



Research article

Combining Off-flow, a Nextflow-coded program, and whole genome sequencing reveals unintended genetic variation in CRISPR/Cas-edited iPSCs



Carole Shum^{a,b}, Sang Yeon Han^a, Bhooma Thiruvahindrapuram^a, Zhuozhi Wang^a, Jill de Rijke^a, Benjamin Zhang^{a,b}, Maria Sundberg^c, Cidi Chen^d, Elizabeth D. Buttermore^d, Nina Makhortova^d, Jennifer Howe^{a,b}, Mustafa Sahin^{c,e}, Stephen W. Scherer^{a,b,f,g,*}, 1,2,3

^a The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON M5G 0A4, Canada

^b Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, ON M5G 0A4, Canada

^c Department of Neurology, FM Kirby Neurobiology Center, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA

^d Human Neuron Core, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA

^e Rosamund Stone Zander Translational Neuroscience Center, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA

^f Department of Molecular Genetics and McLaughlin Centre, University of Toronto, Toronto, ON M5S 1A8, Canada

^g Lead contact

ARTICLE INFO

Keywords:

Induced pluripotent stem cells (iPSCs)
Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)
Genome editing
Whole genome sequencing (WGS)
Off-target detection

ABSTRACT

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-Cas nucleases and human induced pluripotent stem cell (iPSC) technology can reveal deep insight into the genetic and molecular bases of human biology and disease. Undesired editing outcomes, both on-target (at the edited locus) and off-target (at other genomic loci) hinder the application of CRISPR-Cas nucleases. We developed Off-flow, a Nextflow-coded bioinformatic workflow that takes a specific guide sequence and Cas protein input to call four separate off-target prediction programs (CHOPCHOP, Cas-Offfinder, CRISPRitz, CRISPR-Offfinder) to output a comprehensive list of predicted off-target sites. We applied it to whole genome sequencing (WGS) data to investigate the occurrence of unintended effects in human iPSCs that underwent repair or insertion of disease-related variants by homology-directed repair. Off-flow identified a 3-base-pair-substitution and a mono-allelic genomic deletion at the target loci, *KCNQ2*, in 2 clones. Unbiased WGS analysis further identified off-target missense variants and a mono-allelic genomic deletion at the targeted locus, *GNAQ*, in 10 clones. On-target substitution and deletions had escaped standard PCR and Sanger sequencing analysis, while missense variants at other genomic loci were not detected by Off-flow. We used these results to filter out iPSC clones for subsequent functional experiments. Off-flow, which we make publicly available, works for human and mouse genomes currently and can be adapted for other genomes. Off-flow and WGS analysis can improve the integrity of studies using CRISPR/Cas-edited cells and animal models.

1. Introduction

CRISPR-Cas and human iPSC technology has tremendous potential to augment our understanding of human genetics and disease. The reprogramming of human somatic cells to iPSCs [1] revolutionized

human stem cell biology research. Just over a decade ago, studies first demonstrated that Cas9 proteins could be loaded with single RNA molecule to cleave DNA targets in human cells [2,3]. Since this time, the combination of engineered Cas proteins and a short sequence of homologous RNA, or guide RNA (gRNA), has been used to target

* Corresponding author at: The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON M5G 9A4, Canada.

E-mail address: stephen.scherer@sickkids.ca (S.W. Scherer).

¹ Genetics and Genome Biology Program, The Hospital for Sick Children, Toronto, ON M5G 9A4, Canada

² Department of Molecular Genetics, University of Toronto, Toronto, ON M5S 1A8, Canada

³ McLaughlin Centre, Toronto, ON M5G 0A4, Canada

previously inaccessible genomic loci [4–8]. Proof-of-concept studies have shown that the combination of the two technologies can enable novel discoveries about the impact and function of genetic variants [9–13].

An important step in the use of CRISPR/Cas technology is the quantification of target modification specificity. In silico tools, in vitro and in vivo experimental techniques have been developed and used for predicting and detecting genome-wide CRISPR/Cas off-target profiles [14–21], but there is currently no standard for assessing the unintended effects of CRISPR/Cas-editing in iPSCs. A study that systematically benchmarked and integrated in silico tools to develop a platform for genome-wide CRISPR off-target cleavage site prediction found that CRISPR cleavage specificity is heterogeneous in different cell types [22]. Yet, most studies use one in silico CRISPR off-target prediction tool and Sanger sequencing to assess only the top predicted CRISPR off-target sites in exons or intron-exon junctions. At most, some studies use one in silico CRISPR off-target prediction tool and whole genome sequencing (WGS) or RNA-sequencing to intersect with predicted off-target sites to assess CRISPR off-target effects [23,24]. Currently available in silico CRISPR off-target prediction tools may be limited in the number of mismatches considered and the inclusion of “bulge”-type mismatches [15]. In addition, in silico prediction tools may miss other unintended genetic variants that arise from the CRISPR/Cas-editing process. Thus, relying solely on one prediction tool may result in a less than comprehensive list of genome-wide predicted off-target sites.

The aim of this study is to determine the rate of unintended effects induced by CRISPR/Cas-editing in iPSCs and to formulate a standard methodology for assessing these effects at a genome-wide level. Sixteen iPSC clones were assessed after homology directed repair (HDR) of seven separate regions in three disease-associated genes, *KCNQ2*, *ASH1L* and *GNAQ*, by CRISPR-Cas9 or CRISPR-Cas12a. *KCNQ2* and *ASH1L* are linked to epilepsy and autism spectrum disorder [25–28]. *GNAQ* is associated with Sturge-Weber syndrome [29]. In six regions, CRISPR-Cas9/Cas12a was delivered as a ribonucleoprotein (RNP) complex of Cas protein and gRNA to repair a disease-associated variant in participant-derived iPSC lines. In the remaining region, CRISPR-Cas9 and gRNA were delivered in plasmids to introduce a disease-associated variant in a control iPSC line. All iPSC clones had normal karyotypes, assessed via g-banded karyotyping, prior to DNA extraction, WGS and variant detection.

To assess the specificity of target modification, we developed Off-flow, a Nextflow-coded bioinformatic workflow program [30]. Off-flow takes a specific Cas protein, gRNA and protospacer adjacent motif (PAM) input to call four separate off-target prediction programs. It uses their output to establish a comprehensive list of in silico predicted off-target sites. Using this list as a filter, WGS data from CRISPR/Cas-targeted iPSC clones are then cross-referenced with Off-flow’s output to determine which de novo variants were likely caused by off-target effects. In parallel, we searched for sequences matching (up to four mismatches) gRNA+ and gRNA-, separately, within a 200-base-pair (bp) window flanking both sides of the unique single nucleotide variant (SNV), and insertion and deletions (indels) from WGS data for each iPSC clone.

Finally, we analyzed all cell-line specific variants in isogenic iPSC clones irrespective of off-target prediction to determine unintended genetic variation from the CRISPR/Cas-editing process. The integrity and editing of the targeted gene and any unintended variants that were found were validated by PCR, agarose gel electrophoresis and Sanger sequencing.

We found a 3-bp substitution and mono-allelic deletions at the target loci, *KCNQ2* and *GNAQ*, in 3 CRISPR/Cas9-edited clones. We also identified 21 editing-induced missense variants at other genomic loci in 7 edited clones and 4 unedited clones, or clones that failed HDR. We did not find any significant differences in the rate of unintended effects in CRISPR/Cas9-edited iPSC clones compared to CRISPR/Cas12a-edited iPSC clones. We observed that unintended on- and off-target variants

are more frequent following delivery of the CRISPR-Cas system by plasmid compared to ribonucleoprotein delivery.

2. Methods

2.1. Generation of isogenic iPSCs

iPSC lines were generated from erythroid progenitors using CytoTune™ iPSC 2.0 Sendai Reprogramming Kit (Thermo Fisher Scientific) according to manufacturer’s protocol. Three iPSC lines were generated from participants with *KCNQ2* variants, three iPSC lines were generated from participants with *ASH1L* variants, and one iPSC line was generated from a control. 48 h after viral delivery, fresh StemSpan SFEM II media with StemSpan Erythroid Expansion Supplement (STEMCELL Technologies) was added to the cells and incubated for 24 h. Cells were then transitioned to ReproTeSR medium (STEMCELL Technologies) for the duration of reprogramming. Once colonies were of an adequate size and morphology, individual colonies were picked and plated onto Geltrex (Thermo Fisher Scientific) and mTeSR1 medium (STEMCELL Technologies). Clones were further expanded and characterized using standard assays for pluripotency, karyotyping (The Center for Applied Genomics (TCAG), SickKids) and mycoplasma.

