RESEARCH ARTICLE SUMMARY

NEUROGENOMICS

In vivo Perturb-Seq reveals neuronal and glial abnormalities associated with autism risk genes

Xin Jin*, Sean K. Simmons, Amy Guo, Ashwin S. Shetty, Michelle Ko, Lan Nguyen, Vahbiz Jokhi, Elise Robinson, Paul Oyler, Nathan Curry, Giulio Deangeli, Simona Lodato, Joshua Z. Levin, Aviv Regev*†, Feng Zhang*†, Paola Arlotta*†

INTRODUCTION: Human genetic studies have revealed long lists of genes and loci associated with risk for many diseases and disorders, but to systematically evaluate their phenotypic effects remains challenging. Without any a priori knowledge, these risk genes could affect any cellular processes in any cell type or tissue, which creates an enormous search space for identifying possible downstream effects. New high-throughput approaches are needed to functionally dissect these large gene sets across a spectrum of cell types in vivo.

RATIONALE: Analysis of trio-based whole-exome sequencing has implicated a large number of de novo loss-of-function variants that contribute to autism spectrum disorder and developmental delay (ASD/ND) risk. Such de novo variants often have large effect sizes, thus providing a key entry point for mechanistic studies. We have developed in vivo Perturb-Seq to allow simultaneous assessment of the individual phenotypes of a panel of such risk genes in the context of the developing mouse brain.

RESULTS: Using CRISPR-Cas9, we introduced frameshift mutations in 35 ASD/ND risk genes in pools, within the developing mouse neocortex in utero, followed by single-cell transcriptomic analysis of perturbed cells from the early postnatal brain. We analyzed five broad cell classes—cortical projection neurons, cortical inhibitory neurons, astrocytes, oligodendrocytes, and microglia—and selected cells that had received only single perturbations. Using weighted gene correlation network analysis, we identified 14 covarying gene modules that represent transcriptional programs expressed in different classes of cortical cells.

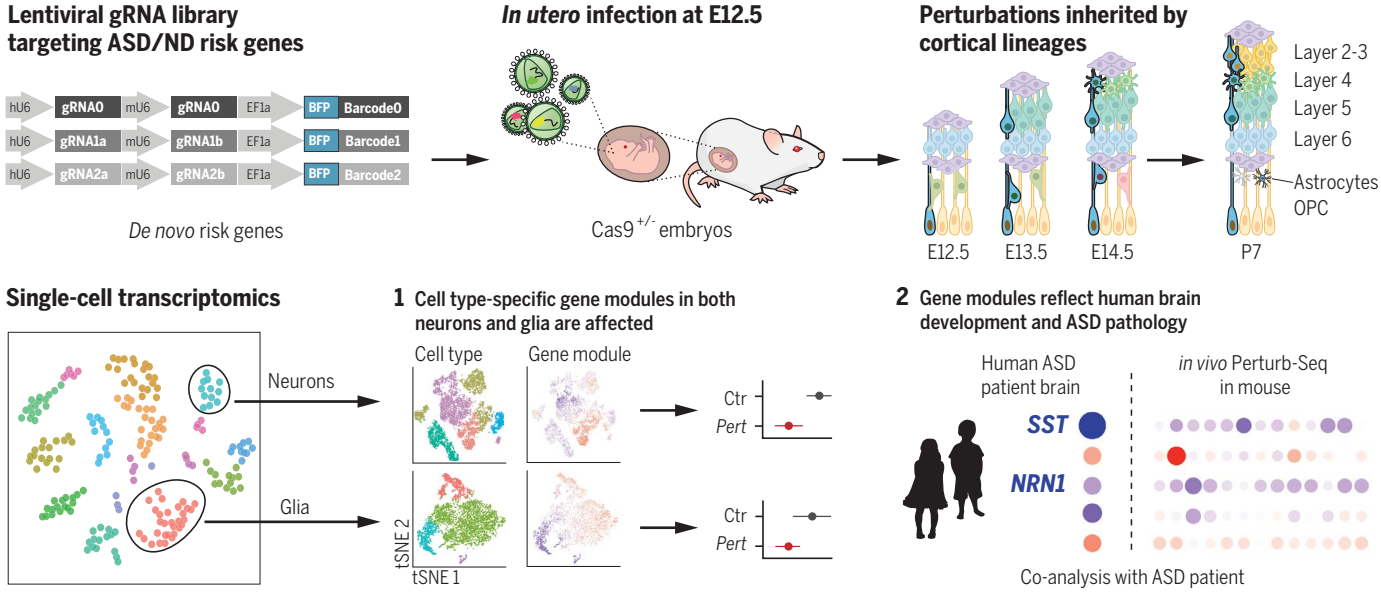
These modules included both those affecting common biological processes across multiple cell subsets and others representing cell type–specific features restricted to certain subsets. We estimated the effect size of each perturbation on each of the 14 gene modules by fitting a joint linear regression model, estimating how module gene expression in cells from each perturbation group deviated from their expression level in internal control cells. Perturbations in nine ASD/ND genes had significant effects across five modules across four cell classes, including cortical projection neurons, cortical inhibitory neurons, astrocytes, and oligodendrocytes. Some of these results were validated by using a single-perturbation model as well as a germline-modified mutant mouse model.

To establish whether the perturbation-associated gene modules identified in the mouse cerebral cortex are relevant to human biology and ASD/ND pathology, we performed co-analyses of data from ASD and control human brains and human cerebral organoids. Both gene expression and gene covariation ("modularity") of several of the gene modules identified in the mouse Perturb-Seq analysis are conserved in human brain tissue. Comparison with single-cell data from ASD patients showed overlap in both affected cell types and transcriptomic phenotypes.

CONCLUSION: In vivo Perturb-Seq can serve as a scalable tool for systems genetic studies of large gene panels to reveal their cell-intrinsic functions at single-cell resolution in complex tissues. In this work, we demonstrated the application of in vivo Perturb-Seq to ASD/ND risk genes in the developing brain. This method can be applied across diverse diseases and tissues in the intact organism. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: xinjin@fas.harvard.edu (X.J.); aregev@broadinstitute.org (A.R.); zhang@broadinstitute.org (F.Z.); paola_arlotta@harvard.edu (P.A.)
†These authors contributed equally to this work.
Cite this article as X. Jin *et al.*, *Science* **370**, eaaz6063 (2020). DOI: 10.1126/science.aaz6063

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.aaz6063



In vivo Perturb-Seq identified neuron and glia-associated effects by perturbations of risk genes implicated in ASD/ND. De novo risk genes in this study were chosen from Satterstrom *et al.* (2018), and co-analysis with ASD patient data at bottom right is from Velmeshev *et al.* (2019); full citations for both are included in the full article online.

RESEARCH ARTICLE

NEUROGENOMICS

In vivo Perturb-Seq reveals neuronal and glial abnormalities associated with autism risk genes

Xin Jin[1,2,3,4]*, Sean K. Simmons[3,5,6], Amy Guo[3], Ashwin S. Shetty[2,3,5], Michelle Ko[2], Lan Nguyen[3,6], Vahbiz Jokhi[2], Elise Robinson[3,5,7], Paul Oyler[2], Nathan Curry[2], Giulio Deangeli[2], Simona Lodato[8], Joshua Z. Levin[3,5,6], Aviv Regev[3,6,9,10]*†‡, Feng Zhang[3,4,10]*†, Paola Arlotta[2,3,5]*†

The number of disease risk genes and loci identified through human genetic studies far outstrips the capacity to systematically study their functions. We applied a scalable genetic screening approach, in vivo Perturb-Seq, to functionally evaluate 35 autism spectrum disorder/neurodevelopmental delay (ASD/ND) de novo loss-of-function risk genes. Using CRISPR-Cas9, we introduced frameshift mutations in these risk genes in pools, within the developing mouse brain in utero, followed by single-cell RNA-sequencing of perturbed cells in the postnatal brain. We identified cell type–specific and evolutionarily conserved gene modules from both neuronal and glial cell classes. Recurrent gene modules and cell types are affected across this cohort of perturbations, representing key cellular effects across sets of ASD/ND risk genes. In vivo Perturb-Seq allows us to investigate how diverse mutations affect cell types and states in the developing organism.

Human genetic studies have uncovered associations between genetic variants in tens of thousands of loci and complex human diseases (*1*–*3*). In particular, analysis of trio-based whole-exome sequencing (WES) has implicated a large number of de novo loss-of-function variants that contribute to the risk of neurodevelopmental pathologies, including autism spectrum disorders and neurodevelopmental delay (ASD/ND) (*4*, *5*). Compared with common variants identified through genome-wide association studies (GWASs), such de novo risk variants often have large effect sizes, are highly penetrant, and occur within a gene's coding region, thus providing a crucial entry point for disease modeling and mechanistic studies.

However, a major challenge remains in identifying the point of action of each of these many risk genes. For example, ASD/ND comprises a broad collection of neurodevelopmental disorders with highly heterogeneous genetic contributions, including hundreds of highly penetrant de novo risk variant genes (*6*). Moreover, there is substantial diversity in the function of the gene products that these genes encode, precluding a clear a priori prediction of the underlying brain cell types, developmental processes, and molecular pathways affected during neurodevelopment (*7*). Few of these risk genes have been studied in animal or cellular models, and thus, their function during brain development remains poorly defined. Because of the labor-intensive and time-consuming nature of generating and analyzing individual knockout animal models for functional investigation, it is crucial to develop phenotyping methods that are scalable, general-purpose, high-resolution, and high-content to identify tissue- and cell type–specific effects of genetic perturbations in vivo.

To address these challenges, we developed in vivo Perturb-Seq, a scalable genetic screen, to investigate the function of large sets of genes at single-cell resolution in complex tissue in vivo. We applied in vivo Perturb-Seq in utero in mice to study the effect of a panel of ASD/ND risk genes on brain development.

