Stem Cell Reports
Resource
ISSCR
OPEN ACCESS

Scaled and efficient derivation of loss-of-function alleles in risk genes for neurodevelopmental and psychiatric disorders in human iPSCs

Hanwen Zhang,[1,8] Ada McCarroll,[1,8] Lilia Peyton,[1,8] Sol Díaz de León-Guerrero,[2,3,8] Siwei Zhang,[1,7] Prarthana Gowda,[2,3] David Sirkin,[1] Mahmoud ElAchwah,[2,3] Alexandra Duhe,[1] Whitney G. Wood,[1] Brandon Jamison,[1] Gregory Tracy,[1] Rebecca Pollak,[4] Ronald P. Hart,[5] Carlos N. Pato,[4] Jennifer G. Mulle,[2,4,6] Alan R. Sanders,[1,7] Zhiping P. Pang,[2,3,*] and Jubao Duan[1,7,9,*]

[1]Center for Psychiatric Genetics, NorthShore University HealthSystem, Evanston, IL, USA
[2]Department of Neuroscience and Cell Biology, Rutgers Robert Wood Johnson Medical School, New Brunswick, NJ, USA
[3]Child Health Institute of New Jersey, Rutgers Robert Wood Johnson Medical School, New Brunswick, NJ, USA
[4]Center for Advanced Biotechnology and Medicine, Rutgers Robert Wood Johnson Medical School, Piscataway, NJ, USA
[5]Department of Cell Biology and Neuroscience, Rutgers University, Piscataway, NJ, USA
[6]Department of Psychiatry, Rutgers Robert Wood Johnson Medical School, New Brunswick, NJ, USA
[7]Department of Psychiatry and Behavioral Neuroscience, The University of Chicago, Chicago, IL, USA
[8]These authors contributed equally
[9]Lead contact
*Correspondence: pangzh@rwjms.rutgers.edu (Z.P.P.), jduan@uchicago.edu (J.D.)
https://doi.org/10.1016/j.stemcr.2024.08.003

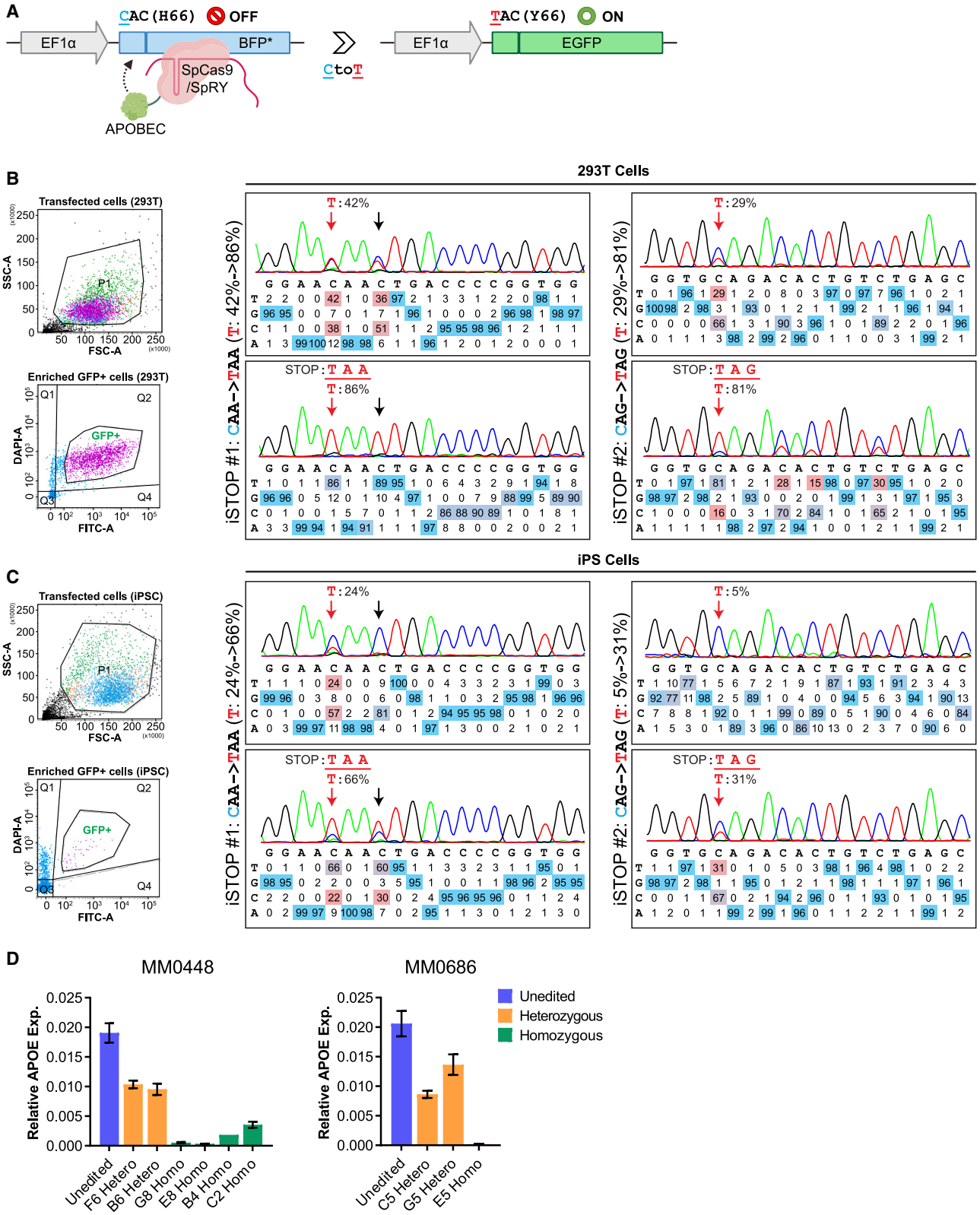
SUMMARY

Translating genetic findings for neurodevelopmental and psychiatric disorders (NPDs) into actionable disease biology would benefit from large-scale and unbiased functional studies of NPD genes. Leveraging the cytosine base editing (CBE) system, we developed a pipeline for clonal loss-of-function (LoF) allele mutagenesis in human induced pluripotent stem cells (hiPSCs) by introducing premature stop codons (iSTOP) that lead to mRNA nonsense-mediated decay (NMD) or protein truncation. We tested the pipeline for 23 NPD genes on 3 hiPSC lines and achieved highly reproducible, efficient iSTOP editing in 22 genes. Using RNA sequencing (RNA-seq), we confirmed their pluripotency, absence of chromosomal abnormalities, and NMD. Despite high editing efficiency, three schizophrenia risk genes (*SETD1A, TRIO*, and *CUL1*) only had heterozygous LoF alleles, suggesting their essential roles for cell growth. We found that *CUL1*-LoF reduced neurite branches and synaptic puncta density. This iSTOP pipeline enables a scaled and efficient LoF mutagenesis of NPD genes, yielding an invaluable shareable resource.

INTRODUCTION

In the past decade, genome-wide association studies (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014; Grove et al., 2019; Howard et al., 2019; Meng et al., 2024; Mullins et al., 2021; Purcell et al., 2009; Ripke et al., 2011, 2013; Shi et al., 2009; Stahl et al., 2019; Stefansson et al., 2009; Trubetskoy et al., 2022; Wray et al., 2018) and whole-exome sequencing studies (Satterstrom et al., 2020; Singh et al., 2022) on neurodevelopmental and psychiatric disorders (NPDs), such as schizophrenia (SZ), autism spectrum disorder, bipolar disorder, and major depression, have identified a growing number of risk genes. However, translating these exciting genetic discoveries into translational actionable biology has been impeded by our limited knowledge of gene function and related disease mechanisms. A bottleneck in the field is that genes are often studied individually, slowing the progress and posing potential bias in functional interpretation. To overcome such limitations, the NIMH (National Institute of Mental Health)-initiated SSPsyGene (Scalable and Systematic Neurobiology of Psychiatric and Neurodevelopmental Disorder Risk Genes) Consortium (sspsygene.ucsc.edu) aims to functionally characterize the contribution of 150–250 NPD genes. The selected NPD genes mostly have disease-associated rare protein-truncating variants (PTVs) that likely cause gene loss of function (LoF) (Palmer et al., 2022; Satterstrom et al., 2020; Singh et al., 2022) and have strong effect sizes (sspsygene.ucsc.edu/resources), which will help interpret their individual biological relevance and determine any convergent or divergent biology across disorders. A large-scale, unbiased, and parallel study of these NPD genes in disease-relevant model systems will substantially deepen our understanding of the pathophysiology of NPDs.

Human induced pluripotent stem cells (hiPSCs) and their derived neural cells empowered by CRISPR-mediated gene editing provide promising cellular models for studying NPD genes (De Los Angeles et al., 2021; Duan, 2023; Michael Deans and Brennand, 2021; Muhtaseb and Duan, 2022; Wang et al., 2020; Wen et al., 2016) and for scaling up the assay. A "cell village" approach (Wells et al., 2023) enables the co-culture of tens to hundreds of hiPSC lines in a dish together, followed by assaying a specific cellular phenotype and being able to genetically infer individual cell identity. Such cell village approach may be combined with pooled screening using CRISPRi (Holtzman and Gersbach, 2018) or CRISPRoff (Nunez et al., 2021) to



(legend on next page)

scale up the number of targeted genes for LoF assay. While an invaluable approach, the pooled CRISPR screening in hiPSC-derived neural models is limited by cell line-specific or LoF allele-specific unequal cellular growth, possible non-autonomous effects, and restrictive phenotypes amenable for screening.

CRISPR-Cas9 editing can be used to systematically create small DNA insertions or deletions (indels) or exon deletions in protein-coding regions through non-homologous end joining repair of double-strand breaks (DSBs) (Ran et al., 2013), resulting in protein-truncating mutations. Alternatively, LoF mutation can be generated by using CRISPR-based cytosine base editors (CBEs) to introduce premature protein stop codons (i.e., nonsense mutations; an iSTOP approach) that lead to mRNA nonsense-mediated decay (NMD) and/or protein truncates (Billon et al., 2017; Cuella-Martin et al., 2021; Hanna et al., 2021; Xu et al., 2021). Compared to the traditional CRISPR-Cas9 gene editing system, the CBE makes "C" to "T" changes in DNAs without creating cell-toxic DSBs as the Cas9 nuclease does (Ran et al., 2013) and with minimized potential off-target DNA editing (Billon et al., 2017; Cuella-Martin et al., 2021; Hanna et al., 2021; Xu et al., 2021). Furthermore, compared to the CRISPR-Cas9 editing-induced small indels that may or may not disrupt a protein sequence reading frame, a CBE can precisely introduce a premature stop codon, which makes the clonal LoF allelic confirmation more straightforward and cost-effective in a scaled LoF mutagenesis workflow. Finally, the CBE-engineered premature stop codon mutations are more reminiscent of the rare patient-specific PTVs or LoF mutations associated with NPD (Satterstrom et al., 2020; Singh et al., 2022). Recently, a DNA base editing reporter gene system has been developed to enrich the edited cells, thereby increasing the base editing efficiency of a target gene (Standage-Beier et al., 2019), including in hiPSCs (Tekel et al., 2021). However, the use of a CBE in editing hiPSC lines has been scarce (Sürün et al., 2020), and its usefulness in developing a scaled and efficient clonal LoF mutagenesis in hiPSCs has not been tested.

As part of the SSPsyGene Consortium, our Assay and Data Generation Center for the Model of iPSC-derived Neurons for NPD (MiNND) aims to employ the CBE-based iSTOP approach to generate isogenic hiPSC lines carrying LoF alleles for about 150–200 NPD genes on multiple donor genetic backgrounds. Here, leveraging an improved reporter gene editing enrichment system that can substantially increase the CBE iSTOP editing efficiency in hiPSCs, we established a semi-automated pipeline for parallel and efficient clonal LoF mutagenesis of a large number of genes. We tested the workflow on 23 NPD genes with 3 donor hiPSC lines (KOLF2.2J, CW20107, and MGS_CD14). We obtained high and reproducible iSTOP editing efficiency across all three hiPSC lines. We systematically characterized the engineered isogenic iSTOP hiPSC lines for pluripotency, karyotyping, neuron differentiation capacity, and the expected NMD and LoF.

RESULTS

The CBEmax DNA base-editing enriching system substantially increases "C" to "T" editing in hiPSCs

A key for generating LoF alleles by using a CBE to introduce premature stop codons (C to T changes; i.e., iSTOP approach) (Billon et al., 2017; Popp and Maquat, 2016) on a large scale is to have sufficiently high gene editing efficiency. Although DNA base editors have high SNP editing efficiency (>50%) in some commonly used cell lines such as HEK293 (Rees and Liu, 2018), hiPSCs are less tested. We opted to employ a base editing reporter gene system to enrich the gene-edited cells (Standage-Beier et al., 2019), thereby increasing iSTOP editing efficiency of a target gene in selected cells. In this CBE editing enriching system (CBEmax_Enrich), a blue fluorescent protein (BFP) reporter on the reporter plasmid pEF-BFP will turn into a functional EGFP reporter when it is edited from CAC (H66) to TAC (Y66) in cells co-transfected with pEF-AncBE4max and single-guide RNAs (sgRNAs) (Figure 1A). We first individually tested the two iSTOP sgRNAs (Table S1) that target the *apolipoprotein E* (*APOE*) gene in HEK293 cells. We transiently co-transfected HEK293 cells with CBEmax_Enrich and sgRNA construct, followed by fluorescence-activated cell sorting

Figure 1. Improved iSTOP base editing efficiency by enriching cells with the reporter gene edited

(A) BFP cassette of the CBEmax_Enrich reporter vector. C to T change turns BFP to EGFP in cells undergoing base editing. APOBEC, apolipoprotein B mRNA editing catalytic polypeptide-like, the base editing enzyme.

(B) High C to T editing efficiency of two iSTOP sgRNAs (iSTOP#1 on the left and iSTOP#2 on the right) in HEK293 cells upon enrichment. Left: the cell gating patterns of the dissociated single cells (transfected; upper) and the editing-enriched GFP^+ cells (lower).

(C) Improved C to T editing efficiency in hiPSCs for the same iSTOP sgRNAs as in (B). Red arrow: intended C to T editing; black arrow: unintended by-standing C to T editing. Editing efficiency in (B) and (C) was calculated by using EditR (Kluesner et al., 2018) that displayed DNA base composition under each sequencing trace peak.

(D) Target gene (APOE) expression knockdown in hiPSC lines homozygous or heterozygous for the T allele after editing using the iSTOP1 and iSTOP2 sgRNAs. Two donor lines are shown, and the mRNA expression was quantified by qPCR and normalized to *GAPDH* expression. $N = 3$ independent cell cultures. Data are presented as Mean±S.E.M.

(FACS) to enrich GFP$^+$ cells (i.e., with the reporter gene edited) for testing editing efficiency by Sanger sequencing (Figure 1B). For each sgRNA, we found a substantial increase in the target gene editing efficiency (C to T) in FACS-sorted GFP$^+$ cells compared to the transfected BFP$^+$ cells (from 42% to 86% and from 29% to 81%, respectively) (Figure 1B).

Next, we similarly tested for the iSTOP editing efficiency in two hiPSC lines (Figure 1C) and whether the introduced iSTOP codons led to the expected NMD (i.e., LoF) (Figure 1D). For both iSTOP sgRNAs, we observed a robust increase, although to a less extent than in HEK293, of the target gene editing efficiency in FACS-sorted GFP$^+$ cells compared to the transfected BFP$^+$ cells (from 24% to 66% and from 5% to 31%, respectively) (Figure 1C). More importantly, as expected from the iSTOP-mediated NMD of mRNAs, we found 86% and 98% of APOE expression reduction in hiPSC clones homozygous for iSTOP1 and iSTOP2, respectively, and ~50% expression reduction in hiPSC clones heterozygous for iSTOP mutations (Figure 1D).

Taken together, these results show that the CBEmax_Enrich system can significantly increase the iSTOP editing efficiency, which enables us to generate LoF alleles on a large scale by introducing premature stop codons.

A scalable workflow for efficiently deriving clonal LoF alleles in hiPSCs using CBEmax_Enrich

Our goal is to develop an efficient pipeline that involves single hiPSC cell sorting for deriving clonal LoF alleles of hundreds of NPD genes in multiple hiPSC lines. To achieve this goal, a key is to obtain a relatively high single hiPSC clonal survival rate after FACS of the enriched GFP$^+$ cells (Figures 1B and 1C). It has been recently shown that the CEPT small molecular cocktail can increase single hiPSC cloning efficiency compared to ROCK inhibitor (Y-27632; ROCK-I) (Tristan et al., 2023). We thus tested the performance of CEPT by treating the hiPSCs with CEPT both during CBEmax_Enrich transfection (for iSTOP sgRNAs of 4 genes) and the FACS sorting of single cells into 96-well plates 48 h post-transfection. However, we observed a very low single hiPSC clonal survivability (~5%) despite a high editing efficiency (~70%) (Figures S1A and S1B). Combining routine ROCK-I treatment of hiPSCs at transfection with CEPT treatment during 48 h post-transfection cell sorting gave us a much higher single hiPSC clonal survivability (~27%), and even higher survivability (~35%) when we sorted cells 72 h post-transfection while maintaining high gene editing efficiency (Figures S1A and S1B).

