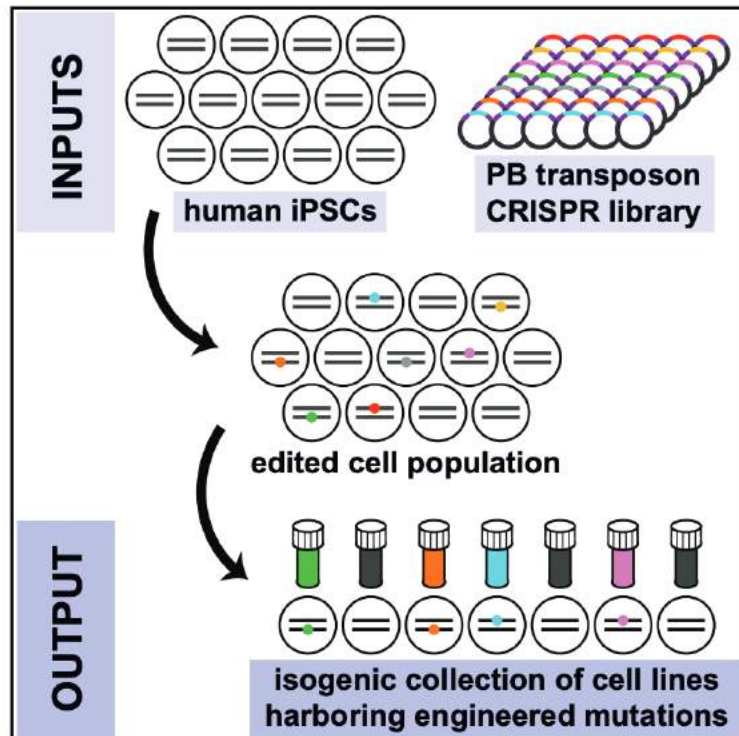## Cell Reports
Methods

Article

Parallelized engineering of mutational models using piggyBac transposon delivery of CRISPR libraries

Graphical abstract

Authors

Xander Nuttle, Nicholas D. Burt, Benjamin Currall, ..., Derek J.C. Tai, James F. Gusella, Michael E. Talkowski

Correspondence

anuttle@mgh.harvard.edu (X.N.), mtalkowski@mgh.harvard.edu (M.E.T.)

In brief

Nuttle et al. present a method for scalable genome engineering that yields dozens to hundreds of cellular models genetically differing only or primarily by CRISPR-mediated mutations. By enabling the development of large isogenic collections of such cell lines, this approach facilitates comparative functional genomic research.

Highlights

- Scalable genome engineering using gRNA libraries with piggyBac transposon delivery

- Efficient indel editing and the ability to make copy-number variants and prime edits

- Scarless transposon excision while editing yields isogenic cellular mutational models

- Large compendiums of such models empower gene and variant functional characterization

Cell Reports Methods

CellPress
OPEN ACCESS

Article

Parallelized engineering of mutational models using piggyBac transposon delivery of CRISPR libraries

Xander Nuttle,[1,2,3,4,8,*] Nicholas D. Burt,[1] Benjamin Currall,[1] Mariana Moysés-Oliveira,[1,2,3,4] Kiana Mohajeri,[1,3,4,7] Riya Bhavsar,[1,2,3,4] Diane Lucente,[1] Rachita Yadav,[1,2,3,4] Derek J.C. Tai,[1,2,3,4] James F. Gusella,[1,4,5,6] and Michael E. Talkowski[1,2,3,4,*]

[1]Center for Genomic Medicine and Department of Neurology, Massachusetts General Hospital, Boston, MA, USA
[2]Department of Neurology, Harvard Medical School, Boston, MA, USA
[3]Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA, USA
[4]Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA
[5]Department of Genetics, Blavatnik Institute, Harvard Medical School, Boston, MA, USA
[6]Harvard Stem Cell Institute, Cambridge, MA, USA
[7]PhD program in Biological and Biomedical Sciences, Harvard Medical School, Boston, MA, USA
[8]Lead contact
*Correspondence: anuttle@mgh.harvard.edu (X.N.), mtalkowski@mgh.harvard.edu (M.E.T.)
https://doi.org/10.1016/j.crmeth.2023.100672

MOTIVATION Technological advances in genome engineering and cellular modeling have catalyzed mechanistic investigations of genetic variants in development and disease. However, implementing CRISPR editing for cellular modeling at scale remains challenging, mostly limiting the functional characterization of large sets of variants to studies employing genetically heterogeneous cell populations. The method we present here produces an isogenic collection of dozens to hundreds of cell lines differing genetically only or primarily by CRISPR-induced mutations in a single experiment. The ability to rapidly generate powerful cellular resources unlocks new possibilities for delineating relationships between genotype and phenotype.

SUMMARY

New technologies and large-cohort studies have enabled novel variant discovery and association at unprecedented scale, yet functional characterization of these variants remains paramount to deciphering disease mechanisms. Approaches that facilitate parallelized genome editing of cells of interest or induced pluripotent stem cells (iPSCs) have become critical tools toward this goal. Here, we developed an approach that incorporates libraries of CRISPR-Cas9 guide RNAs (gRNAs) together with inducible *Cas9* into a piggyBac (PB) transposon system to engineer dozens to hundreds of genomic variants in parallel against isogenic cellular backgrounds. This method empowers loss-of-function (LoF) studies through the introduction of insertions or deletions (indels) and copy-number variants (CNVs), though generating specific nucleotide changes is possible with prime editing. The ability to rapidly establish high-quality mutational models at scale will facilitate the development of isogenic cellular collections and catalyze comparative functional genomic studies investigating the roles of hundreds of genes and mutations in development and disease.

INTRODUCTION

A major challenge in human genomics is translating human disease and trait associations of genetic variants into functional biological insights and pathogenic mechanisms.[1] The increasing scale of genome and exome sequencing studies has produced a deluge of gene and variant associations across Mendelian and complex disorders,[2–5] yet forging mechanistic links from geno...

organisms.[6,7] These approaches can have limitations in the number of patients that can be recruited with the desired variants, access to clinically relevant tissues, and difficulties in interpretation arising from variable genetic backgrounds or uncertain generalizability of findings in animals for humans.[8,9] Consequently, there is a critical need for approaches that can rapidly model disease-associated variants and perturb functional networks to inform therapeutic targeting and development.

The emergence of induced pluripotent stem cell (iPSC) and...

CellPress
OPEN ACCESS

Cell Reports Methods
Article

unprecedented opportunities to experimentally interrogate mutations of interest.[10,11] An expanding compendium of *in vitro* differentiation protocols has enabled researchers to routinely cultivate a wide variety of cell types in a dish, including several distinct populations of neurons.[12,13] Three-dimensional cell culture systems have also yielded organoids that recapitulate some hallmark properties of their *in vivo* counterparts, including cerebral organoids as models of brain development.[13,14] Genome editing tools, most notably CRISPR-Cas9 and its relatives,[15,16] provide methods to engineer a growing catalog of precisely targeted mutations, including insertions or deletions (indels),[17–19] point mutations,[20,21] copy-number variants (CNVs),[22,23] and inversions.[23,24] Continued technological developments such as prime editing promise to enhance editing efficiencies, augment targeting possibilities, and expand the range of programmable mutations.[25,26]

To fully capitalize on these advances for cellular mutational modeling, particularly for strategies involving genome engineering in iPSCs to establish clonal mutant lines, approaches to parallelize and scale the experimental workflow are needed. It is currently possible to generate multiple programmed edits by performing separate transfections with different gRNAs, but time, cost, and labor considerations preclude using this approach for generating hundreds of individual mutations. Saturation genome editing has yielded cell populations with 3,893 coding single-nucleotide variants (SNVs), including 96.5% of all possible SNVs within thirteen targeted *BRCA1* exons.[27,28] This strategy is extremely powerful for introducing desired genetic changes at a single locus, though modifying other sites, even within the same gene, has required separate gRNA transfections. Lentiviral gRNA libraries have been successfully employed in genome-scale genetic screens, including attempts to knock out 19,050 human genes and 20,611 mouse genes.[29–32] However, resulting edited cells harbor integrated lentiviral sequence in addition to any CRISPR-mediated mutations, complicating efforts to infer their functional consequences. Moreover, none of these large-scale editing strategies has yet been coupled with isolating individual mutant cells to create cellular models, and thus functional characterization of large sets of engineered variants has been mostly limited to population-based assays and single-cell omics.[33–35]

Here, we leveraged the piggyBac (PB) transposon, a mobile genetic element that can be scarlessly removed from the cellular genome,[36] to deliver *Cas9*[37–40] together with a gRNA library to iPSCs and perform parallelized genome engineering. We employed this approach for introducing indels, CNVs, and targeted substitutions and evaluated editing outcomes by using molecular inversion probes (MIPs)[41–43] and massively parallel sequencing to genotype clonal lines expanded from single cells. The method produces hundreds of cellular models in a single experiment, with indel formation and large CNV modeling[22] showing the highest editing efficiencies.

RESULTS

Genome engineering strategy
We developed a workflow for engineering cellular models in parallel (Figure 1) to produce iPSC lines genetically differing only by

sulting transposon containing a gRNA gene, inducible *Cas9*, and a puromycin resistance gene. Transfecting this library into iPSCs together with a plasmid encoding PB transposase yields a mixed population of cells, with several integrating one or more PB transposons and constitutively expressing one or more gRNAs. Puromycin treatment selects for PB-containing cells, and addition of doxycycline to the culture medium induces editing. Scarlessly excising PB transposons[37] yields isogenic iPSC models; however, we initially omitted the PB excision process during most experiments (Table S1) to compare integrated gRNA genes with mutations generated and evaluate editing efficiency. After editing and optional PB excision, we perform fluorescence-activated cell sorting (FACS) to isolate single cells and allow colonies to grow in 96-well tissue culture plates for 2 weeks. Colonies are then split into replicate plates for propagation and DNA extraction. MIP genotyping at gRNA target sites identifies indels and point mutations, while data from MIPs targeting SNVs across regions of interest allow detection of CNVs based on altered allele balance.[41–43]

PB library delivery and generation of indels
To evaluate this approach, we cloned a library encoding 40 gRNAs targeting exonic sequence in 40 genes into our PB vector and attempted editing as described above. We delivered this library to iPSCs by nucleofecting 1 million cells with 5 μg library plasmid DNA together with 1 μg PB transposase plasmid. Puromycin selection resulted in numerous resistant surviving colonies. After selection, we treated cells with doxycycline for 4 days to induce editing, sorted for live cells, and cultivated resultant colonies to obtain 308 clonal lines for analysis. Our genotyping assay included MIPs that successfully captured targets within our plasmids, gRNA cleavage sites, and integrated gRNA genes, enabling estimation of PB integration copy numbers, assessment of gRNA representation, and comparisons of potential edits with observed mutations for each line. Furthermore, we designed the gRNA constructs to include molecular tags, random 10 bp sequences following the gRNA poly-T transcription termination sequence. The genotyping assay also captured these tags, allowing us to distinguish distinct integrations of the same gRNA gene.