In vivo Perturb-Seq to assess the function of ASD/ND risk genes

We chose ASD/ND candidate genes from a recently published WES study of 11,986 cases with 6430 ASD/ND probands (table S1) (*4*). We initially prioritized 38 candidate genes (of which 35 were retained in the final analysis) (table S1) that harbor de novo variants specific to ASD/ND patients within the broader class of neurodevelopmental disability (fig. S1A and table S1). These ASD/ND risk genes are expressed in human brain tissue, as assessed by the BrainSpan bulk RNA-seq dataset (*8*); some are highly expressed at embryonic stages, and others are highly expressed from early postnatal to adult stages (fig. S1B). On the basis of mouse cortical single-cell RNA-sequencing (scRNA-seq) data, the orthologs of these ASD/ND risk genes are expressed in diverse cell types (fig. S2) [embryonic day 18.5 (E18.5) data are from the 10x Genomics public dataset (*9*); P7 data are from this work]. Thus, these ASD/ND genes could in principle act in many different cell types and temporal frames, requiring scalable methods to test gene function across a range of cell types and developmental events.

For in vivo Perturb-Seq, we used Cas9-mediated genome editing (*10*–*12*) in a pooled approach to introduce mutations in each of the ASD/ND risk genes within progenitor cells of the mouse developing forebrain in utero, followed by scRNA-seq at P7 to read out both a barcode identifying the perturbation and the expression profile of the perturbed cells (Fig. 1A). Specifically, we used a transgenic mouse line that constitutively expresses Cas9 (*13*) and delivered pools of guide RNAs (gRNAs) targeting the different risk genes through lentiviral infection into the lateral ventricles of the developing embryo in utero. Each lentiviral vector contained two different gRNAs targeting the 5′-end coding exons of one ASD/ND gene (to enhance knockout efficiency) and a blue fluorescent protein (BFP) reporter with a distinct barcode corresponding to the perturbation identity (*10*–*12*). To minimize vector recombination, we packaged each lentivirus separately and then pooled viruses at equal titers.

We injected a pool of lentiviruses with equal gRNA representation into the ventricles of the developing forebrain at E12.5 (Fig. 1A). In this approach, lentiviral injection leads to infection of neural progenitors lining the lateral ventricle of the developing forebrain, including progenitors of the neocortex and the ganglionic eminences. Because lentiviral vectors integrate into the genome, the progeny of the infected progenitors are labeled by BFP and carry a perturbation barcode corresponding to the target ASD/ND gene.

Both immunohistochemical analysis and scRNA-seq of BFP$^+$ cells at P7 showed that the Perturb-Seq vectors were expressed across a variety of neuronal and glial cell types in the cortex (Fig. 1, B and C, and fig. S3, A and B). Although microglia originate mostly from outside the targeted germinal zones, we detected lentiviral vector expression in cortical microglia, indicated by the presence of BFP as well as perturbation barcode expression, across multiple individual experiments (fig. S4, E and F). Although it is unclear how microglia are labeled in our experimental procedure, it is possible that the in utero injection could

[1]Society of Fellows, Harvard University, Cambridge, MA, USA. [2]Department of Stem Cell and Regenerative Biology, Harvard University, MA, USA. [3]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [4]McGovern Institute of Brain Science, Department of Brain and Cognitive Science, Department of Biological Engineering, Massachussetts Institute of Technology (MIT), Cambridge, MA, USA. [5]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [6]Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [7]Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA, USA. [8]Department of Biomedical Sciences and Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS) Humanitas Clinical and Research Center, Humanitas University, Milan, Italy. [9]Koch Institute of Integrative Cancer Research, Department of Biology, MIT, Cambridge, MA, USA. [10]Howard Hughes Medical Institute, Chevy Chase, MD, USA.
*Corresponding author. Email: xinjin@fas.harvard.edu (X.J.); aregev@broadinstitute.org (A.R.); zhang@broadinstitute.org (F.Z.); paola_arlotta@harvard.edu (P.A.)
†These authors contributed equally to this work.
‡Present address: Genentech, 1 DNA Way, South San Francisco, CA.

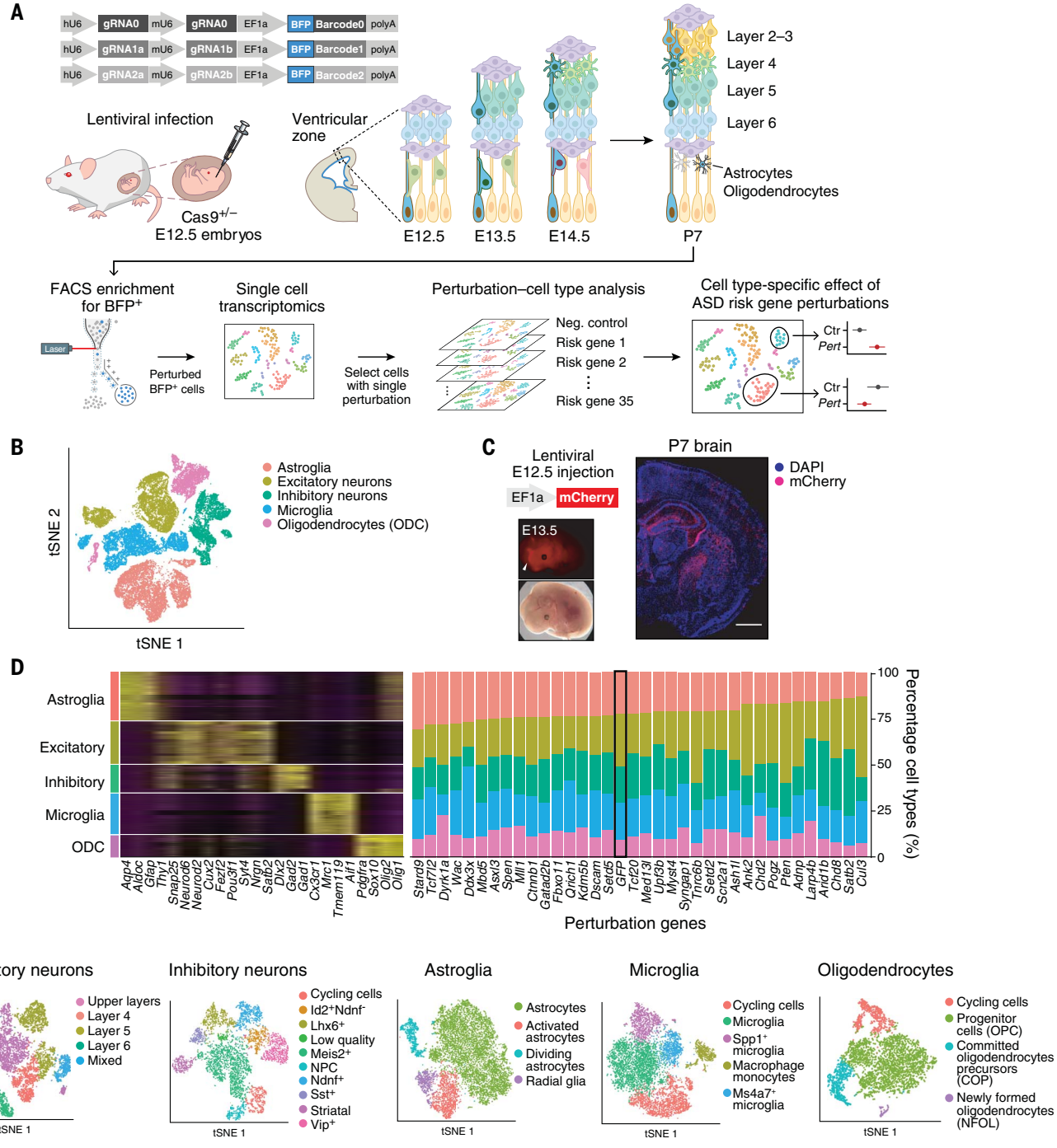


Fig. 1. In vivo Perturb-Seq to investigate functions of a panel of ASD/ND risk genes harboring de novo variants. (**A**) Schematics of the in vivo Perturb-Seq platform, which introduces mutations in individual genes in utero at E12.5, followed by transcriptomic profiling of the cellular progeny of these perturbed cells at P7 by means of scRNA-seq. (**B**) *t*-distributed stochastic neighbor embedding (*t*-SNE) of five major cell populations identified in the Perturb-Seq cells. (**C**) In vivo Perturb-Seq lentiviral vector carrying an mCherry reporter drives detectable expression within 24 hours and can sparsely infect brain cells across many brain regions. Scale bar, 1000 μm. (**D**) Cell-type analysis of in vivo Perturb-Seq of ASD/ND de novo risk genes. Canonical marker genes were used to identify (left) major cell clusters and (right) cell-type distribution in each perturbation group. Negative control (GFP) is indicated with a black rectangle. (**E**) *t*-SNEs showing the subclusters of each of the five major cell types, identified by reclustering each cell type separately.

have led to either local lesions that recruited and expanded the number of microglia along the injection tract or that microglia were labeled within the parenchyma along the same tract. Overall, our approach allowed us to examine the effects of each perturbation across a wide range of cell types from distinct brain regions (such as cortical projection neurons, interneurons, astroglia, and oligodendroglia), and under sparse labeling conditions in which less than 0.1% of cells in the cortex were perturbed, and thus development of individual perturbed cells was highly unlikely to be affected by perturbed neighboring cells (fig. S3, A to C).

In vivo Perturb-Seq targets diverse cell types without affecting overall cell type composition

We performed the experiment with 18 different cohorts of pregnant mice, for a total of 163 embryos, each subjected to the entire pool of perturbations. We microdissected and dissociated cortical tissues at P7, enriched the perturbed cells by means of fluorescence-activated cell sorting (FACS), and used droplet-based scRNA-seq to obtain each cell's expression profile along with its perturbation barcode. The cell survival rate after FACS was 78%, and we confirmed a 40 to 70% frameshift insertion or deletion for each gRNA target among the infected cells (fig. S3, D and E).

