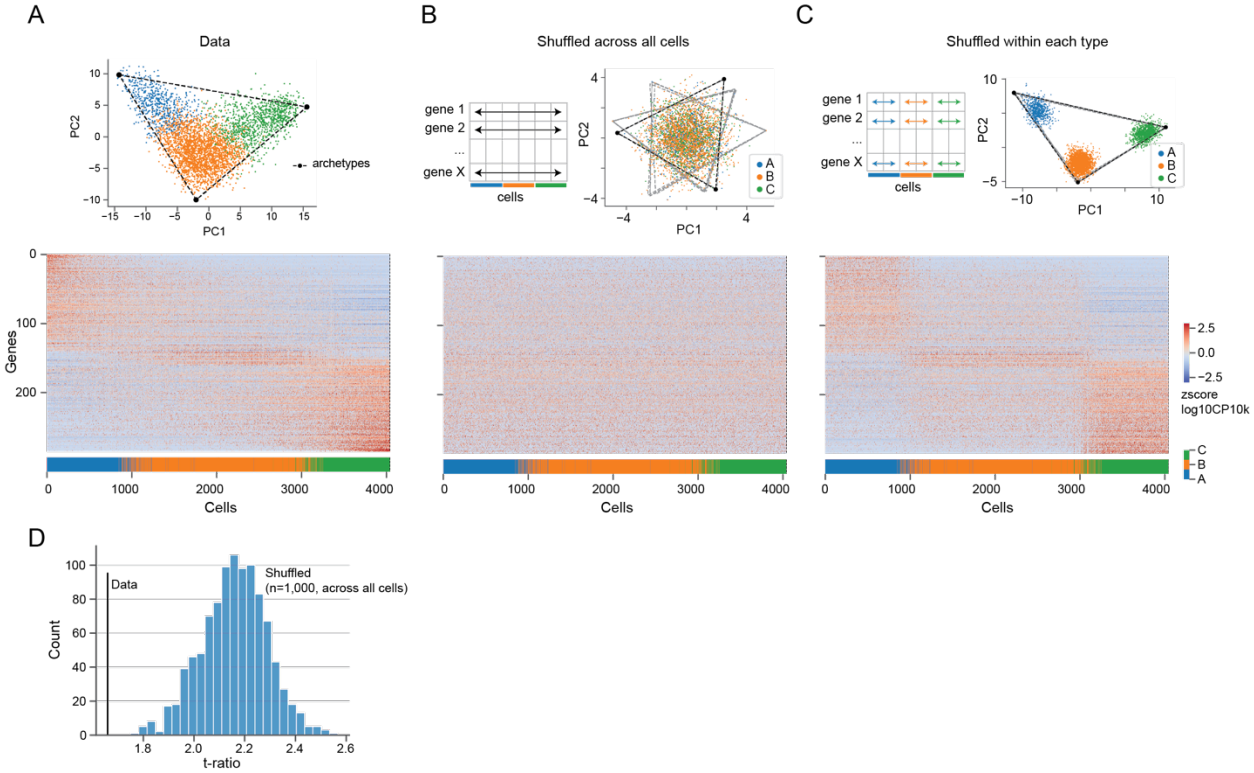


Figure S3. Related to Figure 2. Low dimensional representation of L2/3 cell types in MERFISH and inference of the bounding triangle from MERFISH data alone.

(A) Scatter plots of PC1 (upper), PC2 (middle) and PC3 (lower) vs. the number of detected genes (left) and transcript count (right panel), respectively. PC1 is calculated using the 170 type-identity genes measured by MERFISH. PC2 correlates with these technical variables much more than PC1 and PC3. (B) V1 L2/3 cells embedded in PC1 and PC3. Cells are colored by the mean z-scored expression levels of type-A genes (upper panel), type B genes (middle panel) and type C genes (lower panel), respectively. (C-D) Stability of the bounding triangle. The bound triangle was inferred in several instances involving random down-sampling to 80% of cells (C, left panel) and 80% of genes (C, right panel). The stability of archetype locations are quantified by the % variation in archetype coordinates relative to the typical inter-archetype distances in these trials. Distances are measured by Euclidean distance in PC1-PC3 space. These inferences were done on MERFISH data, recapitulating results from snRNA-seq showed in Figure 2.



503

504 **Figure S4. Related to Figure 2. Data shuffling disrupted the geometry of the L2/3**
505 **continuous transcriptomic manifold.**

506 (A) *Upper panel:* PCA embedding of L2/3 glutamatergic neurons in snRNA-seq using 286 type-
507 identity genes. Cells were colored by type assignment based on the original clustering (9). The
508 bounding triangle is inferred from archetypal analysis. *Lower panel:* expression profiles of L2/3
509 type-identity genes across all L2/3 glutamatergic neurons (n=4,044; P28NR). Expressions were
510 quantified as z-scored, log- and size-normalized counts from scRNA-seq. Cells were ranked by
511 diffusion pseudo-time (DPT; (19, 20)), and colored by type assignment. (B) Same as (A) but
512 after shuffling each gene independently across all cells. (C) Same as (A) but after shuffling each
513 gene independently across cells within each type. (D) Histogram of T-ratios for the data and
514 shuffled data (n=1,000; shuffled across all cells). T-ratio is the ratio between the area of the
515 convex hull and that of the principal convex hull (triangular fit).

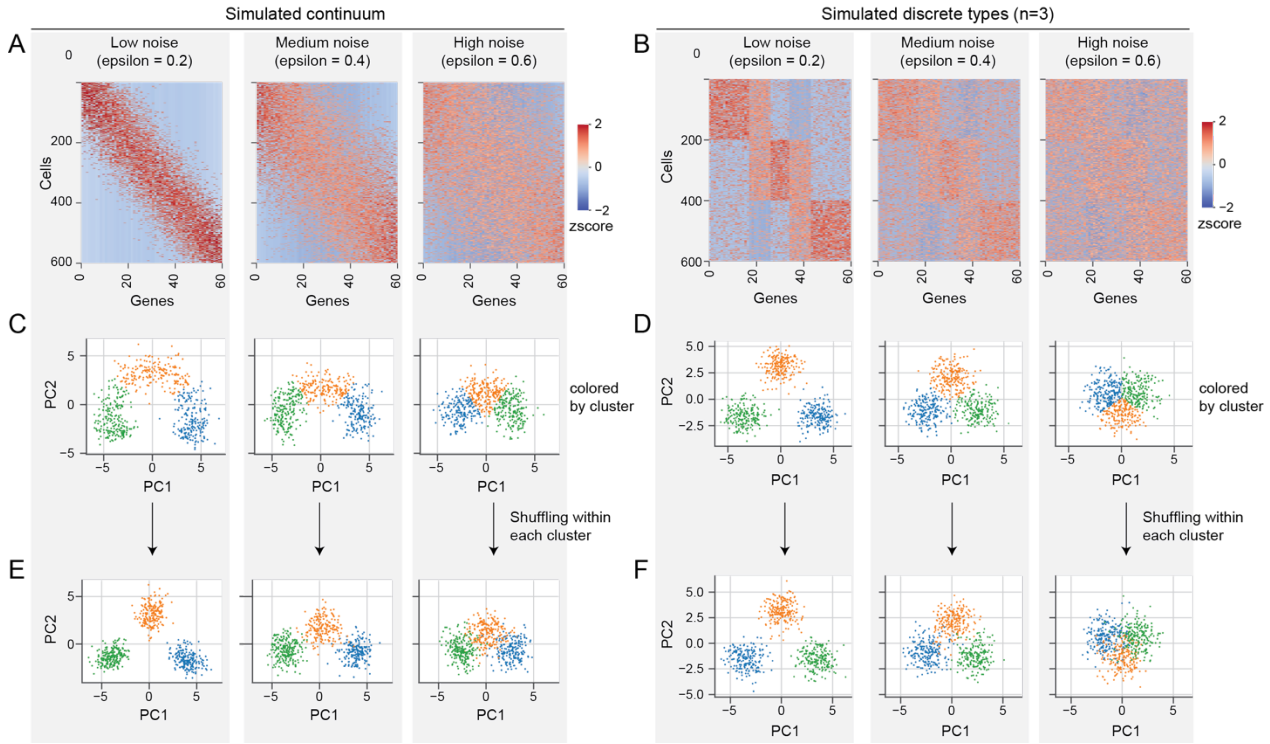


Figure S5. Related to Figure 2. A data shuffling procedure that split a continuum while preserving discrete types.

(A-B) Expression profiles (gene by cell) for simulated continua (A) and simulated discrete types (B) each with varying degrees of noise level (parameterized by epsilon; see **Methods**). Expression was quantified as z-scores. (C-D) PCA embeddings (PC1 and PC2) of simulated continua (C) and simulated discrete types (D). Cells were colored by type. (E-F) Same as (C-D) but after shuffling each gene independently across all cells within each type. Shuffling genes within a cluster splits a continuum into separate clusters when noise is low (epsilon = 0.2 ~ 0.4). The same procedure has no effect on discrete types whatsoever.

A

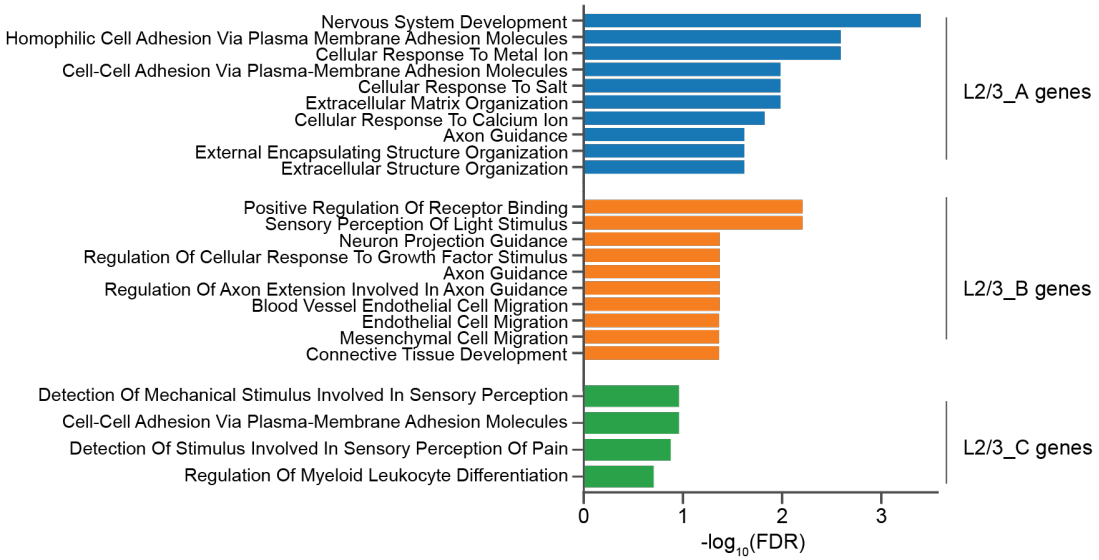


Figure S6. Related to Figure 2. Gene Ontology analysis identify enriched biological processes for type-identity genes.

(A) Enriched biological processes for types A, B and C identity genes, respectively.

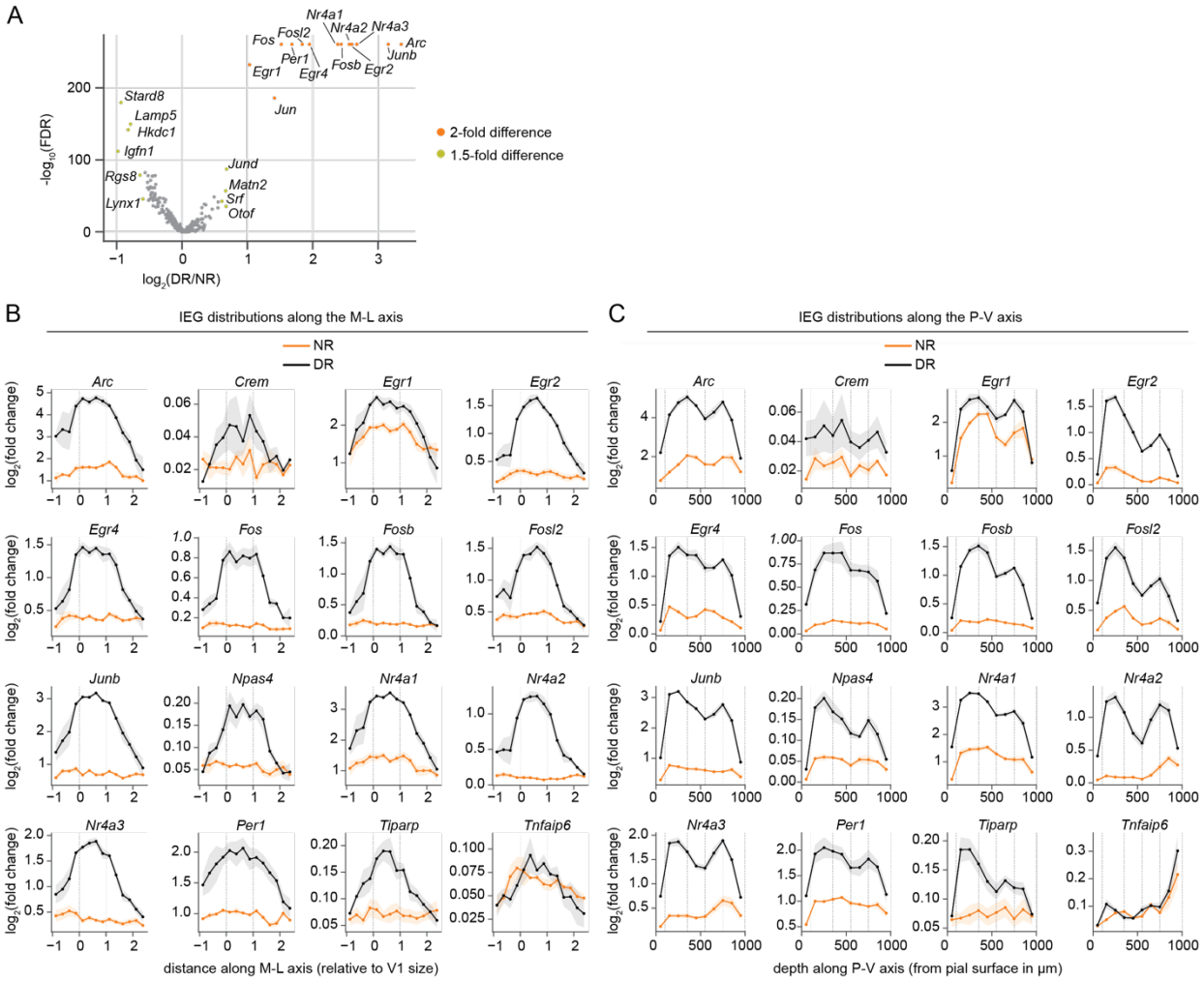


Figure S7. Related to Figure 3. Vision-dependent genes profiled by MERFISH.

(A) Violin plot showing differentially expressed genes between NR and DR L2/3 glutamatergic neurons profiled by MERFISH. (B-C) Mean expression levels of individual IEGs along the M-L axis (B) with V1 in the middle and its flanking regions on the two sides, and along the P-V axis (C) across cortical layers. Lines demarcate V1 (B) and cortical layers (C).

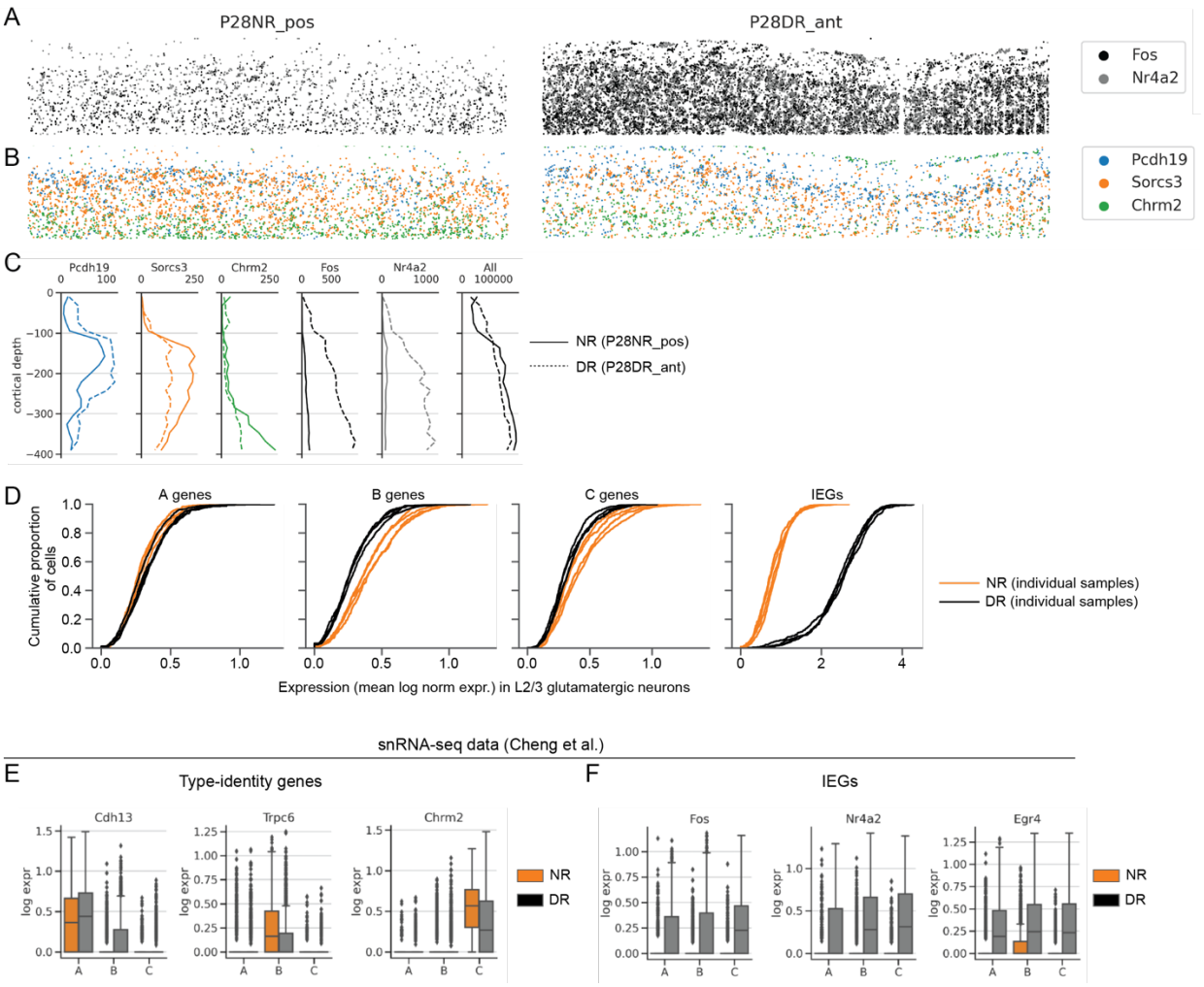


Figure S8. Related to Figure 3. MERFISH detected transcripts in V1 L2/3 for specific genes and gene groups.

(A-B) Detected transcripts of Fos and Nr4a2 (A) and of *Pcdh19*, *Sorcs3*, and *Chrm2* (B) in V1 L2/3 for NR and DR samples. The entire V1 L1-3 area is shown. (C) Quantification of the number of detected transcripts at different cortical depths for genes shown in panels (A-B). (D) Cumulative distribution of gene expression across different gene groups and samples. Samples are colored by NR and DR. Group-level expression is defined as the mean expression level across genes within the group. (E-F) Gene expression distributions (in box plots) of selected type-identity genes (E) and IEGs (F) in L2/3 types from NR and DR mice.

