



Published in final edited form as:

*Cell Stem Cell*. 2023 March 02; 30(3): 312–332.e13. doi:10.1016/j.stem.2023.01.010.

## Natural variation in gene expression and viral susceptibility revealed by neural progenitor cell villages

Michael F. Wells<sup>1,2,3,4,10</sup>, James Nemesh<sup>1,5,10</sup>, Sulagna Ghosh<sup>1,5,10</sup>, Jana M. Mitchell<sup>1,2,6,10</sup>, Max R. Salick<sup>6</sup>, Curtis J. Mello<sup>1,5</sup>, Daniel Meyer<sup>1,5</sup>, Olli Pietilainen<sup>1,2</sup>, Federica Piccioni<sup>7</sup>, Ellen J. Guss<sup>2</sup>, Kavya Raghunathan<sup>1,2</sup>, Matthew Tegtmeier<sup>1</sup>, Derek Hawes<sup>1</sup>, Anna Neumann<sup>1</sup>, Kathleen A. Worringer<sup>8</sup>, Daniel Ho<sup>8</sup>, Sravya Kommineni<sup>8</sup>, Karrie Chan<sup>8</sup>, Brant K. Peterson<sup>8</sup>, Joseph J. Raymond<sup>8</sup>, John T. Gold<sup>2,9</sup>, Marco T. Siekmann<sup>2</sup>, Emanuela Zuccaro<sup>2</sup>, Ralda Nehme<sup>1,2</sup>, Ajamete Kaykas<sup>6</sup>, Kevin Egan<sup>1,2,\*</sup>, Steven A. McCarroll<sup>1,5,11,\*</sup>

<sup>1</sup>Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

<sup>2</sup>Department of Stem Cell and Regenerative Biology, and Harvard Stem Cell Institute, Harvard University, Cambridge, MA 02138 USA

<sup>3</sup>Department of Human Genetics, David Geffen School of Medicine at the University of California Los Angeles, Los Angeles, CA 90095 USA

<sup>4</sup>Molecular Biology Institute, University of California Los Angeles, Los Angeles, CA 90095 USA

<sup>5</sup>Department of Genetics, Harvard Medical School, Boston, MA 02115, USA

<sup>6</sup>Insitro, South San Francisco, CA 94080, USA

<sup>7</sup>Genetic Perturbation Platform, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

<sup>8</sup>Department of Neuroscience, Novartis Institute for BioMedical Research, Cambridge, MA 02139, USA

<sup>9</sup>Department of Biology, Davidson College, Davidson, NC 28035, USA

<sup>10</sup>These authors contributed equally to this work

<sup>11</sup>Lead Contact

### SUMMARY

Human genome variation contributes to diversity in neurodevelopmental outcomes and vulnerabilities; recognizing the underlying molecular and cellular mechanisms will require scalable approaches. Here, we describe a “cell village” experimental platform we used to analyze

\*Correspondence: kevin.egan@bmrn.com (K.E.), smccarro@broadinstitute.org (S.M.).

#### AUTHOR CONTRIBUTIONS

Conceptualization and design: MFW, JN, SG, JMM, FP, AK, KE, SAM; Experimentation: MFW, JN, SG, JMM, MRS, CM, DM, EJG, KR, MT, JTG, MTS, EZ; hPSC line expansion, QC, and sequencing: MT, DH, supervised by RN. Dropulation: JN, SAM; Data analysis: MFW, JN, SG, DM, OP, DH, BKP, SAM; H1-Cas9 generation and validation: SK, KC, JJR, KAW; Writing: MFW, JN, SG, JMM, KE, SAM. Supervision: KE, SAM; Project administration: AN.

#### DECLARATION OF INTERESTS

K.E. is a founder of Q-State Biosciences, Quralis, and Enclear Therapies; an employee and shareholder of BioMarin; and a member of *Cell Stem Cell*'s advisory board.

genetic, molecular, and phenotypic heterogeneity across neural progenitor cells from 44 human donors cultured in a shared *in vitro* environment using algorithms (Dropulation and Census-seq) to assign cells and phenotypes to individual donors. Through rapid induction of human stem cell-derived neural progenitor cells, measurements of natural genetic variation, and CRISPR-Cas9 genetic perturbations, we identified a common variant that regulates antiviral *IFITM3* expression and explains most inter-individual variation in susceptibility to the Zika virus. We also detected expression QTLs corresponding to GWAS loci for brain traits and discovered novel disease-relevant regulators of progenitor proliferation and differentiation such as *CACHD1*. This approach provides scalable ways to elucidate the effects of genes and genetic variation on cellular phenotypes.

---

## INTRODUCTION

Humans harbor immense diversity in biological traits and disease risk, affecting almost all organs and physiological functions. Reservoirs of natural variation allow populations to adapt to existential crises and selective pressures, such as viral outbreaks. In the brain, variation in neurodevelopmental processes—such as the proliferation and differentiation of neural progenitor cells (NPCs)—creates variation in propensities for learning, socializing, and responding to environmental stressors; disruption of these processes can lead to autism spectrum disorder (ASD), cancer, and Congenital Zika Syndrome<sup>1,2</sup>. How genetic variation acts through molecular, cell, and developmental biology to shape trait variation and disease risk remains largely unknown.

Genetic variants can influence neural phenotypes through the regulation of gene expression, which has unknown effects on signaling pathways, cell migration, and cell-cell interactions. Abundant human genetic variation has been catalogued by the International HapMap<sup>3</sup> and 1000 Genomes projects<sup>4</sup>. Efforts such as the Genotype-Tissue Expression (GTEx)<sup>5</sup> Consortium have identified thousands of expression quantitative trait loci (eQTLs)—associations of single nucleotide polymorphisms (SNPs) to RNA expression of nearby genes—in every adult organ analyzed. *In vitro* human pluripotent stem cell (hPSC) models have also proven useful for eQTL detection through approaches that maintain cells from many human donors in separate culture environments followed by preparation of individual bulk RNA sequencing (RNA-seq) libraries<sup>6,7</sup>. This method results in considerable technical variation that can mask biologically relevant effects, and requires substantial resources, hands-on activities, and costs.

Though thousands of eQTLs have been found in numerous brain regions and cell types, we know little about how these effects percolate through cell biology to influence phenotypes. Biological pathways are robust to many kinds of perturbation and may buffer the effects of many genetic variants, even those that affect a gene's expression. Thus, it is essential to understand the relationships among genetic variants, gene expression, and the physiological phenotypes of cells. This has been difficult for neurodevelopmental phenotypes, in substantial part because NPCs are no longer present when individuals' traits are ascertained or when postnatal tissue is sampled for analysis<sup>8,9</sup>.

NPC production can be accomplished by suspending hPSCs as embryoid bodies prior to transfer to an adherent surface and manual or enzymatic selection of neural rosettes<sup>10–12</sup>, or through application of small molecule inhibitors of SMAD<sup>13</sup>. These techniques produce neural cells by mimicking embryonic events and are attractive because of their presumed developmental fidelity. However, they require 11–50 days of induction to produce stable NPC cultures<sup>14</sup> and are variably effective across donor cell lines, in which rounds of differentiation can fail outright or produce heterogeneous cell types<sup>15,16</sup>. Recently, the forced expression of the transcription factor Neurogenin-2 (NGN2) has been shown to robustly generate homogenous cultures of mature post-mitotic cortical neurons<sup>17,18</sup>. Fast and reliable NPC induction techniques are still needed for scaled modeling of the developing brain.

Here, we describe advances in two technologies that can help dissect interactions among alleles, molecules, and cellular phenotypes in human NPCs. The first is genetic multiplexing, in which thousands of cells from scores of donors are pooled in a shared *in vitro* environment—a “cell village”—and then analyzed simultaneously by single-cell RNA-seq; transcribed SNPs are used to assign individual cells to individual donors. The second is an NGN2-based scheme that requires only 48 hours for NPC induction and is effective in over 100 hPSC lines. We coordinated these approaches with “Census-seq”, a rapid, inexpensive method for relating cellular phenotypes to natural genetic variation by sequencing the genomic DNA from cell villages<sup>19</sup>. We also incorporated functional CRISPR-Cas9 screens to explore thousands of artificial genetic perturbations simultaneously<sup>20,21</sup>.

Using this experimental platform, we detected NPC eQTLs in neurodevelopmental disorder (NDD) genes and brain trait genome-wide association study (GWAS) loci. We also uncovered a SNP that explains more than 50% of inter-individual variation in NPC susceptibility to the Zika virus (ZIKV). Genome-wide CRISPR-Cas9 screens aided in the discovery of this functional QTL and revealed new regulators of NPC growth and viability that were significantly enriched for NDD genes. This includes *CACHD1*, which enhanced proliferation and disrupted differentiation in 2D and 3D neural models upon ablation. Our findings establish an integrated experimental format that uses natural and synthetic perturbations to identify genes and genetic variants that change a cell’s phenotype in a meaningful way.

## RESULTS

### Assignment of cells to individual donors using transcribed SNPs

To ascertain how natural genetic variation shapes cellular phenotypes, we sought to eliminate technical sources of variation by culturing cells from many donors in a shared environment and analyzing them together. This requires re-identifying the donor of each cell during single-cell analysis. The combination of hundreds of transcribed SNPs can identify the donor of an individual cell<sup>22</sup>. We further developed such analysis to: (i) to address challenges inherent to scRNA-seq experiments, such as ambient RNA; (ii) to utilize unique molecular identifiers (UMIs) rather than reads as the informative analytical unit; and (iii) to enable scalability up to hundreds of potential donors. We implemented a maximum

likelihood approach that calculates for each donor (utilizing pre-existing whole genome sequencing or SNP data) the likelihood that the observed single-cell level combination of transcribed alleles arose from that donor's genome sequence. We incorporated uncertainty to address sequencing error or ambient RNA. We provide the resulting software "Dropulation" (Droplet-based sequencing of populations) in an open-source format (<https://github.com/broadinstitute/Drop-seq>).

To evaluate the accuracy with which cells were assigned to donors by Dropulation, we first analyzed scRNA-seq data from five human embryonic stem cell (hESC) lines pooled *in silico* (Figure S1a–b). The analysis conclusively identified a donor for 97.6% of cells, doing so with 99.8% accuracy and distinguishing among closely related (genetic sibling) donors (Figure S1c–d). On average, individual cells contained hundreds to thousands of transcripts with sequences that varied among the donors (Figure S1e); 20–50% of all UMIs contained such sites (Figure S1f). Only 2.4% of cells—generally low-quality single-cell profiles, in which few UMIs had been ascertained—were determined to be "unassigned" due to low assignment confidence (Figure S1d). The frequency of donor mis-assignment by Dropulation was low (0.2%), suggesting that the assignment confidence was well-calibrated.

Allelic information provides powerful ways to detect cell-cell doublets<sup>22</sup>, an important challenge in scRNA-seq analyses. Dropulation detects doublets by asking whether an *in silico* mixture of two donors' genotype data (Figure S1g) generates a single cell's data with a higher likelihood than any one donor's genotype data does (Figure S1h). In *in silico* evaluations, doublet detection by Dropulation had a true positive rate of 98.3% and a false positive rate of 1.5% when donor data was mixed in a 1:1 ratio. Misclassification of singlets as doublets was rare (Figure S1i); singlets tended to be misassigned when they had fewer than 100 informative UMIs (Figure S1j). Larger numbers of UMIs were needed to recognize unequal donor mixtures, as might arise from doublets of two cells of distinct sizes. Still, detecting donor mixtures of 4:1 required only about 240 informative UMIs (Figure S1k). These data indicate that doublets are successfully recognized when they involve cells from two donors and the depth of UMI ascertainment is adequate.

We tested these algorithms for their ability to identify the 36 donors present (among 142 candidates) in a cell village; analysts were blind to the number, identity, and proportion of donor lines included in the village. Analysis successfully identified all (36/36) donors with cells present in the village without incorrectly assigning any cells to the 106 other candidates (Figure S1l).

To further test Dropulation in real-world experimental contexts, we constructed a 104-donor cell village that included whole genome sequenced human induced pluripotent stem cells (iPSCs) derived from male and female skin and blood cells (Figure 1a–b). We profiled 86,185 cells from this village by scRNA-seq five days after pooling. We sampled an average of 104,160 UMIs per cell, which was well powered to assign cells to donors (Figure S1m). We then measured the relative proportion of each donor in the village using both Dropulation and Census-seq, a low-coverage-WGS-based computational tool we developed to infer the donor composition of cell mixtures from bulk DNA<sup>19</sup>. We detected high

concordance between these single-cell RNA and bulk-DNA methods, validating both (Figure S1n).

### Analysis of biological variation in “Dropulations”

The ability to measure mRNA expression in a shared culture environment made it possible to quantify effects that have long been of great concern in hiPSC research, including effects of cell source and donor sex. The iPSCs from 104 donors exhibited highly similar RNA expression patterns (Figure 1c). The primary source of variation in donor transcriptional profiles was their progress through the cell cycle (Figure 1d). Heterogeneity in cell state or identity also impacted gene expression, though as expected, the majority of cells expressed pluripotency markers *NANOG* and *OCT4* at high levels (Figure 1e). A subset of cells expressed NPC markers indicative of spontaneous neural differentiation, while others displayed the lower UMI counts and higher percentages of nascent transcripts with intronic reads that are typical of nuclei rather than intact cells (Figure 1f, Figure S1o–p).

There are fundamental methodological questions regarding the comparability of iPSCs created from different tissues. We identified differentially expressed genes (DEGs) across all donors and found only four DEGs (all noncoding RNAs; FDR < 5%) that distinguished the 56 skin-derived from the 45 blood-derived lines (Figure 1g–h; Data S1, File 1). This finding suggests that there is modest retention of epigenetic memory inherited from the parental cell source of origin in the iPSC lines, but that few protein-coding genes are strongly affected by this memory.

Many human phenotypes exhibit sex differences. RNA expression levels of many genes differ on average between males and females in various tissues<sup>23</sup>, but the extent to which cell-autonomous biology contributes to such differences is unknown. RNA expression profiles of individual cells initially appeared to be strongly distinguished by donor sex (Figure 1i). This difference disappeared, however, when we limited analysis to autosomal genes (Figure 1j): it arose almost entirely from Y-linked genes and the X-linked genes that escape X chromosome inactivation and did not appear to involve broader effects on cells' biology. When we compared pairs of donors within and between sex to generate pairwise distributions of DEGs, we found similar numbers of autosomal DEGs in same-sex comparisons (XX vs XX; XY vs XY) as we did in across-group (XX vs. XY) comparisons (Figure 1k), indicating that on average iPSCs from XX and XY individuals were roughly as different from each other as same-sex individual pairs regardless of cell source. Consistent with earlier observations from tissue-level analysis<sup>23</sup>, sex effects were small (median log<sub>2</sub>FC upregulated genes = 0.15, downregulated genes = -0.10; Figure 1l).

Collectively, these results suggest that differences in gene expression generated by donor sex and source cell-type are small compared to the effects of inter-individual variation, and that sex differences in expression of sex-chromosome genes do not lead to broader effects on cells' biology in this context. The ability to remove culture-to-culture sources of variation and quantify sources of molecular variation in a cell village allowed these relationships to emerge clearly and enabled similar experiments in other cell types.

## Brief NGN2-mediated neuralization of hPSCs produces human dorsal telencephalic NPCs

Constructing villages of NPCs requires quick, dependable induction. Overexpression of NGN2 efficiently neuralizes hPSCs, but it was suggested that these post-mitotic neurons bypass the progenitor stage<sup>24</sup>. To determine if NGN2 induction creates progenitor-like cells, we re-analyzed published RNA-seq data from cells harvested during NGN2-directed differentiation to post-mitotic neurons<sup>17</sup>. The expression of forebrain NPC genes increased while pluripotency genes decreased over the first few days (Figure S2a). Most cells were positive for the proliferative marker MKI67 during the first two days of induction, but not at Day 3 (Figure S2b; Table S1), and two days of NGN2 overexpression yielded substantially more progeny over a week of subsequent expansion than did cells subjected to four days of overexpression (Figure S2c). Thus, NGN2 overexpression beyond 48 hours promotes cell cycle exit.

We designed an approach to create and maintain self-renewing human NPCs. SW7388-1 iPSCs were transduced with two separate lentiviruses encoding TetO::*Ngn2:T2A:PURO* and Ubq::rtTA to enable doxycycline (DOX)-inducible expression of mouse NGN2 and the linked puromycin resistance gene (Figure 2a). We initiated NGN2 induction by adding DOX and small molecule inhibitors of the SMAD (LDN-193189 and SB431542) and WNT (XAV939) signaling pathways, which dorsalizes early neural cells<sup>25–27</sup>. After 24 hours, puromycin was added to eliminate non-transduced cells. At 48 hours, these cells (hereby referred to as *Stem cell-derived NGN2-accelerated Progenitors* or “SNaPs”) showed a bipolar morphology characteristic of NPCs (Figure 2b) and expressed forebrain NPC protein markers PAX6, NESTIN, and FOXG1—which were absent in hPSCs (Figure S2d)—and neural stem cell proteins SOX1 and SOX2 (Figure 2c–d). In line with exit from pluripotency, OCT4 levels declined precipitously (Figure S2e). SNaP induction efficiency was independent of seeding density (Figure S2f), which is an improvement over SMAD inhibition methods that require specific, high hPSC confluencies for successful conversion<sup>13</sup>.

To maintain SNaPs in a proliferative, self-renewing state, we passaged into growth factor-containing media. We observed self-organized neural rosette structures and still-higher percentages of NPC marker positivity (Figure 2e–g). SNaPs displayed a mean doubling time of 41.02 hours (Figure S2g) similar to that of other proliferative NPC models<sup>11</sup>.

The hPSC-to-NPC transition is accompanied by changes in molecular landscapes. We used qPCR, bulk RNA-seq, and ChIP-seq to evaluate SNaP transcriptional signatures. SNaPs displayed a marked increase in *PAX6*, *FOXG1*, and *SOX1* levels relative to hPSCs at 48 hours post-induction while *OCT4* declined; this pattern was similar to NPCs generated using established protocols (Figure S2h–j). Bulk RNA-seq revealed upregulation of NPC markers *OTX1* and *EMX2* and downregulation of pluripotency genes *LEFTY1* and *NANOG* (Figure S2k, Data S1, File 2). SNaPs also displayed high expression of genes expressed in dorsal progenitors and low levels of genes expressed in posterior and ventral brain regions (Figure 2h–i). SNaPs therefore resemble dorsal telencephalic NPCs.

We performed ChIP-seq analysis to identify the DNA targets of overexpressed mouse NGN2 protein during SNaP differentiation to gain insight into the regulatory mechanisms



guiding neural conversion. Interaction peaks were detected 24–48 hours post-induction in or near proneural genes that are targets of mammalian NGN2, including *NEUROD1/4* and *ELAVL4*<sup>28,29</sup>, as well as NPC marker genes like *NESTIN* and *PAX6* (Figure S2i; Data S1, File 3). We observed peaks in 8,950 genes, including in the promoter and/or UTR regions of 1,300 genes. These data—along with our qPCR and immunostaining results that showed high expression of NPC markers at 48 hours post-induction—suggest that SNaP identity is driven by the direct activation of NPC gene loci and the remodeling of the chromatin landscape to favor neural differentiation.

### SNaPs self-renew and differentiate into neurons and glia

During fetal development, NPCs differentiate into the cortical excitatory neurons and astrocytes that populate the brain<sup>30–32</sup>. To assess developmental potency, we cultured SNaPs under conditions that promote differentiation into neuronal and glial lineages. Spontaneous differentiation in growth factor-free base media or 1% fetal bovine serum (FBS) resulted in a mixture of neurons (HuC/D<sup>+</sup>) and glia (CD44<sup>+</sup>/S100b<sup>+</sup>; Figure 2j–m). Astrocyte media<sup>33</sup> (AM) produced GFAP<sup>+</sup> glial cells after 60 days in culture. SNaP multipotency was confirmed by scRNA-seq that revealed 7 cell clusters, including intermediate progenitors, excitatory neurons, and glial progenitors after 2 weeks in base media (Figure S3a–b). Comparison to an integrated reference of human fetal and adult brain cells<sup>34,35</sup> showed that differentiated SNaPs resembled many fetal *in vivo* cell types, primarily NPCs, intermediate progenitors, and glutamatergic neurons (Figure S3c–d). SNaPs were capable of self-renewal—defined by the ability to divide while maintaining a multipotent state—as depicted by clonal proliferation and differentiation assays (Figure 2n–q). This feature is characteristic of the neuroepithelium, suggesting that SNaPs functionally resemble the earliest neural cell type to populate the developing brain.

Differentiated excitatory neurons form synaptic connections and display network activity *in vivo*, so we stained 30-day-old spontaneously differentiated cultures with protein markers of cortical glutamatergic neurons. Most neurons co-expressed upper cortical layer markers BRN2 and CUX2 (85.77% and 82.90%) and a few expressed deep layer marker CTIP2 (3.70%; Figure S3e). Inhibitory neuron marker GAD67 was not detected (Data not shown). Neurons co-cultured with mouse glia displayed characteristic punctate expression of the synaptic marker Synapsin I (Figure S3f) and formed functional synaptic networks over several weeks (Figure S3g–m). Like fetal NPCs, SNaPs can spawn mature neurons that organize into active synaptic networks.

### Villages confirm consistent SNaP identity across different genetic backgrounds

We assessed whether the SNaP method is amenable to cell village experiments by testing the reproducibility of induction on 46 hESC and 60 iPSC lines (Table S1). The vast majority (102/106; 96.2%) of lines were successfully converted to stable NPC cultures as defined by 75% NESTIN<sup>+</sup>/PAX6<sup>+</sup>/SOX1<sup>+</sup> and 0.1% OCT4<sup>+</sup> cells at Passage 1–3 (Figure S4a–c). To test intra-line reproducibility, we induced three hiPSC and one hESC lines in duplicate and observed high NPC marker expression in all replicates (Figure S4d). All SNaP lines assayed for multipotency (30/30) differentiated into neurons and glia (Figure S4e–f). Thus, the SNaP protocol readily and consistently generates human NPCs from many disparate donors.

We built two SNaP villages from 21 and 44 hESC donors (SNaP-21 and SNaP-44 villages respectively; Figure 3a) that passed NPC protein marker-based QC metrics. We processed 123,988 SNaPs for scRNA-seq, re-identifying donor-of-origin for each sequenced cell via Dropulation. We simultaneously compared the cellular identities of each donor's cells to a reference *in vivo* human brain dataset<sup>34,35</sup> using UMAP plots; SNaPs predominantly clustered with fetal NPCs (Figure 3b). *In vitro* and *in vivo* NPCs expressed progenitor markers *HES1* and *SOX2* but exhibited minimal expression of differentiated neuronal or astrocytic genes (Figure 3c). SNaPs displayed high *in vivo* fetal NPC similarity scores (>0.8; Figure 3d) using cell-type classification methods<sup>36</sup>. Most (84.1%) SNaP-21 village cells resembled human fetal NPCs most closely, and this degree of similarity was consistent across all lines (Figure 3e–f; Figure S4g). Comparable results were observed in the SNaP-44 village (Figure S4h–i).