This multiplexed experimental design allowed us to test the cell-autonomous effect of all perturbations against the effect of a negative control construct targeting the endogenous green fluorescent protein (GFP) in the Rosa26 locus, thus controlling for effects related to viral infection and DNA double-strand break, among other confounders. To minimize batch-dependent variation, the control construct was included in the same pool as that of the perturbation vectors (fig. S3F). After quality control, we retained for further analysis a total of 46,770 neocortical cells across 17 high-quality experimental batches. We partitioned the cells into major cell classes using Louvain clustering (*14*) and annotated them by known marker gene expression (Fig. 1D) (*15*, *16*).

We focused on five broad cell populations from this cortical dataset for downstream analysis: cortical projection neurons (8450 cells), cortical inhibitory neurons (5532 cells), astrocytes (9526 cells), oligodendrocytes (4279 cells), and microglia/macrophages (8070 cells) (thus excluding vascular, endothelial, and contaminant hippocampal and striatal cells). We further filtered out some remaining low-quality cells in these five major cell classes, retaining 35,857 high-quality cells (median of 2436 detected genes per cell overall, and a median of 4084 genes in the projection neuron cluster, as expected from their large size and known high RNA content) (fig. S3G). We subclustered each of the five major cell types separately and annotated biologically meaningful subclusters (Fig. 1E and fig. S6).

From inspecting the perturbation barcodes from the lentiviral constructs, 92% of the cells (33,231 cells) in these five major cell classes had at least one perturbation read assigned to them, and 50% had barcodes for a single gene (18,044 cells) (fig. S4, A to C), reflecting the low multiplicity of infection (fig. S4D). Because it is rare for multiple ASD/ND loss-of-function risk gene mutations to co-occur in patients, we focused on the 18,044 cells that carried a single perturbation. We found a median of 338 cells per perturbation. After excluding perturbations with <70 perturbed cells, we retained 35 ASD/ND risk gene perturbations. BFP from the lentiviral vector was robustly detected as one of the most highly expressed genes in all retained cells (fig. S4E). The BFP detection rate in each cell type correlated with the average number of genes detected (fig. S4F), further supporting the reliability of the readout.

ASD/ND risk gene perturbations had a very modest effect on the presence and proportions of these five major cell types relative to the negative control (targeting the *GFP* gene). Only loss of *Dyrk1a* had a significant effect on cell type composition, increasing the proportion of oligodendrocytes and reducing the proportion of microglia and macrophages [false discovery rate (FDR)–corrected $P < 0.05$ using Poisson regression (*17*)] (Fig. 1D and fig. S5).

Covarying gene modules associate with cell states

To assess whether ASD/ND genetic perturbations caused molecular changes and alterations in cell states, we first sought to define gene modules that covary within each of the five broad cell classes. As previous work has shown (*10*–*12*, *18*), focusing on gene modules instead of individual genes provides more statistical power to detect biologically meaningful perturbation effects by using fewer cells than would be required for single-gene–level analysis and can capture diversity both within and across cell types.

We first tested whether the expression of known Gene Ontology (GO) gene sets (*19*) was affected by calculating a gene-set expression score for each cell and fitting a linear regression model to this score. After correcting for multiple hypothesis testing, no GO terms were significantly altered by any perturbation (table S8). However, this approach is limited by the large number of tests performed (one test per GO term per cell type per perturbation, for a total of 510,265 tests), as well as the limited number of GO terms relevant to the developing cortex.

We therefore sought to identify gene modules de novo in our data using two approaches: weighted gene correlation network analysis (WGCNA), which identifies "modules" of genes with correlated expression, and structural topic modeling (STM), which attempts to reduce the dimensionality of the gene expression matrix and returns "topics" corresponding to the components of this representation (Fig. 2A, figs. S6 to S8, and tables S2 and S3) (*20*, *21*). We performed these analyses for each of the five major cell clusters separately, to better identify effects associated with specific cell types; our nomenclature for the modules incorporates the cell cluster analysis it is derived from (for example, PN1 represents a module identified through analysis of projection neurons). Each of these analyses used the full set of perturbations to identify effects shared across multiple perturbations. We focused our subsequent analysis on the 14 modules identified with WGCNA because they were highly correlated with one or more topics returned by STM (fig. S7).

The 14 WGCNA modules comprised two broad categories. Some reflected common biological processes and were present across multiple cell subsets (such as cell cycle, differentiation, and maturation). For example, module PN2 is associated with genes involved in neurite development and varied across cells in multiple projection neuron subclusters (fig. S6A). Others represented cell type–specific features specific to only some subsets (such as subcluster-specific features of a neuronal subtype). For example, module PN1 is a module associated with two defined subclusters of neurons of layer 4 and layer 5 (fig. S6A).

ASD/ND gene perturbations affect cell states in multiple cell classes

Because the WGCNA analysis is expected to recover gene modules associated with many kinds of variation across the data, we next tested the association of each risk gene perturbation with the 14 individual WGCNA gene modules. We estimated the effect size of each perturbation on each gene module by fitting a joint linear regression model, estimating how module gene expression in cells from each perturbation group deviated from the GFP control cells (Fig. 2, A and B). To ensure that no single perturbation or batch dominated the linear model, we down-sampled the cells in each cell category so that no perturbation had more than two times the median number of cells over all perturbations. This linear regression analysis was performed on mean-centered and standard deviation–scaled module scores, so effect sizes can be interpreted in terms of standard deviations from the population mean (Fig. 2B). Our modeling approach assumes that module expression in individual cells is independent after conditioning on the experimental batch and that noise is normally distributed. To evaluate the effects of these assumptions, we also compared alternative approaches, including a linear mixed model–based approach and a permutation-based approach (fig. S9 and table S9).

Perturbations in nine ASD/ND genes (*Adnp*, *Ank2*, *Ash1l*, *Chd8*, *Gatad2b*, *Pogz*, *Scn2a1*, *Stard9*, and *Upf3b*) had significant effects across five modules (compared with the GFP control, FDR-corrected $P < 0.05$) (Fig. 2B, indicated circles, and table S4): a module associated with neurons of layers 4 and 5 (PN1, affected by perturbations in *Adnp*, *Ash1l*, *Scn2a1*, and *Stard9*); modules representing two distinct homeostatic signatures in astrocytes (Astro1 affected by perturbation of *Scn2a1*, and Astro3 affected by perturbations of *Chd8*, *Pogz*, and *Upf3b*); a module associated with oligodendrocyte progenitor cells (ODC1, affected by perturbations of *Chd8*

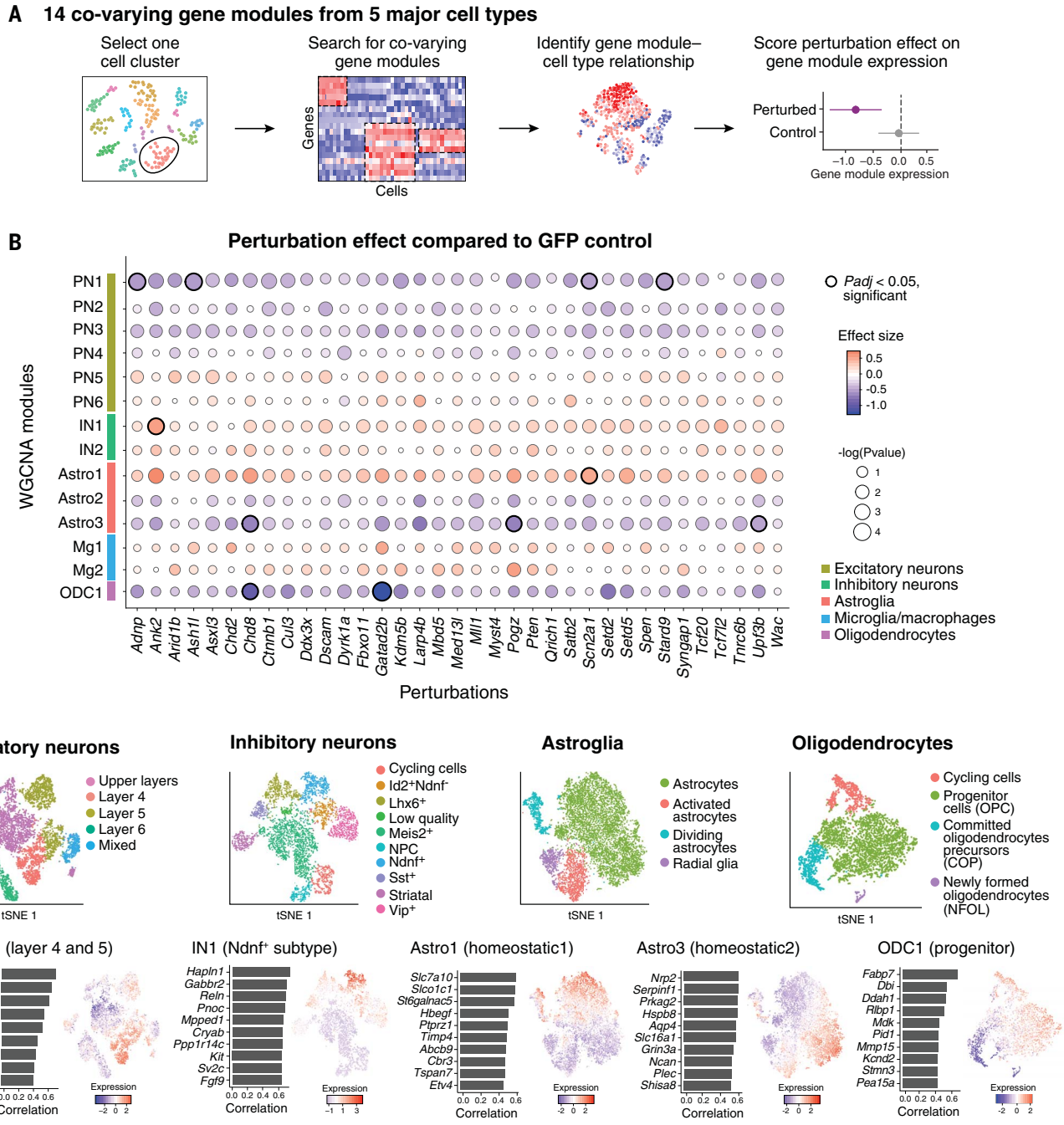


Fig. 2. In vivo Perturb-Seq reveals cell type–specific effects of ASD/ND risk gene perturbations. (**A**) Schematic illustration of the Perturb-Seq analysis pipeline. (**B**) ASD/ND risk gene perturbation effects in different WGCNA gene modules compared with GFP controls. Dot color corresponds to effect size, and dot size corresponds to negative base 10 log(*P* value). Module gene lists are presented in table S2. *P* values were calculated from linear modeling, and P_{adj} was calculated by means of Benjamini and Hochberg FDR correction. (**C**) The four cell types and five gene modules that were altered by ASD risk gene perturbations. (Top) Subcluster *t*-SNE of each cell class (repeated from Fig. 1E for ease of comparison). (Bottom) Feature plots of gene module expression scores and the top correlated genes within each module across the relevant cell class.

and *Gatad2b*); and a module associated with *Ndnf*$^+$ interneurons (IN1, affected by the perturbation of *Ank2*) (Fig. 2C and fig. S6).