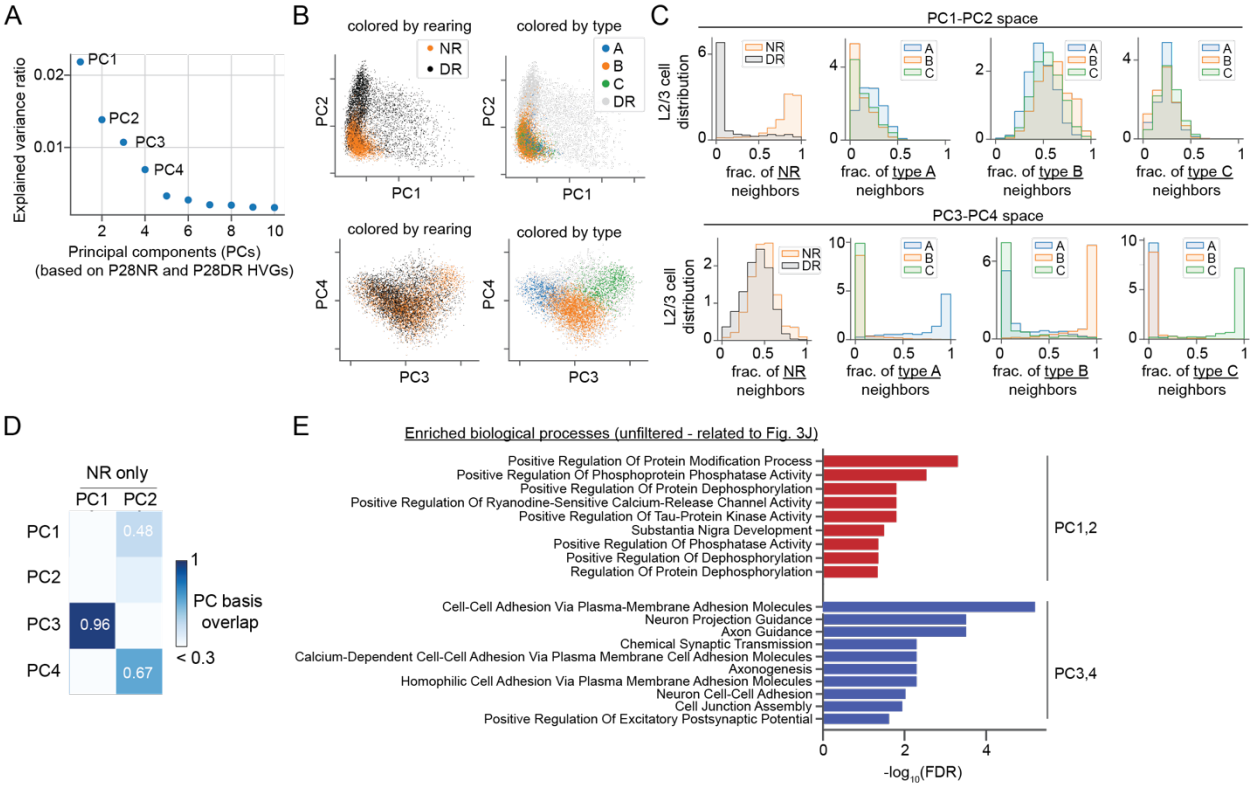


Figure S9. Related to Figure 3. PCA analysis revealed orthogonal components of transcriptomic variations in L2/3 transcriptomes from NR and DR mice.

(A) Fraction of total variance captured within the top 10 principal components (PCs) in NR and DR (P28) L2/3 cells. (B) Distribution of NR and DR cells in PC1-PC2 space (upper) and PC3-PC4 space (lower). PCs are computed from n=6,360 HVGs. Cells are colored by rearing condition (NR vs DR) (left) and by cell type (A, B and C) for NR cells (right). (C) Distribution of neighbor identities. Column 1 shows the fraction of NR neighbors for NR cells (in orange) and DR cells (in black) in PC1-PC2 space (upper panel) and PC3-PC4 space (lower panel). Columns 2-4 show the fraction of types A, B and C neighbors respectively. (D) Pairwise overlap between PCs derived from using both NR and DR cells versus using NR cells only. (E) Enriched biological processes for PC1-PC2 and PC3-PC4 driving genes. This is related to main Figure 3J.

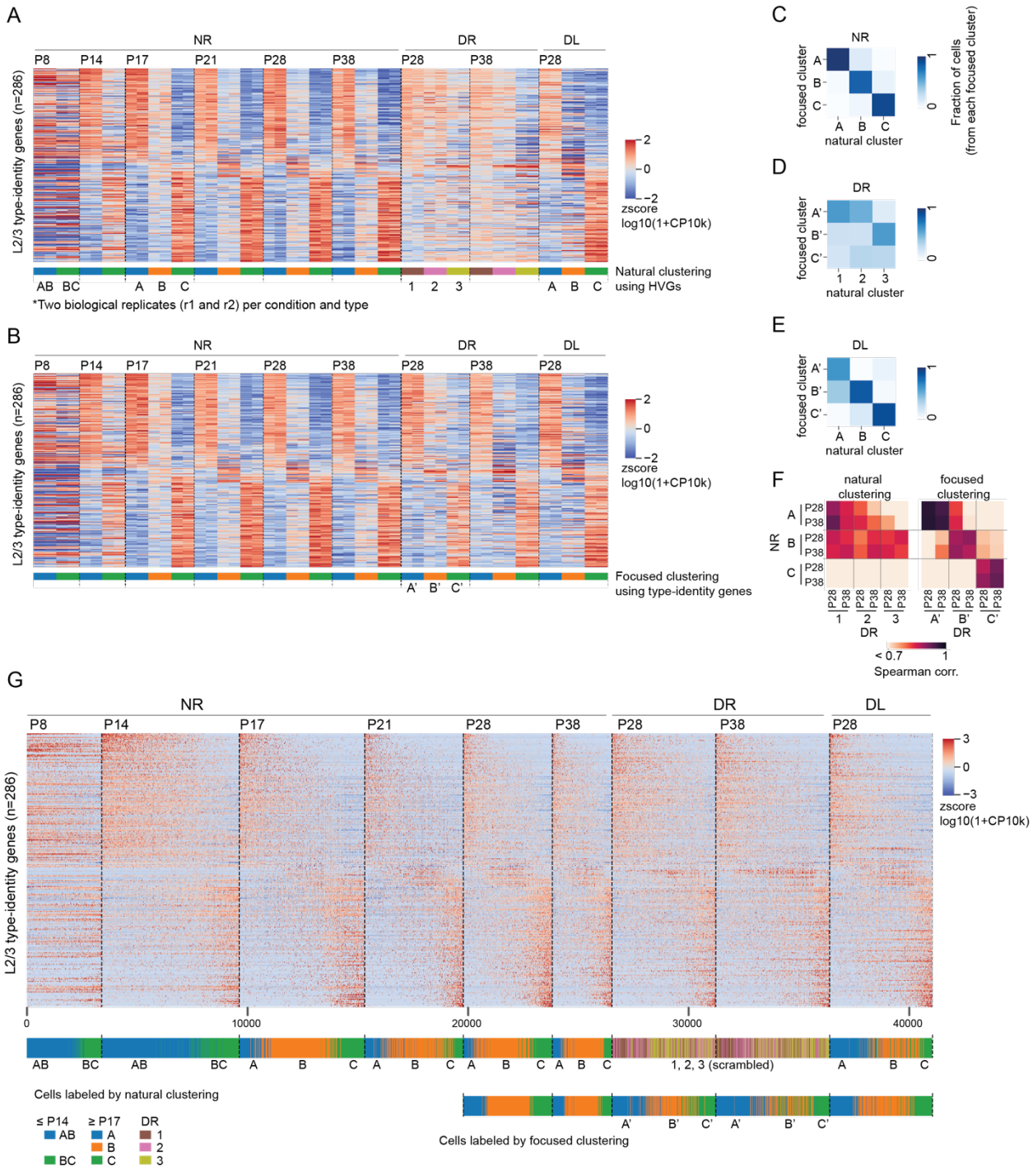


Figure S10. Related to Figure 3. The L2/3 transcriptomic continuum across time and conditions.

(A-B) Expression profiles of L2/3 type-identity genes for clusters based on natural clustering using HVGs (A), and based on focused clustering using type-identity genes (B). Natural clustering using HVGs mask the cell-type signatures in DR, while focused clustering using type-identity genes recover the cell-type signatures. (C-E) Confusion matrix of natural clustering vs focused clustering for NR (C), DR (D) and DL (E), respectively. (F) Pairwise Spearman's

567 correlation coefficients between NR and DR cell clusters based on L2/3 identity genes. Focused
568 clustering has 1:1 correspondence between NR and DR clusters, while natural clustering does
569 not. (G) Expression profiles of L2/3 type-specific genes for individual cells. Top: Gene
570 expression profiles. Bottom: cells colored by HVG-based and focused clustering respectively.
571 Cells were ordered by diffusion pseudo-time for each condition.

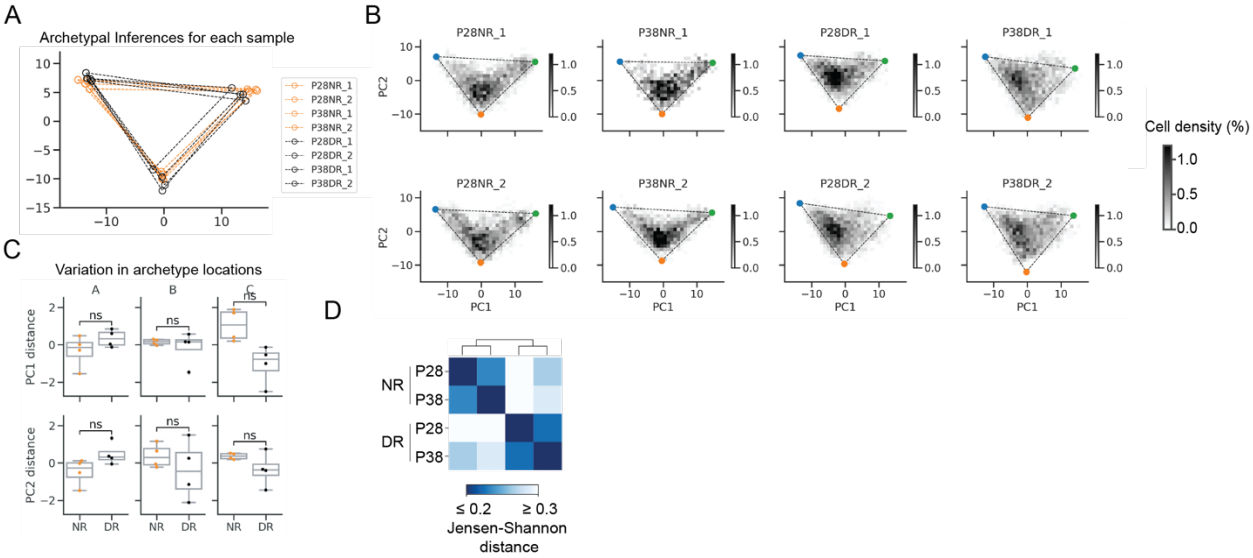


Figure S11. Related to Figure 4. Dark-rearing shifted cells along the continuous transcriptomic manifold of L2/3 cell types (snRNA-seq data).

(A) Inferred L2/3 transcriptomic triangle for individual biological samples. (B) L2/3 cell density plot in PCA embeddings (PC1 and PC2 based on type-identity genes). L2/3 cells from all samples are embedded into the same PC1-PC2 space using type-identity genes. Cell density and the archetypal inference are done separately for each sample. (C) Boxplots showing the variations of archetype locations in PC1-PC2 space. (D) Pairwise Jensen-Shannon distances between samples of different rearing conditions.

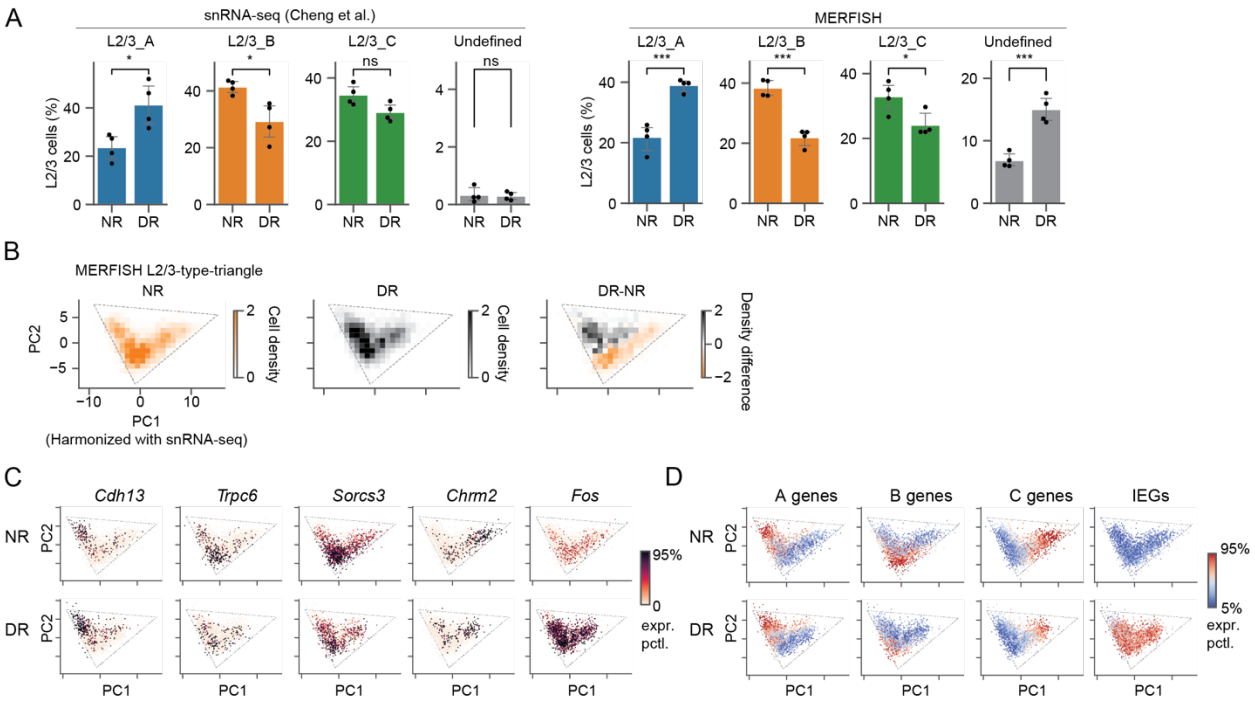
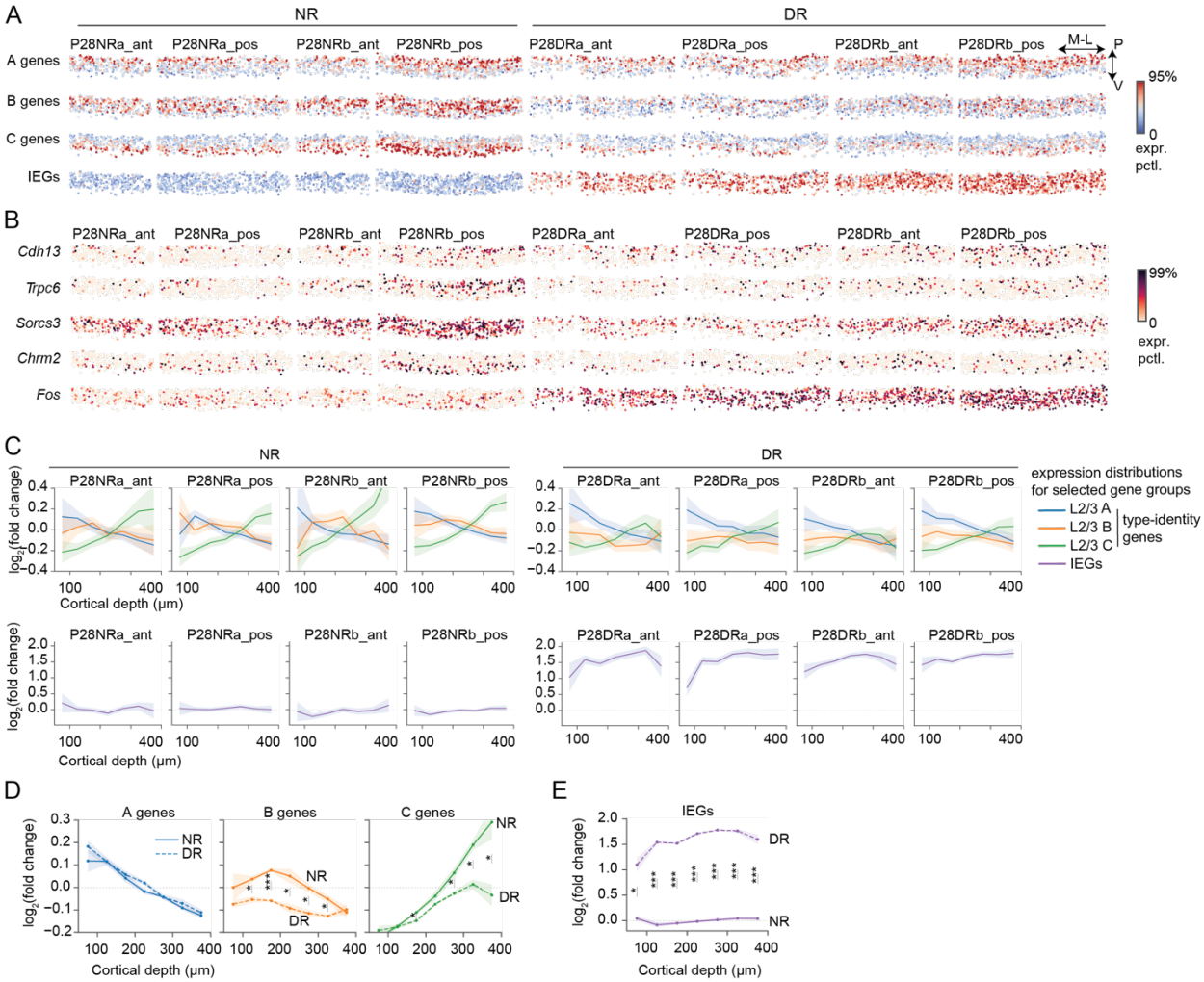


Figure S12. Related to Figure 4. Dark-rearing shifted cells along the transcriptomic manifold of L2/3 cell types (MERFISH data).

(A) Proportion of L2/3 cell types from NR and DR samples profiled by snRNA-seq (left panel) and MERFISH (right panel). Each dot represents a biological sample. Cell type labels are based on the expression levels of type-identity genes (see Methods). (B) Distribution of L2/3 cells within its bounding triangle for NR and DR mice profiled by MERFISH. (C-D) Gene expression patterns for individual genes (C) and groups of genes (D) on the L2/3 type-triangle.



589

590 **Figure S13. Related to Figure 4. Characterization of *in situ* gene expression in V1 L2/3.**

591 (A-B) in situ expression patterns of different gene groups (A) and individual genes (B) in V1
592 L2/3. Each row shows one gene (or gene group) and each column one sample. (C) Line plots
593 showing the mean expression levels of type-identity genes (upper panel) and IEGs (lower
594 panel) along the cortical depth for individual samples. (D-E) Line plots comparing the expression
595 levels of type-identity genes (D) and IEGs (E) in NR and DR across cortical depths spanning
596 L2/3. FDR < 0.05 (*) and FDR < 0.001 (***).

