Natural variation in gene expression and Zika virus susceptibility revealed by villages of neural progenitor cells

Michael F. Wells^{1,2*}, James Nemesh^{1,3*}, Sulagna Ghosh^{1,3*}, Jana M. Mitchell^{1,2*}, Curtis J. Mello^{1,3}, Daniel Meyer^{1,3}, Kavya Raghunathan^{1,2}, Matthew Tegtmeier¹, Derek Hawes¹, Anna Neumann¹, Kathleen A. Worringer⁴, Joseph J. Raymond⁴, Sravya Kommineni⁴, Karrie Chan⁴, Daniel Ho⁴, Brant K. Peterson⁴, Federica Piccioni⁵, Ralda Nehme¹, Kevin Eggen^{1,2,6}, Steven A. McCarroll^{1,3,6}

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

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

³ Department of Genetics, Harvard Medical School, Boston, MA 02115, USA

⁴ Department of Neuroscience, Novartis Institute for BioMedical Research, Cambridge, MA 02139, USA

⁵ Genetic Perturbation Platform, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

⁶ Lead contacts

*These authors contributed equally

SUMMARY

Variation in the human genome contributes to abundant diversity in human traits and vulnerabilities, but the underlying molecular and cellular mechanisms are not yet known, and will need scalable approaches to accelerate their recognition. Here, we advanced and applied an experimental platform that analyzes genetic, molecular, and phenotypic heterogeneity across cells from very many human donors cultured in a single, shared *in vitro* environment, with algorithms (Dropulation and Census-seq) for assigning phenotypes to individual donors. We used natural genetic variation and synthetic (CRISPR-Cas9) genetic perturbations to analyze the vulnerability of neural progenitor cells to infection with Zika virus. These analyses identified a common variant in the antiviral *IFITM3* gene that regulated *IFITM3* expression and explained most inter-individual variation in NPCs' susceptibility to Zika virus infectivity. These and other approaches could provide scalable ways to recognize the impact of genes and genetic variation on cellular phenotypes.

HIGHLIGHTS

- Measuring cellular phenotypes in iPSCs and hPSC-derived NPCs from many donors
- Effects of donor sex, cell source, genetic and other variables on hPSC RNA expression
- Natural genetic variation and synthetic perturbation screens both identify *IFITM3* in NPC susceptibility to Zika virus
- A common genetic variant in *IFITM3* explains most inter-individual variation in NPC susceptibility to Zika virus

Format: Article

Figures/Supplemental figures/Supplemental data files: 6/6/4

Words/Characters of main text (not including references): 7,238/48,922

INTRODUCTION

Humans and indeed all natural populations harbor immense diversity in biological traits, affecting almost all organ systems and physiological processes. Reservoirs of natural variation allow populations to adapt to existential crises and other strong selective pressures. Studies of quantitative traits and disease risk in populations – whether of humans, animals, or plants – reveal a large role for standing variation in shaping phenotypic diversity and enabling adaptation. And yet, the mechanisms by which natural genetic variation acts through molecular, cellular and developmental biology to shape trait variation are largely unknown.

Genetic variation in developmental processes has an important but still poorly understood role in shaping phenotypes throughout life. The spatial and cell-proliferative events that establish the cellular compositions of tissues – the cell types, their relative abundances, and their spatial arrangements – almost certainly affect those organs' physiological properties years later. Consistent with this idea, many genetic discoveries in autism and schizophrenia point to genes that are most highly expressed during fetal brain development (Griesi-Oliveira et al., 2021; Mao et al., 2009; Stein et al., 2014). However, further understanding of the genetic and cell-biological mechanisms of such effects has been limited by the lack of experimental systems and clinical specimens relevant to the transient developmental window in which neural progenitor cells (NPCs) shape the cellular makeup of the brain; NPCs are no longer present when individuals' phenotypes are ascertained or when tissue is sampled for analysis.

One mechanism by which genetic variants can lead to phenotypic diversity is through the regulation of gene expression, which has unknown – but almost certainly important – effects on signaling pathways, cell migration, and cell-cell interactions. Abundant human genetic variation has been ascertained and catalogued by the International HapMap Project (Belmont et al., 2005) and 1000 Genomes Project (Auton et al., 2015); collaborations such as the Genotype-Tissue Expression (GTEx; Aguet et al., 2017) project have identified thousands of expression QTLs (eQTLs) – associations of single nucleotide polymorphisms (SNPs) to RNA expression of nearby genes – in every adult organ analyzed. The relationships of these and other haplotypes to gene expression during development is largely unknown.

Relationships between common genetic variants and gene expression can also be revealed using *in vitro* stem cell-derived models, which have become increasingly adopted with advancements in differentiation protocols and the demonstrated similarities of these cells to corresponding cell types in tissues. Fast, reliable induction techniques for generating NPCs, neurons, glia, and oligodendrocytes from human embryonic stem cells (hESCs) and induced pluripotent stem cells (iPSCs) now enable investigations into gene expression variation. However, maintaining cells from many human donors in separate culture environments results in technical variation that can obscure biologically-relevant effects, and requires substantial resources and effort.

Though thousands of eQTLs have been found in every tissue, in iPSCs, and in iPSC-derived cells of various types, we know little today about the ways in which these effects percolate through cells to influence their functional phenotypes. Which of these genetic variants actually change a cell's phenotype in a meaningful way? Biological pathways are built to be robust to many kinds of perturbation and may contain the effects of many genetic variants, even those that regulate a

gene's expression. The precise expression dosage of most genes is of unknown biological significance. Thus, it is essential to understand the relationships among genetic variants, gene expression, and the physiological phenotypes of cells that shape disease processes.

Here we describe the coordinated application of three approaches for dissecting these aspects of a cell's life and their relationships to one another. One is genetic multiplexing, in which tens of thousands of cells from scores of donors are analyzed simultaneously by single-cell RNA-seq, revealing effects of genetic variants, cell-donor and cell-source properties on RNA expression at population scale. Another is "Census-seq", a rapid, inexpensive method for relating cellular phenotypes to natural genetic variation by sequencing the genomic DNA from such "cell villages" (Mitchell et al., 2020). A third is functional CRISPR-Cas9 screens, which explore thousands of artificial genetic perturbations simultaneously (Joung et al., 2017; Ophir et al., 2014).

By applying these experimental platforms in a coordinated way, we find high levels of inter-individual variation in susceptibility of NPCs to the neurotropic Zika virus (ZIKV), then uncover a common, large-effect SNP in the antiviral *IFITM3* gene that explains more than 50% of this physiological variation and appears to act by regulating expression levels of *IFITM3* in NPCs. Analyses of the high-volume data from these same experiments helped us understand genetic, biological and technical sources of variation on gene expression in human pluripotent stem cells (hPSCs) and NPCs. The population-scale nature of these experiments made it possible to describe and quantify the effects of donor sex, cell source, and cell reprogramming methods. We also identify hundreds of NPC eQTLs, some of which affect expression levels of critical neurodevelopmental genes.

We hope that these results enable better understanding of sources of variation in hPSC-derived systems while suggesting possibilities for integrated experimental approaches that use natural and synthetic genetic perturbations to understand disease-relevant functional variants and their effects on specific cell types.

RESULTS

Analysis of cellular variation in "Dropulations"

To understand how natural genetic variation and donor-specific biological effects shape cellular phenotypes, we sought to eliminate other sources of variation by culturing cells from very many donors in a common environment and analyzing them all together. Transcribed SNPs contain abundant information that can be used to identify the donor of a biological sample using the alleles that are present in RNA transcripts. The combination of hundreds of transcribed SNPs can act as a cell-intrinsic barcode that uniquely identifies the donor of an individual cell (Kang et al., 2018). We further developed genetic multiplexing analysis: (i) to address challenges inherent to scRNA-seq experiments, such as ambient RNA; (ii) to utilize unique molecular indicators (rather than reads) as the basis for demultiplexing; and (iii) to allow scalability up to hundreds of potential donors (Methods). We provide the resulting software ("Dropulation", for droplet-based sequencing of populations) in an open-source format (<https://github.com/broadinstitute/Drop-seq>).

We performed a variety of *in silico* and experimental analyses to confirm the accuracy with which cells were correctly assigned to donors by this approach (Methods, Fig. 1a, Supplementary Note, Fig. S1).

The ability to measure mRNA expression in a common culture environment made it possible to quantify effects that have long been of great concern (but uncertain practical impact) in stem cell research, including the effects of donor sex and cell source. To characterize such effects, we established a 104-donor human iPSC “cell village” (Figure 1a) that included iPSCs derived from multiple cell sources (skin and blood cells) and from male and female donors (Figure 1b, Figure S2a). We profiled this cell village by scRNA-seq (10X Genomics platform) five days after pooling the 104 iPSC lines. We analyzed 86,185 cells by scRNA-seq, sampling their RNAs to an average depth of 104,160 UMIs per cell and assigning each cell to its donor-of-origin using Dropulation (Figure S2b-c).