HDR was performed in participant-derived iPSC lines using ribonucleoprotein complexes (RNPs) of CRISPR/Cas9 or CRISPR/Cas12a and an ssODN donor template to target variants in *KCNQ2* and *ASH1L* (Table 1). CRISPR-Cas9 or CRISPR-Cas12a were selected based on their ability to target the genetic loci of interest. Guide RNA sequences were devised using Benchling and CRISPick. At least four guide RNAs with high predicted efficacy and specificity scores to target within 25 bp of each genetic locus were required. Four of the highest off-target and on-target scored gRNAs were selected. Alt-R® modified Cas9 sgRNAs or Cas12a crRNAs, Alt-R® S.P. HiFi Cas9 Nuclease V3 or Alt-R® A.S. Cas12a (Cpf1) Ultra and Alt-R® HDR Donor Oligos were obtained from Integrated DNA Technologies (IDT). Prior to nucleofection, iPSCs were treated with 10 μM Y-27632 for at least 60 min at 37 °C. 81ul of 10 μM sgRNA or crRNA was incubated with 61 μM Cas9 or 63 μM Cas12a enzyme in 100 μl Human Stem Cell Nucleofection Solution 1 and Supplement (Lonza) for 10 min or Resuspension buffer R (Neon) for 20 min at room temperature to form Cas:gRNA RNP complexes. 22 μl of 10 μM ssODN was added to the complexes and nucleofected into 1×10^6 iPSCs using Nucleofector™ 2b (Lonza, program A-023) or the Neon Transfection System (Thermo Fisher Scientific, 1400 V, 20 ms, 1 pulse). Following nucleofection, iPSCs were maintained in StemFlex, 27 μM HDR enhancer V2 (IDT) and 10 μM Y-27632. Digital droplet PCR was performed on half of the pool of iPSCs from one well of a 6-well plate, approximately 1–2 million cells, to detect HDR. Editing efficiency ranged from 6.4–15.4%. Single cell clones were manually isolated and expanded. PCR and Sanger sequencing was performed to confirm HDR in clonal edited lines.

Insertion of disease-related variants was performed by the UConn Human Genome Editing Core. Briefly, SPY-Cas9 was used to target and cut an intron of *GNAQ* and repaired with the targeting vector including the neo cassette and loxp sites to introduce heterozygous and homozygous single nucleotide variants. The Cre recombinase was later introduced to recombine the loxp sites that flank the selection cassette to remove it.

2.2. Whole genome sequencing and variant calling

DNA extracted from frozen iPSC pellets were submitted to TCAG for genomic library preparation and WGS. DNA was quantified and analyzed using Qubit High Sensitivity Assay and Nanodrop OD260/280. 700 ng of DNA was used as input for library preparation using Illumina TruSeq PCR-free DNA Library Prep. Each validated library was sequenced on two lanes of a high throughput V4 flow cell on a NovaSeq 6000 platform following Illumina’s recommended protocol to generate

Table 1
List of CRISPR/Cas engineered iPSC lines.

Gene	Locus (hg19)	Method	Position of target site (hg19)	Cell line ID
KCNQ2	Chr20:62071001-62071003 (c.875_877delTCCinsCCT, p.L292_L293delinsPF)	Cas9-V3-Hifi RNP	Chr20:62,071,008-62,071,027	HNDS0068-01 #B CNC14 HNDS0068-01 #B CC3
KCNQ2	Chr20:62073808 (c.766 G>T, p.G256W)	Cas9-V3-Hifi RNP	Chr20:62,073,830-62,073,848	HNDS0078-01 #D CNC2 HNDS0078-01 #D CC8 HNDS0078-01 #D CC18
KCNQ2	chr20:62071057 (c.821 C>T, p.T274M)	Cas9-V3-Hifi RNP	Chr20:62,071,037-62,071,056	HNDS0072-01 #C CNC87 HNDS0072-01 #C CC80 HNDS0072-01 #C CC20
ASH1L	Chr1: 155451888 (c.773Gdel; p.G258fs)	Cas12a Ultra RNP	Chr1:155451881-155451901	1-1134-003_CNC5 1-1134-003_CC10
ASH1L	Chr1: 155447758 (c.4902_4903TTdel; p.S1635fs)	Cas9-V3-Hifi RNP	Chr1:155,447,747-155,447,768	1-1217-003_CNC36 1-1217-003_CC37
ASH1L	Chr1: 155450703 (c.1958dup; p.P654fs)	Cas12a Ultra RNP	Chr1:155,450,703-155,450,723	1-1006-003_CNC35 1-1006-003_CC9
GNAQ	chr9:80412493 (c.548 G>A; p.R183Q)	Cas9 plasmid	chr9:80,412,499-80,412,518	31 CC-het 31 CC-hom

Table 2
List of in silico CRISPR off-target prediction tools and parameters used.

Tool	Maximum mismatch number	Maximum DNA bulge size
CHOPCHOP	3	n/a
Cas-OffFinder	5	2
CRISPRitz	5	2
CRISPR-offfinder	5	n/a

pair-end reads of 150-bases in length. Filtered reads were mapped to the reference genome (build GRCh37) using the Burrows-Wheeler Aligner (BWA) algorithm (0.7.15). GATK (GenotypeConcordance) [31] was used to identify indel and SNV calls unique to the isogenic line genomes, as previously described [10]. MuTect2 (v4.1.9.0) was used to identify SNVs and indels unique to the isogenic line genomes with respect to the original participant-derived iPSC line genomes.

2.3. Detection of SNVs and indels

Small variants were annotated, filtered, and detected as described previously [32]. We considered a variant to be a potential unintended variant when there was a heterozygous alternative genotype in the CRISPR-derived clone and homozygous reference genotype in the parental cell line. SNVs that did not pass the Mutect2 filter, of unknown zygosity, below read depth of 10x, below an allele fraction of 0.2 (as computed by Mutect2) were further excluded. We validated all unintended SNVs from all clones by Sanger sequencing.

2.4. CNV and SV analysis

We detected CNVs for each sample using two algorithms, CNVnator [33] and ERDS [34], as previously described [32,35]. Algorithms were run using their default parameters. We retained CNVs with size > 1 kb. We also performed a manual inspection on the quality of CNVs by inspecting reads from the BAM for confirmation. We defined unintended and de novo CNVs as those not observed in the parental cell line and resulting in chromosome abnormalities, large rare CNVs between 3 and 10 Mb in size and CNVs impacting coding exons.

2.5. Off-flow: Genome-wide detection of CRISPR off-target genetic variants

We used Nextflow to develop a bioinformatic workflow program, Off-flow, to automate the process of CRISPR/Cas off-target detection. Four available genome-wide CRISPR off-target cleavage site in silico prediction tools, CHOPCHOP [19], Cas-Offfinder [14], CRISPRitz [20], CRISPR-Offfinder [36] were selected for use in this study (Table 1). These

tools were selected based on a previously published comprehensive comparison and assessment of CRISPR off-target cleavage site algorithms [22] and their command-line availability. Taking a specific Cas protein, guide RNA, PAM input, and a specific number of mismatches and bulges to consider in silico off-target prediction challenges, Off-flow is an ensemble-based approach that simultaneously runs and aggregates results from the four off-target prediction tools to establish a comprehensive list of *in-silico* predicted off-target effects. Using this list as a filter, genetic variants called by Mutect2, within a 200-bp window flanking both sides of each unique variant call, are then cross-referenced with Off-flow's output to determine which genetic variants were likely caused by off-target effects (Fig. 1).

2.6. BWA string search for detection of CRISPR off-target genetic variants

In parallel, BWA [37] string search was performed for sequences matching (up to four mismatches) gRNA+ and gRNA-; for each iPSC clone, we then searched for SNVs and indels from WGS data within 200 bp of the BWA hit. CRISPR off-target variants detected by Off-flow and BWA string search were confirmed by bidirectional Sanger sequencing or PCR and gel electrophoresis.

2.7. Statistical analysis

Data are expressed as mean \pm standard deviation. Statistical analysis was performed using RStudio® (Version 2023.09.1 +494, RStudio, Inc.) equipped with R (Version 4.3.0, R Foundation for Statistical Computing). T-tests were used for intergroup comparisons of continuous variables. Statistical significance was set at $P < 0.05$.