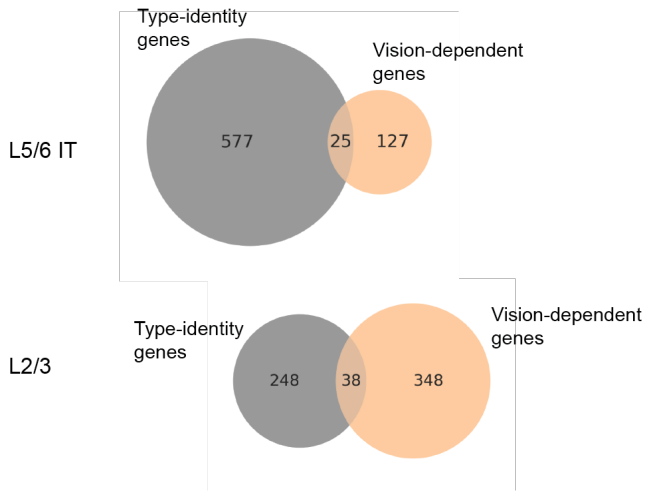


Figure S14. Related to Figure 4. Comparing type-identity gene programs and vision-dependent gene programs in superficial (L2/3) and deep (L5/6) layer intra-telencephalic (IT) neurons.

Venn diagrams showing the numbers and the degree of overlaps between type-identity genes and vision-dependent genes identified in L5/6 IT (upper panel) and L2/3 IT (lower panel) neurons.

Supplemental Tables

Table S1. L2/3 type identity genes (n=286), in which 38 were up- or down-regulated in dark reared mice.
Table S2. MERFISH gene panels (n=500).
Table S3. Enriched GO biological processes in L2/3 type-identity genes.
Table S4. Enriched GO biological processes in PC1-PC2 and PC3-PC4 top loading genes.
Table S5. Previously identified L2/3 stimulus-response genes (from ref (11)).

Methods

Mice for MERFISH

Mouse breeding and husbandry procedures were carried out in accordance with the animal care and use committee protocol number 2009-031-31A at the University of California, Los Angeles. Mice were given food and water *ad libitum* and lived in a 12-hr day/night cycle with up to four adult animals per cage. Only virgin male C57BL/6J wild-type mice were used for MERFISH experiments to match with Ref (9).

Visual deprivation paradigm

Mice were dark-reared starting postnatal day 21 (P21), which marks the beginning of the ocular dominance critical period. Prior to P21, mice were reared in cages in a room with a 12-hour light/dark cycle. At P21, mouse cages were moved to a separate room and placed inside a dark box lined on the inside and the outside with black rubberized fabric (ThorLabs Cat# BK5) with edges sealed by tape and fabric to avoid any light entry. Mice were dark-reared in this setup for 7 days (P21-28DR group). During this period, any handling of the cages was performed in the dark with room lights off, door crevices sealed from exterior light, and in dim red light, which is invisible to the mice. A light meter was used to measure the amount of light inside the dark-rearing box before visual deprivation paradigm. All mice were dark-reared in 0-0.01 lux conditions. At the end of dark-rearing, mice were euthanized and harvested for MERFISH experiments under red light.

Normally reared mice were housed in a 12hr light-ON, 12hr light-OFF cycle and were harvested during a range of 6-8 hours into the light-OFF phase. The cages of normally reared mice were wrapped with the same black rubberized fabric and carried to dark room with only red light on to minimize light exposure before euthanasia.

Tissue preparation and MERFISH experiments

All mice were anesthetized with isoflurane and then perfused transcardially with 10mL heparinized PBS. Following perfusion, the brains were harvested, imbedded in pre-chilled Tissue-Tek® OCT mounting medium, and flash-frozen in dry ice-prechilled methylbutane. The frozen blocks were then kept on dry ice and stored at -80℃ until time of sectioning. To prepare cryosections for MERFISH, two entire OCT blocks (two brains) or three bisected OCT blocks (three hemibrains) were combined and sectioned at -20 °C in a Leica CM1850 cryostat. Two 12-mm-thick coronal slices (one anterior and one posterior containing binocular V1 with ~550um apart in the anterior-posterior axis) were directly collected onto a specially coated 4cm-diameter MERFISH glass slides (merslides, Vizgen Item# 10500001). For locating the visual cortex, atlas coordinates from (Franklin and Paxinos, 2012) and Allen Brain Atlas were used. Collected cryosections on the MERFISH slides (merslides) were fixed in 4% paraformaldehyde in PBS

647 (15min at RT) in a 6cm petri dish, rinsed with cold RNase-free PBS, and stored in 70% ethanol
648 at 4 °C until the step of MERFISH probe hybridization.
649
650 MERFISH was performed according to Vizgen's instructions. Briefly, merslides with brain sections
651 stored in 70% ethanol were rinsed with Vizgen Sample Preparation Buffer (SPB) after aspiration
652 of 70% ethanol, incubated with Vizgen's Formamide Wash Buffer (FWB, 30 min at 37 °C) followed
653 by MERFISH probe labeling with a customized mouse gene panel containing 500 mouse genes
654 (VZGCP0991, **Table S2**, ~40h at 37 °C) in a moist chamber, and washed with FWB twice (30 min
655 at 47 °C). The brain sections were be embedded with Vizgen gel mix after removal of FWB,
656 cleared in 5mL clearing mix solution with 50 mL protease K (overnight at 37 °C), stained with
657 Vizgen DAPI/poly(T) reagent included in the Vizgen 500-gene imaging kit (10 min at RT) after
658 rinse with SPB and FWB (10 min at RT). Then the merslide was thoroughly rinsed with SPB,
659 carefully assembled into an imaging gasket chamber, and uploaded into the MERSCOPE for
660 imaging. The MERFISH imaging was done on the MERSCOPE with an activated Vizgen 500-
661 gene imaging kit after adding the imaging buffer activator and RNase inhibitor (100 mL) as
662 instructed. The imaging process was conducted under the direction of Vizgen's MERSCOPE
663 program (Software version 233.230615.567) with the default settings (both polyT and DAPI
664 channels "on", scan thickness: 10mm). Once the MERFISH imaging process was completed, the
665 output files were transferred for in-depth analysis with MERFISH Visualizer and in-house custom
666 bioinformatic pipeline.

667 **snRNA-seq data processing and normalization**

668 Cell-by-gene count matrices from the previous study were downloaded from Gene Expression
669 Omnibus (GEO) repository GSE190940 (9). Cell type labels were downloaded from the
670 associated Github repository https://github.com/shekharlab/mouseVC. Raw count matrices were
671 normalized as described before. Transcript counts within each cell were rescaled to sum up to
672 10,000. A pseudo-count of 1 was added to the normalized transcript counts for each gene within
673 each cell, followed by log10-transformation. For PCA and clustering, log10-transformed counts
674 were z-scored across cells for each gene.
675
676 We reproduced a list of L2/3 type-identity genes (n=286, **Table S1**) by following the differential
677 expression analysis used in the previous study. Briefly, the expression of each gene was
678 compared in one type versus others (L2/3 A vs B-C; L2/3 B vs A-C; and L2/3 C vs A-B). We
679 used *scanpy.tl.rank_genes_groups* to identify differentially expressed genes if a gene meets the
680 following criteria: a) false discovery rate (FDR) < 0.05 based on Wilcoxon rank-sum test; b) fold
681 change > 2, and c) the gene was expressed in > 30% cells in the up-regulated type.

682 **MERFISH data processing and normalization**

683 Starting from a cell-by-gene matrix, in which each entry is an integer representing the number of
684 transcripts detected in each cell and for a specific gene by MERFISH, we first removed low
685 quality cells. We kept cells with a volume between 50~2000, total number of detected
686 transcripts > 10, and false positive rate of transcript detection <5% for the initial analysis. (Later
687 to identify L2/3 cell types A, B and C, we further selected high-quality cells with total number of
688 detected transcripts > 50). Notably, the false positive rate is a unique feature of MERFISH. In
689 MERFISH experiments, transcripts were identified by binary barcodes extracted from repeated
690 rounds of hybridization, imaging and puncta detection. A few barcodes were designed to be
691 blank barcodes unmapped to any genes, which allowed us to estimate false positive transcript
692 detection rate. Out of the eight MERFISH sections, four were done on the full coronal sections

693 while the rest were done on hemibrain coronal sections only. For consistency, we
694 computationally cut out a hemi-brain section for every sample for downstream analysis.
695
696 Following the same procedure employed in a previous MERFISH study (18), we normalized the
697 raw count matrix first by cell volume and then by mean transcript count across samples such
698 that each section had the same mean number of transcripts (n=250) detected per cell. This
699 procedure effectively ameliorated cell volume-associated variation and batch effects across
700 samples. After the above cell-to-cell and sample-to-sample size normalization, we normalized
701 the data by log (+1) transformation and z-scored every gene.
702
703 Notably, we found a massive up-regulation of a panel of IEGs in sections from DR mice
704 compared to NR (see main text). To remove this confounding factor, sample-to-sample
705 normalization was done using 484 out of the 500 genes profiled by MERFISH (excluding the 16
706 IEGs) to compute the total transcript counts. We found that the identification of major cell
707 subclasses was unaffected by the inclusion/exclusion of IEGs. However, IEG removal helped
708 quantify gene expression changes in NR vs DR within L2/3 types more accurately.

709 **Spatial domain analysis**

710 To identify the major anatomical areas and molecular organization within our MERFISH
711 sections, we performed a clustering analysis on each section based on cell-to-cell gene
712 expression similarity and spatial proximity. Using the normalized data as features, we performed
713 a principal component analysis (PCA) and generated a gene-expression-similarity graph based
714 on coordinates in the space of the top 50 principal components (PCs). We also generated a
715 spatial-proximity graph based on spatial locations of the cells. Both graphs were generated
716 using *scanpy.pp.neighbors* following default parameters. We generated a spatial-and-gene
717 hybrid graph by blending the spatial and the gene graphs with equal weighting. We applied
718 Leiden clustering on the hybrid graph using *scanpy.tl.leiden* following default parameters. The
719 resulting spatial domains recover major anatomical structures in the mouse brain, which were
720 visualized in **Figures 1** and **S1**. Clusters were annotated based on anatomical and molecular
721 features.

722 **Establishing curved cortical coordinates and identifying V1**

723 We developed an automated iterative procedure to fit a smooth curve along the pial surface,
724 which served as a reference line to establish curved coordinate system along the vertical (pia-
725 ventricular; P-V) and the tangential (medial-lateral; M-L) axis of the cortex.
726
727 We took advantage of meningeal cells (VLMCs), which specifically express *Slc6a13* and are
728 located along the pial surface. Using these *Slc6a13*+ cells as anchors, we first fit a fourth-order
729 polynomial using *numpy.polyfit*. Subsequently, we used the fitted line as a reference to calculate
730 the depth for all the anchor cells. Anchor cells whose depth that are >1/2 of the maximum depth
731 (robust maximum taken as the 95% max depth) among all anchors were removed from the set
732 of anchor cells for another round of more refined fitting. This procedure was repeated for 5~10
733 times until the inferred reference line was stable and visually matched the cortical surface.
734 Using the reference line, all cortical cells were assigned a depth along the P-V axis and a width
735 along the M-L axis.
736
737 Using the curved coordinates, cortical cells, defined as having cortical depth less than 1100
738 microns from the pial surface, were extracted from the full coronal sections for downstream

analysis. V1 cells were identified based on tangential locations based on areal-specific markers mentioned in the main text.

Identification of subclasses of V1 cells and integration with published snRNA-seq data

To systematically identify major cell populations, V1 cells were extracted and merged from all MERFISH sections (8 in total for this study; NR and DR) as mentioned above and were integrated with our previously published snRNA-seq data (9) using Harmony (14). We did this at two levels – the class level and the subclass level within the excitatory and inhibitory classes (**Figure S1**).

To integrate the MERFISH and snRNA-seq data (P28NR), we intersected MERFISH genes (n=484, excluding IEGs) and highly variable genes (HVGs; n=7,340) identified by snRNA-seq, which resulted in 361 overlapping genes between the two datasets. For this analysis, we only used high-quality MERFISH cells, defined as cells with at least 50 transcripts. We computed PCA (n=20 components) using cells from both datasets based on the overlapping genes (z-scored separately for each dataset). We then applied Harmony using *scanpy.external.pp.harmony_integrate* to generate harmonized PCs.

These harmonized PCs were used for downstream visualization and cell-type assignment. UMAP were generated using *scanpy.pl.umap* with the default parameters. Cell type labels were transferred from the published snRNA-seq data to MERFISH based on k-nearest-neighbor-based assignment using *sklearn.neighbors.NearstNeighbors*. A MERFISH cell was assigned to the most frequent cell class out of its k=100 snRNA-seq neighbors. To assign cell subclasses with a cell class, k=30 neighbors were used.

A small fraction of MEFISH cells could not be confidently assigned a cell class (**Figure S1C**). These cells were defined as those whose k=100 neighbors in the harmonized PCs contains ≤ 2 cells from snRNA-seq, indicating an unsuccessful data integration for these cells. These cells were assigned label “NA” (**Figure S1**), most likely being low-quality non-neuronal cells with lower number of transcripts, and unspecific spatial signatures. They were mostly grouped with microglia and endothelial cells in the UMAP embedding (**Figure S1C**).

Identification of L2/3 glutamatergic cell types

The procedures above identified all cell subclasses, among which we focused on one subclass, the L2/3 glutamatergic neurons, for further analysis. Two strategies were used to further identify the three types of L2/3 neurons (types A, B and C). The first was to repeat the analysis above at the cell type level, which assigned each MERFISH cell an identity based on its k-nearest neighboring snRNA-seq cells in the harmonized PC space.

We also used a second strategy, which offers more biological interpretability. Given the continuous nature of L2/3 types bounded by the three archetypes A, B and C, we developed continuous A-B-C scores based on the expression levels of type-A, B and C identity genes (n=170 profiled by MERFISH). Each cell was assigned three scores (p_a, p_b, p_c), with $p_a + p_b + p_c = 1$, representing the probabilities that this cell belongs to type A, B and C respectively. For a cell i, we first characterized its overall expression levels of type-A specific genes by the mean z-scored type-A-identity gene expression,

$$z_{ia} = \frac{1}{|G_a|} \sum_{g \in G_a} z_{ig}$$

where G_a represents the set of type-A-identity genes and $|G_a|$ represents its size. Type B and C scores were similarly defined as

$$z_{ib} = \frac{1}{|G_b|} \sum_{g \in G_b} z_{ig}$$

$$z_{ic} = \frac{1}{|G_c|} \sum_{g \in G_c} z_{ig}$$

These mean z-scores were normalized to be bounded between [0, 1].

$$s_{ia} = \frac{|z_{ia} - z_{\min,a}|}{z_{max,a} - z_{min,a}}$$

Where $z_{min,a}$ is defined as the 40th percentile of z_{ia} among L2/3 cells and $z_{max,a}$ is defined as the 95th percentile. Next, we normalized these scores such that they sum to 1,

$$p_{ia} = \frac{s_{ia}}{s_{ia} + s_{ib} + s_{ic}}$$

These scores can be directly visualized at single-cell level using an additive blending between three (archetypal) colors. To assign discrete cell type labels, one can also assign a cell to the type that has the highest ABC scores. Cells with clear A, B or C identity, i.e. archetypal cells, are defined as those whose highest ABC scores were greater than 0.6 ($\max(p_a, p_b, p_c) > 0.6$).

Pseudotime analysis for the L2/3 transcriptomic continuum

We computed pseudo-temporal coordinates for L2/3 cells to understand their continuous organization. We first computed the top 50 PCs using z-scored L2/3 type-identity genes, and generated a k-nearest-neighbor (k=50) graph between cells in the PC space. The graph was built using the function *Scanpy.pp.neighbors*. Next, we ran a diffusion map (19) using *Scanpy.tl.diffmap* and ran diffusion pseudotime using *scanpy.dpt* following default parameters. The cell with the smallest PC1 was assigned as the as root cell and serves as the starting point of diffusion. As a result, each cell was assigned a pseudo-temporal coordinate, and cells were ranked by the increasing order of pseudotime.

We also ranked L2/3 type-identity genes according to their expression along the pseudotime coordinates. We defined the pseudotime of a gene (T_g) as the weighted average of cell pseudotime,

$$T_g = \sum_c w_{gc} t_c \,,$$

where t_c is the pseudotime of the cell (c), and w_{gc} is the weight of the cell c contributing to the gene g. The weight of a gene is defined by its expression level and sums to 1,

$$w_{gc} = \frac{x_{gc}}{\sum_{c'} x_{gc'}} \,,$$

where x_{gc} is the size and log normalized expression. This gene pseudotime was defined at P28NR and was kept fixed to give a consistent representation for comparisons between conditions.

Archetypal analysis (the multi-tasking framework)

We used the python package *py_pcha*: https://github.com/ulfaslak/py_pcha to determine the geometry of the L2/3 cell type continuum in the PC1-PC2 space derived from the L2/3 type-identity genes. The tool uses archetypal analysis, also known as the principal convex hull

analysis, to infer a triangular boundary to the L2/3 transcriptomes. The same procedure (with parameters *delta=0 and noc=3*) was applied to different combinations of data (snRNA-seq, MERFISH, different conditions, harmonized data, and shuffled data; see main text).

We evaluated the significance of the triangular fit by t-ratio test as proposed by the multi-tasking framework (10). T-ratio is the ratio between the area of the convex hull of the data and that of the principal convex hull (the triangular bound in our case). We calculated the area of the convex hull using the python package *scipy* and its function *scipy.spatial.ConvexHull*. We tested the significance of the observed t-ratio by comparing it to t-ratios of shuffled data by permuting each gene independently across all cells. P-value was computed based on 1000 shuffles.

Simulations of continuous and discrete types

First, we describe our procedure to simulate transcriptomic continua. Let cells and genes be ranked, such cells with a particular ranking express genes with a similar ranking. Let i be an integer representing the cell ranking, which ranges from 1 to N_c, with N_c denoting the total number of cells. Let j be an integer representing the gene ranking, which ranges from 1 to N_g, with N_g denoting the total number of genes. We use $p_i = \frac{i}{N_c}$ and $p_j = \frac{j}{N_g}$ to represent the normalized rankings such that both the gene rankings and cell rankings range from 0 to 1. We consider a model where the expected expression level of gene j in cell i is a Gaussian such that:

$$\lambda_{ij} = e^{-\left(\frac{p_i - p_j}{\varepsilon}\right)^2},$$

where ε denotes the level of noise. Gene counts are then drawn from a scaled Poisson distribution parametrized by λ_{ij}:

$$c_{ij} \sim s \cdot Poisson(\lambda_{ij}).$$

For our simulation, we chose $N_g = 60$, $N_c = 600$, $s = 100$, with different ε values between 0.1 and 1. Random numbers were generated using the python package *numpy.random*.

Second, for simulating discrete types, we assumed a model wherein each type is distinguished by a set of "marker" genes. Let i be a cell and $C(i)$ be the cell type it belongs to. Let j be a gene and $C(j)$ be the cell type of which it is a marker. The expected expression level of gene j in cell i was modeled by a binary matrix:

$$\lambda_{ij} = \begin{cases} 1 & if\ C(i) = C(j) \\ \varepsilon & if\ C(i) \neq C(j) \end{cases},$$

where ε denotes the amount of leaky expression, i.e., the level of noise. ε takes values between 0 and 1. The larger ε is, the noisier the types are. The actual count matrix was drawn from a scaled Poisson distribution parametrized by λ_{ij},

$$c_{ij} \sim s \cdot Poisson(\lambda_{ij}).$$

We simulated three discrete types (A, B, C), each with 20 marker genes and containing 200 cells, with different ε values between 0.1 and 1. To assign an order between types, among the 20 marker genes, 6 were shared between neighboring types, such that 6 type-A markers were also expressed in type B, and 6 type-B markers were also expressed in type C, and so on.