After these optimizations to achieve high gene editing efficiency and single hiPSC clonal survivability, we designed a semi-automated pipeline for deriving clonal LoF alleles in hiPSCs for 23 NPD genes for each batch (Figure 2A). Briefly, the CBEmax_Enrich vector, the reporter BFP plasmid, and the sgRNA vector carrying the reporter sgRNA and a targeting sgRNA were transiently transfected into hiPSCs in a 24-well plate, each well with one of the 23 targeted LoF mutations or a non-transfection (sgRNA)-control (NTC) for 1 donor hiPSC line. We then sorted out GFP$^+$ cells that were enriched for base editing and distributed 96 single cells per gene/LoF in a 96-well plate. A handful of single hiPSC colonies from each 96-well plate were further subjected to Sanger sequencing to verify the C to T changes (LoF allele). Then 2–3 hiPSC clones, preferably homozygous for a LoF allele, were banked. The selected hiPSC clones were also subjected to RNA sequencing (RNA-seq) to confirm the absence of chromosomal abnormality by eSNP-Karyotyping and pluripotency test. With this pipeline, we have generated LoF alleles, mostly homozygous, for 22 of the 23 selected SSPsyGene Consortium-prioritized NPD genes (no editing found for *HERC1*, Table S3), including 9 (*ARID1B*, *CACNA1G*, *CHD8*, *DLL1*, *GABRA1*, *KMT2C*, *SCN2A*, *SHANK3*, and *SMARCC2*) out of the 10 "capstone genes" (genes prioritized by the SSPsyGene Consortium to be tested for technical consistency across all consortium sites) for which we could design sgRNAs, on 3 donor lines of European ancestry (KOLF2.2J, CW20107, and MGS_CD14) (Figure 2B; Table S2). Our MiNND project within the SSPsyGene Consortium aims to produce LoF alleles for about 150–200 NPD genes on 6 different hiPSC lines. The derivation of many iSTOP LoF alleles enables us to systematically evaluate the performance of iSTOP base editing on hiPSCs and its efficiency in leading to LoF.

iSTOP CBE base editing in hiPSCs is efficient and reproducible in different hiPSC lines

The performance of the CBE in hiPSCs, especially in the context of the iSTOP design and with reporter gene editing enrichment, has not been systematically evaluated. With data from the iSTOP base editing of 23 genes across 3 donor hiPSC lines (Figure 2; Table S3), we found, on average, the post-editing single-cell clonal viability to be 35%–47% (Figure 3A) and the reporter gene editing efficiency to be 31%–50% (Figure S1C). After reporter gene editing enrichment, the average target gene editing efficiency was ~60%, with a strong correlation among different cell lines (Pearson's R = 0.91–0.95) (Figures 3B, S1D, and S1E), demonstrating the highly efficient and reproducible iSTOP CBE editing across all three hiPSC lines. About half of the genes showed editing efficiency higher than 90%, and only 5 genes with editing efficiency less than 10% (including the one without editing) (Figure 3B). Despite the robust increase of target gene editing efficiency after reporter gene editing enrichment (Figure 1), there was a weak correlation between reporter gene editing efficiency and target gene editing efficiency (Figure S1F), suggesting target gene editing

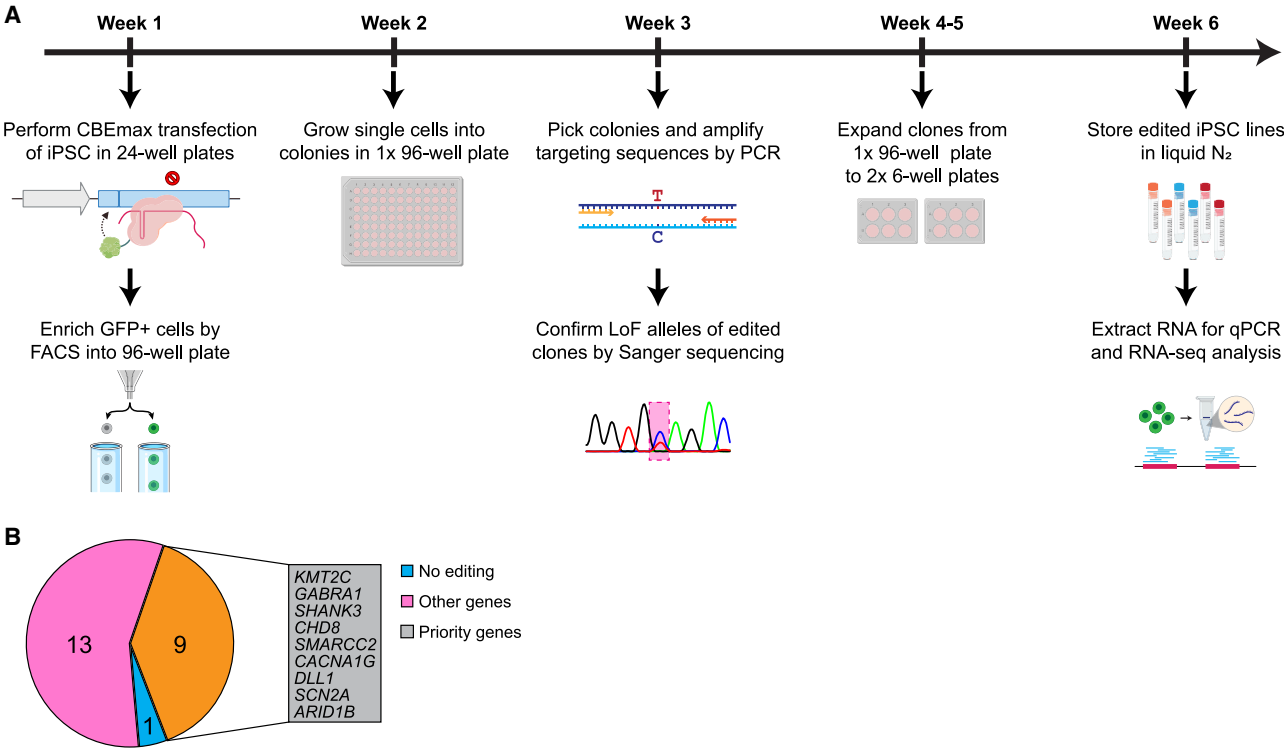


Figure 2. The efficient iSTOP base editing pipeline introduces LoF alleles on a large scale

(A) The workflow that enables the iSTOP editing in batches of 24 (23 target genes +1 control). The hiPSC transfection was performed on a 24-well plate, followed by single hiPSC sorting, single clone expansion on a 96-well plate, clonal sequencing confirmation, and hiPSC banking.

(B) Genes and hiPSC lines were used in the current study. Nine prioritized genes (capstone genes) by the SSPsyGene Consortium are listed in the gray box. Other genes are those NPD risk genes selected by the SSPsyGene Consortium to have the strongest disease associations and highest priorities for creating LoF alleles.

efficiency is predominately determined by gene-specific sgRNA performance.

Overall, we obtained clonal hiPSC lines carrying putative LoF alleles for 22 targeted genes (no editing found for *HERC1*), of which 15 are homozygous (Figure 3B). We found that although the genes with heterozygous LoF alleles tended to have low editing efficiency (<10%) (*ANKRD11*, *KMT2C*, *GABRA1*, and *AKAP11*), some had high editing efficiency (*SETD1A* with 29%–38%, *TRIO* with 46%–71%, and *CUL1* with 33%–50%) (Figure 3B), suggesting that for some NPD genes homozygous LoF alleles may have deleterious effects on hiPSC survival or growth. It is noteworthy that all three genes (*SETD1A*, *TRIO*, and *CUL1*) with only heterozygous LoF clones, despite their relatively high editing efficiency, are top-ranking SZ risk genes found by the SZ Exome Sequencing Meta-Analysis (SCHEMA) Consortium to have rare and highly penetrant SZ-associated PTVs (Singh et al., 2022). Of these genes, *TRIO* was found to initially have homozygous hiPSC clones grown in the 96-well plate post sorting; however, only heterozygous clones (Figure 3C) were found to show sustained normal hiPSC growth, which is consistent with the known necessary role of *TRIO* for cell migration and growth (Deinhardt et al., 2011; Seipel et al., 1999).

The CBE-edited iSTOP hiPSC clones are pluripotent and have minimal chromosomal abnormalities

Next, we characterized the selected iSTOP hiPSC clones for stem cell pluripotency, chromosomal abnormalities, and neuron differentiation capability. Immunofluorescence staining of stem cell pluripotency markers (OCT4, SOX2, and TRA-1-60) of the engineered hiPSC lines for 6 selected LoF alleles confirmed their pluripotency (Figures 4A and S2A). To further evaluate the pluripotency of all the selected hiPSC LoF clones at the genomic and molecular level, we carried out RNA-seq for each hiPSC clone and used CellNet to quantify how closely the engineered hiPSC populations transcriptionally resembled human embryonic stem cells (ESCs) compared to other non-ESC somatic cells (Cahan et al., 2014). All hiPSC clones exhibited high stemness scores (0.93–0.97) and no traces of other somatic cell types (Figures 4B, S2B, and S2C). With the same

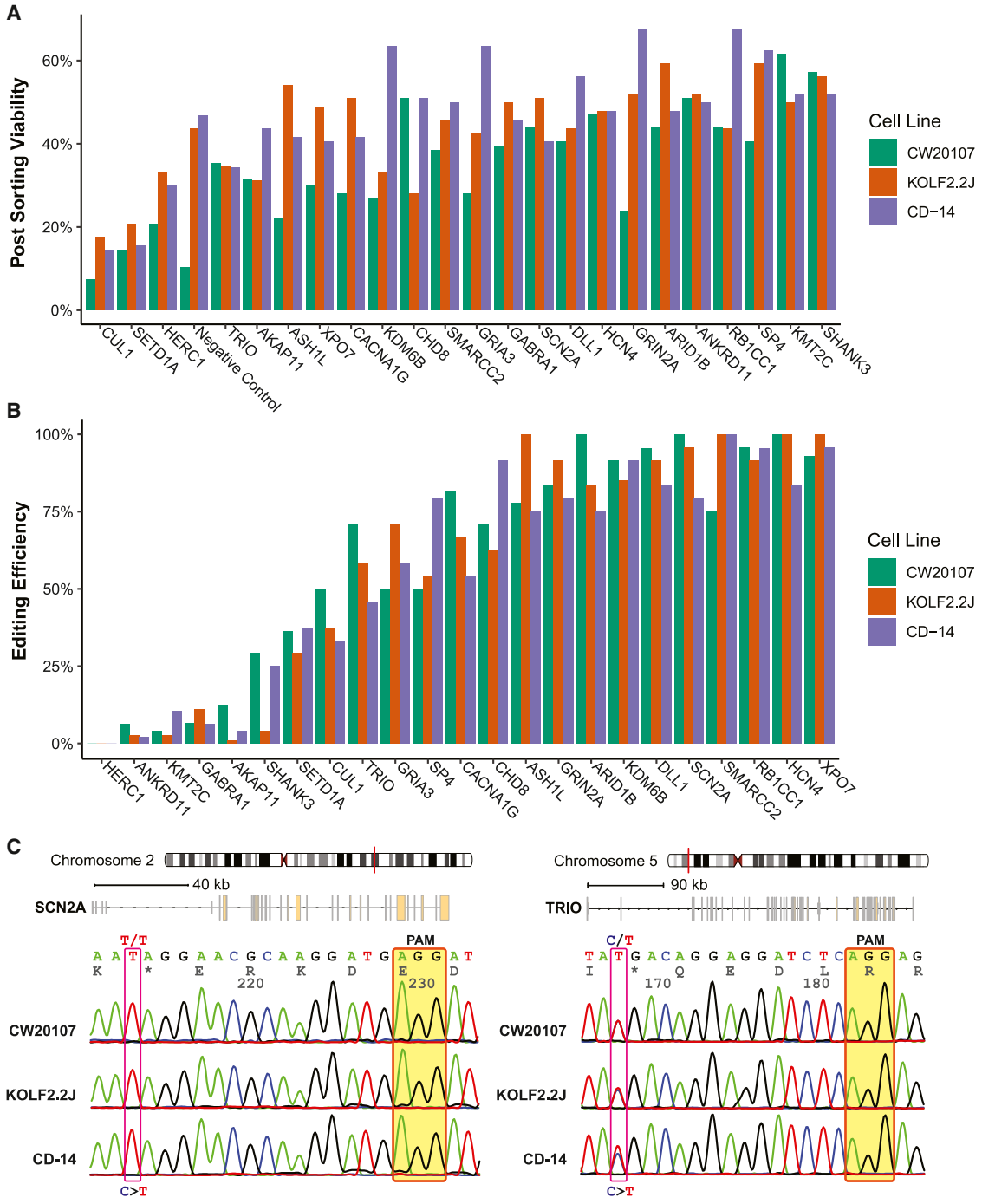


Figure 3. High and reproducible C to T base editing efficiencies across genes and hiPSC lines

(A) High rate of single hiPSC clonal survivability after post-transfection cell sorting.

(B) High iSTOP LoF allele editing efficiency and reproducibility. The genotypes were confirmed by Sanger sequencing for the selected individual hiPSC clones from each gene editing.

(C) Examples of Sanger sequencing traces to confirm LoF alleles of two genes, *SCN2A* (left) and *TRIO* (right), in all three hiPSC lines. The shown sequencing traces are near the iSTOP-sgRNA region, with the PAM sequence highlighted in transparent yellow boxes and the genotype of the LoF mutation site marked in red box.

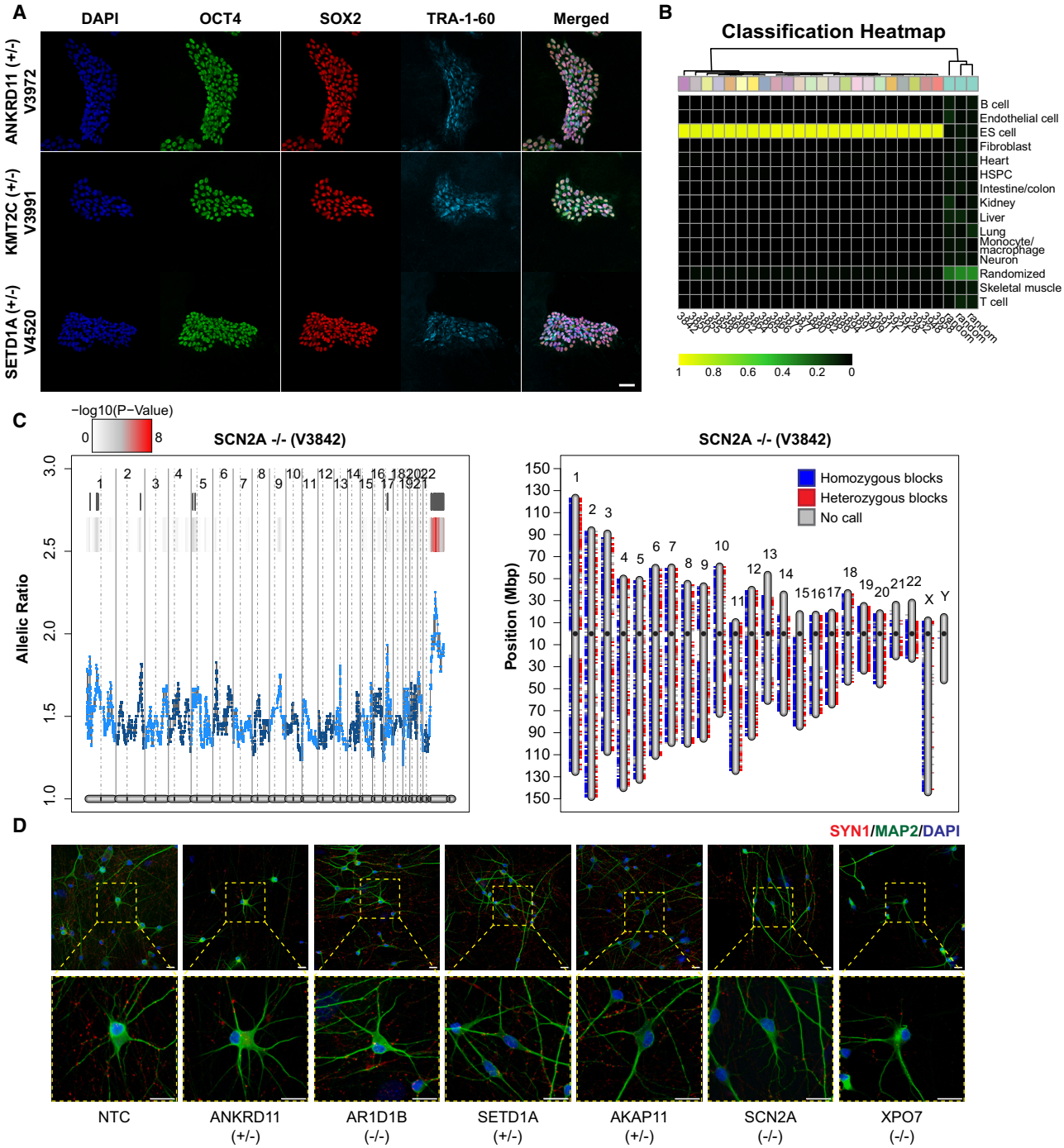


Figure 4. Characterization of isogenic base-edited hiPSC lines carrying iSTOP LoF alleles

(A) The iSTOP mutant lines were stained positive for pluripotent stem cell markers (OCT4, SOX2, and TRA-1-60). Scale bar: 50 μm.

(B) CellNet analysis of RNA-seq data of hiPSC lines confirmed their pluripotency. Pluripotency scores showed transcriptional similarity of the edited iSTOP LoF hiPSC lines to ESC but not other non-ESC cell types. Only one batch of hiPSC lines is shown, and the data for the other two batches are in Figure S2.

(legend continued on next page)

RNA-seq data, we also confirmed the absence of large chromosomal abnormalities due to base editing using eSNP-Karyotyping (Weissbein et al., 2016; Zhang et al., 2023) (Figures 4C and S3) by analyzing the moving average of the allelic ratio of each expressed SNP along the genome (left panel) and stretches of common SNP heterozygosity on each chromosome for each batch of hiPSC lines. In total, for the assayed 69 hiPSC lines of all three batches, eSNP-Karyotyping identified only one hiPSC line from batch 3 (NTC line with barcode V4167) showing chromosomal abnormality on Chr12 (57759165–76048101 bp; hg38) (Figure S3 and https://zenodo.org/records/11591445; Table S4). Finally, we confirmed that all the selected iSTOP hiPSC lines (n = 6) could be successfully induced into neurons ($MAP2^+/Syn^+$) after Ngn2 transduction (Figure 4D).

Most iSTOP hiPSC lines show the expected mRNA or protein reduction with the confirmation of SHANK3 LoF phenotype in Ngn2-induced neurons

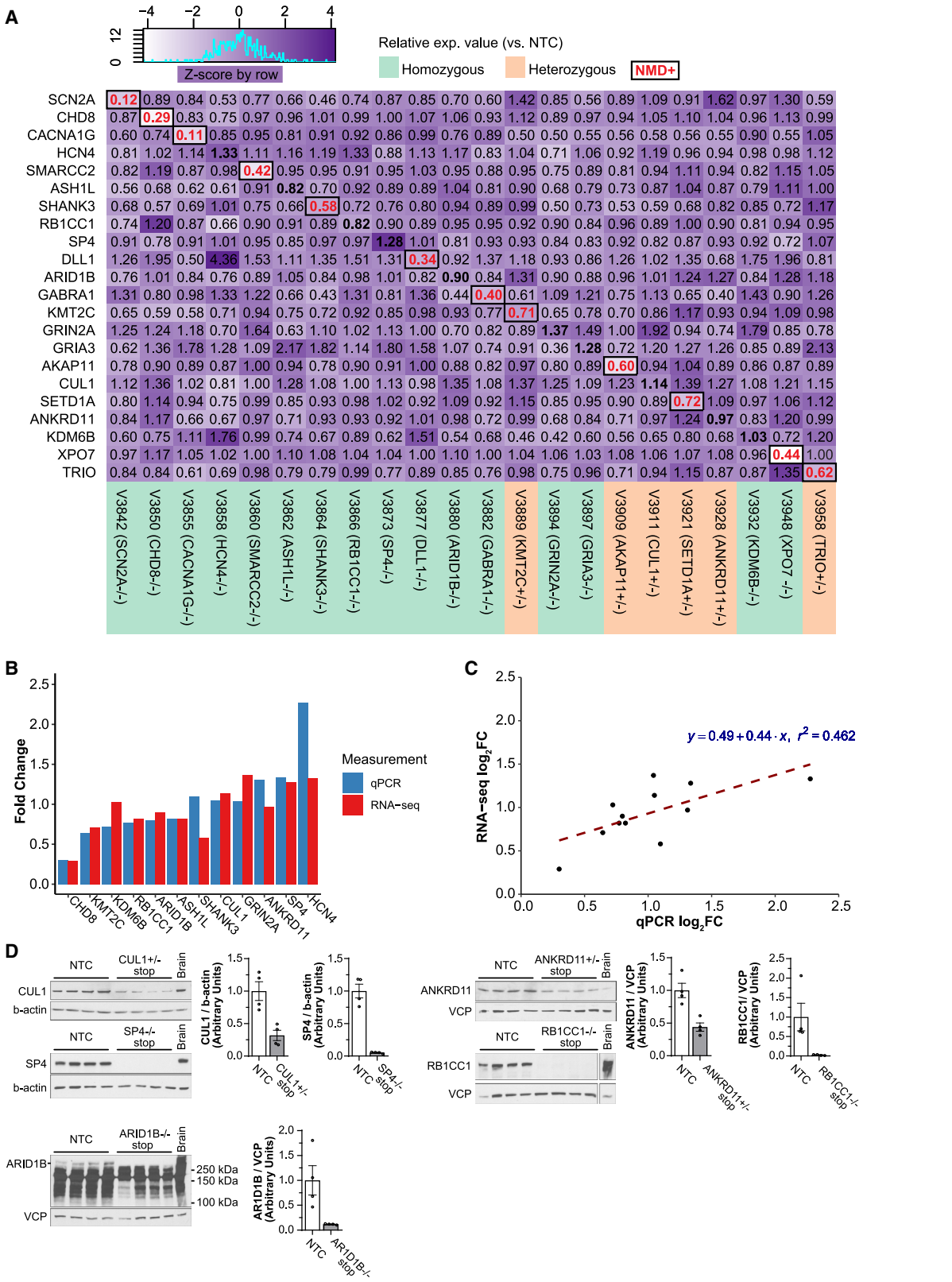
Because we have employed the iSTOP approach to introduce premature stop codons that are predicted to cause NMD (Table S1), we first tested whether we could observe the expected expression reduction for each NPD gene in the engineered hiPSC lines using RNA-seq data. Compared to the unedited cell line, the iSTOP lines for about 12 genes showed partial or near-complete expression knockdown (KD) (compared to the unedited line) as expected for NMD (Figure 5A). Comparing the Z-scored expression values (counts per million reads) of each gene in each iSTOP LoF hiPSC line for all three batches (Figure S4) showed similar patterns of reduced gene expression of those 12 genes as in Figure 5A. The strongest expression KD was observed for hiPSC lines homozygous for *SCN2A*, *CHD8*, and *CACNA1G* iSTOP LoF alleles, exhibiting a 70%–90% expression reduction. The lack of the expected NMD for some genes may be due to possible cell type-specific NMD regulation (Huang et al., 2011), incomplete mRNA degradation, or inaccurate NMD prediction in sgRNA design. Our qPCR further confirmed the incomplete mRNA degradation or even increased mRNA production (e.g., *HCN4*, *SP4*) for genes that did not show the expected NMD in RNA-seq (Figure 5B).