The oligodendrocyte progenitor module (ODC1) also had a significant amount of its variation across the oligodendrocyte cell cluster explained by the perturbation state overall [van der Waerden test, a nonparametric alternative to analysis of variance (ANOVA), FDR-corrected $P < 0.05$] (fig. S5C), suggesting that this module represents most shared effects across different perturbed genes. Collectively, the data indicate that a selected group of perturbations was able to affect recurrent gene modules with cell-type specificity and point to some

convergent effects across diverse ASD/ND risk genes.

Single perturbation of *Ank2* confirms Perturb-Seq effect on an interneuron gene module

In our multiplex in vivo Perturb-Seq results, *Ank2* perturbation led to increased expression of an interneuron module (IN1) (FDR-corrected $P < 0.05$) (fig. S10). This module was strongly correlated with a subcluster of inhibitory interneurons expressing *Ndnf* (fig. S10, C and D) and contains genes such as *Kcnq5* (a voltage-gated potassium channel) and *Gabbr2* [γ-aminobutyric acid (GABA) receptor subunit] (fig. S6B and table S2). To validate our finding from the pooled Perturb-Seq experiment, we performed a single perturbation targeting either *Ank2* or *GFP* (control), followed by scRNA-seq of neocortical cells at P7, resulting in 2943 and 1716 high-quality cells, respectively.

The individual perturbation experiment confirmed the results from the pooled Perturb-Seq screen. *Ank2*-perturbed cells were present across all cell types, and overall proportions of cells were not significantly changed (fig. S10B). Within the *Ndnf*$^+$ interneurons, *Ank2* perturbation led to up-regulation of the IN1 module (FDR-corrected $P < 0.05$) (fig. S10E), confirming the Perturb-Seq result. This finding indicates that multiplexing perturbations in the pooled approach does not significantly distort the results observed for an individually perturbed gene.

Ank2 encodes an ankyrin protein and is expressed broadly in excitatory and inhibitory neurons as well as glial cells in the brain (*22*). Studies examining *Ank2* loss of function suggest that it is involved in axonal morphology, connectivity, and calcium signaling in excitatory neurons (*23*–*26*). Our Perturb-Seq data suggests a role of *Ank2* in the *Ndnf*$^+$ interneuron subtype during cortical development, in addition to its known roles in excitatory neurons.

The ASD/ND risk genes *Chd8* and *Gatad2b* alter gene modules in oligodendrocyte progenitors

In our Perturb-Seq experiment, *Chd8* and *Gatad2b* perturbations significantly decreased the expression of the ODC1 module in the oligodendrocyte cluster (FDR-corrected $P < 0.05$) [Fig. 3, A to D; an alternative measurement of effect size is provided in fig. S11A, estimated by log transcripts per million (TPM) gene expression differences]. The ODC1 module is highly expressed in cycling cells and oligodendrocyte precursor cells (OPCs) and lowly expressed in committed oligodendrocyte progenitor cells (COPs) and newly formed oligodendrocytes (NFOLs), suggesting that this module is linked to oligodendrocyte maturation (Fig. 3A) and therefore that perturbation in *Chd8* and *Gatad2b* might alter oligodendrocyte maturation. This is consistent with recent reports that *Chd8* loss of function potentiates an impaired OPC development phenotype caused by deletion of *Chd7* (*27*).

We further investigated and validated this result by examining oligodendrocyte development in a *Chd8* germline heterozygous mutant model [because homozygous mutation is embryonic lethal (*28*)], using several orthogonal methods. First, we used in situ hybridization for two canonical OPC markers that are involved in fate specification, *Cspg4* (a member of the ODC1 module) and *Pdgfra* (platelet-derived growth factor receptor A). Both were down-regulated in P7 *Chd8*$^{+/-}$ cortex (Fig. 3E and fig. S11, B to D), which is consistent with our in vivo Perturb-Seq results. Second, we used immunohistochemistry to examine a later developmental time point, P11. OPC cell number (such as PDGFRA$^+$ cells) did not show significant differences between the wild-type and *Chd8*$^{+/-}$ littermates, which is also consistent with in vivo Perturb-Seq; however, cells positive for myelin basic protein (MBP), a marker of myelinating oligodendrocytes, were increased in number and displayed increased MBP amounts in the *Chd8*$^{+/-}$ mutant (FDR-corrected $P < 0.05$, nonparametric ANOVA test) (Fig. 3F). In combination with the Perturb-Seq result that showed reduction in the signature of oligodendrocyte progenitor-expressed ODC1 module in *Chd8*-perturbed cells, this suggests that C*hd8* perturbation may result in acceleration of the increase in MBP amounts that occurs postnatally. These data further demonstrate that in vivo Perturb-Seq has the power to identify cell type–specific molecular changes similar to those observed in a single-gene, germline-modified mouse model.

Perturb-Seq gene modules are conserved between human and mouse

To establish whether the perturbed gene modules identified in the mouse cerebral cortex are conserved in human cells, we examined the expression of each module across multiple scRNA-seq and single-nuclei RNA-seq (snRNA-seq) datasets from human tissues: adult human cortex (*29*), ASD donor cortex with matched controls (*31*), fetal human cortex (*31*), and 3- and 6-month-old human brain organoids (Fig. 4, A and B) (*32*). In the fetal brain and the 3-month-old brain organoid samples, glial cell types were sparsely represented owing to the early developmental stages of the samples (fig. S12A). We identified human genes that had 1:1 orthologs to the mouse genes in each module and asked whether the modules were conserved, using two metrics: whether the orthologous genes were also expressed in the corresponding cell type in the human datasets and whether the expression of the genes in each module covaried across single cells (as estimated from correlation), reflecting the degree of "modularity" of these mouse gene programs in humans.

The expression of each module was largely conserved in all human datasets, with different modules showing distinct levels of conservation of expression in each dataset (Fig. 4A). Some modules—such as PN1, PN2, and PN5—displayed high levels of conservation of expression (with at least 75% of the genes in these modules being expressed by at least 5% of cells in the corresponding associated cell type) across all datasets. The proportions of the genes expressed in the corresponding cell types in human tissues were generally lower than in mouse tissues (fig. S12B).

We further calculated whether the covariation of expression of the genes in each module (their "modularity") was also comparable in humans. To do so, for each module and each dataset we calculated the average pairwise expression correlation coefficient between the genes in a given module and compared it with a module-specific null-distribution based on random gene sets with similar expression levels, to calculate both a *P* value for the correlation of our modules and a normalized correlation coefficient. Out of 14 modules, eight showed greater intramodule correlation than that of a comparable random gene set in the adult human brain dataset from Hodge *et al.* (*29*) (Fig. 4B). Correlation also increased with the age of the human samples across brain regions of the BrainSpan dataset (Fig. 4, C to E, and fig. S13) (*8*). As a control, we used the same approach to calculate the expression and modularity of each gene module in nonassociated cell types. We found that the modularity was decreased in nonassociated cell types (fig. S12, D and E), reflected by both the proportion of comparisons with significant correlation and by the strength of the significant correlations, suggesting that our modules reflect cell type–specific effects.

Altogether, our results suggest that expression and modularity of most gene modules in the mouse are conserved in human brain tissue, pointing at potential shared functions and further suggesting that processes identified as affected in our Perturb-Seq experiments are relevant to biological processes that may be developmentally regulated in the human brain.