The question remained, however, exactly which type of fetal NPC our SNaPs resembled. We repeated this analysis using an integrated scRNA-seq dataset that contained only fetal NPC subtypes. The example output (GENEA43) shows cell identity scores that are highest for the earliest detected fetal NPC subtype known as “RG-early” (5-8 post-conception weeks or pcw; Figure 3g). This was consistent across the SNaP-21 village, in which 90.0% of all cells most closely resembled these neuroepithelial RG-early cells and 8.0% were assigned to the “RG-div1” identity (11-21 pcw; Figure 3h–i). Thus, our method can reproducibly generate an *in vitro* cell type that is transcriptionally similar to *in vivo* neuroepithelial cells.

To assess the reproducibility of SNaP multipotency in a village, we pooled 37 hESC-derived SNaP lines and allowed for 16 days of spontaneous differentiation prior to scRNA-seq-based *in vivo* comparison (Figure S5a–b). As expected, all lines produced neural and glial cells though at highly variable ratios (Figure S5c). These findings suggest that some lines may be better than others for generating certain differentiated brain cell types, which has important implications for the experimental design of future comparative differentiation assays.

### Effects of common genetic variation on RNA expression

Cell village analysis, by neutralizing most cell-extrinsic forms of variation in cultures, facilitates the identification of natural genetic variants that affect gene expression. To map expression QTLs (eQTLs) in human NPCs, we tested all genes for associations between expression levels and nearby SNPs among the donors in the SNaP-44 village using a linear regression model<sup>37</sup> (Figure 4a). Using a 10kb window around each gene and a minor allele frequency of 0.20, we detected 24,130 nominally significant eQTLs ( $p < 1e-05$ ), including 993 genes that exhibited analysis-wide significant association to one or more SNPs (“eGenes”,  $q$ -value  $< 0.05$ ). We validated our findings through comparisons to eQTLs detected by bulk RNA-seq of human brain samples and found high concordance (>90%) with prenatal and adult brain eQTLs<sup>38,39</sup> by sign test, which asks if the direction of association for variant-gene pairings is the same in both datasets (Figure 4b). The list of eGenes includes neurodevelopmental disorder candidate genes *MAPK3* and *DMPK*, as well as the fetal brain transcriptional activator *CHURC1* (Figure 4c). Filtering out cells that were classified as non-NPCs had minimal effect on the number and identity of eGenes detected



(Data S1, File 4). Cell village eQTL analysis therefore uncovered genetic influencers of transcript levels in human NPCs that also affect neurodevelopmental phenotypes *in vivo*.

Several of the eQTLs we discovered in SNaPs corresponded to previously discovered genetic effects on cognitive traits and brain disorders. For example, genotype at the SNP rs79600142 associated with expression of the readthrough transcript *LINC02210-CRHR1*—which encodes a protein that shares sequence identity with corticotropin releasing hormone receptor 1—in SNaPs (Figure 4d–e); this SNP was the top hit from a genome-wide association study (GWAS) for human cortical surface area<sup>40</sup>. The schizophrenia risk SNP rs4523957<sup>41</sup> associates in our analysis with expression of the *SRR* gene encoding serine racemase, which converts L-serine to D-serine to modulate NMDA receptors<sup>42</sup> (Figure 4f). We also identified eQTLs for which the same SNPs have been found to associate with cognitive performance (rs1131017), Parkinson’s disease (rs2942168), and depression (rs61990288; Data S1, File 4). The fact that cell village analysis even at this modest sample size detected influential variants in important neurodevelopmental genes suggests that larger cell villages may illuminate more disease-influencing biology in NPCs.

### Differential susceptibility to Zika virus infection across donor lines

Using SNaPs, CRISPR screens, and cell villages, we sought to identify genetic effects on a specific cellular phenotype—the vulnerability of human NPCs to the neurotropic ZIKV that was made famous by recent outbreaks in the Americas<sup>43</sup>. Prenatal exposure to this mosquito-borne virus can result in Congenital Zika Syndrome (CZS), which is characterized by microcephaly and developmental delay<sup>44–46</sup>. Interestingly, only 30% of prenatal infections resulted in CZS<sup>47,48</sup>. While several factors could potentially explain this observation<sup>49,50</sup>, human genetic diversity may significantly contribute to differential responses to this pathogen<sup>51–54</sup>.

ZIKV preferentially targets proliferating NPCs<sup>2,55</sup>. We verified that SNaPs model ZIKV neuropathogenesis by immunostaining for ZIKV envelope protein (4G2) at 54 hours post-infection (hpi) with the original MR-766 Uganda (ZIKV-Ug) and modern Puerto Rico (ZIKV-PR) isolates (Figure 5a–b). ZIKV-Ug killed more cells than ZIKV-PR (Figure S6a), consistent with previous reports<sup>56,57</sup>. ZIKV RNA levels increased over time in the culture media of infected SNaPs, suggesting multiplication of infectious particles (Figure S6b–c). The expression of several immunity-related genes in the host cells was increased at 60 hpi, and there was significant overlap between our results and a study that used the embryoid body induction method to generate NPCs<sup>58</sup> (Figure S6d, Data S1, File 5). Thus, SNaPs model the expected transcriptional and functional responses to ZIKV.

To quantify inherent differences in ZIKV susceptibility across donors, we infected 24 hESC-derived SNaP lines cultured in an arrayed format with ZIKV-Ug (multiplicity of infection or MOI = 1; Figure 5c). We observed a surprisingly high degree of variability (mean infectivity rates = 0.7%–99.4%), suggesting that cell-intrinsic factors make a large contribution to inter-individual variation in susceptibility. We sought to understand this variation using natural genetic variation and synthetic genetic perturbations, which we hypothesized might lead to the same genes.

## Genome-wide CRISPR-Cas9 screens reveal ZIKV host factors in human NPCs

Viruses commandeer cellular proteins at each stage of their life cycles to propagate and spread to additional cells. The specific host biological processes affected by these manipulations differ based on virus and cell type. We performed genome-wide CRISPR-Cas9 positive selection survival screens to determine which genes are important for ZIKV infection in SNaPs and to limit the number of potential genes that could explain variation in ZIKV susceptibility. We transduced SW7388-1 SNaPs with the Brunello human CRISPR knockout pooled library packaged in the LentiCRISPRv2.0 vector (Figure 5d). Transduced SNaPs were treated with mock-infection media, ZIKV-Ug (MOI = 1), or ZIKV-PR (MOI = 5; higher MOI was used to account for decreased virulence). DNA was extracted from samples 10 days later for sequencing, and gRNA enrichment/depletion analysis across conditions was completed. High gRNA coverage (<0.15% missing gRNAs across replicates) and a strong correlation among replicates testified to the quality of the screen (Figure S6e; Data S1, File 6).

Using the redundant siRNA activity (RSA) statistical method<sup>59</sup>, we detected 102 and 765 candidate host factor genes in the ZIKV-PR and ZIKV-Ug infections, respectively (Figure 5e–f). Strong effects were observed in genes detected in whole-genome experiments performed using conventional hPSC-derived NPCs<sup>60</sup>, including heparan sulfate biosynthesis, endoplasmic reticulum membrane complexes, and the oligomeric Golgi complex. Consistent with recent reports, our screen nominated *ITGAV* and *ITGB5* (Integrin  $\alpha$ V $\beta$ 5)—but not the TAM receptor *AXL*—as entry factors in NPCs<sup>61,62</sup>. We also identified 195 and 147 genes that rendered cells more sensitive to death by ZIKV-PR and ZIKV-Ug, respectively, including Type I interferon-responsive genes *IFNAR1-2* and *IFITM3* (Figure 5g–h). In total, 125 genes were significantly enriched (host factor) or depleted (restriction factor) in both screens.

Several hits were confirmed through arrayed infectivity and viability assays performed in CRISPR-edited SNaPs that were generated from a constitutive-Cas9 H1 hESC line (Figure 5i–j, Figure S6f–k). Along with the discoveries from recent drug screens in human NPCs<sup>63,64</sup>, our results should be considered when developing novel clinical therapies against CZS. These data also nominate a small number of genes as potential mediators of differential SNaP viral susceptibility.

### rs34481144 is a functional QTL influencing NPC sensitivity to ZIKV infection

Common genetic variation that strongly affects the expression of a viral-restricting host gene could also affect host resistance. We hypothesized that our screens of synthetic perturbations and natural genetic variation might converge on one or more genes. Cross-referencing the 993 eGenes and the 125 shared ZIKV genetic factors revealed an overlap of 7 genes (Figure 6a). Of these, the genetic variation at *IFITM3* had the strongest effect size: cells from donors homozygous for the low-expression allele expressed *IFITM3* at less than 1/5 the level of cells homozygous for the high-expression allele (Figure S7a). *IFITM3* is believed to protect cells by shuttling viral-containing endosomes to the lysosome for degradation<sup>65</sup>. *IFITM3* knockdown enhances ZIKV infection rates, while overexpression virtually eliminates viral

replication in cancer cell lines<sup>66,67</sup>. The influence of natural variation in *IFITM3* on flavivirus infectivity, however, has yet to be explored.

We examined which *IFITM3* SNP was most likely responsible for the varied expression levels through pairwise linkage disequilibrium (LD) analysis. Four SNPs were in high LD with the index SNP and associated to *IFITM3* expression with similarly large effect size ( $|E| > 0.6$ ; Figure S7b). The only SNP located in a non-intronic region—rs34481144—is in the 5'-UTR promoter segment of *IFITM3* and has been shown to influence expression of this gene<sup>68</sup>, suggesting that this is the specific variant responsible for the effect. SNaPs from donors homozygous for the reference allele (C) of rs34481144 expressed *IFITM3* at levels 4.8-fold higher than donors homozygous for the alternate allele (T; Figure 6b). We hypothesized that the reference allele of rs34481144 may confer protection from ZIKV infection by enhancing expression of this antiviral gene (Figure 6c).

To test this idea, we exposed SNaP Village-44 to ZIKV-Ug (MOI = 1) or mock media (Figure 6d). At 54 hpi, we used Fluorescence-activated Cell Sorting (FACS) to partition cells into four fractions based on ZIKV envelope protein signal intensity (ZIKV-Negative, -Low, -Mid, -High) before harvesting pellets for DNA extraction. Census-Seq<sup>19</sup> estimated each donor's cellular contribution to the different fractions and showed that rs34481144<sup>TT</sup> donor cells were greatly over-represented in the ZIKV-positive populations relative to the ZIKV-negative pool, whereas rs34481144<sup>CC</sup> cells displayed the opposite relationship (Figure 6e–f, Figure S7c–e). No such relationship was observed with other antiviral/host factor SNPs. The *IFITM3* SNP had no association with the cell type composition of NPC cultures before infection, meaning that genotype was unlikely to influence SNaP induction efficiency (Figure S7f). These data support the hypothesis that the rs34481144-C allele renders NPCs less vulnerable to ZIKV compared to the rs34481144-T allele.

We next asked whether the strong effect of this SNP's genotype on ZIKV susceptibility is cell-intrinsic or arises in an unexpected way from the village design. We analyzed the aforementioned arrayed SNaP infectivity data (Figure 3b) and again found the same significant relationship between rs34481144 genotype and ZIKV infectivity (Figure 6g–h). The cell village and arrayed culture experiments exhibited strong concordance (Figure S7g). In a separate replication set of 36 arrayed SNaP lines, rs34481144 genotype explained 58.8% and 29.4% of the inter-individual variation in ZIKV-Ug and ZIKV-PR infectivity, respectively (Figure 6i, Figure S7h).

The focus on rs34481144 was driven by our CRISPR-Cas9 and eQTL screens. We wondered whether this SNP's effect on ZIKV infectivity was sufficiently strong to have been identifiable in an unbiased search, and performed genome-wide association analysis on the arrayed infectivity data. rs34481144 was the genome's top-scoring SNP in the ZIKV-Ug dataset, reaching genome-wide significance even in this modest 36-donor sample ( $p = 9.0e-10$ ; Figure 6j). rs34481144 also associated strongly with ZIKV-PR infectivity, but just below genome-wide significance ( $p = 3.2e-07$ ; Figure S7i). These results demonstrate that a genetically variable, cell-intrinsic property of human NPCs is the major source of inter-individual variation in their susceptibility to a viral pathogen.

## CRISPR screening enumerates NPC fitness genes

While ZIKV contributes to NDD etiology, the majority of cases are of genetic origin—especially for ASD. Recent studies have highlighted dysregulated NPC proliferation in ASD disease mechanisms<sup>69–73</sup>. However, the full set of genes that influence this phenotype is not known. To identify NPC growth/fitness genes, we compared the gRNA enrichment in Day 10 versus Day 0 mock-treated cells from our whole-genome CRISPR-Cas9 screen (Figure 5d).

We identified 123 genes that enhanced growth upon ablation including known proliferation regulators *NF2* and *PTEN*, and novel factors *KIAA1109* and *CACHD1* (Figure 7a). At a functional level, the genes identified in the screen contribute to primary cilia formation, neural tube development, and WNT signaling (Data S1, File 7). We validated 23 genes in a secondary arrayed screen, in which 39/42 (92.9%) of the nominated gRNAs resulted in a significant decrease in population doubling times; we validated 3 of 3 genes in an additional cell line (H1; Figure S8a–b). Our screen also detected 1,449 genes essential for cell viability (BF > 10).

SNaP proliferation genes were significantly enriched for association with developmental delay<sup>74</sup> (*CHD7*), ASD risk<sup>75–77</sup> (*TSC2*), and tumor suppression<sup>78</sup> (*PTEN*; Figure 7b). There was no enrichment of genes associated with inflammatory bowel disease (IBD), suggesting that enrichment is restricted to disorders of early brain development<sup>41,79–81</sup>. These results validate the importance of NPC proliferation genes in ASD and cancer.

Genome-wide screens have been performed in many human cell types<sup>82</sup>, but it is unclear which fitness/essential genes are shared across tissues. A majority of our NPC proliferation genes (101/123) were unique compared to hESCs and neurons<sup>83–85</sup> (Figure S8c). These cell types shared expression of only 4 genes, including *NF2* and *PTEN*, while *TAOK1* was common to neurons and NPCs; 92 genes were non-tumor suppressor, pro-growth NPC-specific genes<sup>86</sup>. Among essential genes, the overlap of NPCs with other cell types was high (990/1,449 between NPCs and core essential genes; Figure S8d)<sup>83–85,87,88</sup>. There were 187 genes exclusively required for NPC survival, including the GTPase *RHOA*, the human-specific cortical neurogenesis gene *NOTCH2NL*<sup>89</sup>, and aquaporins *AQP5/7*. Thus, these large-scale fitness screens uncovered NPC-specific genetic regulators of growth and maintenance.

Finally, we set out to understand how genetic variation may influence the expression of these NPC growth and survival genes identified in our screen by cross-referencing CRISPR screen results with Village-44 eQTL analysis. We detected eQTLs in 88 SNaP essential genes and 7 proliferation genes (Figure S8e–f). Future experiments investigating the implications of these variants on NPC cellular traits are warranted.

## CACHD1 regulates NPC proliferation and differentiation

The voltage-gated calcium channel regulator *CACHD1*<sup>90</sup> has recently been associated with severe brain malformations and NDDs<sup>91</sup>, and our CRISPR screen results suggested a role for *CACHD1* in NPC proliferation. This protein serves as a modulator of the T-type  $\text{Ca}^{2+}$  channel  $\text{CaV}_{3.1}$ <sup>92</sup>, which is the protein product of the NDD risk gene *CACNA1G*<sup>93–95</sup>.

CaV<sub>3.1</sub> plays critical roles in Ca<sup>2+</sup> homeostasis, cell division, and synaptic plasticity<sup>96</sup>. Given the potential links between CACHD1 and CaV<sub>3.1</sub>-mediated neural functions, we studied the effects of *CACHD1* gene disruption on neurogenesis using cerebral organoids<sup>97</sup>. We quantified size over 28 days and detected a significant increase in the size of *CACHD1*-edited organoids as early as Day 9, when NPCs are the predominant cell type (Figure 7c–d). There were no differences in the number of SOX2<sup>+</sup> NPCs at Day 28, though one of the two tested *CACHD1* gRNAs increased the number of cycling KI67<sup>+</sup> NPCs at this stage, suggesting possible delays in differentiation and cell cycle exit (Figure 7e–g). *CACHD1* editing drastically reduced the number of TBR1<sup>+</sup> subplate neurons and TBR1<sup>+</sup>:SOX2<sup>+</sup> cell ratio compared to NT gRNA controls (Figure 7h–j), indicating significant defects in neurogenic potential<sup>98</sup>. These findings highlight the importance of *CACHD1* in neurogenesis, while establishing SNaPs as an *in vitro* human model for future mechanistic studies of relevant disease mechanisms.

## DISCUSSION

Here, we tried to advance *in vitro* experimental systems to improve our understanding of the relationships among genes, molecules, and cellular phenotypes. Our work using cell villages uncovered a common SNP allele (rs34481144-T) in the *IFITM3* promoter that accounted for nearly 60% of the variation in ZIKV infectivity of human NPCs across dozens of genetic backgrounds. *IFITM3* SNPs rs34481144 is associated with the severity of influenza infections<sup>68</sup>, suggesting that the biomedical significance of this genetic variation may be far broader than resistance to ZIKV. In PBMCs, the presence of the T allele at rs34481144 alters binding of transcription factors and reduces promoter activity, leading to reduced baseline expression of *IFITM3*<sup>68</sup>. We believe that the rs34481144-T allele confers similar vulnerabilities to developing human brain cells exposed to ZIKV.

Intriguingly, frequencies of the rs34481144-T allele are highly variable across different global ancestries, ranging from 46% in Europe to 0.6% in regions in which flaviviruses are endemic<sup>4,99</sup>. It is possible that evolutionary selection has historically favored the rs34481144-C allele in places with high rates of mosquito-borne RNA virus infections. Climate change-induced spread of mosquito vectors into Europe<sup>100</sup> has the potential to introduce flaviviruses to populations in which the *IFITM3* risk allele is far more common. Screening for genotype at the rs34481144 locus might be useful in identifying at-risk individuals during these predicted future outbreaks. Enhancement of *IFITM3* expression levels is worth consideration as a potential therapeutic approach<sup>101,102</sup>.

These results were facilitated by our ability to make human NPCs using a 48-hour induction protocol, which is considerably faster than conventional methods (e.g. dual SMAD). The advantages of the SNaP protocol can be largely attributed to the activities of the neuralization factor NGN2<sup>103–105</sup>. Exogenous expression of NGN2 generates neurons *in vitro* from hPSCs, though the applicability of these cells to disease modeling was recently challenged<sup>24</sup>. We present evidence supporting the developmental fidelity of NGN2-neurons and show, contrary to this recent report, that they pass through a proliferative progenitor stage.

Outside the realm of human-viral interactions, our SNaP experiments provided biological insight into neurodevelopment. The detection of eQTLs affecting critical neurodevelopmental genes and GWAS loci sets the stage for subsequent investigations into the molecular and cellular mechanisms connecting these variants to specific traits and diseases. Furthermore, our CRISPR screen highlighted the importance of NPC expansion in NDD etiology and could help explain the link between ASD and brain overgrowth. Interestingly, the effects of *CACHD1* genetic disruption on NPC proliferation, differentiation, and cerebral organoid size and morphology are strikingly similar to recent observations made in the *PTEN* mutant model of ASD<sup>106</sup>, hinting at potential convergence in underlying disease cellular mechanisms. The clinical implications of our findings are further buttressed by recent human patient data connecting *CACHD1* to a novel syndromic NDD<sup>91</sup>. Studies of the mechanisms by which *CACHD1* disruption affects neurogenesis are needed.

Population-scale *in vitro* culture systems provide promising ways to capture the influence of genetic variation on a wide range of cell types and phenotypes<sup>19,22,69,107</sup>. We hope that these and other new approaches open opportunities to find and characterize the many genetic and environmental factors that shape human development.

### Limitations of the study

The SNaP approach efficiently generates cells resemble NPCs at the transcript, protein, and functional level. These cells show strong preference for differentiating into excitatory rather than inhibitory cortical neurons, which is to be expected given the propensity for neurogenenin-2 overexpression to produce glutamatergic neurons<sup>18</sup> and the fact that most cortical interneurons originate from subcortical structures<sup>108</sup>. We therefore cannot support the use of this cellular model for investigations centered on interneuron development and function. Furthermore, we overexpressed mouse *Ngn2* to generate NPCs; we have yet to show that SNaPs can be made using human *NGN2*. Finally, we should note that while the Dropulation algorithm confidently distinguishes among close relatives, it is not able to distinguish between monozygotic twins or clonal lines derived from the same donor. These limitations should be considered when designing village explorations.

## STAR METHODS

### RESOURCE AVAILABILITY

**Lead Contact**—Steven McCarroll (smccarro@broadinstitute.org)

**Materials availability**—Reagents generated in this study are available from the Lead Contact Steven McCarroll (smccarro@broadinstitute.org) with a completed Materials Transfer Agreement.

### Data and code availability

- All codes and algorithms necessary for Dropulation analysis are available at <https://zenodo.org/badge/latestdoi/128078084>



- Data from this publication, including read-level whole genome and single-cell RNA sequencing data, is organized at [https://app.terra.bio/#workspaces/convergentneuro-mccarroll-anvil/Broad\\_ConvergentNeuroscience\\_McCarroll\\_Nehme\\_SupplementaryVillageData](https://app.terra.bio/#workspaces/convergentneuro-mccarroll-anvil/Broad_ConvergentNeuroscience_McCarroll_Nehme_SupplementaryVillageData).