When cultured together, the iPSCs from the 104 donors exhibited highly similar RNA-expression patterns, with cells from the 104 donors distributing similarly across a two-dimensional projection of the single cells’ RNA-expression patterns (Figure 1c). The primary source of variation in the individual cells’ expression profiles appeared to be their progress through the cell cycle (Figure 1d). Heterogeneity in cell state or identity can also impact gene expression (Figure S2d). As expected, the great majority of the hiPSCs expressed pluripotency markers *NANOG* and *OCT4* at high levels (Figure 1e); an exception (cluster 10, 4.3% of all cells) involved cells with higher-than expected expression of neural progenitor markers *RIPPLY3* and *SIX6*, suggesting that they have undergone spontaneous neural differentiation (Figure S2e). Another set of cells (clusters 6 and 7, 3.1% of all cells) exhibited high expression of *UTF1* and *MTG1*, suggesting poor differentiation potential; notably, these cells came from only a handful of cell lines (Figure S2f). A subset of single-cell expression profiles exhibited lower UMI counts and higher percentages of nascent transcripts with intronic reads (Figure S2g-h); we interpret these profiles to have arisen from nuclei (likely generated by unintentional cell lysis during handling) rather than intact cells.

iPSC lines are routinely created from skin (fibroblasts) or from blood (PBMCs); fundamental methodological questions involve the comparability of iPSCs created from different tissue sources. While it is reasonable to hypothesize that tissue source for iPSC reprogramming influences gene expression, our data suggests that such effects were small: Only four RNAs (all noncoding) exhibited genome-wide-significant differences in expression between the 56 skin-derived and 45 PBMC-derived iPSC lines (Figure 1fg; Data File 1). A more-sensitive enrichment analysis of sub-significant differential expression measurements indicated a modest enrichment associated with the somatic cell of origin (Data File 1): genes preferentially expressed in blood-derived iPSCs were nominally enriched in a myeloma related pathway, while genes preferentially expressed in skin-derived iPSCs were nominally enriched in a gene set associated with melanoma relapse (Figure S2i). This finding suggests that there is modest retention of epigenetic memory inherited from the parental cell source of origin in the iPSC cell lines, but that few protein-coding genes are strongly affected by this memory.

Many human phenotypes show sex-biased differences, and RNA-expression levels of many genes differ on average between males and females in various tissues (Oliva et al., 2020). The

Figure 1: scRNA-seq characterization of human iPSC village. **A**, Schematic depicting cell village workflow. Human induced pluripotent stem cell lines from 104 donors were pooled into a single *in vitro* environment prior to scRNA-seq analysis; the donors of individual cells were identified from combinations of many (typically 200+) transcribed common SNPs. **B**, Composition of hiPSC village by donor sex and the tissue source of the cells used for reprogramming. **C-E**, Factors influencing variation in gene expression. (C) tSNE projection of scRNA-seq data color-coded by (C) donor of origin, as inferred from allelic information on scRNA-seq reads; and (D) cell cycle stage, as inferred from RNA expression levels. (E) Expression of pluripotency markers *OCT4* and *NANOG*. **F-G**, Modest effect of cell source (fibroblast vs. PBMC) is seen in (F) tSNE projection of the scRNA-seq data and (G) volcano plot of differential gene expression analysis grouped by donor cell source. **H-L**, Analyses of relationship of donor sex-chromosome status to RNA expression through a series of tSNE projections involving subsets of the scRNA-seq data consisting of (H) all genes, (I) X-chromosome plus autosomal genes, and (J) autosomal genes only. (K) Numbers of differentially expressed genes 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 differential gene expression analysis grouped by donor sex.

extent to which cell-autonomous biology – as opposed to, for example, circulating hormones – contributes to such differences is unknown. The experimental design made it possible to isolate and quantify cell-autonomous effects, since (i) environmental and non-cell-autonomous effects were controlled across cells growing in the same culture, and (ii) very many male and female donors were sampled. The RNA-expression profiles of the individual cells initially appeared to be strongly distinguished by donor sex (Figure 1h), though this difference largely disappeared when we limited analysis to autosomal genes (Figure 1i-j), suggesting that it arose largely from Y-linked genes and genes that escape X chromosome inactivation.

How strong is sex-biased expression in iPSCs relative to the routine effects of inter-individual variation? We compared pairs of donors within and between sex to generate pairwise distributions of differentially expressed genes. We found similar numbers of differentially expressed autosomal genes in same-sex comparisons (XX vs XX; XY vs XY) as we did in across-group (XX vs. XY) comparisons, 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. These results suggest that differences in gene expression generated by donor sex and source cell-type are small compared to the routine effects of inter-individual variation, and that differences in expression of sex-chromosome genes do not lead to broader effects on cells' biology in this context.

To further explore these modest, quantitative effects that become apparent in a population-scale analysis, we computed differential gene expression across all 104 donors using the limma/voom software package (Ritchie et al., 2015). As a point of reference, we performed similar analysis on reprogramming method cell source (Figure 1i). Consistent with earlier observations from tissue-level analysis (Oliva et al., 2020), effects were small (median log2FC upregulated genes = 0.15, downregulated genes = -0.10). To assess potential biological significance of the small transcriptional differences driven by cell source and donor sex, we performed gene set enrichment analysis using CAMERA (Wu and Smyth, 2012) and curated gene sets (C2) from the Molecular Signatures Database (MSigDB; (Subramanian et al., 2005). The most significant gene sets associated with differences across male and female donor lines mapped to pathways related to X chromosome inactivation and imprinting (Figure S2j). This signal disappeared when the analysis was limited to autosomal genes, indicating that differences in gene expression due to donor sex was largely driven by genes on the sex chromosomes (Data File 1).

The small effects of donor sex and cell source, relative to the routine effects of inter-individual variation, suggest that population-scale iPSC experiments can successfully utilize blood- and fibroblast-derived iPSCs together, though will ideally use designs in which these factors do not confound the effects of other variables of interest (such as genotype or case/control status).

Effects of common genetic variation on RNA expression in hiPSCs

To map expression QTLs, we tested for associations between donors' genotypes and their gene expression (summed across their individual iPSCs) using a linear regression model that included independent variables for biological and technical covariates, utilizing the MatrixEQTL analytical pipeline (Shabalin, 2012) (Figure 2a). Some 6,247 genes exhibited significant association to one

Figure 2: Detection of cis-expression QTLs (eQTLs). **A**, eQTL detection workflow. scRNA-seq measurements for individual cells were summed into meta-cells (one per donor), which were then cross-referenced to SNP genotypes for eQTL analysis by MatrixEQTL. Red font denotes example data used for generating the eQTL plot for SNP#4. **B**, The number of eQTL-regulated gene (eGenes) detected by eQTL analysis in the 104-donor hiPSC village is shown in the context of other eQTL studies, including analyses (by GTEx) of 46 different human tissue types (blue circles) and earlier analysis of hPSCs in an arrayed format (orange squares). **C**, Optimization of eQTL search parameters including search windows (genomic distance searched around each gene) and minor allele frequencies (MAF).

or more SNPs (“eGenes”, at a q-value < 0.05) across the 104 hiPSC lines (Figure 2b, Data File 2).

To critically evaluate the utility of the genetically multiplexed format for eQTL discovery, we compared rates of eQTL detection across studies of diverse tissue and cell types, including GTEx and a report discovering eQTLs in human stem cells that were cultured and analyzed in an arrayed format (DeBoever et al., 2017). Across these studies, power to discover eQTLs increased almost linearly in relation to the number of donors sampled (with a slope of 29.2 eQTLs per donor). Analysis of cells in a “cell village” format identified 60 eQTLs per donor (Figure 2b), suggesting that the multiplexed approach helped increase the sensitivity of eQTL detection, likely by controlling for technical and environmental factors that introduce noise into most eQTL discovery efforts.

To assess the relationship of these iPSC eQTLs to eQTLs discovered by large-scale analysis of human tissues, we performed a sign test, asking what fraction of these SNPs exhibited the same direction of association to the same gene-expression phenotypes in other tissues and experiments. We found an 88% concordance (by sign test) between the eQTLs we found, and eQTLs found by the arrayed analysis of many individual iPSC lines. In addition, we observed similar sign test results when comparing to GTEx tissue and cell type data, with the strongest concordance coming from tissues with many mitotic cells (Figure S3a).

Finally, we explored the significance of eQTL search parameters, such as the search window (around each gene) used for cis-eQTL discovery, and the allele frequencies of the SNPs considered; both involve a tradeoff between the increased sensitivity of a broader search, and the reduction in sensitivity when multiple hypothesis testing is fully addressed in such a search (Figure 2c). We hope that this information is useful in guiding other eQTL discovery studies.

Differential susceptibility to Zika virus infection across donor lines

hPSCs and hPSC-derived cell types offer a unique opportunity to analyze genetic influences upon the physiological phenotypes of living cells. They also present an opportunity to learn from both natural genetic variation and synthetic genetic perturbations. We sought to identify and understand genetic effects upon a physiological phenotype, the vulnerability of human NPCs to the neurotropic Zika virus.