Regardless of any detectable NMD from RNA-seq or qPCR, we expected those premature stop codons at the first half of a target gene would result in protein truncations (i.e., the loss of full-length proteins). To confirm this hypothesis, we performed western blotting for 5 selected genes that did not show the expected mRNA NMD (*CUL1*, *ANKRD11*, *SP4*, *RB1CC1*, and *ARID1B*) (Figures 5A and 5B) using cell lysates of their respective iSTOP hiPSC lines ($CUL1^{+/-}$, $ANKRD11^{+/-}$, $SP4^{-/-}$, $RB1CC1^{-/-}$, and $ARID1B^{-/-}$) (Figure 5D). Here, we have included both homozygous and heterozygous lines to examine whether we can observe the expected dosage-dependent LoF effect on protein expression. We found that compared to the unedited hiPSC line (NTC), all 5 LoF lines showed the expected protein reduction based on their genotype, with heterozygous LoF lines showing ~50% decrease of the intact proteins (CUL1 and ANKRD11). In contrast, homozygous LoF lines exhibited near-complete KD (SP4, ARID1B, and RB1CC1) (Figure 5D). To confirm whether the antibody used for western blot targets the major transcript isoforms that did not show the expected NMD of mRNA, we reanalyzed the RNA-seq data using a pseudo-alignment-based model (Kallisto) (Bray et al., 2016) to obtain isoform expression in each hiPSC line (Table S5). We found that each antibody targets all major isoforms that did not show the expected NMD for each gene except for *SP4*, in which the antibody targets only one major isoform (Figure S5; Table S5); however, the *SP4* transcript isoform targeted by the antibody accounts for ~61% of the *SP4* expression and also showed an increase of RNA expression in the $SP4^{-/-}$ line (Figures 5A and 5B). Thus, despite the absence of the expected NMD for *CUL1*, *ANKRD11*, *SP4*, *RB1CC1*, and *ARID1B*, the detected protein KD of these genes was unlikely due to alternative isoforms not detected by an antibody. All together, these results strongly suggest that most LoF alleles engineered by our iSTOP base editing approach effectively led to the expected gene expression KD or complete abolishment of the gene expression, i.e., a LoF effect, at the protein level.

To further corroborate the LoF effect by the introduced iSTOP mutation, we assessed whether we could replicate the previously reported morphological phenotypes in *SHANK3*-deficient human neurons (Yi et al., 2016). We derived excitatory (Ex) and inhibitory (Inh)-induced neurons by ectopic expression of the Ngn2 or Ascl1/Dlx2 transcription factors (Halikere et al., 2020; McGowan et al., 2018; Yang et al., 2017; Zhang et al., 2013) (see Methods) from the hiPSC lines that carry homozygous iSTOP LoF allele of *SHANK3*. We also included an edited hiPSC line that carried the iSTOP LoF allele of *CUL1*, a strong SZ risk gene

(C) eSNP-Karyotyping showed no large chromosomal abnormalities. Left: moving average of the allelic ratio of each expressed SNP along the genome; right: stretches of homozygosity on each chromosome. An example of one hiPSC line is shown. Data for all the lines are available at https://zenodo.org/records/11591445.

(D) Some selected iSTOP LoF hiPSC lines were successfully differentiated into excitatory neurons ($Syn^+/MAP2^+$). Images were taken with two different magnifications (20× and 63×). Scale bar: 20 μm. NTC, non-transfected control line. In (A) and (C), the gene name for the LoF allele and the cell line number (starting with "V") were listed.



(legend on next page)

(Singh et al., 2022) that only had heterozygous clones (Figure 3B). With the co-cultures of Ex- and Inh neurons, we assayed the neurite outgrowth, branches, and synaptic puncta density of the tdTomato-labeled Ex neurons using high content imaging (HCI). The built-in neurite outgrowth module on the ImageXpress system was used for cell segmentation for assaying neurite outgrowth and branches (Figure S6A), and a customized synaptic puncta module with binary masks was used for assaying puncta density (Figure S6B). Compared to the Ex neurons from the isogenic control hiPSC line, the *SHANK3* iSTOP LoF line showed ~2/3 reduction of neurite outgrowth and branches but no significant change of synaptic puncta (Synapsin1^{+}) density (Figures 6 and S6), which are partially consistent with the reported cellular phenotypes of *SHANK3*-haploinsufficiency in human neurons: *SHANK3*$^{-/-}$ neurons showed a reduction of both neurite length/branches and synaptic puncta, while *SHANK3*$^{+/-}$ neurons only showed a reduction of neurite length/branches (Yi et al., 2016). For the SZ risk gene *CUL1*, we observed a significant reduction of Ex neuronal neurite outgrowth and branches by ~60% as well as a reduced synaptic puncta (Synapsin1^{+}) density by ~40% (Figures 6 and S6). It is noteworthy that neurons with Cul1 deficiency exhibit severe dendrite pruning defects in *Drosophila* (Wong et al., 2013), while small interfering RNA KD of *Cul1* in rat hippocampal neurons increases synaptic F-actin but decreases Synapsin1 (Falkovich et al., 2023).

To validate the observed neurite deficits (Figure 6) in a sparse neuronal culture, we performed an independent experiment with a lower cell density (Figures S6C and S6D). We found that the patterns of the neurite outgrowth/branch differences for *CUL1* and *SHANK3* LoF lines (compared to NTC line) were overall similar between "higher" and "lower" density neuron groups (Figures S6C and S6D). However, the phenotypes were less pronounced in lower density cultures, highlighting the importance of specific experimental conditions for detecting the dramatic deficits observed in higher density cultures in HCI. Taken together, our observed NMD of mRNA, protein expression reduction, and neural phenotypic changes in the assayed iSTOP hiPSC lines collectively suggest that most iSTOP hiPSC lines for NPD genes are expected to show LoF effect on protein expression.

DISCUSSION

Although CRISPR editing of individual NPD risk genes/variants in hiPSCs has been widely used in the past decade (De Los Angeles et al., 2021; Duan, 2023; Michael Deans and Brennand, 2021; Muhtaseb and Duan, 2022; Wang et al., 2020; Wen et al., 2016; Zhang et al., 2020, 2023), a scaled and efficient pipeline for clonal LoF mutagenesis in hiPSCs has not been established. Our reported CBE iSTOP editing workflow benefited from the improved gene editing efficiency of the CBEmax_Enrich system, the precision of iSTOP mutagenesis, the streamlined RNA-seq-based assays for pluripotency, eSNP-Karyotyping, and iSTOP-mediated NMD and/or LoF. These factors simplified the workflow and made it more amenable for automation to increase throughput while keeping the pipeline cost-effective. Moreover, because our pipeline only involved transient transfection of hiPSCs, the engineered iSTOP hiPSC lines were genome-integration-free, as opposed to CRISPR pooled screening that often entails hiPSC genome integration with exogenous virus fragments that may confound downstream phenotypic assay readouts. The derived iSTOP hiPSC lines carrying LoF alleles for the current list of 22 (out of 23) edited genes on 3 donor hiPSC lines, and many more engineered LoF hiPSC lines to be generated, will be an invaluable sharable resource for the NPD genetics research community.

We observed high editing efficiency for most genes that were highly reproducible across all three hiPSC lines, suggesting CBE iSTOP performance was not hiPSC line-specific but rather mainly determined by sgRNAs. As expected, genes with high editing efficiency tended to have more clones homozygous for iSTOP LoF alleles. Interestingly, for three strong SZ risk genes (*SETD1A*, *CUL1*, and *TRIO*) identified by SCHEMA, we only obtained heterozygous LoF hiPSC clones despite high editing efficiency, suggesting the likely deleterious effect of LoF on stem cell survival. Indeed, possible lethal effects of LoF of the three genes are

Figure 5. Characterization of NMD and LoF for iSTOP LoF hiPSC lines

(A) Heatmap of relative expression (Z-scored) of each mutant line (vs. NTC) using normalized RNA-seq expression value (CPM, count per million). The values in the diagonal boxes show the fold change of a specific line with LoF mutation. The fold change values in red fonts indicate those showing NMD.

(B) qPCR confirmation of the expression fold change in RNA-seq (vs. NTC).

(C) Strong correlation of expression fold changes (vs. NTC) between RNA-seq and qPCR data.

(D) Western blots showed the expected reduction of protein abundance for the LoF alleles of 5 selected genes that did not exhibit NMD in (A) and (B). NTC, non-transfected control. Mouse brain protein extracts were used as a positive control for each blot. Note that non-specific signals below 250 kDa, expected sizes for ARID1B, were apparent in the blots; only the putative signal of ARID1B was quantified. Each lane represents a cell lysate from 4 independent cell cultures of 2 different passages (2 for each passage). The parent line is CW20107. Data are presented as: Mean±S.E.M.

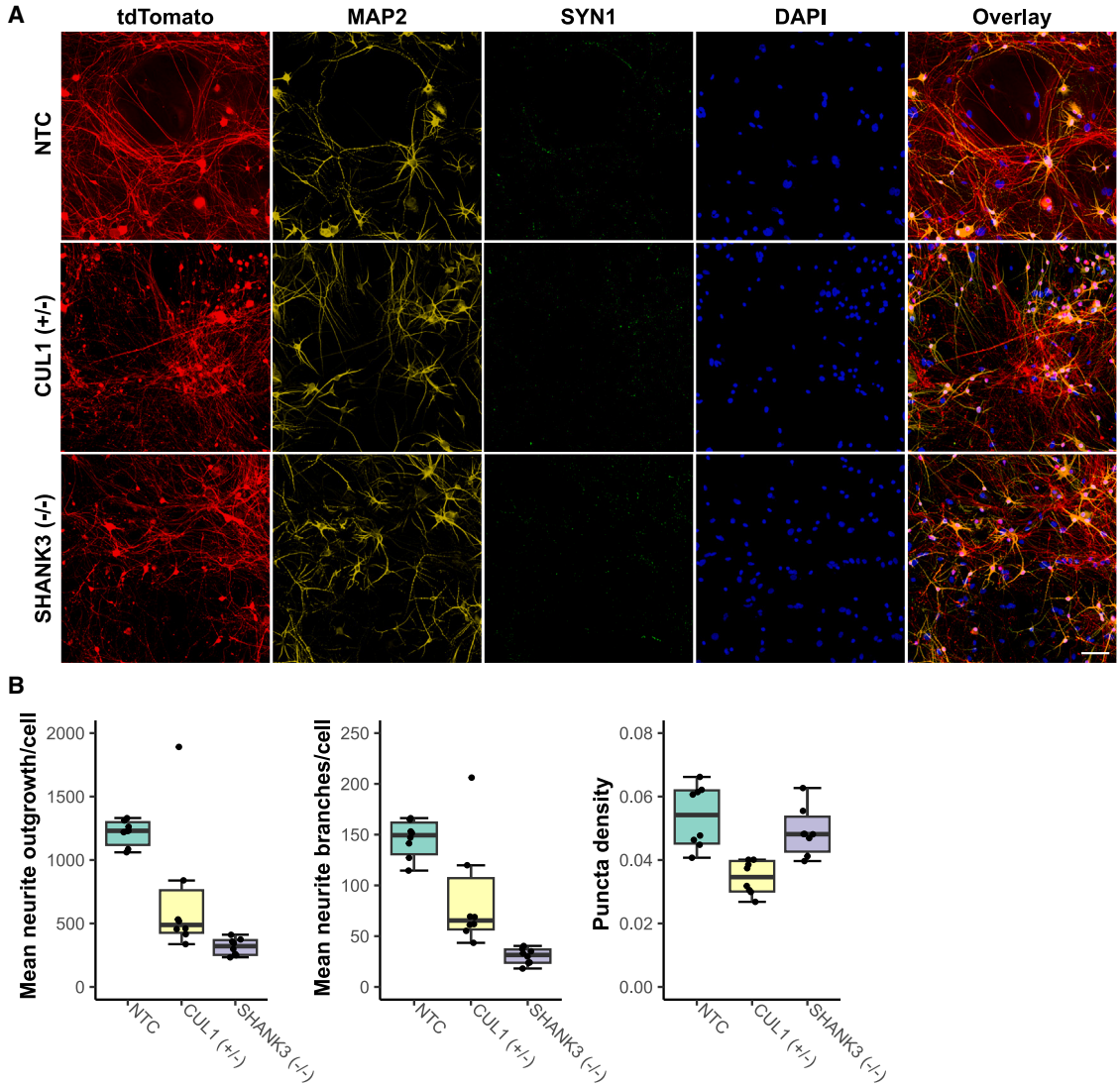


Figure 6. High-content imaging of excitatory and inhibitory neurons co-cultured with mouse glia for iSTOP LoF hiPSC lines

(A) Representative images of neural cultures with hiPSC-differentiated excitatory neurons labeled by tdTomato (red). Neurons were stained for MAP2, SYN1, tdTomato, and DAPI. TdTomato and DAPI staining were used to quantify neurite growth/branches; tdTomato, MAP2, and SYN1 staining were used to analyze synaptic puncta. Scale bar: 20 μm.

(B) Summarized imaging result for neurite outgrowth and branches and synaptic puncta (left to right) for SHANK3 (−/−) and CUL1 (+/−) LoF alleles on donor hiPSC line CW20107. Each data point represents the measurement of an independent well of neuron culture on a 96-well plate; $N = 8$ wells. The upper limit of the horizontal line is 5%, and the lower limit is 95%; the box has three limites: 25%, 50% and 75%. All %s are fractions as defined by the Box and Whisker plot function in Prism 7.

supported by the existing body of literature: *Setd1a*, encoding a histone methyltransferase, was found to be required for embryonic and neural stem cell survival (Bledau et al., 2014), and only heterozygous *SETD1A*-haploinsufficiency hiPSC lines (Chong et al., 2022; Wang et al., 2022; West et al., 2019) or mouse models (Mukai et al., 2019; Nagahama et al., 2020) have been reported for functional characterization; *CUL1* has E3 ubiquitin-protein ligase activity and homozygous deletion of *Cul1* in mice causes arrest in early embryogenesis and embryonic lethality at E6.5 (Wang et al., 1999); and *TRIO* functions as a guanosine diphosphate to guanosine triphosphate exchange factor and is necessary for cell migration and growth (Deinhardt et al., 2011; Seipel et al., 1999). It is noteworthy that highly penetrant patient-specific PTVs in these NPD genes are all heterozygous; thus, the obtained heterozygous LoF hiPSC clones can be valuable for ascertaining more disease-relevant cellular phenotypes.

Our clonal LoF mutagenesis pipeline leverages the precise control of the CBE iSTOP editing approach to introduce a premature stop codon that is predicted to cause NMD of mRNAs or lead to a protein truncation. However, only 12 out of the 22 edited NPD genes showed partial or nearly complete NMD in hiPSC. Other than possible inaccuracy of NMD prediction, the lack of the expected NMD for some genes is likely due to: (1) RNA-seq may have still detected partially degraded mRNAs; (2) an intricate feedback network maintains both RNA surveillance and the homeostasis of normal gene expression in mammalian cells (Huang et al., 2011); and (3) some cells do escape NMD either by translational readthrough at the premature stop codon or by a failure of mRNA degradation after successful translation termination (Sato and Singer, 2021). However, our western blotting analysis showed that most iSTOP mutations, even without detectable NMD, likely led to LoF by yielding a truncated protein too short to be detected. Although a truncated protein may arguably have normal function or even gain of function, the sgRNA in our iSTOP design often targets the protein N-terminal, thus more likely to show LoF. It is reassuring that even for genes that do not show NMD (e.g., *CUL1*) or even increased mRNA level (likely due to negative feedback regulation, e.g., *SP4*) in hiPSCs, our western blotting showed the expected protein expression KD, suggesting most iSTOP mutant alleles are likely to show LoF in neurons.

Among other limitations, our iSTOP design relies on NMD or protein truncation to achieve LoF, where the extent of LoF may not be as complete as obtained from complete knockout (KO) using CRISPR-Cas9. However, a complete gene KO may need multiple sgRNAs targeting multiple genomic regions, which poses a challenge for clonal LoF mutagenesis on a large scale and, more importantly, may cause large chromosomal arrangements. Moreover, although our scaled pipeline is suitable for studying gene LoF, it may not apply to some patient-specific mutations (or natural alleles) or copy-number variants that do not involve C to T changes but are also important for NPD. Furthermore, we acknowledge the limitation of our eSNP-Karyotyping in its inability to detect indels of noncoding regions or balanced translocations as well as its low sensitivity for genes with low expression in hiPSCs, which may leave some genetic lesions undetected in our present study. Future G-band karyotyping or genome-wide genotyping will provide a more comprehensive characterization of different types of genomics abnormalities in these hiPSC lines. Lastly, our clonal iSTOP LoF mutagenesis pipeline may be complemented by pooled CRISPR screening in combination with the rapidly evolving spatial transcriptomics and phenotyping to ascertain LoF allelic effects at single-neuron resolution simultaneously. Albeit limitations, our scaled and efficient clonal LoF mutagenesis pipeline showed robust performance in generating easily sharable individual hiPSC lines carrying LoF alleles for a large number of NPD genes. In addition to engineering LoF alleles, the pipeline can be easily adopted for precise SNP editing (C to T changes by CBE, and A to G changes by adenine base editors [ABE]) in hiPSC for studying coding or noncoding disease risk variants. The scaled LoF mutagenesis pipeline thus empowers hiPSC as a promising cellular model for understanding the disease biology of NPD and other complex genetic disorders.

EXPERIMENTAL PROCEDURES

Methods

hiPSC lines and cell culture

CW20107 was from the California Institute for Regenerative Medicine. KOLF2.2J was an updated version of KOLF2.1J (Pantazis et al., 2022) and is available at The Jackson Laboratory via special request. The other hiPSC line (CD14) is specific to the MiNND project and is from the Duan lab (Shi et al., 2009; Zhang et al., 2020, 2023). CD14 was originally derived from the lymphocytes of a healthy donor of the Molecular Genetics of Schizophrenia cohort (Shi et al., 2009). Detailed information is in Table S2. The institutional review board of NorthShore University HealthSystem approved the study.

Gene selection and iSTOP base editing design

The reported 23 genes were part of the ~250 NPD genes selected by SSPsyGene Consortium (sspsygene.ucsc.edu). For designing iSTOP sgRNAs, the iSTOP web tool was used (Billon et al., 2017).

iSTOP base editing pipeline

The LoF mutagenesis was performed in batches, each containing 23 genes and an NTC on a 24-well plate format. The sorted single cells were cultured on a 96-well plate for Sanger sequencing genotyping.

hiPSC characterization

LoF mutant lines were characterized for stem cell pluripotency by both immunofluorescence staining and by using CellNet analysis of RNA-seq data (Cahan et al., 2014). For chromosomal abnormality, we used eSNP-Karyotyping (Weissbein et al., 2016) as described (Zhang et al., 2020, 2023).

Neuron differentiation from hiPSCs

We used the two commonly used methods: NGN2 + rtTA for excitatory neuron differentiation (Zhang et al., 2013) and ASCL1 + DLX2 + rtTA for inhibitory neuron differentiation (Yang et al., 2017).

