Article

Meso-seq for in-depth transcriptomics in ultra-low amounts of FACS-purified neuronal nuclei

Graphical abstract

Highlights

- Meso-seq reliably identifies over 10,000 genes in 50 FACS-sorted neuronal nuclei

- Meso-seq works well for neurons labeled sparsely with antibodies and genetic markers

- Meso-seq is simple, robust, and cost effective

- Meso-seq is ideal for mechanistic studies in ultra-sparse neuronal populations

Authors

Daniella Apelblat, Ori Roethler, Lidor Bitan, Hadas Keren-Shaul, Ivo Spiegel

Correspondence

ivo.spiegel@weizmann.ac.il

In brief

Apelblat et al. introduce Meso-seq, a simple and robust transcriptomic method that identifies >10,000 reliably expressed genes in as little as 50 FACS-sorted neuronal nuclei. Meso-seq works well in multiple neuroscience applications and is ideal for molecular-mechanistic studies with large sample numbers in ultra-sparse neuronal populations in the brain.

CellPress

Article

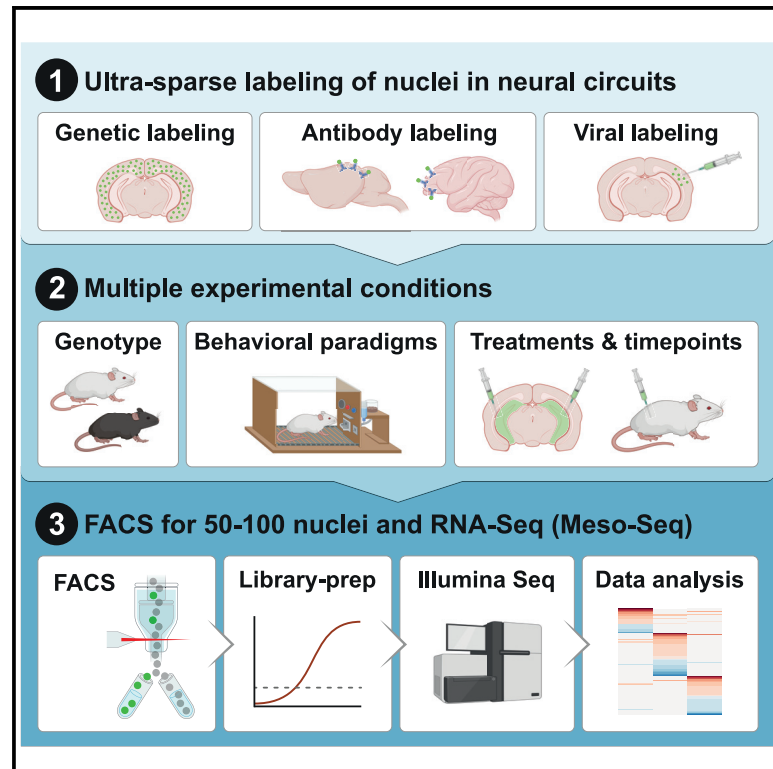# Meso-seq for in-depth transcriptomics in ultra-low amounts of FACS-purified neuronal nuclei

Daniella Apelblat,[1,2,5] Ori Roethler,[1,2,5] Lidor Bitan,[1] Hadas Keren-Shaul,[3,4] and Ivo Spiegel[1,2,6,*]
[1]Department of Brain Sciences, Weizmann Institute of Science, Rehovot, Israel
[2]Department of Molecular Neuroscience, Weizmann Institute of Science, Rehovot, Israel
[3]Life Science Core Facility, Weizmann Institute of Science, Rehovot, Israel
[4]The Nancy & Stephen Grand Israel National Center for Personalized Medicine, Weizmann Institute of Science, Rehovot, Israel
[5]These authors contributed equally
[6]Lead contact
*Correspondence: ivo.spiegel@weizmann.ac.il
https://doi.org/10.1016/j.crmeth.2022.100259

MOTIVATION RNA-seq in sparse populations of genetically defined neurons is essential for dissecting the molecular mechanisms that control the development and plasticity of neural circuits. However, current transcriptomic approaches are ill suited for detailed mechanistic studies in sparse neuronal populations, as they either are technically complex and relatively expensive (e.g., single-cell RNA-seq) or require large amounts of input material (e.g., traditional bulk RNA-seq). Thus, we sought to establish a simple, robust, and cost-effective method for in-depth RNA-seq in multiple samples derived from ultra-sparse neuronal populations in the brain.

SUMMARY

Profiling of gene expression in sparse populations of genetically defined neurons is essential for dissecting the molecular mechanisms that control the development and plasticity of neural circuits. However, current transcriptomic approaches are ill suited for detailed mechanistic studies in sparse neuronal populations, as they either are technically complex and relatively expensive (e.g., single-cell RNA sequencing [RNA-seq]) or require large amounts of input material (e.g., traditional bulk RNA-seq). Thus, we established Meso-seq, a meso-scale protocol for identifying more than 10,000 robustly expressed genes in as little as 50 FACS-sorted neuronal nuclei. We demonstrate that Meso-seq works well for multiple neuroscience applications, including transcriptomics in antibody-labeled cortical neurons in mice and non-human primates, analyses of experience-regulated gene programs, and RNA-seq from visual cortex neurons labeled ultra-sparsely with viruses. Given its simplicity, robustness, and relatively low costs, Meso-seq is well suited for molecular-mechanistic studies in ultra-sparse neuronal populations in the brain.

INTRODUCTION

The brain is built from hundreds of molecularly distinct cell types, and key questions in neuroscience concern the cell-type-specific molecular mechanisms that control the developmental assembly of different cell types into neural circuits and their experience-dependent modulation, e.g., during learning. Cell-type-specific profiling of gene expression by RNA sequencing (RNA-seq) has revolutionized our understanding of these mechanisms by providing a precise census of the cell types in the brain and by identifying a large number of candidate genes and putative molecular mechanisms that may control the development and function of specific neural circuits. Thus, major challenges now lie in testing the hypotheses emerging from transcriptomic analyses via targeted molecular manipulations in specific cell types across multiple developmental and physiological conditions. However, such mechanistic experiments are technically challenging, since they often require in-depth transcriptomic analyses in sparse genetically defined cell types across multiple conditions (e.g., multiple genotypes, timepoints, etc.). Thus, there is a need for a simple, robust, and cost-effective method to RNA-seq in-depth multiple samples derived from sparse populations of genetically defined neurons.

RNA-seq from single cells or nuclei (single-cell RNA-seq [scRNA-seq] or single-nucleus RNA-seq [snRNA-seq]) has transformed neuroscience, since it allows for precise high-throughput measurements of the cellular heterogeneity of neural circuits, even from relatively sparse neuronal populations. Indeed, scRNA-seq- and snRNA-seq-based approaches are currently used in a wide range of neuroscience applications
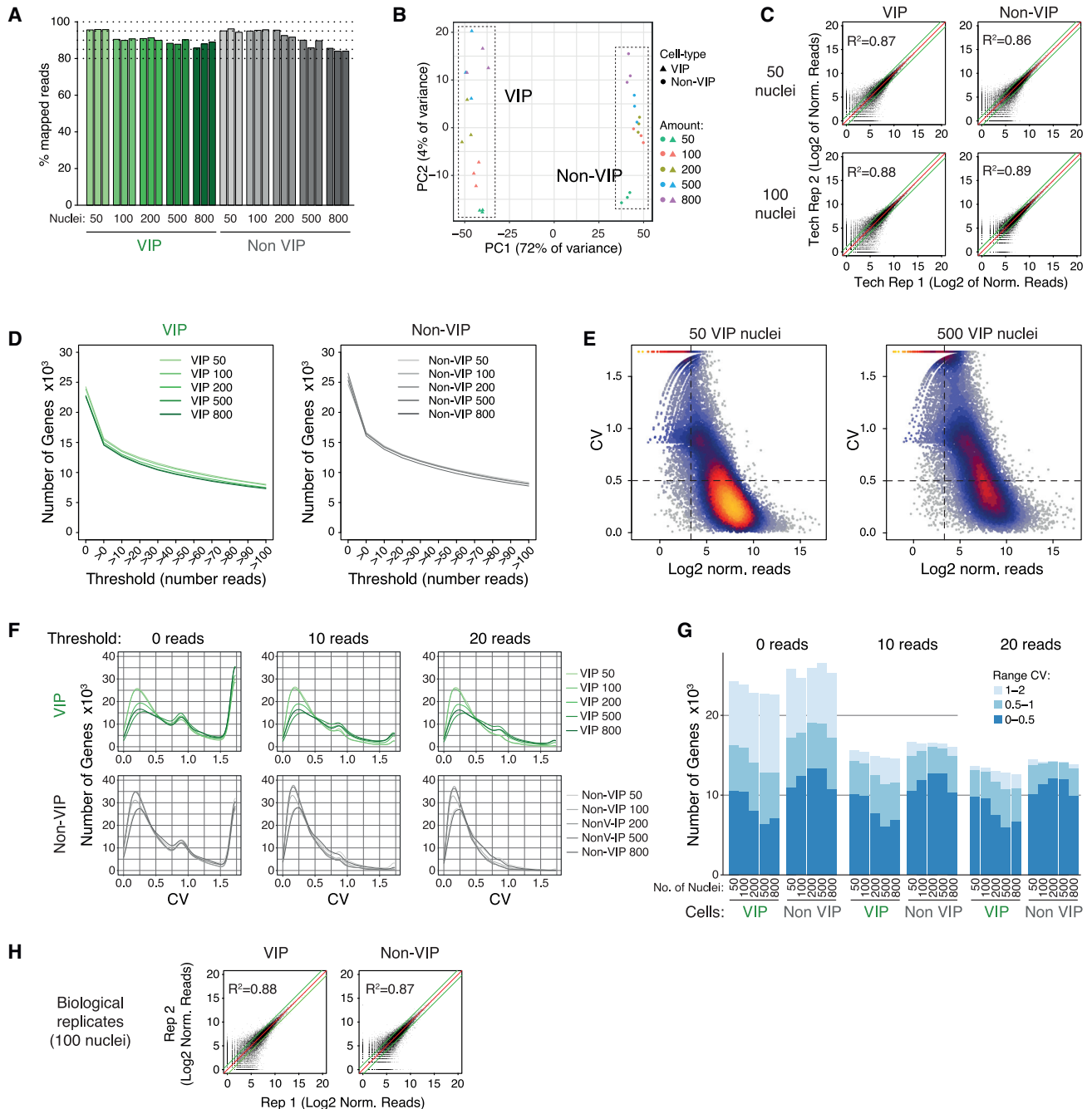
Figure 1. High-quality RNA-seq data can be obtained from 50–100 FACS-sorted neuronal nuclei

(A–C) Technical variability across RNA-seq libraries generated from increasing amounts of FACS-sorted nuclei of VIP INs and non-VIP cells in the cortex. (A) Increasing amounts of FACS-sorted nuclei of VIP INs and non-VIP cells (i.e., nuclei of all cells that are not VIP INs) were collected and used for building RNA-seq libraries in three technical replicates; plotted are percentage of uniquely mapped reads in each library. (B) Principal-component analysis (PCA) reveals that libraries prepared from the same FACS fraction and the same number of nuclei are most similar to each other. (C) Pairwise comparison of gene expression levels across technical replicate libraries is shown (dot represents one gene).

(D) Number of genes identified when applying increasing expression thresholds in technical replicate RNA-seq libraries generated from nuclei of VIP INs and non-VIP cells.

(E) Density scatterplot of the coefficient of variation (CV) (SD/mean) of normalized read counts as function of gene expression level (Log2 normalized read counts) in three technical replicate libraries prepared from 50 or 500 FACS-sorted VIP IN nuclei.

(legend continued on next page)

(Ecker et al., 2017; Habib et al., 2020; Jin et al., 2020; Keren-Shaul et al., 2017) and have identified a wealth of genetically defined neural cell types (Usoskin et al., 2015; Zeisel et al., 2015). However, scRNA-seq and snRNA-seq still suffer from important technical limitations, such as the relatively low number of genes reliably identified, e.g., when using droplet-based approaches (2,000–4,000 genes per cell; Ding et al., 2020). Similarly, even though the bioinformatic tools for analyzing scRNA-seq and snRNA-seq data are becoming increasingly streamlined (Andrews et al., 2021; Soneson and Robinson, 2018; Tian et al., 2019; Vieth et al., 2019), proper analyses of such data still require a very high level of expertise. Finally, scRNA-seq and snRNA-seq experiments are relatively expensive, since typically, a large number of cells and nuclei need to be sequenced to obtain robust and reliable data. Thus, scRNA-seq and snRNA-seq are not ideal for experiments that require transcriptomic analyses of many samples.