Interactions with viruses have significantly altered the human genetic landscape in the form of an arms race between pathogen and host (Enard et al., 2016). Several recent reports have chronicled SNP associations with clinical outcomes of influenza, HIV, and SARS-CoV-2 viral infections, among others (Helminen et al., 1999; McLaren et al., 2015; Zhang et al., 2020, 2013). Many of these SNPs may affect the expression of human genes that play prominent roles in viral entry or replication mechanisms, as well as those that mediate the innate immune response (Kenney et al., 2017), and in doing so influence the probability of productive infection and spread to neighboring host cells. While numerous relationships between specific SNPs and viruses have been identified across a range of cell types, our understanding of the interplay among human genetic variation, viral susceptibility, and the brain is still in its infancy.

The Zika virus is a global health concern, with an outbreak that originated in South America and spread to over 50 countries between 2015 and 2016 (Barbeito-Andrés et al., 2018; Rossi et al., 2018; Saxena et al., 2018). While most healthy adults are asymptomatic after infection, prenatal exposure to this mosquito-borne virus can result in Congenital Zika Syndrome (CZS), which is characterized by severe neurological outcomes including microcephaly and neurodevelopmental delay (Brasil et al., 2016; Cauchemez et al., 2016; Mlakar et al., 2016). Interestingly, only 30% of prenatal infections during this outbreak resulted in CZS, and 95% of these cases were reported in only one country—Brazil (Nielsen-saines et al., 2019; de Oliveira et al., 2017). While many factors have been hypothesized to explain these observations, including maternal nutrition and socioeconomic status (Barbeito-Andrés et al., 2020; Souza et al., 2018), other potential factors – including human genetic diversity – can contribute to differential responses to this viral pathogen (Borda et al., 2021; Gomes et al., 2021; Santos et al., 2019).

ZIKV infection leads to CZS primarily through its preferential targeting of proliferating NPCs and glial cell types over post-mitotic neurons (Retallack et al., 2016; Tang et al., 2016). We therefore used the Stem cell-derived NGN2-accelerated Progenitor (SNaP) *in vitro* model of human dorsal telencephalic neural progenitors, which we recently showed are capable of supporting the complete ZIKV life cycle (Wells et al., 2018), to investigate ZIKV pathogenesis in a microcephaly-relevant cell type. To quantify inherent differences in ZIKV susceptibility *in vitro*, we infected 24 independently-cultured hESC-derived SNaP lines with ZIKV-Ug (MOI = 1) and measured infectivity levels at 54 hpi (Figure 3a). We observed a surprisingly high degree of variability in viral sensitivity, with mean infectivity rates ranging from 0.7% (RUES1 cell line) to 99.4% (Mel2; Figure 3b).

These data suggest that cell-intrinsic factors make a large contribution to the inter-individual variation in vulnerability to ZIKV. We sought to understand this through both natural genetic variation and synthetic genetic perturbations, which we hypothesized might lead us to the same genes.

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

Viruses commandeer cellular proteins and pathways at each stage of their life cycle in an effort to propagate and spread to additional cells. The exact host biological processes subjected to these manipulations differ based on virus and cell type. Genome-wide screens have identified ZIKV host genetic factors in various cell types (Savidis et al., 2016a; Wang et al., 2020), including hPSC-derived NPCs (Li et al., 2019). We wanted to run similar genome-wide CRISPR-Cas9 positive selection survival screens in SNaPs to determine if the same genes are important for ZIKV infection of our cellular model, as well as to limit the number of genes that could potentially explain the differential ZIKV susceptibility we observed across donor lines.

We transduced SW7388-1 SNaPs with the Brunello human CRISPR knockout pooled library packaged in the all-in-one (Cas9 + gRNA) LentiCRISPRv2.0 vector (Figure 3c). LentiCRISPR-expressing SNaPs underwent 7 days of puromycin selection and expansion prior to infection with ZIKV-Ug (MOI = 1), ZIKV-PR (MOI = 5), or mock-infection media. Ten days later, all samples were harvested and prepared for DNA extraction and next generation sequencing (NGS) before the number of gRNA reads in each sample was counted and used for subsequent comparative

Figure 3: ZIKV infectivity varies across human NPC donor lines. **A**, Schematic depicting arrayed infection of 24 hESC-derived SNaP lines with ZIKV-Ug (MOI = 1). Representative immunostaining image of ZIKV-Ug 4G2 envelope protein at 54 hpi. Scale bar = 25 μ m. **B**, Quantification of arrayed infectivity assays (ZIKV-Ug, MOI = 1) at 54 hpi. **C**, Design of whole-genome CRISPR-Cas9 screens. SW7388-1 hiPSCs were induced to SNaPs before transduction of the Brunello gRNA library via the LentiCRISPRv2.0 system. SNaPs used for the ZIKV survival screen were infected with ZIKV-Ug (MOI = 1), ZIKV-PR (MOI = 5), or mock infection media, and harvested 10 days later for DNA sequencing. ZIKV-infected samples were compared to Day 10 mock controls to identify genes that enhance ZIKV survival. **D-E**, RSA plots depicting gene level results of the enriched host factor genes from (D) ZIKV-PR and (E) ZIKV-Ug screens. Dashed line denotes hit cutoff for genes enriched in Day 10 ZIKV samples relative to mock controls (adjusted p-value < 0.05). **F-G**, RSA plots depicting gene level results of the depleted restriction factor genes from (F) ZIKV-PR and (G) ZIKV-Ug screens. Dashed line denotes hit cutoff for genes depleted in Day 10 ZIKV samples relative to mock controls (adjusted p-value < 0.05). **H-I**, Confirmation of primary screen hits. H1-Cas9 SNaPs were transduced with individual gRNAs, expanded, and exposed to ZIKV-Ug (MOI = 1). At 54hpi, infectivity was measured relative to mock controls. (H) Representative images and (I) quantification. Dashed line denotes infectivity levels for non-targeting gRNA controls. One-way ANOVA with Dunnett's test for multiple comparisons. Data presented as mean $\pm$ SD. N.S. = not significant, ****p<0.0001.

analysis. Our results showed high gRNA coverage (<0.15% missing gRNAs across replicates), a high correlation among replicates (Pearson r values >0.92), and the expected limited effects of NT gRNA controls (Figure S4a-b; Data File 3).

The enrichment of sequencing reads for a gRNA targeting a particular gene in the ZIKV-infected samples relative to mock controls at Day 10 would suggest that disruption of said gene confers protection from ZIKV-mediated cell death. Using the redundant siRNA activity (RSA) statistical method (König et al., 2007), we detected 102 and 765 significantly-enriched candidate host factor genes in the ZIKV-PR and ZIKV-Ug samples, respectively, relative to mock controls (adjusted p -value < 0.05; Figure 3d-e). Despite the disparity in the numbers of significantly enriched genes, which is likely due to the higher noise associated with the reduced levels of cell death in the ZIKV-PR screen (ZIKV-Ug: 3.83% cell viability, ZIKV-PR: 26.0% cell viability; Data not shown), the two screens otherwise showed strong concordance: 75% (77/102) of the ZIKV-PR hits were also genome-wide hits in the ZIKV-Ug screen (Data File 3), indicating that the mechanisms that mediate ZIKV-induced NPC death are grossly similar between the two strains.

As expected, the strongest-effect genes and pathways from the survival screens corresponded to genes involved in known flaviviral infection and replication mechanisms, including heparan sulfate biosynthesis (*EXT1*, *SLC35B2*, *B4GALT7*, *EXT2*), endosomal acidification (*ATP6AP2*, *ATPV1B2*, *ATP6V0B*), the oligosaccharyltransferase (*OSTC* and *STT3A*) and endoplasmic reticulum membrane complexes (*EMC1-7*), and the oligomeric Golgi complex (*COG3*, *COG4*, *COG8*, *COG1*). Many of these genes were detected in the whole-genome experiments performed using conventional hPSC-derived NPCs (Li et al., 2019). Our screen identified additional genes not previously attributed to ZIKV pathogenesis, including the viral translation regulator *RACK1* (Majzoub et al., 2014) and the mitochondrial translocase *TIMM29*, thus highlighting inhibitors of these proteins as potential therapeutics. Previously nominated ZIKV entry factors *AXL*, *TYRO3*, *MERTK*, *CD209*, *HAVCR1*, and *TIMD4* were not significantly enriched in our primary screen, consistent with our previous findings (Wells et al., 2016). Our screen did however highlight the *ITGAV* and *ITGB5* genes, which together encode the transmembrane $\alpha V\beta 5$ protein and have previously been shown to serve as an entry factor for other viruses (Ballana et al., 2011).

We also detected 195 and 147 genes that were significantly depleted in the ZIKV-PR and ZIKV-Ug screens, respectively, suggesting that their ablation rendered cells more sensitive to ZIKV-mediated cell death (adjusted p -value < 0.05; Figure 3f-g). The 48 depleted genes that were detected in both screens included the Type I interferon-responsive genes *IFNAR1-2*, *STAT2*, and *IFITM3*.