Neuron morphological characterization

The neurons were imaged using Molecular Devices' (San Jose, CA) ImageXpress Micro Confocal High-Content Imaging System. The neurite phenotypes were analyzed using the built-in module. We used a customized synaptic assay module to assay synapse density.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Jubao Duan (jduan@uchicago.edu).

Materials availability

The hiPSC lines will be made available as part of the SSPsyGene Consortium to fulfill the NIMH (National Institute of Mental Health) material/data-sharing commitment.

Data and code availability

All the reported data and code used during analysis, including the full hiPSC eSNP-Karyotyping results are deposited at https://doi.org/10.5281/zenodo.13273149. The RNA-seq data's GEO accession number is GSE262442.

ACKNOWLEDGMENTS

Data were generated as part of the SSPsyGene Consortium, supported by RM1MH133065 awarded to Z.P.P., J.D., and J.G.M. We thank SSPsyGene Consortium members for selecting NPD genes and providing hiPSC lines CW20107 (from CIRM) and KOLF2.2J (from the Jackson Lab). We thank Molecular Genetics of SZ (MGS) investigators for collecting samples that were used to derive MGS hiPSC lines. Funding was provided by NIH grants R01AA023797 and R01MH125528 (to Z.P.P.), R01MH106575, R01MH116281, and R01AG063175 (to J.D.), and RM1MH133065 (to Z.P.P., J.D., and J.G.M.). A full BioRender license to the University of Chicago was used for visual depiction and graph making.

AUTHOR CONTRIBUTIONS

H.Z., L.P., and A.M. performed the main experiments; S.D.d.L.G. performed the neuron differentiation and characterization; S.Z. analyzed the RNA-seq data and created the figures, tables, and the graphic abstract; P.G. and M.E. performed neuron differentiation and characterization; D.S. and G.T. performed the neuron imaging analyses; A.D. assisted with RNA-seq data analysis; W.G.W. and B.J. assisted with hiPSC derivation and culture; R.P. assisted with the project coordination; R.P.H., C.N.P., J.G.M., and A.R.S. assisted with project design and result interpretation; Z.P.P. and J.D. conceived and supervised the project and wrote the manuscript. All authors contributed to the manuscript writing and editing.

DECLARATION OF INTERESTS

The authors declare no competing interests.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.stemcr.2024.08.003.

Received: March 26, 2024
Revised: August 8, 2024
Accepted: August 10, 2024
Published: September 12, 2024

REFERENCES

Billon, P., Bryant, E.E., Joseph, S.A., Nambiar, T.S., Hayward, S.B., Rothstein, R., and Ciccia, A. (2017). CRISPR-Mediated Base Editing Enables Efficient Disruption of Eukaryotic Genes through Induction of STOP Codons. Mol. Cell *67*, 1068–1079.e4. https://doi.org/10.1016/j.molcel.2017.08.008.

Bledau, A.S., Schmidt, K., Neumann, K., Hill, U., Ciotta, G., Gupta, A., Torres, D.C., Fu, J., Kranz, A., Stewart, A.F., and Anastassiadis, K. (2014). The H3K4 methyltransferase Setd1a is first required at the epiblast stage, whereas Setd1b becomes essential after gastrulation. Development *141*, 1022–1035. https://doi.org/10.1242/dev.098152.

Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. *34*, 525–527. https://doi.org/10.1038/nbt.3519.

Cahan, P., Li, H., Morris, S.A., Lummertz da Rocha, E., Daley, G.Q., and Collins, J.J. (2014). CellNet: network biology applied to stem cell engineering. Cell *158*, 903–915. https://doi.org/10.1016/j.cell.2014.07.020.

Chong, Z.-S., Khong, Z.J., Tay, S.H., and Ng, S.-Y. (2022). Metabolic contributions to neuronal deficits caused by genomic disruption of schizophrenia risk gene SETD1A. Schizophrenia *8*, 115. https://doi.org/10.1038/s41537-022-00326-9.

Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014). Biological insights from 108 schizophrenia-associated genetic loci. Nature *511*, 421–427. https://doi.org/10.1038/nature13595.

Cuella-Martin, R., Hayward, S.B., Fan, X., Chen, X., Huang, J.W., Taglialatela, A., Leuzzi, G., Zhao, J., Rabadan, R., Lu, C., et al. (2021). Functional interrogation of DNA damage response variants with base editing screens. Cell *184*, 1081–1097.e19. https://doi.org/10.1016/j.cell.2021.01.041.

De Los Angeles, A., Fernando, M.B., Hall, N.A.L., Brennand, K.J., Harrison, P.J., Maher, B.J., Weinberger, D.R., and Tunbridge, E.M. (2021). Induced Pluripotent Stem Cells in Psychiatry: An Overview and Critical Perspective. Biol. Psychiatr. *90*, 362–372. https://doi.org/10.1016/j.biopsych.2021.04.008.

Deinhardt, K., Kim, T., Spellman, D.S., Mains, R.E., Eipper, B.A., Neubert, T.A., Chao, M.V., and Hempstead, B.L. (2011). Neuronal growth cone retraction relies on proneurotrophin receptor signaling through Rac. Sci. Signal. *4*, ra82. https://doi.org/10.1126/scisignal.2002060.

Duan, J. (2023). Human stem cell modeling of neuropsychiatric disorders: from polygenicity to convergence. Med. Rev. *3*, 347–350. https://doi.org/10.1515/mr-2023-0016.

Falkovich, R., Danielson, E.W., Perez de Arce, K., Wamhoff, E.C., Strother, J., Lapteva, A.P., Sheng, M., Cottrell, J.R., and Bathe, M. (2023). A synaptic molecular dependency network in knockdown of autism- and schizophrenia-associated genes revealed by multiplexed imaging. Cell Rep. *42*, 112430. https://doi.org/10.1016/j.celrep.2023.112430.

Grove, J., Ripke, S., Als, T.D., Mattheisen, M., Walters, R.K., Won, H., Pallesen, J., Agerbo, E., Andreassen, O.A., Anney, R., et al. (2019). Identification of common genetic risk variants for autism spectrum disorder. Nat. Genet. *51*, 431–444. https://doi.org/10.1038/s41588-019-0344-8.

Halikere, A., Popova, D., Scarnati, M.S., Hamod, A., Swerdel, M.R., Moore, J.C., Tischfield, J.A., Hart, R.P., and Pang, Z.P. (2020). Addiction associated N40D mu-opioid receptor variant modulates synaptic function in human neurons. Mol. Psychiatr. *25*, 1406–1419. https://doi.org/10.1038/s41380-019-0507-0.

Hanna, R.E., Hegde, M., Fagre, C.R., DeWeirdt, P.C., Sangree, A.K., Szegletes, Z., Griffith, A., Feeley, M.N., Sanson, K.R., Baidi, Y., et al. (2021). Massively parallel assessment of human variants with base editor screens. Cell *184*, 1064–1080.e20. https://doi.org/10.1016/j.cell.2021.01.012.

Holtzman, L., and Gersbach, C.A. (2018). Editing the Epigenome: Reshaping the Genomic Landscape. Annu. Rev. Genom. Hum. Genet. *19*, 43–71. https://doi.org/10.1146/annurev-genom-083117-021632.

Howard, D.M., Adams, M.J., Clarke, T.K., Hafferty, J.D., Gibson, J., Shirali, M., Coleman, J.R.I., Hagenaars, S.P., Ward, J., Wigmore, E.M., et al. (2019). Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. Nat. Neurosci. *22*, 343–352. https://doi.org/10.1038/s41593-018-0326-7.

Huang, L., Lou, C.H., Chan, W., Shum, E.Y., Shao, A., Stone, E., Karam, R., Song, H.W., and Wilkinson, M.F. (2011). RNA homeostasis governed by cell type-specific and branched feedback loops acting on NMD. Mol. Cell *43*, 950–961. https://doi.org/10.1016/j.molcel.2011.06.031.

Kluesner, M.G., Nedveck, D.A., Lahr, W.S., Garbe, J.R., Abrahante, J.E., Webber, B.R., and Moriarity, B.S. (2018). EditR: A Method to Quantify Base Editing from Sanger Sequencing. CRISPR J. *1*, 239–250. https://doi.org/10.1089/crispr.2018.0014.

McGowan, H., Mirabella, V.R., Hamod, A., Karakhanyan, A., Mlynaryk, N., Moore, J.C., Tischfield, J.A., Hart, R.P., and Pang, Z.P. (2018). hsa-let-7c miRNA Regulates Synaptic and Neuronal Function in Human Neurons. Front. Synaptic Neurosci. *10*, 19. https://doi.org/10.3389/fnsyn.2018.00019.

Meng, X., Navoly, G., Giannakopoulou, O., Levey, D.F., Koller, D., Pathak, G.A., Koen, N., Lin, K., Adams, M.J., Rentería, M.E., et al. (2024). Multi-ancestry genome-wide association study of major depression aids locus discovery, fine mapping, gene prioritization and causal inference. Nat. Genet. *56*, 222–233. https://doi.org/10.1038/s41588-023-01596-4.

Michael Deans, P.J., and Brennand, K.J. (2021). Applying stem cells and CRISPR engineering to uncover the etiology of schizophrenia. Curr. Opin. Neurobiol. *69*, 193–201. https://doi.org/10.1016/j.conb.2021.04.003.

Muhtaseb, A.W., and Duan, J. (2022). Modeling common and rare genetic risk factors of neuropsychiatric disorders in human induced pluripotent stem cells. Schizophr. Res. https://doi.org/10.1016/j.schres.2022.04.003.

Mukai, J., Cannavò, E., Crabtree, G.W., Sun, Z., Diamantopoulou, A., Thakur, P., Chang, C.-Y., Cai, Y., Lomvardas, S., Takata, A., et al. (2019). Recapitulation and Reversal of Schizophrenia-Related Phenotypes in Setd1a-Deficient Mice. Neuron *104*, 471–487.e12. https://doi.org/10.1016/j.neuron.2019.09.014.

Mullins, N., Forstner, A.J., O'Connell, K.S., Coombes, B., Coleman, J.R.I., Qiao, Z., Als, T.D., Bigdeli, T.B., Børte, S., Bryois, J., et al. (2021). Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. Nat. Genet. *53*, 817–829. https://doi.org/10.1038/s41588-021-00857-4.

Nagahama, K., Sakoori, K., Watanabe, T., Kishi, Y., Kawaji, K., Koebis, M., Nakao, K., Gotoh, Y., Aiba, A., Uesaka, N., and Kano, M. (2020). Setd1a Insufficiency in Mice Attenuates Excitatory Synaptic Function and Recapitulates Schizophrenia-Related Behavioral Abnormalities. Cell Rep. *32*, 108126. https://doi.org/10.1016/j.celrep.2020.108126.

Nunez, J.K., Chen, J., Pommier, G.C., Cogan, J.Z., Replogle, J.M., Adriaens, C., Ramadoss, G.N., Shi, Q., Hung, K.L., Samelson, A.J., et al. (2021). Genome-wide programmable transcriptional memory by CRISPR-based epigenome editing. Cell *184*, 2503–2519.e17. https://doi.org/10.1016/j.cell.2021.03.025.

Palmer, D.S., Howrigan, D.P., Chapman, S.B., Adolfsson, R., Bass, N., Blackwood, D., Boks, M.P.M., Chen, C.Y., Churchhouse, C., Corvin, A.P., et al. (2022). Exome sequencing in bipolar disorder identifies AKAP11 as a risk gene shared with schizophrenia. Nat. Genet. *54*, 541–547. https://doi.org/10.1038/s41588-022-01034-x.

Pantazis, C.B., Yang, A., Lara, E., McDonough, J.A., Blauwendraat, C., Peng, L., Oguro, H., Kanaujiya, J., Zou, J., Sebesta, D., et al. (2022). A reference human induced pluripotent stem cell line for large-scale collaborative studies. Cell Stem Cell *29*, 1685–1702.e22. https://doi.org/10.1016/j.stem.2022.11.004.

Popp, M.W., and Maquat, L.E. (2016). Leveraging Rules of Nonsense-Mediated mRNA Decay for Genome Engineering and Personalized Medicine. Cell *165*, 1319–1322. https://doi.org/10.1016/j.cell.2016.05.053.

International Schizophrenia Consortium, Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., Sullivan, P.F., and Sklar, P. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature *460*, 748–752. https://doi.org/10.1038/nature08185.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. *8*, 2281–2308. https://doi.org/10.1038/nprot.2013.143.

Rees, H.A., and Liu, D.R. (2018). Base editing: precision chemistry on the genome and transcriptome of living cells. Nat. Rev. Genet. *19*, 770–788. https://doi.org/10.1038/s41576-018-0059-1.

Ripke, S., O'Dushlaine, C., Chambert, K., Moran, J.L., Kähler, A.K., Akterin, S., Bergen, S.E., Collins, A.L., Crowley, J.J., Fromer, M., et al. (2013). Genome-wide association analysis identifies 13 new risk loci for schizophrenia. Nat. Genet. *45*, 1150–1159. https://doi.org/10.1038/ng.2742.

Ripke, S., Sanders, A.R., Kendler, K.S., Levinson, D.F., Sklar, P., Holmans, P.A., Lin, D.Y., Duan, J., Ophoff, R.A., Andreassen, O.A., et al. (2011). Genome-wide association study identifies five new schizophrenia loci. Nat. Genet. *43*, 969–976. https://doi.org/10.1038/ng.940.

Sato, H., and Singer, R.H. (2021). Cellular variability of nonsense-mediated mRNA decay. Nat. Commun. *12*, 7203. https://doi.org/10.1038/s41467-021-27423-0.

Satterstrom, F.K., Kosmicki, J.A., Wang, J., Breen, M.S., 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. https://doi.org/10.1016/j.cell.2019.12.036.

Seipel, K., Medley, Q.G., Kedersha, N.L., Zhang, X.A., O'Brien, S.P., Serra-Pages, C., Hemler, M.E., and Streuli, M. (1999). Trio amino-terminal guanine nucleotide exchange factor domain expression promotes actin cytoskeleton reorganization, cell migration and anchorage-independent cell growth. J. Cell Sci. *112*, 1825–1834. https://doi.org/10.1242/jcs.112.12.1825.

Shi, J., Levinson, D.F., Duan, J., Sanders, A.R., Zheng, Y., Pe'er, I., Dudbridge, F., Holmans, P.A., Whittemore, A.S., Mowry, B.J., et al. (2009). Common variants on chromosome 6p22.1 are associated with schizophrenia. Nature *460*, 753–757. https://doi.org/10.1038/nature08192.

Singh, T., Poterba, T., Curtis, D., Akil, H., Al Eissa, M., Barchas, J.D., Bass, N., Bigdeli, T.B., Breen, G., Bromet, E.J., et al. (2022). Rare coding variants in ten genes confer substantial risk for schizophrenia. Nature *604*, 509–516. https://doi.org/10.1038/s41586-022-04556-w.

Stahl, E.A., Breen, G., Forstner, A.J., McQuillin, A., Ripke, S., Trubetskoy, V., Mattheisen, M., Wang, Y., Coleman, J.R.I., Gaspar, H.A., et al. (2019). Genome-wide association study identifies 30 loci associated with bipolar disorder. Nat. Genet. *51*, 793–803. https://doi.org/10.1038/s41588-019-0397-8.

Standage-Beier, K., Tekel, S.J., Brookhouser, N., Schwarz, G., Nguyen, T., Wang, X., and Brafman, D.A. (2019). A transient reporter for editing enrichment (TREE) in human cells. Nucleic Acids Res. *47*, e120. https://doi.org/10.1093/nar/gkz713.

Stefansson, H., Ophoff, R.A., Steinberg, S., Andreassen, O.A., Cichon, S., Rujescu, D., Werge, T., Pietiläinen, O.P.H., Mors, O., Mortensen, P.B., et al. (2009). Common variants conferring risk of schizophrenia. Nature *460*, 744–747. https://doi.org/10.1038/nature08186.

Sürün, D., Schneider, A., Mircetic, J., Neumann, K., Lansing, F., Paszkowski-Rogacz, M., Hänchen, V., Lee-Kirsch, M.A., and Buchholz, F. (2020). Efficient Generation and Correction of Mutations in Human iPS Cells Utilizing mRNAs of CRISPR Base Editors and Prime Editors. Genes *11*, 511. https://doi.org/10.3390/genes11050511.

Tekel, S.J., Brookhouser, N., Standage-Beier, K., Wang, X., and Brafman, D.A. (2021). Cytosine and adenosine base editing in human pluripotent stem cells using transient reporters for editing enrichment. Nat. Protoc. *16*, 3596–3624. https://doi.org/10.1038/s41596-021-00552-y.

Tristan, C.A., Hong, H., Jethmalani, Y., Chen, Y., Weber, C., Chu, P.H., Ryu, S., Jovanovic, V.M., Hur, I., Voss, T.C., et al. (2023). Efficient and safe single-cell cloning of human pluripotent stem cells using the CEPT cocktail. Nat. Protoc. *18*, 58–80. https://doi.org/10.1038/s41596-022-00753-z.

Trubetskoy, V., Pardiñas, A.F., Qi, T., Panagiotaropoulou, G., Awasthi, S., Bigdeli, T.B., Bryois, J., Chen, C.-Y., Dennison, C.A., Hall, L.S., et al. (2022). Mapping genomic loci implicates genes and synaptic biology in schizophrenia. Nature *604*, 502–508. https://doi.org/10.1038/s41586-022-04434-5.

Wang, M., Zhang, L., and Gage, F.H. (2020). Modeling neuropsychiatric disorders using human induced pluripotent stem cells. Protein Cell *11*, 45–59. https://doi.org/10.1007/s13238-019-0638-8.

Wang, S., Rhijn, J.-R.v., Akkouh, I., Kogo, N., Maas, N., Bleeck, A., Ortiz, I.S., Lewerissa, E., Wu, K.M., Schoenmaker, C., et al. (2022). Loss-of-function variants in the schizophrenia risk gene SETD1A alter neuronal network activity in human neurons through the cAMP/PKA pathway. Cell Rep. *39*, 110790. https://doi.org/10.1016/j.celrep.2022.110790.

Wang, Y., Penfold, S., Tang, X., Hattori, N., Riley, P., Harper, J.W., Cross, J.C., and Tyers, M. (1999). Deletion of the Cul1 gene in mice causes arrest in early embryogenesis and accumulation of cyclin E. Curr. Biol. *9*, 1191–1194. https://doi.org/10.1016/S0960-9822(00)80024-X.

Weissbein, U., Schachter, M., Egli, D., and Benvenisty, N. (2016). Analysis of chromosomal aberrations and recombination by allelic bias in RNA-Seq. Nat. Commun. *7*, 12144. https://doi.org/10.1038/ncomms12144.

Wells, M.F., Nemesh, J., Ghosh, S., Mitchell, J.M., Salick, M.R., Mello, C.J., Meyer, D., Pietilainen, O., Piccioni, F., Guss, E.J., et al. (2023). Natural variation in gene expression and viral susceptibility revealed by neural progenitor cell villages. Cell Stem Cell *30*, 312–332.e13. https://doi.org/10.1016/j.stem.2023.01.010.

Wen, Z., Christian, K.M., Song, H., and Ming, G.L. (2016). Modeling psychiatric disorders with patient-derived iPSCs. Curr. Opin. Neurobiol. *36*, 118–127. https://doi.org/10.1016/j.conb.2015.11.003.

West, S., Zhang, H., Zhang, S., Kozlova, A., Sanders, A., Pang, Z., Gejman, P., and Duan, J. (2019). 46 modelling the schizophrenia-associated loss-of-function mutation of SETD1A in human stem cell-derived BRAIN organoids. Eur. Neuropsychopharmacol *29*, S84–S85. https://doi.org/10.1016/j.euroneuro.2019.07.187.

Wong, J.J.L., Li, S., Lim, E.K.H., Wang, Y., Wang, C., Zhang, H., Kirilly, D., Wu, C., Liou, Y.C., Wang, H., and Yu, F. (2013). A Cullin1-based SCF E3 ubiquitin ligase targets the InR/PI3K/TOR pathway to regulate neuronal pruning. PLoS Biol. *11*, e1001657. https://doi.org/10.1371/journal.pbio.1001657.