Bulk-based approaches for RNA-seq are technically robust, provide high transcriptomic coverage, and reliably identify up to 15,000 genes in a given cell type (Mardinly et al., 2016); furthermore, analyses of bulk RNA-seq data are well established and are rather accessible even to non-experts (Love et al., 2014; Robinson et al., 2010). However, a major drawback of bulk approaches is that they require relatively large amounts of input material, usually at least several thousand cells or nuclei (Amamoto et al., 2020; Mardinly et al., 2016). These large amounts of input material are a severe limitation when using bulk RNA-seq for transcriptomic analyses in sparse neuronal populations, since it is often not feasible to collect sufficient amounts of input material in multiple experimental conditions. This limitation applies also to RNA-seq of hand-sorted labeled neurons (Sugino et al., 2019), which is very labor intensive and can be applied only to live tissue—thus, it is not well suited for routine parallel processing of many biological samples. Taken together, there is a need for a simple method that preserves the advantages of bulk RNA-seq (deep coverage, robustness, and simple analyses) but that works for meso-scale amounts (i.e., several dozen) of genetically defined cells or nuclei.

RESULTS

Our goal was to establish a simple and robust method for in-depth RNA-seq from meso-scale amounts of genetically identified neurons (i.e., 50–100 cells per sample) that is applicable to frozen tissues and that allows for the precise isolation and transcriptomic profiling of genetically defined neurons via their labeling with antibodies, genetic tags, and/or viral constructs. Thus, we decided to focus on the fluorescence-activated cell sorting (FACS)-based isolation of nuclei rather than whole cells, since nuclei—unlike whole cells—remain intact even after snap freezing, thereby preserving the nuclei's transcriptional state, e.g., during precisely timed experiments.

Robust in-depth RNA-seq from meso-scale amounts of FACS-sorted nuclei

To establish our meso-scale method for RNA-seq in genetically identified neurons, we focused on GABAergic interneurons (INs) in the mouse visual cortex that express the vasoactive intestinal peptide (VIP INs): these neurons are relatively sparse (1% to 2% of all cortical neurons, mostly concentrated in layer 2/3 [L2/3] of the cortex; Fishell and Kepecs, 2020; Tremblay et al., 2016) and accessible with cell-type-specific genetic tools (He et al., 2016; Taniguchi et al., 2011), and their transcriptome has been profiled extensively (Hrvatin et al., 2018; Mardinly et al., 2016; Yao et al., 2021). Thus, we generated transgenic mice in which all cortical VIP INs are labeled by a nuclear-bound version of GFP (Vip-Cre::INTACT, "VIP INTACT" mice; Mo et al., 2015; Figure S1A) and validated that the GFP-labeled nuclei are found mainly in L2/3 (Figure S1B). Finally, we prepared nuclei from the visual cortices of these mice (using a modified protocol based on Bakken et al., 2018) and validated by FACS that an appropriate fraction of nuclei is GFP labeled (1.2% in VIP INTACT mice and 0% in wild-type control mice; Figure S1C).

Next, we sought to determine whether meso-scale amounts (i.e., tens to hundreds) of FACS-sorted nuclei suffice for in-depth transcriptomics. For this, we prepared nuclei from the visual cortex of VIP INTACT mice, FACS sorted them into GFP-positive and negative fractions ("VIP INs" and "non-VIP cells," respectively), and collected samples of 50, 100, 200, 500, or 800 nuclei from each fraction. To simplify the subsequent generation of RNA-seq libraries, we collected the sorted nuclei into a lysis buffer that is suitable for reverse transcription, thereby avoiding a separate step of RNA purification. RNA-seq libraries were then generated with a modified version of the SMART-Seq2 protocol and sequenced at a depth of 5–10 million reads per library. After an initial quality check, we mapped reads to the mm10 mouse genome: this revealed that the majority of reads were uniquely aligned (>80% uniquely mapped reads) and that libraries prepared from 200 or less nuclei exhibit the highest percentage of mapped reads (>90% uniquely mapped reads) (Figure 1A).

We then sought to characterize the variability between technical replicate libraries prepared from the same number of nuclei, as this would help us to determine the number of nuclei that yields libraries with the largest number of reliably identified genes. Principal-component analysis (PCA) revealed that the libraries prepared from the same number of nuclei of the same FACS fraction tended to cluster together (Figure 1B). Furthermore, pairwise comparisons of these technical replicates revealed that libraries prepared from 50 and 100 VIP INs have the highest correlation coefficient (CC) (0.87 and 0.88, respectively; Figure 1C; Table S1). In contrast, libraries prepared from 500–800 VIP INs were the least similar to each other (CC = 0.80 and 0.81, respectively). In non-VIP cells, libraries prepared from 200 or 500 nuclei were most similar to each other (CC = 0.91), followed by libraries prepared from 100, 800, and 50 nuclei

(F) Density plots for the number of genes as a function of the CV across three technical replicate libraries generated from 50–800 FACS-sorted nuclei of VIP INs and non-VIP cells. Plots are presented for three read-count thresholds (≥ 0, >10, and >20 reads).

(G) Number of genes with low, middle, and high variability (CV = 0–0.5, CV = 0.5–1, and CV = 1–2, respectively) when applying three different expression thresholds (≥ 0, >10, and >20 reads).

(H) Pairwise comparison of gene expression levels across biological replicate libraries generated from 100 FACS-sorted nuclei.

(CC = 0.89, 0.88, and 0.86, respectively). We therefore conclude that the technical variability of RNA-seq libraries prepared from 50–100 FACS-sorted GFP-labeled neuronal nuclei with our meso-scale protocol is relatively low, especially when compared with libraries derived from similar numbers of human cells (Wang et al., 2019).

Next, we assessed the number of genes whose expression can reliably be detected in the libraries prepared from different amounts of nuclei. For this, we determined the number of genes that cross increasing expression thresholds (i.e., number of normalized reads; Figure 1D): this revealed that ∼22,000 genes are identified when no threshold is applied and 15,000–17,000 genes with a threshold of 10 reads. Since one might expect that using fewer input nuclei would yield more variable libraries, we next analyzed the variance in gene expression across the different libraries (Figures 1E–1G): for this, we first calculated the coefficient of variation (CV) (standard deviation/mean) of the expression of all genes in all conditions and then plotted for each condition the CV of each gene as a function of its expression. This revealed that libraries prepared from fewer nuclei are less variable than libraries prepared from larger amounts of nuclei (Figure 1E) and that the variable genes are mostly expressed at low levels. Indeed, libraries prepared from 50 or 100 VIP IN nuclei contain more low variability genes than the libraries prepared from 200–800 VIP IN nuclei, while the optimal amount of non-VIP nuclei is between 200 and 500 (Figure 1F). Notably, applying an expression threshold of 10 or 20 reads greatly reduces the number of intermediate and high-variance genes, but not the number of low-variance genes (∼10,000 genes). Based on these analyses, we then binned for each library the genes into low, medium, and highly variable expression (CV = 0–0.5, CV = 0.5–1, and CV = 1–2, respectively) (Figure 1G). This revealed that, by applying a threshold of 10 reads, we consistently identify more than 10,000 low-variability genes in libraries prepared from 50–100 VIP IN nuclei and an even higher number of such genes in libraries prepared from 100–500 nuclei of non-VIP cells; importantly, the number of low-confidence genes (with CV = 0.5–1 and CV = 1–2) is relatively low under these settings. Thus, we conclude that, for sparse neuronal populations, the ideal number of nuclei collected for library preparation lies between 50 and 100 nuclei and that setting an expression threshold of 10 reads greatly reduces the number of high-variability genes without substantially affecting the number of low-variability genes. Since 100 FACS-sorted nuclei yield high-quality libraries also for non-VIP cells, we decided in subsequent experiments to collect 100 nuclei for each condition and to apply a 10-read expression threshold in our analyses.

Having determined the technical variability between the RNA-seq libraries (Figures 1A–1C), we then assessed how this compares with biological variability of such libraries (i.e., variability between different biological samples). For this, we FACS isolated 100 VIP and 100 non-VIP nuclei from the visual cortices of two additional VIP INTACT mice, performed RNA-seq, and calculated the correlation coefficients obtained by pairwise comparisons of the respective libraries (Figure 1H; Table S2). This revealed that the variability between biological replicates is indistinguishable from the variability between technical replicates. We therefore conclude that our meso-scale protocol allows for

consistent in-depth transcriptomic comparisons of sparse, genetically defined cell types across different animals.

Taken together, these experiments demonstrate that our meso-scale protocol generates high-quality transcriptomic data: using as little as 50 FACS-sorted nuclei, our protocol is robust with regards to technical variability and allows for reliable analyses of the expression levels of 10,000–12,000 genes, depending on the heterogeneity of the cellular population profiled. Notably, these numbers of identified genes are similar to the numbers obtained with standard protocols for bulk RNA-seq. Since our method uses meso-scale amounts of nuclei we call it "Meso-seq."

Meso-seq libraries are cell type specific

An intended key application of Meso-seq is cell-type-specific, in-depth gene profiling in sparse populations of genetically defined neurons. Thus, we next analyzed whether the Meso-seq data obtained from VIP INs are cell type specific when compared with non-VIP cells (Figure 2). Focusing on the biological replicate libraries (Figure 1H), we first analyzed the expression of well-described housekeeping genes, pan-neuronal genes, and known cell-type-specific genes (Figure 2A). Indeed, all these genes show the expected expression patterns, with de-enrichment or enrichment of the cell-type-specific genes in the anticipated fraction of nuclei (e.g., *Vip* and *Gad1* are enriched in VIP IN nuclei, while the glutamatergic markers *Neurod6* and *Slc17a* are de-enriched in VIP IN nuclei). Consistent with this and similar to previous reports (Hrvatin et al., 2018; Mardinly et al., 2016; Mo et al., 2015; Yao et al., 2021), transcriptome-wide analyses revealed several hundred genes that are de-enriched or enriched in VIP INs (406 and 223 genes, respectively; Figure 2B). Finally, we compared the VIP-IN-enriched genes identified by our Meso-seq approach with the VIP-IN-enriched genes identified in previous experiments that used a large number of VIP IN nuclei purified from the cortices of several VIP INTACT mice for standard bulk RNA-seq (Mo et al., 2015; Figure 2C): indeed, most of the genes identified by Meso-seq as VIP IN specific (i.e., >5-fold enriched) are VIP IN specific also in these data (171 of 233 of the genes [71%]) or at least enriched in VIP INs (i.e., less than 5-fold; 53 of 233 of the genes [23%]). We therefore conclude that Meso-seq allows for cell-type-specific gene profiling from sparse populations of genetically defined neurons.

Meso-seq is compatible with antibody labeling of nuclei

Thus far, we have demonstrated that Meso-seq is suitable for profiling of ultra-low amounts of neuronal nuclei labeled with exogenous fluorescent tags, such as GFP. However, for many applications—e.g., transcriptomic profiling of sparse cell populations in the brains of humans or of non-human primates (NHPs)—it would be useful to isolate and profile nuclei based on the endogenous expression of cell-type-specific proteins. Thus, we next tested whether Meso-seq can be used for transcriptomic profiling of ultra-low amounts of antibody-labeled neuronal nuclei. For this, we obtained frozen cortical tissue of an adult cynomolgus monkey (*Macaca fascicularis*), prepared a single-nuclei suspension, and labeled the nuclei with antibodies against the nuclear proteins RBFOX3 (i.e., anti-NeuN) and PROX1: since RBFOX3 is a pan-neuronal marker and since expression of PROX1 is highly enriched
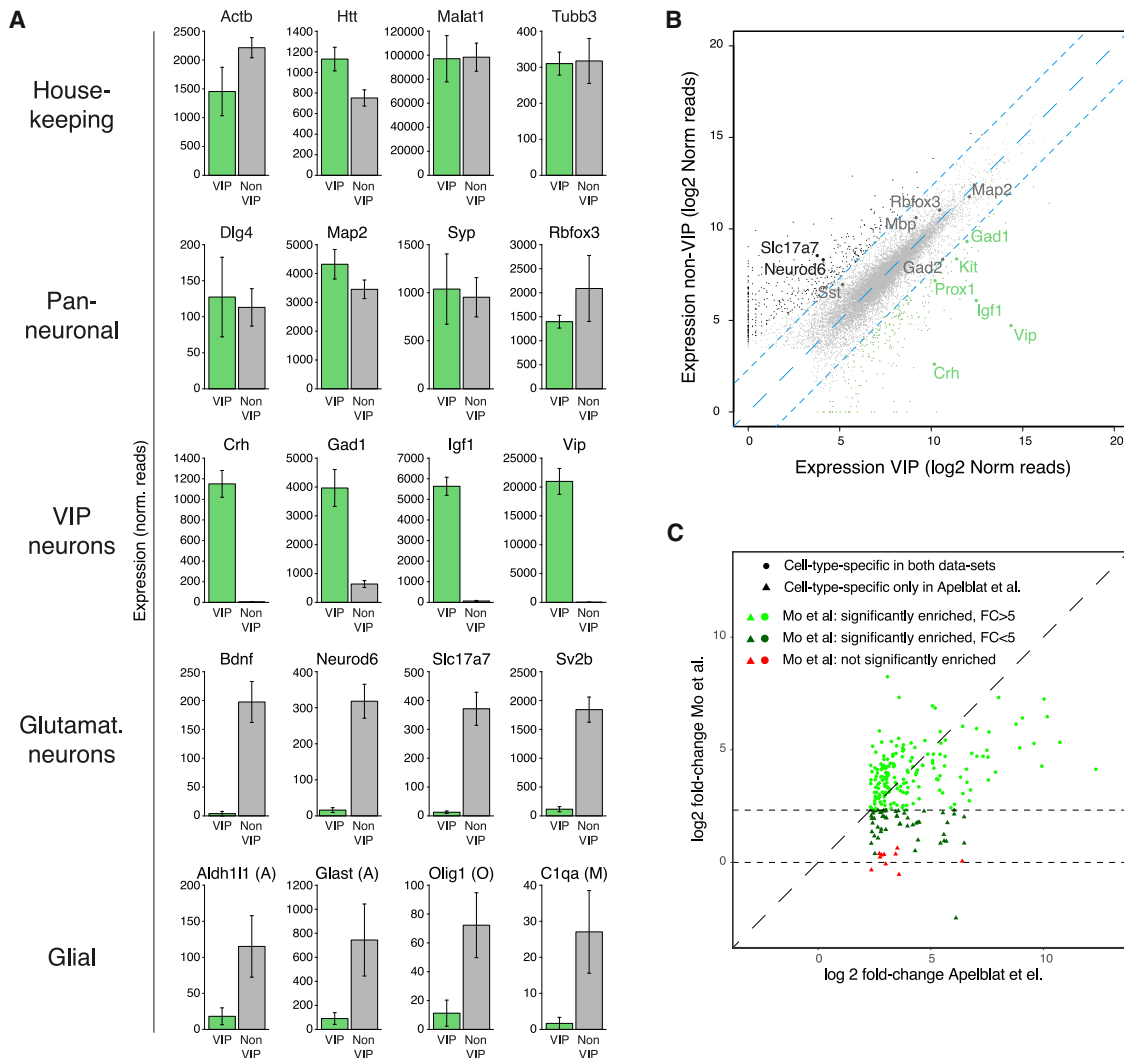
Figure 2. Meso-seq is suitable for in-depth analyses of cell-type-specific gene programs