Counting integrated gRNA constructs in each line indicated that transfection delivered 0–8 different gRNA genes to each cell (corresponding to 0–9 distinct gRNA constructs, i.e., gene-tag combinations), with 84.1% of lines (259 of 308) having received three or fewer different gRNA genes (Figure 2A). Lines where no gRNA gene was detected may have originated from cells protected from puromycin by nearby resistant colonies. MIP targets within the PB transposon (and no plasmid MIP targets outside the PB transposon) were captured in 85.4% of lines (263 of 308), indicating that most integrations were mediated by PB transposition. The likelihood of engineering a specific mutation of interest using this system depends in part on the relative integration frequency of the corresponding gRNA gene. For example, an inefficient gRNA might generate more mutant lines than a highly active gRNA if the former were expressed in more cells than the latter. Thus, to assess gRNA representation, we

Cell Reports Methods
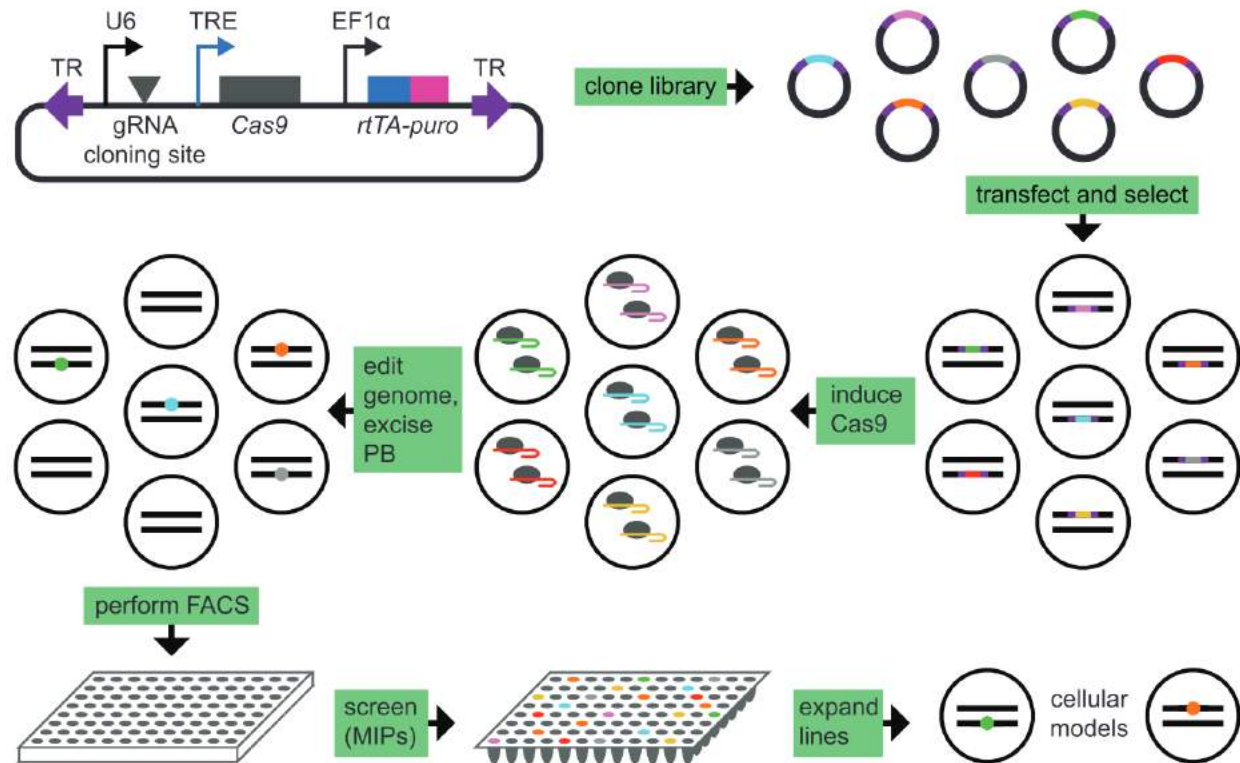## Article

CellPress
OPEN ACCESS

Figure 1. PB system for parallelized genome engineering
Schematic outlines experimental workflow. (Top) Cloning a gRNA library into the PB plasmid vector. (Middle) Delivering and removing CRISPR components via PB transposition enables multiplexed, temporally controlled editing. (Bottom) Isolating single cells establishes clonal iPSC lines, which are genotyped using MIPs to identify mutational models. TR, terminal repeat; TRE, tetracycline response element; rtTA, reverse tetracycline-controlled transactivator.

15.5; Figure 2B), and their distribution was relatively consistent but statistically non-uniform (exact multinomial goodness of fit $p < 1 \times 10^{-5}$, 1,000,000 simulations). Among 660 PB integrations, 596 distinct gRNA constructs were identified, with over 90% of these (537 of 596) detected only once (Figure S1). Most remaining constructs (43 of 59) probably only integrated once, as they were only observed in likely clonally related lines sharing two or more integrated constructs. Taken together, these data indicate that our PB method can efficiently deliver a gRNA library to iPSCs without substantial bottlenecking during cloning, PB integration, selection, or FACS, resulting in good gRNA representation among recovered lines.

We next examined editing efficiency and mutational outcomes. Sequence analysis of gRNA target sites revealed that 95% (38 of 40) had inserted and/or deleted bases in at least one iPSC line (Figure 2B). Considering only lines containing at least one integrated gRNA gene, 51.8% (144 of 278) were edited at one or more corresponding target sites, with most of these lines (68.8%, 99 of 144) acquiring such indels from activity of only one gRNA. Lines harboring more integrated gRNA and *Cas9* genes were more likely to have acquired at least one indel, as the proportion of edited lines generally increased with transgene copy number (Figure 2A). Given that several gRNAs mediated

Overall, we identified 236 editing events, defined as indel editing at a target locus in a single line. Interestingly, we found 35 events where the corresponding gRNA gene was not detected as integrated, suggesting that either these indels were mediated by gRNAs expressed from residual transfected plasmids or that not all integrated gRNA genes were successfully captured during MIP genotyping. Edits were most commonly heterozygous for the indel (42.4%, 100 of 236, with indel allele balances from 0.4 to 0.6), with smaller fractions homozygous for the indel (7.2%, 17 of 236), compound heterozygous for different indels (21.6%, 51 of 236, with allele balances from 0.4 to 0.6), or likely mosaic (28.8%, 68 of 236). As expected for Cas9 editing, some indels were accompanied by one or more substitution mutations. In aggregate, this single experiment generated 153 mutant iPSC lines harboring coding indels in 38 different genes. For 31 of these genes (or 26 if considering only heterozygous edits), we obtained at least one single-gene mutational iPSC model lacking edits at all other targets. Importantly, this approach was reproducible, yielding similar results (Figure S2) from an experiment using starting iPSC lines derived from two individuals (1 male, 1 female) with lipofection delivery of a PB library containing the same 40 gRNAs without molecular tags. These results highlight the power of PB genome engineering for high-throughput production of cellular

CellPress
OPEN ACCESS
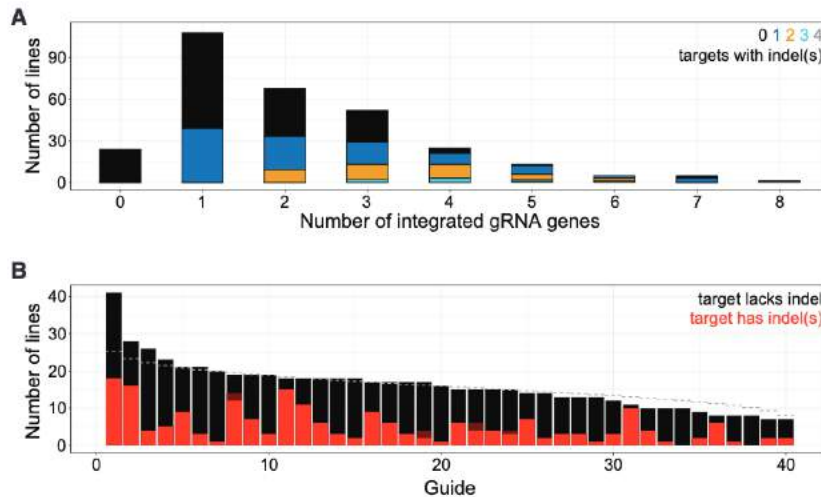
Cell Reports Methods
Article

Figure 2. Characterization of PB integrations and indel editing

(A) Distribution of the number of distinct integrated gRNA genes per iPSC line and resultant indel editing. Bar heights indicate counts of lines containing from 0–8 different gRNA genes. Fill color within each bar signifies proportions of lines with indel editing at one (blue) or multiple (yellow, cyan, gray) corresponding target sites. Seven lines were excluded from the plot because they lacked confident gRNA gene integration genotypes.

(B) gRNA representation and guide efficiencies for indel mutagenesis. Bar heights indicate counts of iPSC lines containing each gRNA gene, with guides ordered based on abundance. Fill color within each bar signifies the proportion of lines harboring each gRNA gene with indel editing at the corresponding target (red), lacking such editing (black), or having an uncertain genotype (maroon) due to poor MIP capture and low sequencing coverage across the target site. Dashed line shows expected gRNA representation based on sampling from a uniform distribution (1,000,000 simulations).

Production of multiple CNVs associated with reciprocal genomic disorders (RGDs)

Editing with a single gRNA targeting duplicated sequences on the same chromosome occasionally results in two double-strand DNA breaks on a single homolog, with imperfect repair leading to deletion or duplication of intervening sequence. We have previously leveraged this strategy to generate cellular models with CNVs equivalent to those found in patients with RGDs.[22] Here, we performed three separate PB experiments (Table S1) to generate these CNVs in a parallelized manner: a pilot experiment including a library encoding 59 gRNAs targeting segmental duplications flanking 19 known RGD loci (as well as three other single gRNAs and 70 gRNA pairs), a CNV-focused experiment using a library encoding the same 62 single gRNAs as the pilot experiment, and an experiment with PB excision that incorporated a library encoding 57 gRNAs conducive to CNV formation as described above at the same 19 regions. We used a maximum likelihood approach[42] utilizing allele balance data for SNVs across targeted loci to generate copy number calls for each target region in each iPSC line recovered after FACS.

Aggregating results from these experiments, PB editing yielded thirteen CNVs at nine RGD loci, including eight microdeletions (~280 kbp to ~5.3 Mbp; Figure 3A) and five microduplications (~280 kbp to ~4.64 Mbp; Figure 3B). Our data indicate that 28 CNV editing events occurred (Table S3), comprising 19 deletion edits and 9 duplication edits, with each instance of dosage alteration at a target locus in a single line considered an event except for two CNVs spanning adjacent targeted regions at chromosome 1q21, which most likely arose spontaneously during cell culture. Remarkably, one line acquired two CRISPR-mediated CNVs: both the ~280 kbp distal chromosome 16p11.2 deletion and the nearby ~740 kbp chromosome 16p11.2 duplication. We estimated deletion and duplication editing efficiencies for individual gRNAs as well as aggregate CRISPR deletion and duplication ef-

portion of lines derived from a single transfection experiment harboring a programmed CNV, considering only lines containing at least one integrated gRNA gene designed for CNV engineering. For the excision experiment, our yield analysis included all lines recovered because based on the experimental design, all those lines should have integrated and expressed at least one such gRNA gene. These yields ranged from 0% (0 of 74 recovered lines) to 11.9% (5 of 42 recovered lines), with the yield from the excision experiment at 5.9% (12 of 202 recovered lines). Collectively, these analyses demonstrate that the PB system can deliver multiple RGD cellular models in parallel, though the efficiencies of generating RGD rearrangements remain low, comparable to our initial publication of the RGD modeling method.[22]

Dual-guide genic deletion, Cas3-mediated deletion, and prime editing

Since our PB method successfully generated coding indels and large CNVs implicated in RGDs, we next explored whether it could be adapted to accommodate engineering focal deletions in unique sequence and specified point mutations and indels. Accordingly, we designed gRNA libraries for dual-guide genic deletion using Cas9, dual-guide genic deletion using Cas12a,[44] deletion of genomic segments using CRISPR-Cas3,[45,46] and prime editing.[25] We cloned them into PB vectors modified from our original vector to encode corresponding editing machinery. Subsequent editing experiments yielded targeted complete deletions of *MAGEL2* (~6.9 kbp, 1 line) and *CHD2* (~133 kbp, 1 line) with the dual-guide Cas9 approach, partial deletions of *CUL3* (~1.8–7.7 kbp, 2 lines) and *NRXN1* (~4.4–11.9 kbp, 2 lines) with the Cas3 strategy, and two heterozygous prime edits: a substitution in *CHD2* (D856G [GRCh38/hg38 chr15:92,974,940 A-to-G], 1 line, allele balance 0.5) and a frameshifting indel in *CHD8* (Y1762fs [GRCh38/hg38 chr14:21,393,671 CAT-to-C], 1 line, allele balance 0.53). Despite these successes, none of these libraries mediated efficient editing

Cell Reports Methods
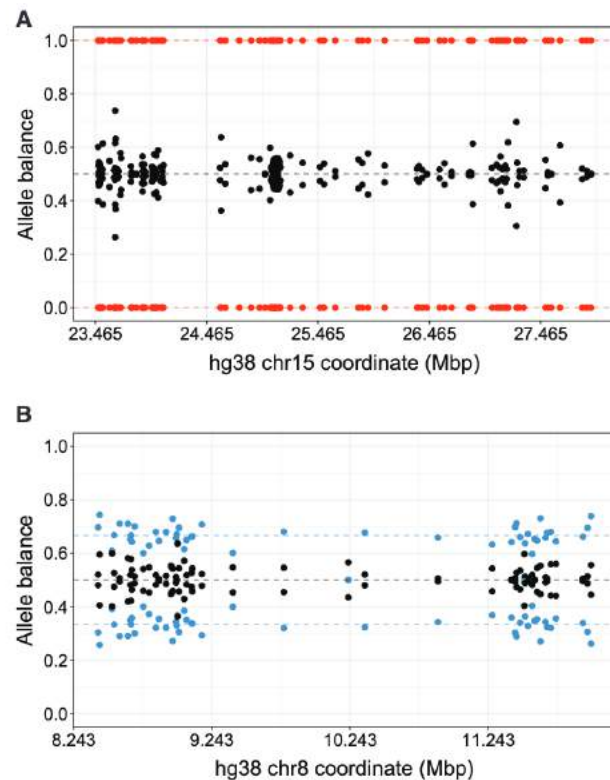## Article

CellPress
OPEN ACCESS

Figure 3. Selected PB CRISPR RGD-associated CNVs

Plots depict allele balance (points) at MIP targets across genomic sequence for lines harboring a Cas9-induced deletion (red) or duplication (blue) and representative unedited controls (black).