The results of the primary survival screen were confirmed through CRISPR- and antibody-mediated secondary assays. We first generated a H1 hESC line that constitutively expresses Cas9 (clone H1-36-23) and validated its genome-editing properties (Figure S4c-f). H1-Cas9 hESCs were then induced to SNaPs and transduced with lenti-gRNAs in an arrayed format for infectivity assays at 54 hpi and cell viability assays at 120 hpi using ZIKV-Ug (MOI = 1). We targeted a subset of genes that were significantly enriched in both SNaP survival screens and represented the spectrum of protein complexes and biological processes identified by these primary screens. As predicted, administering guides that targeted the EMC genes, *SLC35B2*,

OST complex components, and vATPase subunits, but not TAM receptors, resulted in significantly improved cell viability and reduced infectivity compared to NT gRNA controls (Figure 3h-i, S4g-i). Interestingly, disruption of the transmembrane-spanning integrin subunit ITGB5 improved cell viability and dramatically reduced levels of infection (Figure S4j). Similar protection was observed when SNaPs were treated with anti-integrin $\alpha V\beta 5$ antibody prior to ZIKV-Ug exposure (Figure S4k-l) in agreement with two recent reports that also nominated integrin $\alpha V\beta 5$ as a ZIKV entry factor in neural cells (Wang et al., 2020; Zhu et al., 2020).

These survival screens implicated the host genes appropriated by ZIKV along every step of its lifecycle, as well as the antiviral effectors utilized by NPCs to fight infection. The outcomes of these screens are largely consistent with previous findings in hPSC-derived NPCs generated using standard induction protocols (Li et al., 2019), providing additional confirmation that SNaPs behave like NPCs. Along with the discoveries from recent large-scale drug screens in human NPCs (Xu et al., 2016; Zhou et al., 2017), our results should be taken into consideration when developing novel clinical therapies against CZS, for which there are currently no approved treatments. Perhaps most importantly, these data nominate a small number of genes as potential mediators of the differential viral susceptibility we observed across SNaP lines.

Identification of expression QTLs in ZIKV host factors

Having identified (in the CRISPR screens) genes that mediate ZIKV infectivity and death of SNaPs, we sought to identify natural genetic variants that affect the expression of these genes in SNaPs. We pooled 44 hESC-derived SNaP lines into a village (Village-44; Figure 4a) and three days later processed these cells for scRNA-seq and donor re-identification. As expected, cell cycle stage and differentiation status contributed to variation in gene expression (Figure 4b-c). We used scRNA-seq data to assess the quality and identity of the cells in this village, finding that SNaPs predominantly clustered with fetal NPCs when compared to an integrated reference dataset composed of human fetal and adult brain cells (Darmanis et al., 2015; Nowakowski et al., 2017) (Figure S5a). SNaPs and fetal NPCs expressed progenitor markers *HES1*, *SOX2*, and *VIM*, while displaying little to no expression of markers of differentiated neurons or astrocytes (Figure S5b). Furthermore, using the cell type classification pipeline in Seurat 3.0 (Stuart et al., 2019) to measure SNaP similarities to *in vivo* cells, we found that 83.1% of the SNaPs were determined to most closely resemble fetal NPCs, suggesting that the SNaPs have an expression profiles similar to those of their *in vivo* counterparts. In addition to confirming the relevance of this cell type to its *in vivo* counterpart, analysis confirmed the consistent production of high-quality SNaPs across all donors (Figure S5c-d).

We identified 961 genes whose expression levels were associated with the genotypes of nearby SNPs ("eGenes", Data File 4). Many of these genes have known roles in fetal brain development. They included genes associated with neurodevelopmental disorders, including *MAPK3*, *PEX6*, and *DMPK* as well as the fetal brain transcriptional activator *CHURC1* and the Williams Syndrome genes *WBSCR16* and *WBSCR27* (Figure S5e, Data File 4). They also included many genes associated with autism spectrum disorder, epilepsy, and microcephaly such as *DPP10*, *HNRNPU*, and *TSEN2*. This analysis provides an inventory of genes regulated by common genetic variation

Figure 4: Differential expression of ZIKV host factors in SNaP villages. **A**, Schematic describing construction of 44-donor SNaP village. **B-C**, scRNA-seq of Village-44. (B) Though largely homogenous, SNaPs aggregate into five tSNE clusters driven primarily by cell cycle and differentiation status. (C) Expression of marker genes for NPCs (*NESTIN*, *SOX2*), pluripotent stem cells (*OCT4*), cycling progenitors (*TOP2A*), excitatory neurons (*STMN2*) and presumptive neural crest (*LUM*). **D**, Workflow combining Village-44 eQTL analysis with ZIKV survival screen results. 961 eGenes were filtered down to 8 potentially ZIKV-related genes with *IFITM3* having the largest effect size. **E**, *IFITM3* expression levels for SNaP lines in Village-44 harboring reference and alternate alleles at locus chr11:320836 (SNP rs34481144). Linear regression line in black with shaded error in gray.

in human NPCs, which could be useful for future investigations of variant effects on a range of neurodevelopmental processes and diseases.

We had hypothesized that screens of the human genome using synthetic (CRISPR) and natural (genetic) perturbations might converge. We cross-referenced the eGenes from our eQTL analysis to a list of 125 genes that were significantly enriched or depleted in both ZIKV survival screens. We found an overlap of 8 genes (Figure 4d) including the replication factors *EMC3* and *COG4*, and the antiviral genes *IFNAR1* and *IFITM3*. Of the 8 overlapping genes, SNPs in the *IFITM3* gene had the strongest effect size (i.e. the slope of the regression line normalized by mean expression of gene; $E = -0.710$ for rs6421983; Figure S5f). *IFITM3* is believed to protect cells against RNA viruses by identifying and shuttling viral-containing endosomes to the lysosome for degradation (Spence et al., 2019). Previous work in immortalized cancer cell lines has shown that *IFITM3* knockdown enhances ZIKV infection rates, while overexpression virtually eliminates viral replication (Brass et al., 2009; Savidis et al., 2016b). The influence of natural variation in *IFITM3* levels on flavivirus infectivity has yet to be explored.

We first aimed to pinpoint which *IFITM3* SNP was most likely responsible for the varied expression levels of this gene across cell lines. Pairwise linkage disequilibrium (LD) analysis of all statistically-significant *IFITM3* SNPs revealed four SNPs that were in high LD with the index SNP and thus associated to ZIKV infectivity with similarly large effect size ($|E| > 0.6$) (Figure S5g; Data File 4). Of these SNPs, rs34481144 (the only SNP located in a non-intronic region) is positioned in the 5'-UTR promoter segment of *IFITM3* and has previously been shown to influence expression of this gene (Allen et al., 2017), suggesting that it is the specific variant responsible for the effect we see. SNaPs from donors homozygous for the reference allele (C) of rs34481144 expressed *IFITM3* at levels 4.82-fold higher than donors homozygous for the alternate allele (T) ($p = 4.28 \times 10^{-6}$, $r^2 = 0.399$; Figure 4e). We hypothesized that by enhancing expression of this antiviral gene the reference allele of rs34481144 confers protection from ZIKV infection, relative to the alternate allele.

rs34481144 is a functional QTL influencing NPC sensitivity to ZIKV infection

To formally test the hypothesis that the reference allele protects SNaPs from ZIKV infection, we exposed SNaP Village-44 to ZIKV-Ug (MOI = 1) or mock media (Figure 5a). At 54 hpi, we FACS-sorted the cells into four fractions based on ZIKV envelope protein 4G2 antibody-stained GFP signal intensity (ZIKV-Negative, -Low, -Mid, -High) before harvesting pellets for DNA sequencing. We then analyzed each cell fraction using Census-Seq (Mitchell et al., 2020) to estimate each donor's cellular contribution to the different fractions. Donors with the rs34481144^{TT} genotype were greatly over-represented relative to rs34481144^{CC} donors in the ZIKV-positive populations relative to the ZIKV-negative pool ($r^2 = 0.225$, $p = 3.04 \times 10^{-3}$; Figure 5b-e). No such relationship was observed with other significant Village-44 eQTLs located in antiviral/host factor genes (Figure S6a). These data support the hypothesis that the rs34481144-T allele renders SNaPs more vulnerable to ZIKV infection compared to cells harboring the rs34481144-C allele (Figure 5f).