Wray, N.R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E.M., Abdellaoui, A., Adams, M.J., Agerbo, E., Air, T.M., Andlauer, T.M.F., et al. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat. Genet. *50*, 668–681. https://doi.org/10.1038/s41588-018-0090-3.

Xu, P., Liu, Z., Liu, Y., Ma, H., Xu, Y., Bao, Y., Zhu, S., Cao, Z., Wu, Z., Zhou, Z., and Wei, W. (2021). Genome-wide interrogation of gene functions through base editor screens empowered by barcoded sgRNAs. Nat. Biotechnol. *39*, 1403–1413. https://doi.org/10.1038/s41587-021-00944-1.

Yang, N., Chanda, S., Marro, S., Ng, Y.H., Janas, J.A., Haag, D., Ang, C.E., Tang, Y., Flores, Q., Mall, M., et al. (2017). Generation of pure GABAergic neurons by transcription factor programming. Nat. Methods *14*, 621–628. https://doi.org/10.1038/nmeth.4291.

Yi, F., Danko, T., Botelho, S.C., Patzke, C., Pak, C., Wernig, M., and Südhof, T.C. (2016). Autism-associated SHANK3 haploinsufficiency causes Ih channelopathy in human neurons. Science *352*, aaf2669. https://doi.org/10.1126/science.aaf2669.

Zhang, S., Zhang, H., Forrest, M.P., Zhou, Y., Sun, X., Bagchi, V.A., Kozlova, A., Santos, M.D., Piguel, N.H., Dionisio, L.E., et al. (2023). Multiple genes in a single GWAS risk locus synergistically mediate aberrant synaptic development and function in human neurons. Cell Genom. *3*, 100399. https://doi.org/10.1016/j.xgen.2023.100399.

Zhang, S., Zhang, H., Zhou, Y., Qiao, M., Zhao, S., Kozlova, A., Shi, J., Sanders, A.R., Wang, G., Luo, K., et al. (2020). Allele-specific open chromatin in human iPSC neurons elucidates functional disease variants. Science *369*, 561–565. https://doi.org/10.1126/science.aay3983.

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. https://doi.org/10.1016/j.neuron.2013.05.029.

Stem Cell Reports, Volume *19*

Supplemental Information

Scaled and efficient derivation of loss-of-function alleles in risk genes for neurodevelopmental and psychiatric disorders in human iPSCs

Hanwen Zhang, Ada McCarroll, Lilia Peyton, Sol Díaz de León-Guerrerro, Siwei Zhang, Prarthana Gowda, David Sirkin, Mahmoud ElAchwah, Alexandra Duhe, Whitney G. Wood, Brandon Jamison, Gregory Tracy, Rebecca Pollak, Ronald P. Hart, Carlos N. Pato, Jennifer G. Mulle, Alan R. Sanders, Zhiping P. Pang, and Jubao Duan

Supplementary Information Table of Contents

Supplementary Figures

Supplementary Tables

Supplemental Experimental procedures

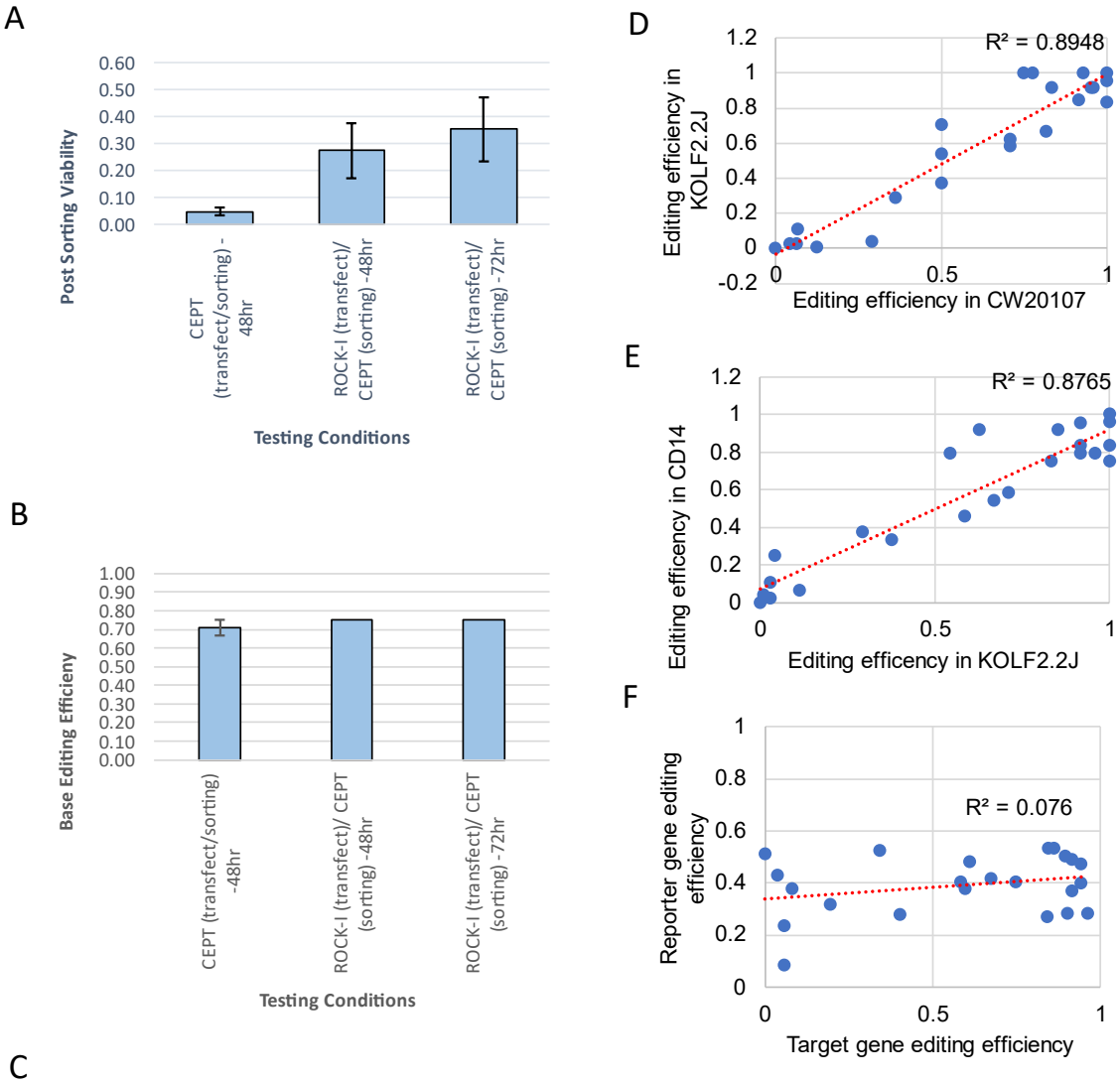
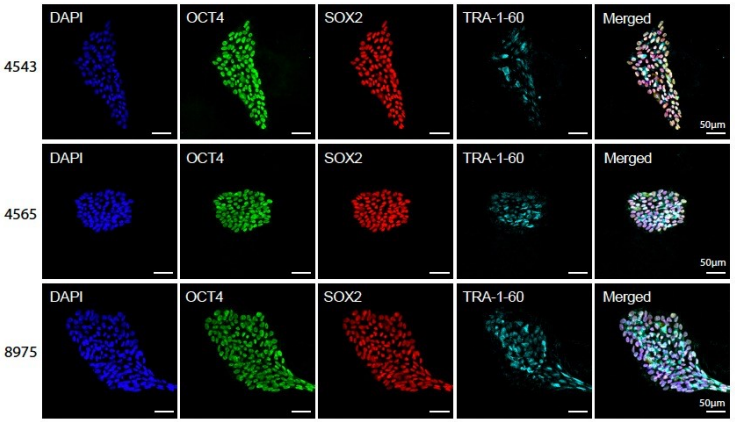
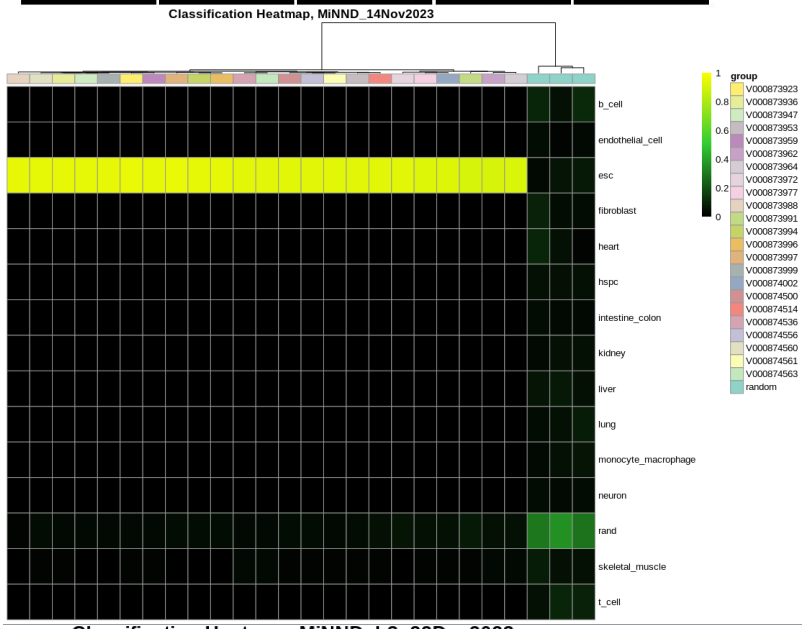

A
Post Sorting Viability
0.60
0.50
0.40
0.30
0.20
0.10
0.00
CEPT (transfect/sorting) - 48hr
ROCK-I (transfect)/ CEPT (sorting) -48hr
ROCK-I (transfect)/ CEPT (sorting) -72hr
Testing Conditions
B
Base Editing Efficieny
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
CEPT (transfect/sorting) -48hr
ROCK-I (transfect)/ CEPT (sorting) -48hr
ROCK-I (transfect)/ CEPT (sorting) -72hr
Testing Conditions
D
R² = 0.8948
Editing efficiency in KOLF2.2J
Editing efficiency in CW20107
E
R² = 0.8765
Editing efficency in CD14
Editing efficency in KOLF2.2J
F
R² = 0.076
Reporter gene editing efficiency
Target gene editing efficiency
C
Reporter Editing Efficiency
GRIA3
ASH1L
HCN4
KDM6B
SMARCC2
KMT2C
GRIN2A
TRIO
CUL1
RB1CC1
SCN2A
SP4
XPO7
ANKRD11
CACNA1G
DLL1
SETD1A
ARID1B
CHD8
AKAP11
HERC1
GABRA1
SHANK3
CW20107
KOLF2.2J
CD14

Figure S1. Reporter gene editing efficiency and optimization for increasing post-sorting single hiPSC clonal survivability (related to Figures 2 and 3). (A) Post-sorting single hiPSC clonal survivability (A) and C to T base editing efficiency (B) (assayed by Sanger sequencing of the LoF mutation site) under different transfection/cell sorting conditions. (C) Reporter gene editing efficiency for all the target genes in three hiPSC lines (CW20107, KOLF2.2J, and CD14). (D) and (E) Strong correlations of target gene editing efficiency across cell lines. (F) Weak correlation between target gene editing efficiency and reporter gene editing efficiency. Target gene editing efficiency was calculated by the genotypes of each single hiPSC clone as confirmed by Sanger sequencing. Reporter gene editing efficiency in (F) was calculated as the proportion of GFP+ cells (reporter gene edited cells) vs. BFP+ cells (cells transfected with sgRNAs) in FACS.

A


4543
4565
8975
DAPI
OCT4
SOX2
TRA-1-60
Merged
50μm

B


Classification Heatmap, MiNND_14Nov2023
b_cell
endothelial_cell
esc
fibroblast
heart
hspc
intestine_colon
kidney
liver
lung
monocyte_macrophage
neuron
rand
skeletal_muscle
t_cell
group
random

C

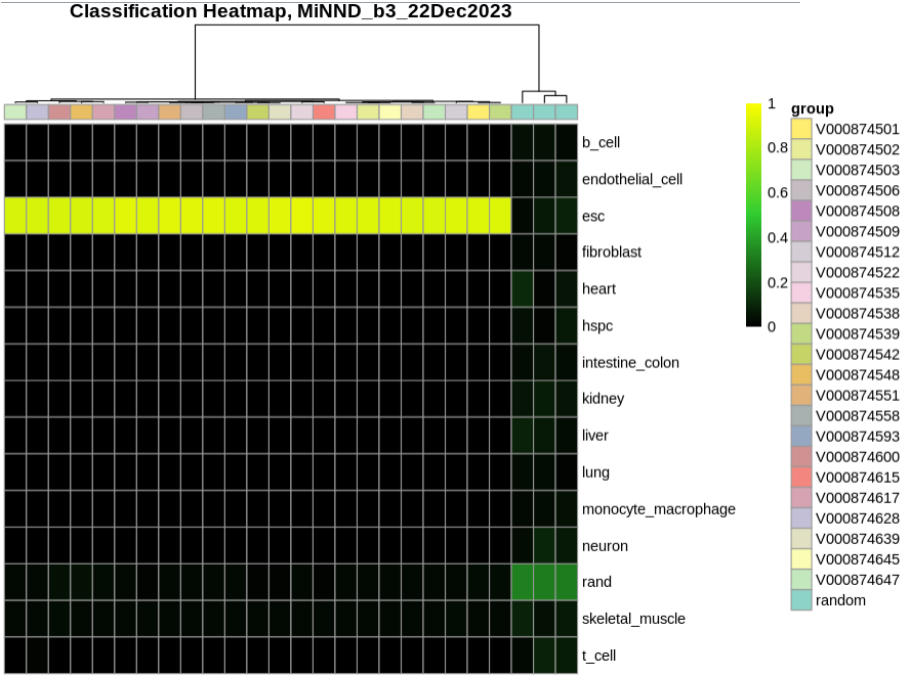

Classification Heatmap, MiNND_b3_22Dec2023
b_cell
endothelial_cell
esc
fibroblast
heart
hspc
intestine_colon
kidney
liver
lung
monocyte_macrophage
neuron
rand
skeletal_muscle
t_cell
group
random

Figure S2. Stem cell pluripotency characterization of iSTOP LoF hiPSC lines (related to Figure 4). (A) The iSTOP mutant lines were stained positive for pluripotent stem cell markers (OCT4, SOX2, TRA-1-60). Scale bar: 50μm. (B) CellNet analysis of RNA-seq data of hiPSC lines confirmed their pluripotency. Pluripotency scores showed the transcriptional similarity of the edited iSTOP LoF hiPSC lines to ESC or other non-ESC cell types. Two batches of edited hiPSC lines that were not listed in the main Figure 4 are shown.

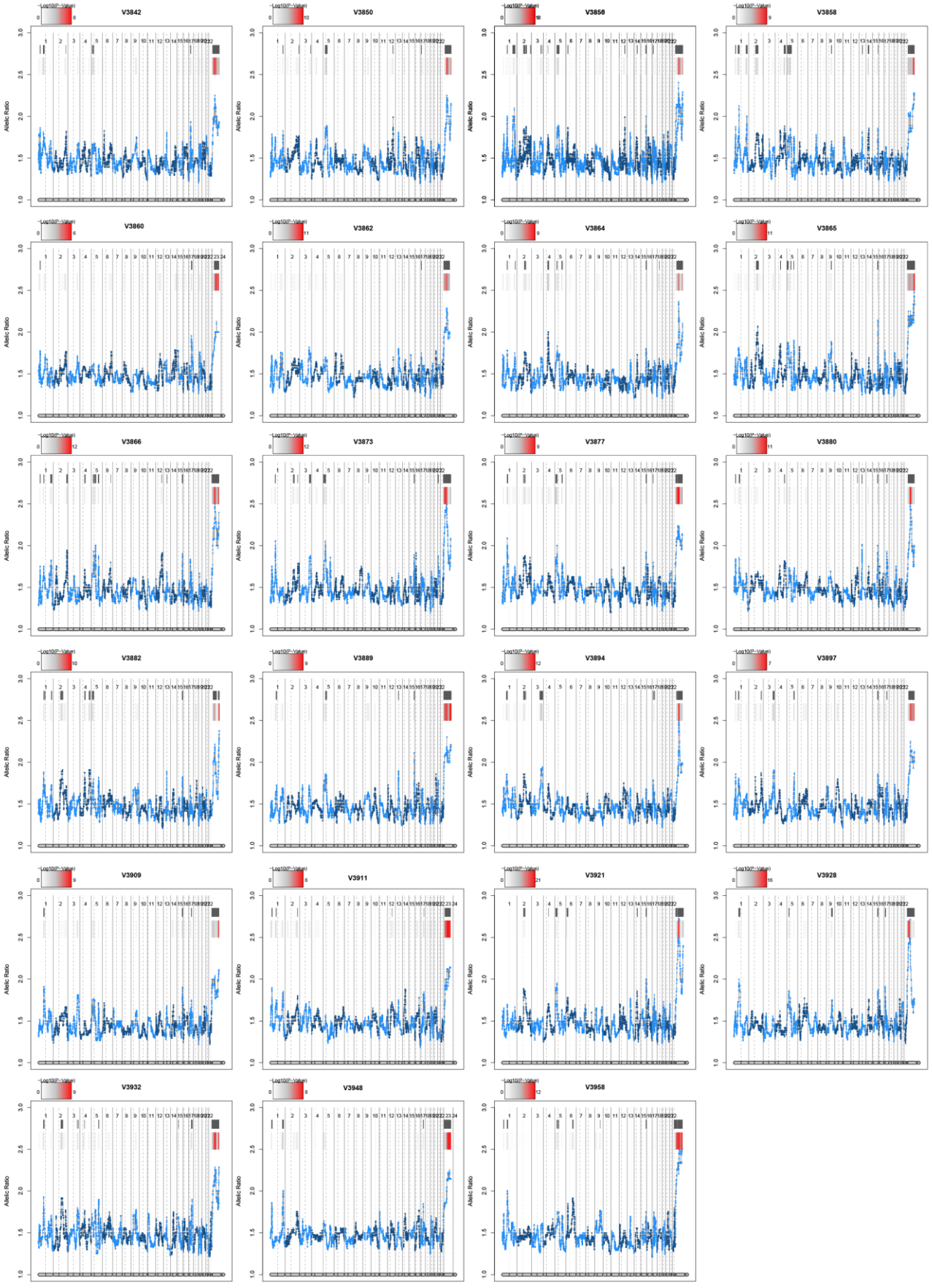
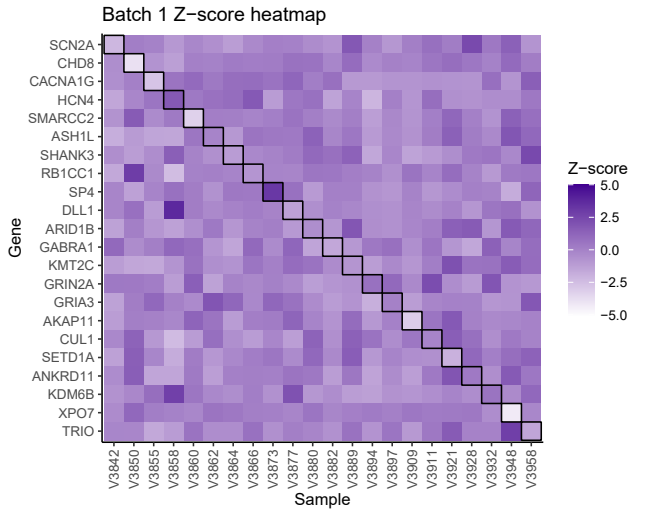
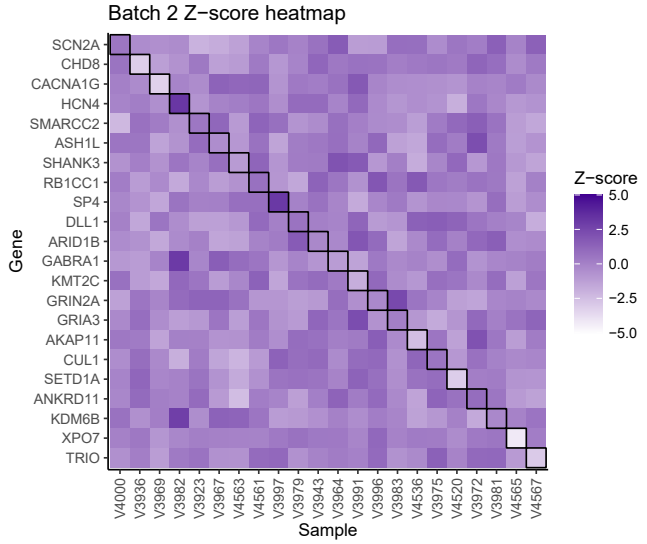

V3842
V3850
V3856
V3858
V3860
V3862
V3864
V3865
V3866
V3873
V3877
V3880
V3882
V3889
V3894
V3897
V3909
V3911
V3921
V3928
V3932
V3948
V3958
Allelic Ratio
Log10(P-Value)

Figure S3. RNA-seq-based eSNP-Karyotyping (related to Figure 4). Each panel showed the moving average of SNP intensity (RNA-seq reads) of the two alleles of heterozygous SNPs. Only the allelic ratio graphs of batch 1 edited lines are shown here; the chromosomal heterozygosity graphs and the results for all three batches are in https://zenodo.org/records/11591445. Note that only an autosomal peak interval marked with a red bar above the peak is considered to be significantly "abnormal", and no iPSC lines showed such abnormality in this batch.