Mouse Perturb-Seq results are correlated with expression changes in ASD patient brain tissues

Last, we explored whether the effects observed in mouse Perturb-Seq may be similar to changes observed in post mortem brains of ASD patients. To this end, we compared our data with a snRNA-seq dataset of post mortem ASD brain

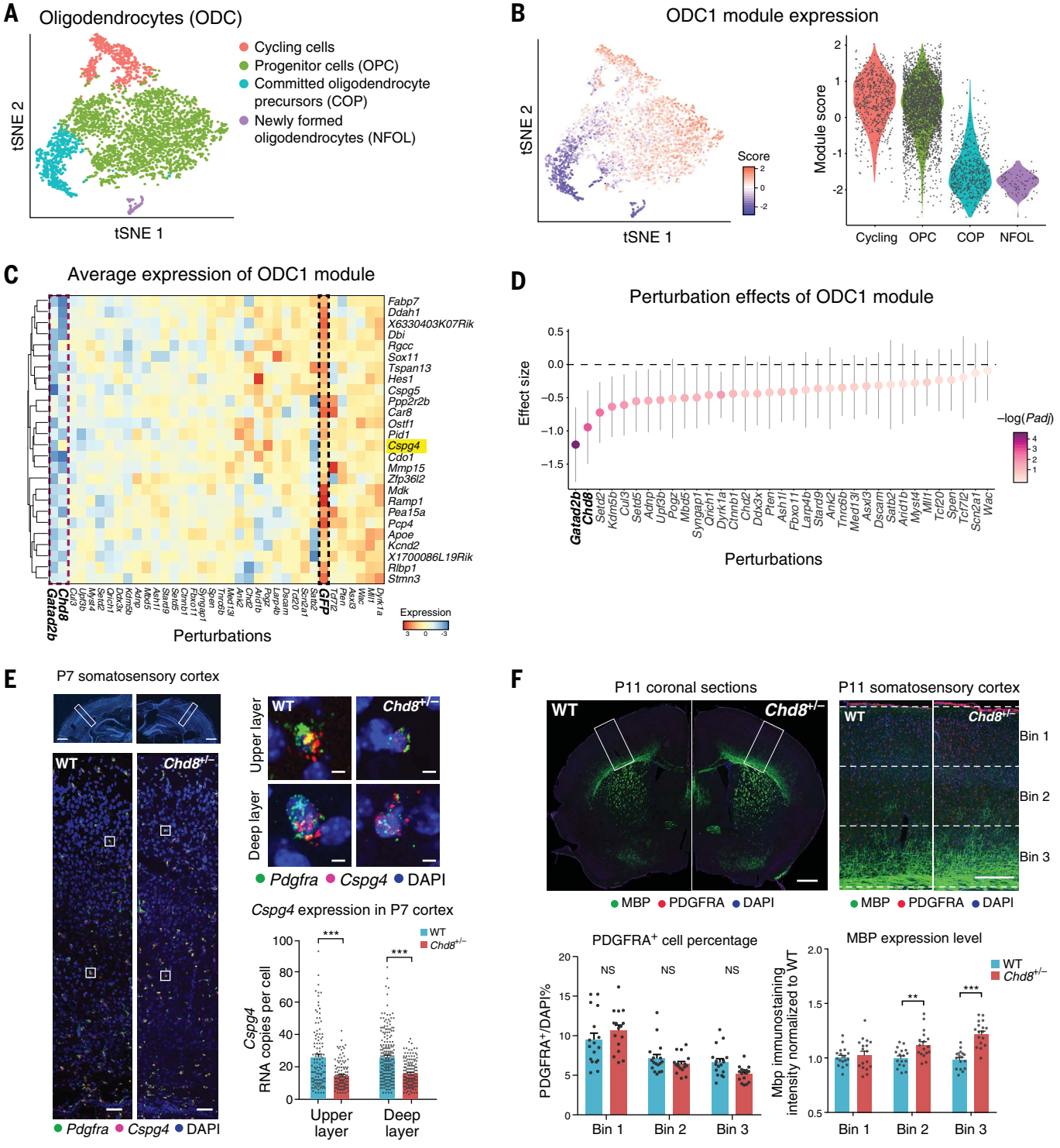
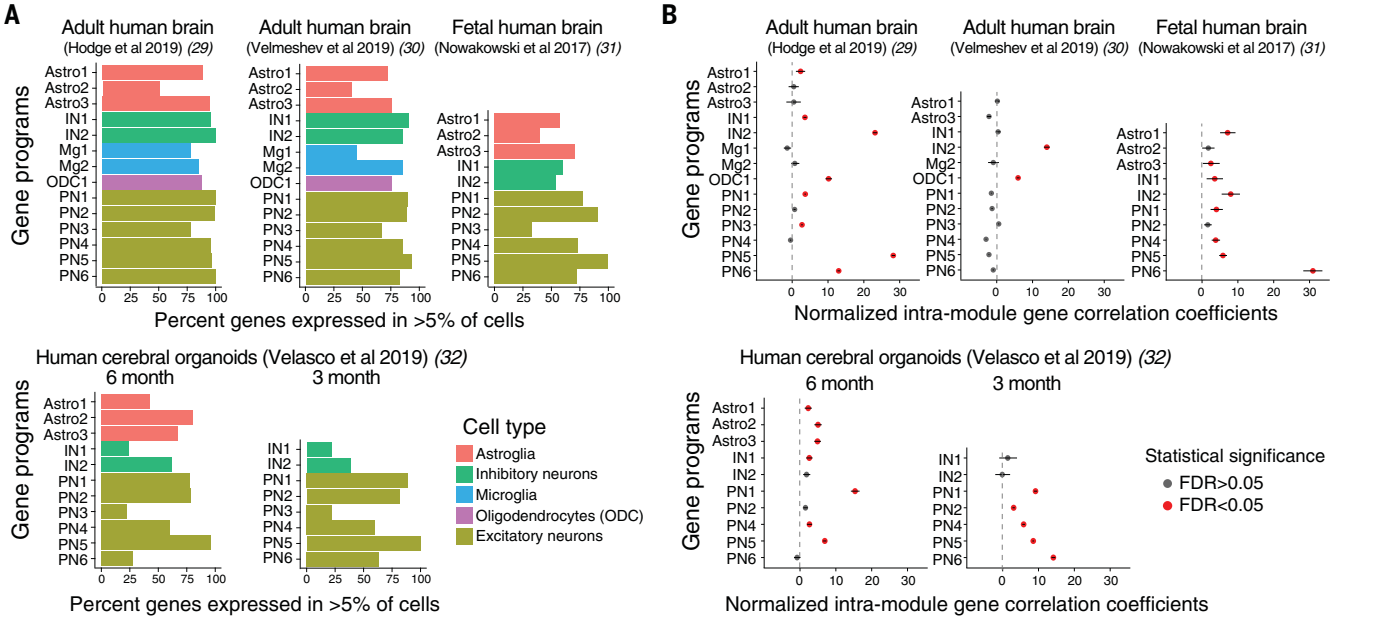
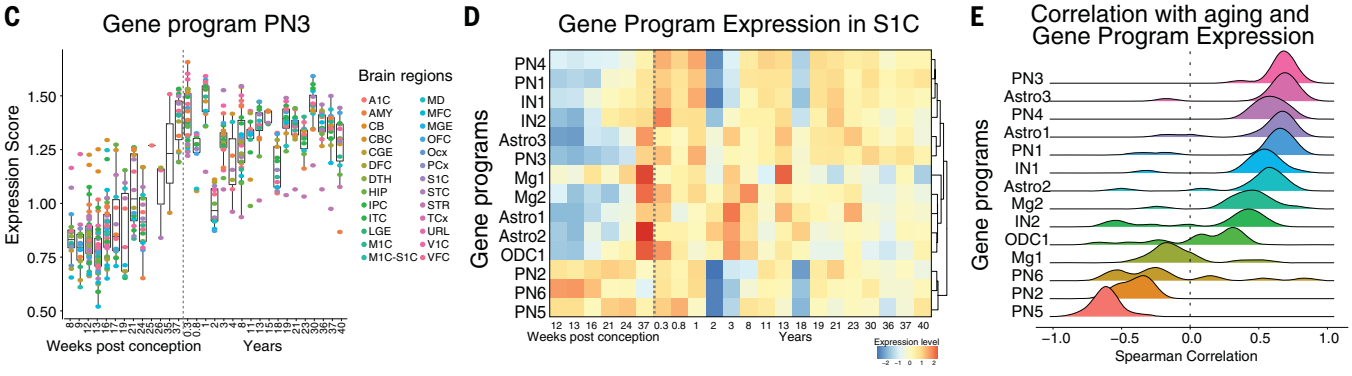


Fig. 3. Perturbation effect in oligodendrocytes and validation in the *Chd8*$^{+/-}$ mouse model. (**A**) *t*-SNE of oligodendrocyte subtypes from the Perturb-Seq data. (**B**) The ODC1 gene module expression score in (left) each celland (right) in each subcluster. (**C**) Average expression of genes in the ODC1 gene module (by row) in each perturbation group (by column), scaled by row. (**D**) Effect size of each perturbation on the ODC1 gene module compared with the control group. The perturbation effects of the different genes present a continuous gradient. Error bars represent 95% confidence intervals. (**E**) In situ hybridization for *Cspg4*, a gene in module ODC1 that is a known marker of oligodendrocyte precursor cells (OPC), in the somatosensory cortex of P7 *Chd8*$^{+/-}$ and wild-type littermates. The bottom images are higher magnifications of top images, and the right images are higher magnifications for each cell. (Right) Quantification of *Cspg4* expression in P7 cortex of *Chd8*$^{+/-}$ and wild-type littermates. Each dot represents the gene expression value from one cell; error bars represent standard error of the mean (n = 3 animals per genotype). Scale bars, 1000 μm (bottom left), 100 μm (top left), and 10 μm (right), respectively. (**F**) Immunohistochemistry for PDGFRA and MBP (markers for immature OPC and mature oligodendrocytes, respectively), PDGFRA^{+} cell counts, and distribution of MBP expression, in the somatosensory cortex of P11 *Chd8*$^{+/-}$ animals and wild-type littermates. Scale bars, 1000 μm (left) and 250 μm (right), respectively.





