#### Check for updates

#### **OPEN ACCESS**

EDITED BY Haiwei Mou, Wistar Institute, United States

#### REVIEWED BY

Francisco Javier Molina-Estevez, Fundación para la Investigación Biosanitaria de Andalucía Oriental (FIBAO), Spain Yacine Cherifi, Genoway, France

· · · · **,**, · · · ·

\*CORRESPONDENCE Xiaoxia Cui, ⊠ x.cui@wustl.edu

<sup>†</sup>These authors have contributed equally to this work

RECEIVED 23 February 2025 ACCEPTED 18 April 2025 PUBLISHED 30 April 2025

#### CITATION

Lagas JS, Sentmanat MF and Cui X (2025) Efficient GBA1 editing via HDR with ssODNs by outcompeting pseudogene-mediated gene conversion upon CRISPR/Cas9 cleavage. *Front. Genome Ed.* 7:1581743. doi: 10.3389/fgeed.2025.1581743

#### COPYRIGHT

© 2025 Lagas, Sentmanat and Cui. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Efficient GBA1 editing via HDR with ssODNs by outcompeting pseudogene-mediated gene conversion upon CRISPR/ Cas9 cleavage

### Joseph S. Lagas<sup>†</sup>, Monica F. Sentmanat<sup>†</sup> and Xiaoxia Cui\*

Department of Genetics, Genome Engineering and Stem Cell Center at the McDonnel Genome Institute (GESC@MGI), School of Medicine, Washington University in St. Louis, St. Louis, MO, United States

**Introduction:** CRISPR/Cas9-edited induced pluripotent stem cells (iPSCs) are valuable research models for mechanistic studies. However, gene conversion between a gene-pseudogene pair that share high sequence identity and form direct repeats in proximity on the same chromosome can interfere with the precision of gene editing. Mutations in the human beta-glucocerebrosidase gene (GBA1) are associated with Gaucher disease, Parkinson's disease, and Lewy body dementia. During the creation of a GBA1 KO iPSC line, we detected about 70% gene conversion from its pseudogene GBAP1. These events maintained the reading frame and resulted from GBA1-specific cleavage by CRISPR/Cas9, without disrupting the GBA1 gene.

**Method:** To increase the percentage of alleles with out-of-frame indels for triggering nonsense-mediated decay of the GBA1 mRNA, we supplied the cells with two single-stranded oligodeoxynucleotide (ssODN) donors as homology-directed repair (HDR) templates.

**Results:** We demonstrate that HDR using the ssODN templates effectively competes with gene conversion and enabled biallelic KO clone isolation, whereas the nonallelic homologous recombination (NAHR)-based deletion rate remained the same.

**Discussion:** Here, we report a generalizable method to direct cellular DNA repair of double strand breaks at a target gene towards the HDR pathway using exogenous ssODN templates, allowing specific editing of one gene in a genepseudogene pair without disturbing the other.

#### KEYWORDS

gene conversion (GC), nonallelic homologous recombination (NAHR), knock out and knock in, gene editing (CRISPR/Cas9), stem cell engineering, pseudogenes

# Introduction

Modeling human disease pathology in preclinical models is an essential step in delineating disease mechanisms and the development of targeted therapies. CRISPR/ Cas9 is routinely used to introduce genetic perturbations in induced pluripotent stem cells (iPSCs) to model disease causing variants in a patient's genetic background. However, in some cases, these gene-editing strategies can be complicated by the existence of

non-coding or non-functional segments of DNA that resemble functional genes, called pseudogenes (Cheetham et al., 2020; Hanss et al., 2020). Pseudogenes predominantly arise from retrotransposition of mRNA or gene duplications and often share a high degree of homology with their functional counterparts despite having accumulated mutations (Cheetham et al., 2020). There are >14,000 pseudogenes in the human genome, many with >50% sequence homology to their protein coding counterpart (Torrents et al., 2003; Zhang and Gerstein, 2004).

Homologous recombination between a gene and its pseudogene can occur via synthesis-dependent strand annealing (SDSA) or alternate resolution of double Holliday junctions by, for example, but not limited to, the break-induced replication (BIR) pathway, whereby short (<500 bp) sequence is unidirectionally replaced with pseudogene-specific sequence, without crossover, and not leading to large structural changes (Chen et al., 2007; 2010). Gene conversion tracts created by SDSA in mammalian mitotic cells average less than 100 bp following I-SceI-mediated double strand breaks (LaRocque and Jasin, 2010). However, longer tracts have been observed in other organisms, such as 280 bp in budding yeast and 471 bp in Drosophila (Cho et al., 1998; Mansai et al., 2011). Gene conversion typically occurs between sequences with >90% identity, is positively correlated to length of sequence similarity and negatively correlated to distance between homologous templates (Schildkraut et al., 2005; Chen et al., 2007; 2010). Gene conversion leads to sequence diversity, contributing to the evolutionary trajectory of gene families like the human leukocyte antigen (HLA) loci that require high variability for environmental adaptability (Adamek et al., 2015) but also contributes to numerous genetic diseases, such as Gaucher Disease, an autosomal recessive lysosome storage disorder caused by mutations in GBA1 (Chen et al., 2007; Hruska et al., 2008).