We next asked whether the apparently strong effect of the rs34481144 genotype on ZIKV susceptibility is cell-intrinsic or arises in an unexpected way from the cell-village experimental design. To do so, we analyzed data from SNaP infectivity assays of independently-cultured

Figure 5: SNaP sensitivity to ZIKV associates to a common SNP in *IFITM3*. **A**, Experimental workflow. SNaP from 44 donors were cultured together, then infected with ZIKV-Ug (MOI = 1) or mock media. At 54 hpi, cells were FACS sorted into 4 bins based on 4G2 signal intensity. Representations of the individual donors in each of the FACS-sorted cell fractions was measured by Census-seq analysis, which infers the quantitative representation of each donor in a mixture from the quantitative representations of hundreds of thousands of common SNP alleles in that mixture. **B-E**, Association between genotype for rs34481144 and the distribution of different donors' SNaPs between the ZIKV-positive fractions (ZIKV-Low, -Mid, -High, as well as an aggregated ZIKV-positive bin), relative to ZIKV-negative fraction. All values were normalized to mock controls to control for differences in intrinsic growth rate differences across donors in the village. Linear regression line in black with shaded error in gray. **F**, Schematic model, in which decreased expression levels of *IFITM3* (which associate with the rs34481144-T allele) result in increased ZIKV infectivity in SNaPs.

individual SNaP lines (Figure 3b) and again found a significant correlation between rs34481144 genotype and the percentage of ZIKV-infected SNaPs (at 54 hpi) in each culture, where the rs34481144^{TT} cells were more prone to infection than rs34481144^{CC} SNaPs ($r^2 = 0.345$, $p = 2.53 \times 10^{-3}$; Figure 6a-b). Plotting these results against the Census-seq data showed agreement of results between the pooled-culture and arrayed-culture experimental designs ($r^2 = 0.703$, $p = 2.63 \times 10^{-5}$; Figure S6b).

To further replicate these results, we infected a separate set of 36 human iPSC-derived SNaP lines in an arrayed format using ZIKV-Ug (MOI = 1) or ZIKV-PR (MOI = 10). In these experiments, rs34481144 genotype explained 58.8% and 29.4% (respectively) of the variation in cell survival (Figure 6c-e).

Our focus on the rs34481144 SNP had been driven by the implication of this gene in 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 genome-wide search for effects of common variants on this cellular phenotype. We thus performed genome-wide association analysis on the survival data from the 36 arrayed cell lines, measuring the association of this phenotype with more than 1 million common SNPs. rs34481144 was the human genome's top-scoring SNP in the ZIKV-Ug dataset, reaching genome-wide significance even in this modest 36-donor sample ($p = 9.0 \times 10^{-10}$; Figure 6f). In an analogous analysis of infectivity with the PR strain of ZIKV, rs34481144 also associated strongly, but just below genome-wide significance ($p = 3.2 \times 10^{-7}$; Figure 6g), and was the genome's second strongest common-variant association.

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.

DISCUSSION

Through analyses of natural and synthetic genetic perturbation on cellular phenotypes, we found that *IFITM3* limited ZIKV infection of NPCs, and that a common SNP allele (rs34481144-T) in the *IFITM3* promoter accounted for 58% of the variation in ZIKV infectivity of human NPCs across dozens of genetic backgrounds. This same allele associated with three-fold reduction in expression of *IFITM3* in our experiments on NPCs from many donors (Figure 4e). Variants in the *IFITM3* gene, including SNPs rs12252 and rs34481144, have previously been associated with the severity of influenza infections (Allen et al., 2017; Pan et al., 2017). In PBMCs, the presence of the T allele at rs34481144 alters binding of transcription factors and reduces *IFITM3* promoter activity, leading to reduced baseline expression of *IFITM3* (Allen et al., 2017). Based on our findings, 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 (Auton et al., 2015; Kim and Jeong, 2020), with frequencies as high as 46% in Europe and as low as 0.6% in regions in which flaviviruses are endemic. An intriguing possibility is that evolutionary selection has historically favored the rs34481144-C allele in places with high rates of mosquito-borne RNA virus infections.

A SNaP array-24

C Mock - or - SNaP array-36

ZIKV-Ug
- or -
ZIKV-PR

F Genome-wide Association Analysis (ZIKV-Ug)

G Genome-wide Association Analysis (ZIKV-PR)

Figure 6: Further validation of the *IFITM3* SNP's role in ZIKV infectivity. **A**, Representative images from 24 SNaP lines infected in arrayed format; rows correspond to donor's rs34481144 genotype. **B**, Association of infectivity to rs34481144 genotype. **C**, Schematic showing arrayed infection of 36 SNaP lines with mock media, ZIKV-Ug (MOI = 1), or ZIKV-PR (MOI = 10). **D-E**, Quantification of (D) ZIKV-Ug and (E) ZIKV-PR arrayed infection phenotype to rs34481144 genotype. Linear regression line in black with shaded error in gray. **F-G**, Genome-wide association analysis of ZIKV infectivity across SNaPs from 36 donors infected with (F) ZIKV-Ug and (G) ZIKV-PR. Red line denotes genome-wide significance.

This presents a concerning though likely scenario in which climate change-induced spread of mosquito vectors into Europe (Kraemer et al., 2019) introduces flaviviruses to populations that lack the natural defenses afforded by the protective *IFITM3* allele. Our results suggest that genotypic screening at the rs34481144 locus might be useful in identifying at-risk individuals during these predicted future outbreaks, and supports the enhancement of *IFITM3* expression levels as a potential therapeutic approach in the fight against this growing threat (Zou et al., 2021). Increases in *IFITM3* expression could potentially be accomplished through FDA- and EMA-approved Type I interferon treatments, some of which were recently determined to be safe for pregnant women (Hellwig et al., 2020).

A clear and timely extension of the kinds of analysis in this work could identify SNPs that affect rates of infection and death of human lung cells by SARS-CoV-2, the pathogen responsible for the global pandemic of coronavirus disease 2019 (COVID-19). Previous work has shown that *IFITM3* activities can restrict SARS-CoV and MERS-CoV infections (Huang et al., 2011; Wrensch et al., 2014; Zhao et al., 2014), and a recent study in a relatively small number of patients found an association between rs12252 genotype and COVID-19 severity (Zhang et al., 2020), though this was not observed in a larger dataset (COVID-19 Initiative, 2021).

Outside the realm of human-viral interactions, the population-scale culturing of SNaPs could make it possible to map genetic underpinnings of variation in neurodevelopmental outcomes. Such SNaP villages could be exposed to effectors of major signaling pathways, such as WNT or Notch, could reveal the ways in which genetic profiles impact some of the earliest molecular events of cortical neurogenesis, while exposure to lead and other heavy metals could explain differential susceptibilities to environmental toxins across individuals. Population-scale analysis of SNaP villages could seek answers to basic questions about the etiology of neurodevelopmental disorders, such as the relationship between abnormal NPC proliferation and autism spectrum disorders (Deshpande et al., 2017; Marchetto et al., 2017; Qureshi et al., 2014), as well as the molecular and cellular mechanisms underlying genome-wide association results for autism spectrum disorders and connecting them to other traits, such as brain size.

Population-scale *in vitro* culture systems provide promising ways to capture the influence of genetic variation on cellular phenotypes (Jerber et al., 2021; Kang et al., 2018; Mitchell et al., 2020). We hope that these and many other new approaches open up new ways to find and characterize the many genetic and environmental factors that shape human development.

ACKNOWLEDGEMENTS

This work was supported by NIH/NIMH grants U01MH105669 and U01MH115727, by the Stanley Center for Psychiatric Research at the Broad Institute and the Harvard University Faculty of Arts and Sciences Dean's Competitive Fund for Promising Scholarship. M.F.W. is supported by the Burroughs Wellcome Fund Postdoctoral Enrichment Program award (1018707) and a K99/R00 Pathway to Independence Award (NIH/NIMH 1K99MH119327-01). We thank Janell Smith and Dr. Martin Berryer (Broad Institute) for their assistance in maintaining cell cultures. We acknowledge Olivia Bare, Adam Brown, and David Root (Broad Institute) for their contributions to the design and execution of the CRISPR-Cas9 screen.

AUTHOR CONTRIBUTIONS

Conceptualization and design: MFW, JN, SG, JMM, FP, KE, SAM; Experimentation: MFW, JN, SG, JMM, CM, DM, KR, MT; hPSC line expansion, QC, and sequencing: MT, DH, supervised by RN. Dropulation code: JN, SAM; Data analysis: MFW, JN, SG, DM, DH, BKP; 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

The authors declare no competing interests.