(A) An ~5.3 Mbp RGD deletion at chromosome 15q11.2-q13.1 and (B) an ~4.64 Mbp RGD duplication at chromosome 8p23.1, each mediated by a single gRNA targeting flanking duplicated sequences (not shown).

of editing modalities, though improving editing efficiencies will be necessary to practically leverage our workflow for these applications.

PB excision

Creating isogenic cellular models differing genetically only by CRISPR-induced mutations using this method requires removal of transgenic editing components. We employed an excision-only PB transposase for this purpose because it achieves >99.7% scarless excision, nearly always restoring integration sites to their original sequences.[36,37] We modified the PB vector to encode *puΔtk*, which confers resistance to puromycin and susceptibility to fialuridine,[47] such that counter-selection after PB excision should kill cells retaining any integrations. Attempting the workflow using this vector, most colonies did not survive puromycin treatment, suggesting that the *puΔtk* used here did not confer sufficient puromycin resistance for use with CRISPR libraries. We thus replaced *puΔtk* with the original puromycin resistance gene fused to *EGFP*, enabling sorting for GFP-negative cells to isolate those where PB copies were excised. We incorporated a

without molecular tags. Nucleofecting iPSCs with this plasmid library together with the PB transposase plasmid yielded numerous colonies that survived puromycin treatment and glowed when viewed under a fluorescence microscope.

We attempted PB excision using three approaches and evaluated them based on FACS analysis, reasoning that higher excision rates would manifest as higher percentages of GFP-negative cells. Transfecting iPSCs with a plasmid (first approach) or mRNA (second approach) encoding excision-only PB transposase after editing did not promote efficient excision, with nearly identical rates of GFP-negative cells between these populations and iPSCs from the same experiment transfected with a control plasmid or mRNA after editing. Anticipating that stronger transposase expression may be necessary for efficient excision, our third strategy incorporated the excision-only PB transposase gene either on a separate PB construct (Figure S3A) or within the same construct, fused to inducible *Cas9* (Figure 4A). We employed this single all-in-one vector to deliver our indel gRNA library to iPSCs, split the puromycin-resistant population into multiple wells, and treated cells in one of the wells with doxycycline for 4 days to simultaneously induce editing and PB excision. Comparison of treated and untreated cell populations revealed a much higher proportion of GFP-negative cells with treatment (55.8% vs. 27.1%; Figures 4B and S3B). These results indicate that the third approach successfully induced PB excision, though some transgene silencing also occurred.

We sorted GFP-negative cells from two replicates of this experiment using different starting iPSC lines into 96-well plates and followed the same workflow to obtain 461 clonal lines for analysis. Overall, 17.1% lines (79 of 461) showed complete PB excision (Figure 4C), including 68 of 367 clones (18.5%) from the first iPSC line and 11 of 94 clones (11.7%) from the second iPSC line. Importantly, this designation combines evidence from two MIPs targeting integrated gRNA genes and fifteen MIPs targeting other transgenic sequences present within our PB construct, providing confidence that these clones truly lack PB integrations. The remaining lines retained 1–5 different integrated gRNA genes, indicating that this population included more clones with low PB copy numbers compared to our indel experiment without excision (Figure 2A) and providing additional evidence that some excision took place. Editing yielded 263 lines harboring coding indels, with most edited lines (62.4%, 164 of 263) modified at a single target (Figure 4D). Among 411 editing events, heterozygous indels were most common (169 events, 41.1%), followed by likely mosaics (116 events, 28.2%), compound heterozygotes (98 events, 23.8%), and homozygotes (28 events, 6.8%), consistent with our previous results. Altogether, this experiment produced 263 iPSC lines collectively containing coding indels in 38 genes. For 33 of these genes (or 26 if considering only heterozygotes), we acquired at least one iPSC model edited exclusively at that target (Figure 4E). These data demonstrate that successful PB excision can be integrated with parallelized editing to generate an isogenic collection of cellular models.

DISCUSSION

CRISPR development has greatly enhanced the possibilities for

CellPress
OPEN ACCESS
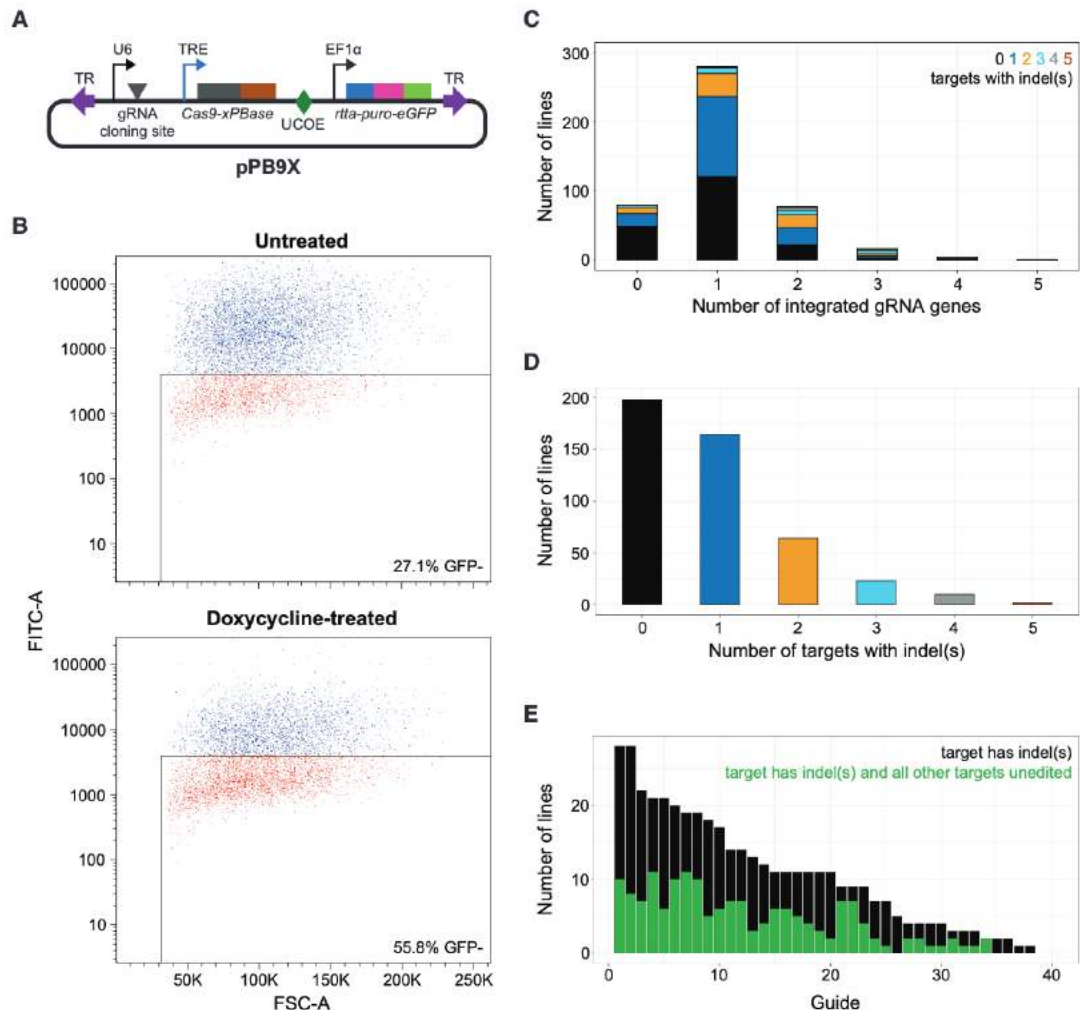
Cell Reports Methods
Article

Figure 4. PB excision

(A) Plasmid containing PB transposon construct for simultaneous editing and excision is shown. TR, terminal repeat; TRE, tetracycline response element; UCOE, ubiquitous chromatin opening element; rtTA, reverse tetracycline-controlled transactivator.

(B) FACS plots depict GFP fluorescence (y axis) and forward scatter area (x axis) from the 488 nm laser for live single cells (points) derived from two wells of transfected cells selected with puromycin, where one well was then treated with doxycycline for 4 days (bottom) and the other well was not treated (top). Gate encompasses cells deemed GFP negative (red), set based on analysis of naive cells from the same experiment.

(C–E) Analyses of iPSC lines recovered from the excision experiment.

(C) Distribution of the number of distinct integrated gRNA genes per iPSC line and resultant indel editing. Bar heights indicate counts of lines containing from 0–5 different gRNA genes. Fill color within each bar signifies proportions of lines with indel editing at one (blue) or multiple (yellow, cyan, gray, dark orange) corresponding target sites. Five lines were excluded from the plot because they had no detected gRNA gene integrations but were not deemed integration-free based on MIP data.

(D) Distribution of the number of distinct gRNA targets (genes) edited per iPSC line. Bar heights indicate counts of lines harboring from 0–5 different mutated targets (genes).

(E) Guide efficiencies for indel mutagenesis. Bar heights indicate counts of iPSC lines that acquired indel(s) at each gRNA target, with guides ordered based on editing frequency. Fill color within each bar signifies the proportion of lines with indel editing at each target exclusively (green) or in combination with indel formation at one or more additional targets (black).

cellular modeling remains challenging, particularly when establishing mutant iPSC lines at scale for further studies. We developed a method for creating such mutational models in parallel and engineered 461 lines with coding indels among 38 different genes in

bor requirements and associated costs, consolidating cell culture experiments that could take many months. PB editing also obviates the need for repeated collection and storage of unedited control lines corresponding to each mutant, provided that obtaining

Cell Reports Methods
Article

CellPress
OPEN ACCESS

and extremely large RGD-associated CNVs at comparable efficiency (3.5% for deletions and 2.5% for duplications) to our initial SCORE strategy.[22] PB CRISPR libraries were also compatible with prime editing and gene deletion, though these efficiencies were below 1% even with extended doxycycline treatment. These data are consistent with mechanistic considerations and previous work showing higher editing rates for indels than CNVs,[49,50] with RGD deletions made using a single gRNA produced more efficiently than comparable deletions mediated by two gRNAs.[22] Importantly, we observed primarily single-copy indel formation, suggesting that we could efficiently obtain heterozygous loss-of-function mutants, which often model human disease more precisely than homozygous knockouts. Furthermore, we recovered 99 lines with indels at multiple loci, demonstrating that the PB system can deliver oligogenic models.

Unlike most transposons, PB can be excised precisely without genetic scarring at integration sites,[36] and a previous study leveraging PB for genome engineering found that all integration sites analyzed after excision had been restored to their original sequences.[37] By incorporating PB excision into our workflow, we capitalized on this property to enable generation of lines differing only via programmed mutations, i.e., isogenic sets. Because only a modest proportion of lines (17.1%) attained excision of all PB integrations, applications requiring such models should be limited to indel-focused projects until efficiencies of other PB mutagenesis options and/or PB excision can be increased. Importantly, a previous study showed that genome engineering in iPSCs with a PB system did not promote substantial off-target editing, with recovered lines showing 1–3 unintended mutations likely related to editor activity.[37] This report also found that PB editing did not lead to an abnormal karyotype, nor did it affect pluripotency markers or differentiation potential.