Recombination events between GBA1 and its pseudogene, GBAP1, 16 kb downstream on the same chromosome, contribute a spectrum of disease-causing recombinant alleles at GBA1 spanning from intron 2 to exon 11 (Hruska et al., 2008). A study of 240 patients with Gaucher Disease found that 12% of alleles at GBA1 were recombinant alleles, most arising from gene conversion, nonreciprocal recombination that contributes short tracts of unidirectional sequence, or reciprocal recombination producing larger structural variation (e.g., deletion events) through nonallelic homologous recombination (NAHR) (Tayebi et al., 2003). Structural variation arising from NAHR involves a double Holliday junction that leads to crossover between high homology sequences that are nonallelic (e.g., pseudogene and parental copy), resulting a byproduct of deletions for intrachromosomal NAHR or concomitant deletions and duplications for inter-chromosomal NAHR when the recombining regions are in the same orientation (Gu et al., 2008). Recurrent copy number variation mediated by NAHR is often flanked by low-copy repeats and are associated with genetic disorders such as Prader-Willi, DiGeorge, and Charcot-Marie-Toothe disease (Dittwald et al., 2013).

GBA1 has also been implicated in Parkinson's disease and certain metabolic disorders when mutated. However, the determination of the exact disease-causing variants has been complicated by the existence of the GBA1 pseudogene, GBAP1, located 16 kb downstream in the same orientation as its functional counterpart. GBAP1 shares >96% identity to GBA1, leading to misalignments during genome assembly and transcript quantification by short read sequencing (Woo et al., 2021). Longread sequencing of transcripts has improved mapping across highly repetitive regions, unveiling pseudogene functional roles, through protein-coding and noncoding transcripts (Troskie et al., 2021; Qian et al., 2022). Recently, GBAP1 was found to generate protein with activity that is independent of the canonical GBA1 lysosomal hydrolase function and transcripts with cell type specificity (Gustavsson et al., 2024). When attempting to edit the GBA1, it is important to avoid perturbing the GBAP1 gene.

Here, we sought to generate a GBA1 KO in an iPSC line by targeting exon 6 using CRISPR/Cas9. We observed that 70% of alleles in the CRISPR RNP transfected pools had no indels, whereas indels are a signature editing outcome via nonhomologous end joining (NHEJ) at double strand breaks (DSBs). Instead, these alleles had single nucleotide variants (SNVs) near the gRNA cut site, largely matching the GBAP1 sequences but not identical, indicative of highly efficient homology-dependent repair of DSBs. We reasoned DNA repair was predominantly via gene conversion, homologous recombination using GBAP1 as template and effectively quenching NHEJ-mediated KO. We sequenced the genomic region spanning GBA1 and GBAP1 in iPS1 using a Nanopore long-read sequencing-based method, LOCK-seq (Sentmanat et al., 2024), and confirmed that SNVs in edited GBA1 alleles match those in GBAP1. We also detected NAHRmediated deletion between the GBA1 and GBAP1 loci.

To compete with gene conversion, i.e., homologous recombination using the pseudogene as template, we cotransfected Cas9/gRNA RNP with two single-stranded oligodeoxynucleotide (ssODN) donors carrying out-of-frame deletions as HDR templates. This approach resulted in >10% knock-in (KI) efficiency in the pool and a reduced gene conversion rate, ultimately enabling the successful isolation of biallelic out-of-frame clones, whereas the rate of large deletions via NAHR was not impacted. Here, we present evidence that GBAP1 serves as a preferential HDR template for repairing DSBs at GBA1 exon 6 and introduce a novel and generalizable strategy to outcompete gene conversion from highly similar pseudogenes and improve editing efficiency of the target genes.

### Materials and methods

#### Cell culture

Human iPSCs used in this study were iPS1, an unpublished line reprogrammed from hepatic fibroblasts from a healthy control, and iPS2, reprogrammed from renal epithelial cells isolated from urine (Chen et al., 2023). Both lines were cultured in mTeSR Plus (Stemcell Technologies, cat#.100-1130) on Matrigel-coated plates (Corning, United States) and passaged using ReLeSR (Stemcell Technologies, cat#.100-0483). HEK293T cells were cultured in Dulbecco's Modified Eagle's Medium (ThermoFisher, cat#.11965092) with 10% FBS (ThermoFisher, cat#.25200056). All cell lines were maintained in tissue culture incubators under conditions of  $37^{\circ}$ C, 95% air, and 5% CO<sub>2</sub> in a humidified incubator. All cultures were routinely tested for the absence of *mycoplasma* and authenticated by STR profiling. iPSCs were confirmed to have normal karyotype using G banding.

#### gRNA and donor ssODN design

gRNAs were designed using an in-house CRISPR design algorithm which combines specificity scores from the Zhang lab (Hsu et al., 2013), activity prediction scores by the Doench lab (Doench et al., 2016), as well as SNP check using dbSNP. Specificity scores are the primary factor considered in gRNA selection. The spacer sequence for the gRNA used to target exon 6 was 5'-CCATTGGTCTTGAGCCAAGT -3', and the gRNA was purchased as a one-piece synthetic molecule, Alt-R CRISPR-Cas9 sgRNA (IDT, Coralville, IA), with standard modifications. The ssODNs were designed with 60 bases of homology arms to the target site upstream and downstream of the gRNA cut site with two phosphorothioate (PTO) bonds to protect from exonuclease activity at each of the terminal 5' and 3' ends of the molecules. The sequence for the 7 bp deletion including PTO modifications denoted by asterisks 5'-c\*c\*ctgcagttggcccagcgtcccgtttcactccttgccagcccctggacatcacccact agaccaatggagcggtgaatgggaaggggtcactcaagggacagcccggagacatctaccac\*c\*a-3' and 10 bp deletion 5'-c\*c\*ctgcagttggcccagcgtcccgtttcactccttgccagcccctgg acatcacccactccaatggagcggtgaatgggaaggggtcactcaagggacagcccggagacatctacc ac\*c\*a-3', both obtained as Ultramers from IDT.