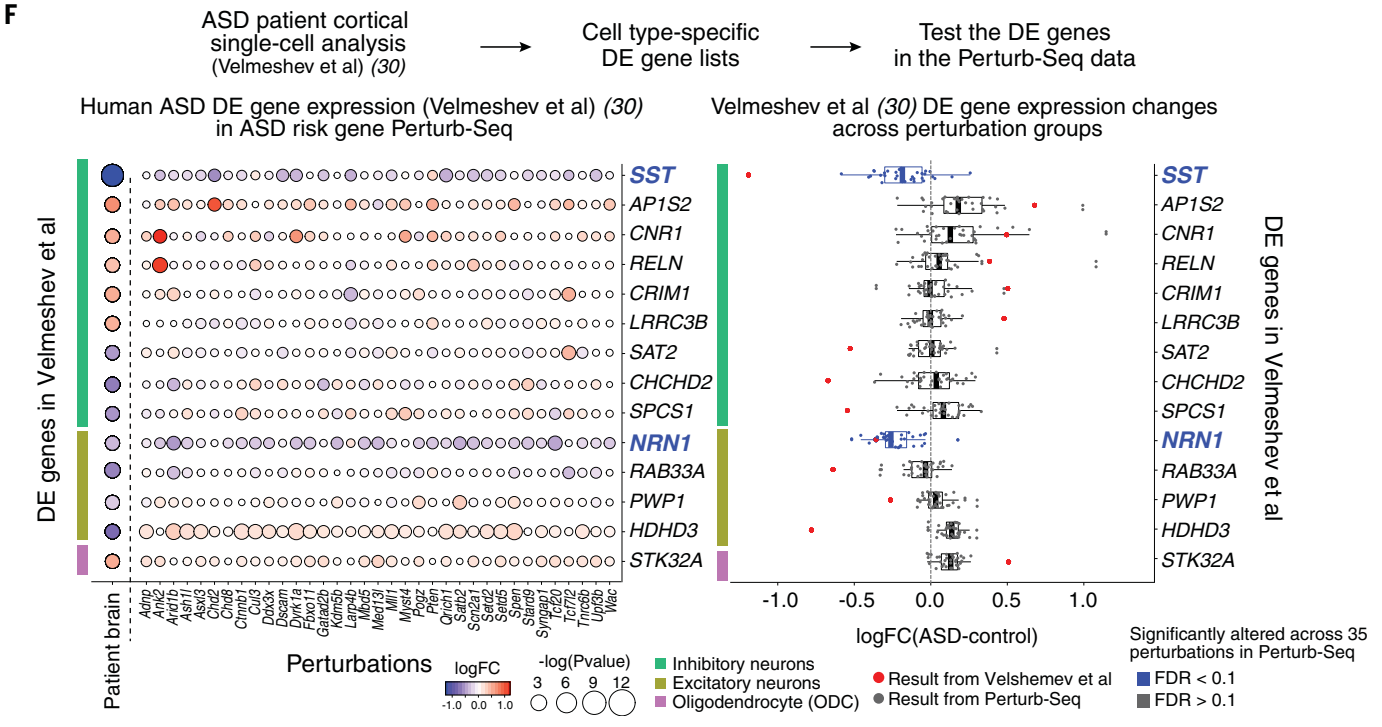


Fig. 4. Cell type–specific gene modules from Perturb-Seq are conserved in developing human brains. (**A**) Percent of genes with a human ortholog expressed in >5% of cells of the associated cell type in scRNA-seq datasets from the human brain or human brain organoids. (**B**) Normalized average pairwise correlation of gene expression within each gene module in the human brain or human brain organoids. Correlation values were normalized to the mean correlation from the background distribution and divided by the standard deviation of the background distribution. Correlations are shown for modules with at least four genes after filtering out genes expressed in <5% of cells. Bars represent 95% confidence intervals. Red color represents statistical significance (FDR < 0.05). (**C**) Expression of module PN3 over developmental time in human brain tissues across regions (BrainSpan data). (**D**) Expression of each module over developmental time in human primary somatosensory cortex S1C (BrainSpan). (**E**) Distribution of the Spearman correlation of module expression with age in human brain data over various brain regions (BrainSpan). (**F**) Differential gene expression analysis of human prefrontal cortical samples from ASD donors and controls. (Left) Expression of DE genes across cell types (color bars) from Velmeshev *et al*. (*30*) (rows) in the Perturb-Seq data across a panel of ASD/ND risk genes (columns). (Right) DE gene expression changes in Perturb-Seq data (black dots; each dot indicates an ASD/ND risk gene perturbation) compared with DE values for the 14 genes found to be DE in ASD patients in the Velmeshev *et al*. dataset (*30*) (FDR < 0.2) (red dots). The two highlighted genes, *SST* and *NRN1*, showed decreased expression in the Perturb-Seq data (FDR < 0.1), which is consistent with the ASD patient dataset.

samples (*30*) and bulk RNA-seq of post mortem psychiatric disorder brain samples from the PsychEncode project (*33*).

Using a dataset of snRNA-seq profiles from 15 ASD donors and 16 controls (*30*), we defined differentially expressed (DE) genes in each cell type using a statistically conservative pseudobulk-based analysis with DESeq2 (*34*, *35*), correcting for age, sex, and patient-to-patient variability. We identified genes that were differentially expressed between patients and controls in at least one of three major cell types (inhibitory neurons, excitatory neurons, or oligodendrocytes) with FDR < 0.2 and selected those that have 1:1 orthologs in mice, resulting in 14 genes (Fig. 4F and table S10).

We then compared these 14 genes with our Perturb-Seq data and asked whether these ASD-patient DE genes were also affected by the 35 ASD risk gene perturbations in our dataset. We aggregated the effects of all 35 perturbations and asked whether the aggregated gene expression changes agreed more strongly with the gene expression changes in the ASD patient data than would be expected by chance. For each ASD patient DE gene, we took its mouse ortholog and calculated the median fold change of expression (logFC) over all perturbations in the Perturb-Seq data. We then compared this logFC with the corresponding logFC in the ASD patient data and generated an agreement score for each gene, defined as a high median logFC and a similar direction of change as in the human data. We binned genes by their expression and compared each ASD patient DE gene with others in the same bin to extract *P* values (with FDR correction). From this analysis, we identified two genes, *SST* in interneurons and *NRN1* in excitatory neurons, both of which showed decreased expression in ASD patients and were likewise significantly decreased in expression across our panel of perturbations (FDR < 0.1), albeit with different effect sizes (Fig. 4F and table S10). This indicates that despite the different developmental stages, high clinical heterogeneity in ASD, and patient genetic diversity, genes and cell types can be identified as affected in both our analyses and in studies of human patient tissue.

We also analyzed the 14 gene modules reported in the PsychEncode study of 700 bulk RNA-seq samples of human cortex from a panel of psychiatric disorders (*33*). Of the 14 modules previously reported to be altered in the ASD patients in the PsychEncode analysis, six were also significantly affected across eight of our ASD/ND risk gene perturbations (fig. S14). Although these analyses are limited by the relatively few available datasets of ASD patient brain samples, they suggest that our Perturb-Seq experiments can identify gene program abnormalities seen in human ASD patients.

Discussion

In vivo Perturb-Seq can serve as a scalable tool for systems genetic studies of large gene panels to reveal their cell-intrinsic functions at single-cell resolution in complex tissues. In this work, we demonstrated the application of in vivo Perturb-Seq to ASD/ND risk genes in the developing brain; this method can be applied across diverse diseases and tissues.

ASD/ND affects brain function profoundly, but its cellular and molecular substrates are not yet defined. The large number of highly penetrant de novo risk genes implicated through human genetic studies offers an entry point to identify the cell types, developmental events, and mechanisms underlying ASD/ND. However, this requires scalable methods to define the function of risk-associated genes with cell-type specificity. Using Perturb-Seq to functionally test large gene sets in the developing embryo, we observed gene expression changes linked to ASD/ND genes in different cell types and processes. Within the power of the analysis that can be achieved with the number of cells that can be reasonably sequenced, we found that some recurrent modules are affected across more than one ASD/ND risk gene perturbation. It is likely that this represents an underestimation of the number of convergent modules across perturbations that might be revealed by larger-scale experiments using greater numbers of cells.

We were particularly interested in validating the observed effects of *Ank2* perturbation because of its known roles in the brain. *Ank2* encodes an ankyrin protein and is expressed broadly in excitatory and inhibitory neurons as well as glial cells in the brain (*22*). Ankyrin homologs interact with ion channels in many neuronal types, and *Ankyrin-G* has been shown to stabilize GABAergic synapses (*36*). The roles of *Ank2* in the brain have largely been studied in the context of excitatory neurons. *Ank2* loss of function results in hypoplasia of the corpus callosum and pyramidal tract, and optic nerve degeneration (*23*)—suggesting that it is required in the maintenance of premyelinated axons in excitatory neurons in early neurodevelopment. *Ank2* mutants showed misregulation of intracellular calcium homeostasis and calcium channel expression in excitatory neurons (*24*, *25*), as well as increased axonal branching and ectopic connectivity (*26*). Our Perturb-Seq data suggests an additional role of *Ank2* in interneurons expressing the *Ndnf* gene along with its known roles in excitatory neurons.

In addition to neurons, oligodendrocytes and astrocytes were also affected by several perturbations. Oligodendrocytes modulate and consolidate neural circuit refinement, and abnormal maturation of oligodendrocytes may be linked to long-lasting changes in neural wiring and brain function (*37*). One of the risk genes, *Chd8*, encodes a protein that binds directly to β-catenin to recruit histone proteins and negatively regulates the Wnt signaling pathway, a critical regulator of neural progenitor proliferation and differentiation in the forebrain (*38*–*41*). Our results showed that *Chd8* modulates gene modules for oligodendrocyte differentiation and maturation, which is consistent with previously reported chromatin immunoprecipitation–sequencing (ChIP-Seq) results showing that CHD8 interacts directly with OPC maturation genes at perinatal stages of development (*27*, *42*).

Although we focused on the perinatal neocortex in this study, in vivo Perturb-Seq can be applied to study gene functions systematically across other tissues and developmental ages, to reveal tissue-specific as well as broadly distributed gene functions. This approach can uncover both the impact of individual

disease-associated genes and of combinations of genes and the overall set of processes that they affect. Our findings underscore the importance of using single-cell profiles as a rich, comprehensive, and interpretable phenotypic readout. With advances in other single-cell profiling approaches [such as single-cell assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq) (*43*), single-cell multi-omics (*44*), and spatial genomics (*45*, *46*)], we expect in vivo Perturb-Seq to be coupled in the near future with diverse readouts to better define the function of disease risk–associated variants, from molecular mechanisms to non–cell autonomous effects in tissues. Spatial transcriptomics in particular should be well suited for in vivo Perturb-Seq and should help uncover non–cell autonomous effects of perturbations. In vivo Perturb-Seq can enable discoveries of pathways and cell types affected in heterogeneous genetic pathologies, directing downstream studies and informing the development of refined models for genetic disorders as we move from genetic variants to function.

Methods summary

In vivo Perturb-Seq experiment

The backbone plasmid contains antiparallel cassettes of two gRNAs (table S5) under mouse U6 and human U6 promoters and the EF1a promoter to express puromycin, BFP, and a polyadenylated barcode specific to each perturbation. Cloning and lentiviral packaging of the 38 vectors were done individually.

All animal experiments were performed according to protocols approved by the Institutional Animal Care and Use Committees (IACUC) of Harvard University and of the Broad Institute of MIT and Harvard. In utero lentiviral injection into the lateral ventricles was performed at E12.5 in Cas9 transgenic mice (*13*) (4 to 6 months old, Jax 026179), and each single-cell library was made by combining the BFP^+ cells from one to three litters (4 to 20 animals) of P7 animals harvested on the same day. Tissue dissociation was performed with the Papain Dissociation kit (Worthington, LK003152). The FACS-purified cells were sorted into cold Hibernate A/B27 medium and subjected to single-cell RNA sequencing library preparation. Our analysis comprises 17 independent libraries of Perturb-Seq cells.

Single-cell RNA-seq libraries were created by using the Chromium Single Cell 3′ Solution v2 kit (10x Genomics) following the manufacturer's protocol. Each library was sequenced with Illumina NextSeq high-output 75-cycle kit with sequencing saturation above 70%. Dial-out polymerase chain reaction (PCR) was performed to extract the perturbation barcode in each cell.

We identified perturbation barcodes by use of two complementary methods. We first used the dial-out sequences to create a cell-by-perturbation UMI (unique molecular identifier) count matrix by means of a modification of from the original Perturb-Seq work (*11*). In addition, we extracted barcode sequences from the 10x Genomics Cell Ranger bam file. Reads were then assigned to the perturbation to which they mapped best. Cell barcodes and UMIs were extracted, and a cell-by-perturbation UMI count matrix was created. We then only kept cells for which either (i) the assigned 10x and dialout perturbations agreed or (ii) the cell was assigned to a perturbation by one method but not assigned to any perturbation in the other.

Perturb-Seq analysis

UMI count data was loaded into R and processed by using the Seurat v 2.2 package (*47*). Clusters were assigned to cell types on the basis of marker genes from the literature, https://mousebrain.org (*15*), and https://DropViz.org (*22*). We focused only on cells of five key types (projection neurons, inhibitory neurons, oligodendrocytes, microglia and macrophages, and astroglia) and removed the rest.

WGCNA and STM were performed for each cell cluster according to the published pipelines (*20*, *21*). Linear regression was used to test the relationship between perturbations and WGCNA gene scores, correcting for batch and number of genes.

RNA in situ hybridization and immunohistochemistry

Multiplexed RNAscope fluorescent in situ hybridization and immunohistochemistry was performed on fixed-frozen tissue. Probes against the following mRNAs were used: *Pdgfra*, *Cspg4*, and *Fezf2* (ACDBio). The antibodies and dilutions were Mouse anti-NeuN antibody (mab377, 1:500; Millipore), Mouse anti-GS antibody (mab302, 1:500; Millipore), Goat anti-Pdgfra antibody (AF1062, 1:200; R&D System), Rabbit Iba1 antibody (019-19741, 1:400; Wako), Chicken anti-GFP antibody (ab16901, 1:500; Millipore), Mouse anti-Satb2 (ab51502, 1:50; Abcam), Rat anti-Ctip2 (ab18465, 1:100, Abcam), Rabbit anti-Sox6 (ab30455, 1:500; Abcam), and Rat anti-Mbp (mab386, 1:100; Millipore). We double-blinded the staining, imaging, and quantifications.

Analysis of human snRNA-seq or scRNA-seq data

For each single cell and nucleus human dataset, the UMI count matrix and metadata were downloaded and processed with Seurat to create Seurat objects. Cell types were extracted from the metadata and combined into more general cell types, namely Microglia, Astroglia (including Radial Glia), Inhibitory neurons, Excitatory neurons, Oligodendrocytes, and other. For differential expression analysis for data from Velmeshev *et al.* (*30*), we removed data from all individuals of <12 years of age and separated prefrontal cortex (PFC) and anterior cingulate cortex (ACC) regions. For each cell type in each region, a pseudobulk profile was constructed, and genes expressed in <5% of cells or with <10 reads were removed. DESeq2 v 1.20.0 (*34*) was then used to perform differential expression analysis between the ASD patients and the controls, correcting for sex and age. We then extracted all genes with 1:1 mouse orthologs (BioMart) and calculated FDR-corrected P values on these genes for both the ACC and PFC. Only analysis on the PFC yielded significant hits, which are presented in Fig. 4F.

To compare these results with the Perturb-Seq data, for each human DE gene, an agreement score was calculated by taking the absolute value of its mouse orthologs' median logFC over all perturbations (calculated with Limma) and giving it a positive sign if its direction agreed with that of the human data, and a negative sign otherwise. Last, genes were binned by expression, and P values were calculated for each gene by comparing the agreement scores with other genes in the same bin.

Detailed procedures for the experiments and data analyses are described in the supplementary materials.

REFERENCE AND NOTES

1. L. de la Torre-Ubieta, H. Won, J. L. Stein, D. H. Geschwind, Advancing the understanding of autism disease mechanisms through genetics. *Nat. Med.* **22**, 345–361 (2016). doi: 10.1038/nm.4071; pmid: 27050589
2. Schizophrenia Working Group of the Psychiatric Genomics Consortium, Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014). doi: 10.1038/nature13595; pmid: 25056061
3. L. Jostins *et al.*, Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119–124 (2012). doi: 10.1038/nature11582; pmid: 23128233
4. F. K. Satterstrom *et al.*, Novel genes for autism implicate both excitatory and inhibitory cell lineages in risk. *Cell* **180**, P568 (2018). doi: 10.1016/j.cell.2019.12.036
5. S. J. Sanders *et al.*, De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 237–241 (2012). doi: 10.1038/nature10945; pmid: 22495306
6. J. A. Chen, O. Peñagarikano, T. G. Belgard, V. Swarup, D. H. Geschwind, The emerging picture of autism spectrum disorder: Genetics and pathology. *Annu. Rev. Pathol.* **10**, 111–144 (2015). doi: 10.1146/annurev-pathol-012414-040405; pmid: 25621659
7. C. Mullins, G. Fishell, R. W. Tsien, Unifying views of autism spectrum disorders: A consideration of autoregulatory feedback loops. *Neuron* **89**, 1131–1156 (2016). doi: 10.1016/j.neuron.2016.02.017; pmid: 26985722
8. J. A. Miller *et al.*, Transcriptional landscape of the prenatal human brain. *Nature* **508**, 199–206 (2014). doi: 10.1038/nature13185; pmid: 24695229
9. Data of 9k brain cells from an E18 mouse, 10x Genomics; https://support.10xgenomics.com/single-cell-gene-expression/datasets/2.1.0/neuron_9k.
10. B. Adamson *et al.*, A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. *Cell* **167**, 1867–1882.e21 (2016). doi: 10.1016/j.cell.2016.11.048; pmid: 27984733
11. A. Dixit *et al.*, Perturb-Seq: Dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* **167**, 1853–1866.e17 (2016). doi: 10.1016/j.cell.2016.11.038; pmid: 27984732