Here, we provide instructions for requesting access to scRNA-seq BAM, VCF, WGS BAM, and genomic array files generated from hiPSCs (e.g., iPSC Village-104). Users should visit <https://anvilproject.org/data/studies/phs002032> and click “Request Access.” This will send the user to dbGAP (Accession number phs002032). Once granted access by dbGAP, the data can be downloaded from this AnVIL workspace: [https://anvil.terra.bio/#workspaces/anvil-datastorage/AnVIL\\_NIMH\\_Broad\\_ConvergentNeuro\\_McCarroll\\_Eggan\\_CIRM\\_GRU\\_VillageData](https://anvil.terra.bio/#workspaces/anvil-datastorage/AnVIL_NIMH_Broad_ConvergentNeuro_McCarroll_Eggan_CIRM_GRU_VillageData)

For controlled access to scRNA-seq BAM, VCF, WGS BAM, genomic array files generated from hESCs (e.g. SNaP Village-44), users can apply for access through DUOS (<https://duos.broadinstitute.org>; Accession number DUOS-000121). Once approved, the data can be downloaded from this Terra workspace: [https://app.terra.bio/#workspaces/convergneuro-mccarroll-anvil/Broad\\_ConvergentNeuro\\_McCarroll\\_Nehme\\_hESC\\_HMB\\_VillageData](https://app.terra.bio/#workspaces/convergneuro-mccarroll-anvil/Broad_ConvergentNeuro_McCarroll_Nehme_hESC_HMB_VillageData)

- Further information requests can be directed to Steven McCarroll (smccarro@broadinstitute.org)

## EXPERIMENTAL MODELS AND SUBJECT DETAILS

**Stem cell culture**—Human ESCs and iPSCs were obtained from various sources. All ESCs are registered in the NIH Human Embryonic Stem Cell Registry. Most iPSCs (e.g. CW1052, CW20012, CW20025, etc). were obtained from CIRM/FujiFilm Cellular Dynamics. Some iPSCs (Mito22, Mito226, NFID1300, NFID1301, NFID1337, NFID\_0176, SW510926-11, SW7388-1) were reprogrammed at the Broad Institute. All lines were authenticated—either by our lab or the providing entity—by genotyping, karyotyping, growth rate measurements, and *in vitro* differentiation to establish pluripotency. The exact lines used for each experiment can be found in Table S1. Human ESCs and iPSCs were maintained in mTeSR media (Stem Cell Technologies, 85850) on Geltrex basement membrane matrix (1:100; Life Technologies, A1413301). Cells were split every 4-5 days (when they reached 80-90% confluency) using a 15 minute/37°C incubation in Accutase (Innovative Cell Technologies, AT104) followed by 1:10 dilution in mTeSR. For each passage, media was supplemented with ROCK inhibitor Y-27632 (10 μM; Stemgent, 04-0012) for 12-24 hours after plating.

**SNaP culture**—Human SNaPs were maintained on Geltrex basement membrane matrix (1:100; Life Technologies, A1413301) in SNaP maintenance media: DMEM/F12, Glutamax (1:100), MEM-NEAA (1:100; Life Technologies, 10370088), B27 minus Vitamin A (1:50; Life Technologies, 12587010), N2 Supplement (1:100; Life Technologies, 17502048), recombinant human EGF (10 ng/mL; R&D Systems, 236-EG-200), recombinant human

basic FGF (10 ng/mL; Life Technologies, 13256029). Cells were split weekly by dissociating with Accutase (Innovative Cell Technologies, AT104) and plating at 120,000 cells/cm<sup>2</sup>. For each passage, media was supplemented with ROCK inhibitor Y-27632 (10 μM; Stemgent, 04-0012) for 12-24 hours after plating.

## METHOD DETAILS

**Human iPSC village**—Human iPSC lines from 104 donors were maintained as independent cultures for one week. Cells were dissociated with Accutase, counted using a Scepter Handheld Automated Cell Counter (Millipore Sigma, PHCC20060), and plated at equal proportions in a 10 cm<sup>2</sup> Geltrex-coated dish at 50,000 cells/cm<sup>2</sup>. For Dropulation experiments, cells were harvested 5 days post-plating and run through the 10X Chromium Single Cell 3' Reagents V2 system to isolate individual cells into droplets per vendor's instructions (10X Genomics; San Francisco, CA). Samples were then sequenced on a NovaSeq 6000 system (Illumina) using a S2 flow cell at 2 × 100bp.

**Lentiviral transduction**—TetO-Ngn2-Puromycin and Ubq-rtTA constructs were obtained from the Wernig lab (Stanford) before being packaged as high-titer lentiviruses (Alstem, Richmond, CA). When hPSCs reached 80-100% confluency, they were dissociated with Accutase before being re-suspended in lentivirus-containing mTeSR media supplemented with Y-27632 at a range of MOI = 1 to MOI = 3. Cells were then plated on Geltrex-coated 12-well plates at 500,000 cells per well in a total volume of 750 μl/well. After 18-24 hours, lentiviral media was aspirated, and cells were fed with mTeSR media and maintained as described above. Transduced cells were maintained for up to 10 passages for inductions, and transduction efficiencies typically ranged from 65-85% across cell lines. Cell lines that failed SNaP induction (Figure S4b) tended to show transduction efficiencies below 30%.

**Induction of SNaPs from human PSCs**—Human PSCs were dissociated and plated at 75,000 cells/cm<sup>2</sup> on Geltrex matrix in mTeSR media supplemented with Y-27632. After 12-24 hours, cells were fed with Induction Media (Day 1): DMEM/F12 (ThermoFisher, 11320082), Glutamax (1:100; ThermoFisher, 10565018), 20% Glucose (1.5% v/v), N2 Supplement (1:100, ThermoFisher, 17502048), Doxycycline (2 μg/mL; Sigma-Aldrich, D9891), LDN-193189 (200 nM; Stemgent, 04-0074), SB431542 (10 μM; Tocris, 1614), and XAV939 (2 μM; Stemgent, 04-00046). After 24 hours in Induction Media, cells were fed with Selection Media (Day 2): DMEM/F12, Glutamax (1:100), 20% Glucose (1.5% v/v), N2 Supplement (1:100), Doxycycline (2 μg/mL), puromycin (5 μg/mL; ThermoFisher, A1113803), LDN-193189 (100 nM), SB431542 (5 μM), and XAV939 (1 μM). After 24 hours in Selection Media, SNaPs were dissociated with Accutase and replated at 120,000 cells/cm<sup>2</sup> on Geltrex-coated plates in SNaP maintenance media supplemented with puromycin and Y-27632 (Day 3): DMEM/F12, Glutamax (1:100), MEM-NEAA (1:100; Life Technologies, 10370088), B27 minus Vitamin A (1:50; Life Technologies, 12587010), N2 Supplement (1:100; Life Technologies, 17502048), recombinant human EGF (10 ng/mL; R&D Systems, 236-EG-200), recombinant human basic FGF (10 ng/mL; Life Technologies, 13256029), puromycin (5 μg/mL), and Y-27632 (10 μM). Starting 8-18 hours after passaging, SNaPs were fed daily with SNaP maintenance media lacking Y-27632 and

puromycin. SNaPs were passaged every week in SNaP maintenance media with Y-27632, and then fed daily with SNaP maintenance media.

**Immunostaining**—SNaPs were washed with 1X PBS and then fixed with 4% paraformaldehyde for 15 minutes at room temperature before three more washes with 1X PBS. Cells were permeabilized with 0.1% Triton for 15 minutes and then blocked with 10% normal donkey serum diluted in 1X PBS for 1 hour at room temperature followed by an overnight 4°C incubation in primary antibody diluted in blocking solution: Mouse anti-Nestin (1:1000; Stem Cell Technologies, 60091), Rabbit anti-PAX6 (1:500; Stem Cell Technologies, 60094), Mouse anti-OCT4 (1:1000; Stem Cell Technologies, 60093), Rabbit anti-SOX1 (1:1000; Stem Cell Technologies, 60095), Mouse anti-SOX2 (1:100; R&D Systems, MAB2018), Rabbit anti-ZO1 (1:200; Life Technologies, 617300), Rabbit anti-FOXG1 (1:400; Abcam, 18259), and/or Rabbit anti-KI67 (1:500; ThermoFisher, MA5-14520). After 3 washes in 1X PBS at room temperature, cells were incubated for 2-4 hours at room temperature in secondary antibody diluted in blocking solution: Donkey anti-mouse Alexa647 (1:1000; Life Technologies, A-31571) and/or Donkey anti-rabbit Alexa555 (1:1000; Life Technologies A-31572). Cells were washed once with 1X PBS followed by a 5-minute incubation in 4', 6-Diamidino-2-Phenylindole Dihydrochloride (DAPI, 1:5000; Life Technologies, D1306). Finally, cells were washed twice more with 1X PBS prior to imaging. For each well of a 96-well plate, 4-8 fluorescent images were captured using the Cytation 3 cell imaging multi-mode reader (BioTek Instruments; Winooski, VT). All images were then processed using the CellProfiler imaging analysis software<sup>109</sup> to quantify the percentage of NPC marker-positive cells.

**Flow cytometry analysis of human pluripotent stem cells and SNaPs**—Human pluripotent stem cells and SNaPs were stained with OCT3/4 antibody (BD Biosciences, 560794) following the manufacturer's instructions contained in the Human Pluripotent Stem Cell Transcription Factor Analysis Kit (BD Biosciences, 560589): hPSCs and SNaPs were dissociated with Accutase and fixed using BD Cytfix at a concentration of  $1 \times 10^7$  cells/mL for 20 minutes at room temperature. Fixed cells were washed twice with 1X PBS and permeabilized using 1X BD Perm/Wash at a concentration of  $1 \times 10^7$  cells/mL for 10 minutes at room temperature.  $1 \times 10^6$  fixed/permeabilized cells were stained with OCT3/4 antibody at a concentration of  $1 \times 10^7$  cells/mL for 20 minutes at room temperature in the dark. Stained cells were washed twice with 1X BD Perm/Wash, resuspended in 1X PBS, and kept on ice in the dark until analysis. Cells were passed through a filter-top 12x75 mm polystyrene tube just before analysis on a BD FACSAria II (BD Biosciences; San Jose, CA). Data was presented with OCT3/4 on the x-axis (PerCP-Cy5.5) and the empty channel mCFP-A on the y-axis.

**SNaP population RNA-seq**—H9 (a.k.a. WA09) hESCs and early passage H9-derived SNaPs were harvested in 350  $\mu$ L of RTL Plus before RNA was extracted using the RNeasy Plus Micro Kit (Qiagen, 74034). For each replicate, 50 ng of purified RNA was used for library construction. Smart-seq2 libraries were then prepared as follows: Total RNA was captured and purified on RNAClean XP beads (Beckman Coulter, A63987). Polyadenylated mRNA was then selected using an anchored oligo(dT) primer and converted to cDNA

via reverse transcription. First strand cDNA was subjected to limited PCR amplification followed by transposon-based fragmentation using the Nextera XT DNA Library Preparation Kit (Illumina, FC-131-1024). Samples were then PCR amplified using barcoded primers such that each sample carried a specific combination of Illumina P5 and P7 barcodes, and then pooled together prior to sequencing. Sequencing was performed on an Illumina NextSeq500 using 2 x 25bp reads (Illumina; San Diego, CA).

**Dual-SMAD neural progenitor cell induction (Protocol #1, Figure S2i)**—When SW7388-1 hiPSCs reached 80-90% confluency (Day 0), neuroectodermal differentiation media A was added to induce NPCs: DMEM/F12 (47% v/v), Neurobasal media (47% v/v; Life Technologies, 21103049), Glutamax (1:50), MEM-NEAA (1:100), B27 (1:50; Life Technologies, 17504044), N2 Supplement (1:100), SB431542 (10  $\mu$ M), LDN (100 nM), and XAV-939 (2  $\mu$ M). Cells underwent complete media exchanges daily. At Day 14, cells were harvested for RNA extraction and subsequent qPCR experiments.

**Dual-SMAD neural progenitor cell induction (Protocol #2, Figure S2j)**—When SW7388-1 hiPSCs reached 80-90% confluency (Day 0), neuroectodermal differentiation media A with retinoic acid (1  $\mu$ M; Sigma, R2625) in place of XAV-939. Cells underwent complete media exchanges daily. Starting on Day 7, cells were fed daily with neuroectodermal differentiation media B: DMEM/F12 (47% v/v), Neurobasal media (47% v/v), Glutamax (1:50), MEM-NEAA (1:100), B27 (1:50), N2 Supplement (1:100), and Retinoic acid (1  $\mu$ M). At Day 21, cells were harvested for RNA extraction for subsequent qPCR experiments.

**SNaP developmental qPCR**—SW7388-1 human iPSCs, SNaPs, and dual-SMAD NPCs were harvested in 350  $\mu$ L of RTL Plus (Qiagen, 1053393) per well of a 24-plate. RNA was extracted from the samples using the RNeasy Plus Micro Kit (Qiagen, 74034). Purified RNA was then used as input for the iScript cDNA Synthesis reaction (Bio-Rad, 1708891) and the product was diluted 1:5 in nuclease-free water. For each sample, 1  $\mu$ L of cDNA was added to iTaq Universal SYBR Green Supermix (Bio-Rad, 1725124) that contained 500 nM of forward and reverse primers in a final volume of 20  $\mu$ L per well of a 384-well plate. Primers were manufactured by Integrated DNA Technologies (*NESTIN*: 5'-CTG CTA CCC TTG AGA CAC CTG-3' and 5'-GGG CTC TGA TCT CTG CAT CTA C-3'; *PAX6*: 5'-AAC GAT AAC ATA CCA AGC GTG T-3' and 5'-GGT CTG CCC GTT CAA CAT C-3'; *SOX1*: 5'-CCA CAT CCT AAT CTT GAG CCA-3' and 5'-CTG ACG TCC ACT CTC AGT CT-3'; *OCT4*: 5'-CCA AGG AAT AGT CTG TAG AAG TGC-3' and 5'-TGC ATG AGT CAG TGA ACA GG-3'; *FOXG1*: 5'-CGT CCA CCA TAT AGT TCC ATG A-3' and 5'-TGA CTG CTT TGC CAT TTC ATT C-3'; *SOX2*: 5'-CTT GAC CAC CGA ACC CAT-3' and 5'-GTA CAA CTC CAT GAC CAG CTC-3'; *GAPDH*: 5'-TTG TCA AGC TCA TTT CCT GGT ATG-3' and 5'-TCC TCT TGT GCT CTT GCT GG-3'). qPCR reactions were run for 40 cycles on the CFX384 Touch Real-Time PCR Detection System (Bio-Rad; Hercules, CA). All samples were run in triplicate, and results were normalized to a GAPDH control run in duplicate. Ct values were calculated and plotted to show relative expression.

**NGN2 ChIP-seq**—SW7388-1 hiPSCs were transfected with V5-tagged constructs: (1) TetO-NGN2-V5 + TetO-GFP or (2) TetO-NGN2 + TetO-GFP-V5. Two technical replicates were included for each time point at Day 0 (stem cells), Day 1, and Day 2 (SNaPs) post-induction. Cells were fixed in 1% paraformaldehyde for 10 minutes at 37°C, lysed, and sonicated (Branson sonicator) for 8 minutes on ice (40% amplitude, 0.7 seconds ON + 1.3 seconds off). Immunoprecipitation was carried out using Anti-V5-tag mAb (100 µg/100 µL; MBL #M167-3). DNA was recovered by reversing crosslinks and purified using AMPure XP beads (Beckman Coulter A63880). DNA libraries were produced with the Illumina TrueSeq library kit as per manufacturer's instruction and sequencing was performed on an Illumina NextSeq500. Bowtie2 was used to align reads to the GRCh38 reference genome<sup>110</sup>. Peak calling was performed using MACS with a bandwidth of 300 bp<sup>111</sup>. The TetO-NGN2 + TetO-GFP-V5 samples were used as input background controls.

**Clonal assay for self-renewal**—SNaPs were plated as single cells in SNaP maintenance media plus Y-27632 on Geltrex-coated 96-well plates using a BD FACSAria II. Cells were then fed daily with SNaP maintenance media for two weeks. At this point, most wells were fixed and stained with Mouse anti-NESTIN (1:1000; Stem Cell Technologies, 60091) and Rabbit anti-PAX6 (1:500; Stem Cell Technologies, 60094) for quantification of proliferation. The remaining wells were dissociated and re-plated as single cells in SNaP maintenance media plus Y-27632. Media was changed the following day to spontaneous differentiation media (base media plus B27/N2) and fed 2-3 times a week for two weeks. SNaPs were then fixed and processed for immunostaining as described above.

**SNaP differentiation single-cell RNA sequencing**—SNaPs were differentiated for 15 days in spontaneous differentiation media [DMEM/F12, Glutamax (1:50), MEM-NEAA (1:100), B27 (1:50), N2 Supplement (1:100)] and dissociated with a 15 minute/37°C Accutase treatment. Samples were filtered via 40 µm tip filters (BelArt, H13680-0040) and centrifuged at 400xg for 5 minutes. Cells were resuspended to 1 million cells/mL and run through the 10X Chromium V2 scRNA-seq pipeline per vendor's instructions. Samples were sequenced on a HiSeq 4000 (Illumina) using 2 x 50-cycle SBS kits (Illumina, FC-410-1001) and clustering was done on a HiSeq 4000 flow cell via cBot2 (Illumina). The library was then sequenced with paired-end reagents, with 26xRead 1 cycles, 8x7 index cycles, and 98xRead 2 cycles.

**SNaP differentiation immunostaining**—SNaPs were plated on Geltrex at 10,000-15,000 cells/cm<sup>2</sup> in spontaneous differentiation media or base media with 10% fetal bovine serum (GE Healthcare, 16777-014) or Astrocyte Media (ScienCell, 1801) in the presence of Y-27632 (10 µM). The media was exchanged the following day to remove Y-27632, and cells were then fed 2-3 times a week for 14 days (base and 10% FBS) or 20-60 days (Astrocyte media). Cells cultured in Astrocyte Media were passaged weekly and re-plated at 15,000 cells/cm<sup>2</sup> without Y-27632. To determine cell identity, SNaP-derived cells were immunostained and quantified as described above. The following primary antibodies were used for these experiments: Mouse anti-HuC/D (1:200; Life Technologies, A-21271), Rat anti-CD44 (1:400; eBioScience, 17-0441-82), Rabbit anti-S100β (1:1000; Sigma Aldrich, S2532), Rabbit anti-GFAP (1:100; Millipore, AB5804), Chicken anti-MAP2



(1:500; Abcam, ab5392), Rabbit anti-Synapsin I (1:1000; Millipore, AB1543), Rabbit anti-BRN2 (1:300; Abcam, ab137469), Rabbit anti-CUX2 (1:200; Abcam, ab130395), and/or Rat anti-CTIP2 (1:1000; Abcam, ab18465). Donkey anti-Mouse Alexa555 (1:1000; Life Technologies, A-31570), Donkey anti-Mouse Alexa647 (1:1000; Life Technologies, A-31571), Donkey anti-Rabbit Alexa555 (1:1000; Life Technologies, A-31572), Goat anti-Rat Alexa555 (1:1000, Life Technologies, A-21434), and/or Goat anti-Chicken Alexa647 (1:1000; Life Technologies, A-21449) were used as secondary antibodies.

**Multi-electrode array (MEA)**—SNaPs were plated on Geltrex at 15,000 cells/cm<sup>2</sup> in base media supplemented with Y-27632. The media was exchanged the following day to remove Y-27632. After 5-6 days in culture, the cells were fed with base media containing DAPT (5 μM; DNSK International). Two days later, the partially differentiated SNaPs were then dissociated and re-plated at 15,000 cells/cm<sup>2</sup> in DAPT-containing base media. One week later, the post-mitotic cells were dissociated and co-cultured with primary mouse glia (23,000 glia + 13,000 neurons per well) on a Geltrex-coated 12-well MEA plate (Axion Biosystems, M768-GL1-30Pt200) in Neurobasal complete media [Neurobasal media (97% v/v; Life Technologies 21103049), Glutamax (1:100), 20% Glucose (1.5% v/v), MEM-NEAA (1:200), B27 (1:50), BDNF (10 ng/mL), CTNF (10 ng/mL), and GDNF (10 ng/mL)]. Cells were fed 2-3 times per week with partial exchanges to reach a final volume consisting of 80% fresh media and 20% conditioned media. Five minutes of neuronal activity was measured weekly using the Maestro 12-well 64 electrodes per well micro-electrode array (MEA) plate system (Axion Biosystems, Atlanta, GA). After approximately 50 days in co-culture, synaptic contents were assessed using pharmacological blockers of neurotransmitter receptors. More specifically, baseline activity was measured for 5 minutes prior to the addition of NBQX (10 μM), D-APV (50 μM), or Picrotoxin (50 μM) directly to the conditioned media. After a brief 5-10 incubation, neuronal activity was again measured for 5 minutes. Data was analyzed using the Axion Integrated Studio 2.4.2 and the Neural Metric Tool (Axion Biosystems).

**SNaP village construction and experimental design**—SNaP lines were generated from dozens of unique hESC donor lines and maintained as independent cultures (Table S1). At Passage 2-3, SNaP lines were dissociated with Accutase and counted using a Countess II FL (ThermoFisher Scientific). An equal number of SNaPs from each cell line were then plated together in a 10 cm<sup>2</sup> Geltrex-coated dish at 120,000 cells/cm<sup>2</sup>. For Dropulation experiments, cells were harvested 2-3 days post-plating and run through the 10X Chromium Single Cell 3' Reagents V2 system to isolate individual cells into droplets per vendor's instructions (10X Genomics; San Francisco, CA). Samples were then sequenced on a NovaSeq 6000 system (Illumina) using a S2 flow cell at 2 x 100bp.

**ZIKV propagation**—Vero cells (ATCC, CCL-81) were plated on uncoated 10 cm<sup>2</sup> dishes in Vero cell growth media (DMEM + 10% heat-inactivated fetal bovine serum). At 80-90% confluency, cells were exposed to ZIKV-Ug (ATCC, VR-1838) or ZIKV-PR (ATCC, VR-1843) diluted in HyClone Earle's 1X Balanced Salt Solution (EBSS; GE Healthcare Life Sciences, SH30029.02) at a low multiplicity of infection (<0.1; based on ATCC manufacturer's quantification of viral titer) in the minimal amount of media to cover the



cells (3 ml). Cells were incubated for 1 hour at 37°C/5% CO<sub>2</sub> with gentle rocking every 15 minutes to prevent the cells from drying. After this infection period, the inoculum was removed and replaced with 12 ml of Vero cell maintenance media (DMEM + 2% heat-inactivated fetal bovine serum) pre-heated to 37°C. Two days after infection, the Vero cell conditioned media was collected and centrifuged for 10 minutes at 2000xg at room temperature to remove cell debris. The virus was aliquoted and stored at -80°C prior to quantification.

**ZIKV quantification via focus forming assay**—Vero cells were plated at 150,000 per well on uncoated 24-well plates in Vero cell growth media and incubated at 37°C/5% CO<sub>2</sub>. One to two days post-plating, cells were rinsed with 1X PBS and then infected with 125 µL of virus diluted in 1X EBSS (10<sup>-4</sup> to 10<sup>-7</sup>) for 1 hour at 37°C/5% CO<sub>2</sub> with gentle rocking of the plate every 15 minutes. After the infection period, cells were rinsed with 1X PBS. Then, 1 mL of pre-warmed overlay media [(2.1% carboxymethylcellulose sodium salt (CMC) in DMEM and 2% HI-FBS)] was slowly added onto the monolayer of infected Vero cells. Thirty-six hours later, cells were rinsed several times with 1X PBS to remove the CMC precipitates and then fixed in 4% paraformaldehyde for 15 minutes at room temperature. Post-fixation, cells were rinsed three times with 1X PBS and then permeabilized with 0.1% Triton in 1X PBS for 10 minutes. Cells were blocked with 10% normal donkey serum for 30 minutes at room temperature followed by a 1 hour/37°C incubation in 150 µL of 1:1000 Mouse monoclonal D1-4G2 anti-flavivirus envelope protein (EMD Millipore, MAB10216) antibody diluted in blocking solution. After two washes in 1X PBS at room temperature, cells were incubated for 1 hour/37°C in 1:1000 Goat anti-Mouse HRP-conjugated secondary antibody (Abcam, ab6789) in blocking solution. Cells were once again washed twice with 1X PBS at room temperature followed by the addition of peroxidase substrate (Vector Laboratories, SK-4600) in 1X PBS. The number of foci were counted and then multiplied by the dilution factor to quantify the viral titer. Dilutions were run in quadruplicate and ZIKV-Ug and ZIKV-PR were quantified at the same time. ZIKV-Ug and ZIKV-PR were quantified as 5.5 x 10<sup>7</sup> and 4.0 x 10<sup>7</sup> focus forming units (FFU) per mL, respectively.