REFERENCES

- Aguet, F., Brown, A.A., Castel, S.E., Davis, J.R., He, Y., Jo, B., Mohammadi, P., Park, Y., Parsana, P., Segrè, A. V, et al. (2017). Genetic effects on gene expression across human tissues. *Nature* 550, 204–213.
- Allen, E.K., Randolph, A.G., Bhangale, T., Dogra, P., Ohlson, M., Oshansky, C.M., Zamora, A.E., Shannon, J.P., 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.
- Auton, A., Abecasis, G.R., Altshuler, D.M., Durbin, R.M., Abecasis, G.R., Bentley, D.R., Chakravarti, A., Clark, A.G., Donnelly, P., Eichler, E.E., et al. (2015). A global reference for human genetic variation. *Nature* 526, 68–74.
- Ballana, E., Pauls, E., Clotet, B., Perron-Sierra, F., Tucker, G.C., and Esté, J.A. (2011). β 5 Integrin Is the Major Contributor to the α v Integrin-Mediated Blockade of HIV-1 Replication . *J. Immunol.* 186, 464–470.
- Barbeito-Andrés, J., Schuler-Faccini, L., and Garcez, P.P. (2018). Why is congenital Zika syndrome asymmetrically distributed among human populations? *PLoS Biol.* 16, 1–11.
- Barbeito-Andrés, J., Pezzuto, P., Higa, L.M., Dias, A.A., Vasconcelos, J.M., Santos, T.M.P., Ferreira, J.C.C.G., Ferreira, R.O., Dutra, F.F., Rossi, A.D., et al. (2020). Congenital Zika syndrome is associated with maternal protein malnutrition. *Sci. Adv.* 6, eaaw6284.
- Borda, V., Junior, R. da S.F., Carvalho, J.B., Morais, G.L., Rossi, Á.D., Pezzuto, P., Azevedo, G.S., Schamber-Reis, B.L., Portari, E.A., Melo, A., et al. (2021). Whole-exome sequencing reveals insights into genetic susceptibility to congenital zika syndrome. *PLoS Negl. Trop. Dis.* 15, 1–17.
- Brasil, P., Pereira, J.P., Moreira, M.E., Ribeiro Nogueira, R.M., Damasceno, L., Wakimoto, M., Rabello, R.S., Valderramos, S.G., Halai, U.-A., Salles, T.S., et al. (2016). Zika Virus Infection in Pregnant Women in Rio de Janeiro. *N. Engl. J. Med.* 375, 2321–2334.
- Brass, A.L., Huang, I.C., Benita, Y., John, S.P., Krishnan, M.N., Feeley, E.M., Ryan, B.J., Weyer, J.L., 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.
- Cauchemez, S., Besnard, M., Bompard, P., Dub, T., Guillemette-Artur, P., Eyrolle-Guignot, D., Salje, H., Van Kerkhove, M.D., Abadie, V., Garel, C., et al. (2016). Association between Zika virus and microcephaly in French Polynesia, 2013-15: A retrospective study. *Lancet*.
- COVID-19 Host Genetics Initiative (2021). Mapping the human genetic architecture of COVID-

19. Nature.

- Darmanis, S., Sloan, S.A., Zhang, Y., Enge, M., Caneda, C., Shuer, L.M., Hayden Gephart, M.G., Barres, B.A., and Quake, S.R. (2015). A survey of human brain transcriptome diversity at the single cell level. *Proc. Natl. Acad. Sci. U. S. A.* 112, 7285–7290.
- DeBoever, C., Li, H., Jakubosky, D., Benaglio, P., Reyna, J., Olson, K.M., 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.
- Deshpande, A., Yadav, S., Dao, D.Q., Wu, Z.-Y., Hokanson, K.C., Cahill, M.K., Wiita, A.P., Jan, Y.-N., Ullian, E.M., and Weiss, L.A. (2017). Cellular Phenotypes in Human iPSC-Derived Neurons from a Genetic Model of Autism Spectrum Disorder. *Cell Rep.* 21, 2678–2687.
- Enard, D., Cai, L., Gwennap, C., and Petrov, D.A. (2016). Viruses are a dominant driver of protein adaptation in mammals. *Elife* 5, 1–25.
- Gomes, J.A., Sgarioni, E., Boquett, J.A., Terças-Trettel, A.C.P., da Silva, J.H., Ribeiro, B.F.R., Galera, M.F., de Oliveira, T.M., de Andrade, M.D.F.C., Carvalho, I.F., et al. (2021). Association between genetic variants in *nos2* and *tnf* genes with congenital zika syndrome and severe microcephaly. *Viruses* 13, 1–13.
- Griesi-Oliveira, K., Fogo, M.S., Pinto, B.G.G., Alves, A.Y., Suzuki, A.M., Morales, A.G., Ezquina, S., Sosa, O.J., Sutton, G.J., Sunaga-Franze, D.Y., et al. (2021). Transcriptome of iPSC-derived neuronal cells reveals a module of co-expressed genes consistently associated with autism spectrum disorder. *Mol. Psychiatry* 26, 1589–1605.
- 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.*
- Helminen, M., Lahdenpohja, N., and Hurme, M. (1999). Polymorphism of the Interleukin-10 Gene Is Associated with Susceptibility to Epstein-Barr Virus Infection. *J. Infect. Dis.* 180, 496–499.
- Huang, I.C., Bailey, C.C., Weyer, J.L., Radoshitzky, S.R., Becker, M.M., Chiang, J.J., Brass, A.L., Ahmed, A.A., Chi, X., Dong, L., et al. (2011). Distinct patterns of IFITM-mediated restriction of filoviruses, SARS coronavirus, and influenza A virus. *PLoS Pathog.* 7.
- Jerber, J., Seaton, D.D., Cuomo, A.S.E., Kumasaka, N., Haldane, J., Steer, J., Patel, M., Pearce, D., Andersson, M., Bonder, M.J., et al. (2021). Population-scale single-cell RNA-seq profiling across dopaminergic neuron differentiation. *Nat. Genet.* 53, 304–312.
- Joung, J., Konermann, S., Gootenberg, J.S., Abudayyeh, O.O., Platt, R.J., Brigham, M.D., Sanjana, N.E., and Zhang, F. (2017). Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nat. Protoc.* 12, 828–863.
- Kang, H.M., Subramaniam, M., Targ, S., Nguyen, M., Maliskova, L., McCarthy, E., Wan, E., Wong, S., Byrnes, L., Lanata, C.M., et al. (2018). Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat. Biotechnol.* 36, 89–94.
- Kenney, A.D., Dowdle, J.A., Bozzacco, L., McMichael, T.M., St. Gelais, C., Panfil, A.R., Sun, Y., Schlesinger, L.S., Anderson, M.Z., Green, P.L., et al. (2017). Human Genetic Determinants of Viral Diseases. *Annu. Rev. Genet.* 51, 241–263.
- Kim, Y.C., and Jeong, B.H. (2020). Ethnic variation in risk genotypes based on single nucleotide polymorphisms (SNPs) of the interferon-inducible transmembrane 3 (IFITM3) gene, a

583 susceptibility factor for pandemic 2009 H1N1 influenza A virus. *Immunogenetics* 72, 447–453.

584 König, R., Chiang, C.Y., Tu, B.P., Yan, S.F., DeJesus, P.D., Romero, A., Bergauer, T., Orth, A.,
585 Krueger, U., Zhou, Y., et al. (2007). A probability-based approach for the analysis of large-scale
586 RNAi screens. *Nat. Methods* 4, 847–849.

587 Kraemer, M.U.G., Reiner, R.C., Brady, O.J., Messina, J.P., Gilbert, M., Pigott, D.M., Yi, D.,
588 Johnson, K., Earl, L., Marczak, L.B., et al. (2019). Past and future spread of the arbovirus
589 vectors *Aedes aegypti* and *Aedes albopictus*. *Nat. Microbiol.* 4, 854–863.

590 Li, Y., Muffat, J., Omer Javed, A., Keys, H.R., Lungjangwa, T., Bosch, I., Khan, M., Virgilio,
591 M.C., Gehrke, L., Sabatini, D.M., et al. (2019). Genome-wide CRISPR screen for Zika virus
592 resistance in human neural cells. *Proc. Natl. Acad. Sci.* 116, 9527 LP – 9532.

593 Majzoub, K., Hafirassou, M.L., Meignin, C., Goto, A., Marzi, S., Fedorova, A., Verdier, Y., Vinh,
594 J., Hoffmann, J.A., Martin, F., et al. (2014). RACK1 controls IRES-mediated translation of
595 viruses. *Cell* 159, 1086–1095.

596 Mao, Y., Ge, X., Frank, C.L., Madison, J.M., Koehler, A.N., Doud, M.K., Tassa, C., Berry, E.M.,
597 Soda, T., Singh, K.K., et al. (2009). Disrupted in Schizophrenia 1 Regulates Neuronal
598 Progenitor Proliferation via Modulation of GSK3 β /Catenin Signaling. *Cell* 136, 1017–1031.

599 Marchetto, M.C., Belinson, H., Tian, Y., Freitas, B.C., Fu, C., Vadodaria, K.C., Beltrao-Braga,
600 P.C., Trujillo, C.A., Mendes, A.P.D., Padmanabhan, K., et al. (2017). Altered proliferation and
601 networks in neural cells derived from idiopathic autistic individuals. *Mol. Psychiatry* 22, 820–835.

602 McLaren, P.J., Coulonges, C., Bartha, I., Lenz, T.L., Deutsch, A.J., Bashirova, A., Buchbinder,
603 S., Carrington, M.N., Cossarizza, A., Dalmau, J., et al. (2015). Polymorphisms of large effect
604 explain the majority of the host genetic contribution to variation of HIV-1 virus load. *Proc. Natl.*
605 *Acad. Sci. U. S. A.* 112, 14658–14663.