#### Nucleofections

Nucleofections were performed using the Lonza Bioscience 4D-Nucleofector P3 kit per the manufacturer's instructions. Briefly, iPSCs were single-cell dissociated with StemPro Accutase (ThermoFisher, cat.#A1110501) to collect one million cells per reaction. Cells were pelleted at  $300 \times g$  and washed once with PBS (ThermoFisher, cat.#10010023). To complex gRNA with Cas9 protein, 2 µL IDT Cas9 protein (10 ug/uL, IDT, cat.#1081058) was combined with 2  $\mu$ L of gRNA (100  $\mu$ M) and incubated at room temperature for 10-30 min. Before adding cells to the Cas9-gRNA RNP, 1 µL of ssODN (100 µM) was added to the reaction. Cells were resuspended in 100 µL P3 nucleofection solution and added to the RNP reaction. The cell suspension was transferred to a large cuvette and nucleofected using the CA137 program. Cells were replated in Matrigel-coated plates in mTeSR Plus with 10 µM ROCK inhibitor (MilliporeSigma, cat.#SCM075). ROCK inhibitor was removed after 24 h and the cells were cultured for an additional 2 days before harvesting for NGS.

#### Next-generation sequencing and analysis

Nucleofected pools of iPSCs were lysed in QuickExtract Solution (Biosearch Technologies, cat#.QE09050) following the manufacturer's instructions. The target region is then amplified by PCR in two steps using universal tails appended to genomicspecific primers. The universal tail sequences are: 5' - CACTCTTTC CCTACACGACGCTCTTCCGATCT - 3' for the forward primer and 5' - GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT -3' for the reverse primer. All primers used in this study are listed in Supplementary Table S4. These tails allow for unique indexes and Illumina P5/P7 adapter sequences to be added for the second round of PCR. The tails were attached to the genomic-specific primer sequences: 5' - CCTGATGTCTGGGGGGTTGAG - 3' forward (full sequence including tail is 5'-CACTCTTTCCCTACACGACGCTCT TCCGATCTCCTGATGTCTGGGGGGTTGAG-3') and 5' - ACA GATCAGCATGGCTAAATGG - 3' reverse (full sequence including tail is 5'- GTGACTGGAGTTCAGACGTGTGCTCTT CCGATCTACAGATCAGCATGGCTAAATGG -3'). Indexing of step 1 product was performed using 0.1X volume from step 1 with indexing primers (list of indexing primers listed in Supplementary Table S4). Products generated from step 2 PCRs were submitted to the sequencing lab at the Center for Genome Sciences and Systems Biology affiliated with Washington University in St. Louis for a 2x250 run on the MiSeq. PCRs were performed using REDTaq ReadyMix PCR Reaction Mix (MilliporeSigma, cat.#R2523-100RXN) according to the manufacturer's instructions. NGS data was analyzed using an in-house script (Connelly and Pruett-Miller, 2019). Briefly, the script uses FASTQ files from ampliconsequencing data as input and searches for short (<30 bp) userprovided sequence proximal and/or overlapping with the primerbinding sites to query reads for the presence of a list of wild-type or knock-in sequence (Supplementary Table S5 lists full data output). The output also parses the most frequently occurring reads for the region of interest, allowing for quantification of percent of total for each population.

### LOCK-seq

LOCK-seq was performed as described in Sentmanat et al. (2024). Briefly, genomic DNA was isolated from tissue and cell pellets using the Monarch Genomic DNA Purification Kit (NEB, cat.#T3010S), according to the manufacturer's instructions. Fragmentation was performed using 500 ng input with the seqWell LongPlex Long Fragmentation kit (seqWell, Beverly, MA). Custom capture probes (JSL001.GBA1-1-10, Supplementary Table S4) were from xGen Custom Hybridization Probe Panels (IDT, Coralville, IA). In-solution hybridization reactions were performed for 16-18 h s at 65°C using the xGen Hybridization and Wash Kit (IDT, cat.#1080577) with Universal Blocker NXT (IDT, cat.#1079584), according to the manufacturer's protocol. Ligation of Oxford Nanopore sequencing adapters was performed using the Ligation Sequencing Kit V14 (Oxford Nanopore Technologies, cat.#SQK\_LSK114) and loaded onto an Oxford Nanopore Technologies flow cell (FLO-MIN114).

#### **Bioinformatics analysis**

Base calling was performed using Dorado (v0.7), and samples were demultiplexed with an in-house python script. Briefly, no trimming was performed, and for the in-house demultiplexing script, exact index matches for both indexes were required. Reads with >1 match for either index or lacking one index were excluded. Capture efficiency was determined using flagstat output (total reads mapped to transgene/total reads) and length of mapped reads calculated using FASTQ files derived from transgene mapped BAM files for each sample. FASTQ files were aligned to the target region (GBAP1-GBA1 at chr1:155,213,240-155,240,028 of the hg38 reference sequence) using Minimap2 (v2.28) and Samtools (v1.20). BAM files were used to build consensus sequences using Canu (v2.2) and visualized using IGV.