We envision that PB genome engineering could be used to genetically dissect pathogenic CNV regions[51] or to disrupt sets of genes involved in a common biological pathway or statistically associated with a disease phenotype.[3] For example, the chromosome 16p11.2 RGD locus contains 25 protein-coding genes.[6] Inactivating each one via frameshifting indels would entail designing a library of 25 gRNAs, or 50 gRNAs if including two gRNAs per gene to control for potential off-target effects. With one or two additional gRNAs targeting segmental duplications flanking the genes, this library would also mediate chromosome 16p11.2 RGD deletions.[22] Our simulations (Figure S4A) indicate that for a library containing 52 gRNA constructs, 1,963 PB integrations on average would be necessary to achieve at least 25 integrations of each construct. Since we observed an average of 2.19 integrations per line, we estimate 897 lines would collectively harbor the desired number of integrations. Our molecular tag analysis indicates that our method can accommodate larger gRNA libraries while likely maintaining good gRNA representation, as >90% of integrated gRNA constructs were observed only once (Figure S1). In addition, PB has successfully been employed to deliver a genome-scale gRNA library to mouse liver.[52] Accordingly, project goals and personnel time considerations should inform library size, as using larger libraries necessitates recovering more lines to acquire multiple replicates of all desired mutant models.

gest that boosting gRNA and/or editor expression may promote more efficient mutagenesis. Using a PB platform akin to that presented here to deliver single gRNA pairs to stem cells and induce editing yielded small deletions (<3 kbp) at estimated rates above 40%.[38] A more recent study described PB-based prime editing that successfully introduced a single targeted substitution in >50% of iPSCs analyzed, though high editing rates were not achieved using all prime editing guide RNAs (pegRNAs).[53] This study leveraged a constitutively expressed prime editor fused to *puΔtk* for dual selection, making it probable that prime editor expression was high in cells surviving puromycin treatment. Based on these data, having multiple integrated copies of a given pegRNA construct per cell and resulting higher pegRNA expression levels appear conducive to efficient editing, provided that pegRNA designs have been optimized. Given increased indel formation with higher PB integration copy numbers (Figure 2A), modifying transfection conditions to attain more integrated PB constructs may augment editing rates. Although in our pilot experiment we found no relationship between PB plasmid library DNA input and numbers of integrations per cell (k-sample Anderson-Darling p = 0.2, 1,000,000 simulations; Figure S4B), multiple studies exploring a wider range of PB DNA input levels have reported higher per-cell PB integration counts.[37–39] Unfortunately, such an approach would make complete PB excision more difficult. Recovering clones lacking PB could potentially be enhanced by engineering the excision-only PB transposase and/or PB terminal repeats[54] to increase excision efficiency and/or by incorporating additional UCOEs[55] into our PB construct, further mitigating transgene silencing and improving enrichment of transgene-free lines via GFP-negative selection. Regardless of further optimization, for many applications, PB removal is unnecessary, especially given that remaining integrations in our final iPSC models are inactive.

Deploying PB CRISPR libraries for parallelized production of cellular mutational models unlocks several exciting new possibilities. First, this method could be leveraged to generate compendiums of isogenic iPSC lines with frameshifting indels across hundreds of genes to investigate consequences of partial or complete loss of function. Second, the approach could be applied to achieve multiplex mutagenesis within cells, rendering oligogenic modeling more tractable and scalable. Third, with further development, we anticipate that our strategy will accommodate base editing[20,21] and prime editing,[25] expediting the creation of variant atlases[56] and cellular models featuring patient-specific point mutations. Given the relatively high efficiency of base editing, it should dovetail well with the PB system without requiring extensive optimization. Fourth, PB libraries increase multiplexing capacity for CRISPR activation/interference applications.[39,40,57] In all these ways, this PB method will promote systematic efforts to compile large collections of cellular models in isogenic backgrounds. Such allelic series will serve as powerful resources for comparative functional genomics,[58] driving research into the impacts of genes and variants at unprecedented scope and scale.

Limitations of the study

As discussed above, the most significant limitations of this

CellPress
OPEN ACCESS

Cell Reports Methods
Article

addition, we were not able to precisely control PB integration copy number, so experiments requiring that each cell express only one gRNA should utilize another approach for library delivery such as lentiviral transduction at a low multiplicity of infection. Lentiviral CRISPR strategies are compatible with large gRNA libraries,[29–32] though unlike PB integrations, lentiviral integrations cannot be scarlessly removed from cellular genomes. Finally, we performed all experiments in two human iPSC lines, but we realize that in some cases, researchers may want to utilize our workflow with other cell types. We anticipate that the method generalizes to any cell type where PB delivery works well, though editing efficiencies would likely differ between cell types. In all cases, we recommend transfecting cells via nucleofection whenever possible since our nucleofection experiments showed greater reproducibility and more even gRNA representation among recovered lines (Figure 2B) than our lipofection experiments (Figure S2B).

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABLILTY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Plasmid construction
  - gRNA library design
  - gRNA library cloning
  - gRNA library transfection
  - Selection and genome engineering
  - PB excision
  - FACS and 96-well iPSC culture
  - 96-Well DNA extraction
  - WGS
  - SNV calling
  - MIP genotyping
  - MIP analyses
  - PB experiment simulations
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.crmeth.2023.100672.

AUTHOR CONTRIBUTIONS

X.N. and M.E.T. designed the study. X.N. designed gRNA libraries and performed cloning experiments, including construction of PB vectors and assembly of gRNA libraries, with assistance from N.D.B. X.N. and N.D.B. performed iPSC experiments with assistance from R.B. M.M.-O., K.M., and D.J.C.T. performed iPSC transfection experiments exploring a wide range of conditions and assisted with troubleshooting during early stages of method development. B.C. performed library construction for whole-genome sequencing (WGS) of the unedited iPSC lines used in experiments. R.Y. performed SNV and indel calling from WGS data. X.N. designed MIPs, performed genotyping experiments, analyzed data, and performed simulations. D.L. provided resources. X.N., J.F.G., and M.E.T. wrote the paper with input and approval from all co-authors.

REFERENCES

1. McCarthy, M.I., and MacArthur, D.G. (2017). Human disease genomics: from variants to biology. Genome Biol. 18, 20.

2. Chong, J.X., Buckingham, K.J., Jhangiani, S.N., Boehm, C., Sobreira, N., Smith, J.D., Harrell, T.M., McMillin, M.J., Wiszniewski, W., Gambin, T., et al. (2015). The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities. Am. J. Hum. Genet. 97, 199–215.

3. 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.

4. Sanders, S.J., Neale, B.M., Huang, H., Werling, D.M., An, J.Y., Dong, S., Whole Genome Sequencing for Psychiatric Disorders WGSPD; Abecasis, G., Arguello, P.A., Blangero, J., et al. (2017). Whole genome sequencing in psychiatric disorders: the WGSPD consortium. Nat. Neurosci. 20, 1661–1668.

5. Taliun, D., Harris, D.N., Kessler, M.D., Carlson, J., Szpiech, Z.A., Torres, R., Taliun, S.A.G., Corvelo, A., Gogarten, S.M., Kang, H.M., et al. (2021). Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. Nature 590, 290–299.

6. Blumenthal, I., Ragavendran, A., Erdin, S., Klei, L., Sugathan, A., Guide, J.R., Manavalan, P., Zhou, J.Q., Wheeler, V.C., Levin, J.Z., et al. (2014). Transcriptional consequences of 16p11.2 deletion and duplication in mouse cortex and multiplex autism families. Am. J. Hum. Genet. 94, 870–883.

7. Kim, Y., Lee, H.M., Xiong, Y., Sciaky, N., Hulbert, S.W., Cao, X., Everitt, J.I., Jin, J., Roth, B.L., and Jiang, Y.H. (2017). Targeting the histone methyltransferase G9a activates imprinted genes and improves survival of a mouse model of Prader-Willi syndrome. Nat. Med. 23, 213–222.

8. Peitz, M., Jungverdorben, J., and Brüstle, O. (2013). Disease-specific iPS cell models in neuroscience. Curr. Mol. Med. 13, 832–841.

9. Erickson, R.P. (1996). Mouse models of human genetic disease: which mouse is more like a man? Bioessays. 18, 993–998.

10. Hockemeyer, D., and Jaenisch, R. (2016). Induced Pluripotent Stem Cells Meet Genome Editing. Cell Stem Cell 18, 573–586.

Cell Reports Methods
Article

 CellPress
OPEN ACCESS

12. Kim, J.E., O'Sullivan, M.L., Sanchez, C.A., Hwang, M., Israel, M.A., Brennand, K., Deerinck, T.J., Goldstein, L.S.B., Gage, F.H., Ellisman, M.H., and Ghosh, A. (2011). Investigating synapse formation and function using human pluripotent stem cell-derived neurons. Proc. Natl. Acad. Sci. USA 108, 3005–3010.

13. 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.

14. Velasco, S., Kedaigle, A.J., Simmons, S.K., Nash, A., Rocha, M., Quadrato, G., Paulsen, B., Nguyen, L., Adiconis, X., Regev, A., et al. (2019). Individual brain organoids reproducibly form cell diversity of the human cerebral cortex. Nature 570, 523–527.

15. Doudna, J.A., and Charpentier, E. (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science (New York, N.Y.) 346, 1258096.

16. Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. Cell 157, 1262–1278.

17. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science (New York, N.Y.) 339, 819–823.

18. Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science (New York, N.Y.) 339, 823–826.

19. Cho, S.W., Kim, S., Kim, J.M., and Kim, J.S. (2013). Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat. Biotechnol. 31, 230–232.

20. Gaudelli, N.M., Komor, A.C., Rees, H.A., Packer, M.S., Badran, A.H., Bryson, D.I., and Liu, D.R. (2017). Programmable base editing of A·T to G·C in genomic DNA without DNA cleavage. Nature 551, 464–471.

21. Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., and Liu, D.R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 533, 420–424.

22. Tai, D.J.C., Ragavendran, A., Manavalan, P., Stortchevoi, A., Seabra, C.M., Erdin, S., Collins, R.L., Blumenthal, I., Chen, X., Shen, Y., et al. (2016). Engineering microdeletions and microduplications by targeting segmental duplications with CRISPR. Nat. Neurosci. 19, 517–522.

23. Kraft, K., Geuer, S., Will, A.J., Chan, W.L., Paliou, C., Borschiwer, M., Harabula, I., Wittler, L., Franke, M., Ibrahim, D.M., et al. (2015). Deletions, Inversions, Duplications: Engineering of Structural Variants using CRISPR/Cas in Mice. Cell Rep. 10, 833–839.

24. Park, C.Y., Kim, D.H., Son, J.S., Sung, J.J., Lee, J., Bae, S., Kim, J.H., Kim, D.W., and Kim, J.S. (2015). Functional Correction of Large Factor VIII Gene Chromosomal Inversions in Hemophilia A Patient-Derived iPSCs Using CRISPR-Cas9. Cell Stem Cell 17, 213–220.

25. Anzalone, A.V., Randolph, P.B., Davis, J.R., Sousa, A.A., Koblan, L.W., Levy, J.M., Chen, P.J., Wilson, C., Newby, G.A., Raguram, A., and Liu, D.R. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. Nature 576, 149–157.

26. Walton, R.T., Christie, K.A., Whittaker, M.N., and Kleinstiver, B.P. (2020). Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. Science (New York, N.Y.) 368, 290–296.

27. Findlay, G.M., Boyle, E.A., Hause, R.J., Klein, J.C., and Shendure, J. (2014). Saturation editing of genomic regions by multiplex homology-directed repair. Nature 513, 120–123.

28. Findlay, G.M., Daza, R.M., Martin, B., Zhang, M.D., Leith, A.P., Gasperini, M., Janizek, J.D., Huang, X., Starita, L.M., and Shendure, J. (2018). Accurate classification of BRCA1 variants with saturation genome editing. Nature 562, 217–222.

29. Wang, T., Wei, J.J., Sabatini, D.M., and Lander, E.S. (2014). Genetic

30. Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelson, T., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G., and Zhang, F. (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. Science (New York, N.Y.) 343, 84–87.