**ZIKV infectivity assay**—SNaPs were infected for 1 hour at 37°C/5% CO<sub>2</sub> with ZIKV-Ug or ZIKV-PR diluted in 1X EBSS at a MOI of 10, 1, 0.1, or 0.01. Vero cell conditioned media from uninfected cells was diluted in 1x EBSS and used for the mock controls. Mock and ZIKV-infected cells were fixed 54 hours post-infection (hpi) with 4% paraformaldehyde for 15 minutes at room temperature and then washed with 1X PBS. Cells were permeabilized with 0.1% Triton for 15 minutes and then blocked with 10% normal donkey serum diluted in 1X PBS for 1 hour at room temperature followed by an overnight 4°C incubation in primary antibody: Mouse monoclonal D1-4G2 anti-flavivirus envelope protein (1:500; EMD Millipore, MAB10216) and Rabbit anti-PAX6 (1:500; Stem Cell Technologies, 60094) antibody diluted in blocking solution. After 3 washes in 1X PBS at room temperature, cells were incubated for 2-4 hours at room temperature in secondary antibody: Donkey anti-Mouse Alexa647 (1:1000) and Donkey anti-Rabbit Alexa555 (1:1000). Cells were washed once with 1X PBS followed by a 5-minute incubation in DAPI (1:5000). For Vero cell infections, an additional 20-minute room temperature incubation with F-Actin CytoPainter Phalloidin-iFluor 555 Reagent (1:10000; Abcam, ab176756) was included. Finally, cells

were washed twice more with 1X PBS prior to imaging. For each well of a 96-well plate, 4-8 fluorescent images were taken using the Cytation 3 cell imaging multi-mode reader. All images were then processed using the CellProfiler imaging analysis software to quantify the percentage of 4G2-positive PAX6 stained cells.

**Cell viability assay**—At 96-120 hpi, cell viability was quantified using the CellTiter Glo 2.0 kit (Promega, G9242). Culture media was removed, and cells were washed once in 100  $\mu$ L 1X PBS. The 1X PBS was removed and replaced with 100  $\mu$ L CellTiter Glo reagent. Plates were rocked gently for 2 minutes to facilitate cell lysis. After a 10-minute incubation at room temperature, luminescence was measured using the Cytation 3 cell imaging multi-mode reader. Data is presented as luminescence as a percentage of mock-infected controls.

**qPCR quantification of ZIKV**—SNaPs were infected with ZIKV-Ug or ZIKV-PR at a MOI of 10 or with mock media for 1 hour at 37°C. At 24 hpi and 72 hpi, 170  $\mu$ L of conditioned media was removed from each well of a 96-well plate. 140  $\mu$ L of this supernatant was flash frozen in dry ice and stored at  $-80^{\circ}\text{C}$  for subsequent RNA extraction experiments, while 30  $\mu$ L of the media was added directly to Vero cells for a 1-hour infection at 37°C. Vero cell infectivity was measured at 24 hpi using the infectivity assay. Viral RNA was prepared from the conditioned media using the QIAamp Viral RNA Mini Kit (Qiagen, 52906). cDNA was prepared from 10  $\mu$ L of viral RNA per sample using the iScript Reverse Transcription Supermix (Bio-Rad, 1708841). At the same time, cDNA from  $6 \times 1:10$  dilutions of the stock viral RNA of known focus-forming units per mL (FFU/mL) was prepared. qRT-PCR was conducted using the CFX96 Touch Real-Time PCR Detection System (Bio-Rad). For each sample, 1  $\mu$ L of the cDNA or quantified stock cDNA was added to 5  $\mu$ L iTaq Universal SYBR Green Supermix (Bio-Rad, 1725124), 400 nM primers (Integrated DNA Technologies; Coralville, IA) designed for ZIKV-Ug or ZIKV-PR (ZIKV-Ug primers: TGG GA G GTT TGA AGA GGT TG and TCT CAA CAT GGC AGC AAG ATC T; ZIKV-PR primers: GGG ACA GTC ACA GTG GAG GT and GGT GGA TCA AGT TCC AGC AT), and enough nuclease-free  $\text{dH}_2\text{O}$  for a final reaction volume of 20  $\mu$ L. Standard curves were established for each strain using the quantified stock dilutions and were used to assign FFU/mL values to tested samples. Due to the high concentration of virus in the supernatant samples, cDNA samples required a 1:100 dilution to fit within the acceptable CT range.

**Human antiviral response**—Mock and ZIKV-infected SNaPs (MOI = 10 for ZIKV-Ug; MOI = 20 for ZIKV-PR) were harvested at 60 hpi using 350  $\mu$ L of RLT Plus reagent. Total RNA was then extracted using the RNeasy Plus Micro kit (Qiagen, 74034). The RT<sup>2</sup> First Strand kit (Qiagen, 330404) was used to prepare cDNA using 400 ng for each sample. The cDNA from each sample was then diluted in 91  $\mu$ L of nuclease-free  $\text{H}_2\text{O}$ . Then, 102  $\mu$ L of cDNA was added to 1,248  $\mu$ L of nuclease-free  $\text{H}_2\text{O}$  and 1,350  $\mu$ L of 2X RT<sup>2</sup> SYBR Green qPCR master mix (Qiagen, 330503). The master mix (10  $\mu$ L) was then added into each well of a 384-well RT<sup>2</sup> Profiler PCR Array Human Antiviral Response plate (Qiagen, PAHS-122ZE-4) that contained primers for 84 genes related to the human antiviral response and five housekeeping genes that were used as internal controls (*ACTB*, *B2M*, *GAPDH*,

*HPRT1*, and *RPL13A*). The data was analyzed using the online portal provided by the kit ([www.SABiosciences.com/pcrarrayprotocolfiles.php](http://www.SABiosciences.com/pcrarrayprotocolfiles.php)).

**Genome-wide CRISPR-Cas9 ZIKV survival screens**—All gRNA and lentiviral reagents for the primary and validation screens were generated at Broad Institute Genetic Perturbation Platform. SW7388-1 SNaPs were transduced with the Brunello barcoded sgRNA library (CP0043 Brunello library containing 77,441 barcoded sgRNAs targeting 19,114 genes and 1,000 not-targeting guides) delivered through the all-in-one LentiCRISPRv2.0 system (pXPR\_BRD023 vector)<sup>112,113</sup>. One hundred million SNaPs per replicate (3 total replicates) were transduced using the spinoculation method, in which cells were cultured in suspension with LentiCRISPRv2.0 (estimated MOI = 0.4) and centrifuged at room temperature for 2 hours at 1,000 rpm before being plated at 120,000 cells/cm<sup>2</sup> on Geltrex coated plates. Transduced SNaPs were then expanded and selected with puromycin (1 µg/mL) for one week, at which point they were passaged onto 15 cm<sup>2</sup> Geltrex-coated dishes at 120,000 cells/cm<sup>2</sup> (40 million cells were plated per replicate to maintain the 500 cells per sgRNA representation). Two days post-plating (one day post-Y27632 removal), SNaPs were either: (1) harvested using Accutase followed by PBS washes (“Pre-infection/Day 0” samples), (2) infected with mock media (for “Mock” samples), (3) infected with ZIKV-Ug (MOI = 1), or (4) infected with ZIKV-PR (MOI = 5) in minimal media for 1 hour at 37°C/5% CO<sub>2</sub> with gentle rocking every 15 minutes to prevent the cells from drying. Cells were then fed every other day starting at 48 hpi by removing all media, washing once with 1X PBS to remove dead cells and debris, and then adding back a 50:50 fresh SNaP maintenance media/conditioned media mixture. On Day 10, all samples (“Mock/Day 10”, “ZIKV-Ug”, and “ZIKV-PR”) were harvested and frozen at –80°C. DNA was then extracted using the QIAmp DNA Blood Maxi kit (Qiagen, 51192). PCR and sequencing were performed as previously described<sup>112,114</sup>. Samples were sequenced on a HiSeq2000 (Illumina).

**H1 constitutive Cas9 stem cell line**—A targeting vector with AAVS1 homology arms and a Flag-Cas9-2A-Blast-BGHpA expression cassette was generated and co-electroporated with AAVS1 TALENS (System Biosciences) into H1 hESCs using the Neon Transfection System (Thermo Fisher Scientific; Waltham, MA). Two days post-electroporation, Blasticidin (4µg/mL; Thermo Fisher Scientific, R21001) was added and emerging clones were picked and analyzed by immunocytochemistry for FLAG-Cas9 using Mouse anti-FLAG antibody (1:300; Sigma Aldrich, F1804) and by PCR for proper integration into the locus across the junctions (5' junction: *AAVS1-F2* AACTCTGCCCTCTAACGCTG and *CAG-R2* CTATGAACTAATGACCCCGTAATTG; 3' junction: *BGH-F1* GGAAGACAATAGCAGGCATGC and *AAVS1-R4* CCACGTAACCTGAGAAGGGAAT; Non-targeted allele: *AAVS1-F3* CCTGGCCATTGTCACCTTTGC and *AAVS1-R4* CCACGTAACCTGAGAAGGGAAT). The H1-36-23 clone was differentiated into neurons using a dual SMAD inhibition protocol and plated into 96-well plates. sgRNAs were delivered by lentiviral vectors that confer puromycin resistance, and the neurons were selected with puromycin (2 µg/mL) for 2 weeks. Neurons were lysed and next-generation sequencing of the gRNA-targeted sites was performed in order to identify and quantify indels generated.

**Validation of CRISPR-Cas9 ZIKV survival screen**—SNaPs were induced from H1-36-23 constitutive Cas9 stem cells before lentiviral transduction via spinoculation with individual sgRNAs (pXPR\_003 and pXPR\_050 vectors) in a 24-well plate format. Cells were then expanded and selected with puromycin (1  $\mu\text{g}/\text{mL}$ ) for one week. Cas9 gRNA-expressing SNaPs were passaged and plated onto Geltrex at 40,000 cells per well of a 96-well plate (120,000 cells/cm<sup>2</sup>). Two days later, SNaPs were infected with ZIKV-Ug (MOI = 1) before conducting the infectivity assay at 54 hpi and the cell viability assay at 120 hpi. Infectivity and cell viability values were compared to Cas9-SNaPs that were transduced with non-targeting gRNAs.

**Census-Seq**—Village-44 SNaPs were exposed to mock media or ZIKV-Ug (MOI = 1) for 1 hour at 37°C. After 54 hours, cells were harvested and fixed for 20 minutes at room temperature in BD Cytifix (BD Biosciences, 554714) at a concentration of  $1 \times 10^7$  cells/mL. Fixed cells were washed twice with 1X PBS and permeabilized using 1X BD Perm/Wash (BD Biosciences, 554723) at a concentration of  $1 \times 10^7$  cells/mL for 10 minutes at room temperature. Cells were then stained with Ms anti-4G2 antibody (1:100) in perm/wash buffer for 1 hr at room temperature in the dark. Stained cells were washed twice with 1X BD Perm/Wash, resuspended in 1X PBS, and kept at 4°C in the dark until analysis. Finally, stained cells were passed through a filter-top 12x75 mm polystyrene tube just before analysis on a BD FACSAria (BD Biosciences; San Jose, CA). Samples were separated based on GFP signal intensity into four bins: ZIKV-Negative (~60% of total cells), ZIKV-Low (~13.3%), ZIKV-Mid (~13.3%), and ZIKV-High (~13.3%).

DNA was unfixed and extracted from each sample using a column-free procedure: First, FAC sorted samples were spun down and resuspended in 300  $\mu\text{L}$  Cell Lysis Solution (Qiagen, 158906) with 2  $\mu\text{L}$  Proteinase K (New England Biolabs, P8107S) and incubated at 56°C overnight. The next day, 1.5  $\mu\text{L}$  of RNase A (Qiagen 158922) was added to each sample prior to a 30-minute incubation at 37°C. Samples were placed on ice for 5 minutes and spun briefly before 200  $\mu\text{L}$  of Protein Precipitation Solution (Qiagen 158910) was added. Samples were vortexed for 20 seconds and then spun down at 13,200 RPM for 10 minutes at 4°C. The supernatant was then transferred to a chilled tube containing 300  $\mu\text{L}$  ice-cold 100% isopropanol with 0.5  $\mu\text{L}$  Glycogen Solution (Qiagen 158930). Samples were again spun down at 13,200 RPM for 10 minutes at 4°C before discarding the supernatant. The DNA pellet was washed in 300  $\mu\text{L}$  of 70% Ethanol and centrifuged at 14,000 RPM for 5 minutes at 4°C. The supernatant was discarded, and the DNA pellet was dried at room temperature for 10 minutes so that all of the ethanol was evaporated. Finally, the pellet was re-hydrated at 15-20  $\mu\text{L}$  dH<sub>2</sub>O and then incubated for 1 hour at 55°C.

Extracted DNA from FAC-sorted bins were processed for low-coverage DNA sequencing, which was performed using either the TruSeq NanoDNA Library Prep for NeoPrep (Illumina Catalog# NP-101-9001DOC) or Nextera DNA Library Prep (Illumina FC-121-1030) setup (Note: any standard kit that generates sequence libraries from DNA can be compatible with the pipeline). Libraries were then sequenced on an Illumina Nextseq500 instrument using a 75-cycle high output kit. The run was setup as a single 85 bp read and an index read when more than one library was pooled. We regularly pool up to 16 Census-Seq samples in one sequencing run.

**Validation of primary CRISPR-Cas9 fitness screen**—To validate the genetic drivers of SNaP proliferation identified in the SNaP fitness screen, SW7388-1 SNaPs were transduced with Cas9-lentivirus (pLX-311-Cas9 vector) via spinoculation in 24-well plates followed by expansion and selection with blasticidin (10 µg/mL) for one week. SNaPs were then transduced with individual lentivirus gRNAs (pXPR\_003 and pXPR\_050 vectors) in the 24-well plate format, before expansion and selection with puromycin (1 µg/mL) for one week. Cas9 gRNA-expressing SNaPs were passaged and plated onto Geltrex at 1,000 cells per well of a 96-well plate (3,333 cells/cm<sup>2</sup>). The following day (Day 1), Hoechst-33342 dye (1:2000 in base media) was added to 1/2 of the wells before Cytation 3 cell imaging multi-mode reader (4X objective; 3 x 3 grid). The dye and imaging process was repeated 9 days later (Day 10) for the remaining wells. All images were then processed using the CellProfiler imaging analysis software to quantify the number of Hoechst-positive cells. Data is presented as calculated doubling rate for each gRNA line using the following equation:

$$\text{Doubling time} = \text{Duration} / \log_2 \left( \frac{\text{cell count at Day 10}}{\text{cell count at Day 1}} \right)$$

where *Duration* refers to time between measurements in hours (216 hours). For all experiments, the average number of cells for a given sgRNA was used as the denominator in the log<sub>2</sub> calculation (i.e. *cell count at Day 1*). Doubling rates were compared to Cas9-SNaPs that were transduced with non-targeting gRNAs.

**Cerebral organoid formation**—Stem cell-derived cerebral organoids were generated as previously described<sup>97</sup>: H1-Cas9 hESCs were dissociated in Accutase and plated at 18,000 cells/well of an Ultra-Low Attachment 96-well Round Bottom plate (Corning, 7007) in 150 µL of mTeSR maintenance media supplemented with Y-27632 (50 µM). Two days later (Day 2), 75 µL of conditioned media was removed from each well and replaced with 150 µL of fresh mTeSR maintenance media supplemented with Y-27632 (50 µM). The following day (Day 3), 125 µL of conditioned media was removed from each well using the “blast” technique and replaced with 150 µL of fresh mTeSR maintenance media without Y-27632. From Day 4 onward, 150 µL of conditioned media was removed every other day from each well using the “blast” technique and replaced with 150 µL of fresh Neural Induction media (DMEM/F12 with Glutamax, N2 supplement (1:100), MEM-NEAA (1:100), Heparin (1 µg/ml). Cerebral organoid size was measured using the Cytation 3 cell imaging multi-mode reader using the 4X bright field objective. All images were then processed using the CellProfiler imaging analysis software to stitch the images and quantify the two-dimensional area of each organoid.

**Cerebral organoid immunohistochemistry**—At Day 28 post-plating, cerebral organoids were washed twice in 1X PBS and then fixed in cold 4% PFA for 30 minutes shaking at 4°C. Organoids were then washed three times in 1X PBS and incubated in 1ml of 30% sucrose in a 1.5ml Eppendorf tube for 1 hour at 4°C (or until the organoids settled to the bottom of the tube). Organoids were then transferred to custom-made molds composed of 2-ply aluminum foil, which were then filled with 900µL of a mix of OCT/30%



sucrose. The molds were placed in a slurry of ethanol and dry ice for 5 minutes to freeze before storage at  $-80^{\circ}\text{C}$ . Frozen organoids were sectioned in  $20\ \mu\text{m}$  slices using a ThermoSci HM550 cryostat and placed onto glass slides.

Organoid sections were then fixed for 5 minutes at room temperature with 4% PFA before a 3X wash in 1X PBS. Sections were blocked and permeabilized (50 mL 10% Donkey Serum in 1X PBS + 0.38g glycine + 150  $\mu\text{L}$  Triton X-100) for 1 hour at room temperature in a humidified chamber. Organoids were stained with the following primary antibodies overnight at  $4^{\circ}\text{C}$  diluted in 1X PBS with 1% Donkey Serum: Mouse anti-SOX2 (1:100; R&D Systems, MAB2018), Rabbit anti-TBR1 (1:200; Cell Signaling Technology 49661S), Rabbit anti-MKI67 (1:500; ThermoFisher, MA5-14520). Sections were washed three times for 5 minutes each with 1X PBS + 0.05% Triton X-100 followed by incubation for 2-3 hours at room temperature in 1% Donkey Serum + 1:1000 DAPI + the following secondary antibodies: Donkey anti-mouse Alexa488 (1:1000; Life Technologies A-21202) and Donkey anti-rabbit Alexa555 (1:1000; Life Technologies A-31572). Sections were washed 3x5 minutes in 1X PBS + 0.05% Triton X-100 followed by two more washes in 1X PBS. Coverslips were placed onto each section, mounted with 50-100  $\mu\text{L}$  Fluoromount, and sealed with clear nail polish. Images were captured using a Zeiss LSM 880 confocal microscope and were analyzed using CellProfiler software.

## QUANTIFICATION AND STATISTICAL ANALYSIS

**Dropulation sequence alignment and donor assignment**—The human iPSC village, SNaP Village-21, SNaP Village-37, and SNaP Village-44 raw scRNA-seq data were subjected to the Dropulation pipeline to re-identify the donor of origin for each sequenced cell. Sequence data was demultiplexed and aligned following the standard Drop-Seq protocol<sup>115</sup> and was aligned to the GRCh38 reference and ensembl v89 gene models. Sequencing reads were then filtered to reads that mapped at high quality ( $\text{MQ} \geq 10$ ) to the human genome. Genotypes in VCF files were called against the GRCh38 reference genome.

For Dropulation to perform accurately, input sequencing and VCF data is filtered on a per-run basis. Sequence reads are filtered to high quality mappings ( $\text{MQ} \geq 10$ ) on the autosomes that have not been flagged as PCR duplicates. VCF sites are considered if they meet all of the following criteria: each site passed GATK's Variant Quality Score Calibration (VQSR) filter, had a mean genotype quality (GQ) score  $\geq 20$ , a mean variant read depth (DP)  $\geq 10$ , a call rate  $> 50\%$ , a Hardy Weinberg Equilibrium p-value  $> 1e^{-3}$ . Variants located in low complexity regions of the genome or in common segmental duplications as annotated by the UCSC genome browser were filtered from the VCF. Individual genotypes with gross allelic imbalances were set to missing and excluded as defined by the following criteria: an allele balance  $\leq 0.25$  for heterozygous sites and  $\leq 0.9$  for homozygous reference and homozygous alternate sites. Samples with a call rate  $< 90\%$  and a mean depth  $< 10$  were removed from the VCF. Additionally, the Dropulation algorithm retains sites that: a) have a GQ score of at least 30 b) are diploid and polymorphic in the subset of donors in the population c) at least 50% of donors have a GQ score  $\geq 30$ . Furthermore, variants on the X, Y, and MT contigs were ignored. For genotype array-based data where site quality scores may not be available, sites where the reference base is ambiguous [A/T, C/G] were not considered.



The Dropulation algorithm analyzes each cell in the data set independently and generates a likelihood of the data having been generated by each of the donors in the VCF (or a subset of them as requested.) At each variant site, the probability of observing the allele at each unique molecular identifier (UMI) for the site is calculated as the probability of the base at that site for the mode observed base, and 1 - probability for reads that disagree with the mode UMI base. This downweights transcripts where the underlying reads disagree on the observed allele. The likelihood of donor is then computed as the diploid likelihood at each UMI, summed across all sites. The diploid genotype is the average of the two haploid genotypes. For homozygous genotypes, this is the same as the haploid genotype, where if the observed base matches the genotype of the donor, the likelihood is 1 - error rate of the UMI. For heterozygous sites, the likelihood is 0.5, regardless of base quality. The probability of the donor is then calculated as the probability of the donor divided by the sum of all donor probabilities.

**Dropulation missing data handling**—As the number of donors in a VCF file increases, the likelihood that at least one donor will not be called at high quality at any genotyping site increases. One way to deal with missing data is to ignore sites with any missing data, but this can exclude a large number of sites. Instead, we filter sites where the majority of donors are missing data, then for other sites missing data we use the remaining members of the population to calculate a per-site likelihood penalty score to use for all donors that have no genotype data. This score is an extension of the donor assignment score, where the likelihood of each genotype class is calculated, then combined as a weighted average score. This replaces the diploid genotype score for each UMI observed. The mixture coefficient is the proportion of the population that has each genotype class in the population, and sums to one.

The diploid likelihood for a single variant site:

$$\Pr\{D|G\} = \prod_j \frac{\Pr\{D_j|H_1\} + \Pr\{D_j|H_2\}}{2}$$

The haploid likelihood

$$\Pr\{D_j|H\} = 1 - e_j \text{ if } D_j = H \text{ else } e_j$$

The missing data penalty for a single UMI

$$PS = \sum_{i \in \{AA, AB, BB\}} \Pr\{D|G_i\} \times M(G_i)$$

D=The list of UMI bases at the site

G=The genotype of the donor at that site.

H1,H2=The haploid genotype

$e_j$ = The error rate of the observed UMI at base  $D_j$ .