FIGURE 1

GBA1 target locus. (A) GBA1 transcripts and gRNA target site (red arrow). (B) Alignment of top 10 off-target sites ordered from highest off-target score (top) to lowest for the gRNA used to target GBA1 exon 6. Score is the MIT specificity score (Hsu et al., 2013) with scores of  $\leq$  50 being poorly specific. Arrow points to GBAP1 off-target site. Dots indicate sequence homology.



Gene conversion is the predominant outcome after CRISPR/Cas9 editing at GBA1. (A) Barplot of percentage of reads with gene conversion (GC, no indels), wild-type, or NHEJ for the ± RNP edited pool and (B) alignments of the five most frequently occurring reads. \*\*P < 0.001 by Student's t-test. The data shown represents the mean  $\pm$  SD (n = 3).

# **Results**

#### CRISPR/Cas9-mediated targeting of GBA1 at exon 6 predominantly results in gene conversion from GBAP1

To disrupt the GBA1 gene, we chose a gRNA that targets a common exon to all isoforms, exon 6, with high specificity (Figures 1A, B) (Doench et al., 2016; Haeussler et al., 2016). iPS cells (line iPS1) were nucleofected with CRISPR/Cas9 RNP and sampled 72 h later. Amplicon sequencing of the transfected pool revealed that despite less than 2% of wild-type (WT) sequence remaining at the target site, instead of indels, greater than 70% of the reads had single nucleotide variants that are present in its pseudogene, GBAP1, indicating DNA repair via gene conversion (GC) (Figure 2A). NHEJ is the predominant repair pathway for CRISPR-mediated double strand breaks in the cell, by which simple KOs are commonly achieved. Alignments revealed that most of the base changes in the reads were proximal to the CRISPR/Cas9 cut site, largely matching the GBAP1 reference sequences, yet not identical (Figure 2B). The mock transfected parental cells produced only wild type reads, supporting that the base changes were not a result of template switching between GBA1 and GBAP1 loci during PCR. The CRISPR-dependent base changes suggest that gene conversion with



GBAP1 is potentially the predominant repair pathway used at this site, not the typical NHEJ that dominates CRISPR/Cas9 edited sites.

GBAP1, a GBA1 pseudogene 16 kb downstream of GBA1, shares 96% sequence identity to GBA1 and interferes with GBA1 genotyping (Woo et al., 2021; Orimo et al., 2024). To confirm that the observed SNPs in the edited pool were products of recombination with GBAP1, we first performed LOCK-seq, a target captured long-read sequencing with probes spanning the GBA1 and GBAP1 genomic region in the parental iPS1 line (Sentmanat et al., 2024). The homologous sequence to the chosen gRNA target site in GBAP1 is 21 kb from the target site in GBA1, and the iPSC line has a C>T point mutation in the GBAP1 gene, compared to the hg38 reference genome, that was also in the edited pool (Figure 3A), explaining the discrepancy we observed in Figure 2B. In total, eight SNPs were identified between GBA1 and GBAP1 in the parental iPSC line, all with allele frequencies consistent with homozygosity (≥90%, Supplementary Table S1). The primers used for amplicon sequencing of edited cells were designed to specifically amplify the GBA1 locus, generating a 411 bp amplicon for Illumina sequencing. The amplicon covers six polymorphisms that distinguish GBA1 and GBAP1: four upstream of the gRNA target site and two within the gRNA target sequence, the latter accountable for the gRNA specificity towards GBA1 (Figure 3B).

NHEJ is usually more efficient than HDR (Certo et al., 2011; Chu et al., 2015; Chien et al., 2020). However, gene conversion is effectively outcompeting NHEJ for the GBA1 gene. We reasoned that many copies of single-stranded oligodeoxynucleotide (ssODN) donors could potentially compete and reduce the gene conversion rate, enabling

more efficient KO. We designed two ssODNs, one with a -7 bp deletion and a second with a -10 bp deletion, flanked by 60 base homology arms. DNA repair through HDR with either donor will result in premature stop codons within exon 6, triggering nonsense-mediated decay of the GBA1 mRNA. In the pools transfected with CRISPR RNP alone, we observed an increasing percentage of GBAP1-specific polymorphisms approaching the gRNA cut site, with gene conversion detected as far as >50 bp upstream from the cut site (Figure 3C). Co-transfection of both ssODNs with CRISPR RNP decreased gene conversion by as much as 15% with a concomitant increase in percentage of alleles with 7 and 10 bp deletions. This data suggests that gene conversion is more efficient than NHEJ at GBA1, likely due to the proximity and length of homology between the sites. The addition of ssODNs for the desired modifications can aid in the suppression of the undesired gene conversion repair outcome.