31. Koike-Yusa, H., Li, Y., Tan, E.P., Velasco-Herrera, M.D.C., and Yusa, K. (2014). Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. Nat. Biotechnol. 32, 267–273.

32. Sanjana, N.E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. Nat. Methods 11, 783–784.

33. Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C.P., Jerby-Arnon, L., Marjanovic, N.D., Dionne, D., Burks, T., Raychowdhury, R., et al. (2016). Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. Cell 167, 1853–1866.e17.

34. Adamson, B., Norman, T.M., Jost, M., Cho, M.Y., Nuñez, J.K., Chen, Y., Villalta, J.E., Gilbert, L.A., Horlbeck, M.A., Hein, M.Y., et al. (2016). A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. Cell 167, 1867–1882.e21.

35. Jaitin, D.A., Weiner, A., Yofe, I., Lara-Astiaso, D., Keren-Shaul, H., David, E., Salame, T.M., Tanay, A., van Oudenaarden, A., and Amit, I. (2016). Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. Cell 167, 1883–1896.e15.

36. Li, X., Burnight, E.R., Cooney, A.L., Malani, N., Brady, T., Sander, J.D., Staber, J., Wheelan, S.J., Joung, J.K., McCray, P.B., Jr., et al. (2013). piggyBac transposase tools for genome engineering. Proc. Natl. Acad. Sci. USA 110, E2279–E2287.

37. Wang, G., Yang, L., Grishin, D., Rios, X., Ye, L.Y., Hu, Y., Li, K., Zhang, D., Church, G.M., and Pu, W.T. (2017). Efficient, footprint-free human iPSC genome editing by consolidation of Cas9/CRISPR and piggyBac technologies. Nat. Protoc. 12, 88–103.

38. Schertzer, M.D., Thulson, E., Braceros, K.C.A., Lee, D.M., Hinkle, E.R., Murphy, R.M., Kim, S.O., Vitucci, E.C.M., and Calabrese, J.M. (2019). A piggyBac-based toolkit for inducible genome editing in mammalian cells. RNA (New York, N.Y.) 25, 1047–1058.

39. Hazelbaker, D.Z., Beccard, A., Angelini, G., Mazzucato, P., Messana, A., Lam, D., Eggan, K., and Barrett, L.E. (2020). A multiplexed gRNA piggyBac transposon system facilitates efficient induction of CRISPRi and CRISPRa in human pluripotent stem cells. Sci. Rep. 10, 635.

40. Li, S., Zhang, A., Xue, H., Li, D., and Liu, Y. (2017). One-Step piggyBac Transposon-Based CRISPR/Cas9 Activation of Multiple Genes. Nucleic acids 8, 64–76.

41. Hardenbol, P., Banér, J., Jain, M., Nilsson, M., Namsaraev, E.A., Karlin-Neumann, G.A., Fakhrai-Rad, H., Ronaghi, M., Willis, T.D., Landegren, U., and Davis, R.W. (2003). Multiplexed genotyping with sequence-tagged molecular inversion probes. Nat. Biotechnol. 21, 673–678.

42. Nuttle, X., Huddleston, J., O'Roak, B.J., Antonacci, F., Fichera, M., Romano, C., Shendure, J., and Eichler, E.E. (2013). Rapid and accurate large-scale genotyping of duplicated genes and discovery of interlocus gene conversions. Nat. Methods 10, 903–909.

43. Hiatt, J.B., Pritchard, C.C., Salipante, S.J., O'Roak, B.J., and Shendure, J. (2013). Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation. Genome Res. 23, 843–854.

44. Campa, C.C., Weisbach, N.R., Santinha, A.J., Incarnato, D., and Platt, R.J. (2019). Multiplexed genome engineering by Cas12a and CRISPR arrays encoded on single transcripts. Nat. Methods 16, 887–893.

45. Dolan, A.E., Hou, Z., Xiao, Y., Gramelspacher, M.J., Heo, J., Howden, S.E., Freddolino, P.L., Ke, A., and Zhang, Y. (2019). Introducing a Spectrum of Long-Range Genomic Deletions in Human Embryonic Stem Cells Using Type I CRISPR-Cas. Mol. Cell 74, 936–950.e5.

46. Morisaka, H., Yoshimi, K., Okuzaki, Y., Gee, P., Kunihiro, Y., Sonpho, E., Xu, H., Sasakawa, N., Naito, Y., Nakada, S., et al. (2019). CRISPR-Cas3

CellPress
OPEN ACCESS

Cell Reports Methods
Article

47. Chen, Y.T., and Bradley, A. (2000). A new positive/negative selectable marker, puDeltatk, for use in embryonic stem cells. Genesis 28, 31–35.

48. Cullmann, K., Blokland, K.E.C., Sebe, A., Schenk, F., Ivics, Z., Heinz, N., and Modlich, U. (2019). Sustained and regulated gene expression by Tet-inducible "all-in-one" retroviral vectors containing the HNRPA2B1-CBX3 UCOE(®). Biomaterials 192, 486–499.

49. Gasperini, M., Findlay, G.M., McKenna, A., Milbank, J.H., Lee, C., Zhang, M.D., Cusanovich, D.A., and Shendure, J. (2017). CRISPR/Cas9-Mediated Scanning for Regulatory Elements Required for HPRT1 Expression via Thousands of Large, Programmed Genomic Deletions. Am. J. Hum. Genet. 101, 192–205.

50. Canver, M.C., Bauer, D.E., Dass, A., Yien, Y.Y., Chung, J., Masuda, T., Maeda, T., Paw, B.H., and Orkin, S.H. (2014). Characterization of genomic deletion efficiency mediated by clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells. J. Biol. Chem. 289, 21312–21324.

51. Golzio, C., Willer, J., Talkowski, M.E., Oh, E.C., Taniguchi, Y., Jacquemont, S., Reymond, A., Sun, M., Sawa, A., Gusella, J.F., et al. (2012). KCTD13 is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant. Nature 485, 363–367.

52. Xu, C., Qi, X., Du, X., Zou, H., Gao, F., Feng, T., Lu, H., Li, S., An, X., Zhang, L., et al. (2017). piggyBac mediates efficient in vivo CRISPR library screening for tumorigenesis in mice. Proc. Natl. Acad. Sci. USA 114, 722–727.

53. Eggenschwiler, R., Gschwendtberger, T., Felski, C., Jahn, C., Langer, F., Sterneckert, J., Hermann, A., Lühmann, J., Steinemann, D., Haase, A., et al. (2021). A selectable all-in-one CRISPR prime editing piggyBac transposon allows for highly efficient gene editing in human cell lines. Sci. Rep. 11, 22154.

54. Luo, W., Hickman, A.B., Genzor, P., Ghirlando, R., Furman, C.M., Menshikh, A., Haase, A., Dyda, F., and Wilson, M.H. (2022). Transposase N-terminal phosphorylation and asymmetric transposon ends inhibit piggyBac transposition in mammalian cells. Nucleic Acids Res. 50, 13128–13142.

55. Rudina, S.S., and Smolke, C.D. (2019). A Novel Chromatin-Opening Element for Stable Long-term Transgene Expression. Preprint at bioRxiv.

56. Fowler, D.M., Adams, D.J., Gloyn, A.L., Hahn, W.C., Marks, D.S., Muffley, L.A., Neal, J.T., Roth, F.P., Rubin, A.F., Starita, L.M., and Hurles, M.E. (2023). An Atlas of Variant Effects to understand the genome at nucleotide resolution. Genome Biol. 24, 147.

57. Chardon, F.M., McDiarmid, T.A., Page, N.F., Martin, B., Domcke, S., Regalado, S.G., Lalanne, J.B.Æ., Calderon, D., Starita, L.M., Sanders, S.J., et al. (2023). Multiplex, single-cell CRISPRa screening for cell type specific regulatory elements. Preprint at bioRxiv..

58. Hilton, I.B., and Gersbach, C.A. (2015). Enabling functional genomics with genome engineering. Genome Res. 25, 1442–1455.

59. Fu, J.M., Satterstrom, F.K., Peng, M., Brand, H., Collins, R.L., Dong, S., Wamsley, B., Klei, L., Wang, L., Hao, S.P., et al. (2022). Rare coding variation provides insight into the genetic architecture and phenotypic context of autism. Nat. Genet. 54, 1320–1331.

60. Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat. Biotechnol. 34, 184–191.

61. Sanson, K.R., Hanna, R.E., Hegde, M., Donovan, K.F., Strand, C., Sullender, M.E., Vaimberg, E.W., Goodale, A., Root, D.E., Piccioni, F., and Doench, J.G. (2018). Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. Nat. Commun. 9, 5416.

62. Coe, B.P., Witherspoon, K., Rosenfeld, J.A., van Bon, B.W.M., Vulto-van Silfhout, A.T., Bosco, P., Friend, K.L., Baker, C., Buono, S., Vissers, L.E.L.M., et al. (2014). Refining analyses of copy number variation identifies specific genes associated with developmental delay. Nat. Genet. 46, 1063–1071.

63. Wang, T., Guo, H., Xiong, B., Stessman, H.A.F., Wu, H., Coe, B.P., Turner, T.N., Liu, Y., Zhao, W., Hoekzema, K., et al. (2016). De novo genic mutations among a Chinese autism spectrum disorder cohort. Nat. Commun. 7, 13316.

64. Stessman, H.A.F., Xiong, B., Coe, B.P., Wang, T., Hoekzema, K., Fenckova, M., Kvarnung, M., Gerdts, J., Trinh, S., Cosemans, N., et al. (2017). Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases. Nat. Genet. 49, 515–526.

65. Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., et al. (2007). Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947–2948.

66. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.

67. Alkan, C., Kidd, J.M., Marques-Bonet, T., Aksay, G., Antonacci, F., Hormozdiari, F., Kitzman, J.O., Baker, C., Malig, M., Mutlu, O., et al. (2009). Personalized copy number and segmental duplication maps using next-generation sequencing. Nat. Genet. 41, 1061–1067.

68. Bae, S., Park, J., and Kim, J.S. (2014). Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. Bioinformatics 30, 1473–1475.

69. Menon, A.V., Sohn, J.I., and Nam, J.W. (2020). CGD: Comprehensive guide designer for CRISPR-Cas systems. Comput. Struct. Biotechnol. J. 18, 814–820.

70. Hsu, J.A.-O., Grünewald, J.A.-O., Szalay, R.A.-O., Shih, J., Anzalone, A.V., Lam, K.C., Shen, M.A.-O., Petri, K.A.-O., Liu, D.R., Joung, J.A.-O., and Pinello, L.A.-O. (2021). PrimeDesign software for rapid and simplified design of prime editing guide RNAs. Nat. Commun. 12, 1034.

71. Aparicio-Prat, E., Arnan, C., Sala, I., Bosch, N., Guigó, R., and Johnson, R. (2015). DECKO: Single-oligo, dual-CRISPR deletion of genomic elements including long non-coding RNAs. BMC Genom. 16, 846.

72. Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

73. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at ArXiv.

74. (2019). Picard Toolkit (Broad Institute, GitHub Repository). http://broadinstitute.github.io/picard/.

75. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079.

76. Poplin, R., Ruano-Rubio, V., DePristo, M.A., Fennell, T.J., Carneiro, M.O., Van der Auwera, G.A., Kling, D.E., Gauthier, L.D., Levy-Moonshine, A., Roazen, D., et al. (2018). Scaling accurate genetic variant discovery to tens of thousands of samples. Preprint at bioRxiv.

77. Nuttle, X., Itsara, A., Shendure, J., and Eichler, E.E. (2014). Resolving genomic disorder-associated breakpoints within segmental DNA duplications using massively parallel sequencing. Nat. Protoc. 9, 1496–1513.

78. Zhang, J., Kobert, K., Flouri, T., Stamatakis, A., and Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 30, 614–620.

79. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760.