3. Results

3.1. Experimental design

We targeted six genetic loci in two genes (*ASH1L1*, *KCNQ2*) that are relevant to autism spectrum disorder and epilepsy [25–28] for repair of disease-related variants in participant iPSC lines and one genetic locus in the third gene (*GNAQ*) linked to Sturge-Weber syndrome [29] for insertion of a disease-related variant in a control iPSC line by HDR (Table 2). Participants with variants in *ASH1L* and *KCNQ2* were recruited to this study to generate iPSCs for research. The *GNAQ* c.584 G>A (p.R183Q, NM_002072) variant is a major determinant genetic factor in Sturge-Weber syndrome [38] and was inserted in a control iPSC line for further study. Repair of disease-related variants was performed by nucleofection of RNP complexes of CRISPR-Cas9 or CRISPR-Cas12a with gRNAs and single stranded oligonucleotide repair templates. Insertion of disease-related variants was performed by the

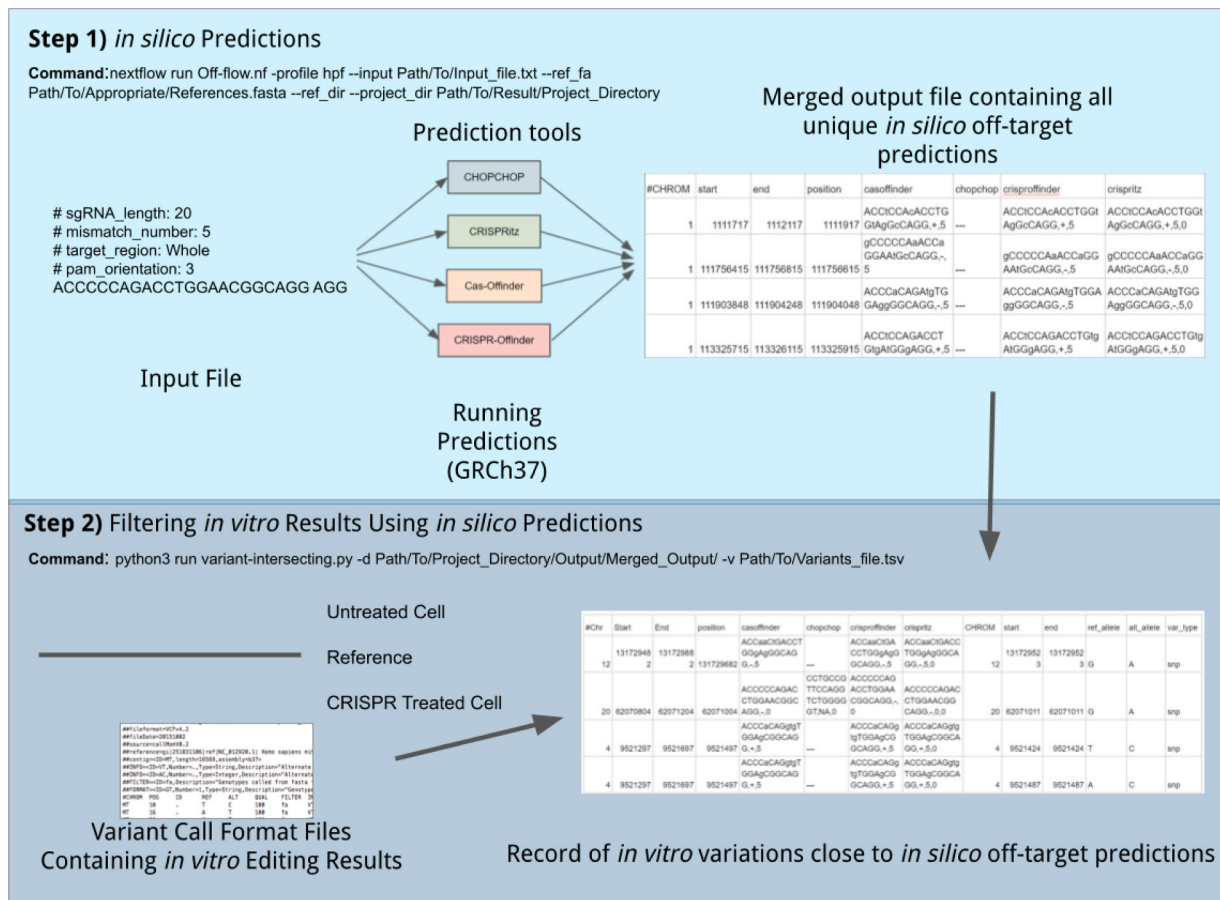


Fig. 1. Off-flow, a Nextflow-coded bioinformatic workflow program to automate CRISPR off-target prediction and detection.

UConn Human Genome Editing Core. We performed WGS in the 7 original iPSC lines and 16 clones generated after targeting these 7 loci. These include 10 CRISPR-corrected (CC) clones, 6 corresponding CRISPR-not-corrected (CNC) clones, defined as clones that were not edited at the intended genetic loci after targeting with gRNA, Cas proteins and repair template.

3.2. Genomic comparisons of isogenic iPSC lines

WGS was performed to investigate the genomic profiles of CRISPR-edited iPSC clones. The average coverage relative to the hg19 reference sequence was 41.3x (Table 3). Mutect2 (v4.1.9.0) was used to call

Table 3

Summary of WGS data and genomic comparisons of CRISPR/Cas engineered iPSC lines.

Cell line ID	Genome coverage	Unique SNVs	Unique indels	Unique SVs
HNDS0068-01 #B CNC14	42.5	202	40	0
HNDS0068-01 #B CC3	37.1	229	24	0
HNDS0078-01 #D CNC2	49.2	217	49	0
HNDS0078-01 #D CC8	49.6	242	40	0
HNDS0078-01 #D CC18	39.9	194	34	0
HNDS0072-01 #C CNC87	36.3	363	36	0
HNDS0072-01 #C CC80	43.8	327	44	1
HNDS0072-01 #C CC20	41.8	343	46	0
31 CC-het	42.8	432	77	0
31 CC-hom	33.4	792	73	1
1-1134-003_CNC5	45.4	190	48	0
1-1134-003_CC10	45.2	220	46	0
1-1217-003_CNC36	32.2	265	60	0
1-1217-003_CC37	39.2	260	64	0
1-1006-003_CNC35	39.4	268	64	0
1-1006-003_CC9	44.3	234	71	0

somatic variants in iPSC clones that underwent HDR given their matched original iPSC line. Following quality control, 298.6 unique SNVs, 51.0 unique indels, and 0.1 unique structural variants (SVs, defined as deletions, duplications, insertions, and inversions ≥ 50 bp) per genome were identified. (Table 3). No copy number variants (CNVs, defined as unbalanced changes >1 kb) were detected in any sample.

3.3. Detection of on-target effects in CRISPR/Cas-targeted iPSC lines

Off-flow detected all intended silent variant sites incorporated for the purpose of preventing recutting by Cas proteins and screening CRISPR-edited clones except for those designed for cell lines 1–1217-003 and

Table 4

List of genetic variants detected by Off-flow and BWA string search.

Cell line ID	Locus (hg19)	Gene	Variant type	Effect
HNDS0068-01 #B CC3	Chr20:62071011	<i>KCNQ2</i>	SNP	synonymous
HNDS0078-01 #D CC8	Chr20:62073825	<i>KCNQ2</i>	SNP	synonymous
	Chr20:62073828	<i>KCNQ2</i>	SNP	Synonymous
	Chr20:62073873-62073875	<i>KCNQ2</i>	MNP	Nonframeshift substitution
HNDS0078-01 #D CC18	Chr20:62073825	<i>KCNQ2</i>	SNP	synonymous
	Chr20:62073828	<i>KCNQ2</i>	SNP	synonymous
HNDS0072-01 #C CC80	Chr20:62071055	<i>KCNQ2</i>	SNP	synonymous
	Chr20:62071037	<i>KCNQ2</i>	insertion	Frameshift insertion
HNDS0072-01 #C CC20	Chr20:62071055	<i>KCNQ2</i>	SNP	synonymous
31 CC-het	Chr9:80412493	<i>GNAQ</i>	SNP	nonsynonymous
31 CC-hom	Chr9:80412493	<i>GNAQ</i>	SNP	nonsynonymous
1-1134-003_CC10	Chr1:155451881	<i>ASH1L</i>	SNP	synonymous
	Chr1:155451898	<i>ASH1L</i>	SNP	synonymous
	Chr1:155451904	<i>ASH1L</i>	SNP	synonymous
1-1217-003_CC37	Chr1:155447750	<i>ASH1L</i>	SNP	synonymous
	Chr1:155447753	<i>ASH1L</i>	SNP	synonymous
	Chr1:155447771	<i>ASH1L</i>	SNP	synonymous
1-1006-003_CC9	Chr1:155450705	<i>ASH1L</i>	SNP	synonymous
	Chr1:155450708	<i>ASH1L</i>	SNP	synonymous
	Chr1:155450723	<i>ASH1L</i>	SNP	synonymous

1–1134-003 (Table 4). This analysis also identified unintended effects at or near the target region in two CC clones. Off-flow revealed a 3-bp substitution in *KCNQ2* at position chr20:62,073,873–62,073,875 (hg19, AGT>CTG), resulting in a single amino acid residue change (p. T234L) in cell line HNDS0078–01 #D CC8 (Fig. 2). It also identified a 71-bp frameshift insertion at position chr20:62,071,037 in cell line HNDS0072–01 #C CC80 (Fig. 2). Further examination of the region in CRAM file from cell line HNDS0072–01 #C CC80 revealed a 1,857-bp mono-allelic deletion (chr20:62,068,445–62,070,301, hg19) (Fig. 2).