Batch 1 Z−score heatmap
Gene
SCN2A
CHD8
CACNA1G
HCN4
SMARCC2
ASH1L
SHANK3
RB1CC1
SP4
DLL1
ARID1B
GABRA1
KMT2C
GRIN2A
GRIA3
AKAP11
CUL1
SETD1A
ANKRD11
KDM6B
XPO7
TRIO
V3842
V3850
V3855
V3858
V3860
V3862
V3864
V3866
V3873
V3877
V3880
V3882
V3889
V3894
V3897
V3909
V3911
V3921
V3928
V3932
V3948
V3958
Sample
Z−score
5.0
2.5
0.0
−2.5
−5.0


Batch 2 Z−score heatmap
Gene
SCN2A
CHD8
CACNA1G
HCN4
SMARCC2
ASH1L
SHANK3
RB1CC1
SP4
DLL1
ARID1B
GABRA1
KMT2C
GRIN2A
GRIA3
AKAP11
CUL1
SETD1A
ANKRD11
KDM6B
XPO7
TRIO
V4000
V3936
V3969
V3982
V3923
V3967
V4563
V4561
V3997
V3979
V3943
V3964
V3991
V3996
V3983
V4536
V3975
V4520
V3972
V3981
V4565
V4567
Sample
Z−score
5.0
2.5
0.0
−2.5
−5.0

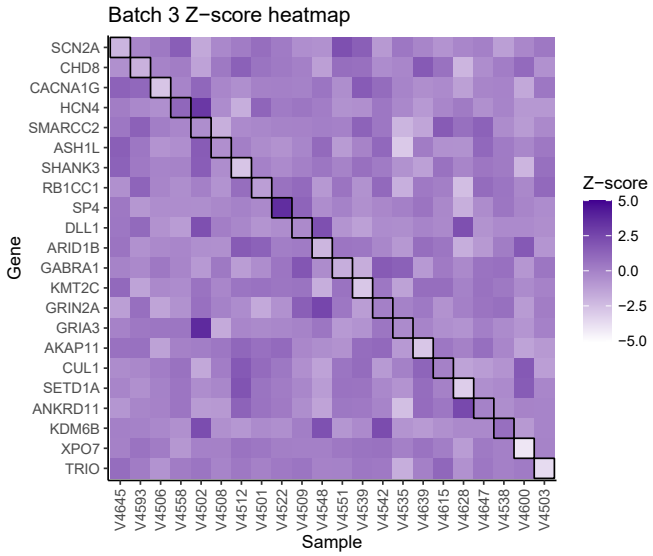
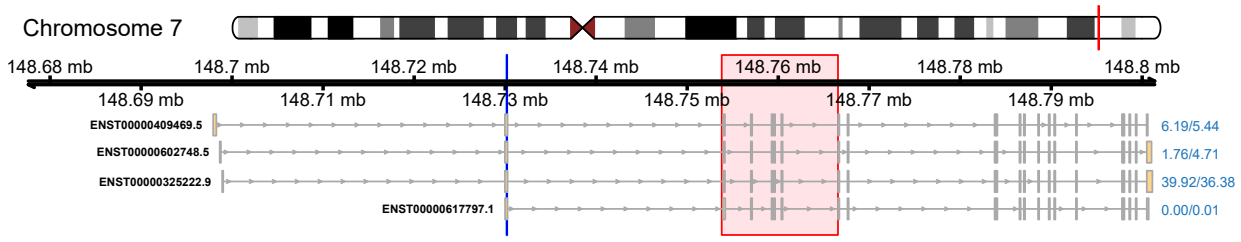
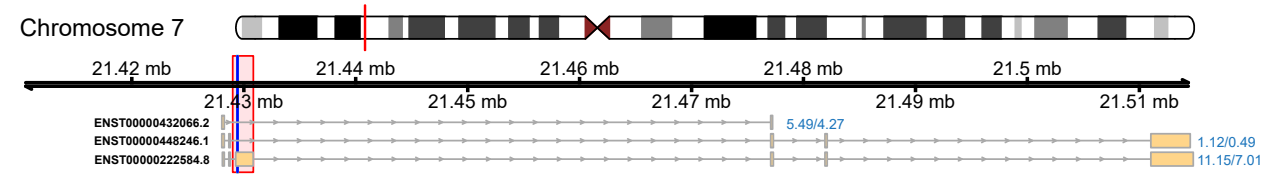

Batch 3 Z−score heatmap
Gene
SCN2A
CHD8
CACNA1G
HCN4
SMARCC2
ASH1L
SHANK3
RB1CC1
SP4
DLL1
ARID1B
GABRA1
KMT2C
GRIN2A
GRIA3
AKAP11
CUL1
SETD1A
ANKRD11
KDM6B
XPO7
TRIO
V4645
V4593
V4506
V4558
V4502
V4508
V4512
V4501
V4522
V4509
V4548
V4551
V4539
V4542
V4535
V4639
V4615
V4628
V4647
V4538
V4600
V4503
Sample
Z−score
5.0
2.5
0.0
−2.5
−5.0

Figure S4. RNA-seq-based estimation of the expected NMD for each iSTOP LoF allele derived from all three donor hiPSC lines (related to Figure 5). Shown are heat maps of Z-scored expression values (counts per million reads, CPM) of each gene in each iSTOP LoF hiPSC line for all three batches. Lined boxes indicate the edited hiPSC line that is expected to show NMD for a specific target gene.

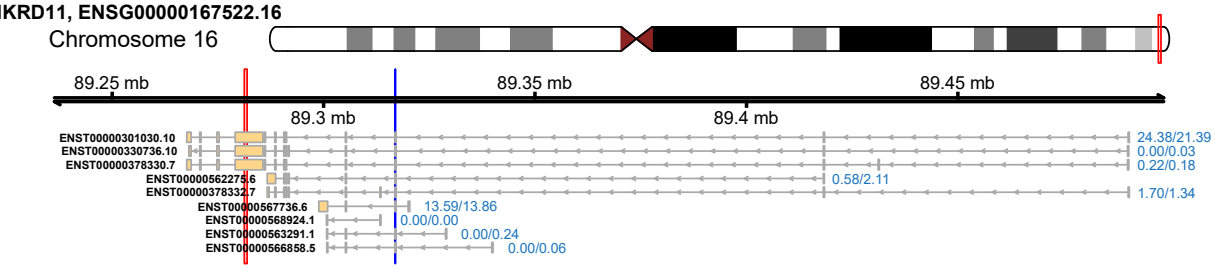
CUL1, ENSG00000055130.17



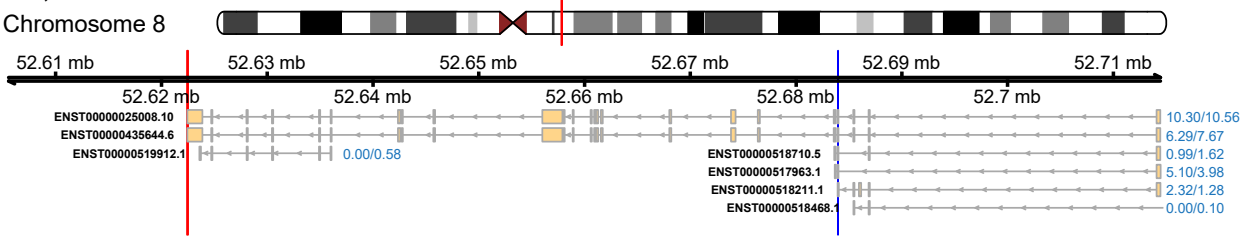
SP4, ENSG00000105866.15



ANKRD11, ENSG00000167522.16



RB1CC1, ENSG00000023287.13



ARID1B, ENSG00000049618.24

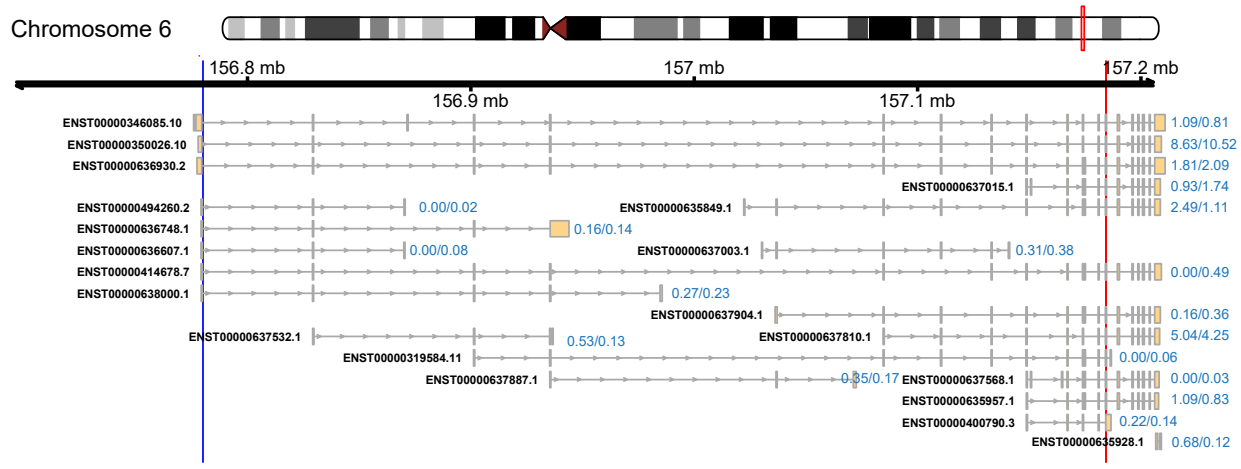


Figure S5. Transcript isoforms and the relative genomics positions of iSTOP sgRNA and antibody for each gene assayed by Western blot assay (related to Figure 5). Red boxes indicate the position of the antibody epitope or part of the protein used for generating the antibody (vendor's information), and the blue line indicates the position of sgRNA used for iSTOP mutagenesis. Only isoforms with non-zero expression (transcript per million reads, TPM) in Table S5 were included. The expression level of each transcript in the iSTOP LoF line for a specific transcript and the average expression of the transcript across all 22 iPSC lines are shown on the right side of each transcript and separated by "/".

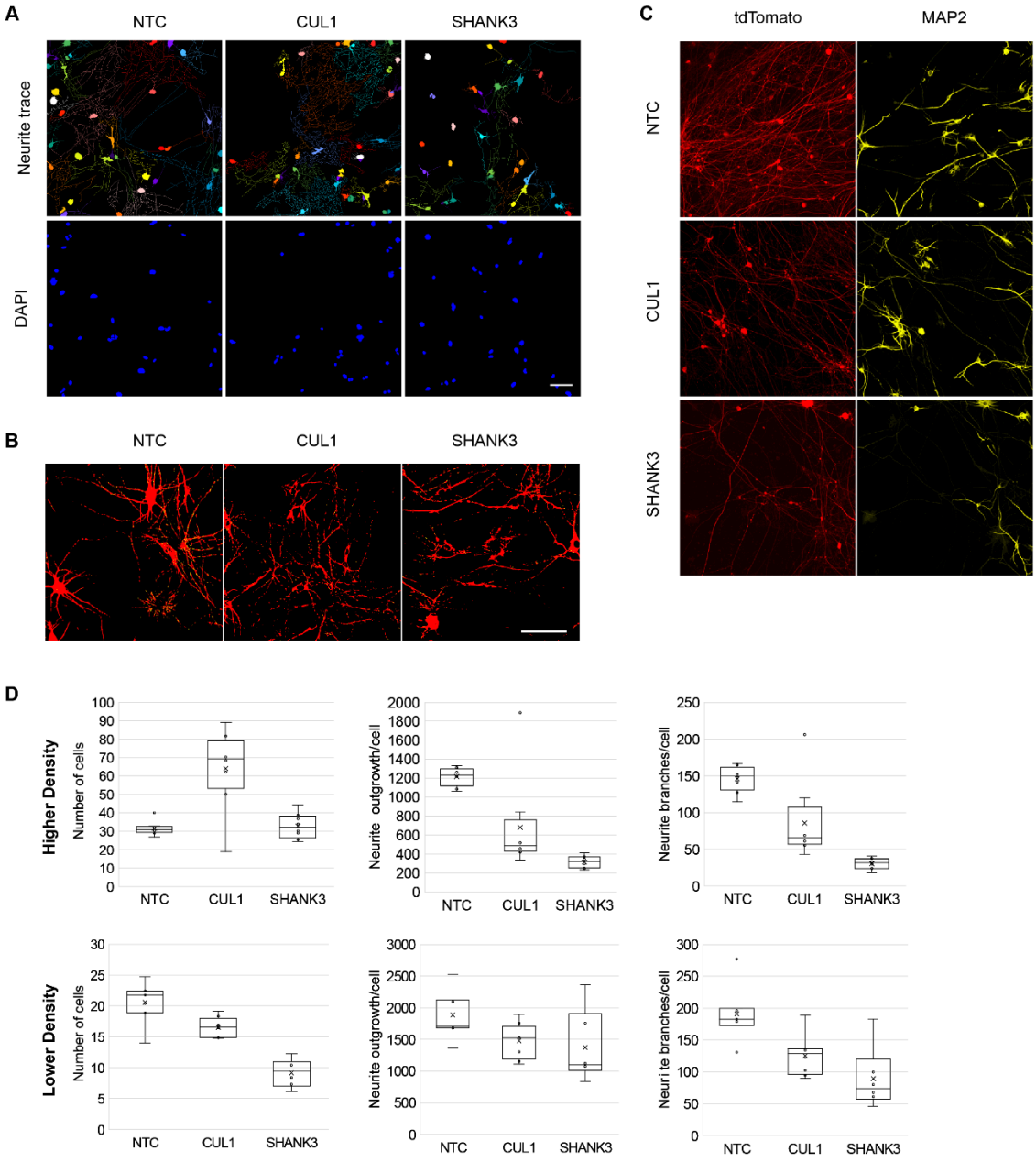

A
NTC
CUL1
SHANK3
Neurite trace
DAPI
B
NTC
CUL1
SHANK3
C
tdTomato
MAP2
NTC
CUL1
SHANK3
D
Higher Density
Number of cells
Neurite outgrowth/cell
Neurite branches/cell
Lower Density
Number of cells
Neurite outgrowth/cell
Neuri te branches/cell
NTC
CUL1
SHANK3

Figure S6. Cell segmentation and binary masks in morphometric analysis using the high content imaging and neurite analysis in low-density neuron/glia co-culture (related to Figure 6). (A) Cell segmentation for assaying neurite outgrowth and branches of LoF alleles of SHANK3 (-/-) and CUL1 (+/-) using the built-in neurite outgrowth module on the ImageXpress system. Single neurons were highlighted in rainbow color. (B) Binary masks for assaying puncta density of LoF alleles of SHANK3 (-/-) and CUL1 (+/-) using a customized synaptic module. MAP2+ and tdTomato+ mask was highlighted in red, and SYN1+ mask was highlighted in yellow. (C) Neurite traces of low-density cell culture in an independent experiment with a second parental line (KOLF2.2J) for validating the observed neurite deficits in (A) and (B). (D) The number of neurons per field of view (FOV) and the assayed neurite outgrowth/branches/cells in higher-density cultures (upper panels) or lower-density cultures (lower panels). Scale bar in all panels: 20 μm.

Supplementary Tables

Table S1. Sequences and other genomics information for iSTOP sgRNAs. Related to the main text.

iSTOP sgRNA sequences and qPCR assays for testing CBE editing efficiency												
APOE-iSTOP1-sgRNA	GGAAcAACTGACCCCGGTGG											
APOE-iSTOP2-sgRNA	GGTGcAGACACTGTCTGAGC											
iSTOP sgRNAs for Batch1-MiNND genes												
Gene name	**chr of gene**	**coordinate (hg38 bp position)**	**Strand**	**Targeted Codon**	**Number Of Isoforms**	**Relative Position in the Largest**	**NMD Prediction (%)**	**PAM: NGG**	**Off-target Prediction PAM: NGG**	**Spacing PAM: NGG**	**9 of the top 10 prioritized genes**	
ARID1B	chr6	156778374	+	CAG	2	0.066	100	GCCTcAGCCCGGCCCCGACA	0	good	ARID1B	
CACNA1G	chr17	50561655	+	CAG	32	0.027	100	TGAGCcAGGACAGCCGCCCG	0	optimal	CACNA1G	
CHD8	chr14	21429305	-	CAG	3	0.113	100	GTCCTCcAGCAGCCACAGTC	0	good	CHD8	
DLL1	chr6 KI270798v1_alt	17326	-	CAG	2	0.024	50	GTGTcAGGTAGGCGGGCAGG	0	good	DLL1	
GABRA1	chr5	161882647	+	CAG	4	0.473	100	CTAAACcAGTATGACCTTCT	0	good	GABRA1	
KMT2C	chr7	152330728	-	CAA	2	0.018	100	CAAAGAAcAATCTGCAGAAG	0	ok	KMT2C	
SCN2A	chr2	165295938	+	CAG	4	0.019	100	CAAAcAGGAACGCAAGGATG	1	good	SCN2A	
SHANK3	chr22	50679171	+	TGG	2	0.145	100	CTCCTGcCAGCCATTCTCGT	0	good	SHANK3	
SMARCC2	chr12	56189395	-	CAG	6	0.018	100	TGACCcAGTTCGACAACGTG	0	optimal	SMARCC2	
AKAP11	chr13	42299467	+	CAG	1	0.126	100	CTCTGGAcAGCAGAAGTCAT	0	ok		
ASH1L	chr1	155481573	-	CAG	3	0.146	66.67	CCCCACTcAGGAACCGCTTA	0	ok		
CUL1	chr7	148730185	+	TGG	4	0.027	100	GGTCGTCcCAGATCTGGTCC	0	ok		
GRIA3	chrX	123184560	+	CAA	6	0.009	100	GGGGcAAAGCGTGCTCCGGG	0	good		
GRIN2A	chr16	10180359	-	TGG	3	0.012	100	CCGCGCcAGACCAGAAGGGC	0	good		
HERC1	chr15	63775588	-	TGG	1	0.002	100	TCAAGcCATTTCAGCTTCAC	0	optimal		
KDM6B	chr17	7845676	+	TGG	2	0.024	100	GCAGCcATGCGCTACGAGGA	0	optimal		
RB1CC1	chr8	52683964	-	CAA	3	0.025	66.67	TATTcAACACCAGGTGCTGG	0	good		
SETD1A	chr16	30963486	+	CAG	1	0.111	100	CCCCTcAGACTGTGCCCACT	0	optimal		
SP4	chr7	21429421	+	CAA	2	0.111	50	TcAACCACAACAGCTAGAAC	0	good		
XPO7	chr8	21966923	+	CAG	2	0.027	100	ACTCcAGGCAGAGAAAGCCT	0	good		
TRIO	chr5	14280387	+	CGA	5	0.032	60	GAATAcGACAGGAGGATCTC	0	optimal		
HCN4	chr15	73368150	-	CAA	1	0.033	100	GCCGCcAAGACCCCAGCCGC	0	optimal		
ANKRD11	chr16	89316947	-	CAG	7	0.009	85.71	GGAGAAGcAGACTGGGAAAA	0	ok		

Table S2. Cell lines for generating LoF alleles (EA=European ancestry).