80. Ferrante, M., and Saltalamacchia, M. (2014). The Coupon Collector's problem. MATerials MATemàtics 2014, 35.

Cell Reports Methods
Article

CellPress
OPEN ACCESS

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
NEB Stable Competent E. coli (High Efficiency)	New England BioLabs	C3040H
Chemicals, peptides, and recombinant proteins		
puromycin dihydrochloride	Gibco	A1113803
doxycycline hydrochloride	Sigma-Aldrich	D3072
rhLaminin-521	Gibco	A29249
CloneR2	Stemcell Technologies	100–0691
Critical commercial assays		
Quick-DNA 96 Kit	Zymo Research	D3012
MIP reagents	Nuttle et al.[77]	N/A
ZymoPURE II Plasmid Maxiprep Kit	Zymo Research	D4202
Human Stem Cell Nucleofector Kit 1	Lonza	VPH-5012
Deposited data		
MIP sequencing data	This paper	National Center for Biotechnology Information Sequence Read Archive (SRA: PRJNA955320)
WGS data for unedited iPSC lines used in this study (derived from two individuals: 1 male, 1 female)	This paper	https://nda.nih.gov/edit_collection.html?id=2304
Experimental models: Cell lines		
female human iPSC line	Massachusetts General Hospital	MGH2069
male human iPSC line	Coriell Institute for Medical Research	GM08330
Oligonucleotides		
gRNA libraries	This paper	N/A
MIPs	This paper	N/A
MIP primers	Nuttle et al.[77]	N/A
MIP primers for dual indexing	This paper	N/A
Recombinant DNA		
PB transposon plasmids (Table S4), except those encoding excision-only PB transposase	This paper	Addgene (catalog numbers 210028, 210029, 210030, 210032, 210033, 210034, 210035, and 210036)
Software and algorithms		
Analysis code for this paper	https://github.com/xnuttle/PB_paper	https://doi.org/10.5281/zenodo.8401815
PEAR	Zhang et al.[78]	https://www.h-its.org/downloads/pear-academic/
BWA	Li and Durbin[79]	https://github.com/lh3/bwa
SAMtools	Li et al.[75]	https://github.com/samtools

RESOURCE AVAILABLILTY

Lead contact
Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Xander Nuttle (anuttle@mgh.harvard.edu).

CellPress
OPEN ACCESS

Cell Reports Methods
Article

Materials availability

Plasmids generated in this study that do not encode PB transposase or excision-only PB transposase have been deposited to Addgene (catalog numbers 210028, 210029, 210030, 210032, 210033, 210034, 210035, and 210036).

Data and code availability

- MIP sequencing data is available at the US National Center for Biotechnology Information Sequence Read Archive under BioProject ID PRJNA955320 (SRA: PRJNA955320). WGS data for unedited iPSC lines used in this study (derived from two individuals: 1 male, 1 female) are available through the NIMH Data Archive (Collection 2304) at https://nda.nih.gov/edit_collection.html?id=2304.
- Programs and scripts for analyzing MIP sequencing data are available on GitHub at https://github.com/xnuttle/PB_paper.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

We used two human iPSC lines, one derived from a 43-year-old female (MGH2069) and the other derived from a 63-year-old male (GM08330). iPSCs were cultured on tissue culture-treated plates (Falcon) coated with Matrigel hESC-Qualified Matrix (Corning) in Essential 8 medium (E8, Gibco) supplemented with penicillin/streptomycin (P/S, Gibco). For routine culture, cells were maintained on 6-well plates in a humidified incubator at 37°C with 5% CO_2 and passaged every 3–4 days (at ~70–80% confluence). For passaging, cells were washed with Dulbecco's Phosphate-Buffered Saline lacking calcium and magnesium (DPBS, Gibco), treated with ReLeSR (Stemcell Technologies) for 7 min at 37°C, resuspended in working medium supplemented with 10 µM ROCK inhibitor (RI, Y-27632, Biological Industries), and plated into new wells after pipetting to reduce the size of cell aggregates. Working medium (E8 + P/S) was refreshed daily. For thawing, cells were resuspended in working medium supplemented with 10 µM RI and plated.

METHOD DETAILS

Plasmid construction

We constructed a collection of PB transposon plasmids (Table S4) to accommodate cloning a variety of gRNA libraries. We also generated plasmids encoding PB transposase and excision-only PB transposase. All plasmid constructs were validated through massively parallel sequencing or long-read nanopore sequencing performed by the MGH DNA core. Brief overviews of molecular cloning steps to create each of these plasmids follow. pPB9v1. We constructed pPB9v1 through a series of cloning steps modifying pPB-rtTA-hCas9-puro-PB (graciously provided by Dr. William Pu).[37] Briefly, a gRNA cloning site derived from pPB-US-ECasE (a gift from Eleanor Chen, Addgene plasmid #83961) was inserted before the *TRE3G* promoter; sequence encoding an SV40 nuclear localization signal (NLS) was added to the beginning of the *Cas9* gene; internal *Cas9* genic sequence was replaced with corresponding sequence from PX459 (a gift from Feng Zhang, Addgene plasmid #48139) to eliminate BbsI restriction sites; sequence encoding the SV40 NLS at the end of the *Cas9* gene was replaced with sequence encoding a nucleoplasmin NLS; and the gRNA cassettes after *Cas9* were removed. pSPBase and pXPBase. We created the PB transposase plasmid pSPBase and the excision-only PB transposase plasmid pXPBase by removing the *ccdB* gene from PB210PA-1 and PB220PA-1 (System Biosciences), respectively, to render them generally propagable. pPB9v2. We developed pPB9v2 by modifying the gRNA cloning site in pPB9v1 to accommodate incorporation of molecularly-tagged single Cas9 guide constructs via Gibson assembly. pPB9P2. We constructed pPB9P2 by modifying the gRNA cloning site in pPB9v2 to accommodate incorporation of PE2 prime editing guide constructs via Gibson assembly and then replacing ending *Cas9* genic sequence with corresponding sequence from pCMV-PE2 (a gift from David Liu, Addgene Plasmid #132775) encoding prime editor 2. pPB1163. We created pPB1163 by modifying pPB9v2 as follows. The puromycin resistance gene was replaced with a blasticidin resistance gene; the gRNA cloning site was altered to accommodate incorporation of dual Cas3 guide constructs via Gibson assembly; and the *Cas9* gene was replaced by a *Cas11-Cas6-Cas3* fusion gene derived from pPV-EF1a-2xNLS-Cse2-Cas6-Cas3-iPA (a gift from Tomoji Mashimo, Addgene plasmid #134925), with the three individual *Cas* genes separated by sequences encoding 2A self-cleaving peptides. pPB758. We developed pPB758 by removing the gRNA cloning site from pPB9v1 and replacing the *Cas9* gene with a *Cas7-Cas5-Cas8* fusion gene derived from pPV-EF1a-2xNLS-Cas7-Cas5-Cse1-iPA (a gift from Tomoji Mashimo, Addgene plasmid #134924), with the three individual *Cas* genes separated by sequences encoding 2A self-cleaving peptides. pPB9P3. We constructed pPB9P3 by removing the U6 promoter from pPB9P2 and modifying the gRNA cloning site to accommodate incorporation of PE3/3b prime editing gRNA constructs via Gibson assembly. pPB12A. We created pPB12A by removing the gRNA cloning site from pPB9v1, replacing the *Cas9* gene with a *Cas12a* gene derived from pY010 (a gift from Feng Zhang, Addgene plasmid #69982), and inserting triplex sequence and a cloning site for a Cas12a gRNA cassette[44] between *Cas12a* coding sequence and its polyadenylation signal sequence. pPB9. We developed pPB9 by inserting the 672 bp human *CBX3* UCOE[48] into pPB9v1 after *Cas9* and before the EF1α promoter and replacing the *rtTA-puro* fusion gene with a *puro-eGFP* fusion gene, with the *puro* and *eGFP* genes separated by sequence encoding a 2A self-cleaving peptide. pPBX. We constructed pPBX as follows. The gRNA cloning site in pPB1163 was altered to accommodate incorporation of single

Cell Reports Methods
Article

CellPress
OPEN ACCESS

and before the EF1α promoter. pPB9X. We modified pPB9 to create pPB9X by appending the codon-optimized excision-only PB transposase gene *xPBase* to *Cas9* and joining the *rtTA* gene to the *puro-eGFP* fusion gene, in both cases leveraging sequence encoding a 2A self-cleaving peptide to separate different genic components.

gRNA library design

We designed several gRNA libraries for different PB experiments attempting a variety of editing strategies (Table S1). Brief descriptions of the gRNA design process for these libraries follow. Indel libraries. We designed two gRNA libraries to perform indel editing in coding sequences. These libraries utilize the same single Cas9 gRNAs, except all gRNAs in the second library have a random 10 nt sequence (i.e., molecular tag) incorporated immediately after the gRNA poly-T transcription termination sequence. Most genes selected for indel editing showed an association with autism spectrum disorder (ASD) in a recent study,[59] and all were amenable to MIP genotyping. Twenty single gRNAs targeting coding sequence for each gene were initially designed using CRISPick,[60,61] which ranks designs based on their predicted editing efficiencies (with higher scores better) and their predicted potentials for off-target editing (with lower potentials better). Guides with cut sites outside of sequence captured by a known[62–64] or predicted high-performing MIP were eliminated from further consideration, as resulting indels would not be detected using our MIP genotyping assay. Guides that cut inside such a MIP target but within 27 bp of its edge (i.e., outside a narrowed MIP target) were similarly discarded, since indels near the edge of a MIP target would prevent successful capture if they disrupted sequence flanking the target complementary to MIP arms. The highest-ranked remaining gRNA for each gene was selected for inclusion in the libraries. SCORE libraries. These libraries employed single Cas9 gRNAs to engineer microdeletions and microduplications of chromosomal regions associated with RGDs. To design these gRNAs, sequences of segmental duplications flanking these regions (and in some cases, also related paralogous sequences) were aligned using Clustal 2.1[65] or MAFFT v7.310.[66] These alignments were parsed to identify all candidate gRNA sequences, defined as 20 bp sequences identical between target segmental duplications that precede a canonical 3′ NGG *S. pyogenes* protospacer adjacent motif (PAM) in both and lack an identical protospacer in other aligned paralogous sequences. gRNA candidates were then scored for predicted efficiency using Azimuth 2.0[60] and initially assessed for potential off-target activity by concatenating protospacers with each possible NGG PAM and mapping them to the human reference genome GRCh38/hg38 using mrFAST.[67] All guide candidates having more perfect matches than a specified threshold were eliminated. To analyze possible off-target sites more carefully, we used Cas-OFFinder[68] to identify all loci in GRCh38/hg38 with up to five mismatches to each candidate protospacer with its PAM, not counting mismatches at the degenerate first PAM position. Cas-OFFinder output was parsed to compute genome-wide counts of sequences having 0–5 mismatches to each candidate as well as the maximum Cutting Frequency Determination (CFD) score[60] considering all identified off-target sequences for each candidate. We manually selected guides from our filtered set of candidates considering both their predicted activities (with higher Azimuth 2.0 scores better) and their likelihoods of specific targeting (with lower maximum CFD scores and fewer off-target sites with 1–2 mismatches better). Dual-guide libraries (Cas9). These libraries leverage pairs of Cas9 gRNAs to generate deletions of single genes linked to ASD[59] (or partial deletions of these genes), genes within a few genomic regions linked with RGDs, or deletions of multigene segments within these regions. To design the first guide pair library (used in the PB pilot experiment), we modified our pipeline for designing SCORE libraries to derive candidate gRNAs from single sequences rather than from sequence alignments, to exclude candidates within genes, and to apply lower perfect match thresholds when filtering candidates. For the second guide pair library (used in the PB genedel experiment), we downloaded gRNA designs from the GRCh38/hg38 UCSC Genome Browser "CRISPR Targets" track and selected designs considering their expected activities and targeting fidelities, as above, as well as their cut sites' compatibility with MIP genotyping. Dual-guide library (Cas12a). This library utilizes cassettes of six Cas12a gRNAs to mediate complete gene deletions for a subset of genes targeted by dual-guide Cas9 editing and a few additional ASD-associated genes. Each cassette encodes three gRNAs targeting intergenic sequence before the gene of interest and three gRNAs targeting intergenic sequence after the gene of interest. For each target gene, sequences on each flank (5,000 nt each) were input into Comprehensive Guide Designer (CGD),[69] which output all possible Cas12a gRNA targets within these sequences and associated editing efficacy scores. As above, Cas-OFFinder[68] was used to compute genome-wide counts of sequences having 0–5 mismatches to each candidate. Finally, single-molecule MIPs were designed[42,43] across the same flanking sequences input to CGD. Cas12a guides were considered for inclusion in each gene's cassette if their cut sites fell within a MIP target, with final choices informed by CGD efficacy scores and specificities based on Cas-OFFinder results. Dual-guide library (Cas3). This library leverages Cas3 gRNAs (expressed in pairs) to produce focal deletions disrupting the same genes targeted by dual-guide Cas12a editing. Sequences flanking each gene of interest (5,000 nt on each side) were parsed to identify candidate Cas3 gRNA sequences, defined as 32 bp target sequences immediately preceded by a 5′ AAG PAM. The fifty candidates closest to the gene on each side oriented such that corresponding Cas3 deletions would impact the gene were mapped to GRCh38/hg38 using mrFAST,[67] and candidates with more than one perfect match throughout the genome were eliminated. Cas-OFFinder[68] was then used to obtain genome-wide counts of sequences having 0–7 mismatches to each candidate, which informed our final gRNA selections. Prime editing libraries. We designed two libraries for prime editing: a library of prime editing guide RNAs (pegRNAs) to mediate PE2 prime editing and a library of prime editing guide pairs (pegRNAs and corresponding nicking guide RNAs [ngRNAs]) to induce PE3/3b prime editing. Most variants selected for prime editing were identified in one or more patients from ASD cohorts and affect one of the genes implicated in ASD in the study described above.[59] All these variants were compatible with

CellPress
OPEN ACCESS

Cell Reports Methods
Article

case, at least one high-performing MIP target (narrowed by 27 bp on each side, as above) includes the interval between the site to be edited and the pegRNA cut site. The PE3/3b library includes the same 20 pegRNAs as the PE2 library coupled with their corresponding ngRNAs. It also incorporates twenty additional pegRNA-ngRNA designs compatible with PE3b prime editing where in each case, at least one narrowed high-performing MIP target includes the widest interval spanning the site to be edited, the pegRNA cut site, and the ngRNA cut site. These design criteria ensure that successful prime edits and nearly all undesired indels arising from pegRNA or ngRNA activity would be detected using our genotyping strategy.

gRNA library cloning

Below we briefly describe steps for cloning each of the various gRNA libraries into corresponding vectors to prepare PB plasmid libraries ready for transfection into iPSCs. In each workflow, PCR reactions for library amplification were cleaned up using AMPure XP magnetic beads (Agencourt). Indel libraries and SCORE libraries. For the PB pilot and PB SCORE experiments, we cloned single Cas9 gRNAs into pPB9v1 as follows. DNA oligonucleotides (oligos) corresponding to single gRNAs were ordered (provided at 10 μM in IDTE buffer, pH 8.0; Integrated DNA Technologies), pooled, diluted in elution buffer to a pool concentration of ~15 nM, and amplified using PCR primers Oligo-Fwd-2 (5′-GGCTTTATATATCTTGTGGAAAGGACGAAACACC-3′) and Oligo-Rev-2 (5′-GCCTTATTTTAACTTGCTATTTCTAGCTC TAAAAC-3′). pPB9v1 was digested with BbsI-HF (New England BioLabs), followed by gel purification of the vector backbone. We cloned our single-guide amplicon library into this vector using NEBuilder HiFi DNA Assembly Master Mix (New England BioLabs), incubating the Gibson assembly reaction at 50°C for 30 min. The resulting plasmid library was transformed into NEB Stable Competent E. coli (New England BioLabs), and 1% of the transformation reaction was plated to estimate library coverage by counting colonies. The remainder was used to inoculate a 100 mL culture (LB with 100 μg/mL carbenicillin) grown in a shaking incubator (250 rpm) at 30°C for 20–24 h. Endotoxin-free plasmid library DNA was prepared using the ZymoPURE II Plasmid Maxiprep Kit (Zymo Research). For the PBINP2C3 and PB excision experiments, we followed the same workflow except starting oligos were ordered as oPool oligo pools (provided with 50 pmol/oligo; Integrated DNA Technologies) and resuspended and diluted in nuclease-free water to a total oligo concentration of 1 nM prior to PCR amplification. In the dual-construct PB excision experiment, two different amplicon libraries were cloned into pPB9 and pPBX, respectively, while the all-in-one construct PB excision experiment (PB excision 2) employed an amplicon library cloned into pPB9X. For the PBINP3C12A experiment, we similarly prepared a starting oPool and generally followed the workflow above, except PCR amplification used primers Oligo-Fwd-2 and XN-DG-AMP-2R (5′- CACTCCTTTCAAGACCTAGCTAG-3′) and the resulting amplicon library was cloned into the pPB9v2 vector backbone prepared via BbsI-HF digestion and gel purification as above. Dual-guide libraries (Cas9). To clone libraries of Cas9 gRNA pairs, we adapted previously published cloning workflows.[49,71] For the PB genedel experiment, we ordered DNA oligos (provided at 10 μM in IDTE buffer, pH 8.0; Integrated DNA Technologies) corresponding to each gRNA as well as a gBlock (Integrated DNA Technologies) containing gRNA backbone sequence and H1 promoter sequence. Each oligo pair was ordered in a separate well, and all oligos shared a common 18 bp of overlap with their partner at their 3′ ends. Oligo pairs were converted to double-stranded fragments containing linked gRNA pairs separated by a BbsI cloning site via a single cycle of PCR. These fragments were then pooled, gel purified, diluted in elution buffer, PCR-amplified using primers Oligo-Fwd-2 and Oligo-Rev-2, and cloned into pPB9v1 using Gibson assembly with the reaction incubated for 1 h. The resulting plasmid library and gBlock were digested with BbsI-HF and gel purified. The gBlock insert was cloned into the vector library using the Quick Ligation Kit (New England BioLabs) with a 7.5-min incubation. Transformation, bacterial culture, and final DNA preparation for this plasmid library followed the workflow above. For the dual guide library used in the PB pilot experiment, we used the same approach as in the PB genedel experiment except dual guide constructs were assembled in an intermediate vector due to the use of Esp3I restriction sites rather than BbsI restriction sites. Full-length dual guide constructs were PCR-amplified from the intermediate vector library and cloned into the final pPB9v1 vector using Gibson assembly as above. For cloning new Cas9 dual guide libraries, we recommend adapting the strategy below for cloning PE3/3b prime editing libraries which replaces the inefficient ligation step with a second round of Gibson assembly. Dual-guide library (Cas12a), dual-guide library (Cas3), and PE2 prime editing library. Starting oligos for these libraries were ordered as oPools and resuspended and diluted in nuclease-free water to 1 nM concentrations as above, then PCR-amplified using corresponding primers: for the Cas12a library, c12a_fwd (5′- CAAGATCCTGGTATTGGTCTGCG-3′) and c12a_rev (5′- GCCTCCCCCGTTTAAACTCATTAC-3′); for the Cas3 library, XN-C3G-AMP-F (5′- ATAAAGTTGGTAGATTGTGACTGGC-3′) and XN-DG-AMP-2R; and for the PE2 library, Oligo-Fwd-2 and XN-DG-AMP-2R. Gibson assembly was used to clone amplicon libraries into corresponding vectors: for the Cas12a library, pPB12A; for the Cas3 library, pPB1163; and for the PE2 library, pPB9P2. Transformation, bacterial culture, and final DNA preparation for these plasmid libraries followed the workflow above. PE3 prime editing library. Starting oligos for this library were ordered as an oPool, resuspended and diluted in nuclease-free water as above, and PCR-amplified using primers XN-P-AMP-F (5′- TTTTGCTGGCCTTTTGCTC-3′) and XN-DG-AMP-2R. The resulting amplicon library was then digested with BbsI-HF to render the amplicon ends compatible with Gibson assembly into pPB9P3, and the digestion reaction was cleaned up using magnetic beads. pPB9P3 was digested with Esp3I (New England BioLabs), and the vector backbone was gel purified. Gibson assembly was then performed to clone the amplicon library into pPB9P3, followed by transformation, bacterial culture, and DNA preparation as above to yield a plasmid library lacking gRNA promoters (the PE3 pre-library). A gBlock containing partial U6 and H1 promoter sequences was cloned into pCR-Blunt II-TOPO using the Zero Blunt TOPO PCR Cloning Kit (Invitrogen). The transformation reaction was plated on an LB + kanamycin plate, eight resulting colonies were picked to inoculate 5 mL cultures (LB with 50 μg/mL kanamycin), and the QIAprep Spin Miniprep Kit (Qiagen) was used to isolate plasmid DNA (pU6rH1)

Cell Reports Methods
Article

CellPress

OPEN ACCESS

were gel purified. Gibson assembly was performed using this vector and insert to produce the final PE3 plasmid library capable of expressing pegRNAs and partner ngRNAs from U6 and H1 promoters in opposite orientations. Transformation, bacterial culture, and final DNA preparation for this plasmid library followed the workflow above.

gRNA library transfection
Some of our experiments used nucleofection for transfecting iPSCs with one or multiple PB plasmid libraries together with PB transposase plasmid pSPBase, while other experiments employed lipofection (Table S1). We found that our nucleofection approach was more robust than our lipofection strategy, with less variability in cell health post-transfection and in the number of resistant colonies surviving selection. Consistent with this observation, another recent report detailed relatively low reproducibility of lipofection for PB delivery.[53] For these reasons, we recommend nucleofection as the best transfection method when working with our PB system. General overviews of both transfection workflows are provided below, with specific details for each transfection (e.g., identities and amounts of PB plasmid libraries included) provided elsewhere (Table S1). Nucleofection. At least 90 min prior to cell detachment, cells were treated with working medium supplemented with 10 µM RI. To detach cells, each input well was washed with DPBS and treated with 750 µL accutase cell detachment solution (Stemcell Technologies) for 10 min at 37°C. The cell and accutase mixture in each well was then pipetted several times to generate a single-cell suspension, each time dispensing the mixture throughout the well to help dislodge cells still attached. The resulting suspension was then pipetted through a 40 µm cell strainer (Falcon) into a 15 mL tube (Falcon) to remove multicell aggregates and centrifuged at 800 rpm for 3 min. The cell pellet was resuspended in Dulbecco's Modified Eagle Medium/Nutrient Mixture F-12 (DMEM/F12, Gibco), and a cell count was obtained using the Countess II Cell Counter (Life Technologies). One million cells were nucleofected with PB plasmid library (or libraries) mixed with pSPBase using the Human Stem Cell Nucleofector Kit 1 (Lonza) and the Amaxa Nucleofector II device (Lonza) set to program B-016. Nucleofected cells were transferred into working medium supplemented with 10 µM ROCK inhibitor and plated into a single well of a 6-well, Matrigel-coated plate. Lipofection. The day before transfection, iPSCs were seeded into a new Matrigel-coated well in a 6-well plate by performing the same steps described above for passaging, except pipetting a few more times to generate smaller cell aggregates. The volume of cell suspension added to the new well was adjusted to attempt to achieve ~20% confluence upon plating. Twenty-four hours later, cells were transfected with PB plasmid library (or libraries) mixed with pSPBase using Lipofectamine Stem reagent (Invitrogen) according to the manufacturer's protocol. Cells were exposed to transfection complexes for 24 h, then medium was refreshed.

Selection and genome engineering
All our PB experiments involved selection for iPSCs that integrated at least one PB transposon followed by induction of editing activity. However, experimental details including selection reagents used, concentrations of these reagents, and time frames for selection and editing sometimes differed between experiments. Selection conditions were sufficient to kill all cells after control transfections including PB plasmid library alone or PB transposase plasmid alone, except for the least stringent condition (0.2 µg/mL puromycin, used in two experiments as a last resort to prioritize cell health when lipofection resulted in excessive cell death). We provide general overviews of our selection and induction procedures after nucleofection and lipofection below, with specifics for each experiment provided elsewhere (Table S1). Nucleofection. Between 30 and 31 h after transfection, each well was treated with working medium containing the selection reagent(s), most commonly puromycin dihydrochloride (Gibco) at a concentration of 0.5 µg/mL. Working medium with the selection reagent(s) was refreshed daily until the end of the selection period, and medium for any passaging during the selection period included the selection reagent(s). Each well of resistant iPSCs was seeded into new wells as described above (i.e., passaging with smaller cell aggregates). The following day, cells were treated with working medium supplemented with doxycycline hydrochloride (Sigma-Aldrich) at a concentration of 2 µg/mL to induce expression of the editor. Working medium with doxycycline was refreshed daily until the end of the editing period, and medium for any passaging during the editing period included doxycycline. Doxycycline induction and editing always occurred after selection, except for PE3 and Cas12a editing, where a lower concentration of puromycin was used during the editing period (except when passaging) to mitigate transgene silencing. Lipofection. Three days after transfection, we performed single-cell passaging using accutase to seed 750,000 cells into a new well in a 6-well plate. Twenty-four hours later, cells were treated with working medium containing the selection reagent(s). Thereafter, selection and doxycycline treatment followed the workflow above except doxycycline was included in the medium used for cell seeding, initiating the editing period that day.