606 Mitchell, J.M., Nemesh, J.C., Mello, C., Ghosh, S., Handsaker, R.E., Eggan, K., and Mccarroll,
607 S.A. (2020). Mapping genetic effects on cellular phenotypes with cell villages. Submitted.

608 Mlakar, J., Korva, M., Tul, N., Popović, M., Poljšak-Prijatelj, M., Mraz, J., Kolenc, M., Resman
609 Rus, K., Vesnaver Vipotnik, T., Fabjan Vodušek, V., et al. (2016). Zika Virus Associated with
610 Microcephaly. *N. Engl. J. Med.* 374, 951–958.

611 Nielsen-saines, K., Brasil, P., Kerin, T., Vasconcelos, Z., Gabaglia, C.R., Damasceno, L., Pone,
612 M., Carvalho, L.M.A. De, Pone, S.M., Zin, A.A., et al. (2019). Delayed childhood
613 neurodevelopment and neurosensory alterations in the second year of life in a prospective
614 cohort of ZIKV-exposed children. *Nat. Med.* 25.

615 Nowakowski, T.J., Bhaduri, A., Pollen, A.A., Alvarado, B., Mostajo-Radji, M.A., Di Lullo, E.,
616 Haeussler, M., Sandoval-Espinosa, C., Liu, S.J., Velmeshev, D., et al. (2017). Spatiotemporal
617 gene expression trajectories reveal developmental hierarchies of the human cortex. *Science*
618 (80-). 358, 1318 LP – 1323.

619 Oliva, M., Muñoz-aguirre, M., Kim-hellmuth, S., Wucher, V., Gewirtz, A.D.H., Cotter, D.J.,
620 Parsana, P., Kasela, S., Balliu, B., Viñuela, A., et al. (2020). The impact of sex on gene
621 expression across human tissues. *Science* (80-). 3066.

622 de Oliveira, W.K., de França, G.V.A., Carmo, E.H., Duncan, B.B., de Souza Kuchenbecker, R.,
623 and Schmidt, M.I. (2017). Infection-related microcephaly after the 2015 and 2016 Zika virus
624 outbreaks in Brazil: a surveillance-based analysis. *Lancet* 390, 861–870.

625 Ophir, S., E., S.N., Ella, H., Xi, S., A., S.D., S., M.T., Dirk, H., L., E.B., E., R.D., G., D.J., et al.
626 (2014). Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science* (80-).

627 343, 84–87.

628 Pan, Y., Yang, P., Dong, T., Zhang, Y., Shi, W., Peng, X., Cui, S., Zhang, D., Lu, G., Liu, Y., et
629 al. (2017). IFITM3 Rs12252-C Variant Increases Potential Risk for Severe Influenza Virus
630 Infection in Chinese Population. *Front. Cell. Infect. Microbiol.* 7, 1–6.

631 Qureshi, A.Y., Mueller, S., Snyder, A.Z., Mukherjee, P., Berman, J.I., Roberts, T.P.L.,
632 Nagarajan, S.S., Spiro, J.E., Chung, W.K., Sherr, E.H., et al. (2014). Opposing Brain
633 Differences in 16p11.2 Deletion and Duplication Carriers. *J. Neurosci.* 34, 11199–11211.

634 Retallack, H., Di Lullo, E., Arias, C., Knopp, K.A., Laurie, M.T., Sandoval-Espinosa, C., Leon,
635 W.R.M., Krencik, R., Ullian, E.M., Spatazza, J., et al. (2016). Zika virus cell tropism in the
636 developing human brain and inhibition by azithromycin. *Proc. Natl. Acad. Sci. U. S. A.* 113,
637 14408–14413.

638 Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma
639 powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic
640 Acids Res.* 43, e47–e47.

641 Rossi, S.L., Ebel, G.D., Shan, C., Shi, P.Y., and Vasilakis, N. (2018). Did Zika Virus Mutate to
642 Cause Severe Outbreaks? *Trends Microbiol.* 26, 877–885.

643 Santos, C.N.O., Ribeiro, D.R., Cardoso Alves, J., Cazzaniga, R.A., Magalhães, L.S., De Souza,
644 M.S.F., Fonseca, A.B.L., Bispo, A.J.B., Porto, R.L.S., Santos, C.A. Dos, et al. (2019).
645 Association between Zika Virus Microcephaly in Newborns with the rs3775291 Variant in Toll-
646 Like Receptor 3 and rs1799964 Variant at Tumor Necrosis Factor- α Gene. *J. Infect. Dis.* 220,
647 1797–1801.

648 Savidis, G., McDougall, W.M., Meraner, P., Perreira, J.M., Portmann, J.M., Trincucci, G., John,
649 S.P., Aker, A.M., Renzette, N., Robbins, D.R., et al. (2016a). Identification of Zika Virus and
650 Dengue Virus Dependency Factors using Functional Genomics. *Cell Rep.* 16, 232–246.

651 Savidis, G., Perreira, J.M., Portmann, J.M., Meraner, P., Guo, Z., Green, S., and Brass, A.L.
652 (2016b). The IFITMs Inhibit Zika Virus Replication. *Cell Rep.* 15, 2323–2330.

653 Saxena, S.K., Kumar, S., Sharma, R., Maurya, V.K., Dandu, H.R., and Bhatt, M.L. (2018). Zika
654 virus disease in India - Update October 2018. *Travel Med. Infect. Dis.*

655 Shabalin, A.A. (2012). Matrix eQTL: Ultra fast eQTL analysis via large matrix operations.
656 *Bioinformatics* 28, 1353–1358.

657 Souza, W.V. De, Albuquerque, M.D.F.P.M. De, Vazquez, E., Bezerra, L.C.A., Mendes,
658 A.D.C.G., Lyra, T.M., Araujo, T.V.B. De, Oliveira, A.L.S. De, Braga, M.C., Ximenes, R.A.D.A., et
659 al. (2018). Microcephaly epidemic related to the Zika virus and living conditions in Recife,
660 Northeast Brazil. *BMC Public Health* 18, 1–7.

661 Spence, J.S., He, R., Hoffmann, H.H., Das, T., Thinon, E., Rice, C.M., Peng, T., Chandran, K.,
662 and Hang, H.C. (2019). IFITM3 directly engages and shuttles incoming virus particles to
663 lysosomes. *Nat. Chem. Biol.* 15, 259–268.

664 Stein, J.L., de la Torre-Ubieta, L., Tian, Y., Parikshak, N.N., Hernández, I.A., Marchetto, M.C.,
665 Baker, D.K., Lu, D., Hinman, C.R., Lowe, J.K., et al. (2014). A quantitative framework to
666 evaluate modeling of cortical development by neural stem cells. *Neuron* 83, 69–86.

667 Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y.,
668 Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell
669 Data. *Cell* 177, 1888–1902.e21.

670 Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A.,

671 Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment
672 analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc.*
673 *Natl. Acad. Sci.* *102*, 15545 LP – 15550.

674 Tang, H., Hammack, C., Ogden, S.C., and Jin, P. (2016). Zika Virus Infects Human Cortical
675 Neural Progenitors and Attenuates Their Growth. *Stem Cell* 1–4.

676 Wang, S., Zhang, Q., Tiwari, S.K., Lichinchi, G., Yau, E.H., Hui, H., Li, W., Furnari, F., and
677 Rana, T.M. (2020). Integrin $\alpha\beta 5$ Internalizes Zika Virus during Neural Stem Cells Infection and
678 Provides a Promising Target for Antiviral Therapy. *Cell Rep.* *30*, 1–15.

679 Wells, M.F., Salick, M.R., Wiskow, O., Ho, D.J., Worringer, K.A., Ihry, R.J., Kommineni, S.,
680 Bilican, B., Klim, J.R., Hill, E.J., et al. (2016). Genetic Ablation of AXL Does Not Protect Human
681 Neural Progenitor Cells and Cerebral Organoids from Zika Virus Infection. *Cell Stem Cell* *19*,
682 703–708.

683 Wells, M.F., Salick, M.R., Piccioni, F., Hill, E.J., Mitchell, J.M., Worringer, K.A., Raymond, J.J.,
684 Kommineni, S., Chan, K., Ho, D., et al. (2018). Genome-wide screens in accelerated human
685 stem cell-derived neural progenitor cells identify Zika virus host factors and drivers of
686 proliferation. *BioRxiv* 476440.

687 Wrensch, F., Winkler, M., and Pöhlmann, S. (2014). IFITM proteins inhibit entry driven by the
688 MERS-Coronavirus Spike protein: Evidence for Cholesterol-Independent Mechanisms. *Viruses*
689 *6*, 3683–3698.

690 Wu, D., and Smyth, G.K. (2012). Camera: a competitive gene set test accounting for inter-gene
691 correlation. *Nucleic Acids Res.* *40*, e133–e133.

692 Xu, M., Lee, E.M., Wen, Z., Cheng, Y., Huang, W.-K., Qian, X., TCW, J., Kouznetsova, J.,
693 Ogden, S.C., Hammack, C., et al. (2016). Identification of small-molecule inhibitors of Zika virus
694 infection and induced neural cell death via a drug repurposing screen. *Nat. Med.* *22*, 1101–
695 1107.