866 **Aligning V1 coordinates across sections**

867 In our samples, V1 spanned about 2 mm (or ~2,000 microns) along the tangential (M-L) axis of
868 the cortex, with different sections having different lengths as defined by the expression domains
869 of V1 marker genes *Scnn1a* and *Igfbp4* (see Main text). To quantify the *in situ* patterns of
870 specific genes within V1 and in other cortical regions across samples, we defined a normalized
871 tangential coordinate system relative to the length of V1. Specifically, for each section we used
872 the tangential locations of the two ends of V1, t_m and t_l, to normalize tangential coordinates
873 along the cortex:

874 $$\tilde{t} = \frac{t - t_m}{t_l - t_m}$$

875 In this normalization, 0 and 1 represent the two ends of V1 (medial-most and lateral-most)
876 respectively. And values <0 and 0>1 represent regions outside of (flanking) V1. Expression
877 levels of specific genes are quantified according to this normalized coordinate system.

878 **Focused clustering**

879 In addition to the cell types identified previously, we applied K-means clustering
880 (*sklearn.cluster.Kmeans)* using only L2/3 type-identity genes as features, to identify focused
881 clusters. As mentioned in the main text, the motivation of this analysis is to uncover L2/3 cell
882 types in DR while ignoring changes to other gene programs related to cell states. This
883 procedure generates matched cell types between NR and DR.

884 **Principal component analysis (PCA)**

885 We applied PCA using *sklearn.decomposition.PCA.* In **Figure 3**, we applied it on L2/3
886 glutamatergic neurons combined from both P28NR and P28DR mice, using a set of unbiasedly
887 identified highly variable genes (HVGs) as features. For HVGs identification, we considered
888 n=21,222 genes that had non-zero counts in at least 10 cells. For each gene, we computed the
889 variance-mean ratio on counts-per-10k-transcripts (CP10k) normalized data. Under a Poisson
890 distribution, this variance-mean ratio is expected to be a constant. Indeed, this value is stable
891 across orders of magnitude difference in mean expression across genes. To select HVGs with
892 different baseline (mean) expressions, we grouped genes into decile-bins according to mean
893 expression, and for each bin we selected the top 30% genes with the most variance-mean ratio.
894 N=6,360 genes were selected, and most (n=270/286) L2/3 type-specific genes are part of the
895 HVGs.
896
897 In **Figure 4**, we applied PCA using L2/3 type-identity genes (n=286; **Table S1**) as features.
898 Focusing on L2/3 type-identity genes, rather than HVGs, allowed us to study the impact of
899 visual deprivation on the L2/3 type-identity gene programs, while ignoring changes to other
900 gene programs.
901
902 We evaluated the degree of overlap between two PCA eigenvectors v_i and v_j as the absolute
903 value of their dot product ($|v_i \cdot v_j|$). As the eigenvectors are orthonormal, this value is bounded
904 between 0 and 1.

905 **Gene ontology (GO) analysis**

906 We used the tool *EnrichR*: https://maayanlab.cloud/Enrichr/ for GO analysis. All enrichment
907 analyses were performed by comparing a gene list of interest with a background set of top 5000

908 most expressed genes in L2/3 (P28NR). Top 10 enriched GO terms of Biological Process (2023
909 version) that were statistically significant with FDR < 0.05 were shown. For terms belonging to
910 the same branch (parents or children), only one term with the highest FDR was shown in the
911 main figures to reduce redundancy. Terms annotated as "obsolete" according to quickGO
912 (https://www.ebi.ac.uk/QuickGO/annotations) were also removed from the main figures. All
913 enriched terms (unfiltered) were shown in supplementary figures.

914 **Previously identified L2/3 stimulus-response genes**

915 Hrvatin et al. (11) reported 611 stimulus-response genes identified from different cell types in
916 V1, among which we filtered out those that are related to L2/3 glutamatergic neurons (labeled
917 as "ExcL23" in Ref (48)), including 42 early-response genes and 37 late-response genes. These
918 genes are listed in **Table S5**.
919
920 We used the Fisher exact test with n=6,360 HVGs as the background to test for enrichment or
921 depletion of the overlap between PC-driving genes and the above activity-regulated genes.

922 **Quantification of cell redistribution in DR**

923 We profiled the cell-density distributions of L2/3 cells in PC1-PC2 space derived from type-
924 identity genes using the python package *seaborn.histplot*. The distributions were robust across
925 a range of bin widths, and we used bin-widths of 1~1.5 for visualization. We used the Jensen-
926 Shannon (JS) divergence (49) to measure the level of difference in cell-density distributions
927 between NR and D and across samples. This metric was calculated using
928 *scipy.spatial.distance.jensenshannon*.
929
930 Optimal transport analysis was performed using the python package *POT*:
931 https://pythonot.github.io/. The program computes the optimal transport map (28, 50) connecting
932 the empirical distributions of NR cells and DR cells in PC1-PC2 space derived from type-identity
933 genes. We first used *ot.dist* to compute pairwise distance matrix between NR and DR cells with
934 default parameters. We then used *ot.emd* to calculate the transport map. The result was a
935 matrix whose elements denote transport couplings between each pair of NR and DR cells, a
936 proxy for their transcriptomic correspondence. We visualized this result by a coarse-grained
937 vector field as follows. First, the PC1-PC2 space was meshed into equal-sized 2D bins. For
938 each 2D bin, we computed a vector representing the mean moving direction and magnitude for
939 all cells that fall into the bin. The vectors are visualized by arrows whose direction represents
940 the average moving directions among local cells, whose lengths are proportional to the
941 movement magnitudes, and whose color (darkness) is proportional to the number of cells
942 represented by the bins.

943 **Identification of differentially expressed genes (DEGs) between NR and DR**

944 In scRNA-seq analysis, it is common to treat individual cells as samples in statistical tests for
945 identifying DEGs. This approach often leads to many statistically significant genes as even
946 modest effect sizes can appear unlikely (low p-values) under the "no effect" null hypothesis
947 owing to large cell numbers. Therefore, as a conservative measure to identify DEGs between
948 NR and DR, we chose to regard the biological samples, rather than single cells, as independent
949 data points.
950
951 We compared NR vs DR across independent biological samples, with each condition having
952 four independent biological samples: P28 rep1, P28 rep2, P38 rep1 and P38 rep2. Raw counts

from cells of the same types and samples were aggregated to produce pseudo-bulk profiles. We only considered genes with mean expression (in counts per million transcripts; CPM) > 10 in at least one type in either NR or DR. We performed independent sample t-test (*scipy.stats.ttest_ind*) on the pseudo-bulk profiles between NR and DR for each type. The resulting p-values were adjusted by the Benjamini-Hochberg procedure to calculate the false discovery rate (FDR). Effect sizes were quantified as the log2 fold change (in CPM) in DR compared with NR. DEGs were defined as those with $FDR < 0.05$ and $|\log_2 FC| > 1$.

Data and code availability

- MERFISH data generated by this study was deposited on Zenodo with DOI: 10.5281/zenodo.13916878
- snRNA-seq data was downloaded from GEO with the accession number GSE190940.
- Code to reproduce L2/3 type specific genes (Table S1, n=286) was obtained from the GitHub repository: https://github.com/shekharlab/mouseVC
- Code to reproduce the figures in the manuscript was deposited in the following GitHub repository: https://github.com/FangmingXie/vision_and_visctx

Acknowledgments

The authors thank members of the Zipursky lab for critical feedback. This work was supported by a grant from the W.M. Keck Foundation to S.L.Z, a NIH grant 1F31 NS131016 to S.B., the Hellman Foundation (K.S.) and McKnight Foundation (K.S.). K.S. is an investigator with the Glaucoma Research Foundation and is supported by the Melza M. and Frank Theodore Barr Foundation. S.L.Z. is an investigator of the Howard Hughes Medical Institute.

Author contributions

F.X., S.J., R.X., K.S., and S.L.Z. conceived the project. R.X., Z.T. and X.X. generated the MERFISH data. F.X. led the analyses. S.J., S.B. and K.S. contributed to analyses. F.X., S.J., S. B., K.S. and S.L.Z wrote the paper.

Declaration of interests

The authors declare no competing interests.

References

1. T. N. Wiesel, D. H. Hubel, Single-cell responses in striate cortex of kittens deprived of vision in one eye. *J. Neurophysiol.* **26**, 1003–1017 (1963).

2. T. K. Hensch, Critical period regulation. *Annu. Rev. Neurosci.* **27**, 549–579 (2004).

3. L. E. White, D. Fitzpatrick, Vision and cortical map development. *Neuron* **56**, 327–338 (2007).

4. J. S. Espinosa, M. P. Stryker, Development and plasticity of the primary visual cortex. *Neuron* **75**, 230–249 (2012).

988 5. B. M. Hooks, C. Chen, Circuitry Underlying Experience-Dependent Plasticity in the Mouse
989 Visual System. *Neuron* **106**, 21–36 (2020).

990 6. J. A. Gordon, M. P. Stryker, Experience-dependent plasticity of binocular responses in the
991 primary visual cortex of the mouse. *J. Neurosci.* **16**, 3274–3286 (1996).

992 7. L. Tan, E. Tring, D. L. Ringach, S. L. Zipursky, J. T. Trachtenberg, Vision Changes the
993 Cellular Composition of Binocular Circuitry during the Critical Period. *Neuron* **108**, 735-
994 747.e6 (2020).

995 8. B.-S. Wang, R. Sarnaik, J. Cang, Critical period plasticity matches binocular orientation
996 preference in the visual cortex. *Neuron* **65**, 246–256 (2010).

997 9. S. Cheng, *et al.*, Vision-dependent specification of cell types and function in the developing
998 cortex. *Cell* **185**, 311-327.e24 (2022).

999 10. M. Adler, Y. Korem Kohanim, A. Tendler, A. Mayo, U. Alon, Continuum of Gene-Expression
1000 Profiles Provides Spatial Division of Labor within a Differentiated Cell Type. *Cell Syst* **8**, 43-
1001 52.e5 (2019).

1002 11. S. Hrvatin, *et al.*, Single-cell analysis of experience-dependent transcriptomic states in the
1003 mouse visual cortex. *Nat. Neurosci.* (2017). https://doi.org/10.1038/s41593-017-0029-5.

1004 12. S.-J. Chou, *et al.*, Geniculocortical input drives genetic distinctions between primary and
1005 higher-order visual areas. *Science* **340**, 1239–1242 (2013).

1006 13. X. Chen, *et al.*, Whole-cortex in situ sequencing reveals input-dependent area identity.
1007 *Nature* 1–10 (2024). https://doi.org/10.1038/s41586-024-07221-6.

1008 14. I. Korsunsky, *et al.*, Fast, sensitive and accurate integration of single-cell data with
1009 Harmony. *Nat. Methods* (2019). https://doi.org/10.1038/s41592-019-0619-0.

1010 15. K. H. Chen, A. N. Boettiger, J. R. Moffitt, S. Wang, X. Zhuang, RNA imaging. Spatially
1011 resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).

1012 16. S. Sahara, Y. Yanagawa, D. D. M. O'Leary, C. F. Stevens, The fraction of cortical
1013 GABAergic neurons is constant from near the start of cortical neurogenesis to adulthood. *J.*
1014 *Neurosci.* **32**, 4755–4761 (2012).

1015 17. Z. Yao, *et al.*, A high-resolution transcriptomic and spatial atlas of cell types in the whole
1016 mouse brain. *Nature* **624**, 317–332 (2023).

1017 18. M. Zhang, *et al.*, Molecularly defined and spatially resolved cell atlas of the whole mouse
1018 brain. *Nature* **624**, 343–354 (2023).

1019 19. L. Haghverdi, M. Büttner, F. A. Wolf, F. Buettner, F. J. Theis, Diffusion pseudotime robustly
1020 reconstructs lineage branching. *Nat. Methods* **13**, 845–848 (2016).

1021 20. R. R. Coifman, *et al.*, Geometric diffusions as a tool for harmonic analysis and structure
1022 definition of data: diffusion maps. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 7426–7431 (2005).

1023 21. N. Eling, M. D. Morgan, J. C. Marioni, Challenges in measuring and understanding
1024 biological noise. *Nat. Rev. Genet.* (2019). https://doi.org/10.1038/s41576-019-0130-6.

1025 22. E. J. Kim, *et al.*, Extraction of Distinct Neuronal Cell Types from within a Genetically
1026 Continuous Population. *Neuron* **107**, 274-282.e6 (2020).

1027 23. B. J. Molyneaux, *et al.*, DeCoN: genome-wide analysis of in vivo transcriptional dynamics
1028 during pyramidal neuron fate selection in neocortex. *Neuron* **85**, 275–288 (2015).

1029 24. Y.-H. Chuang, *et al.*, Parkinson's disease is associated with DNA methylation levels in
1030 human blood and saliva. *Genome Med.* **9**, 76 (2017).

1031 25. Y. Z. Kurmangaliyev, J. Yoo, J. Valdes-Aleman, P. Sanfilippo, S. L. Zipursky,
1032 Transcriptional Programs of Circuit Assembly in the Drosophila Visual System. *Neuron* **108**,
1033 1045-1057.e6 (2020).

1034 26. M. N. Özel, *et al.*, Neuronal diversity and convergence in a visual system developmental
1035 atlas. *Nature* **589**, 88–95 (2021).

1036 27. R. Flamary, *et al.*, POT: Python Optimal Transport. *J. Mach. Learn. Res.* **22**, 1–8 (2021).

1037 28. G. Schiebinger, *et al.*, Optimal-Transport Analysis of Single-Cell Gene Expression Identifies
1038 Developmental Trajectories in Reprogramming. *Cell* **176**, 928-943.e22 (2019).

1039 29. K. D. Harris, G. M. G. Shepherd, The neocortical circuit: themes and variations. *Nat.*
1040 *Neurosci.* **18**, 170–181 (2015).

1041 30. N. A. Morgenstern, J. Bourg, L. Petreanu, Multilaminar networks of cortical neurons
1042 integrate common inputs from sensory thalamus. *Nat. Neurosci.* **19**, 1034–1040 (2016).

1043 31. Q. Lo Giudice, R. J. Wagener, P. Abe, L. Frangeul, D. Jabaudon, Developmental
1044 emergence of first- and higher-order thalamic neuron molecular identities. *bioRxiv*
1045 2024.01.22.576610 (2024).

1046 32. A. M. Xavier, Q. Lin, C. J. Kang, L. Cheadle, A single-cell transcriptomic atlas of sensory-
1047 dependent gene expression in developing mouse visual cortex. *bioRxiv* 2024.06.25.600673
1048 (2024).

1049 33. M. E. Diamond, W. Huang, F. F. Ebner, Laminar comparison of somatosensory cortical
1050 plasticity. *Science* **265**, 1885–1888 (1994).

1051 34. D. E. Feldman, M. Brecht, Map plasticity in somatosensory cortex. *Science* **310**, 810–815
1052 (2005).

1053 35. S. Xu, *et al.*, Behavioral state coding by molecularly defined paraventricular hypothalamic
1054 cell type ensembles. *Science* **370** (2020).

1055 36. C. Condylis, *et al.*, Dense functional and molecular readout of a circuit hub in sensory
1056 cortex. *Science* **375**, eabl5981 (2022).

1057 37. J. W. Kim, *et al.*, Molecular recording of calcium signals via calcium-dependent proximity
1058 labeling. *Nat. Chem. Biol.* **20**, 894–905 (2024).

1059 38. Z. Jiao, *et al.*, Projectome-defined subtypes and modular intra-hypothalamic subnetworks of
1060 peptidergic neurons. *bioRxiv* 2023.05.25.542241 (2023).

1061 39. O. Shoval, *et al.*, Evolutionary trade-offs, Pareto optimality, and the geometry of phenotype
1062 space. *Science* **336**, 1157–1160 (2012).

1063 40. M. Zhang, *et al.*, A molecularly defined and spatially resolved cell atlas of the whole mouse
1064 brain. *bioRxiv* 2023.03.06.531348 (2023).

1065 41. M. Zhang, *et al.*, Spatially resolved cell atlas of the mouse primary motor cortex by
1066 MERFISH. *Nature* **598**, 137–143 (2021).

1067 42. G. Stanley, O. Gokce, R. C. Malenka, T. C. Südhof, S. R. Quake, Continuous and Discrete
1068 Neuron Types of the Adult Murine Striatum. *Neuron* **105**, 688-699.e8 (2020).

1069 43. B. Tasic, *et al.*, Shared and distinct transcriptomic cell types across neocortical areas.
1070 *Nature* **563**, 72–78 (2018).

1071 44. K. D. Harris, *et al.*, Classes and continua of hippocampal CA1 inhibitory neurons revealed
1072 by single-cell transcriptomics. *PLoS Biol.* **16**, e2006387 (2018).

1073 45. V. Kozareva, *et al.*, A transcriptomic atlas of mouse cerebellar cortex comprehensively
1074 defines cell types. *Nature* **598**, 214–219 (2021).

1075 46. Y. Li, *et al.*, Distinct subnetworks of the thalamic reticular nucleus. *Nature* **583**, 819–824
1076 (2020).

1077 47. M. S. Cembrowski, *et al.*, Spatial Gene-Expression Gradients Underlie Prominent
1078 Heterogeneity of CA1 Pyramidal Neurons. *Neuron* **89**, 351–368 (2016).

1079 48. S. Weiler, *et al.*, Functional and structural features of L2/3 pyramidal cells continuously
1080 covary with pial depth in mouse visual cortex. *Cereb. Cortex* **33**, 3715–3733 (2023).

1081 49. Thomas M. Cover, J. A. Thomas, *Elements of Information Theory: Cover/elements of*
1082 *information theory, second edition*, 2nd Ed. (John Wiley & Sons, 2006).

1083 50. C. Villani, *Optimal Transport: Old and New* (Springer Berlin Heidelberg, 2008).