PS=The penalty score for missing data at a site

M=mixture coefficient [proportion of genotype in population]

**Dropulation doublet detection**—Doublet Detection uses the same read and variant filtering as the donor assignment algorithm, with the exception of missing data, where only sites with at least 90% complete data are accepted. The Doublet detection algorithm analyzes each cell in the data set independently and generates a likelihood of the data having been generated by each possible pair of donors. To limit the number of possible tests, doublet detection is more restricted than donor assignment. The first donor of the pair is fixed as the most likely donor based on the single donor assignment, and the second donor of the pair is limited to a set of donors expected in the experiment. This limits the number of combinations to [number of donors – 1] tests per cell.

For each donor pair, we optimize the mixture component of donor 1 to donor 2 to maximize the likelihood of that donor pair. The mixture score is the fraction of the data that arises from the first donor of the pair and is bounded to [0.8-0.2]. If the mixture score is unbounded, sequencing errors, ambient RNA, and genotyping errors will almost always generate mixtures of two donors that are very close to one, with a higher likelihood than the single donor likelihood, resulting in most cells being classified as doublets.

To select the donor pair that best explains the data, we first calculate the maximum likelihood each donor pair by selecting the maximum likelihood of the optimal mixture, the likelihood of the pair with a mixture of 1 (all data arises from donor 1) and the likelihood of the pair with a mixture of 0 (all data arises from donor 2). The donor pair with maximum likelihood is then selected as the best pair.

To classify the pair as a singlet or doublet, we calculate the probability of the data being a doublet as the doublet likelihood divided by the sum of the doublet likelihood and mixture=1,0 likelihoods. We classify cells as doublets if their probability is greater  $\geq$  0.9. The vast majority of doublet probabilities are bimodally distributed at approximately 1 and 0.

Doublet Likelihood

$$Pr\{D|S_1, S_2\} = Pr\{D|G_1\} * M + Pr\{D|G_2\} * (1 - M)$$

$S_1, S_2$ = Donor 1, Donor 2

$G_1, G_2$ =Genotype Donor 1, Genotype Donor 2

M=Fraction of data arising from donor 1

**Dropulation assignable single donors**—To perform downstream analysis at a donor level, the set of cell barcodes in the experiment need to be filtered to the subset where

cells are assigned to a single donor confidently. Cell barcodes are assigned to donors via single donor assignment. Doublet detection is then run, and all cells that are likely doublets (p-value > 0.9) are filtered from the data set. Cell barcodes are filtered if the single donor assignment p-value > 0.05. Given the remaining cells, the relative proportions of each donor are validated to determine if there are significant numbers of cells assigned to donors in the genotype backbone, but not expected in the experiment. Donors with very few assigned cells (less than 0.2%) are removed from the experiment.

The Dropulation software was able to analyze the 104-donor hiPSC village dataset (25,080 cells sequenced to 1,755 average UMIs with 459M total reads) and assign cells to 142 possible donors in 1.72 hours using 512,235 SNPs, and detect doublets in an additional 1.4 hours, using 8g of memory on a single core processor.

**Single cell expression analysis of iPSC village**—The digital gene expression matrices were normalized, and variable gene selection was performed<sup>116</sup>. Clusters were identified using independent component analysis (ICA) based dimension reduction and Louvain community detection algorithms<sup>117</sup>.

**Differential gene expression and Geneset enrichment**—To detect sex-biased or cell source dependent gene expression, we summed the UMI counts of all assignable single cells per donor to generate a donor by gene matrix. Differential expression was run using voom-limma<sup>118</sup> while adjusting for covariates including age and cell source where applicable and additional surrogate variables determined by smartSVA<sup>119</sup>. Gene set enrichment was performed using the C2 (literature curated) and C5 (Gene Ontology Annotations) available in the Molecular Signatures Database (MSigDB) and CAMERA<sup>120</sup> on the list of genes ranked by the voom t-statistics.

**Differentiation scRNA-seq**—The 10X Cell Ranger 1.3.1 pipeline was utilized to convert raw BCL files to cell-gene matrices. The Illumina bcl2fastq script conducted the initial demultiplexing. FASTQ files were then aligned, UMI-filtered, and barcodes were matched via the CellRanger count script. The GRCh37.75 human reference genome was used for alignment. After filtering out barcodes with very few matching transcripts, a total of 2,167 SNaP-derived cells were adequately sequenced. An average of 144,318 reads were mapped per neurosphere cell and 207,402 reads were mapped per SNaP-derived monolayer cell. Resulting scRNA-seq datasets were analyzed in R using Seurat2. Cell-gene matrices were log normalized, and cells with >10% mitochondrial reads of >7000 unique genes were filtered out to reduce the number of dead or doublet cells within the dataset. Variable genes were identified and used to determine the top 15 principal components, which were used for the subsequent analysis. Graph-based clustering was used to approximate different cell groups, and t-stochastic neighborhood embedding (TSNE) analysis used for 2-dimensional representation. Differential expression between clusters was determined by the Wilcoxon rank sum test. Cells were compared to *in vivo* cell types as described below.

**SNaP cell type classification**—scRNA-seq based comparisons between the *in vitro* SNaPs and *in vivo* human brain cells were conducted using the Seurat 3.0 R package<sup>36</sup>. First, a custom script was composed based on the “Multiple Dataset Integration and

Label Transfer: Reference-based” vignette (<http://sajitalab.org/seurat/v3.1/integration.html>; accessed July 15, 2019). Then, gene expression matrices from SNaPs were merged with two reference datasets: 257 cells from 16-18 week post-conception (wpc) fetal and 21- to 63-year-old adult brain tissue<sup>34</sup> and 3396 cells from 5.85 wpc to 37 wpc fetal brain tissue<sup>35</sup>. Similar to a recent report<sup>121</sup>, we condensed the large number of cell types identified in these reference datasets into 8 groups: Fetal Astrocyte (Nowakowski: “Astrocyte”), Fetal Excitatory Neuron (Nowakowski: “EN-PFC-1”, “EN-PFC-2”, “EN-PFC-3”, “EN-V1-1”, “EN-V1-2”, “EN-V1-3”, “nEN-early-1”, “nEN-early-2”, “nEN-late”), Fetal Inhibitory Neuron (Nowakowski: “nIN-1”, “nIN-2”, “nIN-3”, “nIN-4”, “nIN-5”, “IN-CTX-CGE-1”, “IN-CTX-CGE-2”, “IN-CTX-MGE-1”, “IN-CTX-MGE-2”), Intermediate Progenitor Cell (Nowakowski: “IPC-div1”, “IPC-div2”, “MGE-IPC-1”, “MGE-IPC-2”, “MGE-IPC-3”, “IPC-nEN-1”, “IPC-nEN-2”, “IPC-nEN-3”), Neural Progenitor Cells (Darmanis: “fetal replicating”; Nowakowski: “RG-div1”, “RG-div2”, “RG-early”, “vRG”, “rRG”, “oRG”, “MGE-RG-1”, “MGE-RG-2”, “MGE-div”), Oligodendrocytes (Darmanis: “oligodendrocytes”), Postnatal Astrocyte (Darmanis: “astrocytes”), and Postnatal Neuron (Darmanis: “neurons”).

After inputting the merged gene expression matrices with the condensed cell identifier metadata, the merged dataset was split into a list with each dataset as an element using the `CreateSeuratObject` function (`min.cells = 3`). Log-normalization was then performed, and variable features were identified using the `NormalizeData` and `FindVariableFeatures` (`selection.method = “vst”; nfeatures = 2000`) functions, respectively. Anchors between the individual datasets were calculated using the `FindIntegrationAnchors` (`dims = 1:30`). The `IntegrateData` function (`dims = 1:30`) was deployed to create a batch-corrected expression matrix for all reference cells. This integrated expression matrix was then used as the reference to which the SNaP expression matrices (“query”) were compared using the `FindTransferAnchors`, `TransferData`, and `AddMetaData` functions (`dims = 1:30`). The output included cell predictions and prediction scores for each SNaP cell.

For Figure 2g–i, this analysis was repeated using only the human fetal NPC class of cells in the Nowakowski dataset (outer radial glia, ventricular radial glia, truncated radial glia, dividing radial glia-1, dividing radial glia-2, and neuroepithelial radial glia) plus the adult cells from the Darmanis dataset as the reference.

**eQTL discovery**—For the set of assignable single donor cells, the UMI counts across cells of the same donor and gene are summed to a single measurement to generate a donor by gene expression matrix. Genes on the Y and MT chromosomes are filtered out, as are gene symbols that have ambiguous genomic mappings. Gene expression is then normalized to be fractional by dividing the gene expression of each gene/donor to the sum of expression for all genes. This fractional representation is then multiplied by a fixed constant of 100,000. Finally, genes are filtered to the top 50% highest expressed genes for eQTL discovery.

Variants are included for analysis if they pass all of the following filters. At least 90% of donors must have a genotype that was called and has a genotype quality  $\geq 30$ . The minor allele frequency of the variant must be between 5% and 95% in the population. The variant Hardy-Weinberg equilibrium (HWE) p-value must be  $> 1e-4$ . Finally, the variant must be

within 10kb of the start or end of the gene for which it is tested. Variant genotypes are encoded by the number of alternate alleles.

The matrix of normalized expression data and genotype matrix per donor is then encoded in the format required by the R package MatrixEQTL<sup>37</sup>. The expression data was then corrected for latent batch effects using PEER<sup>122</sup>. MatrixEQTL was then used on the corrected expression data to generate empiric p-values for all variant/gene interactions. False discovery rate (FDR) is then controlled hierarchically at two levels. At the gene level, the SNP with the best p-value is selected as the index SNP, and FDR is controlled by using the R package eigenMT<sup>123</sup>, which uses the linkage disequilibrium of SNPs to determine the number of independent tests within a gene. The distribution of index SNP p-values is then transformed into q-values via the R package qvalue. We consider all genes with a q-value < 0.05 to be eGenes.

eQTLs were compared to public GWAS datasets to identify overlapping loci using the SNPnexus online portal<sup>124</sup>. GWAS plots were generated using Enigma-Vis<sup>125</sup> (Figure 4e) or LocusZoom<sup>126</sup> (Figure 4f).

**Genome-wide CRISPR-Cas9 screens**—Gene-level analysis was executed using RSA<sup>125</sup> and BAGEL<sup>127</sup>. For RSA analysis, DESeq2<sup>128</sup> was used to generate log<sub>2</sub> fold change from gRNA read counts. Subsequently, z scores were computed in R Studio and RSA scores were generated. For additional significance thresholding, Benjamini Hotchberg correction was performed on RSA values which were plotted against Quantile 3 (Q3) and Quantile 1 (Q1) values. BAGEL computations were performed using CRISPRAnalyzer<sup>129</sup> with read counts as the input and the pre-set “Brunello” library. For this analysis, gRNAs with fewer than 20 reads were eliminated from the analysis pipeline. For ZIKV survival screens, Day 10 infected samples were compared to Day 10 mock samples. For fitness screens, Day 10 mock samples were compared to Day 0 mock samples.

**Census-seq**—The Census-seq analytical pipeline was executed<sup>19</sup>: DNA sequencing output was run through an alignment protocol using the Picard tools ExtractIlluminaBarcodes and IlluminaBasecallsToSam. The de-multiplexed libraries were then aligned to a human reference genome with BWA. Prior to running the Census-seq algorithm, VCF files were processed to filter variants and add additional site-level information. Variants were first normalized to their appropriate reference sequence using BCFTools. Variants that were monomorphic were dropped, as well as those without a PASS filter, where the site was flagged as problematic during VCF generation. Sites without rsID annotations were updated using information from dbSNP when possible, and otherwise site names were changed to chromosome:position:ref\_allele:alt\_allele.

Input sequencing and VCF data was then filtered on a per-run basis. Sequence reads were filtered to high quality mappings (MQ $\geq$ 10) on the autosomes that were not flagged as PCR duplicates. VCF sites were considered if they met all of the following criteria: each site has GQ score of at least 30, is a diploid site, is polymorphic in the subset of donors in the population, and at least 90% of donors have a genotype quality score  $\geq$ 30. In addition, for genotype array-based data where site quality scores may not be available, sites where the

reference base is ambiguous [A/T, C/G] were not considered. Only variant sites with ~5% allele frequency were included in analysis. A matrix of donor genotypes and the counts of the reference and alternate allele at each variant were generated. The algorithm initializes with the donor proportions set to equal values (1/number of donors), then runs through an estimation maximization (EM) procedure. The allele frequency of each site is calculated from the genotypes of the donors and their relative proportion in the pool. The initial likelihood of the sequencing data given the starting donor ratios is calculated at each SNP by the likelihood function and the results summed across all sites. To determine how to change the donor ratios to explain the data, an adjustment term is calculated for every donor/site, and the results summed across sites for each donor. This adjustment factor is then scaled by an additional parameter and added to each donor's representation. To determine this scaling value the algorithm employs a univariate optimizer to maximize the donor likelihood. The adjustment is then applied to the data, and the algorithm repeats the adjustment/likelihood optimization loop until convergence.

***In vitro* genome-wide association**—Whole-genome sequence data for the 36 donors were jointly-called along with other CIRM-donor whole-genomes to generate a joint VCF. The joint VCF was filtered to use only high-quality variants for genome-wide analysis. In particular, we required that variants are bi-allelic, pass VQSR, have a combined read depth >10 across samples, have < 1% chance of being incorrectly called, are present in more than half the samples in the VCF, have a probability of deviating from Hardy-Weinberg Equilibrium less than 0.001, and that have minor allele frequency of 0.1 or greater. After choosing the final set of variants that we were going to test, we tested for a linear relationship between what allele a donor had for each variant and the percent of their cells that were infected by each of the Zika virus strains. We used in-house R code that leveraged the MatrixQTL package to regress each variant against ZIKV-Ug and ZIKV-PR infectivity. To compute empirical p-values for these associations, we ran adaptive permutation testing for up to 10<sup>10</sup> iterations.

**CRISPR-Cas9 fitness screen disease gene enrichment**—The list of SNaP screen proliferation hits was compared to disease lists curated from various sources. The ASD list was downloaded from the Simons Foundation Autism Research Initiative (SFARI) website (<https://gene.sfari.org/database/human-gene/>; accessed July 18, 2019). The cancer gene census was downloaded from the COSMIC website (<https://cancer.sanger.ac.uk/census/>; accessed July 18, 2019). Statistical significance of overlap was determined by Fisher's exact test calculated using the GeneOverlap package in R. Gene sets were analyzed for GO term statistical overrepresentation using the PANTHER Classification System (<http://patherdb.org>) with default settings. The *GO biological process complete* annotation dataset was used for the RSA-enriched proliferation gene list, while *PANTHER GO-Slim biological process* annotation dataset was used for the BAGEL (BF>10) fitness gene set. Fisher's Exact test was used for statistical analysis. Proliferation and fitness gene sets were analyzed for KEGG pathway enrichment using the Gene Set Enrichment Analysis (<http://software.broadinstitute.org/gsea>).



**Data analysis**—All data were analyzed and plotted using the Prism version 7.03 software (GraphPad; La Jolla, CA) or R 3.5.3.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGEMENTS

This work was supported by NIH/NIMH grants U01MH105669 and U01MH115727, the Stanley Center for Psychiatric Research at the Broad Institute, and the Harvard University Faculty of Arts and Sciences Dean's Competitive Fund. M.F.W. is supported by the Burroughs Wellcome Fund (1018707) and the NIMH (K99MH119327). We thank Janell Smith, Dr. Martin Berryer, Olivia Bare, Adam Brown, and David Root (Broad Institute) for their assistance in cell cultures and CRISPR screen design.

## REFERENCES

- Hemmati HD, Nakano I, Lazareff JA, Masterman-Smith M, Geschwind DH, Bronner-Fraser M, and Kornblum HI (2003). Cancerous stem cells can arise from pediatric brain tumors. *Proceedings of the National Academy of Sciences* 100, 15178–15183. 10.1073/pnas.2036535100.
- Tang H, Hammack C, Ogden SC, Wen Z, Qian X, Li Y, Yao B, Shin J, Zhang F, Lee EM, et al. (2016). Zika virus infects human cortical neural progenitors and attenuates their growth. *Cell Stem Cell* 18, 587–590. 10.1016/j.stem.2016.02.016. [PubMed: 26952870]
- Belmont JW, Boudreau A, Leal SM, Hardenbol P, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Gao Y, et al. (2005). A haplotype map of the human genome. *Nature* 437, 1299–1320. 10.1038/nature04226. [PubMed: 16255080]
- Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, Clark AG, Donnelly P, Eichler EE, et al. (2015). A global reference for human genetic variation. *Nature* 526, 68–74. 10.1038/nature15393. [PubMed: 26432245]
- Aguet F, Brown AA, Castel SE, Davis JR, He Y, Jo B, Mohammadi P, Park Y, Parsana P, Segrè AV, et al. (2017). Genetic effects on gene expression across human tissues. *Nature* 550, 204–213. 10.1038/nature24277. [PubMed: 29022597]
- DeBoever C, Li H, Jakubosky D, Benaglio P, Reyna J, Olson KM, Huang H, Biggs W, Sandoval E, D'Antonio M, et al. (2017). Large-Scale Profiling Reveals the Influence of Genetic Variation on Gene Expression in Human Induced Pluripotent Stem Cells. *Cell Stem Cell* 20, 533–546.e7. 10.1016/j.stem.2017.03.009. [PubMed: 28388430]
- Kammers K, Taub MA, Rodriguez B, Yanek LR, Ruczinski I, Martin J, Kanchan K, Battle A, Cheng L, Wang ZZ, et al. (2021). Transcriptional profile of platelets and iPSC-derived megakaryocytes from whole-genome and RNA sequencing. *Blood* 137, 959–968. 10.1182/blood.2020006115. [PubMed: 33094331]
- Aygun N, Elwell AL, Liang D, Lafferty MJ, Cheek KE, Courtney KP, Mory J, Hadden-Ford E, Krupa O, de la Torre-Ubieta L, et al. (2021). Brain-trait-associated variants impact cell-type-specific gene regulation during neurogenesis. *Am. J. Hum. Genet* 108, 1647–1668. 10.1016/j.ajhg.2021.07.011. [PubMed: 34416157]
- Liang D, Elwell AL, Aygun N, Krupa O, Wolter JM, Kyere FA, Lafferty MJ, Cheek KE, Courtney KP, Yusupova M, et al. (2021). Cell-type-specific effects of genetic variation on chromatin accessibility during human neuronal differentiation. *Nat. Neurosci.* 24, 941–953. 10.1038/s41593-021-00858-w. [PubMed: 34017130]
- Elkabetz Y, Panagiotakos G, Al Shamy G, Socci ND, Tabar V, and Studer L (2008). Human ES cell-derived neural rosettes reveal a functionally distinct early neural stem cell stage. *Genes Dev.* 22, 152–165. 10.1101/gad.1616208. [PubMed: 18198334]
- Koch P, Opitz T, Steinbeck JA, Ladewig J, and Brüstle O (2009). A rosette-type, self-renewing human ES cell-derived neural stem cell with potential for in vitro instruction and synaptic

- integration. *Proc. Natl. Acad. Sci. U. S. A* 106, 3225–3230. 10.1073/pnas.0808387106. [PubMed: 19218428]
12. Zhang S-C, Wernig M, Duncan ID, Brüstle O, and Thomson JA (2001). In vitro differentiation of transplantable neural precursors from human embryonic stem cells. *Nature Biotechnology* 19, 1129–1133. 10.1038/nbt1201-1129.
  13. Chambers SM, Fasano CA, Papapetrou EP, Tomishima M, Sadelain M, and Studer L (2009). Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nat. Biotechnol.* 27, 275–280. 10.1038/nbt.1529. [PubMed: 19252484]
  14. Kelava I, and Lancaster MA (2016). Stem Cell Models of Human Brain Development. *Cell Stem Cell* 18, 736–748. 10.1016/j.stem.2016.05.022. [PubMed: 27257762]
  15. Muratore CR, Srikanth P, Callahan DG, and Young-Pearse TL (2014). Comparison and optimization of hiPSC forebrain cortical differentiation protocols. *PLoS One* 9, e105807. 10.1371/journal.pone.0105807. [PubMed: 25165848]
  16. Engel M, Do-Ha D, Muñoz SS, and Ooi L (2016). Common pitfalls of stem cell differentiation: a guide to improving protocols for neurodegenerative disease models and research. *Cellular and Molecular Life Sciences* 73, 3693–3709. 10.1007/s00018-016-2265-3. [PubMed: 27154043]
  17. Nehme R, Zuccaro E, Ghosh SD, Li C, Sherwood JL, Pietilainen O, Barrett LE, Limone F, Worringer KA, Kommineni S, et al. (2018). Combining NGN2 Programming with Developmental Patterning Generates Human Excitatory Neurons with NMDAR-Mediated Synaptic Transmission. *Cell Rep.* 23, 2509–2523. 10.1016/j.celrep.2018.04.066. [PubMed: 29791859]
  18. Zhang Y, Pak C, Han Y, Ahlenius H, Zhang Z, Chanda S, Marro S, Patzke C, Acuna C, Covy J, et al. (2013). Rapid single-step induction of functional neurons from human pluripotent stem cells. *Neuron* 78, 785–798. 10.1016/j.neuron.2013.05.029. [PubMed: 23764284]
  19. Mitchell JM, Nemes J, Ghosh S, Handsaker RE, Mello CJ, Meyer D, Raghunathan K, de Rivera H, Tegtmeyer M, Hawes D, et al. (2020). Mapping genetic effects on cellular phenotypes with “cell villages.” *bioRxiv*, 2020.06.29.174383. 10.1101/2020.06.29.174383.
  20. Joung J, Konermann S, Gootenberg JS, Abudayeh OO, Platt RJ, Brigham MD, Sanjana NE, and Zhang F (2017). Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nat. Protoc.* 12, 828–863. 10.1038/nprot.2017.016. [PubMed: 28333914]
  21. Ophir S, E. SN, Ella H, Xi S, A. SD, S. MT, Dirk H, L. EB, E. RD, G. DJ, et al. (2014). Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science* 343, 84–87. 10.1126/science.1247005. [PubMed: 24336571]
  22. Kang HM, Subramaniam M, Targ S, Nguyen M, Maliskova L, McCarthy E, Wan E, Wong S, Byrnes L, Lanata CM, et al. (2018). Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat. Biotechnol.* 36, 89–94. 10.1038/nbt.4042. [PubMed: 29227470]
  23. Oliva M, Muñoz-aguirre M, Kim-hellmuth S, Wucher V, Gewirtz ADH, Cotter DJ, Parsana P, Kasela S, Balliu B, Viñuela A, et al. (2020). The impact of sex on gene expression across human tissues. *Science* 3066. 10.1126/science.aba3066.
  24. Schafer ST, Paquola ACM, Stern S, Gosselin D, Ku M, Pena M, Kuret TJM, Liyanage M, Mansour AAF, Jaeger BN, et al. (2019). Pathological priming causes developmental gene network heterochronicity in autistic subject-derived neurons. *Nat. Neurosci.* 22, 243–255. 10.1038/s41593-018-0295-x. [PubMed: 30617258]
  25. Hemmati-Brivanlou A, and Melton D (1997). Vertebrate embryonic cells will become nerve cells unless told otherwise. *Cell* 88, 13–17. 10.1016/S0092-8674(00)81853-X. [PubMed: 9019398]
  26. Nadadhur AG, Leferink PS, Holmes D, Hinz L, Cornelissen-Steijger P, Gasparotto L, and Heine VM (2018). Patterning factors during neural progenitor induction determine regional identity and differentiation potential in vitro. *Stem Cell Res.* 32, 25–34. 10.1016/j.scr.2018.08.017. [PubMed: 30172094]
  27. Smith JR, Vallier L, Lupo G, Alexander M, Harris WA, and Pedersen RA (2008). Inhibition of Activin/Nodal signaling promotes specification of human embryonic stem cells into neuroectoderm. *Dev. Biol.* 313, 107–117. 10.1016/j.ydbio.2007.10.003. [PubMed: 18022151]
  28. Gohlke JM, Armant O, Parham FM, Smith MV, Zimmer C, Castro DS, Nguyen L, Parker JS, Gradwohl G, Portier CJ, et al. (2008). Characterization of the proneural gene regulatory network