Cell Line	Source	Clinical Diagnosis	Age	Sex	Ethnicity
CW20107	CIRM	control	21	female	Caucasian
KOLF2.2J	Jackson Lab	control	55-59	male	EA
CD14	MGS	control	33	female	EA

Table S3. Genotypes at the iSTOP site for each LoF hiPSC line (NA=not available; no editing)

Cell line ID	Cell line ID_Alias	Batch	Source iPSC lines	Gene	Chr	Position (bp; hg38)	iSTOP genotype
g1b1c1	V3909	1	CW20107	AKAP11	chr13	42299467	C/T
g2b1c1	V3862	1	CW20107	ASH1L	chr1	155481573	A/A
g3b1c1	V3855	1	CW20107	CACNA1G	chr17	50561655	T/T
g4b1c1	V3911	1	CW20107	CUL1	chr7	148730185	G/A
g5b1c1	V3897	1	CW20107	GRIA3	chrX	123184560	T/T
g6b1c1	V3894	1	CW20107	GRIN2A	chr16	10180359	T/T
NA	NA	1	CW20107	HERC1	chr15	63775588	NA
g8b1c1	V3932	1	CW20107	KDMB6	chr17	7845676	A/A
g9b1c1	V3866	1	CW20107	RB1CC1	chr8	52683964	A/A
g10b1c1	V3921	1	CW20107	SETD1A	chr16	30963486	C/T
g11b1c1	V3873	1	CW20107	SP4	chr7	21429421	T/T
g12b1c1	V3948	1	CW20107	XPO7	chr8	21966923	T/T
g13b1c1	V3958	1	CW20107	TRIO	chr5	14280387	C/T
g14b1c1	V3858	1	CW20107	HCN4	chr15	73368150	A/A
g15b1c1	V3928	1	CW20107	ANKRD11	chr16	89316947	G/A
g16b1c1	V3864	1	CW20107	SHANK3	chr22	50679170	A/A
g17b1c1	V3850	1	CW20107	CHD8	chr14	21429305	A/A
g18b1c1	V3889	1	CW20107	KMT2C	chr7	152330728	G/A
g19b1c1	V3860	1	CW20107	SMARCC2	chr12	56189395	A/A
g20b1c1	V3842	1	CW20107	SCN2A	chr2	165295938	T/T
g21b1c1	V3877	1	CW20107	DLL1	chr6	170290088	A/A
g22b1c1	V3880	1	CW20107	ARID1B	chr6	156778374	T/T
g23b1c1	V3882	1	CW20107	GABRA1	chr5	161882647	C/T
g1b2c2	V4536	2	KOLF2.2	AKAP11	chr13	42299467	C/T
g2b2c2	V3967	2	KOLF2.2	ASH1L	chr1	155481573	A/A
g3b2c2	V3969	2	KOLF2.2	CACNA1G	chr17	50561655	T/T
g4b2c2	V3975	2	KOLF2.2	CUL1	chr7	148730185	G/A
g5b2c2	V3983	2	KOLF2.2	GRIA3	chrX	123184560	T/T
g6b2c2	V3996	2	KOLF2.2	GRIN2A	chr16	10180359	T/T
NA	NA	2	KOLF2.2	HERC1	chr15	63775588	NA
g8b2c2	V3981	2	KOLF2.2	KDMB6	chr17	7845676	A/A
g9b2c2	V4561	2	KOLF2.2	RB1CC1	chr8	52683964	A/A
g10b2c2	V4520	2	KOLF2.2	SETD1A	chr16	30963486	C/T
g11b2c2	V3997	2	KOLF2.2	SP4	chr7	21429421	T/T
g12b2c2	V4565	2	KOLF2.2	XPO7	chr8	21966923	T/T
g13b2c2	V4567	2	KOLF2.2	TRIO	chr5	14280387	C/T
g14b2c2	V3982	2	KOLF2.2	HCN4	chr15	73368150	A/A
g15b2c2	V3972	2	KOLF2.2	ANKRD11	chr16	89316947	G/A
g16b2c2	V4563	2	KOLF2.2	SHANK3	chr22	50679170	G/A

g17b2c2	V3936	2	KOLF2.2	CHD8	chr14	21429305	A/A
g18b2c2	V3991	2	KOLF2.2	KMT2C	chr7	152330728	G/A
g19b2c2	V3923	2	KOLF2.2	SMARCC2	chr12	56189395	A/A
g20b2c2	V4000	2	KOLF2.2	SCN2A	chr2	165295938	T/T
g21b2c2	V3979	2	KOLF2.2	DLL1	chr6	170290088	A/A
g22b2c2	V3943	2	KOLF2.2	ARID1B	chr6	156778374	T/T
g23b2c2	V3964	2	KOLF2.2	GABRA1	chr5	161882647	T/T
g1b3c3	V4639	3	CD14	AKAP11	chr13	42299467	C/T
g2b3c3	V4508	3	CD14	ASH1L	chr1	155481573	A/A
g3b3c3	V4506	3	CD14	CACNA1G	chr17	50561655	T/T
g4b3c3	V4615	3	CD14	CUL1	chr7	148730185	G/A
g5b3c3	V4535	3	CD14	GRIA3	chrX	123184560	T/T
g6b3c3	V4542	3	CD14	GRIN2A	chr16	10180359	T/T
NA	NA	3	CD14	HERC1	chr15	63775588	NA
g8b3c3	V4538	3	CD14	KDMB6	chr17	7845676	A/A
g9b3c3	V4501	3	CD14	RB1CC1	chr8	52683964	A/A
g10b3c3	V4628	3	CD14	SETD1A	chr16	30963486	C/T
g11b3c3	V4522	3	CD14	SP4	chr7	21429421	T/T
g12b3c3	V4600	3	CD14	XPO7	chr8	21966923	T/T
g13b3c3	V4503	3	CD14	TRIO	chr5	14280387	C/T
g14b3c3	V4558	3	CD14	HCN4	chr15	73368150	A/A
g15b3c3	V4647	3	CD14	ANKRD11	chr16	89316947	G/A
g16b3c3	V4512	3	CD14	SHANK3	chr22	50679170	A/A
g17b3c3	V4593	3	CD14	CHD8	chr14	21429305	A/A
g18b3c3	V4539	3	CD14	KMT2C	chr7	152330728	A/A
g19b3c3	V4502	3	CD14	SMARCC2	chr12	56189395	A/A
g20b3c3	V4645	3	CD14	SCN2A	chr2	165295938	T/T
g21b3c3	V4509	3	CD14	DLL1	chr6	170290088	A/A
g22b3c3	V4548	3	CD14	ARID1B	chr6	156778374	T/T
g23b3c3	V4551	3	CD14	GABRA1	chr5	161882647	C/T

Table S4. Chromosomal abnormality detected by eSNP-karyotyping in iPSC line V4167. Related to the main text, Fig. S4

chr	SNP position	-Log-*p*
chr12	57759165	3.85054963
chr12	57796735	3.85054963
chr12	59325835	3.92460732
chr12	59326071	3.89474465
chr12	59326091	4.03153324
chr12	62466937	4.20311241
chr12	62644140	4.16558805
chr12	62646492	4.24742666
chr12	62647908	4.3456919
chr12	64194285	4.45671364
chr12	64449074	4.45671364
chr12	64449075	4.36892906
chr12	64449084	4.40883795
chr12	64449140	4.48379216
chr12	64449350	4.50717128
chr12	64450281	4.48379216
chr12	64450354	4.4834706
chr12	64450369	4.50717128
chr12	64695205	4.52307202
chr12	64695505	4.52307202
chr12	64696039	4.51382747
chr12	64697060	4.52307202
chr12	64697251	4.52307202
chr12	64713715	4.64571987
chr12	64714129	4.64571987
chr12	64716472	4.64571987
chr12	65824905	4.64571987
chr12	65964288	4.64571987
chr12	65965972	4.62044753
chr12	66137398	4.52307202
chr12	66234536	4.50717128
chr12	67313337	4.51382747
chr12	67658398	4.52307202
chr12	67661419	4.52307202
chr12	67662032	4.52307202
chr12	68764959	4.48379216
chr12	68765256	4.40883795
chr12	68765355	4.44099113
chr12	68808384	4.50717128
chr12	68840106	4.50717128

chr12	68841207	4.45671364
chr12	68841626	4.40883795
chr12	68843093	4.40883795
chr12	68843224	4.45671364
chr12	68843230	4.23215593
chr12	68843233	4.38638609
chr12	68843263	4.40883795
chr12	68843698	4.40883795
chr12	68843754	4.24168138
chr12	69272155	4.24168138
chr12	69272869	4.24246392
chr12	69273496	4.24168138
chr12	69601562	3.9660467
chr12	69601565	3.86867891
chr12	70431058	3.85054963
chr12	71610121	3.81615486
chr12	71787256	3.89479184
chr12	74540073	4.06395829
chr12	75496341	4.21492583
chr12	75496553	4.2737851
chr12	75496872	4.24742666
chr12	75498671	4.23294743
chr12	75498993	4.24742666
chr12	75499601	4.40883795
chr12	75499619	4.40883795
chr12	75499706	4.48379216
chr12	76026505	4.51382747
chr12	76027761	4.51382747
chr12	76027977	4.48379216
chr12	76029027	4.48379216
chr12	76029219	4.48379216
chr12	76045879	4.48379216
chr12	76046117	4.48379216
chr12	76046488	4.48379216
chr12	76047371	4.40883795
chr12	76047374	4.2016076
chr12	76047427	3.89474465
chr12	76048101	3.85054963

Table S5. Transcript isoform expression level for all the targeted genes in LoF iPSC lines (as xlsx file). Related to the main text, Fig. 5.

Table S6. Oligos, primers, and antibodies were used for sgRNA cloning, sequencing confirmation, qPCR, and Western blot assays (as xlsx file). Related to the main text, Fig. 5, Fig. 6.

Supplemental Experimental procedures

Source hiPSC lines

The use of CW20107 (from California Institute for Regenerative Medicine - CIRM) and KOLF2.2J (from The Jackson Laboratory; an updated version of KOLF2.1J and not officially released but made available for advanced access to SSPsyGene) was part of the SSPsyGene consortium agreement on the common cell lines. KOLF2.2J, an updated version of a well-characterized KOLF2.1J, was chosen for its improved features and potential impact on the study. (Pantazis et al., 2022), with a targeted correction of a known splice-site disruption in *COL3A1* in KOLF2.1J by the MorPhic consortium (https://morphic.bio/). The original KOLF2.1J is currently listed in the catalogue of the Jackson Laboratory, but KOLF2.2J will be available soon and can already be obtained via special request. The other hiPSC line (CD14) was specific to the MiNND project and was from the Duan lab (Shi et al., 2009; Zhang et al., 2023; Zhang et al., 2020). CD14 was originally derived from the lymphocytes of a healthy donor of the Molecular Genetics of Schizophrenia (MGS) cohort (Shi *et al.*, 2009). Detailed cell line information is described in Table S2. The hiPSCs were maintained in mTeSRPlus (StemCell #100-0276) with Primocin (Invitrogen #ant-pm-1) on tissue culture plates coated with Matrigel (Fisher Scientific #08-774-552) or Geltrex (Fisher Scientific #A1413202). The study was approved by the Institutional Review Board (IRB) of NorthShore University HealthSystem, ensuring the ethical conduct of the research. More detailed information on the donor cell lines can be found in Table S2.

Chemicals and reagents

The chemicals, media, reagents used in cell culture, PCR, Sanger sequencing, and other main experiments include: BbsI-HF (NEB: R3539S), NEBuilder® HiFi DNA Assembly reagent (NEB: E2621S), mTeSR Plus (StemCell: 100-0276), Primocin (InvivoGen: ant-pm-1), Matrigel (FisherScientific: 08-774-552), Geltrex (Fisher Scientific: A1413202), DPBS (no calcium, no magnesium) (Fisher Scientific: 14-190-144), ReLeSR (StemCell: 100-0483), ROCK-Inhibitor (Tocris: 1254), mFreSR (Stem Cell: 05855), Accutase (StemCell: 07920), Trypan Blue Stain (FisherScientific: T10282), Gibco Opti-MEM I Reduced Serum Medium (Fisher Scientific: 31-985-062), Lipofectamine Stem Transfection Reagent (Fisher Scientific: STEM00003), Chroman 1 (R&D Systems: 7163/10), Emricasan (R&D Systems: 7310/5), transISRIB (R&D Systems: 5284/10), Polyamine Supplement (R&D Systems: 7739/1), QuickExtract DNA Extraction Solution 1.0 (Fisher Scientific:

NC9904870), PCRx Enhancer System (Fisher Scientific: 11-495-017), Deoxynucleoside Triphosphate Set (Sigma Aldrich: 3622614001), Shrimp Alkaline Phosphatase and buffer (Fisher Scientific: 783905000UN), E.coli exonuclease I (Fisher Scientific: 70073X5000UN), BigDye Terminator v3.1 Cycle Sequencing Kit (Fisher Scientific: 4337455), HiDi Formamide (Fisher Scientific: 4311320), RNeasy Plus Mini Kit (Qiagen: 74134), QIAshredder (Qiagen: 79654), DMSO (Sigma Aldrich: D2650-100ML), Fetal Bovine Serum (Fisher Scientific: A3160501), Mineral Oil (Sigma Aldrich: M5310), 16% Formaldehyde (Fisher Scientific: PI28908), DAPI (Fisher Scientific: EN62248). The antibodies used in immunofluorescence staining include: Synapsin I antibody (Synaptic Systems: 106-011), goat anti-tdTomato antibody (Fisher Scientific: 50-167-1115), anti-MAP2 antibody (Sigma Aldrich: AB5543), Donkey Anti-Mouse 488 (Fisher Scientific: A21202), Donkey Anti-Goat IgG (H+L) Cross Adsorbed Secondary Antibody, Alexa Fluor 568 (Fisher Scientific: A11057), Donkey Anti-Chicken IgY (H+L) Highly Cross Adsorbed Secondary Antibody, Alexa Fluor 647 (Fisher Scientfic: A78952). The main lab supplies include: 96-well non skirted PCR plate (DotScientific: 650-PCR), Corning™ Internally Threaded Cryogenic Vials 2 ML (FisherScientific: 03-374-21), 96-well plates (Corning: 353072), 24-well plates (ThermoFisher: 142475), 6-well plates (StemCell: 38016).

iSTOP gRNA design and cloning

We first retrieved the best pre-computed iSTOP-gRNA for each selected NPD gene using the iSTOP web-based tool (https://www.ciccialab-database.com/istop/#/) (Billon et al., 2017), requiring >50% NMD rate and in >50% transcript isoforms. NMD prediction was determined based on whether the targeted base was 55 nucleotides upstream of the final exon-exon junction (Billon *et al.*, 2017; Popp and Maquat, 2016). Whenever possible, the iSTOP-gRNA location was placed to the first half of the gene to ensure the resultant protein truncation (likely to be LoF) even without causing NMD. Finally, to minimize any possible off-target editing, all the selected iSTOP-gRNAs are free of any predicted off-target site via aligning to the human genome, allowing up to two mismatches in the first eight bases of the guide sequence (Billon *et al.*, 2017; Popp and Maquat, 2016). A total of 23 genes were designed for iSTOP-gRNAs in the current study (See Table S1). For cloning the designed sgRNAs, pDT-sgRNA (Addgene # 138271) vector was selected as gRNA carrier. The vector was digested with BbsI-HF. After gel purification, a single strand oligo with a prefix, gRNA of interest and a postfix was introduced into the vector backbone through Gibson assembly using NEBuilder® HiFi DNA

Assembly reagent. After cloning, miniprep plasmids were sequenced using M13 Rev primer. M13 Rev: 5' – caggaaacagctatgac – 3'. Example of a single strand oligo: 5' – atatcttgtggaaaggacgaaacaccgXXXXXXXXXXXXXXXXXXXXgttttagagctagaaatagcaagtta – 3'. After genotyping, correct clones were expanded, and transfection-grade plasmid was prepared using an endo-free plasmid kit (QIAGEN: 12362).

HEK293 culture, transfection, and editing evaluation

HEK293T cells were purchased from ATCC (Cat# CRL-3216) and maintained in DMEM with 10% FBS following the vendor's instructions. For transfection and editing efficiency evaluation, 90% confluent 293T culture was dissociated with Accutase at 37°C for 5 min. About 2×10^5 cells were replated into one well on 12 w tissue culture plates. 24 hr post replating, 2 μg of selected CBE plasmid, 1 μg of pEF-BFP (Addgene# 138272) plasmid and 1 μg of pDT-sgRNA plasmid carrying selected gRNA were transfected using Fugene HD (Promega # E2311) reagent with 1:3 DNA:Reagent ratio following vendor's instructions. 48hr post-transfection, BFP+/GFP+ and BFP+/GFP- cells were sorted through BD Aria Fusion Flow Cytometer and replated. 120 hr post-transfection, replated cells were collected after Accutase dissociation and 30-75 μl of QuickExtract DNA Extraction Solution (FisherSci # NC9904870) was added to the cell pellet for DNA extraction on a thermocycler. Extracted DNA was subsequently amplified for Sanger sequencing genotyping to evaluate editing efficiency at loci of interest. To evaluate the base editing efficiency by Sanger sequencing, we utilized an open-source tool EditR (https://moriaritylab.shinyapps.io/editr_v10/) (Kluesner et al., 2018) that takes Sanger sequencing .ab1 file and gRNA sequence as input. The percentage of each base under the sequencing peak (without sequencing noise reduction) will be calculated.

hiPSC culture and transfection

hiPSCs were maintained in mTeSRPlus (StemCell #100-0276) with Primocin (Invivogen #ant-pm-1) on tissue culture plates coated with Matrigel (Fisher Scientific #08-774-552) or Geltrex (Fisher Scientific #A1413202) throughout the mutagenesis process. The medium was changed every other day, and colonies were passaged every 4-6 days when cells reached 70% confluence. For DNA transfection, cells were plated at a density of $1.2\sim1.5\times10^5$ per well on a 24-well plate (ThermoFisher # 142475) in mTeSR Plus with 5μM ROCK-Inhibitor (Tocris #1254). The next day, antibiotics-free mTeSR Plus with 5μM ROCK-Inhibitor was changed on

the plate after ensuring appropriate cell density (60-70% confluence) and survival. Shortly after, each well was transfected with 750 ng pEF-AncBE4max (Addgene # 138270), 300 ng pEF-BFP, and 300 ng pDT-sgRNA containing the variant-specific gRNA using LipofectamineSTEM with 1:2.5 DNA:reagent ratio. In total, 23 different pDT-sgRNAs were used (one per well), and the 24th well was used as a negative control. Media was refreshed with regular mTeSR Plus containing Primocin at 24 hr and 48 hr post-transfection. At 72 hr post-transfection, cells were prepped for single-cell sorting.