Off-flow did not detect any off-target missense variants or off-target indels in the iPSC clones assessed. A parallel BWA string search for sequences matching (up to four mismatches) gRNA+ and gRNA-, within a 200-bp window flanking both sides of the unique SNV and indels for each iPSC clone detected the same on-target variants identified by Off-flow and the intended silent variant sites for cell lines 1–1217-003 and 1–1134-003. Unintended on-target variants were validated by bidirectional Sanger sequencing or PCR and gel electrophoresis (Fig. 2).

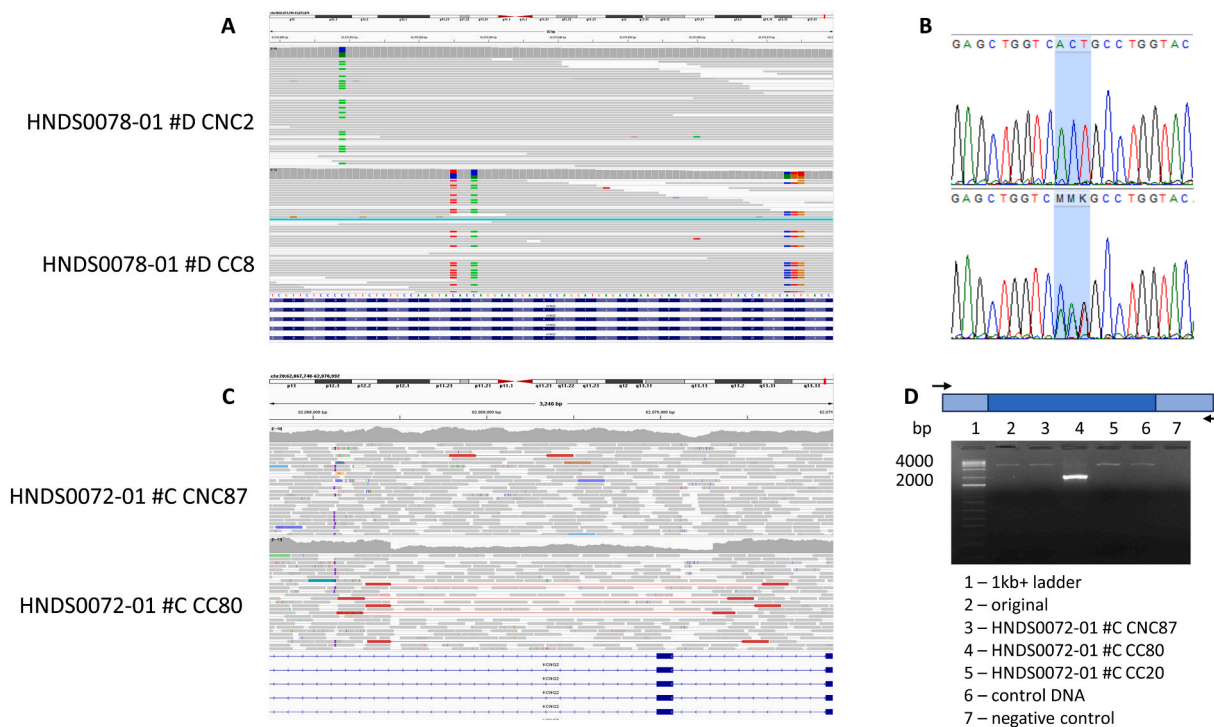


Fig. 2. On-target 3-bp substitution detected by Off-flow in CRISPR/Cas-edited clone HNDS0078–01 #D CC8 and on-target mono-allelic deletion detected by Off-flow in CRISPR/Cas-edited clone HNDS0072-01 #C CC80. (A) IGV visualization. The colored portions of reads represent mismatched bases. Each color represents a different nucleotide, green represents A, blue represents C, red represents T and orange represents G. If all reads match the reference genome, the coverage bar is gray. If a nucleotide differs from the reference sequence in greater than 20% of quality weighted reads, the bars are colored in proportion to the read count of each base. (B) Sequence trace showing the 3 bp substitution in *KCNQ2* in cell line HNDS0078–01 #D CC8. (C) IGV visualization. Deletions are displayed with a red bar. The length of the bar indicates the size of the deletion. (D) Schematic of PCR amplicons. Dark blue rectangle shows location of deletion. Arrows show the primers used for PCR (top panel). Agarose gel electrophoresis of ~4000 bp PCR amplified fragment reveals a single band of ~2000 bp from cell line HNDS0072–01 #C CC80 that was not present in other clones (bottom panel).

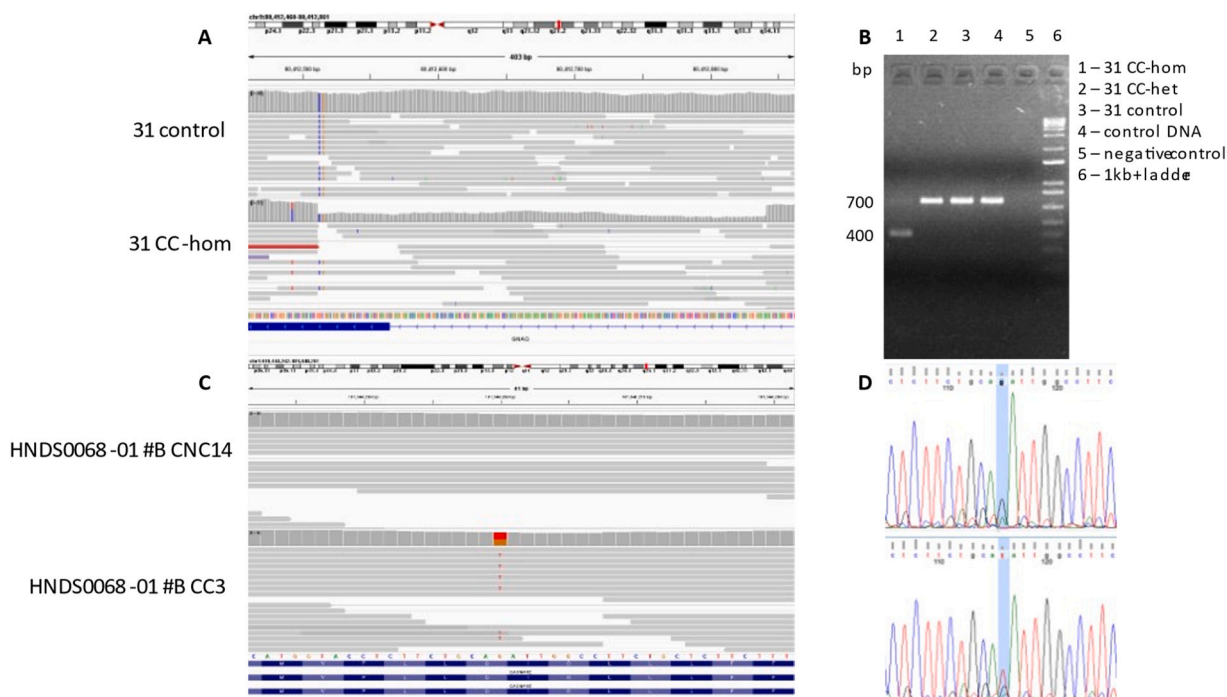


Fig. 3. On- and off-target effects identified by WGS analysis and undetected by Off-flow in CRISPR/Cas-edited clones. (A) IGV visualization. Deletions are displayed with a gray bar. The length of the bar indicates the size of the deletion. (B) Agarose gel electrophoresis of ~700 bp PCR amplified fragment reveals a band of ~400 bp from cell line CC-hom that was not present in other clones. (C) IGV visualization. Each color represents a different nucleotide. Red represents T and orange represents G. The bars are colored in proportion to the read count of each base. (D) Sequence trace showing the missense variant in *CACNA1E* in cell line HNDS0068–01 #B CC3.