- during mouse telencephalon development. *BMC Biol.* 6, 15. 10.1186/1741-7007-6-15. [PubMed: 18377642]
29. Karow M, Camp JG, Falk S, Gerber T, Pataskar A, Gac-Santel M, Kageyama J, Brazovskaja A, Garding A, Fan W, et al. (2018). Direct pericyte-to-neuron reprogramming via unfolding of a neural stem cell-like program. *Nat. Neurosci.* 21, 932–940. 10.1038/s41593-018-0168-3. [PubMed: 29915193]
  30. Franco SJ, and Müller U (2013). Shaping Our Minds: Stem and Progenitor Cell Diversity in the Mammalian Neocortex. *Neuron* 77, 19–34. 10.1016/j.neuron.2012.12.022. [PubMed: 23312513]
  31. Kriegstein A, and Alvarez-Buylla A (2009). The Glial Nature of Embryonic and Adult Neural Stem Cells. *Annu. Rev. Neurosci.* 32, 149–184. 10.1146/annurev.neuro.051508.135600. [PubMed: 19555289]
  32. Nowakowski TJ, Pollen AA, Sandoval-Espinosa C, and Kriegstein AR (2016). Transformation of the Radial Glia Scaffold Demarcates Two Stages of Human Cerebral Cortex Development. *Neuron* 91, 1219–1227. 10.1016/j.neuron.2016.09.005. [PubMed: 27657449]
  33. Tcw J, Wang M, Pimenova AA, Bowles KR, Hartley BJ, Lacin E, Machlovi SI, Abdelaal R, Karch CM, Phatnani H, et al. (2017). An Efficient Platform for Astrocyte Differentiation from Human Induced Pluripotent Stem Cells. *Stem Cell Reports* 9, 600–614. 10.1016/j.stemcr.2017.06.018. [PubMed: 28757165]
  34. Darmanis S, Sloan SA, Zhang Y, Enge M, Caneda C, Shuer LM, Hayden Gephart MG, Barres BA, and Quake SR (2015). A survey of human brain transcriptome diversity at the single cell level. *Proc. Natl. Acad. Sci. U. S. A.* 112, 7285–7290. 10.1073/pnas.1507125112. [PubMed: 26060301]
  35. Nowakowski TJ, Bhaduri A, Pollen AA, Alvarado B, Mostajo-Radji MA, Di Lullo E, Haeussler M, Sandoval-Espinosa C, Liu SJ, Velmeshev D, et al. (2017). Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science* 358, 1318 LP–1323. 10.1126/science.aap8809. [PubMed: 29217575]
  36. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, Hao Y, Stoeckius M, Smibert P, and Satija R (2019). Comprehensive Integration of Single-Cell Data. *Cell* 177, 1888–1902.e21. 10.1016/j.cell.2019.05.031. [PubMed: 31178118]
  37. Shabalin AA (2012). Matrix eQTL: Ultra fast eQTL analysis via large matrix operations. *Bioinformatics* 28, 1353–1358. 10.1093/bioinformatics/bts163. [PubMed: 22492648]
  38. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, Hasz R, Walters G, Garcia F, Young N, et al. (2013). The Genotype-Tissue Expression (GTEx) project. *Nat. Genet* 45, 580–585. 10.1038/ng.2653. [PubMed: 23715323]
  39. Werling DM, Pochareddy S, Choi J, An J-Y, Sheppard B, Peng M, Li Z, Dastmalchi C, Santpere G, Sousa AMM, et al. (2020). Whole-Genome and RNA Sequencing Reveal Variation and Transcriptomic Coordination in the Developing Human Prefrontal Cortex. *Cell Rep.* 31, 107489. 10.1016/j.celrep.2020.03.053. [PubMed: 32268104]
  40. Grasby KL, Jahanshad N, Painter JN, Colodro-Conde L, Bralten J, Hibar DP, Lind PA, Pizzagalli F, Ching CRK, McMahon MAB, et al. (2020). The genetic architecture of the human cerebral cortex. *Science* 367, eaay6690. 10.1126/science.aay6690. [PubMed: 32193296]
  41. Ripke S, Neale BM, Corvin A, Walters JTR, Farh KH, Holmans PA, Lee P, Bulik-Sullivan B, Collier DA, Huang H, et al. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511, 421–427. 10.1038/nature13595. [PubMed: 25056061]
  42. Wong JM, Folorunso OO, Barragan EV, Berciu C, Harvey TL, Coyle JT, Balu DT, and Gray JA (2020). Postsynaptic Serine racemase regulates NMDA receptor function. *J. Neurosci* 40, 9564–9575. 10.1523/JNEUROSCI.1525-20.2020. [PubMed: 33158959]
  43. Lowe R, Barcellos C, Brasil P, Cruz OG, Honório NA, Kuper H, and Carvalho MS (2018). The Zika Virus Epidemic in Brazil: From Discovery to Future Implications. *Int. J. Environ. Res. Public Health* 15. 10.3390/ijerph15010096.
  44. Brasil P, Pereira JP Jr., Moreira ME, Ribeiro Nogueira RM, Damasceno L, Wakimoto M, Rabello RS, Valderramos SG, Halai U-A, Salles TS, et al. (2016). Zika Virus Infection in Pregnant Women in Rio de Janeiro. *N. Engl. J. Med.* 375, 2321–2334. 10.1056/nejmoa1602412. [PubMed: 26943629]

45. Cauchemez S, Besnard M, Bompard P, Dub T, Guillemette-Artur P, Eyrolle-Guignot D, Salje H, Van Kerkhove MD, Abadie V, Garel C, et al. (2016). Association between Zika virus and microcephaly in French Polynesia, 2013-15: A retrospective study. *Lancet*. 10.1016/S0140-6736(16)00651-6.
46. Mlakar J, Korva M, Tul N, Popovi M, Poljšak-Prijatelj M, Mraz J, Kolenc M, Resman Rus K, Vesnaver Vipotnik T, Fabjan Vodušek V, et al. (2016). Zika Virus Associated with Microcephaly. *N. Engl. J. Med* 374, 951–958. 10.1056/NEJMoa1600651. [PubMed: 26862926]
47. Nielsen-saines K, Brasil P, Kerin T, Vasconcelos Z, Gabaglia CR, Damasceno L, Pone M, Carvalho LMAD, Pone SM, Zin AA, et al. (2019). Delayed childhood neurodevelopment and neurosensory alterations in the second year of life in a prospective cohort of ZIKV-exposed children. *Nat. Med* 25. 10.1038/s41591-019-0496-1.
48. de Oliveira WK, de França GVA, Carmo EH, Duncan BB, de Souza Kuchenbecker R, and Schmidt MI (2017). Infection-related microcephaly after the 2015 and 2016 Zika virus outbreaks in Brazil: a surveillance-based analysis. *Lancet* 390, 861–870. 10.1016/S0140-6736(17)31368-5. [PubMed: 28647172]
49. Barbeito-Andrés J, Pezzuto P, Higa LM, Dias AA, Vasconcelos JM, Santos TMP, Ferreira JCCG, Ferreira RO, Dutra FF, Rossi AD, et al. (2020). Congenital Zika syndrome is associated with maternal protein malnutrition. *Science advances* 6, eaaw6284. 10.1126/sciadv.aaw6284. [PubMed: 31950075]
50. Souza WVD, Albuquerque MDFPMD, Vazquez E, Bezerra LCA, Mendes ADCG, Lyra TM, Araujo TVBD, Oliveira ALSD, Braga MC, Ximenes RADA, et al. (2018). Microcephaly epidemic related to the Zika virus and living conditions in Recife, Northeast Brazil. *BMC Public Health* 18, 1–7. 10.1186/s12889-018-5039-z.
51. Borda V, Junior R. da S.F., Carvalho JB, Morais GL, Rossi ÁD, Pezzuto P, Azevedo GS, Schamber-Reis BL, Portari EA, Melo A, et al. (2021). Whole-exome sequencing reveals insights into genetic susceptibility to congenital zika syndrome. *PLoS Negl. Trop. Dis* 15, 1–17. 10.1371/journal.pntd.0009507.
52. Gomes JA, Sgarioni E, Boquett JA, Terças-Trettel ACP, da Silva JH, Ribeiro BFR, Galera MF, de Oliveira TM, de Andrade MDFC, Carvalho IF, et al. (2021). Association between genetic variants in *nos2* and *tnf* genes with congenital zika syndrome and severe microcephaly. *Viruses* 13, 1–13. 10.3390/v13020325.
53. Santos CNO, Ribeiro DR, Cardoso Alves J, Cazzaniga RA, Magalhães LS, De Souza MSF, Fonseca ABL, Bispo AJB, Porto RLS, Santos CAD, et al. (2019). Association between Zika Virus Microcephaly in Newborns with the rs3775291 Variant in Toll-Like Receptor 3 and rs1799964 Variant at Tumor Necrosis Factor- $\alpha$  Gene. *J. Infect. Dis* 220, 1797–1801. 10.1093/infdis/jiz392. [PubMed: 31352487]
54. Han Y, Tan L, Zhou T, Yang L, Carrau L, Lacko LA, Saeed M, Zhu J, Zhao Z, Nilsson-Payant BE, et al. (2022). A human iPSC-array-based GWAS identifies a virus susceptibility locus in the *NDUFA4* gene and functional variants. *Cell Stem Cell* 29, 1475–1490.e6. 10.1016/j.stem.2022.09.008. [PubMed: 36206731]
55. Retallack H, Di Lullo E, Arias C, Knopp KA, Laurie MT, Sandoval-Espinosa C, Leon WRM, Krencik R, Ullian EM, Spatazza J, et al. (2016). Zika virus cell tropism in the developing human brain and inhibition by azithromycin. *Proc. Natl. Acad. Sci. U. S. A* 113, 14408–14413. 10.1073/pnas.1618029113. [PubMed: 27911847]
56. Simonin Y, Loustalot F, Desmetz C, Foulongne V, Constant O, Fournier-Wirth C, Leon F, Molès JP, Goubaud A, Lemaitre JM, et al. (2016). Zika Virus Strains Potentially Display Different Infectious Profiles in Human Neural Cells. *EBioMedicine* 12, 161–169. 10.1016/j.ebiom.2016.09.020. [PubMed: 27688094]
57. Simonin Y, van Riel D, Van de Perre P, Rockx B, and Salinas S (2017). Differential virulence between Asian and African lineages of Zika virus. *PLoS Negl. Trop. Dis* 11, e0005821. 10.1371/journal.pntd.0005821. [PubMed: 28934211]
58. Mesci P, Macia A, Moore SM, Shiryayev SA, Pinto A, Huang C-T, Tejwani L, Fernandes IR, Suarez NA, Kolar MJ, et al. (2018). Blocking Zika virus vertical transmission. *Sci. Rep* 8, 1218. 10.1038/s41598-018-19526-4. [PubMed: 29352135]

59. König R, Chiang CY, Tu BP, Yan SF, DeJesus PD, Romero A, Bergauer T, Orth A, Krueger U, Zhou Y, et al. (2007). A probability-based approach for the analysis of large-scale RNAi screens. *Nat. Methods* 4, 847–849. 10.1038/nmeth1089. [PubMed: 17828270]
60. Li Y, Muffat J, Omer Javed A, Keys HR, Lungjangwa T, Bosch I, Khan M, Virgilio MC, Gehrke L, Sabatini DM, et al. (2019). Genome-wide CRISPR screen for Zika virus resistance in human neural cells. *Proceedings of the National Academy of Sciences* 116, 9527 LP–9532. 10.1073/pnas.1900867116.
61. Wang S, Zhang Q, Tiwari SK, Lichinchi G, Yau EH, Hui H, Li W, Furnari F, and Rana TM (2020). Integrin  $\alpha\beta 5$  Internalizes Zika Virus during Neural Stem Cells Infection and Provides a Promising Target for Antiviral Therapy. *Cell Rep.* 30, 1–15. 10.1016/j.celrep.2019.11.020. [PubMed: 31914378]
62. Wells MF, Salick MR, Wiskow O, Ho DJ, Worringer KA, Ihry RJ, Kommineni S, Bilican B, Klim JR, Hill EJ, et al. (2016). Genetic Ablation of AXL Does Not Protect Human Neural Progenitor Cells and Cerebral Organoids from Zika Virus Infection. *Cell Stem Cell* 19, 703–708. 10.1016/j.stem.2016.11.011. [PubMed: 27912091]
63. Xu M, Lee EM, Wen Z, Cheng Y, Huang W-K, Qian X, Tcw J, Kouznetsova J, Ogden SC, Hammack C, et al. (2016). Identification of small-molecule inhibitors of Zika virus infection and induced neural cell death via a drug repurposing screen. *Nat. Med* 22, 1101–1107. 10.1038/nm.4184. [PubMed: 27571349]
64. Zhou T, Tan L, Cederquist GY, Fan Y, Hartley BJ, Mukherjee S, Tomishima M, Brennand KJ, Zhang Q, Schwartz RE, et al. (2017). High-Content Screening in hPSC-Neural Progenitors Identifies Drug Candidates that Inhibit Zika Virus Infection in Fetal-like Organoids and Adult Brain. *Cell Stem Cell* 21, 274–283.e5. 10.1016/j.stem.2017.06.017. [PubMed: 28736217]
65. Spence JS, He R, Hoffmann HH, Das T, Thinon E, Rice CM, Peng T, Chandran K, and Hang HC (2019). IFITM3 directly engages and shuttles incoming virus particles to lysosomes. *Nat. Chem. Biol* 15, 259–268. 10.1038/s41589-018-0213-2. [PubMed: 30643282]
66. Savidis G, Perreira JM, Portmann JM, Meraner P, Guo Z, Green S, and Brass AL (2016). The IFITMs Inhibit Zika Virus Replication. *Cell Rep.* 15, 2323–2330. 10.1016/j.celrep.2016.05.074. [PubMed: 27268505]
67. Brass AL, Huang IC, Benita Y, John SP, Krishnan MN, Feeley EM, Ryan BJ, Weyer JL, van der Weyden L, Fikrig E, et al. (2009). The IFITM Proteins Mediate Cellular Resistance to Influenza A H1N1 Virus, West Nile Virus, and Dengue Virus. *Cell* 139, 1243–1254. 10.1016/j.cell.2009.12.017. [PubMed: 20064371]
68. Allen EK, Randolph AG, Bhangale T, Dogra P, Ohlson M, Oshansky CM, Zamora AE, Shannon JP, Finkelstein D, Dressen A, et al. (2017). SNP-mediated disruption of CTCF binding at the IFITM3 promoter is associated with risk of severe influenza in humans. *Nat. Med* 23, 975–983. 10.1038/nm.4370. [PubMed: 28714988]
69. Cederquist GY, Tchieu J, Callahan SJ, Ramnarine K, Ryan S, Zhang C, Rittenhouse C, Zeltner N, Chung SY, Zhou T, et al. (2020). A Multiplex Human Pluripotent Stem Cell Platform Defines Molecular and Functional Subclasses of Autism-Related Genes. *Cell Stem Cell* 27, 35–49.e6. 10.1016/j.stem.2020.06.004. [PubMed: 32619517]
70. Marchetto MC, Belinson H, Tian Y, Freitas BC, Fu C, Vadodaria K, Beltrao-Braga P, Trujillo CA, Mendes APD, Padmanabhan K, et al. (2017). Altered proliferation and networks in neural cells derived from idiopathic autistic individuals. *Mol. Psychiatry* 22, 820–835. 10.1038/mp.2016.95. [PubMed: 27378147]
71. Pucilowska J, Vithayathil J, Tavares EJ, Kelly C, Karlo JC, and Landreth GE (2015). The 16p11.2 Deletion Mouse Model of Autism Exhibits Altered Cortical Progenitor Proliferation and Brain Cytoarchitecture Linked to the ERK MAPK Pathway. *Journal of Neuroscience* 35, 3190–3200. 10.1523/JNEUROSCI.4864-13.2015. [PubMed: 25698753]
72. Hazlett HC, Gu H, Munsell BC, Kim SH, Styner M, Wolff JJ, Elison JT, Swanson MR, Zhu H, Botteron KN, et al. (2017). Early brain development in infants at high risk for autism spectrum disorder. *Nature* 542, 348–351. 10.1038/nature21369. [PubMed: 28202961]
73. Courchesne E, Carper R, and Akshoomoff N (2003). Evidence of brain overgrowth in the first year of life in autism. *JAMA* 290, 337–344. 10.1001/jama.290.3.337. [PubMed: 12865374]



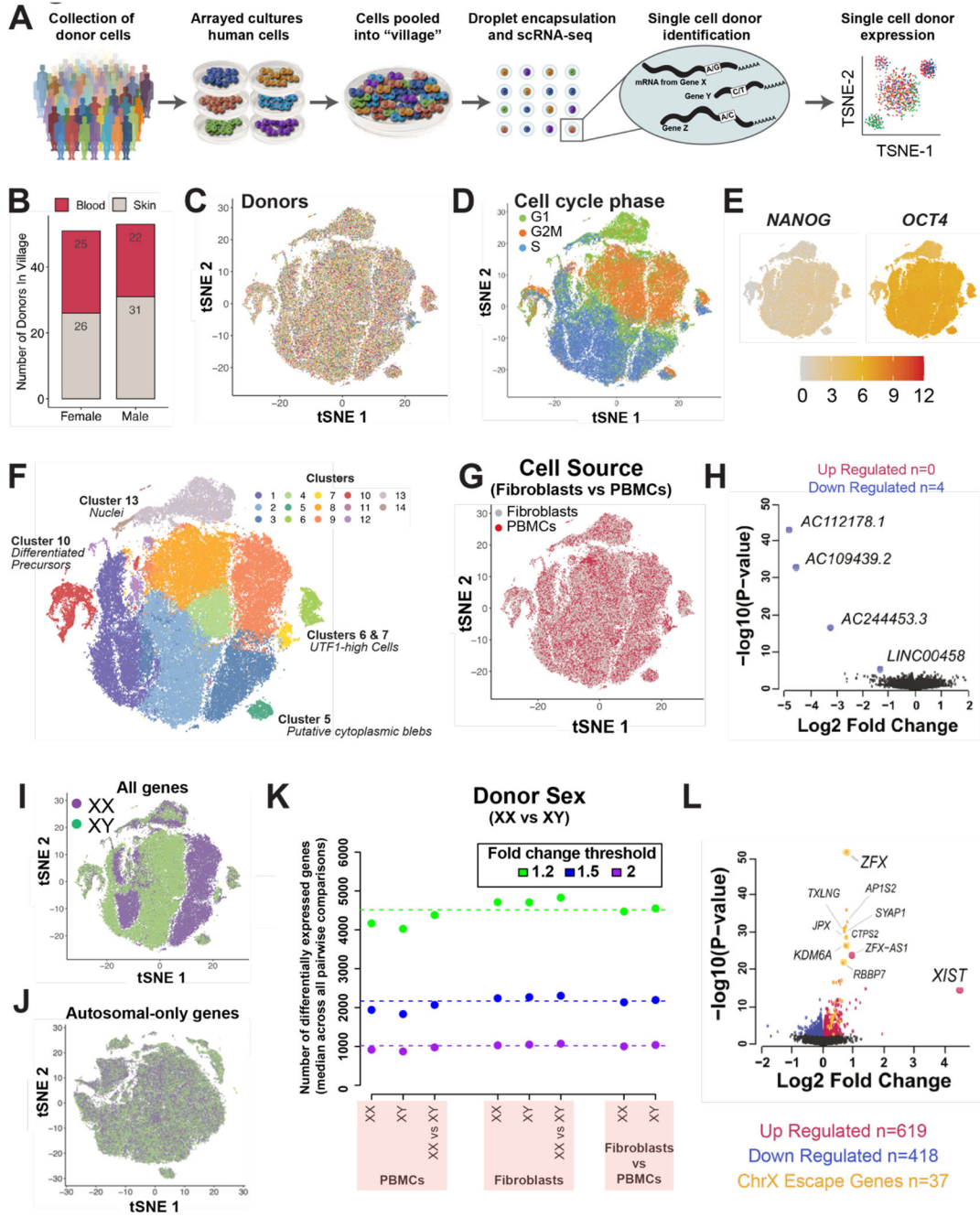
74. McRae JF, Clayton S, Fitzgerald TW, Kaplanis J, Prigmore E, Rajan D, Sifrim A, Aitken S, Akawi N, Alvi M, et al. (2017). Prevalence and architecture of de novo mutations in developmental disorders. *Nature* 542, 433. 10.1038/nature21062. [PubMed: 28135719]
75. Abrahams BS, Arking DE, Campbell DB, Mefford HC, Morrow EM, Weiss LA, Menashe I, Wadkins T, Banerjee-basu S, and Packer A (2013). SFARI Gene 2.0: a community-driven knowledge base for the autism spectrum disorders (ASDs). *Mol. Autism* 4, 2–4. [PubMed: 23347615]
76. Sanders SJ, He X, Willsey AJ, Ercan-Sencicek AG, Samocha KE, Cicek AE, Murtha MT, Bal VH, Bishop SL, Dong S, et al. (2015). Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron* 87, 1215–1233. 10.1016/j.neuron.2015.09.016. [PubMed: 26402605]
77. Satterstrom FK, Kosmicki JA, Wang J, Breen MS, De Rubeis S, An J-Y, Peng M, Collins R, Grove J, Klei L, et al. (2020). Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell* 180, 568–584.e23. 10.1016/j.cell.2019.12.036. [PubMed: 31981491]
78. Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, Cole CG, Ward S, Dawson E, Ponting L, et al. (2017). COSMIC: Somatic cancer genetics at high-resolution. *Nucleic Acids Res.* 45, D777–D783. 10.1093/nar/gkw1121. [PubMed: 27899578]
79. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, Lee JC, Philip Schumm L, Sharma Y, Anderson CA, et al. (2012). Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* 491, 119–124. 10.1038/nature11582. [PubMed: 23128233]
80. Marshall CR, Howrigan DP, Merico D, Thiruvahindrapuram B, Wu W, Greer DS, Antaki D, Shetty A, Holmans PA, Pinto D, et al. (2017). Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nat. Genet* 49, 27–35. 10.1038/ng.3725. [PubMed: 27869829]
81. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, Adams MJ, Agerbo E, Air TM, Andlauer TMF, et al. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet* 50, 668–681. 10.1038/s41588-018-0090-3. [PubMed: 29700475]
82. Bock C, Datlinger P, Chardon F, Coelho MA, Dong MB, Lawson KA, Lu T, Maroc L, Norman TM, Song B, et al. (2022). High-content CRISPR screening. *Nature Reviews Methods Primers* 2, 8. 10.1038/s43586-021-00093-4.
83. Tian R, Gachechiladze MA, Ludwig CH, Laurie MT, Hong JY, Nathaniel D, Prabhu AV, Fernandopulle MS, Patel R, Abshari M, et al. (2019). CRISPR Interference-Based Platform for Multimodal Genetic Screens in Human iPSC-Derived Neurons. *Neuron* 104, 239–255.e12. 10.1016/j.neuron.2019.07.014. [PubMed: 31422865]
84. Tian R, Abarientos A, Hong J, Hashemi SH, Yan R, Dräger N, Leng K, Nalls MA, Singleton AB, Xu K, et al. (2021). Genome-wide CRISPRi/a screens in human neurons link lysosomal failure to ferroptosis. *Nat. Neurosci* 24, 1020–1034. 10.1038/s41593-021-00862-0. [PubMed: 34031600]
85. Ihry RJ, Salick MR, Ho DJ, Sondey M, Kommineni S, Paula S, Raymond J, Henry B, Frias E, Wang Q, et al. (2019). Genome-Scale CRISPR Screens Identify Human Pluripotency-Specific Genes. *Cell Reports* 27, 616–630.e6. 10.1016/j.celrep.2019.03.043. [PubMed: 30970262]
86. Zhao M, Kim P, Mitra R, Zhao J, and Zhao Z (2016). TSGene 2.0: An updated literature-based knowledgebase for Tumor Suppressor Genes. *Nucleic Acids Res.* 44, D1023–D1031. 10.1093/nar/gkv1268. [PubMed: 26590405]
87. Hart T, Chandrashekhar M, Aregger M, Steinhart Z, Brown KR, MacLeod G, Mis M, Zimmermann M, Fradet-Turcotte A, Sun S, et al. (2015). High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell* 163, 1515–1526. 10.1016/j.cell.2015.11.015. [PubMed: 26627737]
88. Mair B, Tomic J, Masud SN, Tonge P, Weiss A, Usaj M, Tong AHY, Kwan JJ, Brown KR, Titus E, et al. (2019). Essential Gene Profiles for Human Pluripotent Stem Cells Identify Uncharacterized Genes and Substrate Dependencies. *Cell Rep.* 27, 599–615.e12. 10.1016/j.celrep.2019.02.041. [PubMed: 30970261]