(A) Expression of cell-type-specific marker genes in nuclei of VIP INs (green) and non-VIP cells (gray). Error bars represent the standard error of the mean (SEM).

(B) Scatterplot of the expression levels of all expressed genes (black dots, genes enriched in non-VIP cells; green dots, genes enriched in VIP IN nuclei; small dotted blue line, 5-fold enrichment).

(C) Comparison of VIP-IN-enriched genes identified by Meso-seq to VIP-IN-enriched genes identified by Mo et al. (2015). Meso-seq identifies 233 genes as significantly enriched in cortical VIP INs; of these, 73% (171 genes) were also found to be significantly enriched in VIP INs by Mo et al. (2015). FC, fold change.

in VIP INs (Miyoshi et al., 2015; Rubin and Kessaris, 2013), this labeling strategy should allow for collection and transcriptomic comparison of VIP IN nuclei (PROX1$^+$, NeuN$^+$, and DAPI$^+$) versus all neuronal nuclei (PROX1$^-$, NeuN$^+$, and DAPI$^+$) and all cortical cells (DAPI$^+$). Similar as with genetically labeled nuclei (Figure 2), we find that Meso-seq reliably identifies over 12,000 genes in each fraction of 100 FACS-sorted nuclei (Figure 3A) and that the replicate libraries in each fraction are very similar to each other (Figure 3B). Next, we assessed the cell type specificity of these libraries by comparing the libraries derived from NHP VIP INs with those derived from all cortical NHP neurons (Figure 3C). Similar as in VIP INs isolated from VIP INTACT mice (Figure 2), we identify a large number of genes that are de-enriched or enriched in VIP INs, including expected VIP-IN-specific genes (e.g., VIP,

PROX1, IGF1, CRH, etc.). Finally, we asked whether the Meso-seq approach can be used for identifying genes that are expressed in a cell-type-specific manner only in certain species. For this, we prepared nuclei from the visual cortices of an adult wild-type mouse and repeated the experiment as was done for the NHP nuclei (i.e., antibody labeling, FACS, and Meso-seq). Comparing the nuclei isolated from mice and NHPs, we find that 164 genes are enriched 5-fold or more in NHP PROX1$^+$ neurons while 109 genes are enriched in mouse PROX1$^+$ neurons (Figure 3D); 10 genes are 5-fold or more enriched in cortical PROX1$^+$ neurons in both species (as compared with PROX1$^-$ nuclei). Since the number of genes that are enriched only in one of the species in PROX1$^+$ neurons is relatively high and includes some unexpected genes (e.g., Vip, Prox1, Igf1, Crh, and Kit), we further validated these
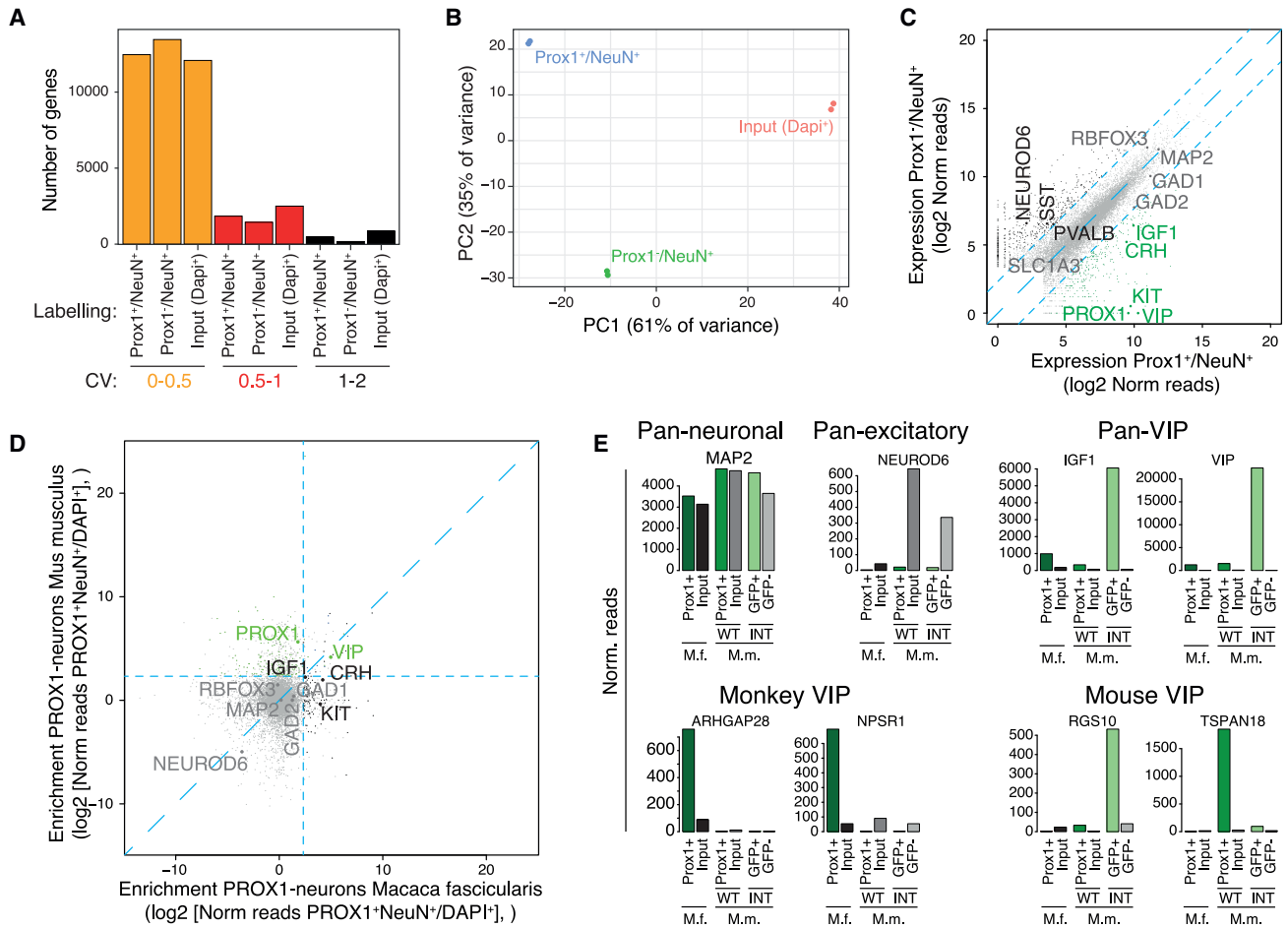
Figure 3. Meso-seq of antibody-labeled nuclei of sparse neuronal subtypes isolated from mice and non-human primate (NHP) cortices yields high-quality RNA-seq libraries
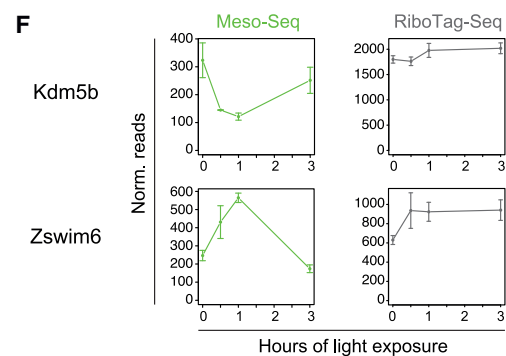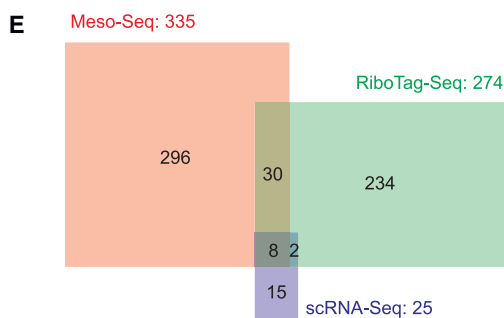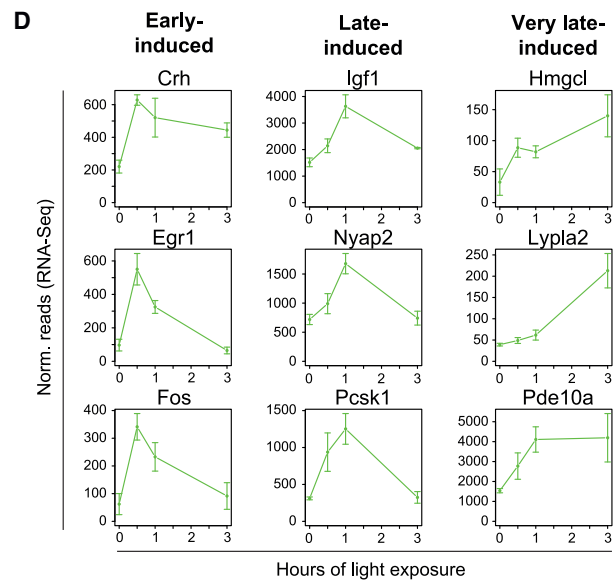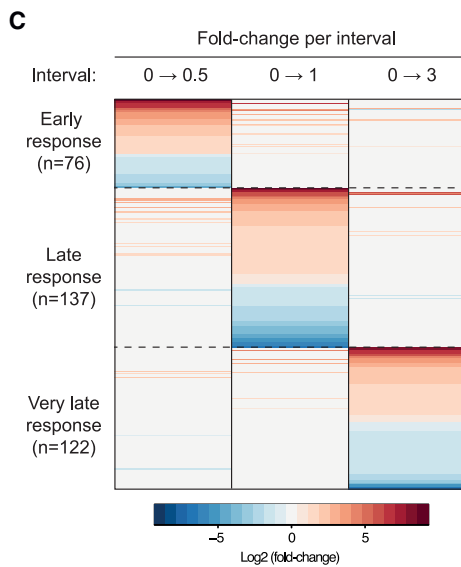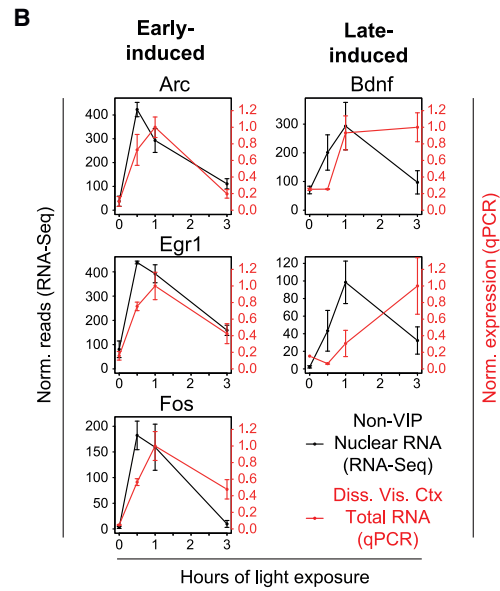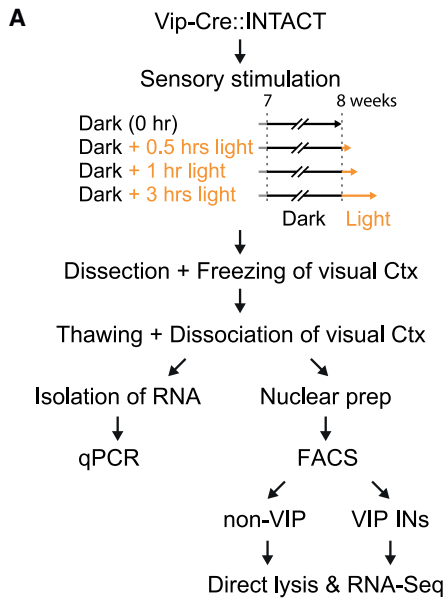
Meso-seq in nuclei FACS sorted from the cortices of an adult mouse (*Mus Musculus*) or an adult cynomolgus monkey (*Macaca fascicularis*); nuclei were FACS sorted based on antibody labeling for the pan-neuronal marker NeuN (i.e., RBFOX3) and the lineage marker PROX1.