696 Zhang, Y., Qin, L., Zhao, Y., Zhang, P., Xu, B., Li, K., Liang, L., Zhang, C., Dai, Y., Feng, Y., et
697 al. (2020). Interferon-induced transmembrane protein-3 genetic variant rs12252-C is associated
698 with disease severity in COVID-19. *J. Infect. Dis.*

699 Zhang, Y.H., Zhao, Y., Li, N., Peng, Y.C., Giannoulatou, E., Jin, R.H., Yan, H.P., Wu, H., Liu,
700 J.H., Liu, N., et al. (2013). Interferon-induced transmembrane protein-3 genetic variant rs12252-
701 C is associated with severe influenza in Chinese individuals. *Nat. Commun.* *4*, 2–7.

702 Zhao, X., Guo, F., Liu, F., Cuconati, A., Chang, J., Block, T.M., and Guo, J. (2014). Interferon
703 induction of IFITM proteins promotes infection by human coronavirus OC43. *111*.

704 Zhou, T., Tan, L., Cederquist, G.Y., Fan, Y., Hartley, B.J., Mukherjee, S., Tomishima, M.,
705 Brennand, K.J., Zhang, Q., Schwartz, R.E., et al. (2017). High-Content Screening in hPSC-
706 Neural Progenitors Identifies Drug Candidates that Inhibit Zika Virus Infection in Fetal-like
707 Organoids and Adult Brain. *Cell Stem Cell* *21*, 274-283.e5.

708 Zhu, Z., Mesci, P., Bernatchez, J.A., Gimple, R.C., Wang, X., Schafer, S.T., Wettersten, H.I.,
709 Beck, S., Clark, A.E., Wu, Q., et al. (2020). Zika Virus Targets Glioblastoma Stem Cells through
710 a SOX2-Integrin $\alpha\beta 5$ Axis. *Cell Stem Cell* *26*, 187-204.e10.

711 Zou, X., Yuan, M., Zhang, T., Zheng, N., and Wu, Z. (2021). EVs Containing Host Restriction
712 Factor IFITM3 Inhibited ZIKV Infection of Fetuses in Pregnant Mice through Trans-placenta
713 Delivery. *Mol. Ther.* *29*, 176–190.

Supplemental Figure 1: In silico evaluation and optimization of the analytical pipeline for cell-to-donor assignment (Dropulation).

A, Schematic of *in silico* experiments. Five human pluripotent stem cell lines were cultured separately (i.e. in an arrayed format) prior to scRNA-seq preparations. Expression data was then pooled computationally for subsequent experiments. **B-F**, Dropulation singlet detection and donor re-identification pipeline. (B) Donor-assignment likelihood calculation, for singlet cells. (C) Plot depicting analysis of donor identity for an example cell. The sequenced cell was correctly assigned to the donor of origin (dark blue circle), distinguishing this donor from two related individuals (light blue circles). (D) Number and (E) fraction of unique molecular identifiers (UMIs) per cell that contain allelic information about at least one informative SNP that varies among the candidate donors. (F) Table depicting the outcomes of donor re-identification attempts when donor genotypes were obtained using whole genome sequencing or SNP arrays. The availability of more-complete genetic data (WGS) increases power in the analysis. **G-I**, Detection of cell-cell doublets. (G) Doublet detection likelihood function used. (H) Example cell in which two donors (red and dark blue circles) show similar likelihoods of assignment before doublets are considered. (I) Doublet detection algorithm successfully recognizes the presence of doublets for exclusion. Color along the line denotes p-value for the discrimination between best singlet and best doublet model. **J-K**, Requirements for successful singlet and doublet detection. Plot showing the number of informative UMIs needed for accurate (J) single cell donor assignment and (K) doublet detection. **L**, Results of blinded experiment identifying donors from a much-larger set of candidates. All 36 cell donors in this test hPSC village were detected and assigned to sequenced cells (blue dots). No cells from >100 other candidate donors (red dots) were detected. **M**, Number of UMIs-per-cell containing allelic information informative SNPs per cell in a 104-donor hiPSC village dataset. **N**, Concordance of estimates of hiPSC village donor composition obtained through Dropulation (by single-cell RNA-seq) and Census-seq (by bulk DNA-seq) analyses.

Supplemental Figure 2: Sources of variation in 104-donor hiPSC village. **A**, PC plots depicting ancestral backgrounds of iPSC donors (orange circles). **B**, Total UMIs per donor for hiPSC village. **C**, tSNE plot of village scRNA-seq data color-coded by scRNA-seq 10X library. **D**, tSNE clusters from village scRNA-seq. Sequenced cells grouped into 14 unique clusters representing proliferative stem cells, differentiated cells, nuclei, and cytoplasmic blebs. **E**, Expression of marker genes for mitotic stem cells (*HIST1H4C*, *TOP2A*), differentiated neural cells (*RIPPLY3*, *SIX6*, *DLK1*, *DPP10*), and differentiation potential-associated genes (*UTF1*, *MT1G*). **F**, Breakdown of donor representation in three representative clusters. Three cell lines represented a majority of low differentiation potential (“UTF1-high”) cluster 6 cells. **G-H**, Cluster 13 cells are likely nuclear fragments as demonstrated by the (G) low numbers of UMIs and (H) high percentage of intronic reads. **I-J**, CAMERA gene enrichment plots highlight signatures detected from (I) cell source and (J) donor sex differential expression analysis.

A

Sign test with GTEx

Supplemental Figure 3: hiPSC village eQTL sign test comparison with human tissue dataset. A, Fraction of SNPs from arrayed and village iPSC datasets that showed the same direction of association to the gene-expression phenotypes in other studies. Similar sign test results were observed between the village and arrayed data when compared to GTEx tissue and cell type data.

Supplemental Figure 4: Quality controls and enrichment analysis results of SNaP ZIKV survival screen. **A**, Correlation matrix of Pearson r values across all replicates (range 0 to 1). **B**, Log2 fold change of individual guides in the ZIKV-Ug (red dots) and ZIKV-PR (blue dots) survival screens. Non-targeting (NT) controls (pink, light blue dots) are also shown. **C-F**, Validation of H1-Cas9 stem cell line. (C) Diagram of AAVS1 targeting vector and site of genomic integration. (D) Immunostaining using an anti-FLAG antibody for clone H1-36-23. Scale bar = 200 μ M. (E) PCR across the junctions for 4 clones, indicating proper targeting into one allele. (F) Fraction indels in neurons differentiated from H1-36-23 clone and infected with lenti-gRNAs, measured by next-generation sequencing. **G-I**, Confirmation of primary screen hits. H1-Cas9 SNaPs were transduced with individual gRNAs, expanded, and exposed to ZIKV-Ug (MOI = 1). At 120hpi, cell viability was measured relative to mock controls. Dashed line denotes viability levels for non-targeting gRNA controls in each batch. **J-L**, Integrin α Vb5 as a ZIKV entry factor. (J) Genetic ablation of *ITGB5* decreases ZIKV-Ug (MOI = 1-10) infectivity relative to controls at 54 hpi. Scale bar = 50 μ M. (K-L) Anti-integrin α Vb5 antibody blockade prevents ZIKV penetration in a dose-dependent manner (IC_{50} = 60 mg/mL). Scale bar = 50 μ M. One-way ANOVA with Dunnett's test for multiple comparisons (G-J, L) was used for statistical analysis. Data presented as mean $\pm$ S.D. N.S. = not significant, * p <0.05, ** p <0.01, *** p <0.001, **** p <0.0001.

Supplemental Figure 5: eQTL analysis reveals variation in gene expression across dozens of SNaP cell lines. A, UMAP cluster plot of Village-44 SNaPs and reference fetal and adult brain cells. **B**, 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. **C-D**, Quantification of computed cell type classification. (C) Seurat 3.0 computed cell type classification for all SNaPs in Village-44 and (D) on a per donor basis. **E**, Village-44 eQTL plots in genes relevant to human neurodevelopment. **F**, Village-44 eQTL plots in genes identified by the survival screens as ZIKV host or antiviral factors. Linear regression line in black with shaded error in gray. **G**, Linkage disequilibrium analysis of 5 statistically-significant eQTLs associated with *IFITM3* gene in Village-44. Box color depicts R^2 value while number within square denotes D' value of pairwise comparisons.

Supplemental Figure 6: Natural variation in *IFITM3*, but not other antiviral genes, correlates with ZIKV infection severity. A, Village-44 eQTLs in ZIKV non-*IFITM3* host factor genes show no relationship with infectivity of SNaPs. B, Synergy between Village-44 Census-seq results and independently-cultured infectivity assay data. Linear regression line in black with shaded error in gray.

SUPPLEMENTAL DATA FILES

Data File 1: Human iPSC village differential gene expression and CAMERA analysis

Data File 2: Human iPSC village eQTL results

Data File 3: ZIKV CRISPR screen results

Data File 4: SNaP village eQTL results