To determine whether using ssODNs to outcompete gene conversion is generally applicable to other cell lines, including non-stem cell lines, we tested a second iPSC line (iPS2) and HEK293T cells using the same gRNA and ssODNs. iPS2 showed the same trend as iPS1 (Figure 3D). Even though the SNP at site 6 is absent in the GBAP1 gene in iPS2, the gRNA was specific and only cleaved GBA1. The GBAP1 site was intact in RNP transfected iPS2 cells (Supplementary Table S2). In both iPSC lines, the closer the SNP is to the cut site, the higher the conversion rate, indicating that the predominant repair pathway was in fact via HDR using GBAP1 as donor template, instead of deletion between the homologous repeats via nonallelic homologous recombination. Interestingly, in iPS1 we observed a drastic drop of conversion



GC and NHEJ in HEK293T. (A) Bar plot of percent GC across sites expected to undergo recombination with GBAP1. (B) Total percent NHEJ  $\pm$  ssODNs with –10 bp and –7 bp indels or unrelated ssODN without homolgy to GBA1 locus with significance in total %NHEJ (C) Alignment of top five reads in transfected pools. \*\*P < 0.001 and ns is P > 0.05 by Student's t-test. The data shown represents the mean  $\pm$  SD (n = 3).



Homology-directed gene fusion between GBAP1 and GBA1. (A) Schematic of NAHR outcomes (adapted from Gu et al., 2008). (B) Schematic of gene fusion product generated from deletion between loci. (C) Barplot of percent of total reads positive for GBA1-specific C>T variant indicative of deletion product (D) Alignment of perfect deletion product detected in CRISPR RNP  $\pm$  ssODN pools of iPS1. ns is P > 0.05 by Student's t-test. The data shown represents the mean  $\pm$  SD (n = 3).

rate between sites 5 and 6, implying the repair process is most efficient within this 12 bp window. iPS2 has no mismatch at site 6 between the pair of loci and cannot be assessed similarly.

The HEK293T cells also lack the mismatch at site 6 and are heterozygous for a 1 bp deletion at Site 4 (Figure 3B). Compared to the iPSCs, the pool of HEK293T cells transfected with CRISPR RNP had a much higher NHEJ rate (over 50%) and a much lower percentage of gene conversion, 7% (vs. over 70% in iPSCs) (Figure 4A). We did not

observe a reduction in conversion alleles when ssODNs were included despite 20% HDR-mediated -7 and -10 alleles (Figure 4B). Alignment of the top five reads in CRISPR RNP  $\pm$  ssONDs supports that indel-containing reads lack nearby GBAP1 variants (Figure 4C).

The resolution of inter- or intra-chromosomal and intrachromatid NAHR results in deletions when the two loci are in the same orientation on a chromosome arm (Figure 5A). To assess whether transfected iPSC pools are undergoing a homology-

directed deletion event that results in a fusion of the GBAP1 and GBA1 loci with deletion of the intervening genomic region, we performed NGS on PCR amplicons targeting the deletion junction. The primers were designed to simultaneously amplify both the deletion product as well as the GBAP1 locus, producing amplicons of the same size but differ at a single T/C SNV (chr1:155,238,392, hg38) (Figure 5B). This allowed us to quantify the percentage of reads of the deletion product and unmodified GBAP1 locus, respectively. The deletion reads contain a GBA1-specific variant 148 bp away from cut site (Supplementary Figures S1, S5B). A single deletion product was detected in 20% of total reads in iPS1 pools with and without cotransfection of ssODNs and was absent in the untransfected parental line (Figure 5C). The same deletion allele was present in all replicates transfected with RNPs, representing an abundant repair product besides those via gene conversion (Figure 5D). We further confirmed the presence of deletion products using locus-specific primers that bind to either GBAP1- and GBA1-specific variants present in HEK293T cells (Supplementary Figures S2A, B). iPS2 lacks the unique GBA1-specific variant within the genotyping window suitable for NGS and could not be assessed the same way. Only the perfect deletion junction was present with the expected variants identified in iPS1 and HEK293T cellswith or without cotransfection of ssODNs (Supplementary Figure S2C). Additionally, the HEK293T cells have two deletion alleles, sharing identical junction but containing two heterozygous SNVs, one within GBAP1 and the other, the GBA1-specific SNV 148 bp from the cut site (Supplementary Figures S3A, B). The absence of indels both at the breakpoint of the single deletion product detected and at the GBAP1 locus in RNP transfected pools implies the deletion more likely resulted from NAHR than from NHEJ-mediated fusion following the gRNA cutting of both GBA1 and GBAP1 (see Supplementary Table S2), although we cannot completely exclude the possibility of the latter. NHEJ-mediated deletions are usually accompanied by indels at the deletion junction (Canver et al., 2014).

#### Introduction of ssODNs containing out-offrame indels for GBA1 enabled successful isolation of GBA1 KO clones

By co-delivery of ssODNs carrying out-of-frame deletions with RNPs, we observed that the indels introduced by the ssODNs were among the most abundant GBA1 alleles in the transfected pools in both iPSCs (Figure 6A). We screened 126 single-cell derived iPS1 clones and identified 22 KO and three wild-type isogenic clones (Figure 6B; Supplementary Table S3). Across clones with alleles from gene conversion, SNV conversion rate correlates to its distance from the cut site. Many clones only possess the GBAP1specific polymorphisms closest to the cut site for at least one allele, in agreement with the data on the edited pool (Figure 3C). As predicted from the edited pool data, 36% (8/22) of KO clones had one or both donor-specified KO alleles, two of which were biallelic clones containing one of each donor sequence (Supplementary Table S3). Indeed, one advantage of using two donors is that loss-ofheterozygosity at the target site, as reported to occur in up to a third of edited clones, can be avoided by selecting clones with distinct indels (Kosicki et al., 2018; Boutin et al., 2021).