(A–C) Meso-seq libraries prepared from PROX1+/NeuN+ nuclei are cell type specific and highly enriched for VIP INs. (A) Quantification of the number of genes with low, middle, and high variability (CV = 0–0.5, CV = 0.5–1, and CV = 1–2, respectively) in the different FACS fractions is shown. Using a threshold of >10 reads, more than 12,000 genes can be identified reliably even in PROX1+/NeuN+ neurons in the cynomolgus monkey cortex. (B) Principal-component analysis reveals that Meso-seq libraries prepared from the same nuclear fraction co-cluster. (C) Scatterplot of all genes expressed (>10 reads) in RNA-seq libraries prepared from PROX1+/NeuN+ nuclei (i.e., enriched for VIP INs) or PROX1−/NeuN+ nuclei (i.e., all other neurons) isolated from the cortex of adult cynomolgus monkeys is shown (dark gray dots, genes 5-fold enriched in PROX1−/NeuN+ nuclei; green dots, genes 5-fold enriched in PROX1+/NeuN+ nuclei; small dotted blue line, 5-fold enrichment).

(D and E) Meso-seq of PROX1−/NeuN+ nuclei isolated from the mouse and cynomolgus monkey cortex reveals species-specific gene expression in VIP INs. (D) Scatterplot of enrichment of expression in PROX1+/NeuN+ nuclei (as compared with expression in DAPI+ nuclei) in the cortex of cynomolgus monkey and mouse is shown (expression threshold >10 reads; black dots, mouse genes significantly enriched 5-fold; green dots, monkey genes significantly enriched 5-fold; small dotted blue line, 5-fold enrichment). (E) Examples of genes that are expressed selectively in VIP INs either in monkeys (*ARHGAP28* and *NPSR1*) or in mice (*Rgs10* and *Tspan18*) are shown (Prox1+, PROX1+/NeuN+ nuclei; input, DAPI+ nuclei; GFP+, VIP IN nuclei isolated from VIP INTACT mice). INT, INTACT; WT, wild type.

findings by asking whether these genes are de-enriched or enriched in a similar manner also in VIP INs in the cortices of VIP INTACT mice (Figure 2). This revealed that only for a few genes the enrichment in PROX1+ neurons in NHP and mouse cortices is consistent with the enrichment in VIP INs in the cortices of VIP INTACT mice (Figure 3E): for example, *ARHGAP28* and *NPSR1* are highly enriched in PROX1+ neurons in the NHP cortex but are essentially absent from these cells in the mouse cortex, while *Rgs10* and *Tspan18* are highly enriched in mouse VIP INs but are

essentially absent from PROX1+ neurons in the NHP cortex. Additional experiments will be required to fully establish the identification of genes that are enriched in a species-specific manner in cortical VIP INs. Nevertheless, our experiments demonstrate that Meso-seq can be used in combination with antibody labeling for deep profiling of sparse neuronal populations of frozen brain tissue of mice and NHPs. Furthermore, using Meso-seq, we provide the first transcriptomic analyses of VIP INs in the cortex of cynomolgus monkeys, thereby identifying genes that seem to be

A Vip-Cre::INTACT → Sensory stimulation → Dark (0 hr), Dark + 0.5 hrs light, Dark + 1 hr light, Dark + 3 hrs light (7 to 8 weeks, Dark/Light) → Dissection + Freezing of visual Ctx → Thawing + Dissociation of visual Ctx → Isolation of RNA → qPCR; Nuclear prep → FACS → non-VIP / VIP INs → Direct lysis & RNA-Seq

B Early-induced and Late-induced gene expression plots (Arc, Bdnf, Egr1, Fos) showing Norm. reads (RNA-Seq) and Norm. expression (qPCR) versus Hours of light exposure. Non-VIP Nuclear RNA (RNA-Seq); Diss. Vis. Ctx Total RNA (qPCR).

C Fold-change per interval heatmap. Intervals: 0 → 0.5, 0 → 1, 0 → 3. Early response (n=76), Late response (n=137), Very late response (n=122). Log2 (fold-change) scale from −5 to 5.

D Early-induced (Crh, Egr1, Fos), Late-induced (Igf1, Nyap2, Pcsk1), Very late-induced (Hmgcl, Lypla2, Pde10a) Norm. reads (RNA-Seq) versus Hours of light exposure.

E Venn diagram: Meso-Seq: 335, RiboTag-Seq: 274, scRNA-Seq: 25. 296, 30, 234, 8, 2, 15.

F Meso-Seq and RiboTag-Seq Norm. reads for Kdm5b and Zswim6 versus Hours of light exposure.

(legend on next page)

expressed in a species-specific manner in cortical VIP INs in mice and NHPs.

Meso-seq allows for in-depth analyses of experience-regulated gene programs in sparse neuronal populations

Next, we tested whether Meso-seq is suitable for analyzing experience-regulated gene programs in sparse neuronal populations. Experience-induced transcription is a key molecular mechanism for regulating the plasticity of neural circuits in a cell-type-specific manner (Gray and Spiegel, 2019; Hrvatin et al., 2018; Mardinly et al., 2016; Nord and West, 2020; Yap and Greenberg, 2018), and mutations in or nearby experience-regulated genes are associated with a variety of neurodevelopmental and psychiatric disorders (Ebert and Greenberg, 2013). We hypothesized that Meso-seq is well suited for analyzing experience-induced transcription in sparse neuronal populations, since such analyses often require that the respective tissues are harvested at very precise time points post-stimulus and Meso-seq works well on ultra-low numbers of nuclei FACS isolated from snap-frozen tissue (Figures 1, 2, and 3). To test this hypothesis, we analyzed the changes in gene expression in VIP INs in the adult visual cortex upon dark housing/light exposure (DH/LE): previous studies used the same stimulation paradigm to identify sensory-induced transcriptomic changes in visual cortex VIP INs (either with RiboTag-seq [Mardinly et al., 2016] or with scRNA-seq [Hrvatin et al., 2018]), thereby allowing us to directly compare between Meso-seq data and these existing data. Thus, we housed adult VIP INTACT mice in complete darkness for 1 week and dissected and snap froze their visual cortices either in the dark or after exposure to light for 0.5, 1, or 3 h (Figure 4A). First, we compared the kinetics of experience-induced gene expression in RNAs isolated from FACS-sorted nuclei of non-VIP visual cortex cells with those isolated from total visual cortices (i.e., nuclear versus total RNA) (Figure 4B). This revealed that DH/LE induces in both RNA preparations a rapid rise in the RNA levels of early-induced genes (e.g., *Arc*, *Fos*, and *Npas4*) and a slower increase in late-induced effector genes (e.g., *Bdnf* and *Nptx2*). However, the kinetics of the changes in RNA levels differ in the two preparations: in purified nuclei, the RNA levels of early-induced genes peak at 30 min after stimulus onset, while in whole-tissue lysates, they reach their peak values after 1 h. Similarly, the RNA levels of late-induced genes peak in purified nuclei at 1 h post-stimulus but 3 h in total RNA.

Next, we analyzed the sensory-induced gene program in VIP INs (Figures 4C and 4D). As before, we observed rapid kinetics

of sensory-induced gene expression in these cells as well: specifically, we identified 76 genes with peak RNA levels at 30 min post-stimulus ("early-response genes," including many well-known early-induced transcription factors, e.g., *Egr1* and *Fos*), 137 late-response genes with peak RNA levels at 1 h post-stimulus (including many well-described late-induced effector genes, e.g., *Igf1* and *Pcsk1*), and 122 very-late-response genes whose RNA levels peak at 3 h after exposure to light. Interestingly, when we compared the genes identified as experience regulated in VIP INs by Meso-seq with those previously identified via RiboTag-seq or scRNA-seq, we observed only a small overlap (Figure 4E): the majority of the genes identified as experience regulated by Meso-seq are not identified as regulated either in the RiboTag-seq or scRNA-seq experiments (296 out of 335 genes), and most experience-regulated genes in the RiboTag-seq data are not found to be regulated by Meso-seq data or scRNA-seq (234 out of 274 genes). This relatively low overlap might be due to (1) differences in the pools of RNAs assayed in the different experiments (Meso-seq: nuclear RNA; RiboTag-seq: ribosome-attached RNA; scRNA-seq: whole-cell RNA) or (2) the higher sample-to-sample variability in the Meso-seq samples than in the RiboTag-seq samples (average CC between libraries is 0.83 in Meso-seq and 0.95 in RiboTag-seq). Indeed, many genes that are called "regulated" in the RiboTag-seq data seem also regulated in the Meso-seq data, but their expression levels are too variable to cross the significance threshold (not shown). Our analyses cannot reveal the source of the higher variability across Meso-seq samples, but they might be due to different sampling strategies in these experiments (Meso-seq: one mouse per library; RiboTag-seq: three mice per library).

Finally, we asked whether the experience-regulated genes identified exclusively in our Meso-seq data include genes associated with autism-spectrum disorders (ASDs)—if so, this might provide novel insights into the etiology of ASDs. Indeed, we found that 10 of the genes identified by Meso-seq as experience regulated are associated with ASDs (according to the SFARI database; https://gene.sfari.org), including the high-confidence ASD-associated genes *Zswim6* and *Kdm5b* (Figure 4F). Thus, our experiments demonstrate that Meso-seq is highly useful for analyzing experience-regulated gene programs in sparse populations of neurons and that this might reveal important aspects of the etiology of neurodevelopmental and neuropsychiatric disorders.

Meso-seq for transcriptomics in ultra-sparse neuronal subpopulations

Next, we tested whether Meso-seq can be used for profiling of ultra-sparse neuronal populations. For this, we focused on

Figure 4. Meso-seq allows for in-depth analyses of experience-regulated gene programs in sparse neuronal populations
(A) Experimental strategy for dark housing/light exposure (DH/LE) in VIP INTACT mice and transcriptomic analyses in nuclei of VIP INs and all other nuclei.
(B) Comparison of the kinetics of sensory-induced gene expression in total visual cortex extracts (prior to FACS, measured via qPCR) and FACS-purified nuclei (after FACS, measured by Meso-seq). Error bars represent the SEM.
(C and D) Sensory-driven gene programs in visual cortex VIP INs nuclei after DH/LE. (C) Heat plots of all sensory-regulated genes in visual cortex VIP INs are shown (one row represents one gene). (D) Line plots of expression of select sensory-induced genes in visual cortex VIP INs are shown. Error bars represent the SEM.
(E) Venn diagram comparing the number of sensory-regulated genes identified by different RNA-seq approaches in visual cortex VIP INs of adult mice upon sensory stimulation via DH/LE.
(F) *Kdm5b* and *Zswim6* are ASD-associated genes newly identified as experience regulated in VIP INs. Error bars represent the SEM.
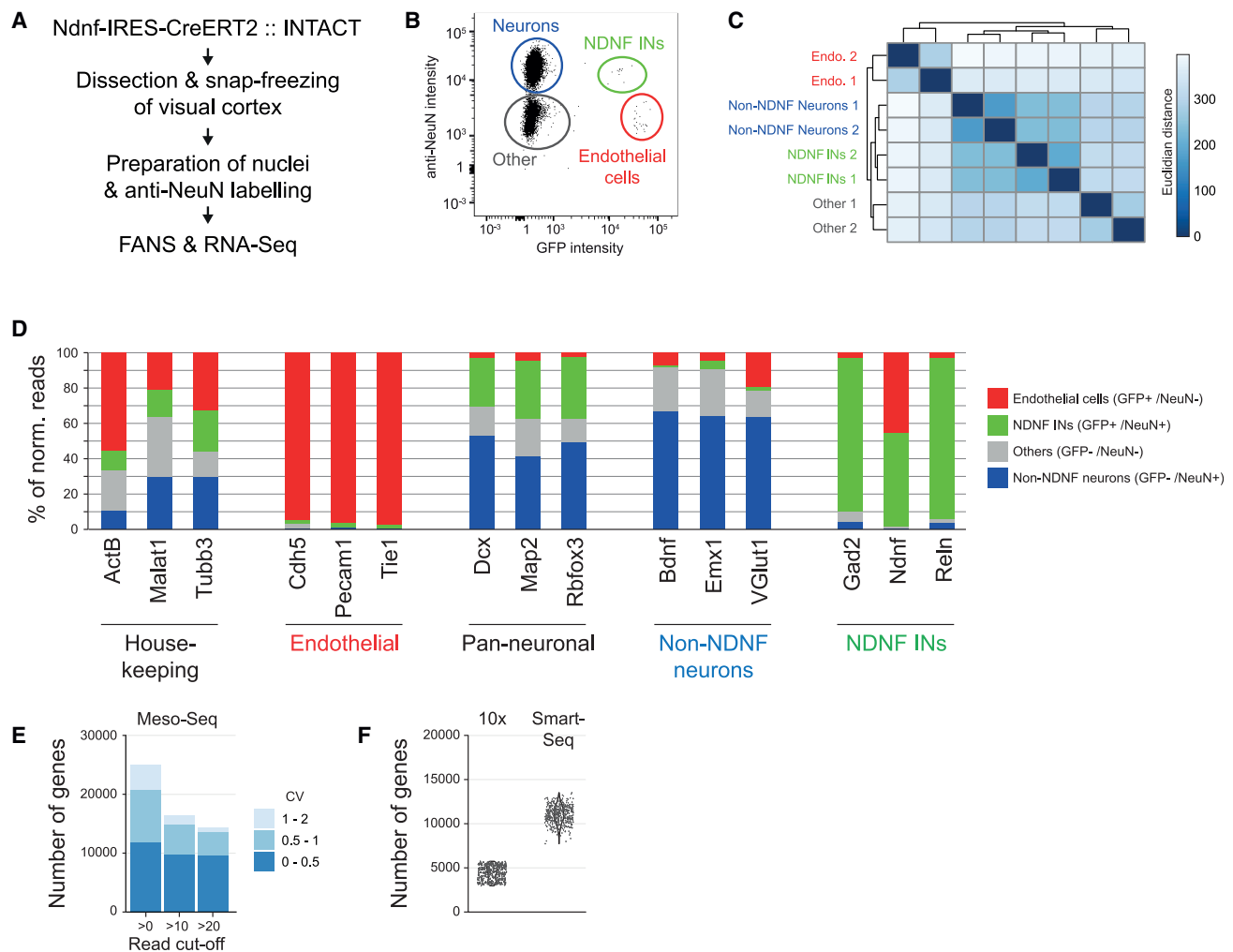
Figure 5. Meso-seq is suitable for in-depth transcriptomics in ultra-sparse neuronal subtypes

Meso-seq of *Ndnf*-expressing GABAergic interneurons (NDNF INs) in the visual cortex.

(A) Experimental strategy for profiling of visual cortex NDNF INs.

(B) FACS strategy for isolating nuclei of NDNF INs and other cell types from the visual cortex of Ndnf-IRES-CreERT2::INTACT mice (NDNF INs, GFP$^+$ and NeuN$^+$; endothelial cells, GFP$^+$ and NeuN$^-$; non-NDNF neurons, GFP$^-$ and NeuN$^+$; all other cells, GFP$^-$ and NeuN$^-$).

(C) Clustering analyses of Meso-seq data derived from all populations FACS purified in (B).

(D) Relative enrichment of cell-type-specific marker genes in the Meso-seq data derived from all populations FACS purified in (B).

(E) Number of genes reliably detected in NDNF INs by Meso-seq when applying different expression thresholds.

(F) Number of genes detected by scRNA-seq in Ndnf-expressing INs with a high-throughput, droplet-based technique (10× Genomics) or a low-throughput method (SMART-Seq4).

GABAergic interneurons in layer 1 of the cortex that express the neuron-derived neurotrophic factor (L1 NDNF INs) and that are less abundant than VIP INs by at least an order of magnitude. For this, we generated mice that are double heterozygous for Ndnf-IRES-CreERT2 (Abs et al., 2018) and INTACT and prepared visual cortex nuclei for FACS (Figure 5A). Since Ndnf-IRES-CreERT2 in combination with a Cre-reporter driven by the CAG promoter labels also endothelial cells (Abs et al., 2018), we immuno-labeled the nuclei prior to FACS sorting with an antibody against the pan-neuronal nuclear marker NeuN, as this allowed us to Meso-seq nuclei of NDNF INs (GFP$^+$/NeuN$^+$; ~0.1% of all DAPI$^+$ nuclei), endothelial cells (GFP$^+$/NeuN$^-$), non-NDNF neu-

rons (GFP$^-$/NeuN$^+$), and all other cells (GFP$^-$/NeuN$^-$) (Figure 5B). Indeed, we find that Meso-seq libraries prepared from the same cell types co-cluster (Figure 5C) and that the expression of cell-type markers is enriched in the appropriate FACS fractions (e.g., *Ndnf* expression is highly enriched in GFP$^+$/NeuN$^+$ and in GFP$^+$/NeuN$^-$ nuclei, while expression of *Gad2* and *Reln* is highly enriched only in GFP$^+$/NeuN$^+$ nuclei) (Figure 5D). Furthermore, as in VIP INs, we identify also in NDNF INs roughly 10,000 low-variance genes (Figure 5E)—this number is similar to the total number of genes identified in NDNF INs via scRNA-seq with the SMART-seq protocol (~11,000 genes) and considerably higher than the number of genes identified by
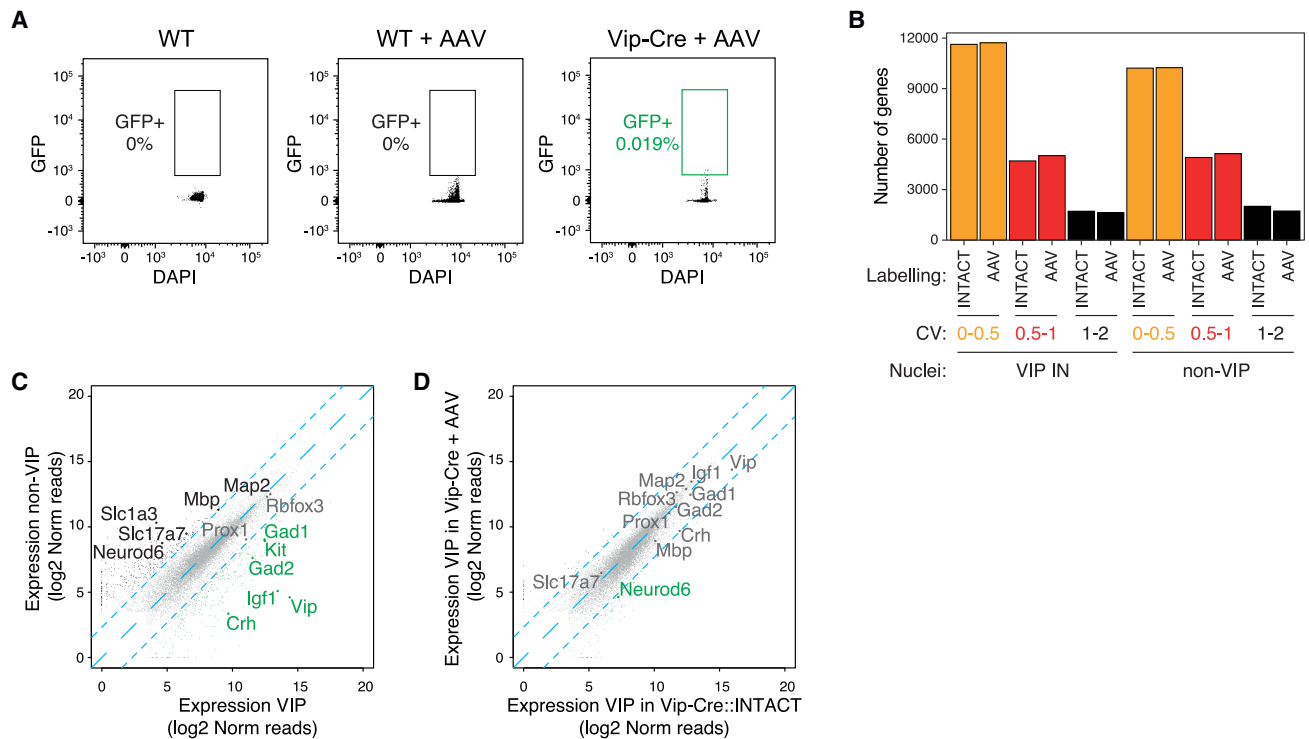
Figure 6. Meso-seq allows for in-depth transcriptomic profiling in virally transduced neurons

AAVs expressing an empty gRNA cassette and EYFP anchored to the nuclear membrane in a Cre-dependent manner (AAV-U6-gRNA-hSyn-DIO-EYFP-KASH) were injected into the visual cortex of Vip-Cre mice, and 100 FACS-purified EYFP-labeled nuclei were analyzed by Meso-seq.

(A) FACS strategy for isolating the nuclei of AAV-infected visual cortex VIP INs.

(B) Quantification of the number of genes with low, middle, and high variability (CV = 0–0.5, CV = 0.5–1, and CV = 1–2, respectively; expression threshold = >10 reads).

(C) Scatterplot of all genes expressed (i.e., genes with >10 reads and CV ≥ 0.5) in RNA-seq libraries prepared from FACS-sorted nuclei of VIP INs and non-VIP cells (black dots, genes significantly 5-fold enriched in non-VIP cells; green dots, genes significantly 5-fold enriched in VIP IN nuclei; small dotted blue line, 5-fold enrichment).

(D) Scatterplot comparing gene expression in AAV-labeled VIP INs and in INTACT-labeled VIP INs (black dots, genes significantly 5-fold enriched in AAV-labeled VIP INs; green dots, genes significantly 5-fold enriched in INTACT-labeled VIP INs; small dotted blue line, 5-fold enrichment).

droplet-based scRNA-seq (e.g., 10×) in these neurons (~4,500 genes) (Figure 5F). Notably, while SMART-seq requires several hundred single-cell libraries for robust gene-expression analyses, Meso-seq requires only three biological replicate libraries to reliably identify ~10,000 genes. Thus, Meso-seq is suitable for profiling of very sparse genetically defined neuronal cell types and can considerably reduce experimental costs.

Meso-seq for transcriptomics in neurons ultra-sparsely labeled with viral constructs

Finally, we tested whether Meso-seq is suitable for in-depth transcriptomic analyses in neurons infected with viral constructs: this would be highly useful for molecular-mechanistic studies, e.g., when testing *in vivo* how virally mediated genome editing via CRISPR-Cas9 or gene knockdown via short hairpin RNAs (shRNAs) affects gene expression in the infected cells. Thus, we injected into the visual cortices of adult Vip-Cre mice an adeno-associated viral (AAV) construct that drives the Cre-dependent expression of a nuclear-bound version of EYFP (KASH-EYFP) and the constitutive expression of a guide RNA (gRNA) cassette (Figure S2). Two weeks after injection, we dissected

the visual cortices, FACS isolated ~100 GFP$^+$ and GFP$^-$ nuclei from each animal, performed Meso-seq, and assessed parameters such as the number of reliably identified genes and the cell type specificity of the RNA-seq libraries. Despite the sparsity of AAV-mediated labeling (only 0.005%–0.02% of all DAPI$^+$ nuclei were EYFP labeled; Figure 6A), we reliably identify a comparable number of genes in these neurons as in GFP-labeled visual cortex VIP INs in VIP INTACT mice (Figure 6B). Similarly, the Meso-seq libraries derived from AAV-infected VIP INs show a high degree of cell type specificity, as evidenced by the cell-type-specific expression of marker genes (Figure 6C). Furthermore, when we compared the transcriptomes of the VIP INs in VIP INTACT mice with those of the AAV-infected VIP INs (Figure 6D), we found that they were very similar to each other and that they differ primarily in lowly expressed genes. This transcriptomic similarity of VIP INs labeled by INTACT and by AAV infection is remarkable insofar as much fewer VIP INs are labeled by AAV infections than by the genetically encoded INTACT allele, and therefore, FACS isolation of 100 VIP INs took up to 80 times longer in AAV-injected mice than in VIP INTACT mice (up to 3 h versus ~2 min)—this is consistent with our observation that the

quality of RNA in our nuclear preps is not affected by the duration of FACS (Figure S3). Taken together, these experiments demonstrate that Meso-seq is suitable for transcriptomic profiling in populations of genetically defined neurons ultra-sparsely labeled by viral infection.

DISCUSSION

Profiling of genetically defined cell types by RNA-seq has provided unprecedented insight into the molecular and cellular heterogeneity of neural circuits and has generated a wealth of hypotheses about the molecular mechanisms that control the development, plasticity, and pathology of these circuits. Now, major challenges lie in testing these hypotheses experimentally, for example, by manipulating specific genes and genomic regions *in vivo* in sparse neuronal populations and by assessing the molecular consequences of these molecular manipulations via RNA-seq. However, current RNA-seq approaches are ill suited for such mechanistic studies: snRNA-seq and scRNA-seq are technically and analytically complicated and relatively expensive, while methods for bulk RNA-seq require large amounts of input material, which often cannot be obtained in such mechanistic experiments. Here, we developed and validated Meso-seq: we demonstrate that this method (1) yields RNA-seq libraries with high mappability from as little as 50 bulk-processed nuclei FACS isolated from snap-frozen tissue, (2) routinely and robustly identifies more than 10,000 low-variance genes per RNA-seq library, (3) works well for genetically and antibody-labeled nuclei, (4) can be used for cell-type-specific profiling in rodents and NHP brain, (5) is suitable for profiling of stimulus-induced gene programs, and (6) can be used for profiling of neurons ultra-sparsely labeled with viral constructs. In the following, we discuss several points that are noticeable about Meso-seq and the result we obtained with it.

Limitations of the study
Optimal number of nuclei per RNA-seq library
Unlike what we expected at the onset of our experiments, we find that, for sparse neuronal populations, the libraries generated from low numbers of FACS-sorted nuclei are of higher quality than the libraries generated from larger numbers of nuclei with regards to percentage of uniquely mapped reads and numbers of low-variability genes identified. This is probably due to our use of a direct lysis protocol that eliminates the need for purifying RNA in a separate step prior to library construction but that requires the FACS-sorted nuclei to be collected in a relatively low volume of lysis buffer (2.35 μL total, including collected nuclei). Thus, since the concentration of the detergent in this lysis buffer is relatively low, the volume of each nucleus-containing FACS droplet is ~3.3 nL in our FACS settings, and the library preparation protocol originally foresees a volume of 330 nL of collected nuclei (i.e., 100 nuclei), our Meso-seq protocol probably works best for volumes of collected nuclei approximately in this range. At higher volumes—i.e., when more nuclei are collected—the protocol is expected to perform worse, be it because the collected nuclei are not fully lysed or the initial steps of the library preparation protocol are inhibited by the additional debris generated by the larger number of

nuclei collected and/or the sub-optimal concentration of reagents in these conditions.

Intronic reads
Unlike RNA-seq data generated from RNA that was isolated from whole cells or tissues and that is mostly fully processed cytosolic RNA, the transcriptomic data generated by Meso-seq are derived from nuclear RNA that is not yet fully processed and contains a relatively large percentage of intronic reads (~70% of the reads). This needs to be considered when aligning the sequenced reads to the genome and quantifying RNA expression by our method.

Profiling of neurons ultra-sparsely labeled upon infection with viral vectors
A central goal for us was to establish a robust method for RNA-seq in populations of neurons that are ultra-sparsely labeled upon infection with viral vectors—this would be highly useful for molecular neuroscience applications, such as transcriptomic profiling after acute gene knockout in sparse neuronal populations. As demonstrated in Figures 5 and 6, Meso-seq is well suited for such applications, even for neurons that are labeled so sparsely by viral construct that FACS sorting can take up to several hours. This is a major advantage of Meso-seq and will be highly useful for a variety of molecular neuroscience applications.

PROX1-positive neurons in cortex of mice and cynomolgus monkeys
Our transcriptomic analyses of mouse and NHP cortical PROX1[+] neurons demonstrate that Meso-seq is suitable for in-depth transcriptomic profiling of antibody-labeled neuronal subtypes. Thus, since most types of neurons can be identified by the expression of only a few transcription factors (Hobert, 2016; Zeisel et al., 2018) and since several antibodies can be multiplexed in FACS, Meso-seq should be useful for deep profiling of sparse neuronal subtypes in brain samples derived from NHPs and humans. Notably, our identification of genes highly enriched in the cortices and/or VIP INs of mice and cynomolgus monkeys is suggestive, but not conclusive: our analyses in the cynomolgus monkey cortex are based on a single biological replicate, and additional replicates are required to strengthen our conclusions. Furthermore, PROX1 in the cortex is highly enriched in VIP INs but is expressed also in a subpopulation of oligodendrocytes (Yao et al., 2021)—thus, validation with orthogonal methods (e.g., single-molecule fluorescent *in situ* hybridization [smFISH]) is required to establish conclusively that the genes identified by us are indeed expressed in a species-specific manner.

Experience-regulated gene programs
Analyses of experience-regulated gene programs often require RNA-seq at precise time points before and after stimuli and experiences, and the ability to snap freeze the dissected tissues prior to transcriptomic analyses greatly improves the feasibility and quality of such experiments. Thus, the fact that Meso-seq works well on nuclei isolated from snap-frozen tissue is beneficial for such studies, particularly when analyzing sparsely labeled cells. Interestingly, Meso-seq identifies experience-induced changes in gene expression with faster kinetics than methods based on cytosolic RNA. Likely, this is due to the fact that the nuclear RNAs analyzed by Meso-seq are still not fully processed and potentially nascent, which is different from cytosolic RNAs that are fully processed and much more abundant (Lake et al., 2017;

Zaghlool et al., 2021). This suggests that Meso-seq provides a closer representation of the process of transcription than RiboTag-seq, in which the levels of the sequenced RNAs are influenced by transcription but also by many other molecular mechanisms (e.g., recruitment of mRNAs to the ribosomes). Notably, methods for directly measuring transcription (e.g., Run-OFF; Greenberg et al., 1986) reveal even faster kinetics than Meso-seq, indicating that Meso-seq does not assay transcription per se but rather abundance of nuclear RNAs. These differences in the pools of RNAs assayed by the different methods probably contribute to the relatively low overlap in the number of genes identified as "sensory induced" by these methods upon DH/LE (Figure 5E). An additional factor contributing to this low overlap probably lies in the higher sample-to-sample variability in the Meso-seq data than in the RiboTag-seq data, and this might be due to different sampling strategies and/or to the very different amounts of input material used in the two approaches. Notably, not all the differences in experience-regulated genes identified by different methods can be ascribed to differences in sample-to-sample variability: as evidenced by our identification of the ASD-associated genes *Kdm5b* and *Zswim6* as experience regulated in VIP INs, different methods indeed reveal different pools of experience-regulated genes, and this can be highly useful, e.g., to gain insights into the etiology of neurodevelopmental disorders.

In summary, we conclude that Meso-seq allows for robust in-depth transcriptomics from ultra-low amounts of genetically defined neurons in a large range of molecular neuroscience applications. We therefore expect that Meso-seq will be highly useful for molecular-mechanistic experiments in sparse neuronal populations and an important addition to the expanding toolbox of molecular neuroscience.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Animals and DH/LE stimulation paradigm
- METHOD DETAILS
  - AAV constructs, AAV production, and stereotactic injections
  - FACS: Preparation of nuclei, antibody staining, FACS-sorting and collection/lysis of sorted nuclei
  - Library preparation and RNA-Seq
  - Bioinformatic analyses
- QUANTIFICATION AND STATISTICAL ANALYSES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.crmeth.2022.100259.

REFERENCES

Abs, E., Poorthuis, R.B., Apelblat, D., Muhammad, K., Pardi, M.B., Enke, L., Kushinsky, D., Pu, D.-L., Eizinger, M.F., Conzelmann, K.-K., et al. (2018). Learning-related plasticity in dendrite-targeting layer 1 interneurons. Neuron *100*, 684–699.e6. https://doi.org/10.1016/j.neuron.2018.09.001.

Amamoto, R., Zuccaro, E., Curry, N.C., Khurana, S., Chen, H.-H., Cepko, C.L., and Arlotta, P. (2020). FIN-Seq: transcriptional profiling of specific cell types from frozen archived tissue of the human central nervous system. Nucleic Acids Res. *48*, e4. https://doi.org/10.1093/nar/gkz968.

Andrews, T.S., Kiselev, V.Y., McCarthy, D., and Hemberg, M. (2021). Tutorial: guidelines for the computational analysis of single-cell RNA sequencing data. Nat. Protoc. *16*, 1–9. https://doi.org/10.1038/s41596-020-00409-w.

Bacher, R., Chu, L.-F., Leng, N., Gasch, A.P., Thomson, J.A., Stewart, R.M., Newton, M., and Kendziorski, C. (2017). SCnorm: robust normalization of single-cell RNA-seq data. Nat. Methods *14*, 584–586. https://doi.org/10.1038/nmeth.4263.

Bakken, T.E., Hodge, R.D., Miller, J.A., Yao, Z., Nguyen, T.N., Aevermann, B., Barkan, E., Bertagnolli, D., Casper, T., Dee, N., et al. (2018). Single-nucleus and single-cell transcriptomes compared in matched cortical cell types. PLoS One *13*, e0209648. https://doi.org/10.1371/journal.pone.0209648.

Challis, R.C., Ravindra Kumar, S., Chan, K.Y., Challis, C., Beadle, K., Jang, M.J., Kim, H.M., Rajendran, P.S., Tompkins, J.D., Shivkumar, K., et al. (2019). Systemic AAV vectors for widespread and targeted gene delivery in rodents. Nat. Protoc. *14*, 379–414. https://doi.org/10.1038/s41596-018-0097-3.

Cohen-Kashi Malina, K., Tsivourakis, E., Kushinsky, D., Apelblat, D., Shtiglitz, S., Zohar, E., Sokoletsky, M., Tasaka, G.-I., Mizrahi, A., Lampl, I., and Spiegel, I. (2021). NDNF interneurons in layer 1 gain-modulate whole cortical columns according to an animal's behavioral state. Neuron *109*, 2150–2164.e5. https://doi.org/10.1016/j.neuron.2021.05.001.

Ding, J., Adiconis, X., Simmons, S.K., Kowalczyk, M.S., Hession, C.C., Marjanovic, N.D., Hughes, T.K., Wadsworth, M.H., Burks, T., Nguyen, L.T., et al. (2020). Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. Nat. Biotechnol. *38*, 737–746. https://doi.org/10.1038/s41587-020-0465-8.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

Ebert, D.H., and Greenberg, M.E. (2013). Activity-dependent neuronal signalling and autism spectrum disorder. Nature *493*, 327–337. https://doi.org/10.1038/nature11860.

Ecker, J.R., Geschwind, D.H., Kriegstein, A.R., Ngai, J., Osten, P., Polioudakis, D., Regev, A., Sestan, N., Wickersham, I.R., and Zeng, H. (2017). The BRAIN initiative cell census consortium: lessons learned toward generating a comprehensive brain cell atlas. Neuron *96*, 542–557. https://doi.org/10.1016/j.neuron.2017.10.007.

Fishell, G., and Kepecs, A. (2020). Interneuron types as attractors and controllers. Annu. Rev. Neurosci. *43*, 1–30. https://doi.org/10.1146/annurev-neuro-070918-050421.

Gray, J.M., and Spiegel, I. (2019). Cell-type-specific programs for activity-regulated gene expression. Curr. Opin. Neurobiol. *56*, 33–39. https://doi.org/10.1016/j.conb.2018.11.001.

Greenberg, M.E., Ziff, E.B., and Greene, L.A. (1986). Stimulation of neuronal acetylcholine receptors induces rapid gene transcription. Science *234*, 80–83. https://doi.org/10.1126/science.3749894.

Habib, N., McCabe, C., Medina, S., Varshavsky, M., Kitsberg, D., Dvir-Szternfeld, R., Green, G., Dionne, D., Nguyen, L., Marshall, J.L., et al. (2020). Disease-associated astrocytes in Alzheimer's disease and aging. Nat. Neurosci. *23*, 701–706. https://doi.org/10.1038/s41593-020-0624-8.

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., 3rd, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.

He, M., Tucciarone, J., Lee, S., Nigro, M., Kim, Y., Levine, J., Kelly, S., Krugikov, I., Wu, P., Chen, Y., et al. (2016). Strategies and tools for combinatorial targeting of GABAergic neurons in mouse cerebral cortex. Neuron *91*, 1228–1243. https://doi.org/10.1016/j.neuron.2016.08.021.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589. https://doi.org/10.1016/j.molcel.2010.05.004.

Hobert, O. (2016). A map of terminal regulators of neuronal identity in Caenorhabditis elegans. Wiley Interdiscip. Rev. Dev. Biol. *5*, 474–498. https://doi.org/10.1002/wdev.233.

Hrvatin, S., Hochbaum, D.R., Nagy, M.A., Cicconet, M., Robertson, K., Cheadle, L., Zilionis, R., Ratner, A., Borges-Monroy, R., Klein, A.M., et al. (2018). Single-cell analysis of experience-dependent transcriptomic states in the mouse visual cortex. Nat. Neurosci. *21*, 120–129. https://doi.org/10.1038/s41593-017-0029-5.

Jin, X., Simmons, S.K., Guo, A., Shetty, A.S., Ko, M., Nguyen, L., Jokhi, V., Robinson, E., Oyler, P., Curry, N., et al. (2020). In vivo Perturb-Seq reveals neuronal and glial abnormalities associated with autism risk genes. Science *370*, eaaz6063. https://doi.org/10.1126/science.aaz6063.

Keren-Shaul, H., Spinrad, A., Weiner, A., Matcovitch-Natan, O., Dvir-Szternfeld, R., Ulland, T.K., David, E., Baruch, K., Lara-Astaiso, D., Toth, B., et al. (2017). A unique microglia type Associated with restricting development of alzheimer's disease. Cell *169*, 1276–1290.e17. https://doi.org/10.1016/j.cell.2017.05.018.

Lake, B.B., Codeluppi, S., Yung, Y.C., Gao, D., Chun, J., Kharchenko, P.V., Linnarsson, S., and Zhang, K. (2017). A comparative strategy for single-nucleus and single-cell transcriptomes confirms accuracy in predicted cell-type expression from nuclear RNA. Sci. Rep. *7*, 6031. https://doi.org/10.1038/s41598-017-04426-w.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.

Mardinly, A.R., Spiegel, I., Patrizi, A., Centofante, E., Bazinet, J.E., Tzeng, C.P., Mandel-Brehm, C., Harmin, D.A., Adesnik, H., Fagiolini, M., and Greenberg, M.E. (2016). Sensory experience regulates cortical inhibition by inducing IGF1 in VIP neurons. Nature *531*, 371–375. https://doi.org/10.1038/nature17187.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.J. *17*, 10. https://doi.org/10.14806/ej.17.1.200.

Miyoshi, G., Young, A., Petros, T., Karayannis, T., McKenzie Chang, M., Lavado, A., Iwano, T., Nakajima, M., Taniguchi, H., Huang, Z.J., et al. (2015). Prox1 regulates the subtype-specific development of caudal ganglionic eminence-derived GABAergic cortical interneurons. J. Neurosci. *35*, 12869–12889. https://doi.org/10.1523/jneurosci.1164-15.2015.

Mo, A., Mukamel, E., Davis, F., Luo, C., Henry, G., Picard, S., Urich, M., Nery, J., Sejnowski, T., Lister, R., et al. (2015). Epigenomic signatures of neuronal diversity in the mammalian brain. Neuron *86*, 1369–1384. https://doi.org/10.1016/j.neuron.2015.05.018.

Nord, A.S., and West, A.E. (2020). Neurobiological functions of transcriptional enhancers. Nat. Neurosci. *23*, 5–14. https://doi.org/10.1038/s41593-019-0538-5.

Picelli, S., Faridani, O.R., Björklund, Å.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nat. Protoc. *9*, 171–181. https://doi.org/10.1038/nprot.2014.006.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140. https://doi.org/10.1093/bioinformatics/btp616.

Rubin, A.N., and Kessaris, N. (2013). PROX1: a lineage tracer for cortical interneurons originating in the lateral/caudal ganglionic eminence and preoptic area. PLoS One *8*, e77339. https://doi.org/10.1371/journal.pone.0077339.

Soneson, C., and Robinson, M.D. (2018). Bias, robustness and scalability in single-cell differential expression analysis. Nat. Methods *15*, 255–261. https://doi.org/10.1038/nmeth.4612.

Spiegel, I., Mardinly, A., Gabel, H., Bazinet, J., Couch, C., Tzeng, C., Harmin, D., and Greenberg, M. (2014). Npas4 regulates excitatory-inhibitory balance within neural circuits through cell-type-specific gene programs. Cell *157*, 1216–1229. https://doi.org/10.1016/j.cell.2014.03.058.

Sugino, K., Clark, E., Schulmann, A., Shima, Y., Wang, L., Hunt, D.L., Hooks, B.M., Tränkner, D., Chandrashekar, J., Picard, S., et al. (2019). Mapping the transcriptional diversity of genetically and anatomically defined cell populations in the mouse brain. Elife *8*, e38619. https://doi.org/10.7554/eLife.38619.

Taniguchi, H., He, M., Wu, P., Kim, S., Paik, R., Sugino, K., Kvitsiani, D., Fu, Y., Lu, J., Lin, Y., et al. (2011). A resource of Cre driver lines for genetic targeting of GABAergic neurons in cerebral cortex. Neuron *72*, 1091. https://doi.org/10.1016/j.neuron.2011.12.010.

Tian, L., Dong, X., Freytag, S., Lê Cao, K.A., Su, S., JalalAbadi, A., Amann-Zalcenstein, D., Weber, T.S., Seidi, A., Jabbari, J.S., et al. (2019). Benchmarking single cell RNA-sequencing analysis pipelines using mixture control experiments. Nat. Methods *16*, 479–487. https://doi.org/10.1038/s41592-019-0425-8.

Tremblay, R., Lee, S., and Rudy, B. (2016). GABAergic interneurons in the neocortex: from cellular properties to circuits. Neuron *91*, 260–292. https://doi.org/10.1016/j.neuron.2016.06.033.

Usoskin, D., Furlan, A., Islam, S., Abdo, H., Lönnerberg, P., Lou, D., Hjerling-Leffler, J., Haeggström, J., Kharchenko, O., Kharchenko, P.V., et al. (2015). Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing. Nat. Neurosci. *18*, 145–153. https://doi.org/10.1038/nn.3881.

Vieth, B., Parekh, S., Ziegenhain, C., Enard, W., and Hellmann, I. (2019). A systematic evaluation of single cell RNA-seq analysis pipelines. Nat. Commun. *10*, 4667. https://doi.org/10.1038/s41467-019-12266-7.

Wang, J., Rieder, S.A., Wu, J., Hayes, S., Halpin, R.A., de Los Reyes, M., Shrestha, Y., Kolbeck, R., and Raja, R. (2019). Evaluation of ultra-low input RNA sequencing for the study of human T cell transcriptome. Sci. Rep. *9*, 8445. https://doi.org/10.1038/s41598-019-44902-z.

Yao, Z., van Velthoven, C.T., Nguyen, T.N., Goldy, J., Sedeno-Cortes, A.E., Baftizadeh, F., Bertagnolli, D., Casper, T., Chiang, M., Crichton, K., et al. (2021). A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. Cell *184*, 3222–3241.e26. https://doi.org/10.1016/j.cell.2021.04.021.

Yap, E.-L., and Greenberg, M.E. (2018). Activity-regulated transcription: bridging the gap between neural activity and behavior. Neuron *100*, 330–348. https://doi.org/10.1016/j.neuron.2018.10.013.

Zaghlool, A., Niazi, A., Björklund, Å.K., Westholm, J.O., Ameur, A., and Feuk, L. (2021). Characterization of the nuclear and cytosolic transcriptomes in human brain tissue reveals new insights into the subcellular distribution of RNA transcripts. Sci. Rep. *11*, 4076. https://doi.org/10.1038/s41598-021-83541-1.

Zeisel, A., Muñoz-Manchado, A.B., Codeluppi, S., Lönnerberg, P., La Manno, G., Juréus, A., Marques, S., Munguba, H., He, L., Betsholtz, C., et al. (2015). Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science *347*, 1138–1142. https://doi.org/10.1126/science.aaa1934.

Zeisel, A., Hochgerner, H., Lönnerberg, P., Johnsson, A., Memic, F., van der Zwan, J., Häring, M., Braun, E., Borm, L.E., La Manno, G., et al. (2018). Molecular architecture of the mouse nervous system. Cell *174*, 999–1014.e22. https://doi.org/10.1016/j.cell.2018.06.021.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
anti-Prox1 conjugated to Alexa Fluor 647	Novusbio	NBP1-30045AF647
anti-Neun conjugated to PE	Millimark	FCMAB317PE; RRID:AB_11212465
Bacterial and virus strains		
AAV-U6-sgRNA-hSyn1-DIO-EYFP-KASH	Spiegel Lab	N/A
Chemicals, peptides, and recombinant proteins		
Tamoxifen	Merck	T5648
Deposited data		
RNA-Seq data	This paper	GSE185221
Experimental models: Organisms/strains		
Mice: Vip-IRES-Cre	The Jackson Laboratory	010908
Mice: LSL-SUN1-2xsfGFP-6xMYC (INTACT)	The Jackson Laboratory	021039
Mice: Ndnf-IRES-CreERT2	The Jackson Laboratory	034875
Recombinant DNA		
AAV-U6-sgRNA-hSyn1-DIO-EYFP-KASH	Spiegel Lab	N/A
Software and algorithms		
MATLAB	MathWorks	N/A
R	www.r-project.org	N/A

RESOURCE AVAILABILITY

Lead contact
Further information and requests for resources and reagents should be directed to and will be fulfilled upon reasonable request by the Lead Contact Ivo Spiegel (ivo.spiegel@weizmann.ac.il).

Materials availability
With the exception of the newly sub-cloned AAV plasmid AAV-U6-sgRNA-hSyn1-DIO-EYFP-KASH, no new reagents were created in this study. This newly created plasmid is available upon request from the lead contact until it will be deposited at Addgene.

Data and code availability
RNA-Seq data has been deposited at GEO (Accession number GSE185221) and will be made freely available upon acceptance of our manuscript. This paper did not report original code. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon reasonable request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Animals and DH/LE stimulation paradigm
All experiments, surgeries and procedures involving mice were conducted according to the procedures approved by the Institutional Animal Care and Use Committee (IACUC) of the Weizmann institute of Science. Mice were housed under standard conditions in a 12-h light/dark cycle with food and water *ad libitum*. For all experiments and analyses, we used tissues/samples derived from mice of both sexes and the tissues/samples of both sexes were processed and/or analyzed together (i.e. not separately for each sex). For FACS-sorting of VIP IN nuclei, we used 2-month old mice that were double-heterozygous for Vip-IRES-Cre (Taniguchi et al., 2011) and LSL-SUN1-2xsfGFP-6xMYC (INTACT; Mo et al., 2015) (JAX strains #010908 and #021039). For transcriptomic analyses in NDNF INs we crossed INTACT mice with Ndnf-IRES-CreERT2 mice (JAX strain #034875); tamoxifen (10 mg/mL) was administered at P35-P37 to induce Cre-activity and to label *Ndnf*-expressing cells (Abs et al., 2018). Visual cortices were collected at P56 and snap-frozen for Meso-Seq preparation.

For FACS purification and RNA-Seq of AAV-infected neurons, we used mice that were heterozygous for VIP-IRES-Cre in combination with stereotactic injection of AAV-U6-sgRNA-hSyn1-DIO-EYFP-KASH (see below).

For dark-housing/light-exposure (DH/LE) experiments, we used the same procedure as previously described (Hrvatin et al., 2018; Mardinly et al., 2016; Spiegel et al., 2014): 2-month old mice were housed in complete darkness for 7 days after which the visual cortices were either dissected in the dark (using night vision goggles) or after turning on regular structured light for 0.5, 1 and 3 h. Dissected tissues were snap-frozen and stored at −80C until further processing.

For analyzing the gene expression in NHP cortex, we used frozen tissue derived from the brain of an adult cynomolgus monkey (*Macaca fascicularis*, male, 5 years old) and generously provided by Dr. Rony Paz at the Weizmann Institute of Science. The use of frozen NHP brain tissue was approved by the IACUC of the Weizmann Institute of Science.

METHOD DETAILS

AAV constructs, AAV production, and stereotactic injections

AAV constructs for Cre-dependent expression of a nuclear-tagged YFP (EYFP-KASH) and expression of a gRNA cassette were cloned using the NEBuilder HIFI DNA assembly master mix (New England BioLabs, catalog no. NEB-E2621) following standard procedures. The plasmid AAV-U6-sgRNA-hSyn1-DIO-EYFP-KASH was generated by replacing the EF1a promoter in the construct AAV-Ef1a-DIO-EYFP-KASH (Addgene #27056) with a U6 driven gRNA cassette and a hSyn1 promoter.

AAV particles were produced essentially as described (Challis et al., 2019; Mardinly et al., 2016). HEK293T cells were grown in DMEM supplemented with 5% FBS, penicillin-streptomycin, NEAA and sodium pyruvate and were plated the day before transfection at a density of 12 million cells on 15cm poly-L-lysine coated plates. On the next day, equal amounts of the pDJ, pHelper and the expression cassette (13.33μg each) were transfected using PEI. Two days post-transfection, the medium containing the released AAV virions was stored at 4C and replaced with fresh medium for an additional two days. Five days post transfection, the cells and medium were collected for purification. PEG was used to precipitate AAV particles in the medium and subsequently added to the cell lysate. Using a salt active nuclease (SRE0015, Sigma), the cells were lysed to release the AAV and pelleted by centrifugation. The supernatant was loaded onto an iodixanol gradient and centrifuged for 2.25 h at 59.000 RPM (70 Ti rotor) in an ultracentrifuge. 4-5mL were collected from the clear 40% layer containing the AAV and further concentrated using Amicon filters to the desired volume. AAV titers were estimated using qPCR, after which the AAVs were aliquoted and kept at −80C for long-term storage.

Surgeries for stereotactic injections were done in 2-month old mice essentially as described (Cohen-Kashi Malina et al., 2021). Mice were anesthetized by isoflurane and secured in the stereotaxic apparatus (Kopf) on a heating pad maintained at 37C. The scalp was cleaned with 70% ethanol and betadine three times before incision and exposure of the skull. The visual cortex was estimated by stereotaxic coordinates (2.7 mm anterior from lambda and 2.5mm lateral) and 2 burr holes were drilled through the skull on either side. A glass capillary filled with AAV was inserted to a depth of 300–400μm to reach layer 2/3 of the cortex. Two min after penetrating the brain, 300nL was injected at a rate of 65 nL/min. Two minutes post-injection the pipette was retracted and repeated three more times (two injections in each hemisphere). After the injections, the scalp was closed with Vetbond (3M) and the mouse was allowed to recover on a heating pad until returned to the home cage. For pain management, animals were given buprenorphine systemically and lidocaine locally during the surgery and an additional dose on the next day.

FACS: Preparation of nuclei, antibody staining, FACS-sorting and collection/lysis of sorted nuclei

Freshly dissected visual cortices were snap-frozen in liquid nitrogen and stored at −80C. Frozen tissue was placed in a pre-cooled 2mL Dounce homogenizer containing 0.5mL of freshly prepared homogenization buffer (10mM Tris pH 8.0, 250mM sucrose, 25mM KCl, 5mM MgCl$_2$, 0.1% Triton X-100, 0.5% RNasin Plus RNase inhibitor (Promega), 1X protease inhibitor (Promega), and 0.1mM DTT). The tissue was homogenized with ten strokes of the loose Dounce pestle followed by 10 strokes with the tight pestle. The homogenate was then passed through a 50μm cell strainer and centrifuged at 700× *g* for 8 min. The pellet was resuspended in 1mL staining buffer (1X PBS, 0.8% nuclease-free BSA and 0.5% RNasin Plus RNase inhibitor). For antibody staining, anti-Prox1 antibody conjugated with Alexa Fluor 647 (Novusbio, NBP1-30045AF647) or anti-Neun conjugated to PE (Millimark, FCMAB317PE) was added at a concentration of 1ug/mL and incubated for 30 min at 4C. After incubation the homogenate was centrifuged at 400× *g* for 8 min and resuspended in 1mL staining buffer. Prior to FACS-sorting, DAPI was added (20uM final concentration) and filtered once more through a 50μm strainer to maintain a single nucleus suspension. A given number of nuclei were captured by gating for DAPI$^+$ events and excluding debris and doublets. Subsequently, the population of target nuclei were determined by fluorescence and collected into 0.2mL low-bind tubes containing lysis buffer (0.2% Triton X-100 in ddW, dNTPs, Oligo-dT30VN primers, ERCC & RNAse inhibitors) such that the final volume including the sorted nuclei was 2.35 μL.

Library preparation and RNA-Seq

Bulk RNA-seq libraries were prepared at the Crown Genomics institute of the Nancy and Stephen Grand Israel National Center for Personalized Medicine, Weizmann Institute of Science (INCPM). Libraries were prepared based on the Smart-Seq2 protocol (Picelli et al., 2014), with modifications. Briefly, 100 nuclei were collected directly into 2.03 μL lysis buffer (0.2% Triton X-100 in ddW) supplemented with dNTPs (2mM each), SMART-dT30VN primers (1mM), ERCC (1:1.25 × 10^7), and RNAse inhibitor. Reverse transcription was performed directly on isolated nuclei, followed by PCR-amplification (17 cycles). After cleanup with Agencourt Ampure XP beads (Beckman Coulter), the amplified cDNA underwent library generation using Nextera XT (Illumina). Libraries were quantified by

Qubit (Thermo Fisher Scientific) and TapeStation (Agilent). Sequencing was done on a NextSeq 500 sequencer (Illumina) using a 75 cycles high output kit, allocating ~10M reads per sample (single read sequencing).

Bioinformatic analyses

For the analysis of mouse RNA expression, Illumina adapter sequences were first removed from the raw read (fastq) files using Cutadapt (Martin, 2011), and library quality was assessed with FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Next, reads were aligned to the mm10 mouse genome using STAR (Dobin et al., 2013). Gene read counts were generated using Homer (Heinz et al., 2010) with a gtf reference file containing whole-gene annotations (thus counting both exonic and intronic reads). On average, ~90% of mapped reads were uniquely mapped to the mm10 genome.

For the analysis of the NHP RNA expression, adapter removal and library quality assessment were performed as described above. Next, reads were aligned to the *Macaca fascicularis* 6.0 monkey genome using STAR (Dobin et al., 2013). Gene read counts were generated using Homer (Heinz et al., 2010) with a gtf reference file containing whole-gene annotations (thus counting both exonic and intronic reads). On average, ~70% of mapped reads were uniquely mapped to the *Macaca fascicularis* genome.

All subsequent analysis was performed in R. Normalization of expression levels, differential expression testing and principal component analyses were performed using the DESeq2 package (Love et al., 2014).

To identify the optimal number of nuclei for collection, three technical replicates of 50/100/200/500/800 VIP IN and non-VIP cell nuclei were collected (for a total of 30 samples) from adult mice visual cortices. R^2 of the gene expression correlation between technical replicates is derived from a fitted linear model of the data. To identify the number of genes passing the basic expression threshold, the average expression and the standard deviation were calculated for the three technical replicates of each cell type and each nuclei amount collected. Genes were flagged as expressed if their average count was over 10 reads and their CV value was <=0.5.

For comparing gene expression between biological replicates, three biological replicates of 100 VIP IN and non-VIP cell nuclei were collected from adult mice visual cortices. Gene expression barplots show the average and standard error of the mean of three biological replicates. Genes were flagged as expressed if their average count was over 10 reads and their CV value was <=0.5. Genes were classified as "cell-type-specific" if they were found to be 5-fold enriched in one cell type compared to the other with an adjusted p-value <=0.05.

For comparing the gene expression between the VIP IN/non-VIP cell 100-nuclei samples and previously-published RNA sequencing data of affinity-purified cortical VIP IN nuclei (Mo et al., 2015), fastq files of VIP IN and cortex samples were downloaded via GEO accession GSM1541954 (SRR1647858, SRR1647859, SRR1647860, SRR1647861- 2 biological replicates of each cell type). For these samples, 5 base pairs were removed from all 5' ends to remove possible adapter contamination, and reads were aligned with STAR and counted with Homer as for the nuclear SmartSeq data. Genes were flagged as expressed if their average count was over 10 reads and their CV value was <=0.5. Genes were classified as "cell-type-specific" if they were found to be 5-fold enriched in one cell type compared to the other with an adjusted p-value <=0.05.

To find experience-regulated genes in VIP IN and non-VIP cell nuclei, three biological replicates of VIP IN and non-VIP cell samples were collected from adult mice visual cortices after dark housing, followed by light exposure of either 0.5, 1 or 3 h. Genes were flagged as expressed if their average count was over 10 reads and their CV value was <=0.5. Genes were classified as "cell-type-specific" if they were found to be 5-fold enriched in one cell type compared to the other with an adjusted p-value <=0.05. Then, expression at the 0.5, 1 and 3- h time points was compared to the expression in the dark condition. Genes were classified as "experience-regulated" if they had a fold change of >=2 between the time points and a p-value of <=0.005. For genes that were significantly enriched in more than one time point, the strongest regulated time point was selected as the regulation time point.

For comparing experience-regulated genes in VIP INs identified by Meso-Seq with those identified by scRNA-Seq, we used our own Meso-Seq data and the relevant gene lists generated from a previously published scRNA-Seq study (Hrvatin et al., 2018) (41593_2017_29_MOESM5_ESM.xlsx).

For comparing the list of experience-regulated genes in VIP INs identified by Meso-Seq with those identified by Ribotag-Seq, we used our own Meso-Seq data and re-analyzed the relevant RiboTag-Seq data generated in a previous study (Mardinly et al., 2016). For this, we downloaded the VIP-Cre IP sequencing samples (GEO accession GSE77243 [GSM2046545-GSM2046559], 3 biological replicates of each of the time points standard/dark/1 h/3 h 7.5 h). Next, reads were aligned to the mm10 mouse genome using STAR (Dobin et al., 2013) and gene read counts were generated using Homer (Heinz et al., 2010) with a gtf reference file containing whole-gene annotations (thus counting both exonic and intronic reads). Genes were flagged as expressed if their average count was over 10 reads and their CV value was <=0.5. Then, expression in the 1, 3 and 7.5- hour time points were compared to the expression in the dark condition. Genes were classified as "experience-regulated" if they had a fold change of >=2 between the time points and a p-value of <=0.005. For genes that were significantly enriched in more than one time point, the strongest regulated time point was selected as the regulation time point.

For analyzing the gene expression in adult cynomolgus monkey (*Macaca fascicularis*), two technical replicates of each cell type (PROX1$^+$/NeuN$^+$, PROX1$^-$/NeuN$^+$ and Input samples) were FACS-collected from the cortex and Meso-Seq'ed. Genes were flagged as expressed if their average count was over 10 reads and their CV value was <=0.5. Genes were classified as "cell-type-specific" if they were found to be 5-fold enriched in one cell type compared to the other with an adjusted p-value <=0.05. Expressed genes were binned according to their CV value into 1 of 3 groups: CV<=0.5, 0.5 < CV < 1, CV>=1. For comparing gene expression between

monkey and mouse, only one-to-one orthologs were selected. Expression level in each organism was normalized within the organism, and species-specific PROX1⁺/NeuN⁺-enriched genes were found, within in each species, as genes with a 5-fold enrichment and an adjusted p-value <=0.05 when comparing PROX1⁺/NeuN⁺ vs. Input samples.

For comparing the gene expression between AAV-injected and non-injected VIP IN and non-VIP cell samples, three biological replicates of 100 nuclei VIP IN and non-VIP cell nuclei were collected. Genes were flagged as expressed if their average count was over 10 reads and their CV value was <=0.5. Genes were classified as "cell-type-specific" if they were found to be 5-fold enriched in one cell type compared to the other with an adjusted p-value <=0.05. Expressed genes were binned according to their CV value into 1 of 3 groups: CV<=0.5, 0.5 < CV < 1, C>=1.

To determine the specificity of FACS-sorting of NDNF IN nuclei, 2 biological replicates of 100 nuclei of distinct FACS fractions were sorted from adult mice visual cortex for comparing gene expression between NDNF INs, non-NDNF neurons, endothelial cells and other nuclei (Figures 5B–5D). To determine the number of genes expressed in NDNF IN nuclei (Figure 5E), three biological replicates of NDNF IN nuclei were collected from adult mice visual cortices. Average expression as well as CV were calculated based on the three samples, and expressed genes were binned according to their CV value into 1 of 3 groups: CV<=0.5, 0.5 < CV < 1, C>=1.

To determine the number of expressed genes identified in NDNF INs by single-cell RNA-Seq techniques, we used publicly available single-cell datasets generated with 10× genomics and SMART-Seq4. For 10× genomics data we retrieved the raw expression matrix from GEO (GSE185862) which contains all the counts for all the single-cell libraries in the study. Based on the metadata of this dataset we extracted only the counts for the Lamp5-clustered libraries and subsequently selected all the single-cell libraries that contained >1 read for *Ndnf*. The resulting libraries were then normalized together using the R-package Seurat (Hao et al., 2021) and gene expression was analyzed as described above for Meso-Seq. For the SMART-Seq4 analysis, we used the raw fasta files from GEO (GSE115746) of the libraries that clustered into the Lamp5 subtype based on the metadata (see SRR accession list in Table S3). Paired-end reads were aligned to the mm10 genome using STAR and counted using HOMER as described above for the Meso-Seq protocol. The resulting count matrix was subsequently filtered for libraries that contained >1 reads for Ndnf and finally the libraries were normalized with the R-package SCnorm (Bacher et al., 2017), and gene expression was analyzed as described above.

QUANTIFICATION AND STATISTICAL ANALYSES

For RNA-seq libraries, statistical tests for assessing differential gene expression were done with the Wald-test in the R-studio package Deseq2 (each experimental group/time-point contained three technical or biological replicates). Specific filtering parameters are described in the text and the corresponding figure legends. Fold-changes in gene expression as measured by qPCR samples was calculated with the Delta-Delta Ct method (three biological replicates per time point). In all the figures, the mean is depicted with the standard error of the mean.