Table 5

List of deletion or missense variants identified from WGS analysis.

iPSC clone	Locus (hg19)	Gene	Variant type	Effect
HNDS0068-01 #B CC3	Chr1: 181548260	<i>CACNA1E</i>	SNP	nonsynonymous
	Chr3:126915614	<i>C3ORF56</i>	SNP	nonsynonymous
	Chr22:26168388	<i>MYO18B</i>	SNP	nonsynonymous
HNDS0078-01 #D CNC2	Chr16:6367034	<i>RBFOX1</i>	SNP	nonsynonymous
HNDS0078-01 #D CC18	Chr8:105368407	<i>DCSTAMP</i>	SNP	nonsynonymous
HNDS0072-01 #C CNC87	Chr7:6692698	<i>ZNF316</i>	SNP	nonsynonymous
HNDS0072-01 #C CC20	Chr17:36486618	<i>GPR179</i>	SNP	nonsynonymous
HNDS0072-01 #C CC80	Chr2:21365297	<i>TDRD15</i>	SNP	nonsynonymous
	Chr16:15878562	<i>MYH11</i>	SNP	nonsynonymous
	Chr8:1616677	<i>DLGAP2</i>	SNP	nonsynonymous
31 CC-het	Chr15:45562441	<i>SLC28A</i>	SNP	nonsynonymous
	Chr1:184692924	<i>EDEM3</i>	SNP	nonsynonymous
31 CC-hom	Chr2:170850922	<i>URB3</i>	SNP	nonsynonymous
	Chr5:140589980	<i>PCDHB12</i>	SNP	nonsynonymous
	Chr9: 80412511	<i>GNAQ</i>	Deletion	
	Chr11:64679322	<i>ATG2A</i>	SNP	nonsynonymous
	Chr15:42492130	<i>VPS39</i>	SNP	nonsynonymous
	Chr16:560714	<i>RAB11FIP3</i>	SNP	nonsynonymous
	Chr16:30001071	<i>TAOK2</i>	SNP	nonsynonymous
1-1134-003_CNC5	Chr4:964783	<i>DGKQ</i>	SNP	nonsynonymous
1-1217-003_CNC36	Chr5:148709327	<i>AFAP11</i>	SNP	nonsynonymous
1-1006-003_CC9	Chr14:42356229-42356230	<i>LRFN5</i>	MNP	Nonframeshift substitution

3.4. Detection of genome-wide unintended variants in CRISPR/Cas-targeted iPSC clones

To determine whether additional genetic variation arose from the CRISPR/Cas-gene editing process, we examined WGS data in isogenic iPSC lines irrespective of off-target prediction using our previous approaches [32]. We identified an on-target 329-bp (chr9:80,412,512–80,412,841, hg19) mono-allelic deletion in the cell line 31 CC-hom (Fig. 3). In addition, 21 missense variants at other genomic loci were observed in 7 CC and 4 CNC clones (Table 5). Unintended variants were validated by bidirectional Sanger sequencing or PCR and gel electrophoresis (Table 6).

3.5. Gene editing by CRISPR/Cas9 plasmids is associated with greater unintended genetic variation

We compared the rates of unintended genetic variation in iPSC clones generated by different Cas proteins or different delivery approaches. The total number of unintended variants were not different in CRISPR-Cas9- or CRISPR-Cas12a-targeted clones (Fig. 4; 1.83 variants/clone vs 0.5 variants/clone, $p = 0.2536$, two sample unpaired t-test). iPSC clones generated by CRISPR/Cas plasmids showed greater unintended variation than those generated by CRISPR/Cas RNPs (5.0 variants/clone vs 1.0 variants/clone, $p = 0.003$, two sample unpaired t-test) (Fig. 4). Based on the small sample size, variance and means of the

Table 6
gRNA and repair template sequences.

Target locus (hg19)	Gene	sgRNA or crRNA	ssODN
Chr20:62071001-62071003	KCNQ2	ACCCCCAGACCTGGAAGGGC (AGG)	AGCGGAAAGGAGACACCCGATGAGGGTGAAGGTTGCGGCAAGGAGCCCTGCCTTTCCAAAGTCTGG GGGTACTTGTCCCGGTAGCCAAATGGTGTGACGGCT
Chr20:62073808	KCNQ2	TCTCATCTCGGGCTCGTTCC (TGG)	GCCTGTACATGGCGTTCTCTCTCCTCTCCTGCGCTGTTCCTTGTATAGTTGGCAGAGAAGGGGG AGAAGGACCACTTTGACACCTACCGCGATGCACTCTGGTGGGGCTGGTGTAGTTGGTCA
Chr20:62071057	KCNQ2	GCTGACCACCACTTGGCTAGC (GGG)	GGGCTCCAGCCCTGCCCTCAGGGGTGAGCAGGCCCTCTGCTGTGACTAGAGCCCTGGGTCCACCA GATCAGTGTACCACCACTTGGCTAGCGGGAACAAGTACCCCCAGACCTGGAACGGGAGGCTCTCT
Chr1: 155451888	ASHL	AITCAGGAAAGCAGTGTGGC (TTTG)	TGGCTTTTATTAAGTCTTATGTAITTTCCAGTACAGATCCAAACAGCTCTTTCCGGATCAGA TCCTTGTAAACCAATCCTGCTGTAGTGCAGTT
Chr1: 155447758	ASHL	GTCCCTGTGCAGAAGGTCCA (GGG)	CGTTTAAAGAGGCTACTCCTTTCTGTGAAGCCGATTTTAGTAAAGTCTCTGGGCAAGAAAAGAGTCT CAGGACTGTCAAGTAAATTTCTCCGGCCACTGGA
Chr1: 155450703	ASHL	CAAGGGAAAGCCTTATTCCTT (TTTC)	GAATCTGGATTTCAGAAAGTCAACTTGGGCTTCTTACCAAGGAAAGAGCTGATTTCTTTGGAAATATCTA TATGGGTAGTTTTGAATCAATTAACCTCTTTATC
Chr9:80412493	GNAQ	CTAAGCACATCTTTTGGCGT (AGG)	n/a
ChrX:133620533-133620553	HPR1	GATGATCTCAACITTAAC (TGG)	n/a
ChrX:133609206-133609228	HPR1	TGTAGGACTGACGCTTGTGCTCG (TTTC)	n/a
Nontarget (CRISPR-Cas9)	n/a	GCACTACGAGAGCTAAGTCA	n/a
Nontarget (CRISPR-Cas12a)	n/a	CGTTAATCGGGTATAATACGG	n/a

aggregated statistics, this study was not sufficiently powered to assess the differences among the different delivery methods or different Cas proteins used for gene editing.

4. Discussion

CRISPR and iPSC technology facilitate research into the genetic and molecular bases of human biology and disease. Engineered CRISPR-Cas systems have improved efficacy and specificity [5,6,39–41], but undesired editing outcomes, both on-target and off-target, remain a concern for broader application of this technology. There is currently no standard for assessing the fidelity of CRISPR-Cas editing in iPSCs. In this study, we developed a standard methodology by using WGS, BWA string search and Off-flow, a bioinformatic pipeline, to evaluate the specificity of CRISPR-Cas9 and -Cas12a HDR in human iPSCs. We observed unintended variants in more than half of the iPSC clones that had undergone gene editing, including on-target substitution and mono-allelic deletions, and off-target missense variants.

Off-flow detected intended on-target synonymous changes incorporated for preventing recutting by Cas proteins and for screening CRISPR-edited clones for all cell lines except for cell lines 1–1217-003 and 1–1134-003 (Table 4). The gRNA sequences for these two cell lines were designed to complement the indels specific to the participant DNA sequence. Off-flow also detected unintended on-target changes in two edited clones, including a 3-bp substitution and a mono-allelic 1,857-bp deletion in *KCNQ2* (Table 4 and Fig. 2). The 3-bp substitution is located at the edge of the repair template, which may be a potential hot spot for unintended effects. These unintended variants had escaped standard PCR and Sanger sequencing analysis. Two recent studies also reported the presence of unintended variants at the target loci in up to 40% of CRISPR-Cas9-edited human iPSC clones [42,43]. Unintended variants were of different types, including large deletions and insertions, which were further shown to have a functional impact on downstream phenotyping studies [42,43]. These and our data show that target loci and up to 2 kb surrounding the target should be scrutinized in detail following gene editing by CRISPR-Cas nucleases.

Off-flow did not detect any off-target missense variants nor small indels in the iPSC clones assessed in this study. The program used a specified number of mismatches for each prediction tool and “bulge” mismatches of 2 (Table 1). CRISPR-Cas nucleases have been shown to cleave at off-target sites containing up to 7 mismatched nucleotides [44], but CRISPR/Cas systems have also shown high fidelity in iPSCs [45]. An important limitation to Off-flow is that it includes only alignment-based in silico off-target prediction tools. Furthermore, one of the prediction tools included in Off-flow, CRISPRitz, is limited by a specific combination of gRNA and PAM sequences. Addition of hypothesis-driven, learning-based, and energy-based prediction tools to the workflow may generate a more comprehensive list of potential off-target sites. Höijer et al. [46] identified difficult to predict and difficult to detect CRISPR-Cas9 off-target sites using long-read sequencing [46]. In addition, this study showed that a SNV can induce allele-specific Cas9 cleavage [46]. Current in silico off-target prediction tools use the reference genome for computational modeling. Thus, differences in the DNA sequence of each participant cell line relative to the reference genome may result in false-positive or false-negative predictions.