### Discussion

Mutations in GBA1 is one of the most common genetic risk factors for Parkinson's disease and Gaucher's disease. Some GBA1 mutations are predicted to be loss-of-function leading to reduced protein levels of beta-glucocerebrosidase, whereas others are suspected of being gain-of-function and require a different therapeutic approach (Smith et al., 2022; Huh et al., 2023). Modeling such mutations in iPSCs is a scalable means to survey the phenotypic outcome across mutations. The generation of GBA1 KO iPSC lines as a disease model was hindered by the existence of GBAP1, a GBA1 pseudogene, which shares 96% identity to the coding sequences of the GBA1 gene. Previous literature has described challenges to CRISPR gene editing of targets with pseudogenes due to limited regions with uniquely targetable sequences and undesired gene conversion or NAHR events (Klatt et al., 2019; Javidi-Parsijani et al., 2020; Wrona et al., 2020; Shaw and Estus, 2021; Yanovsky-Dagan et al., 2022).

Here we showed that double strand breaks at GBA1 by CRISPR/ Cas9 were repaired predominantly by base changes instead of indels. Using LOCK-seq, we identified eight polymorphisms across GBA1 and GBAP1 that are unique to this iPSC line, emphasizing the need to carefully sequence the parental line across this highly variable region. The base changes observed match perfectly to sequences of GBAP1, a GBA1 pseudogene 16 kb downstream on the same chromosome. Bases closer to the cut site were converted more frequently than those further away (Figure 3), suggesting high-efficiency gene conversion utilizing GBAP1 as a repair template. In addition, we also detected repair products via the NAHR pathway, which results in a deletion between the direct repeats with the fusion junction exactly at the cut sites.

To generate a GBA1 KO model, we designed two ssODNs containing out-of-frame indels to serve as HDR template and conucleofected the cells with CRISPR RNP to outcompete GBAP1mediated gene conversion. The inclusion of these ssODNs increased the indel percentage in the pool by ~17% (Figure 3C), which proved to be sufficient to produce multiple GBA1 KO clones (Figure 6B; Supplementary Table S3). We also show that both donors can serve as templates for HDR in the same cell, assisting with the recovery of KOs with distinct alleles that ensure copy number is intact. We outline here a novel strategy for achieving desired CRISPR/ Cas9 gene editing complicated by high frequency pseudogenemediated gene conversion events.

The interference we observed is unlikely to be limited to GBA1 and GBAP1 genes but potentially applicable to other genepseudogene pairs that share high sequence identity and are in proximity to each other on the same chromosome. Thus far, no methods have been reported for preventing or lowering the efficiency of pseudogene-mediated conversion events to ensure only the target gene is edited as desired. To generate a GBA1 KO *in vitro* iPSC model we utilized a novel approach involving the use of ssODNs to outcompete the GBAP1 as HDR template.

The approach described herein is technically simple and costeffective. ssODNs carrying the desired mutations flanked by two 60base homology arms can be designed and ordered from various vendors quickly and included in nucleofection reactions along with Cas9-RNP. The inclusion of ssODNs reduced the pseudogenemediated gene conversion events by ~10% at the GBA1 locus,



allowing ready isolation of single-cell clones with two out-offrame alleles.

To address if the gene conversion observed was generalizable to other cell lines, we next tested competition between ssODNmediated HDR and gene conversion in a second iPSC line and HEK293T cells, an SV40 transformed clone of the human embryonic kidney 293 cell line. The second iPSC line produced similar results as in iPS1 line, however, we did not observe significant gene conversion rate in HEK293T cells. Further studies are needed to demonstrate whether this is common in transformed cells or specific for the GBA1/GBAP1 pair in HEK293T cells.

Our report highlights the complexity of targeting genes with pseudogene counterparts for CRISPR-Cas9 gene editing. DSBs in GBA1 by CRISPR-Cas9 resulted in a high frequency of GBAP1 pseudogene-mediated gene conversion events that can be circumvented by the inclusion of ssODNs specifying the mutation of interest, out-of-frame indels in this case, as competing templates, for the successful creation of GBA1 KO. Here we have shown that the novel approach of utilizing ssODN donors is a simple and straightforward method to overcome the interference of pseudogenemediated gene conversion events during CRISPR-Cas9 gene editing and that LOCK-seq is a useful targeted sequencing method to verify precise editing at the gene of interest even in the presence of highly homologous pseudogene nearby.

### Data availability statement

The data presented in this study are included in the article/ Supplementary Materials. Further inquiries can be directed to the corresponding author.

### Ethics statement

Ethical approval was not required for the studies on humans in accordance with the local legislation and institutional requirements because only commercially available established cell lines were used.

# Author contributions

JL: Data curation, Formal Analysis, Investigation, Writing – original draft. MS: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Supervision, Visualization, Writing – review and editing. XC: Funding acquisition, Methodology, Resources, Supervision, Writing – review and editing.

# Funding

The author(s) declare that financial support was received for the research and/or publication of this article. MS and XC are partially supported by the NCI cancer center supporting grant P30 CA091842-03.

# Acknowledgments

We thank Wan Ching Chan and Rusudan Turabelidze for technical assistance. and Zi Teng Wang for technical assistance and helpful discussion. Jessica Hoisington-Lopez at DNA Sequencing Innovation Lab at The Edison Family Center for Genome Sciences and Systems Biology, Washington University in St. Louis for amplicon NGS analysis. Figures created with BioRender.com.

# **Conflict of interest**

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

# **Generative AI statement**

The author(s) declare that no Generative AI was used in the creation of this manuscript.

# References

Adamek, M., Klages, C., Bauer, M., Kudlek, E., Drechsler, A., Leuser, B., et al. (2015). Seven novel HLA alleles reflect different mechanisms involved in the evolution of HLA diversity: description of the new alleles and review of the literature. *Hum. Immunol.* 76, 30–35. doi:10.1016/j.humimm.2014.12.007

Boutin, J., Rosier, J., Cappellen, D., Prat, F., Toutain, J., Pennamen, P., et al. (2021). CRISPR-Cas9 globin editing can induce megabase-scale copy-neutral losses of heterozygosity in hematopoietic cells. *Nat. Commun.* 12, 4922. doi:10.1038/s41467-021-25190-6

Canver, M. C., Bauer, D. E., Dass, A., Yien, Y. Y., Chung, J., Masuda, T., et al. (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. doi:10.1074/jbc.m114.564625

Certo, M. T., Ryu, B. Y., Annis, J. E., Garibov, M., Jarjour, J., Rawlings, D. J., et al. (2011). Tracking genome engineering outcome at individual DNA breakpoints. *Nat. Methods* 8, 671–676. doi:10.1038/nmeth.1648

Cheetham, S. W., Faulkner, G. J., and Dinger, M. E. (2020). Overcoming challenges and dogmas to understand the functions of pseudogenes. *Nat. Rev. Genet.* 21, 191–201. doi:10.1038/s41576-019-0196-1

Chen, J.-M., Cooper, D. N., Chuzhanova, N., Férec, C., and Patrinos, G. P. (2007). Gene conversion: mechanisms, evolution and human disease. *Nat. Rev. Genet.* 8, 762–775. doi:10.1038/nrg2193

Chen, J.-M., Cooper, D. N., Férec, C., Kehrer-Sawatzki, H., and Patrinos, G. P. (2010). Genomic rearrangements in inherited disease and cancer. *Seminars Cancer Biol.* 20, 222–233. doi:10.1016/j.semcancer.2010.05.007

Chen, Y.-H., Connelly, J. P., Florian, C., Cui, X., and Pruett-Miller, S. M. (2023). Short tandem repeat profiling via next generation sequencing for cell line authentication. *Dis. Model. Mech.* 16, dmm050150. doi:10.1242/dmm.050150

Chien, J. C.-Y., Tabet, E., Pinkham, K., da Hora, C. C., Chang, J. C.-Y., Lin, S., et al. (2020). A multiplexed bioluminescent reporter for sensitive and non-invasive tracking of DNA double strand break repair dynamics *in vitro* and *in vivo*. *Nucleic Acids Res.* 48, e100. doi:10.1093/nar/gkaa669

Cho, J. W., Khalsa, G. J., and Nickoloff, J. A. (1998). Gene-conversion tract directionality is influenced by the chromosome environment. *Curr. Genet.* 34, 269–279. doi:10.1007/s002940050396

Chu, V. T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K., et al. (2015). Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat. Biotechnol.* 33, 543–548. doi:10.1038/nbt.3198

Connelly, J. P., and Pruett-Miller, S. M. (2019). CRIS.py: a versatile and high-throughput analysis program for CRISPR-based genome editing. *Sci. Rep.* 9, 4194. doi:10.1038/s41598-019-40896-w

Dittwald, P., Gambin, T., Szafranski, P., Li, J., Amato, S., Divon, M. Y., et al. (2013). NAHR-mediated copy-number variants in a clinical population: mechanistic insights into both genomic disorders and Mendelizing traits. *Genome Res.* 23, 1395–1409. doi:10. 1101/gr.152454.112

Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* 34, 184–191. doi:10.1038/nbt.3437

Gu, W., Zhang, F., and Lupski, J. R. (2008). Mechanisms for human genomic rearrangements. *PathoGenetics* 1, 4. doi:10.1186/1755-8417-1-4

# Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

# Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgeed.2025.1581743/ full#supplementary-material

Gustavsson, E. K., Sethi, S., Gao, Y., Brenton, J. W., García-Ruiz, S., Zhang, D., et al. (2024). The annotation of GBA1 has been concealed by its protein-coding pseudogene GBAP1. *Sci. Adv.* 10, eadk1296. doi:10.1126/sciadv.adk1296

Haeussler, M., Schönig, K., Eckert, H., Eschstruth, A., Mianné, J., Renaud, J.-B., et al. (2016). Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol.* 17, 148. doi:10.1186/s13059-016-1012-2

Hanss, Z., Boussaad, I., Jarazo, J., Schwamborn, J. C., and Krüger, R. (2020). Quality control strategy for CRISPR-Cas9-based gene editing complicated by a pseudogene. *Front. Genet.* 10, 1297. doi:10.3389/fgene.2019.01297

Hruska, K. S., LaMarca, M. E., Scott, C. R., and Sidransky, E. (2008). Gaucher disease: mutation and polymorphism spectrum in the glucocerebrosidase gene (GBA). *Hum. Mutat.* 29, 567-583. doi:10.1002/humu.20676

Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* 31, 827–832. doi:10.1038/nbt.2647

Huh, Y. E., Usnich, T., Scherzer, C. R., Klein, C., and Chung, S. J. (2023). GBA1 variants and Parkinson's disease: paving the way for targeted therapy. *J. Mov. Disord.* 16, 261–278. doi:10.14802/jmd.23023

Javidi-Parsijani, P., Lyu, P., Makani, V., Sarhan, W. M., Yoo, K. W., El-Korashi, L., et al. (2020). CRISPR/Cas9 increases mitotic gene conversion in human cells. *Gene Ther.* 27, 281–296. doi:10.1038/s41434-020-0126-z

Klatt, D., Cheng, E., Philipp, F., Selich, A., Dahlke, J., Schmidt, R. E., et al. (2019). Targeted repair of p47-CGD in iPSCs by CRISPR/Cas9: functional correction without cleavage in the highly homologous pseudogenes. *Stem Cell Rep.* 13, 590–598. doi:10. 1016/j.stemcr.2019.08.008

Kosicki, M., Tomberg, K., and Bradley, A. (2018). Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* 36, 765–771. doi:10.1038/nbt.4192

LaRocque, J. R., and Jasin, M. (2010). Mechanisms of recombination between diverged sequences in wild-type and BLM-deficient mouse and human cells. *Mol. Cell. Biol.* 30, 1887–1897. doi:10.1128/mcb.01553-09

Mansai, S. P., Kado, T., and Innan, H. (2011). The rate and tract length of gene conversion between duplicated genes. *Genes* 2, 313–331. doi:10.3390/genes2020313

Orimo, K., Mitsui, J., Matsukawa, T., Tanaka, M., Nomoto, J., Ishiura, H., et al. (2024). Association study of GBA1 variants with MSA based on comprehensive sequence analysis -Pitfalls in short-read sequence analysis depending on the human reference genome-. J. Hum. Genet. 69, 613–621. doi:10.1038/s10038-024-01266-1

Qian, S. H., Chen, L., Xiong, Y.-L., and Chen, Z.-X. (2022). Evolution and function of developmentally dynamic pseudogenes in mammals. *Genome Biol.* 23, 235. doi:10.1186/s13059-022-02802-y

Schildkraut, E., Miller, C. A., and Nickoloff, J. A. (2005). Gene conversion and deletion frequencies during double-strand break repair in human cells are controlled by the distance between direct repeats. *Nucleic Acids Res.* 33, 1574–1580. doi:10.1093/nar/gki295

Sentmanat, M. F., Wang, Z. T., Kouranova, E., Peters, S., Chan, W. C., Lin, J., et al. (2024). Efficient creation and characterization of multi-kilobase knockins by CRISPR/ Cas9 and multiple rAAVs. *bioRxiv* 2024. doi:10.1101/2024.11.26.625457 Shaw, B. C., and Estus, S. (2021). Pseudogene-mediated gene conversion after CRISPR-Cas9 editing demonstrated by partial CD33 conversion with SIGLEC22P. CRISPR J. 4, 699–709. doi:10.1089/crispr.2021.0052

Smith, L. J., Lee, C.-Y., Menozzi, E., and Schapira, A. H. V. (2022). Genetic variations in GBA1 and LRRK2 genes: biochemical and clinical consequences in Parkinson disease. *Front. Neurol.* 13, 971252. doi:10.3389/fneur.2022.971252

Tayebi, N., Stubblefield, B. K., Park, J. K., Orvisky, E., Walker, J. M., LaMarca, M. E., et al. (2003). Reciprocal and nonreciprocal recombination at the glucocerebrosidase gene region: implications for complexity in gaucher disease. *Am. J. Hum. Genet.* 72, 519–534. doi:10.1086/367850

Torrents, D., Suyama, M., Zdobnov, E., and Bork, P. (2003). A genome-wide survey of human pseudogenes. *Genome Res.* 13, 2559–2567. doi:10.1101/gr.1455503

Troskie, R.-L., Jafrani, Y., Mercer, T. R., Ewing, A. D., Faulkner, G. J., and Cheetham, S. W. (2021). Long-read cDNA sequencing identifies functional

pseudogenes in the human transcriptome. Genome Biol. 22, 146. doi:10.1186/ s13059-021-02369-0

Woo, E. G., Tayebi, N., and Sidransky, E. (2021). Next-generation sequencing analysis of GBA1: the challenge of detecting complex recombinant alleles. *Front. Genet.* 12, 684067. doi:10.3389/fgene.2021.684067

Wrona, D., Pastukhov, O., Pritchard, R. S., Raimondi, F., Tchinda, J., Jinek, M., et al. (2020). CRISPR-directed therapeutic correction at the NCF1 locus is challenged by frequent incidence of chromosomal deletions. *Mol. Ther.* 17, 936–943. doi:10.1016/j. omtm.2020.04.015

Yanovsky-Dagan, S., Frumkin, A., Lupski, J. R., and Harel, T. (2022). CRISPR/Cas9induced gene conversion between *ATAD3* paralogs. *Hum. Genet.* 3, 100092. doi:10. 1016/j.xhgg.2022.100092

Zhang, Z., and Gerstein, M. (2004). Large-scale analysis of pseudogenes in the human genome. *Curr. Opin. Genet. Dev.* 14, 328–335. doi:10.1016/j.gde.2004.06.003