12. D. A. Jaitin *et al*., Dissecting immune circuits by linking CRISPR-pooled screens with single-cell RNA-seq. *Cell* **167**, 1883–1896.e15 (2016). doi: 10.1016/j.cell.2016.11.039 pmid: 27984734
13. R. J. Platt *et al*., CRISPR-Cas9 knockin mice for genome editing and cancer modeling. *Cell* **159**, 440–455 (2014). doi: 10.1016/j.cell.2014.09.014; pmid: 25263330
14. V. D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks. *J. Stat. Mech.* **2008**, P10008 (2008). doi: 10.1088/1742-5468/2008/10/P10008
15. A. Zeisel *et al*., Molecular Architecture of the Mouse Nervous System. *Cell* **174**, 999–1014.e22 (2018). doi: 10.1016/j.cell.2018.06.021; pmid: 30096314
16. S. Mancinelli, S. Lodato, Decoding neuronal diversity in the developing cerebral cortex: From single cells to functional networks. *Curr. Opin. Neurobiol.* **53**, 146–155 (2018). doi: 10.1016/j.conb.2018.08.001; pmid: 30165269
17. A. L. Haber *et al*., A single-cell survey of the small intestinal epithelium. *Nature* **551**, 333–339 (2017). doi: 10.1038/nature24489; pmid: 29144463
18. B. Duan *et al*., Model-based understanding of single-cell CRISPR screening. *Nat. Commun.* **10**, 2233 (2019). doi: 10.1038/s41467-019-10216-x; pmid: 31110232
19. The Gene Ontology Consortium, The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **47** (D1), D330–D338 (2019). doi: 10.1093/nar/gky1055; pmid: 30395331
20. M. Roberts, B. Stewart, D. Tingley, stm: R package for structural topic models. *J. Stat. Softw.* **91**, 1–40 (2019). doi: 10.18637/jss.v091.i02
21. P. Langfelder, S. Horvath, WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008). doi: 10.1186/1471-2105-9-559; pmid: 19114008
22. A. Saunders *et al*., Molecular diversity and specializations among the cells of the adult mouse brain. *Cell* **174**, 1015–1030.e16 (2018). doi: 10.1016/j.cell.2018.07.028; pmid: 30096299
23. P. Scotland, D. Zhou, H. Benveniste, V. Bennett, Nervous system defects of AnkyrinB$^{-/-}$ mice suggest functional overlap between the cell adhesion molecule L1 and 440-kD AnkyrinB in premyelinated axons. *J. Cell Biol.* **143**, 1305–1315 (1998). doi: 10.1083/jcb.143.5.1305; pmid: 9832558
24. S. Tuvia, M. Buhusi, L. Davis, M. Reedy, V. Bennett, Ankyrin-B is required for intracellular sorting of structurally diverse Ca^{2+} homeostasis proteins. *J. Cell Biol.* **147**, 995–1008 (1999). doi: 10.1083/jcb.147.5.995; pmid: 10579720
25. C. F. Kline, J. Scott, J. Curran, T. J. Hund, P. J. Mohler, Ankyrin-B regulates Cav2.1 and Cav2.2 channel expression and targeting. *J. Biol. Chem.* **289**, 5285–5295 (2014). doi: 10.1074/jbc.M113.523639; pmid: 24394417
26. R. Yang *et al*., *ANK2* autism mutation targeting giant ankyrin-B promotes axon branching and ectopic connectivity. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 15262–15271 (2019). doi: 10.1073/pnas.1904348116; pmid: 31285321
27. C. Marie *et al*., Oligodendrocyte precursor survival and differentiation requires chromatin remodeling by Chd7 and Chd8. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E8246–E8255 (2018). doi: 10.1073/pnas.1802620115; pmid: 30108144
28. M. Nishiyama *et al*., Early embryonic death in mice lacking the beta-catenin-binding protein Duplin. *Mol. Cell. Biol.* **24**, 8386–8394 (2004). doi: 10.1128/MCB.24.19.8386-8394.2004; pmid: 15367660
29. R. D. Hodge *et al*., Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61–68 (2019). doi: 10.1038/s41586-019-1506-7; pmid: 31435019
30. D. Velmeshev *et al*., Single-cell genomics identifies cell type-specific molecular changes in autism. *Science* **364**, 685–689 (2019). doi: 10.1126/science.aav8130; pmid: 31097668
31. T. J. Nowakowski *et al*., Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science* **358**, 1318–1323 (2017). doi: 10.1126/science.aap8809; pmid: 29217575
32. S. Velasco *et al*., Individual brain organoids reproducibly form cell diversity of the human cerebral cortex. *Nature* **570**, 523–527 (2019). doi: 10.1038/s41586-019-1289-x; pmid: 31168097
33. M. J. Gandal *et al*., Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science* **359**, 693–697 (2018). doi: 10.1126/science.aad6469; pmid: 29439242
34. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014). doi: 10.1186/s13059-014-0550-8; pmid: 25516281
35. A. T. L. Lun, J. C. Marioni, Overcoming confounding plate effects in differential expression analyses of single-cell RNA-seq data. *Biostatistics* **18**, 451–464 (2017). doi: 10.1093/biostatistics/kxw055; pmid: 28334062
36. W. C. Tseng, P. M. Jenkins, M. Tanaka, R. Mooney, V. Bennett, Giant ankyrin-G stabilizes somatodendritic GABAergic synapses through opposing endocytosis of GABAA receptors. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 1214–1219 (2015). doi: 10.1073/pnas.1417989112; pmid: 25552561
37. K. K. Bercury, W. B. Macklin, Dynamics and mechanisms of CNS myelination. *Dev. Cell* **32**, 447–458 (2015). doi: 10.1016/j.devcel.2015.01.016; pmid: 25710531
38. R. J. Platt *et al*., Chd8 mutation leads to autistic-like behaviors and impaired striatal circuits. *Cell Rep.* **19**, 335–350 (2017). doi: 10.1016/j.celrep.2017.03.052; pmid: 28402856
39. Y. Katayama *et al*., CHD8 haploinsufficiency results in autistic-like phenotypes in mice. *Nature* **537**, 675–679 (2016). doi: 10.1038/nature19357; pmid: 27602517
40. O. Durak *et al*., Chd8 mediates cortical neurogenesis via transcriptional regulation of cell cycle and Wnt signaling. *Nat. Neurosci.* **19**, 1477–1488 (2016). doi: 10.1038/nn.4400; pmid: 27694995
41. I. Sakamoto *et al*., A novel beta-catenin-binding protein inhibits β-catenin-dependent Tcf activation and axis formation. *J. Biol. Chem.* **275**, 32871–32878 (2000). doi: 10.1074/jbc.M004089200; pmid: 10921920
42. C. Zhao *et al*., Dual requirement of CHD8 for chromatin landscape establishment and histone methyltransferase recruitment to promote CNS myelination and repair. *Dev. Cell* **45**, 753–768.e8 (2018). doi: 10.1016/j.devcel.2018.05.022; pmid: 29920279
43. A. J. Rubin *et al*., Coupled single-cell CRISPR screening and epigenomic profiling reveals causal gene regulatory networks. *Cell* **176**, 361–376.e17 (2019). doi: 10.1016/j.cell.2018.11.022; pmid: 30580963
44. S. Bian *et al*., Single-cell multiomics sequencing and analyses of human colorectal cancer. *Science* **362**, 1060–1063 (2018). doi: 10.1126/science.aao3791; pmid: 30498128
45. S. G. Rodriques *et al*., Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019). doi: 10.1126/science.aaw1219; pmid: 30923225
46. S. Simmons, seanken/ivPerturbSeq: ForPub. Zenodo (2020). doi: 10.5281/zenodo.4019534

ACKNOWLEDGMENTS

We thank C. Dulac, W. Gilbert, M. Meselson, and C. I. Bargmann for critical reading of our manuscript; L. Gaffney, A. Hupalowska, R. Macrae, J. Brown, S. Smith, as well as members of the Levin laboratory, Regev laboratory, Zhang laboratory, and Arlotta laboratory for technical and intellectual support. **Funding:** This work is supported by the Stanley Center for Psychiatric Research at the Broad Institute, NIH grants (U01MH115727, R01MH096066, and P50MH094271 Conte Center to P.A.); NARSAD Young Investigator Grant from the Brain and Behavior Research Foundation and Harvard William F. Milton Grant (to X.J.); Stanley Center for Psychiatric Research at the Broad Institute (to J.Z.L.); The Klarman Cell Observatory, Howard Hughes Medical Insititue (HHMI) and an NHGRI Center for Cell Circuits CEGS grant (to A.R.); and NIH grants (1R01-HG009761, 1R01-MH110049, and 1DP1-HL141201), HHMI, New York Stem Cell and Mathers Foundations, the Poitras Center for Affective Disorders Research at MIT, the Hock E. Tan and K. Lisa Yang Center for Autism Research at MIT, and J. and P. Poitras (to F.Z.). F.Z. is a New York Stem Cell Foundation–Robertson Investigator. **Author contributions:** X.J., A.R., F.Z., and P.A. conceived the project; X.J., A.G., and M.K. performed the experiments; X.J., S.K.S., and G.D. analyzed the data; A.S.S., V.J., and S.L. helped with cell subtype identification; L.N. assisted with sequencing library preparations; E.R. contributed human genetics interpretations; P.O. and N.C. assisted with molecular cloning; X.J., S.K.S., J.Z.L., A.R., F.Z., and P.A. wrote and revised the manuscript. **Competing interests:** P.A. is a SAB member in System 1 Biosciences and Foresite Labs and is a cofounder of FL60. A.R. is a cofounder of and equity holder in Celsius Therapeutics, Equity holder in Immunitas. Until 31 July 2020, A.R. was an SAB member of ThermoFisher Scientific, Syros Pharmaceuticlas, Asimov, and Neogene Therapeutics. From August 2020, A.R. is an employee of Genetech. F.Z. is a cofounder of Editas Medicine, Beam Therapeutics, Pairwise Plants, Arbor Biotechnologies, and Sherlock Biosciences. X.J., S.K.S., A.R., F.Z., and P.A. are co-inventors on in vivo Perturb-Seq and CRISPR inventions filed by the Broad Institute relating to the work in this manuscript. **Data and materials availability:** Data generated for this study are available through the Gene Expression Omnibus (accession no. GSE157977) as well as the Broad single cell portal (https://singlecell.broadinstitute.org/single_cell/study/SCP1184). The analysis pipeline is deposited on GitHub repository (https://github.com/klarman-cell-observatory/ivPerturbSeq) and Zenodo (*46*). All other data are available in the manuscript or the supplementary materials. Material used in this study will be available through Addgene under the Uniform Biological Material Transfer Agreement.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6520/1057/suppl/DC1
Materials and Methods
Figs. S1 to S14
References (*47*–*54*)
Tables S1 to S10
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

26 September 2019; resubmitted 24 May 2020
Accepted 9 October 2020
10.1126/science.aaz6063

In vivo Perturb-Seq reveals neuronal and glial abnormalities associated with autism risk genes

Xin Jin, Sean K. Simmons, Amy Guo, Ashwin S. Shetty, Michelle Ko, Lan Nguyen, Vahbiz Jokhi, Elise Robinson, Paul Oyler, Nathan Curry, Giulio Deangeli, Simona Lodato, Joshua Z. Levin, Aviv Regev, Feng Zhang and Paola Arlotta

Science **370** (6520), eaaz6063.
DOI: 10.1126/science.aaz6063

An in vivo analysis of autism risk genes

CRISPR targeting in vivo, especially in mammals, can be difficult and time consuming when attempting to determine the effects of a single gene. However, such studies may be required to identify pathological gene variants with effects in specific cells along a developmental trajectory. To study the function of genes implicated in autism spectrum disorders (ASDs), Jin *et al.* applied a gene-editing and single-cell–sequencing system, Perturb-Seq, to knock out 35 ASD candidate genes in multiple mice embryos (see the Perspective by Treutlein and Camp). This method identified networks of gene expression in neuronal and glial cells that suggest new functions in ASD-related genes.

Science, this issue p. eaaz6063; see also p. 1038