Irrespective of off-target prediction, WGS analysis of SNVs, indels, CNVs and SVs was performed to study the impact of the gene editing process on genetic variation. This revealed a mono-allelic 329-bp deletion in *GNAQ*, near the target and SNVs at other genomic loci in 10/16 iPSC clones that was not detected by Off-flow (Table 5). Some of these variants are found in genes relevant to ASD and epilepsy, including *RBFOX1*, *DLGAP2* and *TAOK2* [47–52]. Based on the study design, it is unclear whether these protein-coding variants arose from routine culturing and expansion of iPSCs or from the gene editing process. Somatic coding variants in human iPSCs have been reported to be enriched

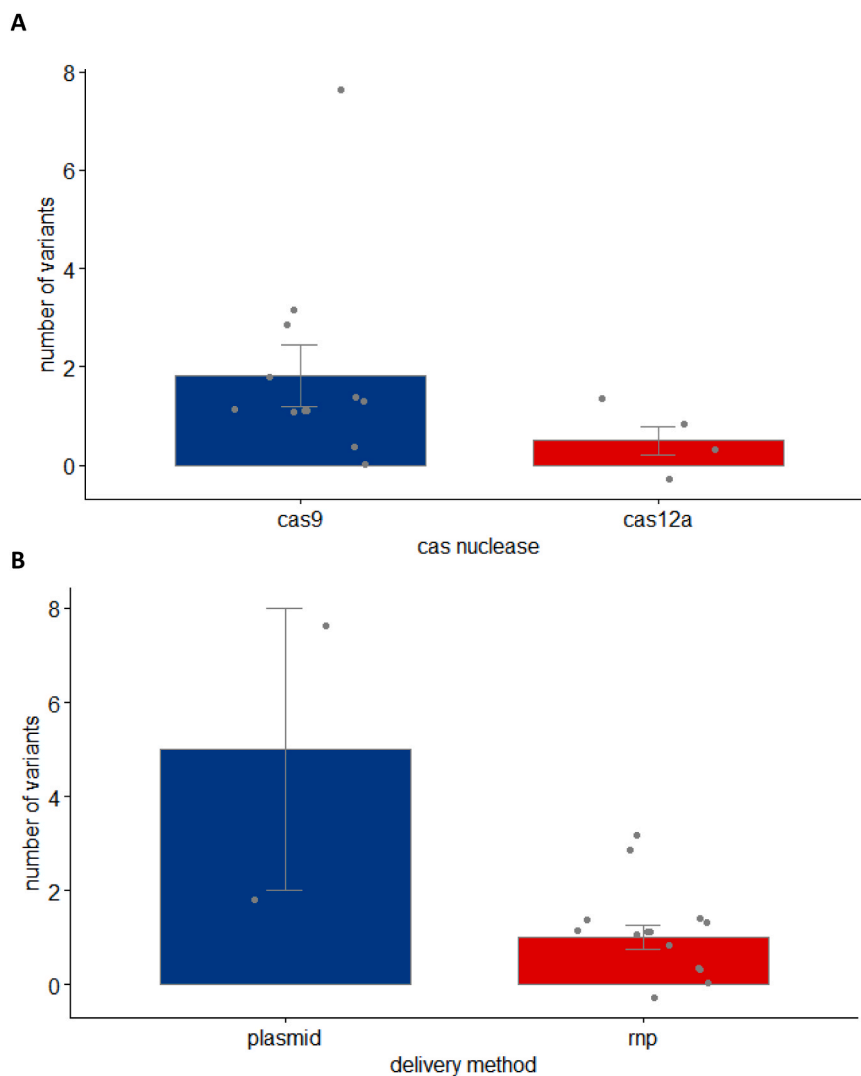


Fig. 4. Rate of unintended effects observed in iPSC clones (A) targeted by CRISPR/Cas9 or Cas12a, (B) delivered by RNPs or plasmids.

in genes mutated or having causative effects in cancers [53,54] or in active promoters [55] or depleted from genic regions [56]. However, another in-depth analysis showed that variants in human iPSCs are generally benign and fall within intergenic or intronic regions [57]. In agreement with the latter study, the subclonal SNVs observed in this show enrichment in intergenic and intronic regions. In addition, none of these SNVs were reported in iPSC tissue in SomaMutDB (accessed 13 Sep 2023), a database of somatic variants [58].

The SNVs observed in this study may be associated with the gene editing process (nucleofection, colony-picking). Of note, one such variant (*CACNA1E*, c.G669T, p.Q223H) was observed in an edited clone following CRISPR-Cas9 editing of a participant cell line with a variant in *KCNQ2* (c.875_877delTCCinsCCT, p.L292_L293delinsPF). Both *CACNA1E* and *KCNQ2* are clinically relevant genes associated with epileptic encephalopathy [25,26,59]. The unintended *CACNA1E* variant is predicted to be likely damaging, thus may have functional consequences. If iPSC gene editing efficiency is improved to a level to bypass colony-picking and clonal expansion, resulting isogenic iPSCs could be free of variants.

iPSC clones edited with CRISPR-Cas12a show less unintended effects than those edited with CRISPR-Cas9, although this difference did not reach significance in the small sample assessed in this study. CRISPR-Cas12a has been reported to be more efficient and precise [60].

Similarly, and in agreement with previous studies, introduction of CRISPR-Cas and repair templates via RNPs is associated with reduced unintended effects compared to introduction via vectors [61,62]. mRNA delivery of CRISPR-Cas systems is another transient delivery method associated with efficient genome-editing, reduced toxicity and off-target activity [63]. Finally, the more recently engineered prime editing system has been shown to be more specific than all other CRISPR-Cas systems, although with lower efficacy [64,65]. Thus, developments in both delivery and engineering of CRISPR-Cas continue to improve the fidelity of gene editing.

Quantitative genomic PCR (qgPCR) and SNP genotyping-based tools have been proposed as an additional quality control measure to detect unintended on-target variants from CRISPR-Cas editing [42]. This approach enables the detection of deleterious on-target variants and has been shown to reliably detect variants that are missed by the traditional approach of Sanger sequencing of the target region. qgPCR and SNP genotyping are reliable and economical tools for revealing large deletions or insertions but may not detect the 3-bp on-target substitution observed in this study.

To our knowledge, WGS and Off-flow is the most robust and stringent in silico and in vivo method for identifying unintended variants from CRISPR-Cas-mediated genome editing. Our findings suggest that HDR of iPSCs using CRISPR-Cas nucleases should be scrutinized for genome-

wide unintended variants prior to downstream functional studies. We propose a standard for evaluating and screening of CRISPR-Cas-edited iPSCs for functional studies: 1. An initial screen by Sanger sequencing the 500 bp surrounding the targeted genomic loci, including potential off-target hot spots at the edge of repair templates, 2. An intermediate step using qPCR to determine deletions or insertions or loss of heterozygosity around the targeted genomic loci, 3. Off-flow and WGS of the edited clone to determine unintended genome-wide CRISPR-Cas effects, 4. Validation of unintended variants using molecular biology approaches (PCR/gel/Sanger sequencing). Alternatively, multiple edited and control unedited clones or multiple gRNAs may be used to facilitate the integrity of iPSC-based studies. As CRISPR-Cas systems continue to improve, quality control strategies should be revisited and standards for evaluating CRISPR-Cas-edited iPSCs updated.

Funding

This work was funded by Autism Speaks, the University of Toronto McLaughlin Center, the Northbridge Chair in Pediatric Research held at the Hospital for Sick Children and University of Toronto, the Canada Foundation for Innovation and the Hospital for Sick Children Foundation.

Author statement

All authors contributed to this research and to drafting this manuscript. All authors approved the submission to CSBJ. The manuscript or any content within it are not currently under consideration or published by another journal.

CRedit authorship contribution statement

Carole Shum: Conceptualization, Formal analysis, Investigation, Writing – Original draft preparation, Visualization; **Yeon Han:** Formal analysis, Investigation, Validation; **Bhooma Thiruvahindrapuram:** Methodology, Software, Formal analysis; **Zhuozhi Wang:** Software, Formal analysis; **Benjamin Zhang:** Software, Formal Analysis; **Jill de Rijke:** Validation, Writing – Original draft preparation; **Maria Sundberg:** Investigation, Resources; **Cidi Chen:** Investigation, Resources, Validation; **Elizabeth D. Buttermore:** Investigation, Resources, Project administration; **Nina Makhortova:** Resources, Project administration; **Jennifer Howe:** Project administration, Writing – Review & Editing; **Mustafa Sahin:** Supervision, Funding acquisition; **Stephen W. Scherer:** Conceptualization, Writing – Review & Editing, Supervision, Funding acquisition.