Single hiPSC sorting and clonal culture

Single hiPSCs from the post-transfection culture above were sorted into 96 well plates with one cell per well using a BD FACSAria Fusion Flow Cytometer. All sorting procedures were done using mTeSR Plus with CEPT cocktail (1:10,000 chroman 1, emricasan, and transISRIB; 1:1,000 polyamine supplement) (Tristan et al., 2023). To prepare for FACS, cells were dissociated into single cells using Accutase (StemCell: 07920) for 7 min at 37°C. Cells were transferred to 15 ml tubes with 1 ml mTeSR plus to inactivate the Accutase and centrifuged at 300 × g for 3 min. The resulting cell pellets were resuspended in 700 μl media and filtered twice using 5 ml corning round bottom tubes with the blue strainer cap (Fisher Scientific: 0877123). Samples were placed on ice immediately to minimize clogging. Samples were processed and analyzed using the BD FACSAria Fusion Flow Cytometer, gating for BFP+/GFP+ cells, one single cell per well on a 96-well plate for each condition; for NTC, BFP+/GFP- cells were sorted. After sorting, the cells were not disturbed for 24 hr. 48 hr post sorting, 50 μl mTeSR plus was added to each well. 72 hr post sorting, 50 μl media refreshment for each well. 96 hr post posting, 120 μl media refreshment for each well. 144 hr post sorting, aspirated 120 μl and added 100 μl media for each well. Afterwards, 100 μl media refreshment every other day (10-14 days) until colonies appeared with an appropriate size to pick for Sanger sequencing genotyping. All media refreshments were performed using Integra's MINI 96 electronic pipette.

PCR and Sanger sequencing for LoF genotype confirmation

Once colonies reached an appropriate size and had stem cell-like morphology, 8-12 colonies were picked for each edited condition. DNA was extracted from the picked colonies using Quick Extract DNA Extraction Solution (Fisher Scientific NC9904870). Following PCR to amplify the DNA for each LoF gene, Sanger sequencing was completed to confirm that the appropriate base was changed at the desired location to create

a stop codon. The sequencing was performed on a 3730xl DNA Analyzer and the sequencing data were imported to SeqScape v2.5 for automatic analysis and genotype calling. Up to 4 colonies with confirmed homozygous editing or heterozygous editing (if there were no homozygous colonies) and good morphology were expanded for RNA isolation and cell cryopreservation. The oligos and primers for sequencing and qPCR were listed in Table S6.

RNA isolation for RNA sequencing (RNA-seq)

Based on sequencing results, two selected clones were expanded from one well on 96-well plates to one well on 6-well plates. Once reaching 70% confluency, cells were expanded a second time from one well to two wells per clone, one for cryopreservation and one for RNA extraction/RNA-seq. For RNA isolation, cells were lysed using 800 μl of Buffer RLT Plus (QIAGEN 1053393). Cell lysates in buffer RLT were stored at -80°C until ready to be isolated using the QIAGEN RNeasy Plus Mini Kit (QIAGEN 74134) following the vendor's instructions. Purified RNAs were sent to Novogene for RNA-seq.

RNA-seq data processing

Bulk RNA-seq was performed by external vendor Novogene and in 2×150 bp paired-end format of 25-30 M reads per sample. Briefly, raw FASTQ files were aligned to the human GRCh38.p14 genome by STAR 2.7.7 and subsequently counted by the built-in function of STAR at the gene level using the GTF file of GENCODE v42 with parameters `--quantMode Genecounts --alignSoftClipAtRefenceEnds No --outFilterScoreMinOverLread 0.30 --outFilterMatchNminOverLread 0.30`. Gene counts from each of the samples were collected by a customized script and collated into a single count matrix. Genes that had 0 counts in all samples were removed prior to analysis.

CellNet analysis for pluripotency

The RNA-seq data of each edited iSTOP hiPSC line were used for pluripotency evaluation by using the R package `CellNet` (Cahan et al., 2014). Briefly, the gene × sample count matrix generated in the previous step was loaded by `EdgeR` and a new count matrix contained log-transformed, library size-normalized CPM (counts-per-million) value was generated by `calcNormFactors()`, `estimateDisp()` and `cpm()`. Subsequently, the script constructed a random forest classifier using the in-built model from the `CellNet`

Package. Finally, the likelihood of each sample-cell type pair (in scores) was evaluated by passing the log-transformed gene × sample count matrix through the classifier and plotting the results in hierarchically clustered heatmaps.

Using RNA-seq data for e-SNP Karyotyping

As part of the high throughput LoF mutagenesis pipeline for evaluating possible hiPSC chromosomal abnormality due to editing or hiPSC clonal growth, we opted to use e-SNP Karyotyping rather than the classical G-band karyotyping. e-SNP Karyotyping detects any potential chromosomal aberrations, including duplications, loss of heterozygosity, and meiotic recombination. As we previously described (Zhang *et al.*, 2023; Zhang *et al.*, 2020), we used the e-Karyotyping R package developed by the Benven lab (github.io/BenvenLab/eSNPKaryotyping) (Weissbein et al., 2016) with customization to our current environment settings. Briefly, 2x150 bp of paired-end raw FASTQ files of all samples were firstly trimmed for adaptors and low-quality reads by Trimmomatic (Bolger et al., 2014). Only paired-end reads were kept post Trimmomatic processing. Trimmed read pairs were aligned to the human hg38 genome by Bowtie2 v2.5.1 to generate BAM files for eSNP Karyotyping package (Weissbein *et al.*, 2016). The original package code in R has been rewritten for GATK 4 and R 4.2 platforms. Only common SNPs (MAF > 0.05) from dbSNP 154 were retained for genome and zygosity block analysis. A rolling window of 151 bp in size was used to smooth the Allele Frequency (AF) curves when plotting the AF ratio and P values by genome coordinates. Each block represented 1.0 MB in zygosity block graphs.

Immunofluorescence staining for hiPSCs

hiPSCs were dissociated with Accutase (innovative cell technologies AT-104) and seeded into Matrigel (Corning 354234) coated round glass coverslips in a 24-well plate and kept in mTESR+ media (Stem cell technology 100-0275). Cells were kept until they formed medium-sized colonies. Cells were washed twice with 1× PBS and fixed with 4% PFA for 30 min. Samples were incubated with blocking buffer 4% BSA (A3803 Sigma), 1% Goat serum (ThermoFisher 16210072), 0.2% triton X-100 (BP151 Fisher BioReagents) in PBS for 1hr at room temperature (RT). Primary antibodies were incubated for 1 hr at RT. Samples were washed 3 times with PBS 0.2% Triton X-100, and secondary antibodies were incubated for 1 hr at RT, making sure samples were protected from light. Samples were washed 3 times with PBS 0.02% Triton X-100 and rinsed

with MiliQ water before mounting with Fluoroshield with DAPI (F6057 Sigma) and placed on a glass slide. The following antibodies were used: rabbit anti-Sox2 (Millipore AB5603), mouse IgG anti-Oct4 (Millipore MAB4401), mouse IgM anti-Tra-1-60 (Millipore MAB4360), goat anti-rabbit Alexa Fluor 546 (Invitrogen A11035), goat anti-mouse IgG Alexa Fluor 488 (Invitrogen A11001), goat anti-mouse IgM Alexa Fluor 647 (Invitrogen A21238). Confocal images were taken using a Zeiss LSM700 laser-scanning confocal microscope with a 20× objective.

Western blot

hiPSCs were grown on 6 well plates as stated above. When the cells reached ~80% confluency, they were washed twice with DPBS and lysed with 100µl RIPA buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS) supplemented with 0.5mM DTT, 1mM PMSF and 1× Protease inhibitor Cocktail (Sigma P8340) using a cell scraper. The lysate was transferred to a tube incubated on ice for 10 min and centrifuged at 14000 × g for 10 min at 4°C. The supernatant was transferred to a new tube and protein concentration was determined using a BCA protein assay (Thermo Scientific 23225) measuring absorbance at 562 nm in a plate reader (SpectraMax i3, Molecular Devices). For the SDS-PAGE, protein samples were prepared with 15µg of protein with 2× Laemmli sample buffer (161-0737 BIORAD) and heated for 5 min at 95°C. Samples were resolved using 7.5% or 10% acrylamide gels (4561023DC, 4561033DC BIO-RAD) and transferred into a 0.45 µm nitrocellulose membrane (1620115 BIO-RAD). Membranes were blocked for 1hr at RT using 5% non-fat milk dissolved in TBS-T (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1% Tween 20). Primary antibodies were diluted in 5% non-fat milk (732-291-1940 LabScientific) or 5% BSA (A3803 Sigma-Adrich) in TBS-T and incubated overnight at 4°C. Membranes were washed 3 times with TBS-T and incubated with an HRP-conjugated secondary antibody diluted in 5% non-fat milk or BSA in TBS-T for 1 hr at RT. Membranes were washed 3 times with TBS-T and protein was visualized using the Clarity Western ECL Substrate (Bio-rad 1705060) and autoradiography films (XAR ALF 1318, LabScientific). Densitometry analysis was performed using ImageJ software, and β-actin or VCP were used as loading control. The following antibodies (Table S6) were used: ANKRD11 (1:1000 Santa Cruz sc-514916), BAF250b/ARID1B (1:3000 Cell Signaling 92964S), β-Actin (1:20000 Sigma-Aldrich A544), CUL1 (1:2000 Santa Cruz SC-17775),

FIP200/RB1CC1 (1:2000 Cell Signaling 12436S, SP4 (1:2000 Santa Cruz, SC-390124), VCP (1:20000, K331).

Lentivirus generation

Lentiviral vectors were generated by transfecting HEK293T cells with lentivirus packaging plasmids (pMDLg/pRRE, VsVG and pRSV-REV) with the desired vectors as previously described (Pang et al., 2011) using lipofectamine 3000. The following plasmids were used: pMDLg/pRRE (Addgene 12251), pRSV-Rev (Addgene #12253), pCMV-VSV-G (Addgene #8454), FUW-M2rtTA (Addgene #20342), FUW-TetO-Ngn2-P2A-puromycin (Addgene #52047), FUW-TetO-Ascl1-T2A-puromycin (Addgene #97329), FUW-TetO-Dlx2-IRES-hygromycin (Addgene #97330), TdTomato (Addgene #197033). Lentiviral particles were collected in mTeSR+ media and stored at -80°C until further use.

Neuron differentiation and coculture

hiPSCs were dissociated with Accutase (innovative cell technologies AT-104), cells were counted and 2×10^5 cells were plated per well in 6-well plates coated with Matrigel (Corning 354234) in mTESR+ (Stem cell technology 100-0275) with CETP cocktail (Chroman 1 50nM MedChemExpress HY-15392, Emricasan 5mM Selleck chem S7775, Polyamine supplement 1× Sigma Aldrich P8483, trans-ISRIB 700nM R&D systems 5284). A mixture of virus was added to the cell media before plating: i) Ngn2 + rtTA was added for excitatory neuron differentiation (Zhang et al., 2013), ii) Ascl1 + Dlx2 + rtTA was added for inhibitory neuron differentiation (Yang et al., 2017). Excitatory neurons were also transduced with a lentivirus with a plasmid expressing TdTomato on day 4. to distinguish them from inhibitory neurons. On day 1, the media was changed to Neurobasal (Gibco 21103-049) supplemented with B27 (Gibco 17504044) and GlutaMAX (Gibco 35050061), doxycycline (2µg/ml, MP biomedicals 198955) was added to the media and kept for 7 days. On days 2 and 3, infected cells were selected with Puromycin (1 µg/ml, Sigma-Aldrich P8833) for excitatory neurons or Puromycin (1µg/ml) and Hygromycin (100 µg/ml, Sigma-Aldrich H9773) for inhibitory neurons. On day 4, 8×10^3 primary mouse glia were plated into Matrigel-coated wells in a 96-well plate. On day 5, induced neurons were dissociated with Accutase and counted, and 12×10^3 excitatory and 6×10^3 inhibitory cells were seeded per well into the coverslips with mouse glia in neurobasal media with 5% FBS (R&D systems S11550) and CEPT cocktail. On day 6 media was changed with neurobasal (with B27 and GlutaMAX) 5% FBS with BDNF

(10ng/ml, PeproTech 10781-164), GDNF (10 ng/ml, PeproTech 10781-226) and NT3 (10 ng/ml, PeproTech 10781-174), Cytosine β-D-arabinofuranoside (AraC 2-4 μM Sigma-Aldrich C1768) was added to the media to stop glia proliferation. Half the media was changed every 5 days with neurobasal 5% FBS with BDNF, GDNF and NT3. On day 35 cells were washed 2 times with PBS 1× and fixed with 4% PFA for 30 min. Cells were left in PBS 0.02% sodium azide until staining.

High-content imaging

For immunofluorescence staining, **h**iPSC-derived neurons were washed twice with 1× PBS and fixed with 4% PFA for 30 min in a 96-well optical bottom plate with a polymer base (Fisher Scientific: 12-566-70) at Rutgers University (New Brunswick, NJ). Fixed neurons were stored at 4ºC in 1× PBS with 0.02% sodium azide and shipped overnight to NorthShore Research Institute (Evanston, IL). Neurons were permeabilized in 1× PBS with 0.5% Triton X-100 for 15 minutes at RT without shaking. After blocking with 3% BSA and 0.1% Triton X-100 in 1× PBS for 1hr at RT, the neurons were stained with primary antibodies, mouse anti-Synapsin 1 (1:500), goat anti-tdTomato (1 μg/ml), and chicken anti-MAP2 (1:5000), in blocking buffer for 1.5hr at RT. The samples were washed three times in 1× PBS with 0.1% Triton X-100 (0.1% PBST) for 5 min each and incubated with the secondary antibodies Donkey anti-mouse Alexa 488 (1:1000), donkey anti-goat Alexa 568 (1:1000), and donkey anti-chicken Alexa 647 (1:1000) in blocking buffer for 1 hr at RT in the dark. Next, the neurons were washed twice with 0.1% PBST for 5 min, and incubated with DAPI (0.5 μg/mL, Fisher Scientific, EN62248) at RT for 10 min. Neurons were washed with 0.05% sodium azide in PBS. The plate was stored at 4℃ and allowed to warm to RT before imaging.

For Image acquisition, the neurons were imaged using Molecular Devices (San Jose, CA) ImageXpress Micro Confocal High-Content Imaging System at both 20× and 40×. The laser wavelengths used were DAPI, FITC, Texas Red, and Cy5. Each well in the 96 well plate was imaged at 8 sites for 40× and 9 sites for 20× with 8-10 *z* stacks at 1 μm step size. For the 40× objective, the pixel size is 0.3438 μm^2 with a pinhole of 60 μm, and the 20× objective pixel size is 0.6842 μm^2 also with a pinhole of 60μm.

For image analyses, the acquired mages were analyzed as 2D maximum projection. The first two morphometrics, the mean number of neurite branches per cell and the mean length of neurite outgrowth per cell were analyzed with the built-in Neurite Outgrowth Application Module within the MetaXPress 6 software,

version 6.7.2.290. Both the mean number of neurite branches per cell and the mean length of neurite outgrowth per cell were calculated based on DAPI stain as a nuclear marker and tdTomato stain, which labels excitatory neurons (see a generation of neuron culture methods), as neurite and cell body marker. Cell bodies were defined with an approximate maximum width of 30 μm, a minimum area of 300 μm^2, and a pixel value of at least 1500 above the local background level. Nuclei were identified with an approximate minimum width of 8μm, an approximate maximum width of 20 μm, and a pixel value of at least 1500 above the local background level. Neurite outgrowths were determined with a maximum width of 2 μm, a minimum projection length of 15 μm from the cell body, and a pixel value of at least 500 above the local background level. The 20× objective images were used for neurite outgrowth analysis. For assaying the third morphometrics, excitatory synapse density, we used an in-house generated custom synaptic assay module with MetaXPress 6 software. Specifically, puncta were identified through Synapsin1 staining with an approximate minimum width of 0.5 μm, an approximate maximum width of 2 μm, and a minimum pixel value of 2500 above the local background level. The number and area of Synapsin1 positive puncta within the colocalized MAP2 and tdTomato signals were used for analysis. The puncta density was generated by the number of total area of puncta within the colocalized MAP2 and tdTomato staining divided by the area of MAP2+& tdTomato+ signal within the neurites. The 40× objective images were used for synaptic puncta density analysis.

Statistical analyses

Pearson's correlation was used to evaluate the correlations between the two groups. For high content imaging, the cellular phenotypic measurements were from 8 replicates (wells with independent cell cultures), and each well's data were averaged from 9 images. For western blot, lysates of 4 different cell cultures of 2 different passages were included. However, per the journal's policy on the type of replicate and upon the editor's request, we did not present the *P*-values from any statistical testing. To determine chromosomal abnormalities in SNP e-Karyotyping, we use the statistical cut-off as described in the method (Weissbein *et al.*, 2016).

Supplementary References

Billon, P., Bryant, E.E., Joseph, S.A., Nambiar, T.S., Hayward, S.B., Rothstein, R., and Ciccia, A. (2017). CRISPR-Mediated Base Editing Enables Efficient Disruption of Eukaryotic Genes through Induction of STOP Codons. Mol Cell *67*, 1068-1079 e1064. 10.1016/j.molcel.2017.08.008.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114-2120. 10.1093/bioinformatics/btu170.

Cahan, P., Li, H., Morris, S.A., Lummertz da Rocha, E., Daley, G.Q., and Collins, J.J. (2014). CellNet: network biology applied to stem cell engineering. Cell *158*, 903-915. 10.1016/j.cell.2014.07.020.

Kluesner, M.G., Nedveck, D.A., Lahr, W.S., Garbe, J.R., Abrahante, J.E., Webber, B.R., and Moriarity, B.S. (2018). EditR: A Method to Quantify Base Editing from Sanger Sequencing. Crispr j *1*, 239-250. 10.1089/crispr.2018.0014.

Pang, Z.P., Yang, N., Vierbuchen, T., Ostermeier, A., Fuentes, D.R., Yang, T.Q., Citri, A., Sebastiano, V., Marro, S., Sudhof, T.C., and Wernig, M. (2011). Induction of human neuronal cells by defined transcription factors. Nature *476*, 220-223. nature10202 [pii]

10.1038/nature10202.

Pantazis, C.B., Yang, A., Lara, E., McDonough, J.A., Blauwendraat, C., Peng, L., Oguro, H., Kanaujiya, J., Zou, J., Sebesta, D., et al. (2022). A reference human induced pluripotent stem cell line for large-scale collaborative studies. Cell Stem Cell *29*, 1685-1702.e1622. 10.1016/j.stem.2022.11.004.

Popp, M.W., and Maquat, L.E. (2016). Leveraging Rules of Nonsense-Mediated mRNA Decay for Genome Engineering and Personalized Medicine. Cell *165*, 1319-1322. 10.1016/j.cell.2016.05.053.

Shi, J., Levinson, D.F., Duan, J., Sanders, A.R., Zheng, Y., Pe'er, I., Dudbridge, F., Holmans, P.A., Whittemore, A.S., Mowry, B.J., et al. (2009). Common variants on chromosome 6p22.1 are associated with schizophrenia. Nature *460*, 753-757. 10.1038/nature08192.

Tristan, C.A., Hong, H., Jethmalani, Y., Chen, Y., Weber, C., Chu, P.H., Ryu, S., Jovanovic, V.M., Hur, I., Voss, T.C., et al. (2023). Efficient and safe single-cell cloning of human pluripotent stem cells using the CEPT cocktail. Nat Protoc *18*, 58-80. 10.1038/s41596-022-00753-z.

Weissbein, U., Schachter, M., Egli, D., and Benvenisty, N. (2016). Analysis of chromosomal aberrations and recombination by allelic bias in RNA-Seq. Nat Commun *7*, 12144. 10.1038/ncomms12144.

Yang, N., Chanda, S., Marro, S., Ng, Y.H., Janas, J.A., Haag, D., Ang, C.E., Tang, Y., Flores, Q., Mall, M., et al. (2017). Generation of pure GABAergic neurons by transcription factor programming. Nat Methods *14*, 621-628. 10.1038/nmeth.4291.

Zhang, S., Zhang, H., Forrest, M.P., Zhou, Y., Sun, X., Bagchi, V.A., Kozlova, A., Santos, M.D., Piguel, N.H., Dionisio, L.E., et al. (2023). Multiple genes in a single GWAS risk locus synergistically mediate aberrant synaptic development and function in human neurons. Cell Genom *3*, 100399. 10.1016/j.xgen.2023.100399.

Zhang, S., Zhang, H., Zhou, Y., Qiao, M., Zhao, S., Kozlova, A., Shi, J., Sanders, A.R., Wang, G., Luo, K., et al. (2020). Allele-specific open chromatin in human iPSC neurons elucidates functional disease variants. Science *369*, 561-565. 10.1126/science.aay3983.

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.