Figure 1

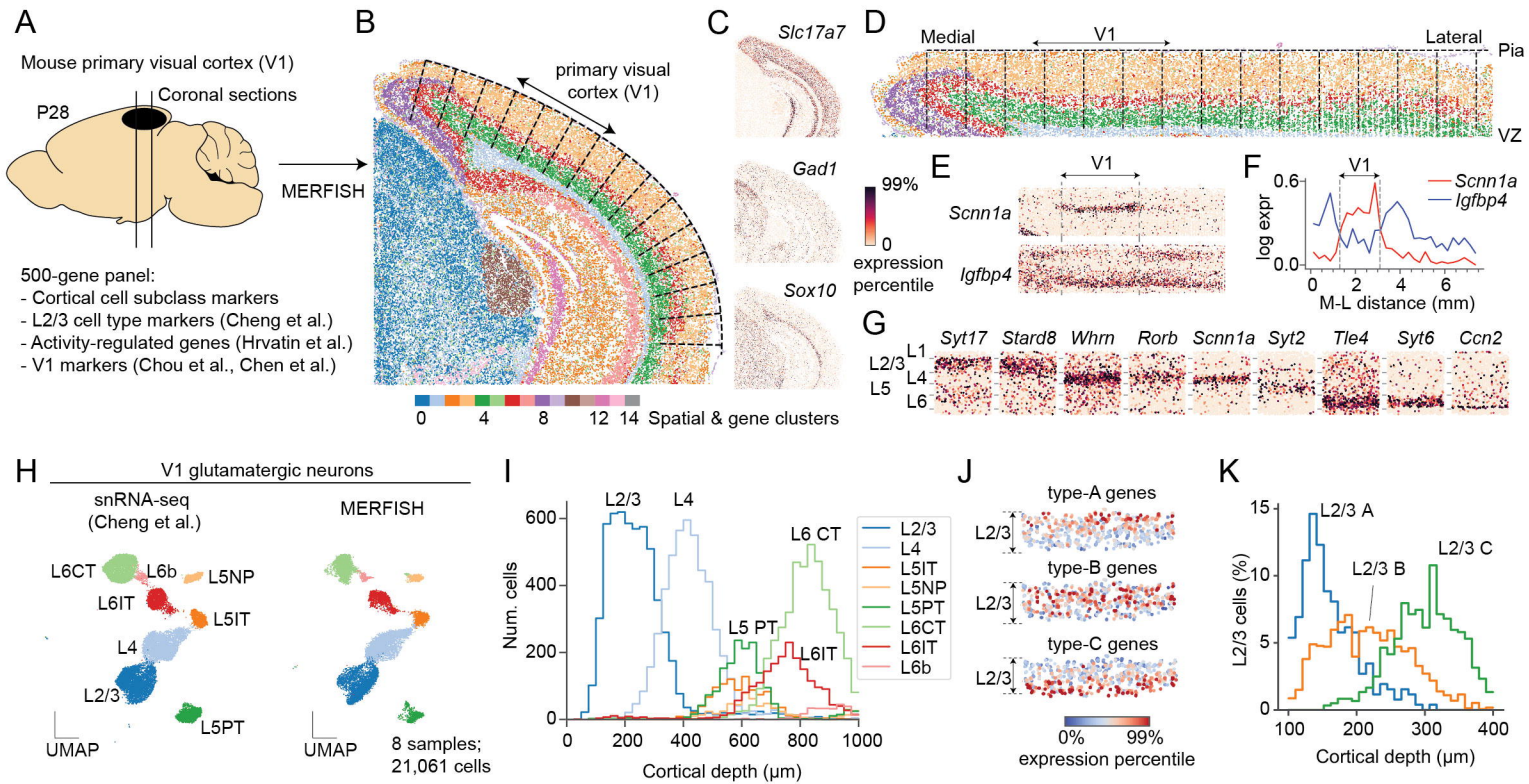

A
Mouse primary visual cortex (V1)
P28
Coronal sections
MERFISH
500-gene panel:
- Cortical cell subclass markers
- L2/3 cell type markers (Cheng et al.)
- Activity-regulated genes (Hrvatin et al.)
- V1 markers (Chou et al., Chen et al.)
B
primary visual cortex (V1)
0 4 8 12 14 Spatial & gene clusters
C
Slc17a7
Gad1
Sox10
99%
0
expression percentile
D
Medial
V1
Lateral
Pia
VZ
E
V1
Scnn1a
Igfbp4
F
V1
Scnn1a
Igfbp4
log expr
0.6
0.0
0 2 4 6
M-L distance (mm)
G
Syt17 Stard8 Whrn Rorb Scnn1a Syt2 Tle4 Syt6 Ccn2
L2/3
L1
L4
L5
L6
H
V1 glutamatergic neurons
snRNA-seq (Cheng et al.)
MERFISH
L6CT
L6b
L5NP
L6IT
L5IT
L4
L2/3
L5PT
UMAP
UMAP
8 samples; 21,061 cells
I
L2/3
L4
L6 CT
L5 PT
L6IT
Num. cells
600
400
200
0
0 200 400 600 800 1000
Cortical depth (μm)
L2/3
L4
L5IT
L5NP
L5PT
L6CT
L6IT
L6b
J
type-A genes
type-B genes
type-C genes
L2/3
0%
99%
expression percentile
K
L2/3 A
L2/3 B
L2/3 C
L2/3 cells (%)
15
10
5
0
100 200 300 400
Cortical depth (μm)

Figure 2

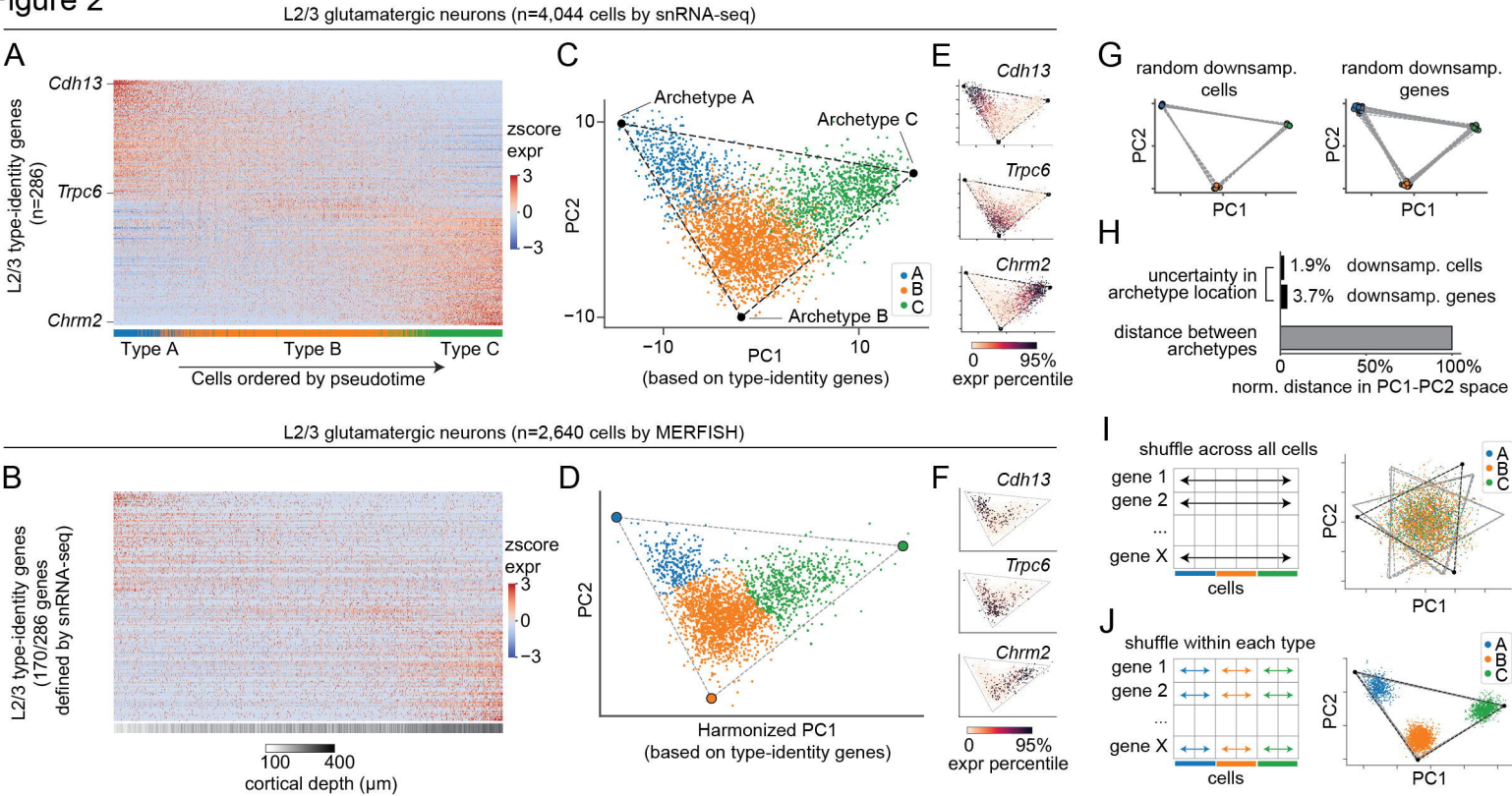

L2/3 glutamatergic neurons (n=4,044 cells by snRNA-seq)
A
Cdh13
Trpc6
Chrm2
L2/3 type-identity genes (n=286)
zscore expr
3
0
−3
Type A
Type B
Type C
Cells ordered by pseudotime
C
Archetype A
Archetype B
Archetype C
PC2
PC1
10
−10
(based on type-identity genes)
A
B
C
E
Cdh13
Trpc6
Chrm2
0
95%
expr percentile
G
random downsamp. cells
random downsamp. genes
PC2
PC1
H
uncertainty in archetype location
1.9% downsamp. cells
3.7% downsamp. genes
distance between archetypes
0
50%
100%
norm. distance in PC1-PC2 space
L2/3 glutamatergic neurons (n=2,640 cells by MERFISH)
B
L2/3 type-identity genes (170/286 genes defined by snRNA-seq)
zscore expr
3
0
−3
100
400
cortical depth (μm)
D
PC2
Harmonized PC1
(based on type-identity genes)
F
Cdh13
Trpc6
Chrm2
0
95%
expr percentile
I
shuffle across all cells
gene 1
gene 2
...
gene X
cells
PC2
PC1
A
B
C
J
shuffle within each type
gene 1
gene 2
...
gene X
cells
PC2
PC1
A
B
C

Figure 3

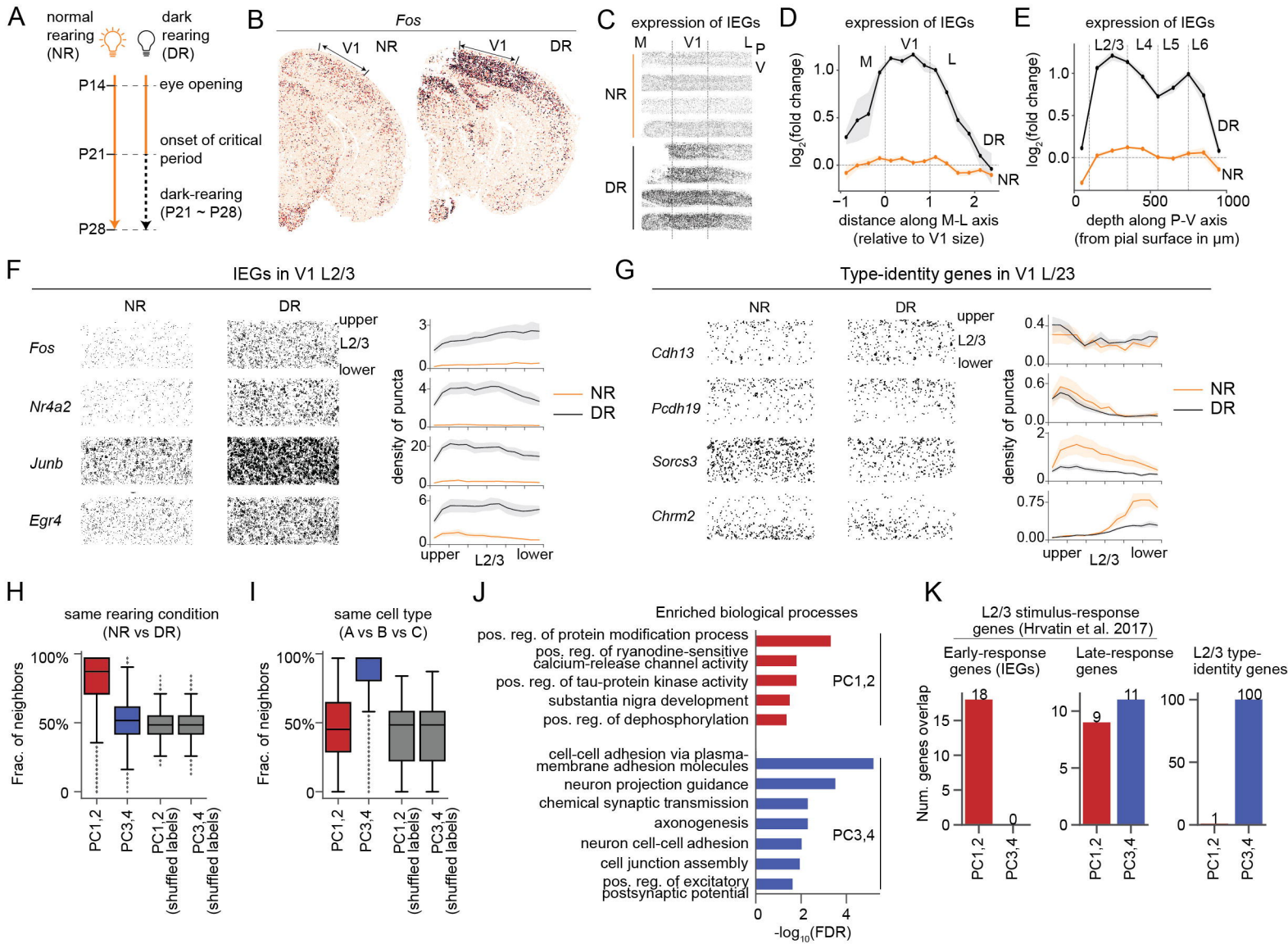

A
normal rearing (NR)
dark rearing (DR)
P14
eye opening
P21
onset of critical period
dark-rearing (P21 ~ P28)
P28
B
Fos
V1
NR
V1
DR
C
expression of IEGs
M
V1
L
P
V
NR
DR
D
expression of IEGs
M
V1
L
DR
NR
log2(fold change)
distance along M-L axis (relative to V1 size)
E
expression of IEGs
L2/3
L4
L5
L6
DR
NR
log2(fold change)
depth along P-V axis (from pial surface in μm)
F
IEGs in V1 L2/3
NR
DR
upper
L2/3
lower
Fos
Nr4a2
Junb
Egr4
density of puncta
NR
DR
upper
L2/3
lower
G
Type-identity genes in V1 L/23
NR
DR
upper
L2/3
lower
Cdh13
Pcdh19
Sorcs3
Chrm2
density of puncta
NR
DR
upper
L2/3
lower
H
same rearing condition (NR vs DR)
Frac. of neighbors
100%
50%
0
PC1,2
PC3,4
PC1,2 (shuffled labels)
PC3,4 (shuffled labels)
I
same cell type (A vs B vs C)
Frac. of neighbors
100%
50%
0
PC1,2
PC3,4
PC1,2 (shuffled labels)
PC3,4 (shuffled labels)
J
Enriched biological processes
pos. reg. of protein modification process
pos. reg. of ryanodine-sensitive calcium-release channel activity
pos. reg. of tau-protein kinase activity
substantia nigra development
pos. reg. of dephosphorylation
PC1,2
cell-cell adhesion via plasma-membrane adhesion molecules
neuron projection guidance
chemical synaptic transmission
axonogenesis
neuron cell-cell adhesion
cell junction assembly
pos. reg. of excitatory postsynaptic potential
PC3,4
-log10(FDR)
K
L2/3 stimulus-response genes (Hrvatin et al. 2017)
Early-response genes (IEGs)
Late-response genes
L2/3 type-identity genes
Num. genes overlap
18
0
9
11
1
100
PC1,2
PC3,4
PC1,2
PC3,4
PC1,2
PC3,4

Figure 4

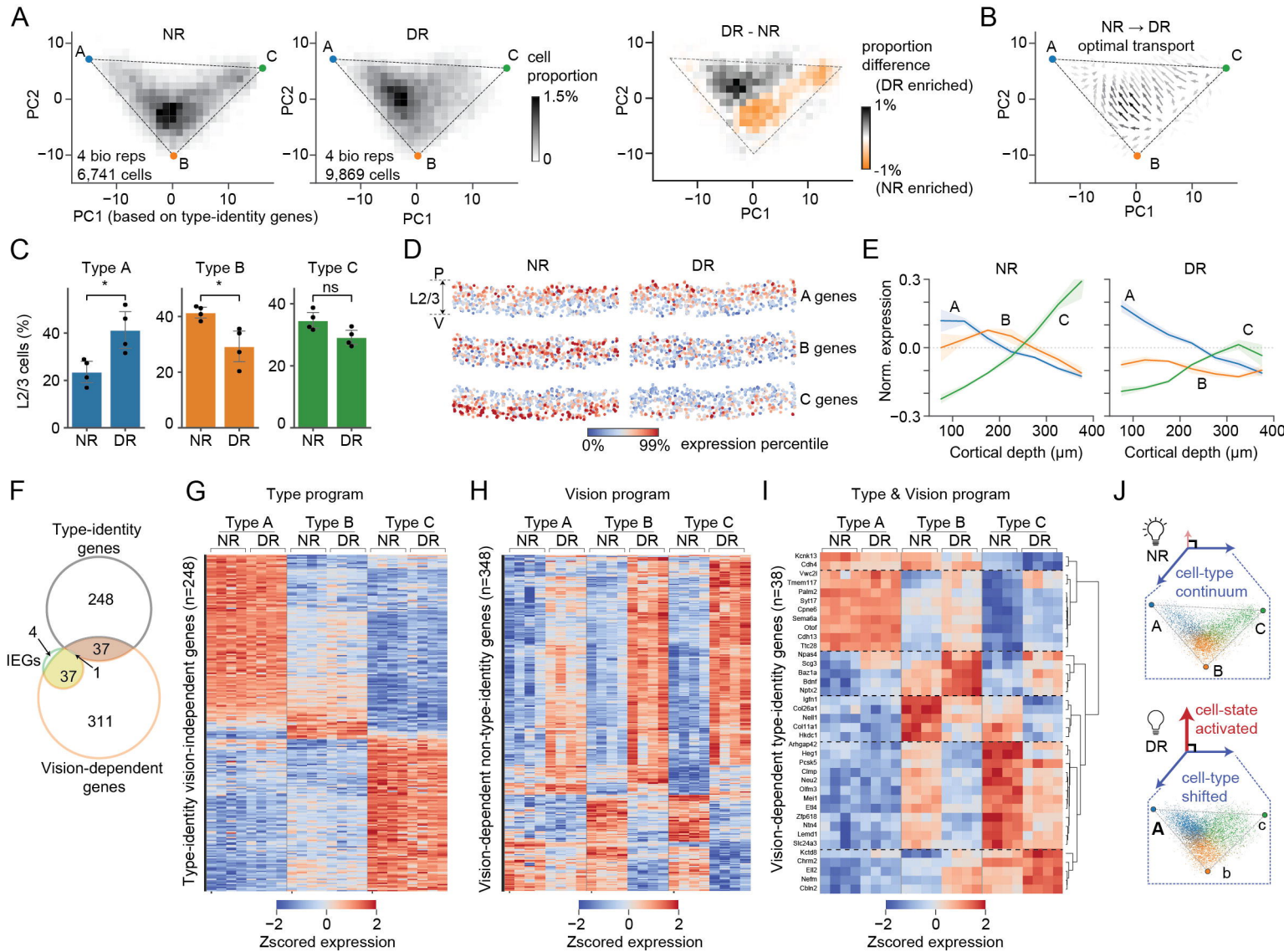

A
NR
DR
DR - NR
cell proportion
1.5%
0
proportion difference
(DR enriched)
1%
-1%
(NR enriched)
4 bio reps
6,741 cells
4 bio reps
9,869 cells
PC1 (based on type-identity genes)
PC1
PC2
B
NR → DR
optimal transport
C
Type A
Type B
Type C
L2/3 cells (%)
NR
DR
ns
D
P
L2/3
V
A genes
B genes
C genes
0%
99%
expression percentile
E
Norm. expression
Cortical depth (μm)
F
Type-identity genes
248
4
37
IEGs
37
1
311
Vision-dependent genes
G
Type program
Type-identity vision-independent genes (n=248)
Zscored expression
H
Vision program
Vision-dependent non-type-identity genes (n=348)
I
Type & Vision program
Vision-dependent type-identity genes (n=38)
J
NR
cell-type continuum
DR
cell-state activated
cell-type shifted

Figure S1

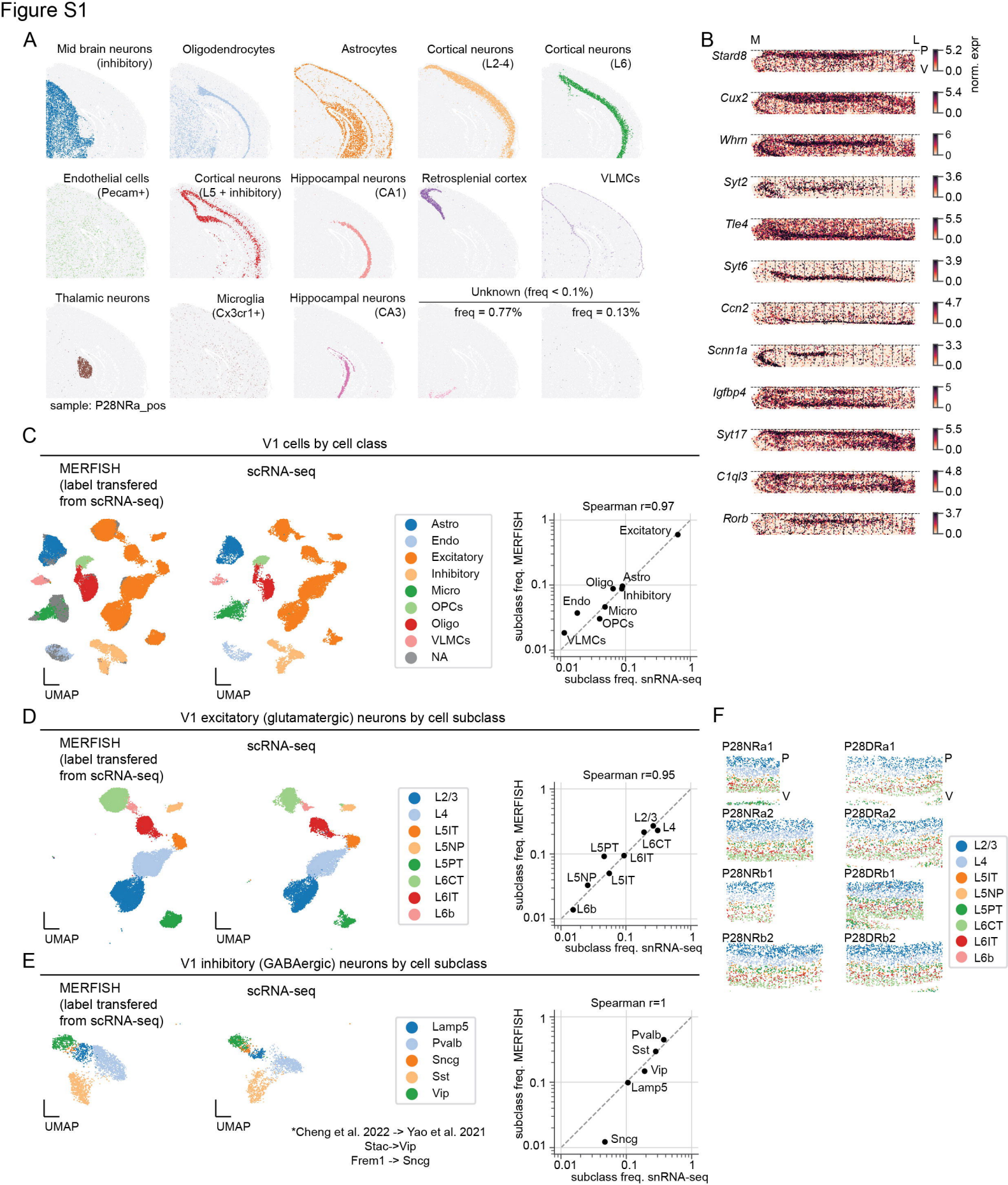

A
Mid brain neurons (inhibitory)
Oligodendrocytes
Astrocytes
Cortical neurons (L2-4)
Cortical neurons (L6)
Endothelial cells (Pecam+)
Cortical neurons (L5 + inhibitory)
Hippocampal neurons (CA1)
Retrosplenial cortex
VLMCs
Thalamic neurons
Microglia (Cx3cr1+)
Hippocampal neurons (CA3)
Unknown (freq < 0.1%)
freq = 0.77%
freq = 0.13%
sample: P28NRa_pos
B
M
L
P
V
Stard8
Cux2
Whrn
Syt2
Tle4
Syt6
Ccn2
Scnn1a
Igfbp4
Syt17
C1ql3
Rorb
norm. expr
5.2
0.0
5.4
0.0
6
0
3.6
0.0
5.5
0.0
3.9
0.0
4.7
0.0
3.3
0.0
5
0
5.5
0.0
4.8
0.0
3.7
0.0
C
V1 cells by cell class
MERFISH (label transfered from scRNA-seq)
scRNA-seq
UMAP
UMAP
Astro
Endo
Excitatory
Inhibitory
Micro
OPCs
Oligo
VLMCs
NA
Spearman r=0.97
subclass freq. MERFISH
subclass freq. snRNA-seq
Excitatory
Oligo
Astro
Inhibitory
Endo
Micro
OPCs
VLMCs
D
V1 excitatory (glutamatergic) neurons by cell subclass
MERFISH (label transfered from scRNA-seq)
scRNA-seq
UMAP
UMAP
L2/3
L4
L5IT
L5NP
L5PT
L6CT
L6IT
L6b
Spearman r=0.95
subclass freq. MERFISH
subclass freq. snRNA-seq
L2/3
L4
L6CT
L5PT
L6IT
L5NP
L5IT
L6b
E
V1 inhibitory (GABAergic) neurons by cell subclass
MERFISH (label transfered from scRNA-seq)
scRNA-seq
UMAP
UMAP
Lamp5
Pvalb
Sncg
Sst
Vip
*Cheng et al. 2022 -> Yao et al. 2021
Stac->Vip
Frem1 -> Sncg
Spearman r=1
subclass freq. MERFISH
subclass freq. snRNA-seq
Pvalb
Sst
Vip
Lamp5
Sncg
F
P28NRa1
P28DRa1
P28NRa2
P28DRa2
P28NRb1
P28DRb1
P28NRb2
P28DRb2
P
V
L2/3
L4
L5IT
L5NP
L5PT
L6CT
L6IT
L6b

Figure S2

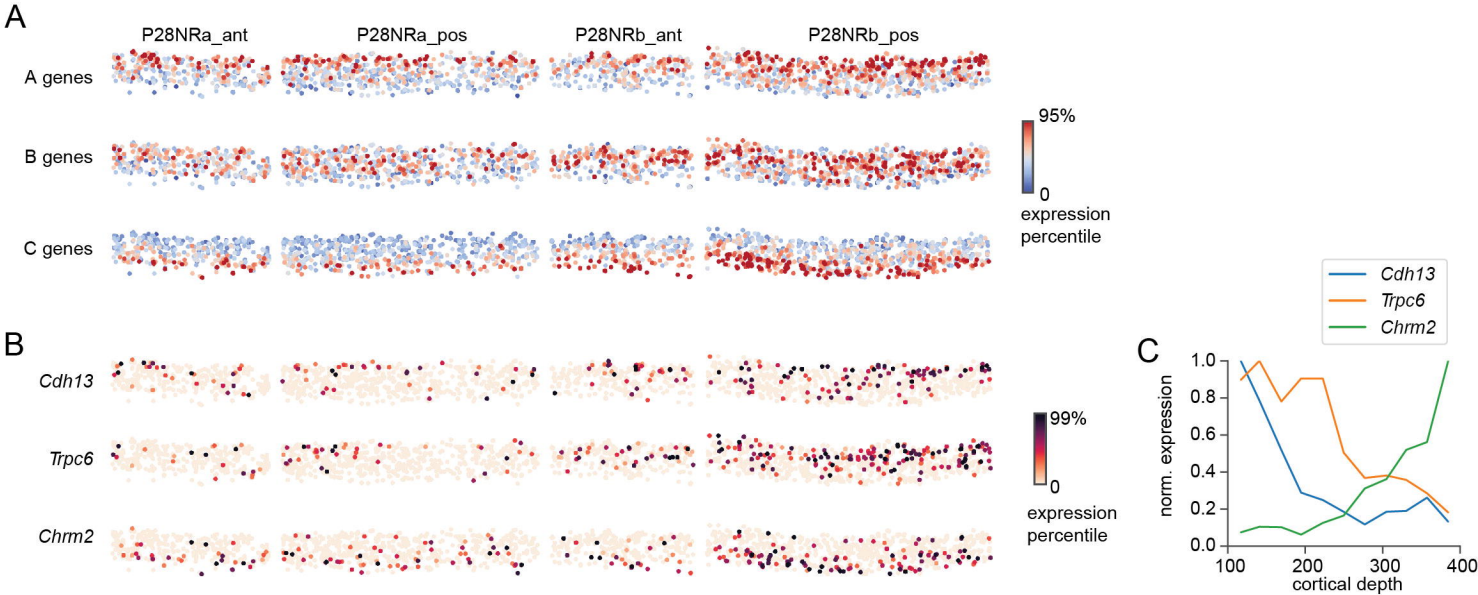

A
P28NRa_ant
P28NRa_pos
P28NRb_ant
P28NRb_pos
A genes
B genes
C genes
95%
0
expression percentile
B
Cdh13
Trpc6
Chrm2
99%
0
expression percentile
C
Cdh13
Trpc6
Chrm2
norm. expression
1.0
0.8
0.6
0.4
0.2
0.0
100
200
300
400
cortical depth

Figure S3

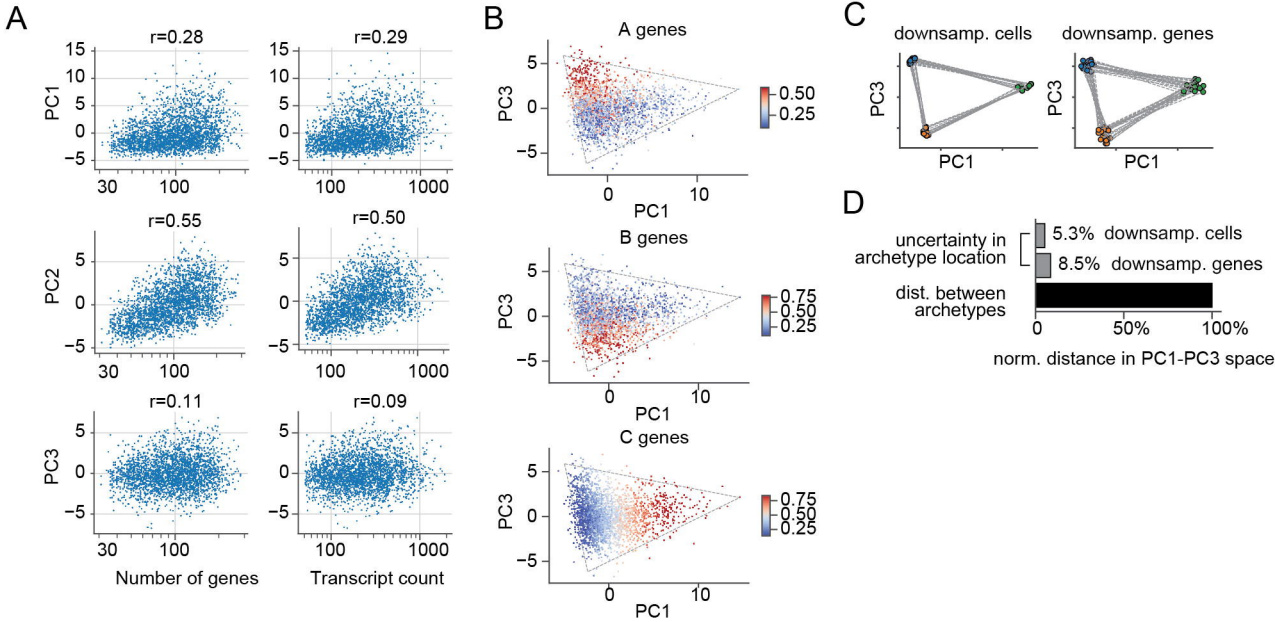

A
r=0.28
r=0.29
r=0.55
r=0.50
r=0.11
r=0.09
PC1
PC2
PC3
Number of genes
Transcript count
B
A genes
B genes
C genes
PC3
PC1
0.50
0.25
0.75
C
downsamp. cells
downsamp. genes
PC3
PC1
D
uncertainty in archetype location
5.3% downsamp. cells
8.5% downsamp. genes
dist. between archetypes
0
50%
100%
norm. distance in PC1-PC3 space

Figure S4

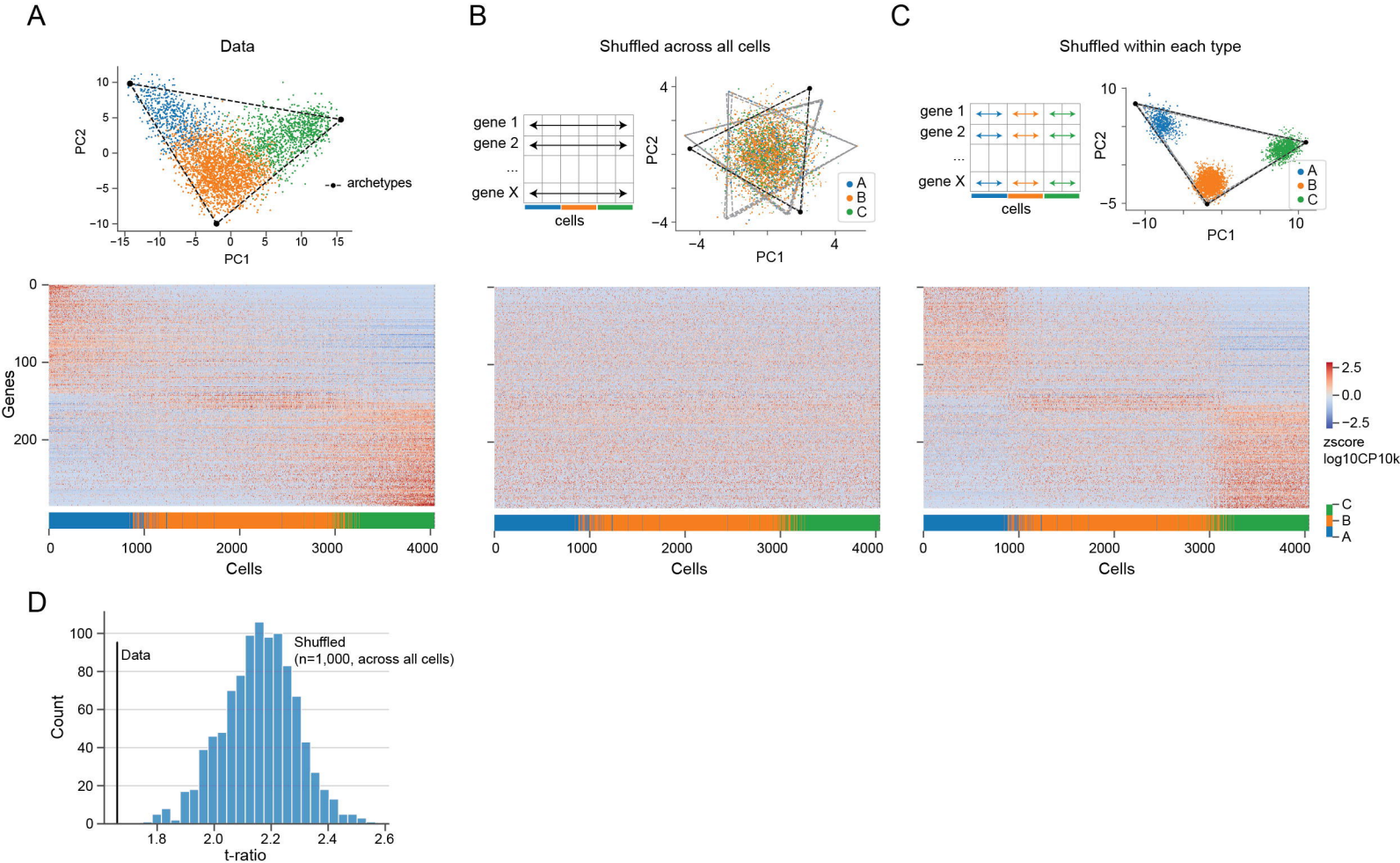

A
Data
PC2
PC1
archetypes
B
Shuffled across all cells
gene 1
gene 2
...
gene X
cells
PC2
PC1
A
B
C
C
Shuffled within each type
gene 1
gene 2
...
gene X
cells
PC2
PC1
A
B
C
Genes
Cells
zscore
log10CP10k
D
Data
Shuffled
(n=1,000, across all cells)
Count
t-ratio

Figure S5

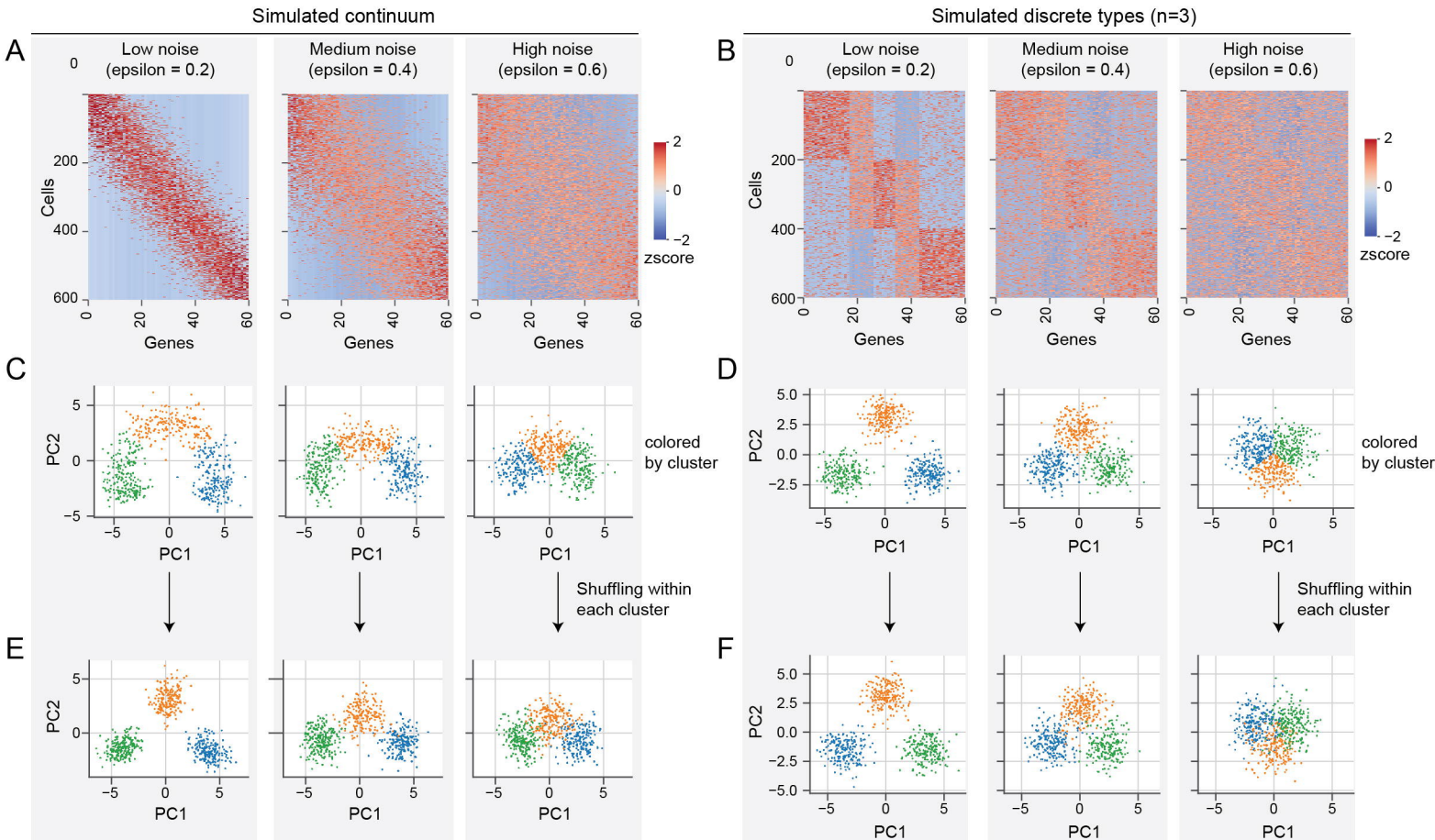

Simulated continuum
Simulated discrete types (n=3)
A
B
C
D
E
F
Low noise (epsilon = 0.2)
Medium noise (epsilon = 0.4)
High noise (epsilon = 0.6)
Cells
Genes
zscore
PC1
PC2
colored by cluster
Shuffling within each cluster

Figure S6

A

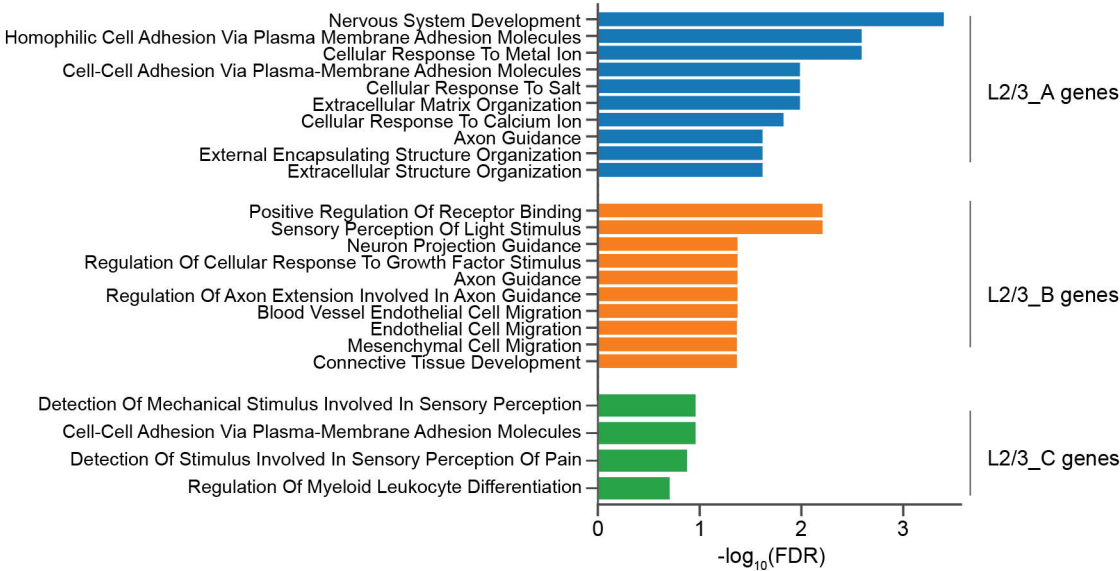

Nervous System Development
Homophilic Cell Adhesion Via Plasma Membrane Adhesion Molecules
Cellular Response To Metal Ion
Cell-Cell Adhesion Via Plasma-Membrane Adhesion Molecules
Cellular Response To Salt
Extracellular Matrix Organization
Cellular Response To Calcium Ion
Axon Guidance
External Encapsulating Structure Organization
Extracellular Structure Organization
L2/3_A genes
Positive Regulation Of Receptor Binding
Sensory Perception Of Light Stimulus
Neuron Projection Guidance
Regulation Of Cellular Response To Growth Factor Stimulus
Axon Guidance
Regulation Of Axon Extension Involved In Axon Guidance
Blood Vessel Endothelial Cell Migration
Endothelial Cell Migration
Mesenchymal Cell Migration
Connective Tissue Development
L2/3_B genes
Detection Of Mechanical Stimulus Involved In Sensory Perception
Cell-Cell Adhesion Via Plasma-Membrane Adhesion Molecules
Detection Of Stimulus Involved In Sensory Perception Of Pain
Regulation Of Myeloid Leukocyte Differentiation
L2/3_C genes
0
1
2
3
-log10(FDR)

Figure S7

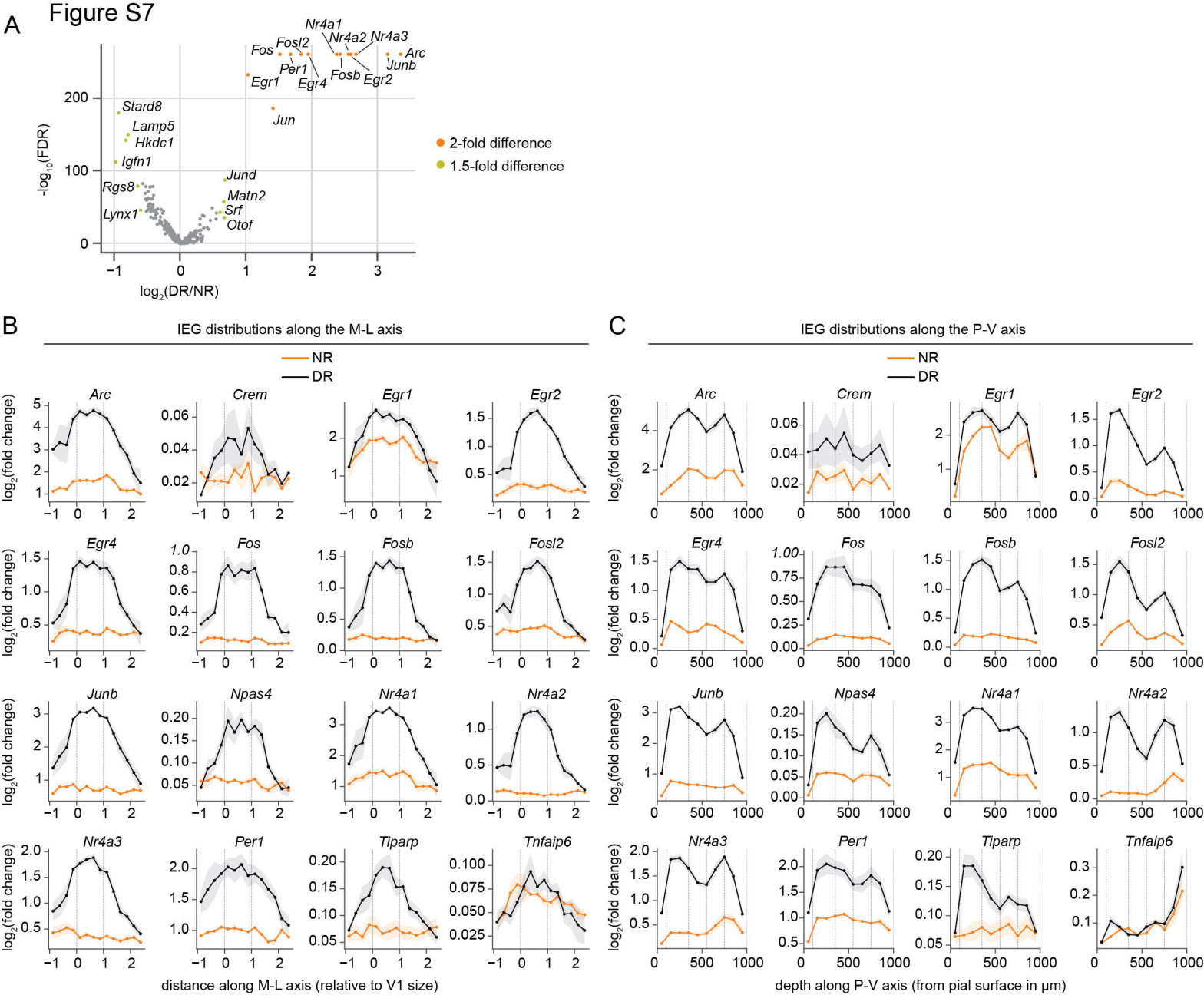

A
Nr4a1
Nr4a2
Nr4a3
Fos
Fosl2
Arc
Junb
Per1
Egr1
Egr4
Fosb
Egr2
Jun
Stard8
Lamp5
Hkdc1
Igfn1
Jund
Rgs8
Matn2
Lynx1
Srf
Otof
-log10(FDR)
log2(DR/NR)
2-fold difference
1.5-fold difference
B
IEG distributions along the M-L axis
NR
DR
Arc
Crem
Egr1
Egr2
Egr4
Fos
Fosb
Fosl2
Junb
Npas4
Nr4a1
Nr4a2
Nr4a3
Per1
Tiparp
Tnfaip6
log2(fold change)
distance along M-L axis (relative to V1 size)
C
IEG distributions along the P-V axis
depth along P-V axis (from pial surface in µm)

Figure S8

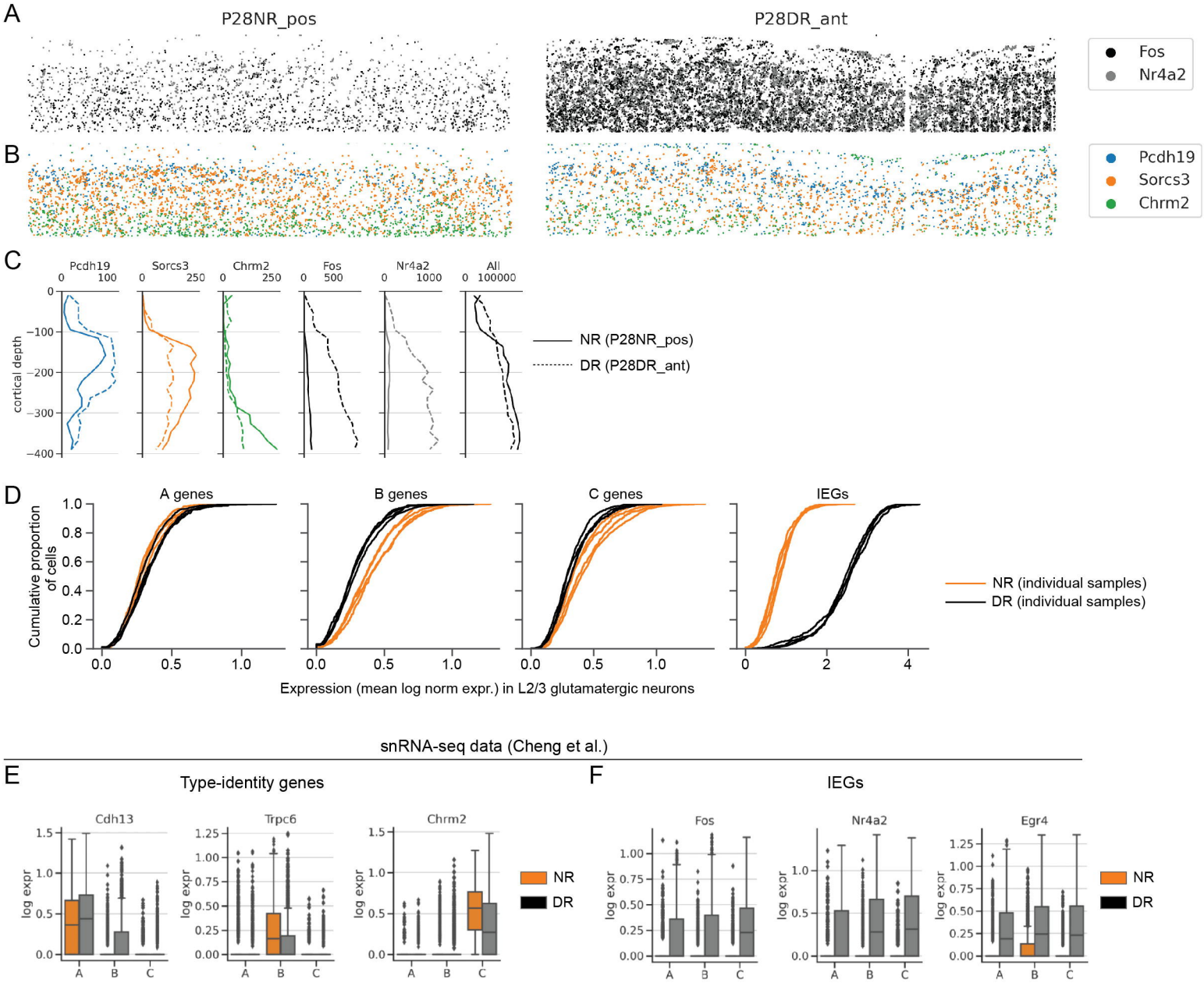

A
P28NR_pos
P28DR_ant
Fos
Nr4a2
B
Pcdh19
Sorcs3
Chrm2
C
Pcdh19
Sorcs3
Chrm2
Fos
Nr4a2
All
cortical depth
NR (P28NR_pos)
DR (P28DR_ant)
D
A genes
B genes
C genes
IEGs
Cumulative proportion of cells
NR (individual samples)
DR (individual samples)
Expression (mean log norm expr.) in L2/3 glutamatergic neurons
snRNA-seq data (Cheng et al.)
E
Type-identity genes
Cdh13
Trpc6
Chrm2
log expr
NR
DR
F
IEGs
Fos
Nr4a2
Egr4
log expr
NR
DR

Figure S9

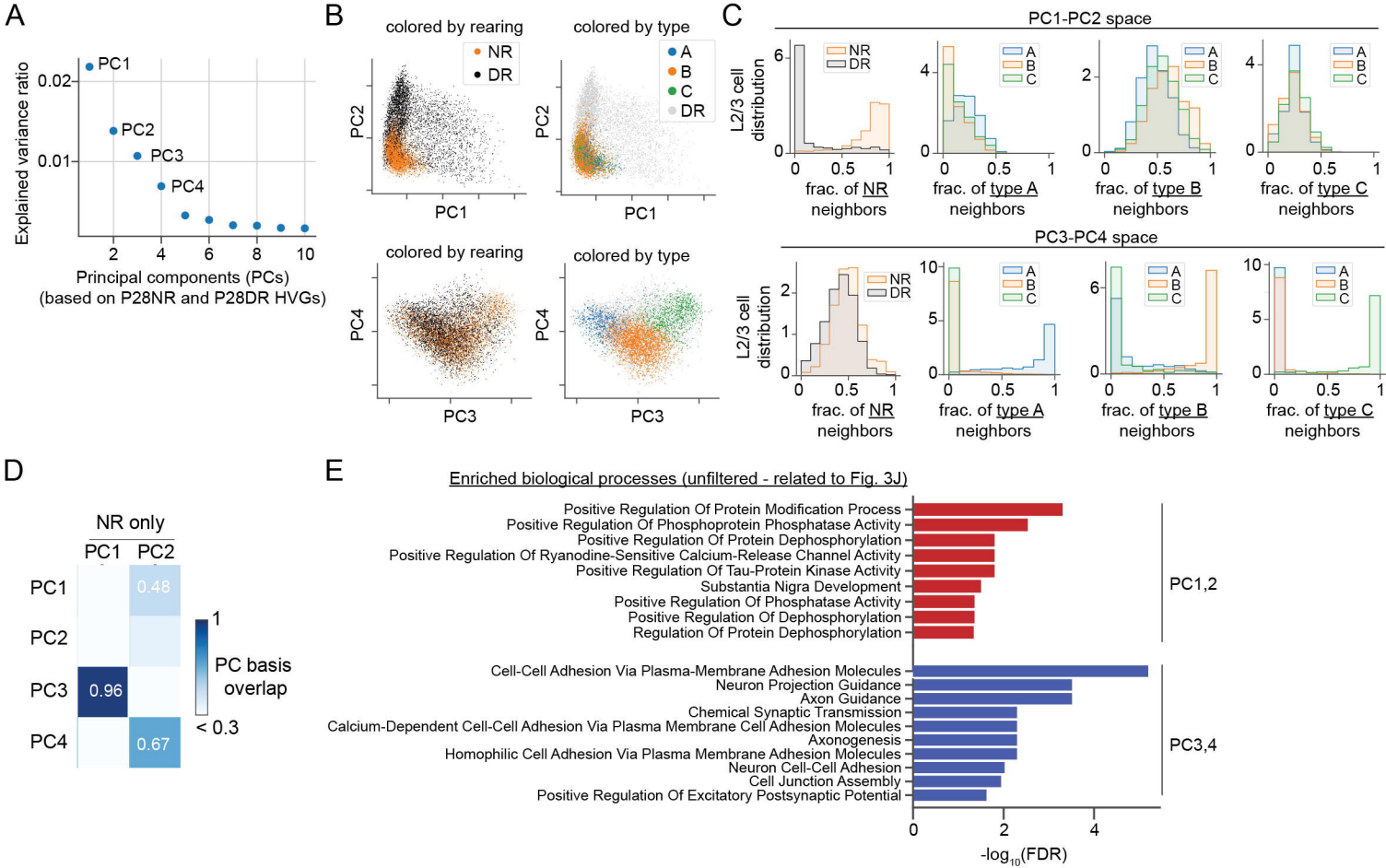

A
Explained variance ratio
0.02
0.01
PC1
PC2
PC3
PC4
2
4
6
8
10
Principal components (PCs)
(based on P28NR and P28DR HVGs)
B
colored by rearing
NR
DR
colored by type
A
B
C
DR
PC2
PC1
PC4
PC3
C
PC1-PC2 space
PC3-PC4 space
L2/3 cell distribution
NR
DR
A
B
C
frac. of NR neighbors
frac. of type A neighbors
frac. of type B neighbors
frac. of type C neighbors
D
NR only
PC1
PC2
PC3
PC4
0.48
0.96
0.67
1
PC basis overlap
< 0.3
E
Enriched biological processes (unfiltered - related to Fig. 3J)
Positive Regulation Of Protein Modification Process
Positive Regulation Of Phosphoprotein Phosphatase Activity
Positive Regulation Of Protein Dephosphorylation
Positive Regulation Of Ryanodine-Sensitive Calcium-Release Channel Activity
Positive Regulation Of Tau-Protein Kinase Activity
Substantia Nigra Development
Positive Regulation Of Phosphatase Activity
Positive Regulation Of Dephosphorylation
Regulation Of Protein Dephosphorylation
PC1,2
Cell-Cell Adhesion Via Plasma-Membrane Adhesion Molecules
Neuron Projection Guidance
Axon Guidance
Chemical Synaptic Transmission
Calcium-Dependent Cell-Cell Adhesion Via Plasma Membrane Cell Adhesion Molecules
Axonogenesis
Homophilic Cell Adhesion Via Plasma Membrane Adhesion Molecules
Neuron Cell-Cell Adhesion
Cell Junction Assembly
Positive Regulation Of Excitatory Postsynaptic Potential
PC3,4
0
2
4
-log10(FDR)

Figure S10

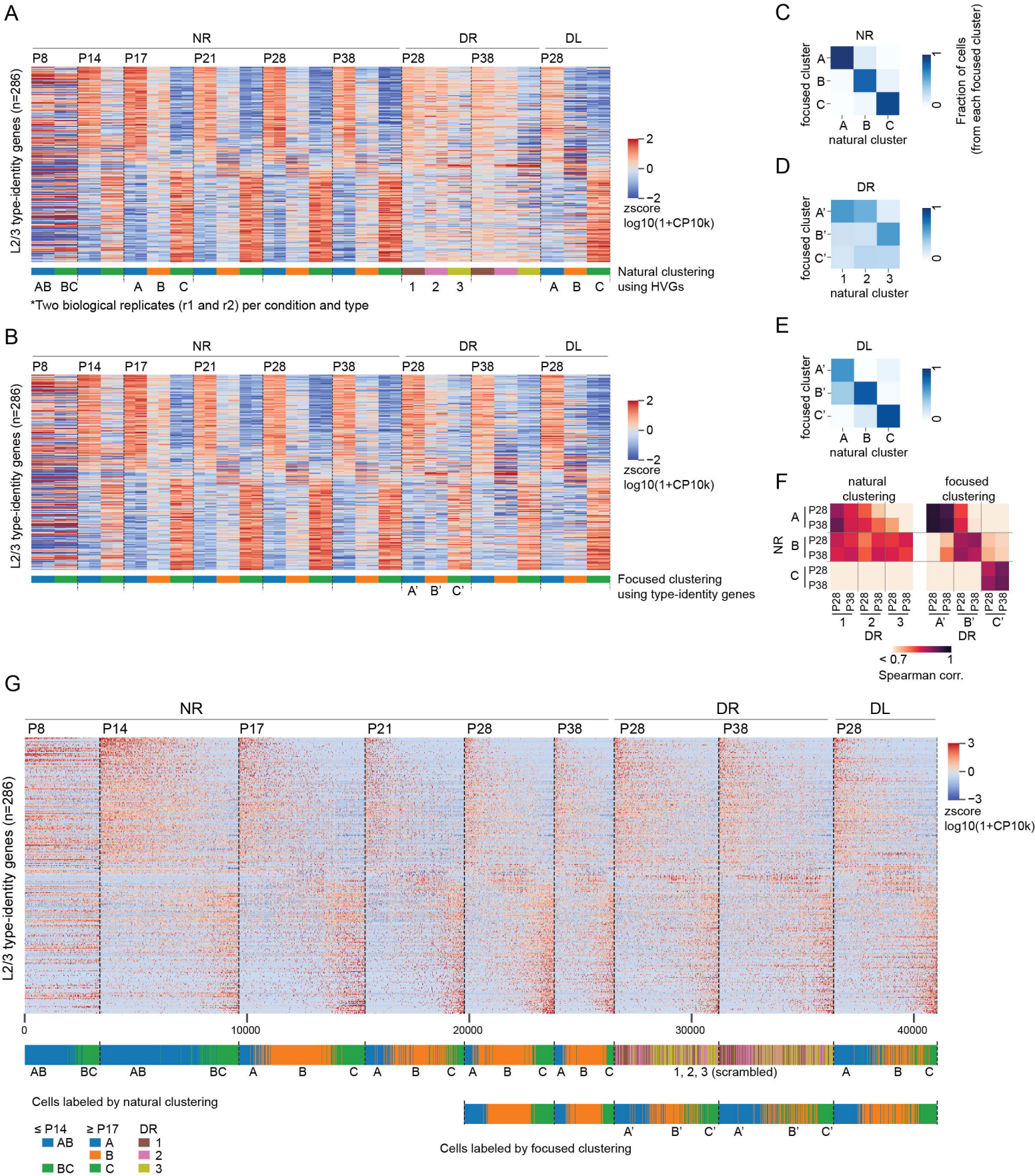
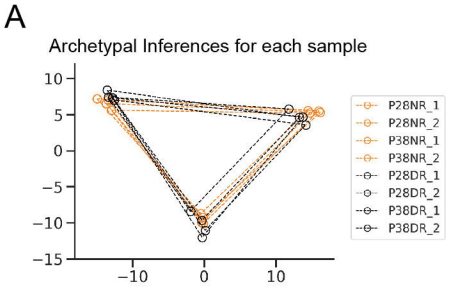
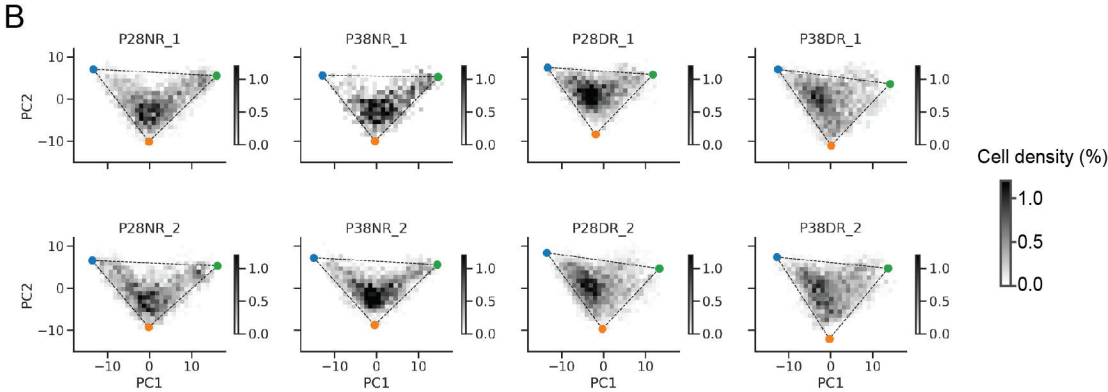
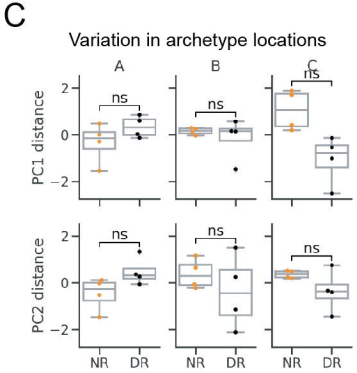

A
NR
DR
DL
P8
P14
P17
P21
P28
P38
P28
P38
P28
L2/3 type-identity genes (n=286)
2
0
−2
zscore
log10(1+CP10k)
Natural clustering
using HVGs
AB BC
A B C
1 2 3
A B C
*Two biological replicates (r1 and r2) per condition and type
B
Focused clustering
using type-identity genes
A' B' C'
C
D
E
focused cluster
natural cluster
Fraction of cells
(from each focused cluster)
F
natural
clustering
focused
clustering
P28
P38
< 0.7
Spearman corr.
G
0
10000
20000
30000
40000
1, 2, 3 (scrambled)
Cells labeled by natural clustering
≤ P14
≥ P17
DR
Cells labeled by focused clustering

Figure S11


A
Archetypal Inferences for each sample
P28NR_1
P28NR_2
P38NR_1
P38NR_2
P28DR_1
P28DR_2
P38DR_1
P38DR_2


B
P28NR_1
P38NR_1
P28DR_1
P38DR_1
P28NR_2
P38NR_2
P28DR_2
P38DR_2
PC1
PC2
Cell density (%)


C
Variation in archetype locations
A
B
C
PC1 distance
PC2 distance
ns
NR
DR

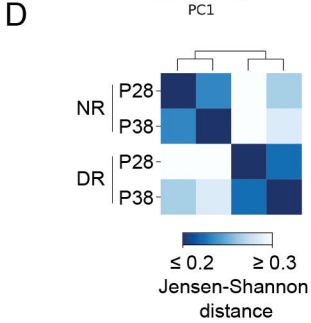

D
NR
DR
P28
P38
≤ 0.2
≥ 0.3
Jensen-Shannon distance

Figure S12

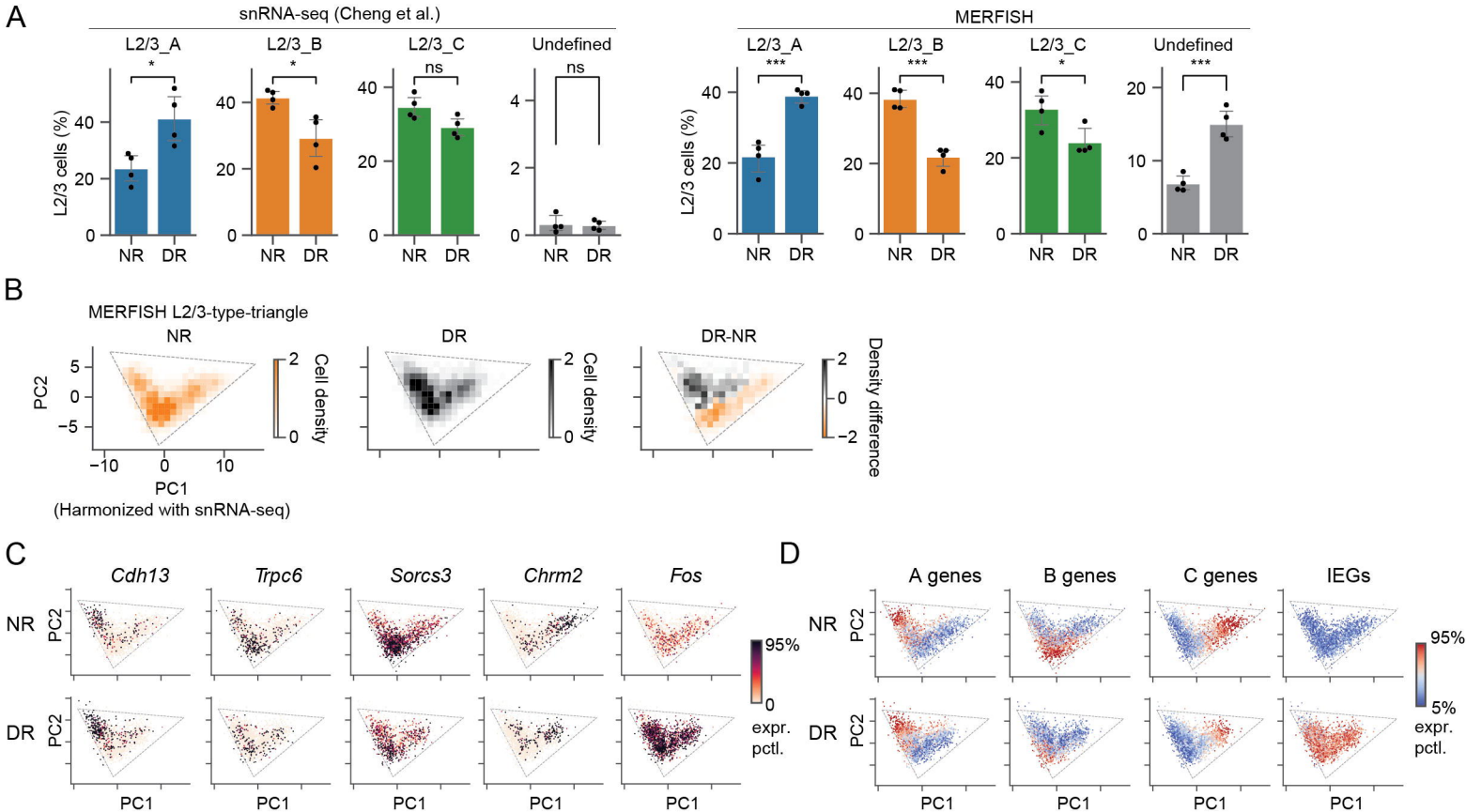

A
snRNA-seq (Cheng et al.)
MERFISH
L2/3_A
L2/3_B
L2/3_C
Undefined
L2/3 cells (%)
NR
DR
B
MERFISH L2/3-type-triangle
NR
DR
DR-NR
PC2
PC1
(Harmonized with snRNA-seq)
Cell density
Density difference
C
Cdh13
Trpc6
Sorcs3
Chrm2
Fos
95%
0
expr. pctl.
D
A genes
B genes
C genes
IEGs
5%

Figure S13

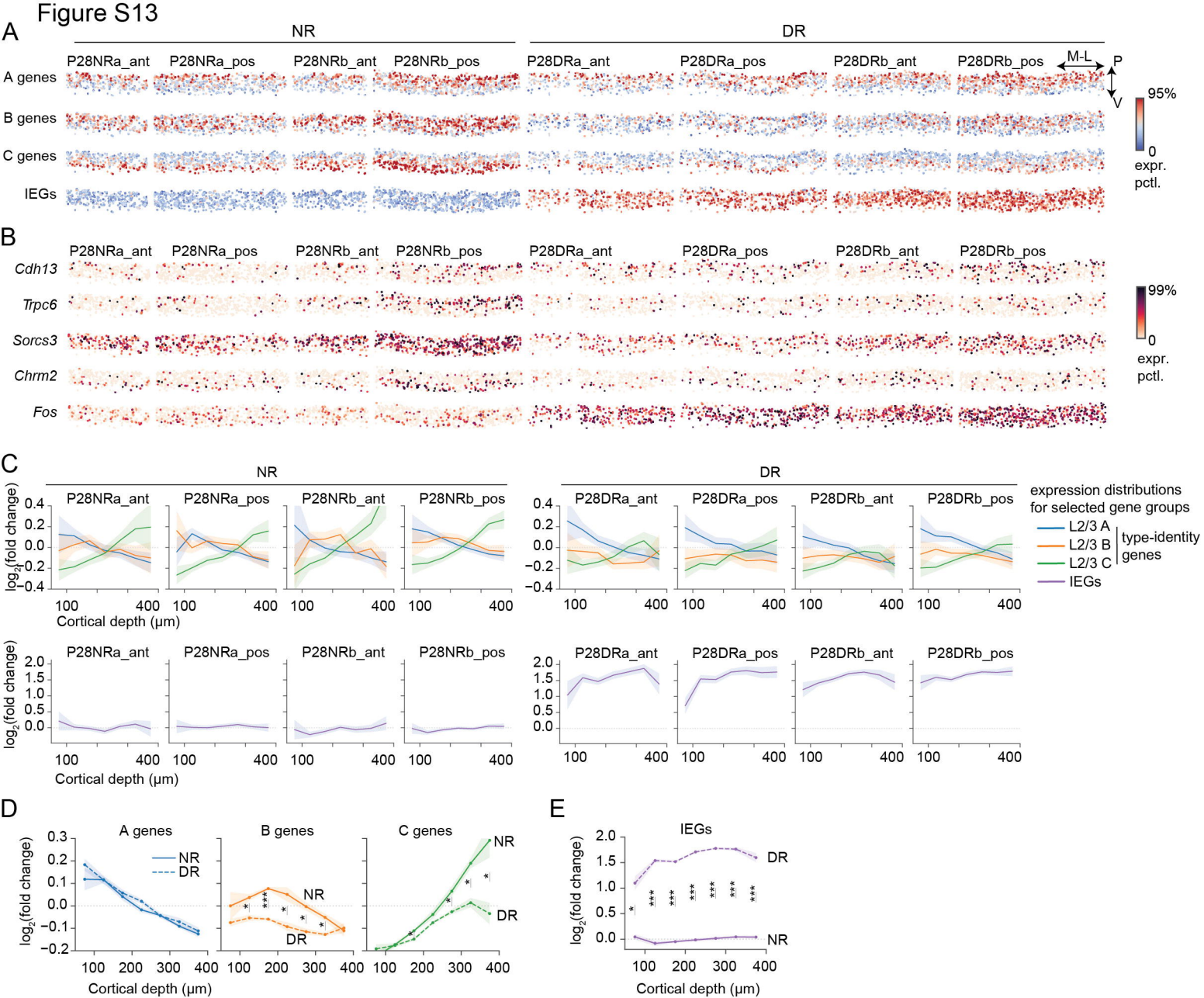
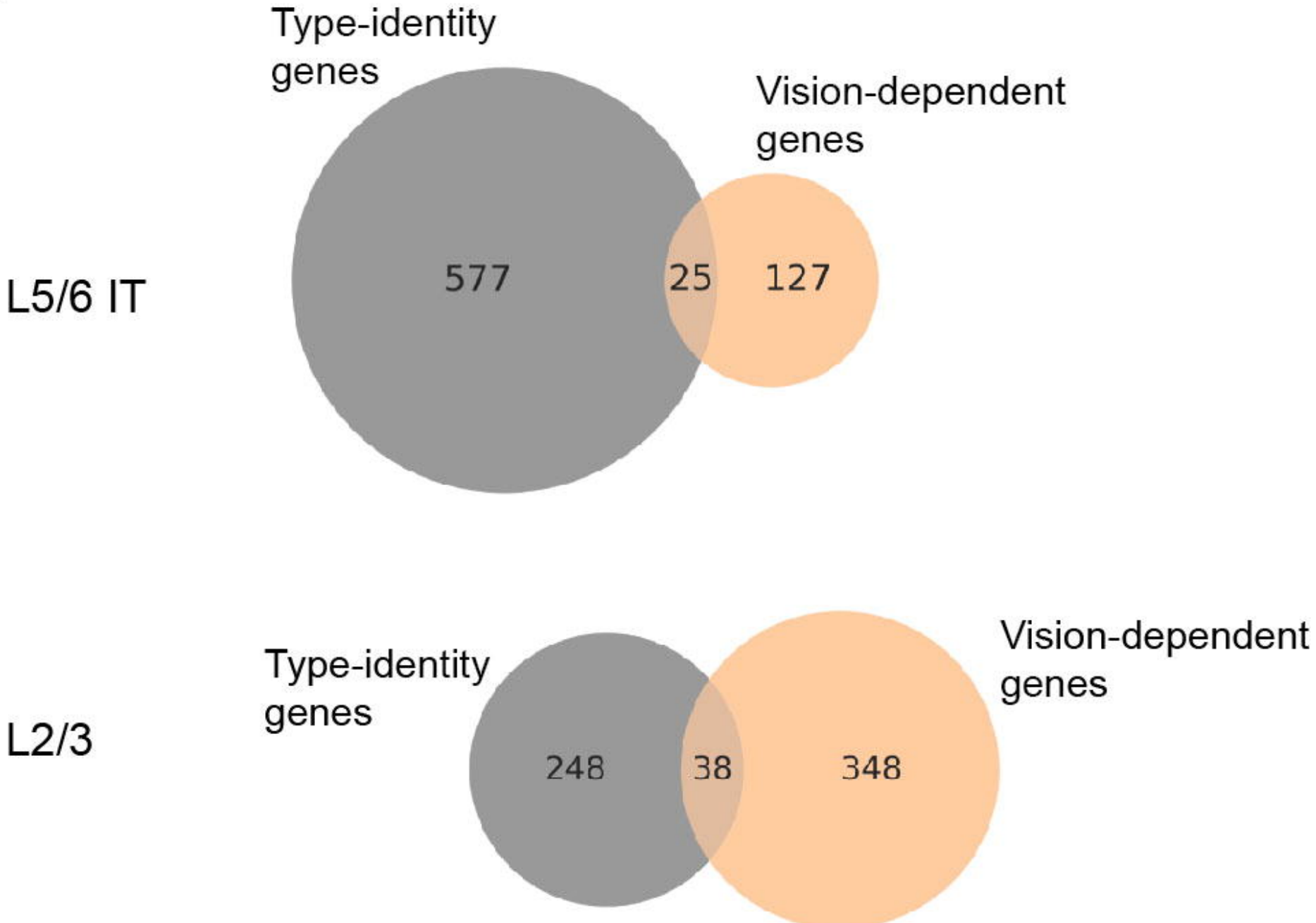

A
NR
DR
P28NRa_ant
P28NRa_pos
P28NRb_ant
P28NRb_pos
P28DRa_ant
P28DRa_pos
P28DRb_ant
P28DRb_pos
M-L
P
V
A genes
B genes
C genes
IEGs
95%
0
expr. pctl.
B
Cdh13
Trpc6
Sorcs3
Chrm2
Fos
99%
C
log2(fold change)
Cortical depth (μm)
expression distributions for selected gene groups
L2/3 A
L2/3 B
L2/3 C
type-identity genes
IEGs
D
A genes
B genes
C genes
NR
DR
E
IEGs

Figure S14


Type-identity genes
Vision-dependent genes
L5/6 IT
577
25
127
Type-identity genes
Vision-dependent genes
L2/3
248
38
348