Declaration of Competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank the participants and their family members for their contributions to this study. We acknowledge the resources of The Center for Applied Genomics.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.csbj.2023.12.036](https://doi.org/10.1016/j.csbj.2023.12.036).

References

- [1] Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 2007;131:861–72.
- [2] Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 2012;337:816–21. <https://doi.org/10.1126/science.1225829>.
- [3] Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, et al. RNA-guided human genome engineering via Cas9. *Science* 2013;339:823–6. <https://doi.org/10.1126/science.1232033>.
- [4] Kleinstiver BP, Prew MS, Tsai SQ, Topkar VV, Nguyen NT, Zheng Z, et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* 2015; 523:481–5. <https://doi.org/10.1038/nature14592>.
- [5] Hu JH, Miller SM, Geurts MH, Tang W, Chen L, Sun N, et al. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* 2018;556:57–63. <https://doi.org/10.1038/nature26155>.
- [6] Kleinstiver BP, Sousa AA, Walton RT, Tak YE, Hsu JY, Clement K, et al. Engineered CRISPR-Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. *Nat Biotechnol* 2019;37:276–82. <https://doi.org/10.1038/s41587-018-0011-0>.
- [7] Miller SM, Wang T, Randolph PB, Arbab M, Shen MW, Huang TP, et al. Continuous evolution of SpCas9 variants compatible with non-G PAMs. *Nat Biotechnol* 2020; 38:471–81. <https://doi.org/10.1038/s41587-020-0412-8>.
- [8] Walton RT, Christie KA, Whittaker MN, Kleinstiver BP. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* 2020;368. <https://doi.org/10.1126/science.aba8853>.
- [9] Firth Amy L, Menon T, Parker Gregory S, Qualls Susan J, Lewis Benjamin M, Ke E, et al. Functional gene correction for cystic fibrosis in lung epithelial cells generated from patient iPSCs. *Cell Rep* 2015;12:1385–90. <https://doi.org/10.1016/j.celrep.2015.07.062>.
- [10] Deneault E, White SH, Rodrigues DC, Ross PJ, Faheem M, Zaslavsky K, et al. Complete disruption of autism-susceptibility genes by gene editing predominantly reduces functional connectivity of isogenic human neurons. *Stem Cell Rep* 2018; 11:1211–25. <https://doi.org/10.1016/j.stemcr.2018.10.003>.
- [11] Lin Y-T, Seo J, Gao F, Feldman HM, Wen H-L, Penney J, et al. APOE4 causes widespread molecular and cellular alterations associated with Alzheimer's disease phenotypes in human iPSC-derived brain cell types. *Neuron* 2018;98:1141–1154. <https://doi.org/10.1016/j.neuron.2018.05.008>.
- [12] Deneault E, Faheem M, White SH, Rodrigues DC, Sun S, Wei W, et al. CNTN5-/+or EHMT2-/+human iPSC-derived neurons from individuals with autism develop hyperactive neuronal networks. *eLife* 2019;8:e40092. <https://doi.org/10.7554/eLife.40092>.
- [13] Ross PJoel, Zhang Wen Bo, Mok Rebecca SF, Zaslavsky K, Deneault E, D'Abate L, et al. Synaptic dysfunction in human neurons with autism-associated deletions in PTCHD1-AS. *Biol Psychiatry* 2020;87:139–49. <https://doi.org/10.1016/j.biopsych.2019.07.014>.
- [14] Bae S, Park J, Kim J-S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* 2014;30:1473–5. <https://doi.org/10.1093/bioinformatics/btu048>.
- [15] Tsai SQ, Zheng Z, Nguyen NT, Liebers M, Topkar VV, Thapar V, et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol* 2014;33:187–97. <https://doi.org/10.1038/nbt.3117>.
- [16] Kim D, Bae S, Park J, Kim E, Kim S, Yu HR, et al. Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat Methods* 2015;12: 237–43. <https://doi.org/10.1038/nmeth.3284>.
- [17] Cameron P, Fuller CK, Donohoue PD, Jones BN, Thompson MS, Carter MM, et al. Mapping the genomic landscape of CRISPR-Cas9 cleavage. *Nat Methods* 2017;14 (6):600. <https://doi.org/10.1038/nmeth.4284>.
- [18] Tsai SQ, Nguyen NT, Malagon-Lopez J, Topkar VV, Aryee MJ, Joung JK. CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets. *Nat Methods* 2017;14:607–14. <https://doi.org/10.1038/nmeth.4278>.
- [19] Labun K, Montague TG, Krause M, Torres Cleuren YN, Tjeldnes H, Valen E. CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Res* 2019;47:W171–4. <https://doi.org/10.1093/nar/gkz365>.
- [20] Cancellieri S, Canver MC, Bombieri N, Giugno R, Pinello L. CRISPRitz: rapid, high-throughput and variant-aware in silico off-target site identification for CRISPR genome editing. *Bioinformatics* 2019;36:2001–8. <https://doi.org/10.1093/bioinformatics/btz867>.
- [21] Wienert B, Wyman SK, Richardson CD, Yeh CD, Akcakaya P, Porritt MJ, et al. Unbiased detection of CRISPR off-targets in vivo using DISCOVER-Seq. *Science* 2019;364:286–9. <https://doi.org/10.1126/science.aav9023>.
- [22] Yan J, Xue D, Chuai Guohui, Gao Y, Zhang G, Liu Q. Benchmarking and integrating genome-wide CRISPR off-target detection and prediction. *Nucleic Acids Res* 2020; 48:11370–9. <https://doi.org/10.1093/nar/gkaa930>.
- [23] Corsi GI, Gadekar VP, Gorodkin J, Seemann SE. CRISPRroots: on- and off-target assessment of RNA-seq data in CRISPR-Cas9 edited cells. *Nucleic Acids Res* 2021; 50. <https://doi.org/10.1093/nar/gkab1131>.
- [24] Simkin DJ, Marshall KA, Vanoye CG, Desai RR, Bustos BI, Piyevsky BN, et al. Dyshomeostatic modulation of Ca²⁺-activated K⁺ channels in a human neuronal model of KCNQ2 encephalopathy. *eLife* 2021;10. <https://doi.org/10.7554/elife.64434>.

- [25] Dedek K, Fusco L, Teloy N, Steinlein OK. Neonatal convulsions and epileptic encephalopathy in an Italian family with a missense mutation in the fifth transmembrane region of KCNQ2. *Epilepsy Res* 2003;54:21–7. [https://doi.org/10.1016/s0920-1211\(03\)00037-8](https://doi.org/10.1016/s0920-1211(03)00037-8).
- [26] Kato M, Yamagata T, Kubota M, Arai H, Yamashita S, Nakagawa T, et al. Clinical spectrum of early onset epileptic encephalopathies caused by KCNQ2 mutation. *Epilepsia* 2013;54:1282–7. <https://doi.org/10.1111/epi.12200>.
- [27] Shen W, Krautscheid P, Rutz AM, Bayrak-Toydemir P, Dugan SL. De novo loss-of-function variants of ASH1L are associated with an emergent neurodevelopmental disorder. *Eur J Med Genet* 2019;62:55–60. <https://doi.org/10.1016/j.ejmg.2018.05.003>.
- [28] Stessman HAF, Xiong B, Coe BP, Wang T, Hoekzema K, Fenckova M, et al. Targeted sequencing identifies 91 neurodevelopmental disorder risk genes with autism and developmental disability biases. *Nat Genet* 2017;49:515–26. <https://doi.org/10.1038/ng.3792>.
- [29] Shirley MD, Tang H, Gallione CJ, Baugher JD, Frelin LP, Cohen B, et al. Sturge–Weber syndrome and port-wine stains caused by somatic mutation in GNAQ. *N Engl J Med* 2013;368:1971–9. <https://doi.org/10.1056/NEJMoa1213507>.
- [30] Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C. Nextflow enables reproducible computational workflows. *Nat Biotechnol* 2017;35:316–9. <https://doi.org/10.1038/nbt.3820>.
- [31] McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297–303. <https://doi.org/10.1101/gr.107524.110>.
- [32] Trost B, Thiruvahindrapuram B, Chan AJS, Engchuan W, Higginbotham EJ, Howe JL, et al. Genomic architecture of autism from comprehensive whole-genome sequence annotation. *Cell* 2022;185:4409–27. <https://doi.org/10.1016/j.cell.2022.10.009>.
- [33] Abyzov A, Urban AE, Snyder M, Gerstein M. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res* 2011;21:974–84. <https://doi.org/10.1101/gr.114876.110>.
- [34] Zhu M, Need AC, Han Y, Ge D, Maia JM, Zhu Q, et al. Using ERDS to infer copy-number variants in high-coverage genomes. *Am J Hum Genet* 2012;91:408–21. <https://doi.org/10.1016/j.ajhg.2012.07.004>.
- [35] Trost B, Walker SP, Sun R, Thiruvahindrapuram B, Macdonald JM, Sung WWL, et al. A comprehensive workflow for read depth-based identification of copy-number variation from whole-genome sequence data. *Am J Hum Genet* 2018;102:142–55. <https://doi.org/10.1016/j.ajhg.2017.12.007>.
- [36] Zhao C, Zheng X, Qu W, Li G, Li X, Miao Y-L, et al. CRISPR-offfinder: a CRISPR guide RNA design and off-target searching tool for user-defined protospacer adjacent motif. *Int J Biol Sci* 2017;13:1470–8. <https://doi.org/10.7150/ijbs.21312>.
- [37] Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 2009;25:1754–60. <https://doi.org/10.1093/bioinformatics/btp324>.
- [38] Nakashima M, Miyajima M, Sugano H, Yasuhiro Iimura, Kato M, Yoshinori Tsurusaki, et al. The somatic GNAQ mutation c.548G>A (p.R183Q) is consistently found in Sturge–Weber syndrome. *J Hum Genet* 2014;59:691–3. <https://doi.org/10.1038/jhg.2014.95>.
- [39] Ran FAnn, Hsu Patrick D, Lin C-Y, Gootenberg Jonathan S, Konermann S, Trevino AE, et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* 2013;154:1380–9. <https://doi.org/10.1016/j.cell.2013.08.021>.
- [40] Guilinger JP, Thompson DB, Liu DR. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat Biotechnol* 2014;32:577–82. <https://doi.org/10.1038/nbt.2909>.
- [41] Nishimasu H, Shi X, Ishiguro S, Gao L, Hirano S, Okazaki S, et al. Engineered CRISPR–Cas9 nuclease with expanded targeting space. *Science* 2018;361:1259–62. <https://doi.org/10.1126/science.aas9129>.
- [42] Weisheit I, Kroeger JA, Malik R, Klimmt J, Crusius D, Dannert A, et al. Detection of deleterious on-target effects after HDR-mediated CRISPR editing. *Cell Rep* 2020;31:107689. <https://doi.org/10.1016/j.celrep.2020.107689>.
- [43] Simkin D, Papakis V, Bustos BI, Ambrosi CM, Ryan SJ, Baru V, et al. Homozygous might be hemizygous: CRISPR/Cas9 editing in iPSCs results in detrimental on-target defects that escape standard quality controls. *Stem Cell Rep* 2022;17:993–1008. <https://doi.org/10.1016/j.stemcr.2022.02.008>.
- [44] Pattanayak V, Lin S, Guilinger JP, Ma E, Doudna JA, Liu DR. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol* 2013;31:839–43. <https://doi.org/10.1038/nbt.2673>.
- [45] Smith C, Gore A, Yan W, Abalde-Atristain L, Li Z, He C, et al. Whole-genome sequencing analysis reveals high specificity of CRISPR/Cas9 and TALEN-Based genome editing in human iPSCs. *Cell Stem Cell* 2014;15:12–3. <https://doi.org/10.1016/j.stem.2014.06.011>.
- [46] Höjjer I, Johansson J, Gudmundsson S, Chin C-S, Bunikis I, Häggqvist S, et al. Amplification-free long-read sequencing reveals unforeseen CRISPR–Cas9 off-target activity. *Genome Biol* 2020;21. <https://doi.org/10.1186/s13059-020-02206-w>.
- [47] Zhao W-W. Intragenic deletion of RBFOX1 associated with neurodevelopmental/neuropsychiatric disorders and possibly other clinical presentations. *Mol Cytogenet* 2013;6:26. <https://doi.org/10.1186/1755-8166-6-26>.
- [48] Lal D, Reinthaler EM, Altmüller J, Tolia MR, Thiele H, Nürnberg P, et al. RBFOX1 and RBFOX3 mutations in rolandic epilepsy. *PLOS One* 2013;8:e73323. <https://doi.org/10.1371/journal.pone.0073323>.
- [49] Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J, et al. Structural variation of chromosomes in autism spectrum disorder. *Am J Hum Genet* 2008;82:477–88. <https://doi.org/10.1016/j.ajhg.2007.12.009>.
- [50] Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R, et al. Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* 2010;466:368–72. <https://doi.org/10.1038/nature09146>.
- [51] C Yuen RK, Merico D, Bookman M, I Howe J, Thiruvahindrapuram B, Patel RV, et al. Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nat Neurosci* 2017;20:602–11. <https://doi.org/10.1038/nn.4524>.
- [52] Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, Fossdal R, et al. Association between microdeletion and microduplication at 16p11.2 and autism. *N Engl J Med* 2008;358:667–75. <https://doi.org/10.1056/nejmoa075974>.
- [53] Gore A, Li Z, Fung H-L, Young JE, Agarwal S, Antosiewicz-Bourget J, et al. Somatic coding mutations in human induced pluripotent stem cells. *Nature* 2011;471:63–7. <https://doi.org/10.1038/nature09805>.
- [54] Ruiz S, Gore A, Li Z, Panopoulos AD, Montserrat N, Fung H-L, et al. Analysis of protein-coding mutations in hiPSCs and their possible role during somatic cell reprogramming. *Nat Commun* 2013;4. <https://doi.org/10.1038/ncomms2381>.
- [55] D’Antonio M, Benaglio P, Jakubosky D, Greenwald WW, Matsui H, Margaret KR, et al. Insights into the mutational burden of human induced pluripotent stem cells from an integrative multi-omics approach. *Cell Rep* 2018;24:883–94. <https://doi.org/10.1016/j.celrep.2018.06.091>.
- [56] Kuijk E, Jager M, van der Roest B, Locati MD, Van Hoeck A, Korzelius J, et al. The mutational impact of culturing human pluripotent and adult stem cells. *Nat Commun* 2020;11:1–12. <https://doi.org/10.1038/s41467-020-16323-4>.
- [57] Bhutani K, Nazor KL, Williams R, Tran H, Dai H, Dzakula Ž, et al. Whole-genome mutational burden analysis of three pluripotency induction methods. *Nat Commun* 2016;7. <https://doi.org/10.1038/ncomms10536>.
- [58] Sun S, Wang Y, Maslov AY, Dong X, Vijg J. SomaMutDB: a database of somatic mutations in normal human tissues. *Nucleic Acids Res* 2021;50:D1100–8. <https://doi.org/10.1093/nar/gkab914>.
- [59] Helbig KL, Lauerer RJ, Bahr JC, Souza IA, Myers CT, Uysal B, et al. De Novo pathogenic variants in CACNA1E cause developmental and epileptic encephalopathy with contractures, macrocephaly, and dyskinesias. *Am J Hum Genet* 2018;103:666–78. <https://doi.org/10.1016/j.ajhg.2018.09.006>.
- [60] Strohkendl I, Saifuddin FA, Rybarski JR, Finkelstein LJ, Russell R. Kinetic basis for DNA target specificity of CRISPR–Cas12a. *Mol Cell* 2018;71:816–824.e3. <https://doi.org/10.1016/j.molcel.2018.06.043>.
- [61] Kim S, Kim D, Cho SW, Kim J, Kim J-S. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res* 2014;24:1012–9. <https://doi.org/10.1101/gr.171322.113>.
- [62] Ramakrishna S, Kwaku Dad A-B, Beloor J, Gopalappa R, Lee S-K, Kim H. Gene disruption by cell-penetrating peptide-mediated delivery of Cas9 protein and guide RNA. *Genome Res* 2014;24:1020–7. <https://doi.org/10.1101/gr.171264.113>.
- [63] Zhang H-X, Zhang Y, Yin H. Genome editing with mRNA encoding ZFN, TALEN, and Cas9. *Mol Ther* 2019;27:735–46. <https://doi.org/10.1016/j.ymthe.2019.01.014>.
- [64] Anzalone AV, Randolph PB, Davis JR, Sousa AA, Koblán LW, Levy JM, et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 2019;576. <https://doi.org/10.1038/s41586-019-1711-4>.
- [65] Nelson JW, Randolph PB, Shen SP, Everette KA, Chen PJ, Anzalone AV, et al. Engineered pegRNAs improve prime editing efficiency. *Nat Biotechnol* 2021;40. <https://doi.org/10.1038/s41587-021-01039-7>.