89. Fiddes IT, Lodewijk GA, Mooring M, Bosworth CM, Ewing AD, Mantalas GL, Novak AM, van den Bout A, Bishara A, Rosenkrantz JL, et al. (2018). Human-Specific NOTCH2NL Genes Affect Notch Signaling and Cortical Neurogenesis. *Cell* 173, 1356–1369.e22. 10.1016/j.cell.2018.03.051. [PubMed: 29856954]
90. Dahimene S, Page KM, Kadurin I, Ferron L, Ho DY, Powell GT, Pratt WS, Wilson SW, and Dolphin AC (2018). The  $\alpha 2\delta$ -like Protein CACHD1 Increases N-type Calcium Currents and Cell Surface Expression and Competes with  $\alpha 2\delta$ -1. *Cell Rep.* 25, 1610–1621.e5. 10.1016/j.celrep.2018.10.033. [PubMed: 30404013]
91. Scala M, Cogné B, Beneteau C, von Hardenberg S, Lim D, Accogli A, Mancardi MM, Torella A, Nobili L, Striano P, et al. (2021). Biallelic loss-of-function variants in CACHD1, encoding an  $\alpha 2\delta$ -like voltage-gated calcium channels regulator, cause a novel syndromic neurodevelopmental condition. In American Society for Human Genetics Virtual Meeting (American Society for Human Genetics).
92. Cottrell GS, Soubrane CH, Hounshell JA, Lin H, Owenson V, Rigby M, Cox PJ, Barker BS, Ottolini M, Ince S, et al. (2018). CACHD1 is an  $\alpha 2\delta$ -like protein that modulates Cav3 voltage-gated calcium channel activity. *Journal of Neuroscience* 38, 9186–9201. 10.1523/JNEUROSCI.3572-15.2018. [PubMed: 30181139]
93. Singh B, Monteil A, Bidaud I, Sugimoto Y, Suzuki T, Hamano S-I, Oguni H, Osawa M, Alonso ME, Delgado-Escueta AV, et al. (2007). Mutational analysis of CACNA1G in idiopathic generalized epilepsy. *Hum. Mutat* 28, 524–525. 10.1002/humu.9491.
94. Singh T, Neale BM, Daly MJ, and Consortium, on B. of T.S.E.M.-A. (schema) (2020). Exome sequencing identifies rare coding variants in 10 genes which confer substantial risk for schizophrenia. medRxiv, 2020.09.18.20192815. 10.1101/2020.09.18.20192815.
95. Strom SP, Stone JL, ten Bosch JR, Merriman B, Cantor RM, Geschwind DH, and Nelson SF (2010). High-density SNP association study of the 17q21 chromosomal region linked to autism identifies CACNA1G as a novel candidate gene. *Mol. Psychiatry* 15, 996–1005. 10.1038/mp.2009.41. [PubMed: 19455149]
96. Weiss N, and Zamponi GW (2020). Genetic T-type calcium channelopathies. *J. Med. Genet* 57, 1. 10.1136/jmedgenet-2019-106163. [PubMed: 31217264]
97. Salick MR, Wells MF, Egan K, and Kaykas A (2017). Modelling Zika Virus Infection of the Developing Human Brain In Vitro Using Stem Cell Derived Cerebral Organoids. *J. Vis. Exp.* 1–10. 10.3791/56404.
98. Albanese A, Swaney JM, Yun DH, Evans NB, Antonucci JM, Velasco S, Sohn CH, Arlotta P, Gehrke L, and Chung K (2020). Multiscale 3D phenotyping of human cerebral organoids. *Sci. Rep* 10, 1–17. 10.1038/s41598-020-78130-7. [PubMed: 31913322]
99. Kim YC, and Jeong BH (2020). Ethnic variation in risk genotypes based on single nucleotide polymorphisms (SNPs) of the interferon-inducible transmembrane 3 (IFITM3) gene, a susceptibility factor for pandemic 2009 H1N1 influenza A virus. *Immunogenetics* 72, 447–453. 10.1007/s00251-020-01188-0. [PubMed: 33174121]
100. Kraemer MUG, Reiner RC, Brady OJ, Messina JP, Gilbert M, Pigott DM, Yi D, Johnson K, Earl L, Marczak LB, et al. (2019). Past and future spread of the arbovirus vectors *Aedes aegypti* and *Aedes albopictus*. *Nature Microbiology* 4, 854–863. 10.1038/s41564-019-0376-y.
101. Zou X, Yuan M, Zhang T, Zheng N, and Wu Z (2021). EVs Containing Host Restriction Factor IFITM3 Inhibited ZIKV Infection of Fetuses in Pregnant Mice through Trans-placenta Delivery. *Mol. Ther* 29, 176–190. 10.1016/j.ymthe.2020.09.026. [PubMed: 33002418]
102. Hellwig K, Geissbuehler Y, Sabidó M, Popescu C, Adamo A, Klinger J, Ornoy A, and Huppke P (2020). Pregnancy outcomes in interferon-beta-exposed patients with multiple sclerosis: results from the European Interferon-beta Pregnancy Registry. *J. Neurol* 10.1007/s00415-020-09762-y.
103. Fode C, Gradwohl G, Morin X, Dierich A, LeMeur M, Golidis C, and Guillemot F (1998). The bHLH protein NEUROGENIN 2 is a determination factor for epibranchial placode-derived sensory neurons. *Neuron* 20, 483–494. 10.1016/S0896-6273(00)80989-7. [PubMed: 9539123]
104. Fode C, Ma Q, Casarosa S, Ang SL, Anderson DJ, and Guillemot F (2000). A role for neural determination genes in specifying the dorsoventral identity of telencephalic neurons. *Genes and Development* 14, 67–80. 10.1101/gad.14.1.67. [PubMed: 10640277]

105. Sommer L, Ma Q, and Anderson DJ (1996). Neurogenins, a Novel Family of atonal-Related bHLH Transcription Factors, are Putative Mammalian Neuronal Determination Genes That Reveal Progenitor Cell Heterogeneity in the Developing CNS and PNS. *Mol. Cell. Neurosci* 241, 221–241.
106. Li Y, Muffat J, Omer A, Bosch I, Lancaster MA, Sur M, Gehrke L, Knoblich JA, and Jaenisch R (2017). Induction of Expansion and Folding in Human Cerebral Organoids. *Cell Stem Cell* 20, 385–396.e3. 10.1016/j.stem.2016.11.017. [PubMed: 28041895]
107. Jerber J, Seaton DD, Cuomo ASE, Kumasaka N, Haldane J, Steer J, Patel M, Pearce D, Andersson M, Bonder MJ, et al. (2021). Population-scale single-cell RNA-seq profiling across dopaminergic neuron differentiation. *Nat. Genet* 53, 304–312. 10.1038/s41588-021-00801-6. [PubMed: 33664506]
108. Lim L, Mi D, Llorca A, and Marín O (2018). Development and Functional Diversification of Cortical Interneurons. *Neuron* 100, 294–313. 10.1016/j.neuron.2018.10.009. [PubMed: 30359598]
109. Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang IH, Friman O, Guertin DA, Chang JH, Lindquist RA, Moffat J, et al. (2006). CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* 7, R100. 10.1186/gb-2006-7-10-r100. [PubMed: 17076895]
110. Langmead B, and Salzberg SL (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. 10.1038/nmeth.1923. [PubMed: 22388286]
111. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. (2008). Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* 9, R137. 10.1186/gb-2008-9-9-r137. [PubMed: 18798982]
112. Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, Smith I, Tothova Z, Wilen C, Orchard R, et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol* 34, 184–191. 10.1038/nbt.3437. [PubMed: 26780180]
113. Sanjana NE, Shalem O, and Zhang F (2014). Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* 11, 783–784. 10.1038/nmeth.3047. [PubMed: 25075903]
114. Piccioni F, Younger ST, and Root DE (2018). Pooled Lentiviral-Delivery Genetic Screens. *Curr. Protoc. Mol. Biol* 121, 32.1.1–32.1.21. 10.1002/cpmb.52.
115. Macosko EZ, Basu A, Regev A, Mccarroll SA, Macosko EZ, Basu A, Satija R, Nemes J, Shekhar K, and Goldman M (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* 161, 1202–1214. 10.1016/j.cell.2015.05.002. [PubMed: 26000488]
116. Saunders A, Huang KW, Vondrak C, Hughes C, Smolyar K, Sen H, Philson AC, Nemes J, Wysoker A, Kashin S, et al. (2021). Ascertaining cells' synaptic connections and RNA expression simultaneously with massively barcoded rabies virus libraries. *bioRxiv*, 2021.09.06.459177. 10.1101/2021.09.06.459177.
117. Krienen FM, Goldman M, Zhang Q, del Rosario CH, Florio R, Machold M, Saunders R, Levandowski A, Zaniewski K, Schuman H, B., et al. (2020). Innovations present in the primate interneuron repertoire. *Nature* 586, 262–269. 10.1038/s41586-020-2781-z. [PubMed: 32999462]
118. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, and Smyth GK (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43, e47–e47. 10.1093/nar/gkv007. [PubMed: 25605792]
119. Chen J, Behnam E, Huang J, Moffatt MF, Schaid DJ, Liang L, and Lin X (2017). Fast and robust adjustment of cell mixtures in epigenome-wide association studies with SmartSVA. *BMC Genomics* 18. 10.1186/s12864-017-3808-1.
120. Wu D, and Smyth GK (2012). Camera: a competitive gene set test accounting for inter-gene correlation. *Nucleic Acids Res.* 40, e133–e133. 10.1093/nar/gks461. [PubMed: 22638577]
121. Velasco S, Kedaigle AJ, Simmons SK, Nash A, Rocha M, Quadrato G, Paulsen B, Nguyen L, Adiconis X, Regev A, et al. (2019). Individual brain organoids reproducibly form cell diversity of the human cerebral cortex. *Nature* 570, 523–527. 10.1038/s41586-019-1289-x. [PubMed: 31168097]

122. Stegle O, Parts L, Piipari M, Winn J, and Durbin R (2012). Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat. Protoc* 7, 500–507. 10.1038/nprot.2011.457. [PubMed: 22343431]
123. Davis JR, Fresard L, Knowles DA, Pala M, Bustamante CD, Battle A, and Montgomery SB (2016). An Efficient Multiple-Testing Adjustment for eQTL Studies that Accounts for Linkage Disequilibrium between Variants. *Am. J. Hum. Genet* 98, 216–224. 10.1016/j.ajhg.2015.11.021. [PubMed: 26749306]
124. Oscanoa J, Sivapalan L, Gadaleta E, Dayem Ullah AZ, Lemoine NR, and Chelala C (2020). SNPnexus: a web server for functional annotation of human genome sequence variation (2020 update). *Nucleic Acids Res.* 48, W185–W192. 10.1093/nar/gkaa420. [PubMed: 32496546]
125. Shatokhina N, Grasby KL, Jahanshad N, Stein JL, Medland SE, and Thompson PM (2021). ENIGMA-Vis: A web portal to browse, navigate & visualize brain genome-wide association studies (GWAS). *Biol. Psychiatry* 89, S136. 10.1016/j.biopsych.2021.02.350.
126. Boughton AP, Welch RP, Flickinger M, VandeHaar P, Taliun D, Abecasis GR, and Boehnke M (2021). LocusZoom.js: Interactive and embeddable visualization of genetic association study results. *Bioinformatics.* 10.1093/bioinformatics/btab186.
127. Hart T, and Moffat J (2016). BAGEL: A computational framework for identifying essential genes from pooled library screens. *BMC Bioinformatics* 17, 1–7. 10.1186/s12859-016-1015-8. [PubMed: 26817711]
128. Love MI, Huber W, and Anders S (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. 10.1186/s13059-014-0550-8. [PubMed: 25516281]
129. Winter J, Schwering M, Pelz O, Rauscher B, Zhan T, Heigwer F, and Boutros M (2017). CRISPRAnalyzeR: Interactive analysis, annotation and documentation of pooled CRISPR screens. *bioRxiv.*



**Figure 1: scRNA-seq characterization of human iPSC village.**

**A**, Schematic of cell village workflow. **B**, Composition of hiPSC village by donor sex and reprogrammed tissue source. **C-E**, Factors influencing variation in gene expression. tSNE projection of scRNA-seq data color-coded by (C) donor and (D) cell cycle stage as inferred from RNA expression profiles. (E) Expression of pluripotency markers *OCT4* and *NANOG*. **F**, Cell groups represent proliferative stem cells, differentiated cells, and nuclei. **G-H**, Minimal effect of cell source (fibroblast vs. PBMC) is seen in (G) tSNE projection of scRNA-seq data and (H) volcano plot of DEG analysis grouped by donor cell source. **I-L**,

Effect of donor sex on RNA expression. tSNE projections of scRNA-seq data from (I) all genes and (J) autosomal genes only. (K) Numbers of DEGs in pairwise comparisons of all iPSC donors, grouped by donor sex and cell source. Three different fold-change thresholds denoted by green (1.2-fold), blue (1.5-fold), and purple (2.0-fold) dotted lines. (L) Volcano plot of DEG analysis grouped by donor sex.

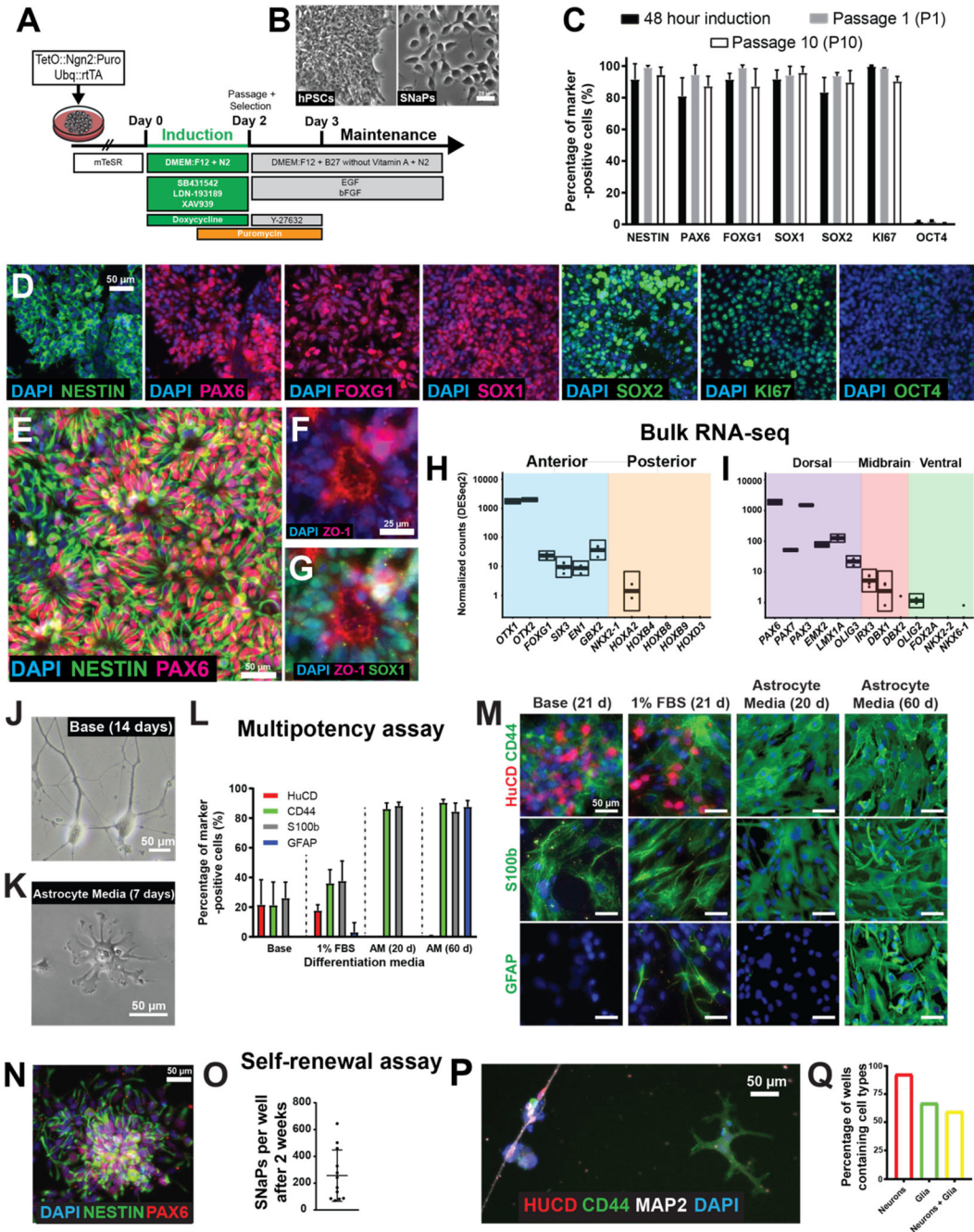
Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript





**Figure 2: Rapid induction of stem cell-derived human NPCs.**

**A**, SNaP induction protocol. **B**, Bright field images of SW7388-1 iPSCs (left) and SNaPs at 48 hours post-induction (right). **C**, Quantification of protein marker expression from immunocytochemistry at 48 hours post-induction, after first passage, and after passage 10. **D**, Immunostaining of forebrain NPC protein markers at 48 hours post-induction. **E-G**, SNaPs self-organize into (E) rosette-like structures 2 days after first passage. Magnified images of a (F) ZO-1<sup>+</sup> and (G) SOX1<sup>+</sup> rosette structure. **H-I**, Bulk RNA-seq of H9 SNaPs. Normalized DESeq2 counts of (H) anterior/posterior genes and (I) dorsal/ventral genes. **J-K**, SNaPs differentiate into (J) neural stem cells (Base, 14 days) and (K) astrocytes (Astrocyte Media, 7 days). **L**, Multipotency assay. **M**, SNaPs differentiate into astrocytes (Astrocyte Media, 20 d and 60 d). **N**, SNaPs self-organize into rosettes. **O**, Self-renewal assay. **P**, SNaPs self-renewal. **Q**, Self-renewal assay.