PB excision
We attempted PB excision using three strategies: transfection of a plasmid encoding excision-only PB transposase, transfection of an mRNA encoding excision-only PB transposase, and genomic integration of an inducible excision-only PB transposase gene contained within a PB construct (either an all-in-one PB construct [Figure 4] or a second PB construct in a two-construct system [Figure S3], with the first construct harboring inducible *Cas9*). These experiments were performed on cells nucleofected on the same day, grown at the same time, and analyzed with FACS on the same day, allowing comparison of the different approaches without potential confounding from batching. For the dual-construct experiment, after excision, we passaged cells into replicate wells, using some wells for initial FACS analysis and remaining replicate wells for sorting three days later. FACS data suggest these additional three

safari-web-extension://C36269C9-CDBA-4674-BFED-E5D046D72A90/…com%2Fcell-reports-methods%2Fpdf%2FS2667-2375(23)00351-X.pdf

Page 16 of 19

CellPress
OPEN ACCESS

Cell Reports Methods
Article

each excision experiment follow. Plasmid-based excision. After PB nucleofection, selection, and induction of editing as described above, cells were seeded for lipofection as above. Twenty-four hours later, cells were lipofected as above with either 2000 ng pSPBase, 2000 ng excision-only PB transposase plasmid pXPBase, or 2000 ng Cas3 editing plasmid pPB758 as a negative control. Cells were exposed to transfection complexes for 24 h, then medium was refreshed. mRNA-based excision. This procedure followed the same workflow as plasmid-based excision, except lipofections included mRNA rather than plasmid DNA. Specifically, cells were lipofected with either 2000 ng excision-only PB transposase mRNA (Hera BioLabs) or CleanCap firefly luciferase mRNA (TriLink Bio-Technologies) as a negative control. Inducible excision. PB nucleofection and selection were performed as above. For the dual-construct experiment, both a pPB9 plasmid library (2500 ng, encoding indel editing guides) and a pPBX plasmid library (2500 ng, encoding CNV editing guides) were included in the transfection along with pSPBase (1000 ng). Transfected cells were selected with 0.5 μg/mL puromycin and 5 μg/mL blasticidin S HCl (Gibco) for 14 days, an extended duration used to synchronize this experiment with the time frame of the plasmid-based and mRNA-based excision experiments. After selection, cells were seeded and treated with doxycycline as above for four days, inducing both editing and PB excision. For the all-in-one construct experiment, a pPB9X plasmid library (5000 ng, encoding indel editing guides) was transfected together with pSPBase (1000 ng). Transfected cells were selected with 0.5 μg/mL puromycin (Gibco) for 7 days. After selection, cells were seeded and treated with doxycycline as above for four days, inducing both editing and PB excision.

FACS and 96-well iPSC culture

After doxycycline treatment, colonies were dissociated into single cells using accutase, resuspended in DPBS supplemented with 10 μM RI, and stained with 1 μL TO-PRO-3 viability dye (Thermo Fisher). Cells were then filtered through a 40 μm cell strainer (Corning) and sorted under sterile conditions for live (TO-PRO-3-negative), single iPSCs. For the dual-construct PB excision experiment, we first performed FACS analysis on cell populations from each excision attempt and on corresponding negative control cell populations. Three days later, we sorted cells from the doxycycline-treated cell population from the inducible excision experiment, since this strategy alone appeared successful (Figure S3B). Here, we first sorted for live, single iPSCs that were GFP-positive and then sorted for live, single iPSCs that were GFP-negative. For the all-in-one construct PB excision experiment, FACS analysis of naive, untreated, and doxycycline-treated cell populations was performed immediately before sorting live, GFP-negative, single iPSCs from the treated populations. For all FACS experiments where sorting was performed, sorters were set to deposit one cell into each well of a 96-well plate filled with 100 μL cloning medium. The sorter, plate coating material, and cloning medium used sometimes differed between experiments (Table S1). We found that the most effective combination (i.e., the setup yielding the highest average colony recovery rate) is sorting using the Bigfoot cell sorter (Invitrogen) into 96-well plates coated with rhLaminin-521 (Gibco) and filled with working medium with 10% CloneR2 cloning supplement (Stemcell Technologies). Plates filled with sorted cells were transferred to the 37°C incubator and left undisturbed for 48 h after sorting. Cloning medium was refreshed on day 2 post-FACS, and then starting on day 4, medium changes (using regular working medium) were performed every other day. On day 14, we passaged visible colonies using ReLeSR as above to consolidate them into new 96-well plates coated with Matrigel. When consolidating, cells can be passaged into replicate plates for colony propagation and DNA extraction if expansion and eventual cryopreservation of cellular models is desired.

96-Well DNA extraction

DNA extraction from clonal iPSC lines cultured in 96-well plates was performed using the Quick-DNA 96 Kit (Zymo Research). Briefly, we treated colonies with ReLeSR to detach them from the plate and resuspended them in genomic lysis buffer. Subsequent steps involving cell lysis, DNA isolation, washing, and elution were performed following the manufacturer's protocol.

WGS

WGS was performed for each starting, unedited iPSC line (two total) using a modified NEBNext Ultra II DNA Library Prep Kit for Illumina (E7645L, New England BioLabs). Briefly, 1 μg of DNA from the unedited iPSC line was fragmented to approximately 350 bp using Covaris shearing (E220, Covaris). DNA fragments were end repaired and A-tailed using Ultra II reagents. Illumina stubby-Y adapters were ligated to fragments and finished with PCR using 8-bp barcoded primers (10005921, Integrated DNA Technologies) using Ultra II reagents. Individual libraries were pooled and sequenced on NovaSeq S4 using paired-end 150 bp chemistry on the Walk-Up Sequencing Platform at the Broad Institute.

SNV calling

WGS data were quality controlled using FastQC.[72] Read pairs obtained from sequencing were aligned to GRCh37v71 with BWA-MEM v0.7.10-r789.[73] Alignments were processed using PicardTools[74] and samtools[75] to check the quality of the alignments. The Genome Analysis Toolkit (GATK)[76] v3.5 was applied for base quality score recalibration, indel realignment, duplicate removal, and SNV and indel discovery and genotyping as per published best practice protocols. GATK-called variants were filtered based on base quality, sequencing read depth, genotype quality, results of a rank-sum test for relative positioning of reference versus alternate alleles within reads, strand bias, and allelic bias.

Cell Reports Methods
Article

♂ **CellPress**
OPEN ACCESS

MIP genotyping

We designed 4,006 molecularly-tagged MIPs[43] to simultaneously genotype sequence at editing target sites, copy number across genes and genomic regions targeted for dosage alteration,[42] the presence/absence of PB transposon integrations, and identities of integrated gRNA genes. A few MIPs also targeted plasmid sequences outside the PB transposon to enable detection of any random plasmid integration if it occurred. All MIPs were designed as previously described,[42,77] except MIPs assaying copy number across unique genomic sequence were designed to target SNVs present in one or both starting iPSC lines. To genotype editing target sites, whenever possible, we used known high-performing MIPs from previous studies.[62–64] MIP pooling and phosphorylation for the PB pilot and PB SCORE experiments was performed as previously described.[42,77] Starting with the PB genedel experiment, MIPs were ordered as oPool oligo pools, resuspended in IDTE Buffer (pH 8.0), phosphorylated as previously described,[42,77] and diluted in elution buffer to their working concentration. Some MIPs (e.g., those targeting integrated gRNA sequences) were included in final MIP pools at a 50X concentration to augment capture of corresponding targets. MIP capture and sequencing were performed as previously described[42,77] using DNA extracted from clonal iPSC lines cultured in 96-well format as input. For some experiments, dual-indexing was performed by using barcoded forward primers during library preparation to enable multiplexing >384 samples in the same sequencing lane.

MIP analyses

Demultiplexing of MIP sequencing data based on sample of origin was performed by the Walk-Up Sequencing Platform at the Broad Institute, leveraging barcodes incorporated into primers used during MIP library preparation and sequenced as index (and sometime also index 2) sequencing reads. A custom program extracted MIP molecular tags from read 2 and prepared reads 1 and 2 for merging using PEAR.[78] Merged reads were mapped to a custom genome using the Burrows-Wheeler Aligner[79] (bwa 0.7.12-r1039, single-end mapping) with the BWA-MEM algorithm. This custom genome contained each CRISPR-targeted gene, each chromosomal region of interest where CNV engineering was attempted, each transfected plasmid sequence, and each gRNA construct sequence included in the plasmid library or libraries transfected. It also included positive control genes (*HYDIN* and *HYDIN2*) to provide quality control for each sample. Mapping output was parsed using custom programs to generate a final set of sequences associated with each MIP target in each sample and corresponding counts of distinct MIP molecular tags associated with each such sequence. Sequences with fewer than 10 associated molecular tags were discarded, as were sequences with allele fractions below 0.1 among those retained after the depth cutoff, as such sequences most likely represent sequencing errors. To determine which gRNA genes were integrated in each iPSC line, numbers of capture events corresponding to each gRNA expressed from a U6 promoter were counted. Minimum count thresholds were then imposed, and gRNA constructs with counts exceeding these thresholds were called as integrated. Using a similar strategy, each line was assessed for random plasmid integrations by counting capture events corresponding to plasmid sequences within and outside the PB transposon. Targets exceeding a minimum count threshold were called as genomically present. Finally, sequence analysis at CRISPR target sites using a custom program was performed to call indels and prime edits, and CNVs were inferred from altered SNV allele balance over targeted genomic intervals. CNVs were called using an automated caller based on a maximum-likelihood, graph-based approach[42] modified to accommodate identifying deletions and duplications in unique sequence rather than in duplicated genes. This program calls copy number for each haplotype across the spatial extent of sequence covered by MIPs targeting SNVs. It assumes the major allele at each MIP target corresponds to the same haplotype, because if a CNV is present, this will be the case, with the major allele present at an allele balance of 1 (within a deletion interval) or ∼0.66 (within a duplication interval). After automated calling, sequence and copy number variants were manually reviewed to ensure accuracy.

PB experiment simulations

We implemented a Markov chain[80] to simulate PB experiments with different gRNA library sizes and different desired integration profiles. Briefly, our model randomly integrates one gRNA construct at a time until each construct in the library reaches a minimum number of integrations. Running this simulation 100,000 times for each experimental scenario and averaging the numbers of integrations required yielded estimates of how many integrations are necessary to achieve prespecified outcomes pertaining to gRNA library integration. Thus, these simulated data are useful for designing new PB gRNA libraries and planning corresponding experiments, e.g., determining how many iPSC lines to isolate.

QUANTIFICATION AND STATISTICAL ANALYSIS

The exact multinomial goodness-of-fit test was performed in R (https://www.r-project.org/) using the xmonte function in the XNomial package with 1,000,000 simulations. This test compared observed numbers of isolated iPSC lines containing each gRNA gene with expected numbers under sampling from a uniform distribution. Repeatedly running this test revealed that the gRNA gene count distribution in the first indel experiment (Figure 2B) was inconsistent with uniformity ($p < 1*10^{-5}$). The k-sample Anderson-Darling test was performed in R using the ad.test function in the kSamples package with 1,000,000 simulations. This test compared distributions of numbers of distinct integrated gRNA constructs per recovered iPSC line between four transfection conditions in the pilot experiment using different amounts of gRNA plasmid library (Figure S4B). The result indicated the data were consistent with all recovered lines having originated from a single population, or in other words, that different plasmid library input levels did not lead to significant

differences in PB integration copy numbers per cell (p = 0.2 using version 2 of the Anderson-Darling test statistic, intended for discrete populations).