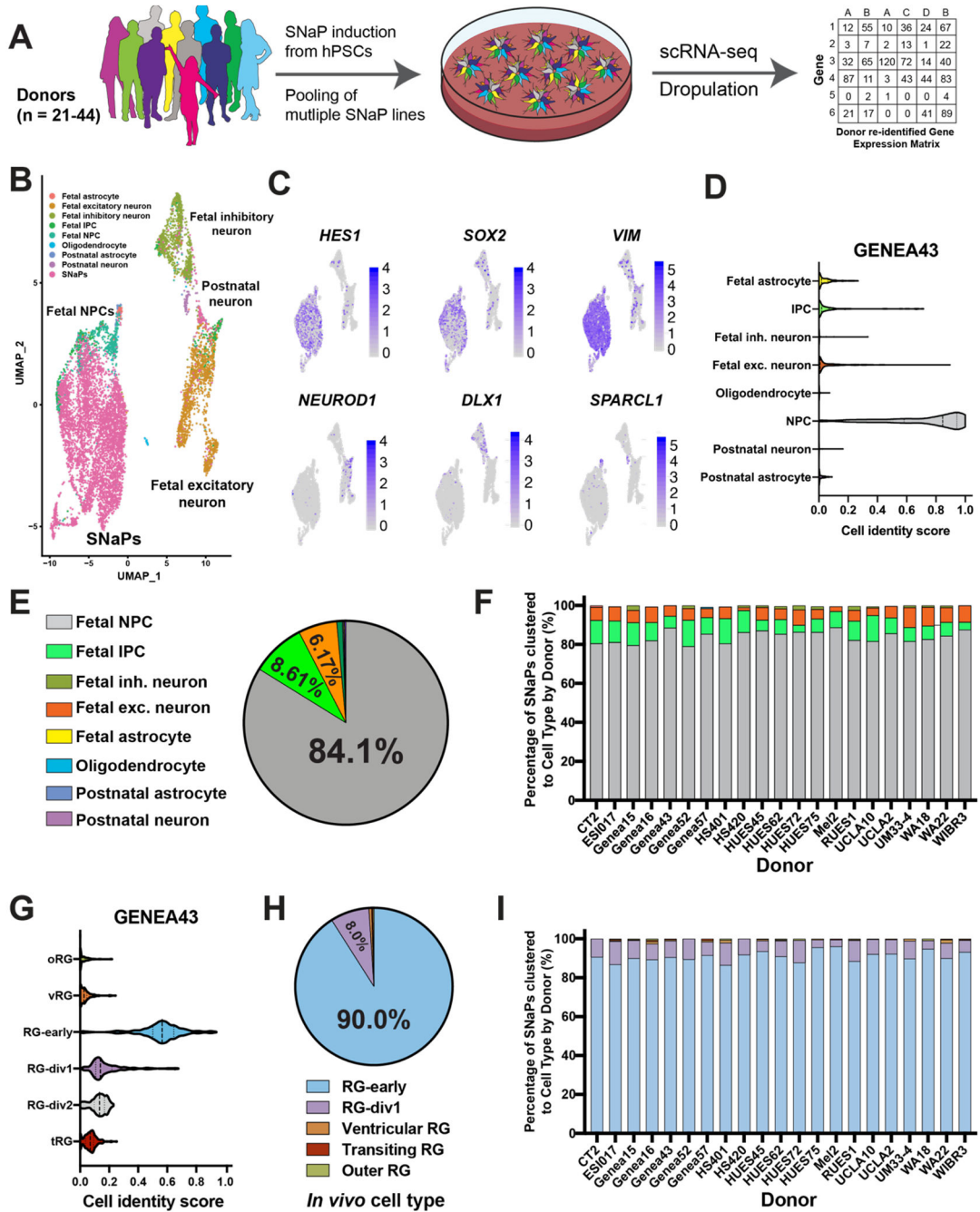
**M**, SNaP multipotency assays. Bright field images of (J) SNaP-derived post-mitotic neurons after 2 weeks in base media and (K) glial cells after 7 days in Astrocyte Media (AM). **L**, Quantification of Panel M. **M**, Immunostaining of neuronal (HuC/D) and glial (CD44, S100b, and GFAP) protein markers in base media or 1% FBS media after 2-3 weeks in culture or after 20-60 days in AM. **N-Q**, SNaPs can self-renew. (N) Representative image of a cluster of PAX6<sup>+</sup>/NESTIN<sup>+</sup> cells 14 days after plating of a single SNaP. (O) Quantification of N. (P) Representative image of well containing both SNaP-derived neurons and glia at 14 days post-differentiation. (Q) Percentage of wells that contain neurons, glia, or both. Data presented as mean  $\pm$  SD.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Figure 3: SNaPs resemble *in vivo* fetal dorsal telencephalic NPCs.**

**A**, Cell village workflow. Multiple donor lines were induced to SNaPs individually before pooling. Donor re-identified gene expression matrix was used for cell type comparisons. **B**, UMAP cluster plot of Village-21 SNaPs and reference fetal and adult brain cells. **C**, NPC marker expression limited to SNaP/Fetal NPC cluster. Excitatory neuron (*NEUROD1*), inhibitory neuron (*DLX1*), and astrocyte (*SPARCL1*) markers are enriched in non-SNaP/Fetal NPC clusters. **D**, Representative data (GENEA43 line) showing high fetal NPC cell identity scores for SNaPs. **E-F**, Quantification of computed cell type classification. (E)

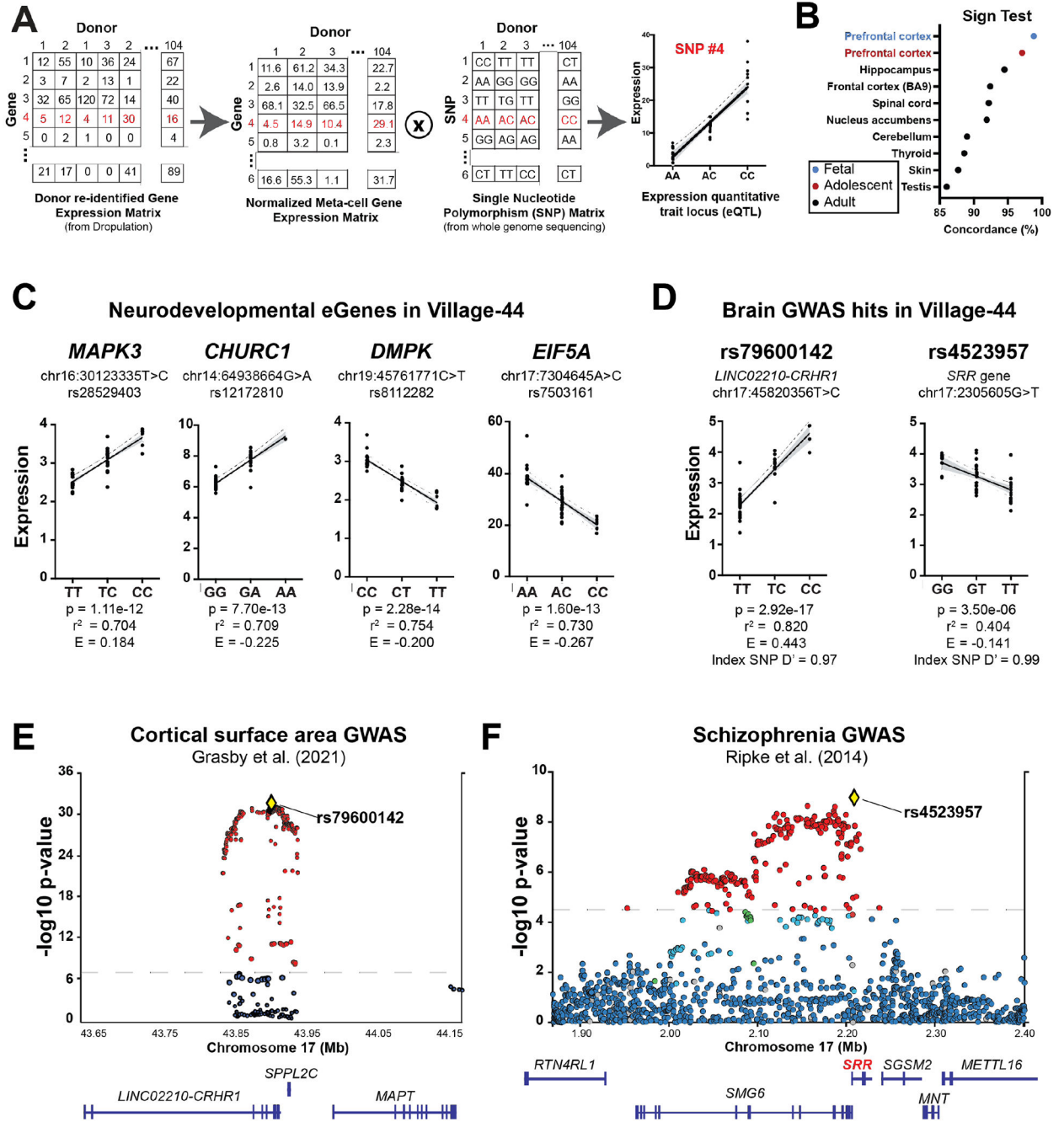
Seurat 3.0 computed cell type classification for all SNaPs in Village-21 and (F) on a per donor basis. **G-I**, Comparison to *in vivo* fetal NPC cell types. (G) Representative data (GENEA43) showing high “RG-early” cell identity scores for SNaPs. (H) Seurat 3.0 computed NPC subtype classification for all cells (5,053 total) in Village-21 and (I) on a per donor basis.

Author Manuscript

Author Manuscript

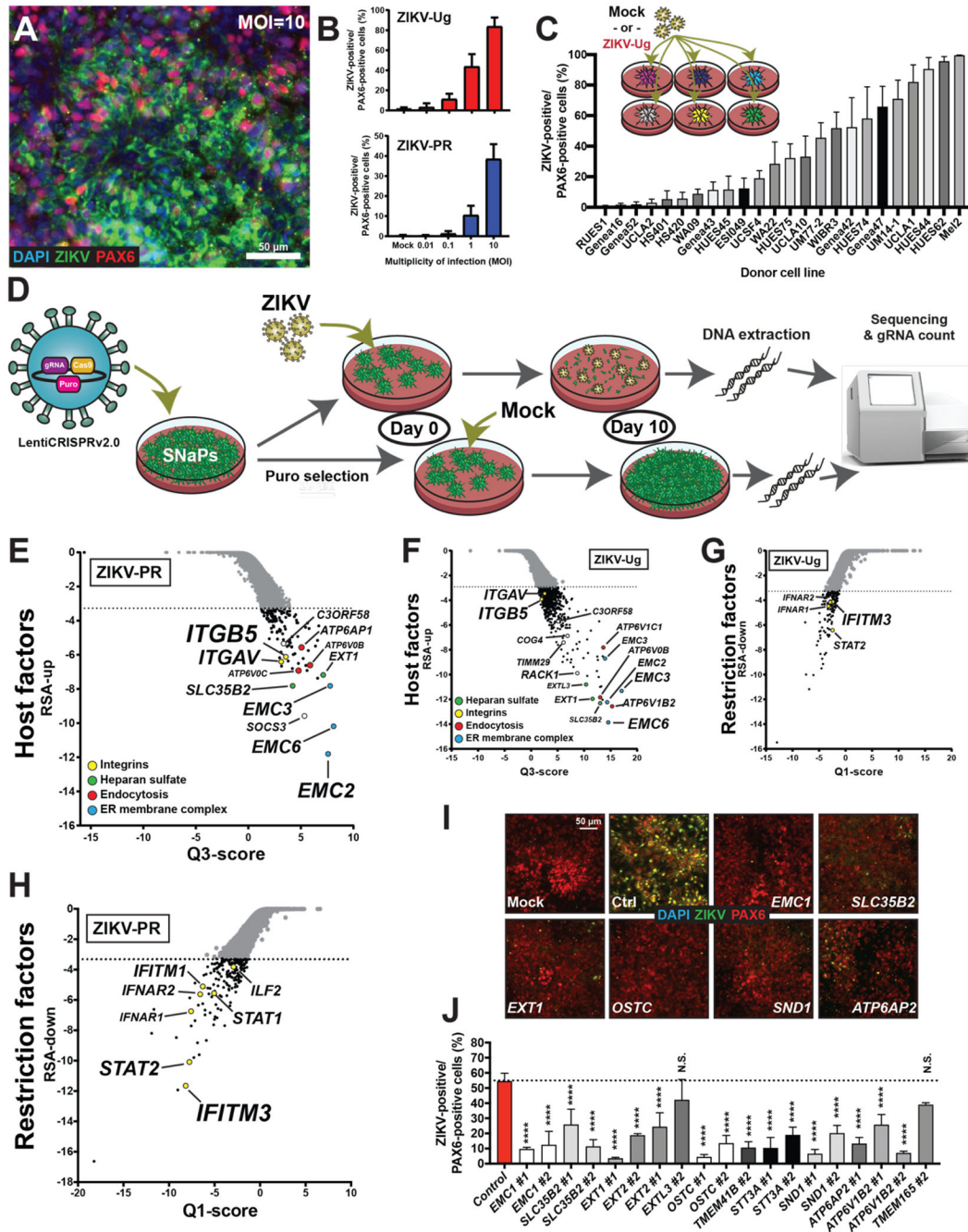
Author Manuscript

Author Manuscript



**Figure 4: eQTL discovery in SNaP villages.**

**A**, eQTL detection workflow. scRNA-seq measurements for individual cells are summed into meta-cells and cross-referenced to SNP genotypes for eQTL analysis. **B**, Village SNaP-44 eQTLs show high concordance by sign test with fetal and adult brain eQTLs. Concordance rates are lower for certain non-brain tissue types. **C**, SNaP village eQTLs detected in neurodevelopmental genes. **D-F**, Overlap with brain GWAS results. (D) Village eQTLs rs79600142 and rs4523957 are *in vivo* GWAS hits for (E) cortical surface area and (F) schizophrenia.



**Figure 5: CRISPR screen identifies potential genetic contributors to ZIKV infectivity variation across donors.**

**A**, Immunostaining of ZIKV 4G2 envelope protein at 54 hpi. **B**, Quantification of SNaP infections with ZIKV-Ug and ZIKV-PR at 54 hpi. **C**, Quantification of arrayed immunocytochemistry-based infectivity assays across donors (ZIKV-Ug, MOI = 1) at 54 hpi. **D**, Design of whole-genome CRISPR-Cas9 screens. **E-F**, RSA plots depicting gene level results of the enriched host factor genes from (E) ZIKV-PR and (F) ZIKV-Ug screens. **G-H**, RSA plots depicting gene level results of the depleted restriction factor genes from

(G) ZIKV-PR and (H) ZIKV-Ug screens. **I-J**, Screen validation. H1-Cas9 SNaPs were transduced with individual gRNAs and exposed to ZIKV-Ug (MOI = 1). (I) Representative images and (J) quantification at 54 hpi. Dashed line denotes adjusted p-value < 0.05 (E-H) or infectivity levels for non-targeting gRNA controls (J). Oneway ANOVA with Dunnett's test for multiple comparisons. Data presented as mean  $\pm$  SD. N.S. = not significant, \*\*\*p<0.0001.

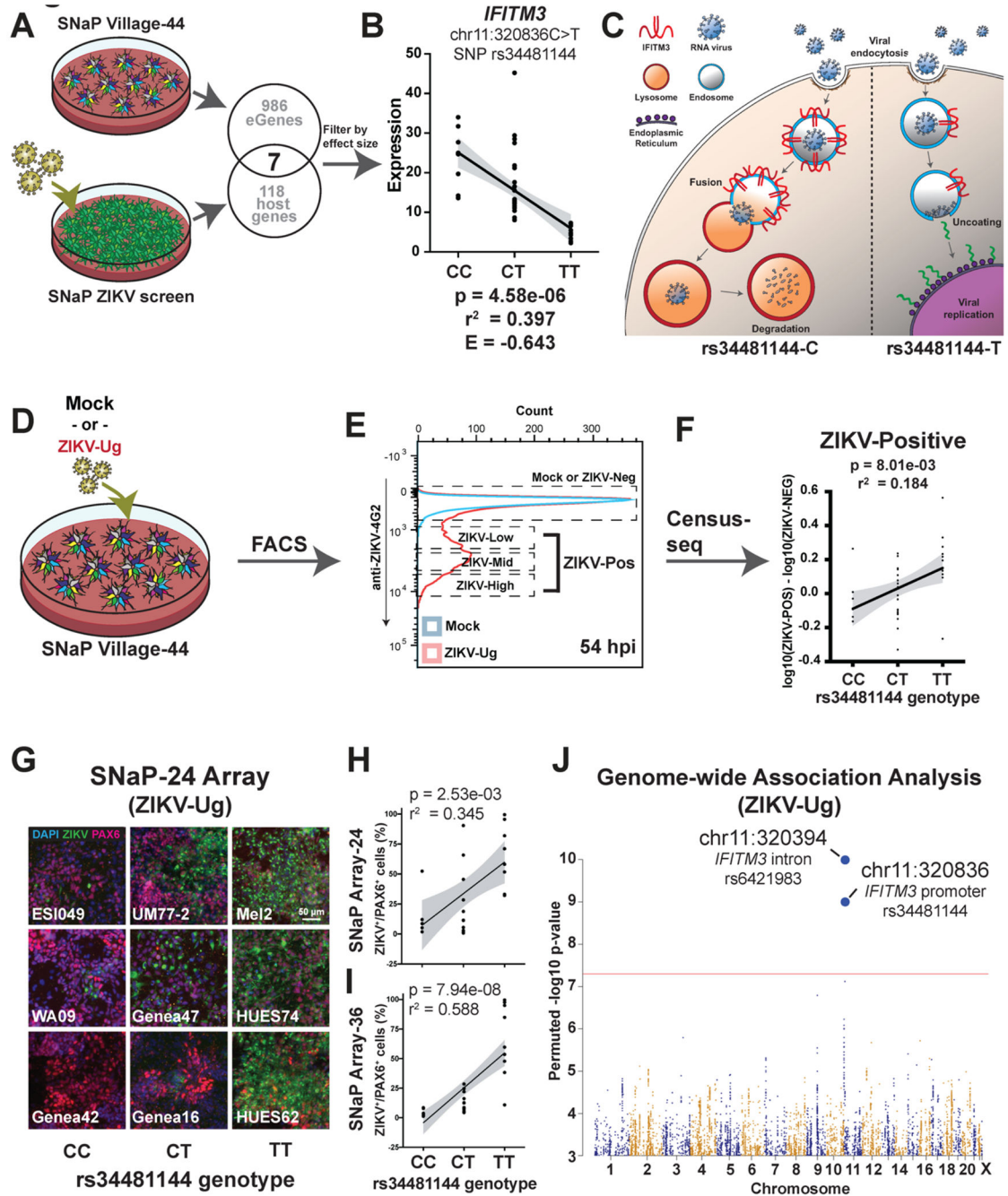
Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript





**Figure 6: SNaP sensitivity to ZIKV associates to a common SNP in *IFITM3*.**

**A**, Workflow combining Village-44 eQTL analysis with CRISPR screen results. **B**, *IFITM3* expression levels in SNaP Village-44 donors harboring reference and alternate alleles at SNP rs34481144. **C**, Schematic model of hypothesis in which decreased expression levels of *IFITM3* result in increased ZIKV infectivity in SNaPs. **D**, SNaP Village-44 was infected with ZIKV-Ug (MOI = 1) or mock media. **E**, At 54 hpi, cells were FACS sorted based on 4G2 signal intensity. **F**, Donor representation in each FACS fraction was measured by Census-seq. Association between genotype for rs34481144 and the distribution of different

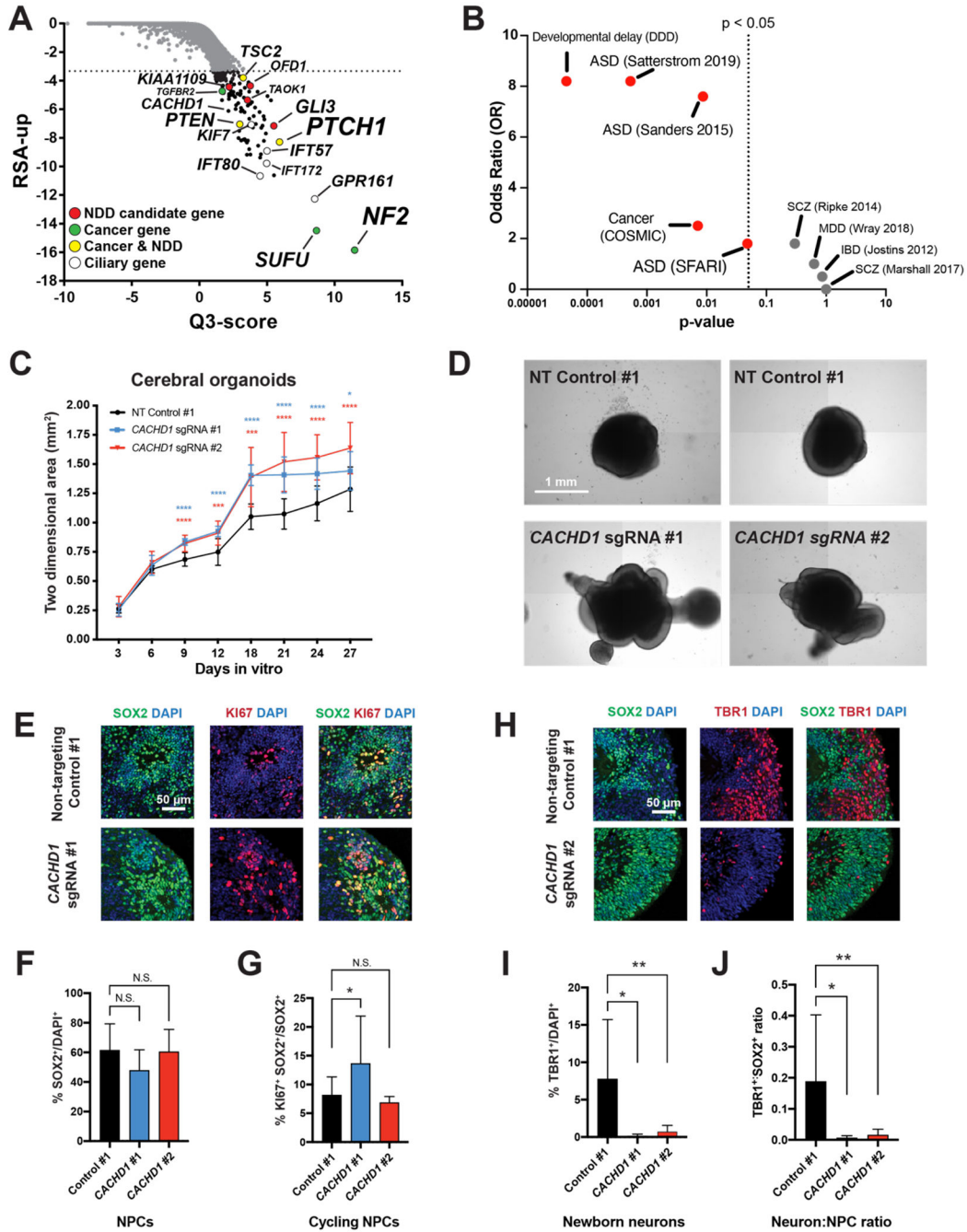
donors' cells between the aggregated ZIKV-positive relative to ZIKV-negative fractions. All values were normalized to mock to control for donor differences in growth rates. **G-I**, Arrayed infectivity assays (ZIKV-Ug, MOI = 1). (G) Representative images from 24 hESC-derived SNaP lines. (H) Quantification of G. (I) Quantification of 36 hiPSC-derived SNaPs. **J**, Genome-wide association analysis of arrayed ZIKV-Ug infectivity (n = 36 donors). Red line denotes genome-wide significance. Linear regression line in black with shaded error in gray (B, F, H-I).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Figure 7: CACHD1 regulates organoid neurogenesis.**

**A**, RSA analysis of genome-wide CRISPR fitness screen. Dashed line denotes adjusted p-value < 0.05. **B**, Summary of disease gene enrichment analysis. Red dots denote gene lists with significant overlap with SNP proliferation hits. **C-D**, Increased size of *CACHD1*-depleted cerebral organoids. (C) Quantification of 2D size over 28 days. (D) Representative brightfield images. **E-J**, Organoid immunohistochemistry. (E) Representative confocal images of sectioned controls and *CACHD1*-edited Day 28 cerebral organoids stained with NPC marker SOX2 and proliferative cell marker KI67. Quantification of (F) NPCs and (G)

cycling NPCs. (H) Representative SOX2 and newborn neuron marker TBR1 immunostains. Quantification of (I) TBR1<sup>+</sup> neurons and (J) neuron-to-NPC ratio. Fisher's exact test (B), repeated measures two-way ANOVA with Dunnett's tests for multiple comparisons (C), and one-way ANOVA with Dunnett's tests for multiple comparisons (F-G, I-J) were used for statistical analysis. Data presented as mean  $\pm$  SD. \*p<0.05, \*\*p<0.01, \*\*\* p < 0.001, \*\*\*\*p < 0.0001, N.S. = not significant.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript