UNIVERSITY OF HELSINKI

Designing and generating isogenic CRISPR/Cas9-edited human iPSC lines for genetic variants in schizophrenia and intellectual disability

Master’s Programme in Genetics and Molecular Biosciences,
Genetics and Genomics study track
Master's thesis

Author:
Panagiotis Lilis

Supervisors:
Olli Pietiläinen, PhD
Nelli Jalkanen, MSc

24.07.2025
Helsinki

Faculty: Biological and Environmental Sciences

Degree programme: Master's Programme in Genetics and Molecular Biosciences

Study track: Genetics and Genomics

Author: Panagiotis Lilis

Title: Designing and generating isogenic CRISPR/Cas9-edited human iPSC lines for genetic variants in schizophrenia and intellectual disability.

Level: Master's thesis

Month and year: July 2025

Number of pages: 73 + 18 (Appendices)

Keywords: CRISPR/Cas9, sgRNA design, genome editing, iPSCs, rare genetic variants, schizophrenia, intellectual disability, *INTS6*

Supervisor or supervisors: Olli Pietiläinen, PhD; Nelli Jalkanen, MSc

Where deposited: Helsinki University Library / Helda

Additional information: NA

Abstract:

The emergence of the clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein 9 (Cas9) genome editing system has changed the way human genetics are studied. This innovative technology, combined with the use of human induced pluripotent stem cells (iPSCs), offers a unique setting for disease modelling. This is particularly useful in brain disorders with genetic basis, such as schizophrenia (SCZ) and intellectual disability (ID), where little is known of the underlying biological processes. Gene editing enables investigating causative effects of disease-associated genetic variants in donor-specific genetic background. Recently identified rare high-risk coding variants that directly nominate the affected gene propose an advantageous experimental setup for gene editing to study the biological mechanisms of these complex disorders.

The main objective of the thesis was to explore the CRISPR/Cas9 system in iPSCs to study the cause-and-effect relationship of rare predicted highly penetrant disease variants. Initially, the *in silico* single guide RNA (sgRNA) and homology-directed repair (HDR) DNA template design methodology was optimized to edit 18 rare variants in 8 genes linked to SCZ. As a proof of concept, isogenic pairs of CRISPR-edited iPSCs were generated to model a novel *INTS6* heterozygous loss-of-function variant. A wild-type iPSC line was used to introduce the pathogenic *INTS6* variant, and a heterozygous patient-derived iPSC line to rescue the genotype. After electroporation with two independent sgRNAs per cell line, T7 Endonuclease 1 (T7E1) assay was employed to assess the editing outcomes at a bulk level. Individual single-cell derived colonies were isolated, expanded in replicate 96-well plates, and subjected for Illumina MiSeq sequencing to identify successfully edited clones.

The sgRNA design approach showed that variants in 6 out of the 8 genes had high potential of successful editing *in silico*, while PAM (protospacer adjacent motif) limitations in the other 2 cases were resolved using an alternative nuclease with more flexible PAM requirements. In the proof-of-concept study of the *INTS6* variant, the T7E1 assay indicated low to moderate Cas9 cleavage activity. The sequencing analysis revealed that one sgRNA introduced HDR edits efficiently in both knock-in and isogenic correction lines. For the efficient sgRNA, 51% of the patient-derived iPSC clones were isogenically rescued, while 51% of the wild-type clones had HDR homozygous knock-in edits. In addition, one successful heterozygous clone (1%) was produced at the knock-in experiment. These findings demonstrated that multiple factors could impact the editing outcomes. Further improvements of the CRISPR design and experimental pipeline to increase the reproducibility and effectiveness of precise genome editing can be implemented. Moreover, the clonality of edited iPSCs must be considered, so further characterization of the edited iPSC lines is required before application in functional assays to study the cellular impact of the pathogenic variant.

Table of contents

List of Abbreviations

BSCC	Biomedicum Stem Cell Center
bp	base pair
Cas	CRISPR-associated
Cas9	CRISPR-associated protein 9
CFD	cutting frequency determination
CNV	copy number variants
CRISPR	clustered regularly interspaced short palindromic repeats
CRISPR/Cas9	clustered regularly interspaced short palindromic repeats/CRISPR-associated protein 9
crRNA	CRISPR RNA
DPBS	Dulbecco's Phosphate-Buffered Saline
DMSO	dimethyl sulfoxide
DSB	double-strand break
EDTA	ethylenediaminetetraacetic acid
GWAS	genome-wide association studies
HDR	homology-directed repair
ID	intellectual disability
IDT	Integrated DNA Technologies
indels	insertions/deletions
INTS6	Integrator complex subunit 6
iPSC	induced pluripotent stem cell
KSR	knock-out serum replacement
NDD	neurodevelopmental disorder
NFID	Northern Finnish Intellectual Disability

NHEJ	non-homologous end-joining
PAM	protospacer-adjacent motif
RNP	ribonucleotide protein
ROCK	rho-associated coiled-coil containing protein kinase
SCHEMA	schizophrenia exome meta-analysis
SCZ	schizophrenia
sgRNA	single-guide RNA
SpCas9	*Streptococcus pyogenes* Cas9
ssDNA	single-stranded DNA
ssODN	single-stranded oligodeoxynucleotide
tracrRNA	trans-activating CRISPR RNA
T7E1	T7 Endonuclease 1 assay
TE	Tris-EDTA
WES	whole exome sequencing

1 Introduction

The inaccessibility of human brain tissues and the extremely complex neurobiology of neurological processes poses a challenge in understanding the molecular mechanisms of complex brain-related disorders, such as schizophrenia (SCZ). Both environmental and heritable genetic factors contribute to the liability of complex disorders (Mitchell, 2012). Complex disorders are polygenic, meaning that most of the genetic liability is explained by common genetic variants with small effects that synergistically contribute to the genetic risk (Bassett, 2017; Kurishev et al., 2023; Mitchell, 2012). So far, the molecular etiology of brain-related disorders is poorly understood, and most of the information is derived from animal model organisms. Genetic association studies, systematically screening the whole genome, have provided new knowledge of the causative biological pathways underlying brain disorders. The identified genetic variants when combined with human cellular models offer completely new ways for studying causative biology in human cell types (Bassett, 2017; Kurishev et al., 2023). Human induced pluripotent stem cells (iPSCs) are embryonic-like stem cells that have the capacity to differentiate into any other cell type, like neurons which might be difficult to access otherwise. Patient cell lines carry the donor's disease-relevant genetic background, generating a clinically relevant disease model (Cerneckis et al., 2024).

Over the last decades, innovations in genome editing technologies have unlocked the ability to precisely alter the genomic sequences in a targeted manner. Genome editing is defined as precise manipulation and "rewriting" of genetic material in a simple, programmable, and efficient manner that primarily relies on the DNA repair mechanisms of the cells (Pacesa et al., 2024). Engineered nucleases, enzymes that cleave DNA in targeted genomic regions, were first developed as gene editing tools since the beginning of the 21st century (Pacesa et al., 2024). These set the ground for the emergence of the Nobel-winning technology called CRISPR (clustered regularly interspaced short palindromic repeats). CRISPR is a versatile, precise, efficient, and easy to design genome editing tool that can be used to study the causality of genetic variation in complex diseases. Although the majority of the genetic liability in polygenic disorders such as SCZ is conferred by the aggregate effect of common variants with individually small effects (Trubetskoy et al., 2022), some individuals carry high-penetrant coding genetic variants that directly point to an affected gene and a molecular mechanism (Genovese et al., 2016; Singh et al., 2022). Therefore, these rare variants provide a strong starting point for gene editing and disease modelling to understand the biological processes underlying complex brain disorders. In this thesis, I will combine CRISPR gene editing with the benefits of iPSC technology for rare high-risk genetic variants that have recently been identified in Finnish patient cohorts.

1.1 CRISPR/Cas9 genome editing technology

1.1.1 History of CRISPR/Cas9 system

CRISPR genetic elements were first discovered in E. coli in 1987 by Ishino et al. (1987), but their biological function was uncovered only a decade later, with similar sequences observed in other prokaryotes (Mojica et al., 2000). These elements consist of an array of short repeated palindromic sequences interrupted by spacer sequences in the bacterial genome (Jansen et al., 2002). Bolotin et al. (2005) and Mojica et al. (2005) noticed homologous sequences between the CRISPR loci and extrachromosomal elements such as bacteriophage DNA and plasmids. This led to the discovery that they play a role in the bacterial defensive mechanism by capturing and storing snippets of foreign genetic material, called protospacers, in the host's CRISPR genomic locus to develop adaptive immunity after phage infection (Barrangou et al., 2007).

Conserved CRISPR-associated (Cas) genes were identified near the CRISPR loci, suggesting a functional association with the integration of the repeated sequences and the CRISPR-mediated immunity mechanism (Jansen et al., 2002; Bolotin et al., 2005). Barrangou et al. (2007) further supported this relationship by revealing that the Cas proteins comprise an enzymatic machinery which is involved in the CRISPR-mediated immunity. The recognition and interference with the foreign DNA were shown to be mediated by Cas proteins guided by short RNAs transcribed from the CRISPR loci, called CRISPR RNAs (crRNAs) (Brouns et al., 2008). Sapranauskas et al. (2011) later confirmed that only the CRISPR-associated protein 9 (Cas9) was necessary for the cleavage of the extrachromosomal genetic material because of its endonuclease activity. Moreover, a second small RNA, trans-activating CRISPR RNA (tracrRNA), was discovered to pair with the crRNA and activate the Cas9 (Deltcheva et al., 2011). Jinek et al. (2012) confirmed that this dual crRNA-tracrRNA can guide the Cas9 in a programmable manner to introduce site-specific DNA breaks. Following this discovery, further studies employed the CRISPR/Cas9 system in human and mouse cells and accomplished to introduce site-specific genetic modifications (Jinek et al., 2013; Cong et al., 2013; Cho et al., 2013; Mali et al., 2013). These discoveries set the groundwork for the development of an RNA-programmable genome editing technology that has a wide range of applications.

1.1.2 CRISPR/Cas9 genome editing molecular mechanism

Successful genome editing with the CRISPR/Cas9 system depends on the occurrence of two key events in the cellular level: the induction of a site-specific double-strand break (DSB) and its subsequent correction by the endogenous DNA repair pathways (Pacesa et al., 2024).

The Cas9 nuclease naturally derived from *Streptococcus pyogenes* (SpCas9) is the most used endonuclease for genome editing purposes (Pacesa et al., 2024). SpCas9 forms a ribonucleoprotein (RNP) complex with a base-paired RNA structure formed by the crRNA-tracrRNA pairing (Jinek et al., 2012). The tracrRNA is loaded on the endonuclease for enzymatic activation, while the 5' end of the crRNA provides a 20-nucleotide long base-pairing complementarity with the target sequence in the genome to guide the SpCas9 (Jinek et al., 2012). The tracrRNA-crRNA complex can be also substituted by a single guide RNA (sgRNA) with combined properties, thus making the system more versatile and easier to design (Jinek et al., 2012).

The site-specific binding of the SpCas9 in a double-strand DNA (dsDNA) molecule is defined by the complementary base-pairing of the sgRNA to a protospacer sequence, which is the target sequence in the genome. The presence of a protospacer adjacent motif (PAM) immediately downstream of the protospacer at the non-target strand is also necessary for this recognition (Jinek et al., 2012; Figure 1). For the SpCas9 endonuclease, the PAM sequence is 5'-NGG-3', where N denotes any nucleotide (Jinek et al., 2012). A PAM sequence is estimated to occur every 8 base pairs (bp) in the genome (Bassett, 2017), providing numerous naturally occurring, potential editing sites in the human genome. When the SpCas9 recognizes the GG dinucleotide, the dsDNA is locally denatured and the sgRNA forms a heteroduplex with the target sequence (Jinek et al., 2012). If complementarity between the sgRNA and the protospacer is achieved, subsequent structural changes of the SpCas9 activate its endonuclease catalytic domains (Jinek et al., 2012; Pacesa et al., 2024). SpCas9 is a large protein consisting of six domains, two of which are nuclease domains that catalyse the DSB DNA cleavage (Xue & Greene, 2021). The HNH-like domain cleaves the target strand at precisely three nucleotides upstream of the PAM site, whereas the RuvC-like domain acts on the non-target strand within three to eight nucleotides upstream of the PAM (Jinek et al., 2012; Xue & Greene, 2021). Therefore, a Cas9-induced DSB is introduced either with blunt ends (both strands cleaved at the same site), or with sticky ends (strands cleaved at different positions) resulting in 5'-end nucleotide overhangs on the non-target strand (Jinek et al., 2012).

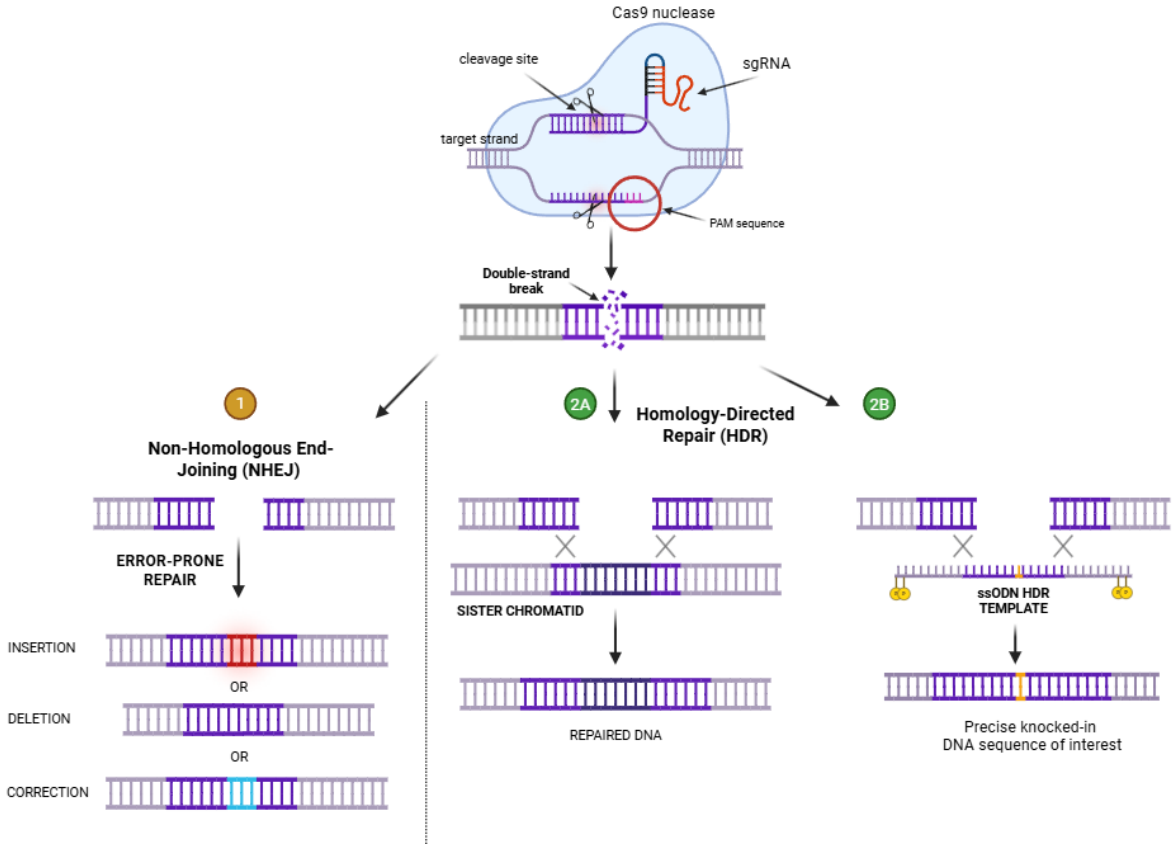

Figure 1. Schematic of the Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-associated (CRISPR/Cas) genome editing mechanism and the subsequent common endogenous DNA repair machineries. The CRISPR-associated protein 9 (Cas9) forms a ribonucleoprotein (RNP) complex with a single guide RNA (sgRNA). The sgRNA recognizes a specific genomic sequence via base-paired complementarity and activates the Cas9, which induces a double strand break (DSB) three base pairs upstream to a protospacer-adjacent motif (PAM) sequence. The DSB then triggers the endogenous DNA repair mechanisms. The error-prone non-homologous end-joining (NHEJ) repair pathway rapidly fixes the DNA lesion, but can result in short random insertions and deletions, whereas the template-dependent homology-directed repair (HDR) pathway is more precise and utilizes the sister chromatid or an exogenous donor DNA template with homology arms to repair the DSB. NHEJ is the predominant repair mechanism, while HDR is mainly active during the G2/S phase of the cell cycle. Figure created with Biorender.com.

1.1.3 Two DNA repair pathways mediate the resolution of the Cas9-induced DSB

The resolution of the Cas9-induced genomic DSB is the second critical point that determines the success of precise gene editing. DSBs are a type of DNA lesions that are highly genotoxic and naturally activate endogenous repair mechanisms of the cell to maintain the integrity of the genome (Xue & Greene, 2021). Unrepaired DSBs can cause cell death, chromosomal rearrangements, and DNA loss (Mao et al., 2008; Xue & Greene, 2021). In mammalian cells, DSBs mainly recruit two major competing DNA repair machineries that differ in precision and template requirements, and which properties can be used to alter the DNA sequences in a targeted manner (Xue & Greene, 2021; Figure 1).

The primary mechanism for repairing a DSB is the error-prone non-homologous end-joining (NHEJ) pathway (Xue & Greene, 2021). Although imprecise, the NHEJ pathway has the advantage of being rapid, active in all cell-cycle phases, and template-independent, which allows the immediate correction of the lesion without risking extensive DNA damage (Mao et al., 2008; Xue & Greene, 2021). The NHEJ mechanism initially involves the recognition of the DSB by a protein heterodimer known as Ku70/80 (Chang et al., 2017). This heterodimer protects the cleaved DNA and recruits additional proteins of the NHEJ machinery, including DNA ligases, to ligate the DNA lesion without using a repair template (Chang et al., 2017). Cas9 can introduce DSBs with either blunt ends or nucleotide overhangs (Jinek et al., 2012). Error-free repair can only be achieved when there are blunt ends that are directly re-ligated by DNA ligases without adding or removing any nucleotides (Elton et al., 2022). If nucleotide overhangs are introduced by the Cas9, additional proteins are then required to process the DNA ends and generate DNA ends that can be ligated (Xue & Greene, 2021). The processing of the DNA ends without a DNA repair template often results in nucleotide insertions or deletions (indels), which can introduce frameshift mutations, affecting the function of genes (Elton et al., 2022). Given the high error rates in the NHEJ pathway including frameshift indels, NHEJ is often the preferred mechanism for editing when the goal is to disrupt gene function by a knock-out mutation (Elton et al., 2022).

The alternative and more precise homology-directed repair (HDR) pathway is mediated by using a donor template DNA as a reference sequence to repair the DNA lesion (Xue & Greene, 2021). This accurate, high-fidelity pathway relies on the complementarity between the cleaved DNA molecule and a sister chromatid during the G2/S phase of the cell cycle (Xue & Greene, 2021). This homology directs the HDR protein machinery to repair the DSB via homologous recombination. With the same principle, exogenous single-stranded or double-stranded DNA (ssDNA or dsDNA) molecules with homology arms that contain a desired genetic modification can be introduced in the cell and used as DNA template by the HDR machinery (Elton et al., 2022). Therefore, when the intention is to precisely modify a defined genomic locus (e.g. targeted mutation or correction of the genotype) via a genetically designed repair template, the error-free HDR pathway should be favoured over the NHEJ pathway. In this project the goal was to either introduce or correct specific genetic variants, so strategies to improve HDR efficiency were employed.

1.1.4 Major challenges of the CRIPSR/Cas9 system

The CRISPR/Cas9 system has been repurposed from a bacterial immune defence mechanism to serve as a genome editing tool in the more intricate genetic landscape of eukaryotic cells

(Jinek et al., 2012). Despite its versatility and robustness, this technology contains several functional constraints that need to be considered before designing a CRISPR experiment.

The most well-studied limitation of the CRISPR/Cas9 system for genome editing purposes is its specificity. CRISPR/Cas9 is not always specific and can inherently tolerate mismatches between the sgRNA and the protospacer in a sequence-dependent manner (Fu et al., 2013; Hsu et al., 2013). A few mismatches (usually up to five) at the PAM-distal nucleotides towards the 5' end of the sgRNA can still result in Cas9 cleavage at unintended off-target sequences (Hsu et al., 2013; Fu et al., 2013). The first 10-12 nucleotides of the sgRNA are the most critical for Cas9 cleavage. Mismatches in this seed region of the sgRNA usually block the activity of the Cas9 (Jinek et al., 2012; Fu et al., 2013). However, Hsu et al. (2013) demonstrated that off-target activity is not exclusively dependent on the position of the mismatches between the sgRNA and the protospacer, but is also defined by the sequence, the nucleotide preference, and the total number and spacing of mismatches. The off-target activity of the CRISPR/Cs9 system may result in chromosomal rearrangements or gene knockouts and leads to cellular stress and DNA damage (Pacesa et al., 2024). Therefore, when designing a sgRNA, off-targets which contain less than 3-4 mismatches, particularly those located in exons, should be avoided.

Another downside is the PAM-dependent genome accessibility of the Cas9. The proximity of the DSB to the target site plays a pivotal role in HDR efficiency. Several studies suggest that this distance should be kept within 10 nucleotides, with this being a flexible guideline rather than a strict threshold (Bruntraeger et al., 2019; Schubert et al., 2021; Liang et al., 2017). However, the accurate positioning of the SpCas9 to the target site is dependent and constrained on the availability of the 5'-NGG-3' PAM sequence nearby (Walton et al., 2020). If no 'NGG' PAM sequence is available near the edit site, the SpCas9 cannot access the genomic locus to induce a DSB.

Furthermore, the Cas9-induced DSB can be toxic for the cell and by triggering p53-mediated apoptosis (Ihry et al., 2018; Pacesa et al., 2024). p53 is a master regulator protein involved in the apoptotic pathway and is mainly activated upon DNA damage and cellular stress (Lane, 1992). Knocking-out or suppressing the p53 gene in CRISPR-edited cells was found to significantly enhance cell viability (Ihry et al., 2018). Both the on-target and off-target activity of Cas9 can lead to DNA damage and genotoxicity, which may impact the success of the CRISPR/Cas9 gene editing (Ihry et al., 2018).

1.2 Strategies to improve HDR efficiency in iPSCs

1.2.1 sgRNA considerations

To overcome the PAM requirement at a proximal position from the edit site, alternative nucleases orthologs from other species (e.g., Cas12) that recognize different PAM sequences can be utilized (Elton et al., 2022; Pacesa et al., 2024). This allows of targeted modification of nearly any site in the genome. Moreover, structurally engineered variants of the SpCas9 endonuclease have been developed over the years (Kleinstiver et al., 2015; Pacesa et al., 2024; Walton et al., 2020). These engineered endonucleases have a relaxed PAM requirement by recognising more PAM sequences (Walton et al., 2020). They can be directed to genomic loci previously inaccessible by the wild-type SpCas9, providing the opportunity to edit more disease-causing genetic variants (Elton et al., 2022; Walton et al., 2020). Nonetheless, their ability to access more genomic loci associates with reduced cleavage efficiency and specificity compared to SpCas9, which needs to be considered when selecting the appropriate nuclease for each CRISPR experiment (Pacesa et al., 2024; Walton et al., 2020). In addition, high fidelity variants of the wild-type SpCas9 (e.g., HiFi SpCas9) have been engineered to substantially reduce the off-target activity of the nuclease (Pacesa et al., 2024).

The cleavage efficiency and specificity are major factors that determine the selection of the sgRNAs. The guide RNA is not only responsible for the target site recognition, but also for activating the Cas9 endonuclease activity (Jinek et al., 2012). There are several online software tools based on algorithms that predict the effectiveness of selected sgRNAs, as detailed in Section 1.2.2.

To date, synthetic, chemically modified sgRNAs that form a RNP complex with the Cas9 have been shown to be less toxic and more stable compared to the crRNA:tracrRNA complex or the *in vitro*-transcribed sgRNAs (Bruntraeger et al., 2019; Elton et al., 2022). Truncated 17-18 nucleotide-long sgRNAs can significantly improve the specificity of the system, although they may reduce efficiency (Fu et al., 2014). For this project, full length 20 nucleotide-long synthetically designed sgRNAs were used to maximize the site-specific Cas9 cleavage efficiency.

1.2.2 In silico computational methods predict sgRNA efficiency and specificity

Even if a PAM sequence is located near a genomic target site, the selection of the corresponding sgRNA is determined by its efficiency and specificity, which vary greatly among sgRNAs (Cui et al., 2018). For this reason, more than 30 online sgRNA design tools that depend on large-scale experimental datasets have been developed over the last decade (Cui et al., 2018). It has

been established that intrinsic features of the sgRNA sequence and the target site, such as nucleotide biases in specific positions, considerably impact the sgRNA activity (Doench et al., 2014). Local chromatin structure and sequence-specific patterns can also hinder sgRNA efficiency by affecting the ability of the Cas9 to recognise the PAM sequence, which suggests that coding regions should be preferred for sgRNA design (Doench et al., 2014). Guanine-rich sgRNAs have been found to be associated with increased structural stability and activity (Moreno-Mateos et al., 2015), whereas very high or low GC content of the sgRNA is linked to reduced sgRNA efficiency (Doench et al., 2014). Building on previous algorithms, Doench et al. (2016) developed a more accurate and robust predictive model by incorporating additional features, such as thermodynamic properties, which have been the most informative for sgRNA design.

Similarly, off-target detection algorithms have been developed to calculate scores for quantifying sgRNA specificity (Cui et al., 2018). The off-target score is correlated with the specificity of the sgRNA and its potential to guide the Cas9 to unintended genome sites due to mismatch tolerance between the sgRNA and the target DNA sequence (Cui et al., 2018). Among several predictive algorithms, the cutting frequency determination (CFD) score developed by Doench et al. (2016) is reported to perform better than other scoring algorithms (Cui et al., 2018; Haeussler et al., 2016).

Despite their significant contribution to the optimization of sgRNA design, these prediction scores offer limited reliability and modest correlation to the actual sgRNA activity when tested *iv vitro* (Bruntraeger et al., 2019; Haeussler et al., 2016). However, they provide a useful estimate on whether certain sgRNAs are predicted to be highly ineffective and can save time and resources by reducing the number of sgRNAs to be screened (Haeussler et al., 2016). Therefore, it is recommended to experimentally test at least 2-3 sgRNAs per edit to maximize the chances of successful Cas9-mediated DSB in the genome by at least one sgRNA (Bruntraeger et al., 2019; Haeussler et al., 2016).

1.2.3 Delivery methods

The delivery of the CRISPR components differs based on the application and the model system (Pacesa et al., 2024). Viral vectors have been widely used *iv vivo,* whereas for *in vitro* applications electroporation/nucleofection and lipofection are the most preferred delivery methods because of their low immunogenicity and higher efficiency (Pacesa et al., 2024). In iPSCs, the *in vitro* pre-assembly of the Cas9-sgRNA RNP complex prior to nucleofection allows for a more rapid and transient delivery compared to transfection via mRNA or DNA plasmids (Skarnes et al., 2019; Schubert et al., 2021; Pacesa et al., 2024). The Cas9 RNP is ready-to-use and is depleted within 24 hours which reduces off-target effects, minimizes toxicity, and

excludes the possibility of DNA integration by plasmids (Pacesa et al., 2024; Skarnes et al., 2019).

1.2.4 HDR template design

Stimulation of the HDR pathway over the error-prone NHEJ is required to introduce precise template-dependent edits into the genome. To enhance HDR, several approaches have been proposed regarding the genetic composition, orientation, length, and chemical properties of the repair DNA template (Pacesa et al., 2024). Single-stranded oligonucleotide (ssODNs) repair templates are generally more efficient and specific compared to their double-stranded counterparts because they facilitate higher recombination rates during HDR, which increases the likelihood of their successful integration in the genome (Liang et al., 2017). Chemical modifications, in particular phosphorylation, at the two first and last nucleotides of the template are associated with increased stability and genome editing efficiency (Liang et al., 2017). Homology arms complementary to the DNA around the target site are essential for successful recombination during HDR. For short indels and point mutations, 80-100 nucleotide-long symmetric ssODN repair templates are more effective in iPSCs (Skarnes et al., 2019; Bruntraeger et al., 2019). Moreover, the optimal strand preference of the repair template relies on the proximity and position of the DSB to the target site in human iPSCs (Bruntraeger et al., 2019; Schubert et al., 2021). Okamoto et al. (2019) reported that the antisense ssODN (complementary to the non-target strand of the sgRNA) is preferred for single-base substitutions, while Bruntraeger et al. (2019) recommended that for edits within 5 base pairs from the Cas9 cleavage site either strand is preferred.

To prevent repeated editing of the same locus after successful HDR, the repair template can be genetically designed to contain blocking silent mutations. Once the site-specific successful integration of the repair template in the genome has been achieved, the Cas9 RNP can still bind to the PAM sequence with base-pairing complementarity and re-cleave at the same locus (Okamoto et al., 2019). This event significantly drops HDR efficiency and is established that silent (synonymous) mutations of the DNA repair template that disrupt the PAM sequence block the activity of the nuclease and enhance HDR (Jinek et al., 2012; Okamoto et al., 2019; Schubert et al., 2021). In case no silent PAM mutations can be introduced, it is sufficient to add at least three silent mutations as close as possible to the PAM to generate mismatches in the seed region of the sgRNA (Okamoto et al., 2019).

1.2.5 Experimental methods to enhance HDR

In mammalian cells, the end joining repair pathway is predominant and suppresses HDR, although during the late S/G2 phase of the cell cycle the HDR can prevail due to the presence

of homologous DNA template (Xue & Greene, 2021). Therefore, strategies to favour HDR over NHEJ focus on the cell culture media supplementation with NHEJ inhibition factors, or HDR enhancers to increase the chances of successful repair template integration (Skarnes et al., 2019). To mitigate the cell death caused by the DSB or the electroporation procedure itself, transient inhibition of apoptotic factors, such as p53, is an efficient method to improve HDR and increase the number of edited iPSCs (Ihry et al., 2018; Pacesa et al., 2024). In addition, a brief "cold shock" treatment typically involving incubation of cells at +32°C for 24-48 hours after electroporation has been shown to further increase Cas9 activity in human iPSCs (Guo et al., 2018; Skarnes et al., 2019). Maurissen and Woltjen (2020) reported that cold shock delays the G2/M phase progression in human iPSCs, which improves HDR, and might contributes to increased stability of the nuclease and the repair templates.

Overall, these improvements, ranging from the sgRNA and the repair template design to pharmacological treatments, have a synergistic effect on the CRISPR/Cas9 mechanism and can significantly improve the success rates of gene editing experiments.

1.3 Disease modeling with human iPSCs

1.3.1 iPSCs

To decipher the molecular mechanisms underlying neuropsychiatric disorders and to identify the causation of specific genetic changes to the disease phenotype, experimental cellular models are required. Model organisms such as rodents and other primates are extensively used to study human diseases because they recapitulate the *in vivo* physiological environment and allow for complex cell-tissue interactions which are not present in cellular *in vitro* models (Bassett, 2017). However, the evolutionary genetic divergence and their inability to capture the intricate genetic, structural, and phenotypic landscape of the human brain necessitates the use of human cellular models of disease (Bassett, 2017; Cerneckis et al., 2024). Primary cells derived from patients offer an advantageous resource, especially for drug discovery and personalized medicine, but they come with critical downsides. Most cell types, particularly in the brain and heart, are largely inaccessible, they have lost their self-renewal properties and can be highly heterogeneous when introducing them in the cell culture, restricting their use for modelling human diseases (Bassett, 2017; Cerneckis et al., 2024). Although human embryonic stem cells overcome these limitations, ethical concerns and technical challenges pose significant hurdles (Bassett, 2017; Cerneckis et al., 2024).

In 2006, the discovery that pluripotent stem cells can emerge by reprogramming human somatic cells back to their stem cell state has brought substantial opportunities in the field of disease modelling (Takahashi & Yamanaka, 2006). iPSCs were first generated from human

fibroblasts after retroviral induction with four transcriptional factors (Sox2, Oct4, c-Myc, Klf4) known for their function as pluripotency markers (Takahashi et al., 2007; Takahashi & Yamanaka, 2006). iPSCs have the capacity to self-renew and differentiate to any cell type that might be inaccessible otherwise, such as neuronal cell types (Bassett, 2017). Patient-derived iPSCs also retain the donor's genetic risk profile and possibly the pathogenic cellular phenotype of the donor, offering a powerful source to study complex genetic disorders (Bassett, 2017; Cerneckis et al., 2024). In addition, iPSC-derived differentiated neurons resemble the maturational stage of human fetal-stage neurons (Bassett. 2017). This is particularly valuable in neurodevelopmental disorders because the disease processes in cells can be studied from the early developmental stages, even in conditions where symptoms typically become apparent later in life (Bassett, 2017; Kurishev et al., 2023).

1.3.2 CRISPR gene editing creates isogenic iPSC-derived cellular models

Patient-derived iPSCs contain the full genetic context of the disease, where genetic variants interact and drive disease progression (Bassett, 2017). Genome engineering with CRISPR/Cas9 offers the powerful advantage of modelling genetic variants while keeping the genetic background untouched (Bassett, 2017; Kurishev et al., 2023). Editing targeted sites in the same genetic background generates isogenic cell line pairs, meaning that they are genetically identical but differ only in a single locus (Bassett, 2017). Isogenic pairs minimize genetic variability by isolating single-gene effects, providing the optimal experimental setting to unravel the causative effect of genetic variants (Bassett, 2017).

The cause-and-effect relationship and the relevant contribution of a genetic variant can be validated by employing two complementary genome editing approaches. To determine whether a specific variant alone is enough to cause the disease, the introduction of the mutation in wild-type genetic background is required (Bassett, 2017). On the other hand, isogenic correction of the variant in patient-derived iPSCs, which already manifest the disease phenotype, informs about its contribution (Bassett, 2017). Eventually, comparing the phenotypic outcomes of healthy and patient-derived cells with and without the mutation establishes the relative effect of a risk variant to the disease progression (Kurishev et al., 2023).

Because of clonal variability and the off-target mutagenesis of the CRISPR/Cas9 gene editing mechanism, edited clones may not be completely isogenic (Panda et al., 2023). Accumulation of mutations during culturing and off-target edits results in mosaicism in iPSCs, which affects the reliability of the results (Panda et al., 2023). One way to counteract this limitation is to include several biological replicates and examine multiple comparisons of samples to determine if the studied genetic variant is indeed causative (Bassett, 2017). In addition, further

characterization of the CRISPR-edited iPSCs and off-target screening should be employed before proceeding to downstream applications such as differentiation protocols (Panda et al., 2023).

1.4 The genetics of SCZ as an example to study causative genetic variants

1.4.1 Gene discovery in SCZ and ID

Schizophrenia (SCZ) is a complex, highly heritable, and polygenic psychiatric disorder (Nakamura & Takata, 2023). SCZ was hypothesized to have a neurodevelopmental component (Weinberger, 1987; Murray & Lewis, 1988). Epidemiological and developmental studies further support this model, suggesting that abnormal molecular events in early brain development lead to increased susceptibility to SCZ, before the symptoms become apparent (Owen et al., 2011). Neurodevelopmental disorders (NDDs) which typically arise earlier in life, such as intellectual disability (ID), autism spectrum disorders, and attention-deficit/hyperactivity disorders, share pathogenic phenotypes with SCZ (Owen et al., 2011; Owen and O'Donovan, 2017). Recent studies have also demonstrated that many of the genetic factors are shared across psychiatric disorders and NDDs (Marshall et al., 2017; Singh et al., 2022). This genetic sharing is further supported by high comorbidity between SCZ and ID in epidemiological studies in line with shared clinical etiology (Kurki et al., 2019; Song et al., 2022).

Epidemiological family studies report that the genetic heritability of SCZ ranges between 60 – 80% (Jauhar, 2022). SCZ follows a polygenic pattern of inheritance with multiple genetic variants of various frequencies and effect sizes contributing to the disorder's genetic susceptibility (Nakamura & Takata, 2023). During the last decade, advancements in genotyping and sequencing technologies have substantially contributed to the understanding of the genetic architecture of SCZ (Nakamura & Takata, 2023). Large-scale genome-wide association studies (GWAS) have identified numerous common single nucleotide polymorphisms (minor allele frequency >1%) with individually small effects, which account for most of the genetic liability in SCZ (Ripke et al., 2014; Trubetskoy et al., 2022). Each of these low-impact common variants contribute only little to the disease risk, but they jointly comprise the individual's polygenic score for the disorder (Trubetskoy et al., 2022). In addition, common low-risk variants in rare severe NDDs, including ID, have a significant genetic overlap with SCZ and cognitive impairment (Niemi et al., 2018).

Rare copy number variants (CNVs) have also been reported to confer substantial risk in SCZ (Rees et al., 2016; Marshal et al., 2017). CNVs are large genomic deletions or duplications, ranging from a few dozen base pairs to hundreds of kilobases. Eight rare recurrent CNVs have

been significantly implicated in SCZ, although they explain less than 1% of the overall genetic risk (Marshal et al., 2017). In the same context, the genetic burden of CNVs in ID and other NDDs also confer substantial risk for SCZ, further supporting the neurodevelopmental basis of SCZ (Rees et al., 2016).

More importantly, rare and ultra rare (minor allele frequency <0.1% and <0.01% respectively) deleterious coding variants with individually large effect sizes from whole exome sequencing (WES) studies have been discovered in SCZ cases (Genovese et al., 2016; Singh et al., 2016; Singh et al., 2017; Singh et al., 2022). Most of these rare variants also show pleiotropic effects, which means that they are involved in other psychiatric conditions and NDDs, including ID (Singh et al., 2017; Niemi et al., 2018; Song et al., 2022). Because of their deleteriousness, negative natural selection combined with the reduced reproductive success of affected individuals keep their frequencies very low in the population level (Singh et al., 2022). Consequently, rare variants only explain a small portion of the genetic heritability, compared to common variants (Singh et al., 2022). Nonetheless, such high-impact loss-of-function genetic variants are particularly informative and directly nominate the affected genes to study their cause-and-effect relationship with the disorder (Kurishev et al., 2023). So far, ultra-rare coding variants in 10 highly constrained genes (*CACNA1G, CUL1, HERC1, GRIN2A, GRIA3, TRIO, RB1CC1, SP4, SETD1A, XPO7*) have been identified with exome-wide significance in SCZ (Singh et al., 2022; Table 1), providing a strong starting point for disease modelling.

2 Research aims

The main scope of this thesis was to explore the utility of CRISPR/Cas9 genome editing technology in disease modelling to investigate the cause-and-effect relationship of high-risk variants in genes associated with SCZ and ID. I initially screened an iPSC resource from carriers with natural rare deleterious variants to design sgRNAs and DNA repair templates for isogenic experiments. Then, as proof of concept, I generated isogenic pairs of CRISPR-edited iPSCs *in vitro*. Hence, my specific research objectives were:

1. To develop an optimized *in silico* CRISPR/Cas9 gene editing design methodology to create isogenic pairs that carry natural rare high-penetrant coding variants in genes associated with SCZ at exome-wide significance.

2. To generate isogenic pairs of CRISPR/Cas9-edited iPSC lines to model one novel deleterious variant. The following edits were performed:

    a. Isogenic correction to restore the wild-type allele in patient-derived iPSCs via the template-dependent HDR pathway.

    b. Isogenic knock-in to introduce the pathogenic loss-of-function variant in wild-type iPSCs in heterozygosity via the HDR pathway. Heterozygous cells can also occur via the error-prone NHEJ pathway.

3 Materials and Methods

3.1 Acquisition of genetic variant data and iPSC lines

The genetic variant data used for the CRISPR design were collected from the Finnish cohort of psychosis patients, called SUPER. In total, 18 rare deleterious variants in 8 out of the top 10 genes associated with SCZ at exome wide significance have been identified (no variants available for *GRIN2A, GRIA3*). Two edits were designed for each variant: one to correct the pathogenic variant back to wild type (isogenic rescue), and one to introduce it into a wild-type genetic background (knock-in). Table 1 lists the variant information from the SUPER cohort.

Table 1. Genetic variant information from the SUPER cohort used to design sgRNAs.

The MPC pathogenicity score (MPC = Missense badness, PolyPhen-2, and Constraint) estimates the deleteriousness of the missense variant. Values greater than 3 indicate deleterious effect.

Gene Name / Ensembl ID	Variant	Variant type	Gene function
CACNA1G / ENSG00000006283	chr17:50624357:CAGG:C	Splice acceptor	Ion transport
CUL1 / ENSG00000055130	chr7:148790376:T:C	Missense (MPC>3)	Ubiquitin ligation
HERC1 / ENSG00000103657	chr15:63637595:CTG:C	Frameshift	Ubiquitin ligation
	chr15:63656291:G:A	Stop-gain	
	chr15:63713615:G:A	Stop-gain	
	chr15:63746986:G:A	Stop-gain	
	chr15:63693963:C:T	Splice donor	
	chr15:63655874:G:A	Stop-gain	
	chr15:63775279:A:C	Stop-gain	
RB1CC1 / ENSG00000023287	chr8:52657117:C:CTGAG	Frameshift	Transcriptional regulation
SETD1A / ENSG00000099381	chr16:30966303:C:T	Stop-gain	Transcriptional regulation
	chr16:30963433:G:T	Missense (MPC>3)	
	chr16:30966112:C:CA	Frameshift	
	chr16:30961466:AG:A	Frameshift	
SP4 / ENSG00000105866	chr7:21429774:TAATC:T	Frameshift	Transcriptional regulation
	chr7:21481973:G:T	Stop-gain	
TRIO / ENSG00000038382	chr5:14472658:G:GGTGTGTA	Frameshift	Neuronal migration and growth
XPO7 / ENSG00000130227	chr8:21976513:G:A	Stop-gain	Nuclear transport

To address the second research aim and as a proof-of-concept, two human iPSC lines were utilized: one derived from a patient which is heterozygous for a deleterious coding mutation in the Integrator complex subunit 6 (*INTS6)* gene, and the other from an unaffected non-carrier of the same family, identified in the Northern Finnish Intellectual Disability (NFID) cohort (Figure 2). In this family, seven carriers of the same *INTS6* variant exhibit mild ID and learning difficulties, suggesting a strong causation of the gene in the etiology of ID.

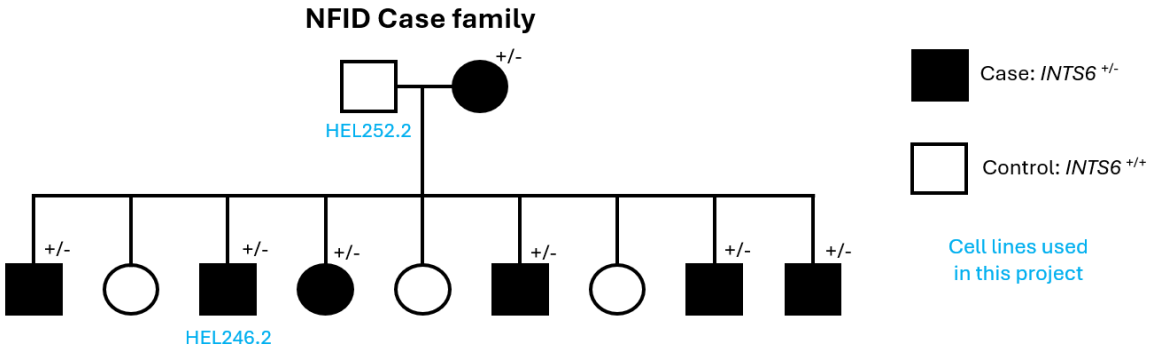

Figure 2. Pedigree (modified for anonymity) of a family identified in the Northern Finnish Intellectual Disability (NFID) cohort with seven heterozygous carriers of a novel rare loss-of-function coding mutation in the *INTS6* gene (chr13:51378376:G:A). Carriers of the pathogenic coding variant exhibit mild intellectual disability (ID) symptoms. Induced pluripotent stem cells (iPSCs) from the unaffected father (HEL252.2) and one affected son (HEL246.2) were used in this project.

HEL252.2 derived from the unaffected father was used as the wild-type (*INTS6*$^{+/+}$) cell line for knock-in of the pathogenic variant. HEL246.2 cell line from an affected son with mild ID symptoms was selected to rescue the pathogenic genotype. HEL246.2 carries the heterozygous (*INTS6*$^{+/-}$) loss-of-function stop-gain mutation chr13:51378376:G:A at the reverse strand of exon 12 of the *INTS6* gene, which results in a premature termination codon. Table 2 provides an overview of the cell lines used in this project.

Table 2. Features of the cell lines used for *in vitro* CRISPR experiment.

The passage number of each cell line is on the day of electroporation.

Cell line ID	Passage	Clinical condition	Genotype	Expected outcome
HEL252.2	27	Unaffected – no ID	*INTS6*$^{+/+}$	Knock-in of the variant allele
HEL246.2	26	Mild ID	*INTS6*$^{+/-}$	Correction of the *INTS6* chr13:51378376:G:A mutation

The iPSC lines were previously reprogrammed from peripheral blood mononuclear cells at Harvard Stem Cell Institute, Cambridge, MA, USA and Biomedicum Stem Cell Center (BSCC), University of Helsinki, Finland. All study participants and donors provided an informed written consent. The studies were approved by the ethical review boards of University of

Helsinki Hospital and the Hospital district of Northern Ostrobothnia. The cell lines had been previously reprogrammed and cultured in feeder-free conditions by the iPS Core Facility at the Biomedicum Stem Cell Center (BSCC). The clones have been characterised by expression of stem cell markers, normal karyotype, morphology and growth characteristics consistent with iPSCs and therefore meet the common criteria for iPSC validation. The clones were Sanger sequenced to confirm the *INTS6* genotype.

3.2 *In silico* CRISPR design

3.2.1 Design of the sgRNAs and the ssODN HDR templates

The sgRNAs and the DNA repair templates for each variant were designed in Benchling using the CRISPR Design Tool (https://www.benchling.com/crispr). All sgRNAs were also screened using the CRISPOR program (https://crispor.gi.ucsc.edu) developed by Concordet and Haeussler, 2018. CRISPOR designs, evaluates, and ranks guide RNA sequences based on their predictive efficiency and specificity for CRISPR/Cas experiments (Concordet & Haeussler, 2018).

The reference sequences for all designs were imported from the human genome GRCh38 (Ensembl genome browser release 113) based on the gene names listed in Table 1. For each knock-in experiment, the imported reference sequence matches the wild-type allele. To design the sgRNAs for the rescue experiments, the same reference sequence of each gene was manually edited to include the pathogenic variant and match the patient genotype to be corrected.

For each edit to be introduced, at least two 20 nucleotide-long sgRNAs were selected. First, all possible PAM sequences around the target site were identified. The sgRNAs were then designed and ranked by order based on i) the proximity of the Cas9 cleavage site and the edit locus, ii) the predicted cleavage efficiency of the Cas9 (on-target score), iii) the predicted specificity of the sgRNA (off-target score), and iv) the ability to synthesize repair template for each sgRNA with the intended mutations. To maximize the occurrence of HDR over NHEJ after the DSB, the sgRNAs that induce Cas9 cleavage within 10bp from the target site were prioritized.

For all selected sgRNAs, both Benchling and CRISPOR programs were utilized to calculate their on-target and off-target scores. To predict the cleavage efficiency, Benchling uses the algorithm developed by Doench et al. (2014), while CRISPOR provides more accurate efficacy scoring methods using the Doench Rule Set 3 score (-200, +200) (Doench et al., 2016). The off-target scores were calculated by Benchling's scoring method based on Hsu et al. (2013), and

by CRISPOR's more accurate CFD specificity score (0-100) developed by Doench et al. (2016). Higher off-target score correlates with a lower probability that the Cas9 will bind and cut other sequences with high similarity in the genome. The CRISPOR program was also used to evaluate potential off-target genomic sequences for each sgRNA. Exonic off-targets and off-targets with fewer than four mismatches were avoided. CFD score greater than 60 (CFD>60) is typically sufficient to select a designed sgRNA.

Upon ranking all possible sgRNAs for a specified target locus, at least two sgRNAs with the closest distance from the edit site and the highest efficiency and specificity scores were selected, regardless of their orientation. In case not enough efficient candidate sgRNAs could be designed with the initial approach of using the 5'-NGG-3' PAM sequence for the Cas9 nuclease, the engineered Cas9 variant SpG was used. This Cas9 variant recognizes the 5'-NGN-3' PAM sequence which enables targeting more binding sites around a genomic locus. Figure 3 provides a graphical illustration of the *in silico* CRISPR design pipeline.

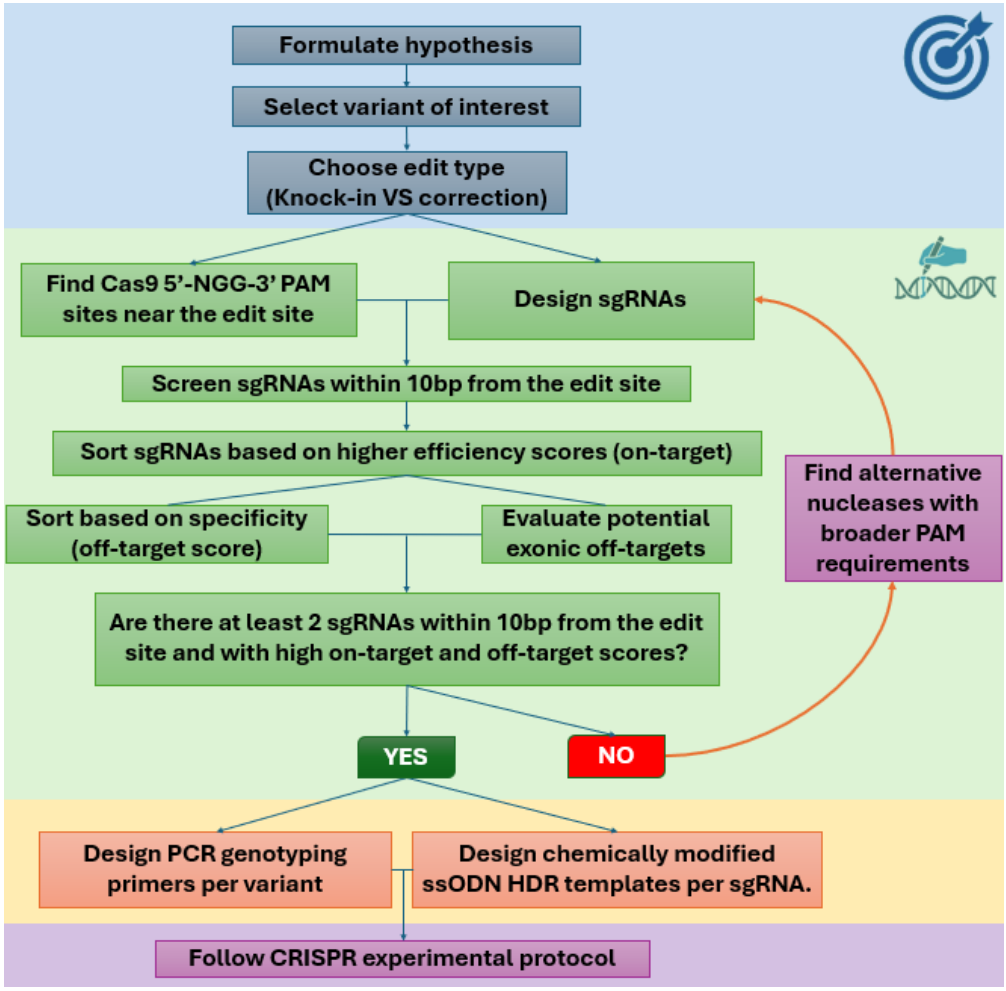

Figure 3. Flow diagram of the CRISPR design pipeline.

After the selection of at least two promising sgRNA sequences for each edit, ssODN HDR templates were designed in Benchling for each of the sgRNAs, because symmetric single-stranded DNA molecules are the most efficient for HDR efficiency (Liang et al., 2017; Bruntraeger et al., 2019; Skarnes et al., 2019; Schubert et al., 2021). The ssODN HDR templates contained the desired edit, with 50 bp homology arms on either side around the edit locus to maximize HDR efficiency (Figure 4). Either strand was used as a template for point mutations within 5 bp of the Cas cleavage site. However, for longer distances between the cut and the edit sites, the choice of the template strand was defined based on the position of the edit in regard to the cleavage site (Bruntraeger et al., 2019). Antisense ssODNs targeting the non-target strand were designed for mutations downstream of the cut site, and sense ssODNs targeting the sgRNA-binding strand for upstream mutations, regardless of sgRNA orientation (Bruntraeger et al., 2019; Schubert et al., 2021).

The ssODN HDR template design included intentional introduction of silent mutations that disrupt the PAM site recognition from the Cas enzyme. These extra modifications aim to prevent re-cutting of the Cas enzyme after a successful integration of the HDR template in the genome. In case that a silent PAM site mutation could not be introduced, at least 3 silent mutations were introduced within 8bp of the PAM site to disrupt the complementarity of the sgRNA's seed region. Figure 4 depicts the designed ssODN HDR templates.

Figure 4. Design of the 100-nucleotide long single-stranded DNA repair template, centred around the intended edit. The protospacer adjacent motif (PAM) site blocking silent mutations are marked in yellow. Phosphorylation marks were added to the first and last two nucleotides. The figure was created with Biorender.com.

All sgRNAs and ssODN HDR donor templates used *in vitro* were ordered from Integrated DNA Technologies, according to the manufacturer's instructions. The HDR template sequences were ordered as chemically modified ssDNA-oligonucleotides by adding two phosphorothioate-modified nucleotides to both ends. Supplementary Table 1 and Supplementary Table 2 list all *in silico* designed sgRNAs and ssODN HDR templates for the SUPER variants, while Table 3 and Table 4 show the designed oligos for the *INTS6* variant.

Table 3. Features of the 20-nucleotide long single guide RNAs (sgRNAs) used for each nucleofection.

Two sgRNAs were designed per edit, each with different orientation (- or + strand). The locus of the edit was within 5bp from the expected nuclease cleavage site, maximizing the likelihood of successful integration of the repair template via the homologous-directed repair (HDR) mechanism. Higher on-target and off-target scores indicate higher efficacy and specificity of the sgRNAs respectively.

sgRNA_ID	strand	Sequence (5' -> 3')	Purpose	Cut – edit site distance	Benchling on-target score	Benchling off-target score	CRISPOR on-target score	CRISPOR CFD score
HEL252_sg01_RV	-	AAAGTCCGGAGCCGATCACA	Knock-in of the pathogenic variant	4 bp	67.3	90.2	86	96
HEL252_sg02_FW	+	ATAAACCATGTGATCGGCTC	Knock-in of the pathogenic variant	1 bp	53.3	73.9	6	97
HEL246_sg03_RV	-	AAAGTCCGGAGCTGATCACA	Correction to wild type	4 bp	69.3	76	90	90
HEL246_sg04_FW	+	ATAAACCATGTGATCAGCTC	Correction to wild type	1 bp	55.7	67.7	22	89

Table 4. The sequences of the 100-nucleotide long single strand oligonucleotide (ssODN) homology-directed repair (HDR) template *INTS6* sequences for each single guide RNA (sgRNA).

The desired edit is highlighted in red letters. The protospacer adjacent motif (PAM) site is denoted in bold blue letters, while the silent mutations to prevent re-cutting of the Cas9 are highlighted in yellow. The asterisks at the first and last two nucleotides indicate the phosphorylated nucleotides.

sgRNA_ID	ssODN HDR template sequence (5' -> 3')	Purpose	Template orientation
HEL252_sg01_RV	G*A*TCTGTAGGCAAAAAAGTAGTACAGGAGACTGGAATAAAAGTCCGGAGC TGAAGTCATGGTTTATCAATGGCATATAGGAAAGATTTTCAACAACTCC*T*C	Knock-in of the pathogenic allele	Sense
HEL252_sg02_FW	G*A*TCTGTAGGCAAAAAAGTAGTACAGGAGACTGGAATAAAAGTTAGGAGC TGATCACATGGTTTATCAATGGCATATAGGAAAGATTTTCAACAACTCC*T*C	Knock-in of the pathogenic allele	Antisense
HEL246_sg03_RV	G*A*TCTGTAGGCAAAAAAGTAGTACAGGAGACTGGAATAAAAGTCCGGAGC CGAAGTCATGGTTTATCAATGGCATATAGGAAAGATTTTCAACAACTCC*T*C	Correction to wild type	Sense
HEL246_sg04_FW	G*A*TCTGTAGGCAAAAAAGTAGTACAGGAGACTGGAATAAAAGTTAGGAGC CGATCACATGGTTTATCAATGGCATATAGGAAAGATTTTCAACAACTCC*T*C	Correction to wild type	Antisense

3.2.2 Design of primer pairs for the *INTS6* variant

The nested primer pairs targeting the locus around the *INTS6* mutation were designed using NCBI Primer-BLAST. The 500bp primer pair was used for the T7 Endonuclease 1 (T7E1) assay as described at Section 3.9 and the first PCR round to amplify the target sequence for genotyping in both knock-in and isogenic correction clones. The 200bp primer pair was required for the nested PCR following the amplification of the 500bp amplicon. Detailed information on the primer pair sequences is presented in Table 5. Figure 5 denotes the primer binding positions on the *INTS6* target sequence. The i5 and i7 primers containing the barcodes and the Illumina sequencing adaptors used for MiSeq are listed in Supplementary Table 3 and Supplementary Table 4.

Table 5. Primer pairs for PCR amplification of the *INTS6* locus of interest.

The amplicon size of the PCR product is calculated including the length of the primer sequences. The 200bp primer sequence complementary to the genome is denoted in bold blue letters. The rest of the sequence of the 200bp primers belongs to the universal appends that are recognized by the i5 and i7 primers containing Illumina sequencing adaptors and barcodes.

Primer_ID	Orientation	Tm (°C)	GC content	Amplicon size	Sequence (5' -> 3')
500bp_fw	Forward	56.5	39.13%	500bp	AGGCCAAAATAGAATCTGATCGA
500bp_rv	Reverse	61.2	45.45%		TCTACCATCCACTACTGCAACT
200bp_fw	Forward	58.25	52.38%	265bp	ACACTCTTTCCCTACACGACGCTCTTCC GATCT**CGAGTCATTGGATCTGTAGGC**
200bp_rv	Reverse	58.13	39.13%		TCGGCATTCCTGCTGAACCGCTCTTCC GATCT**ACCTTATTCAGCAAAGCAACTTG**

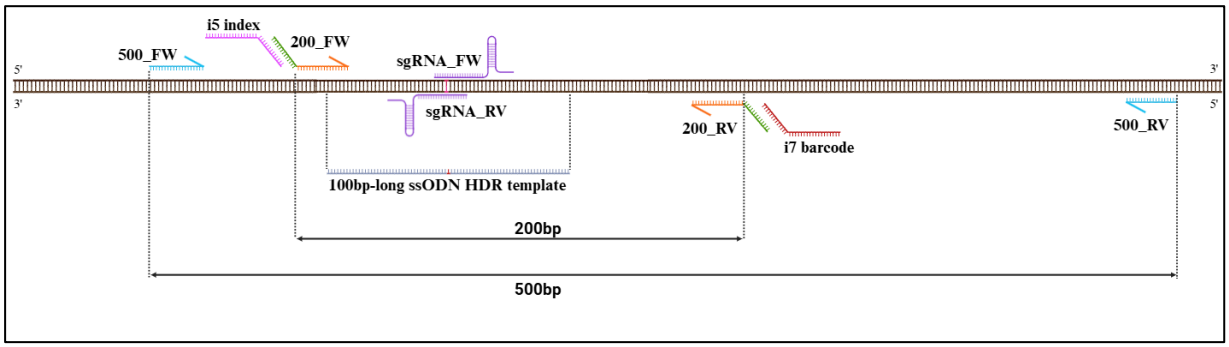

Figure 5. Primer pair and sgRNA binding positions around the *INTS6* mutation. The 500bp and 200bp primer pairs are not precisely centred around the *INTS6* mutation. The binding sites of the 500_fw and 200_fw primers in the gene are very close to each other and to the edit site, while the reverse primer sequences bind further downstream. The i5 and i7 primers introduce the Illumina sequencing adaptors and unique barcodes for the identification of individual sets of amplicons. Figure was created with Biorender.com

3.3 Graphical overview of the CRISPR experimental workflow

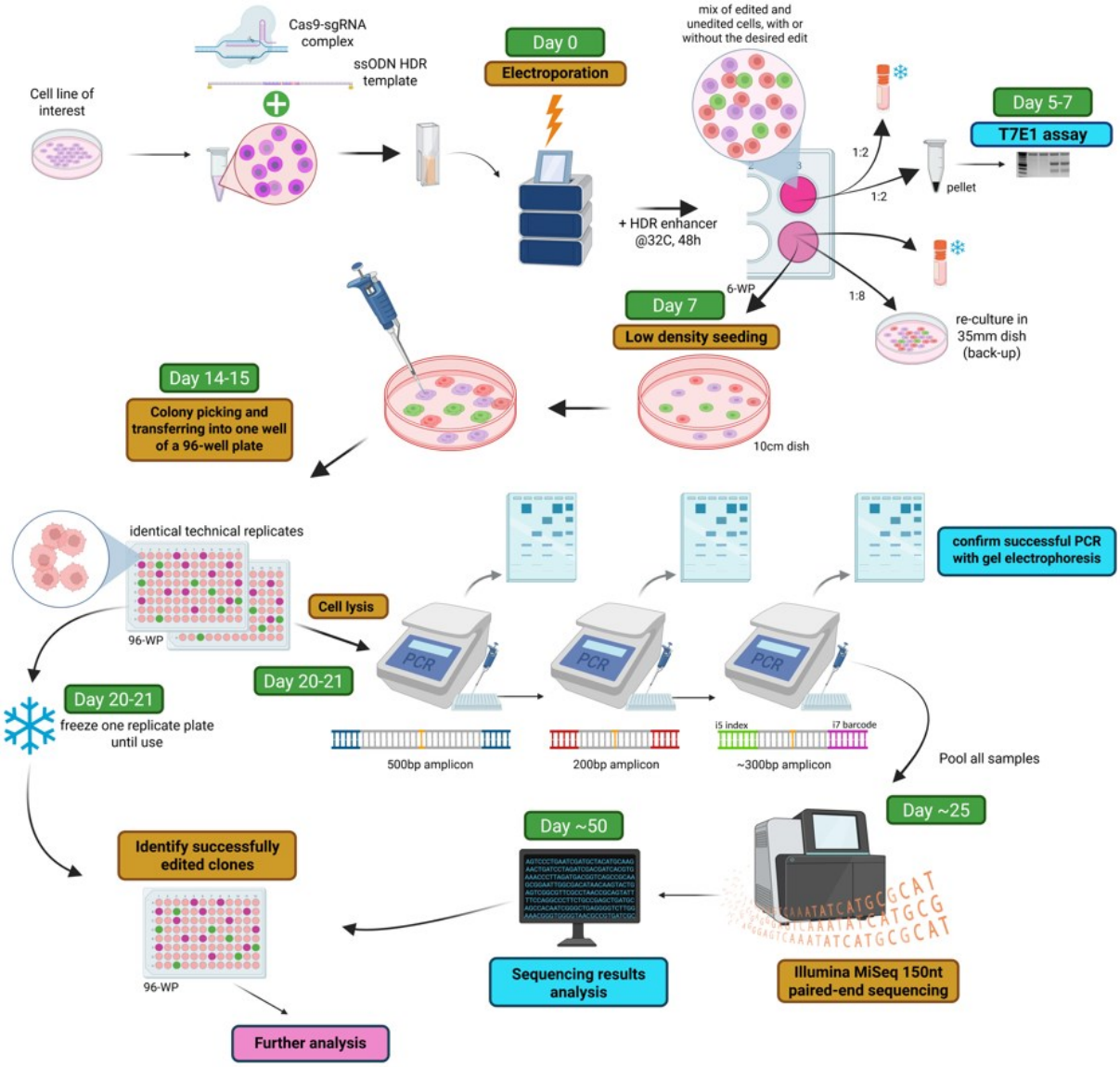

Figure 6. Experimental workflow of the CRISPR/Cas9 editing process followed in this project. The Cas9-single guide RNA (sgRNA) complex and single-stranded DNA (ssDNA) repair templates were delivered into the human induced pluripotent stem cells (iPSCs) via electroporation. To improve the activation of the homology-directed repair (HDR) pathway, cold shock treatment and the addition of an HDR enhancer were utilized. Upon recovery, the editing outcomes were evaluated with a T7 Endonuclease 1 assay. The electroporated cells were sub-cultured in 10cm dishes at low density to isolate single-cell-derived colonies. Individual clonal colonies were picked and expanded in duplicate 96-well plates. One plate was cryopreserved as a back-up, and the other was used for PCR genotyping and Illumina MiSeq to identify successfully edited clones. The workflow is simplified for one sgRNA condition. Figure was created with Biorender.com.

The experimental protocol was adapted from pre-existing literature and previously optimized in the lab (Figure 6). More specifically, the T7E1 assay, the cryopreservation of clones from a 96-well plate, the cell lysis and genotyping protocols were modified from Bruntraeger et al. (2019). The electroporation protocol, as well as the low density seeding and single-cell-derived colony picking steps were adjusted from Skarnes et al. (2019).

3.4 Agarose gel electrophoresis

Agarose gel electrophoresis confirmed the success of the PCRs and evaluated the outcomes of the T7E1 assay described in Section 3.9. The gel was prepared by diluting 1 – 2.5 g of agarose powder (Bioline, cat. # BIO-41025) in 100 ml of 1x Tris-acetate-ethylenediaminetetraacetic acid (EDTA) buffer and heating until complete dissolution. 4 µl of Midori Green Advance DNA stain (Nippon Genetics, cat. # MG04) was added to the solution before casting. The gel was loaded on a casting tray with the appropriate size of well combs in place and left to solidify for 30 minutes. The gel was then placed into the chamber, covered in 1x Tris-acetate-EDTA running buffer. To evaluate the size of the PCR products, 3-4 µl of a 50bp Gene Ruler DNA ladder (Thermo Fisher Scientific, cat. # SM0373) was used. 6X gel blue loading dye (New England Biolabs, cat. # B7021S) was mixed with 4 µl of each sample to a final concentration of 1X (1:6 dilution). The samples were then loaded on the gel and the electrophoresis was performed at 100V for 40-45 minutes. The gel was imaged with an Alpha Imager HP (Alpha Innotech) using the AlphaEaseFC software.

3.5 Cell culture – iPSC maintenance.

3.5.1 Cell culture media

iPSC lines were cultured in sterile-filtered complete E8 medium in 35mm dishes, incubated in a 37°C, 5% CO_2 humidified incubator. Complete E8 medium was prepared by supplementing Essential 8™ basal medium (Gibco, cat. # A15169-01) with Essential 8™ Supplement (50X, Gibco, cat. # A15171-01) to a final concentration of 1X (50:1 dilution). E8 was filter-sterilized using a Nalgene Rapid-Flow Filter Unit (Thermo Fisher Scientific, cat. # 569-0020) and stored at +4°C covered with aluminum foil due to its light sensitivity.

E8 medium was changed daily by carefully aspirating the old medium and adding fresh 2 ml complete E8 medium in the dish. In case of increased cell death, the cells were washed once with 1ml sterile-filtered DPBS (Dulbecco's Phosphate-Buffered Saline, without Ca^{2+} and Mg^{2+}, 1X, Gibco, cat. # 14190-144) before adding fresh media. All cell culture media were sterile filtered through a 0.2µm sterile syringe filter (Thermo Fisher Scientific, cat. # 15206869).

3.5.2 Coating of cell culture plates

iPSC lines were cultured in Matrigel-coated or Synthemax-coated cell culture plates to support cell attachment and growth. The coating solution was selected based on application.

Matrigel (Corning, cat. # 356231) was diluted in sterile filtered +4°C Dulbecco's Modified Eagle Medium/Nutrient Mixture (DMEM) F-12 (Gibco, cat. # 21331-020) at a working concentration

of 1:200. The plates were coated with an appropriate volume of coating solution according to Table 6, and incubated at either +4°C overnight, room temperature for 2 hours, or 37°C for 30 minutes before use. The plates could also be stored at +4°C with the coating solution for up to one week, sealed with parafilm.

To prepare the Synthemax coating solution, Synthemax II stem cell synthetic substrate (Corning, cat. # 3535, stock concentration: 1mg/mL) was diluted at a 1:40 ratio in sterile cell culture water (Fresenius Kabi AG, cat. # B230523) to a working concentration of 0.025mg/ml. The plates were incubated for 2 hours at room temperature, coating solution was aspirated, and plates stored dry at +4°C sealed with parafilm for up to 3 months until use.

Table 6. Volumes of coating solutions and culture media feed for different cell culture vessels.

Cell culture vessel	Supplier	Cat. #	Volume of coating solution	Volume range of culture medium feed
10cm plate	Corning	430167	8 ml	8-12 ml
35mm plate	SARSTEDT AG&Co.KG	83.3900	1.5 ml	2-3 ml
6-well plate	Corning	3506	1 ml	2-3 ml
24-well plate	Corning	3527	250 µl	500 µl
96-well plate	Corning	3596	50 µl	100-150 µl

3.5.3 Thawing, culturing and cryopreservation of iPSC lines

Prior to thawing, sterile-filtered complete E8 medium and Matrigel-coated 35mm dishes (SARSTEDT AG &Co.KG, cat. # 83.3900) were pre-warmed at room temperature.

2 ml cryogenic vials (Corning, cat. # 430488) from liquid nitrogen storage containing frozen iPSCs were rapidly thawed in a beaker with warm water at 37°C while continuously monitoring the temperature with a thermometer. Once only a small ice fragment remained, the vial was immediately decontaminated by wiping with 80% ethanol and transferred to the biosafety cabinet. The cell suspension (1 ml) was collected into a 15-ml Falcon tube containing 8 ml of E8 medium supplemented with 10µM ROCK (Rho-associated coiled-coil containing protein kinase, Y-27632, HelloBio, cat. # HB2297, 10mM stock solution) inhibitor to prevent dissociation-induced apoptosis. The tube was centrifuged at 200 x g (RCF) for 3 minutes at room temperature (Eppendorf, 5804 R centrifuge). During centrifugation time, the pre-coated 35mm plates were prepared by removing the Matrigel solution and adding 1 ml of pre-warmed E8 medium supplemented with 1 µl of 10mM ROCK inhibitor.

After centrifugation, the supernatant was carefully aspirated, and the cell pellet was mechanically resuspended in 1 ml of E8 medium supplemented with ROCK inhibitor as stated

above. The resuspended cells were plated into the prepared 35mm Matrigel-coated dishes containing 1 ml of fresh prepared medium, resulting in a final volume of 2 ml per dish. Cells were incubated at 37°C with 5% CO_2 and maintained under standard conditions. The E8 medium was changed every day without the addition of the ROCK inhibitor, unless required due to poor cell survival.

The cell lines were passaged once they reached 70-80% confluence, to ensure optimal growth and expansion. Cells were detached using a mild dissociation solution, UltraPure™ 0.5M EDTA (pH 8.0, Life Technologies, cat. # 15575-038) diluted 1:1000 in 1X DPBS and sterile-filtered before use.

The culture medium was removed using a P1000 micropipette to minimize loss of cells. The cells were washed once with 1 ml of 0.5mM EDTA/DPBS and incubated with 1ml of the same solution at 37°C for 3 minutes. Upon incubation, the dissociation solution was removed, and cells were mechanically detached in 1ml of pre-warmed E8 medium. The cells were then plated at a desired ratio onto new Matrigel-coated dishes containing 2 ml of pre-warmed complete E8 medium. Plates were immediately returned to a humidified incubator at 37°C with 5% CO_2. The cell cultures were observed daily using a microscope to monitor their status. The cell culture medium was replaced every 24 hours as described in the previous section.

Once the cell culture reached 70-80% confluency, the cells were detached as described above and resuspended in 1 ml of freezing medium, consisting of Knock-out Serum Replacement (KSR, Gibco, cat. # 10828-010) supplemented with 10% (v/v) DMSO (dimethyl sulfoxide, Sigma-Aldrich, cat. # D2650-100ML). The cell suspension was then collected into a 2 ml cryogenic vial and stored at -80°C overnight in a cryobox. After 24 hours, the cryovials were transferred to liquid nitrogen for long-term storage.

3.6 Preparation prior to electroporation

3.6.1 Reconstitution of sgRNAs and HDR ssODN templates in TE buffer

The sgRNAs and the ssODN HDR templates for the experiments were ordered from Integrated DNA Technologies (IDT) company and shipped in lyophilised form. The principles of their design are explained in detail in Section 3.2.

Tris-EDTA (TE) buffer (pH=8.0) was prepared under sterile conditions for the reconstitution of the sgRNAs and the ssODN HDR templates at the desired concentration to maintain constant pH for oligo stability, according to the supplier's recommendations. For a total volume of 10 ml, 2 µl of UltraPure™ 0.5M EDTA (pH 8.0, Life Technologies, cat. #15575-038) and 33.3 µl of 3M Tris-HCl buffer (prepared by Institute of Biotechnology media kitchen,

University of Helsinki) were added in 9,88 ml of nuclease-free water (Macherey-Nagel, cat. # 15879077). The solution was mixed in a 15-ml falcon tube inside a biosafety cabinet and sterile-filtered before use. The day before electroporation, the DNA and RNA oligonucleotides (shipped dry) were briefly centrifuged for a few seconds to minimize yield loss and resuspended in 16.2 µl of TE buffer overnight at +4°C at a final concentration of 4 µg/µl and stored at -20°C until use. Similarly, the Alt-R HDR modified ssODNs were resuspended in 25.13 µl of TE buffer overnight at room temperature at a final concentration of 5 µg/µl and stored at -20°C until use.

3.6.2 Preparation of the nucleofection solution, the cell culture medium, and the Matrigel-coated plates

22.73 µl of the Nucleofector® Supplement 1 (Lonza, cat. # PBP3-00675, lot. # S10641) was added to 102.27 µl of the P3 Primary Cell Solution (Lonza, cat# PBP3-00675, lot# 10667) to obtain the complete 4D-Nucleofection solution per reaction (120 µl + 5 µl for pipetting error, P3 primary solution/supplement ratio = 4.5:1).

For each nucleofection, 6 ml of cell culture medium were prepared by diluting 600µl of CloneR (CloneR V2 Cloning 10X Supplement, StemCell Technologies, cat. # 100-0691) and 4.35 µl of 0.69mM Alt-R HDR enhancer V2 (IDT, cat. # 10007921, 1:1380 dilution, working concentration = 0.5µM) in 5.396 ml of complete filtered-sterile E8 medium.

For each nucleofection, 2 wells of one 6-well plate (Corning, cat# 3506) were incubated with 1 ml Matrigel according to the instructions at Section 3.5.2 and incubated for 60 minutes at 37°C. Following incubation, the coating solution was aspirated, and the freshly prepared cell culture medium was added in both wells, 4 ml and 2 ml respectively. The plates were then incubated at 37°C until use.

3.6.3 Pre-assembly of the Cas9 RNP complex

For each nucleofection, 4 µl of the sgRNA (= 16µg = 500pmol, stock concentration = 4 µg/µL) were mixed with 2 µl of the Alt-R HiFi Cas9 V3 nuclease (= 20 µg = 125 pmol, stock concentration = 10 µg/µL, IDT, cat. # 1081061). The solution was incubated at room temperature for 30-45 minutes before adding to the cells immediately prior to electroporation.

3.6.4 Preparation of the iPSCs

Approximately 10-15 minutes after mixing the sgRNAs with the Cas9 to form the RNP complex, the cells cultured in a 35mm plate were washed once with sterile 1x DPBS and incubated with 1 ml of pre-warmed StemPro Accutase (Gibco, cat. # A11105-01) for up to 5 minutes at 37°C.

Cells were then resuspended and collected in a 15-ml falcon containing 3 ml of complete E8 medium to inactivate the chelating agent. A cell pellet was formed after a centrifugation step at 200 x g (5804 R) for 3 minutes at room temperature and resuspended in 1 ml of E8 medium. 9 µl of the suspension were mixed with 1 µl of Acridine Orange stain (Logos Biosystems, cat. # F23002) and counted using Luna FX7 automated cell counter (Logos Biosystems, LOG-L70001). The live cell count was used to calculate the volume of suspension needed for the electroporation.

For each nucleofection, 8 x 10^5 cells were aliquoted and centrifuged at the same speed for 3 minutes at room temperature. During centrifugation and immediately before nucleofection, 1 µl of the Alt-R HDR ssODN template repair (=5 µg = 200pmol, stock concentration = 5 µg/µl) was added to the pre-assembled Cas9 RNP to form the final Cas9 RNP component. After centrifugation the supernatant was carefully removed, and the cells were gently resuspended in 120 µl of the prepared 4D-Nucleofection Solution. Subsequently, the cell suspension was added onto the RNP complex containing the Cas9 RNP and the ssODNs. The solution was then mixed twice and 120 µl were transferred to a nucleocuvette™ vessel (Lonza, cat. # PCK-2005). The cuvette was tapped on a flat surface to eliminate any air bubbles and electroporated immediately.

3.7 Electroporation

The electroporation was performed using the Lonza 4D-Nucleofector X Core Unit (Lonza). The following program was selected and executed:

Device: X	Vessel: 1	Experiment: A1
Solution: P3	Pulse: C	Pulse Code: CA-137

Immediately after the electroporation, the cells were gently transferred via a sterile disposable Pasteur pipette (Lonza) into one well of the Matrigel-coated 6-well plate, which contained 4 ml of the prepared cell medium. To evenly distribute the cells in the well, the plate was rocked back and forth a couple of times. 1 ml of the cell suspension from that well was then transferred to the second well containing 2 ml of media, resulting in a total volume of 3 ml per well. The two wells were technical replicates and had different cell densities, as expected. The plates were incubated in a humidified incubator at 32°C and 5% CO_2 for 48 hours. The lower temperature slows cell cycle progression, allowing cells more time to repair DSBs via the HDR pathway (Maurissen & Woltjen, 2020). 16 hours post-electroporation, the cell medium was replaced with regular complete E8 medium to remove the reagents and small molecules that may increase cell toxicity. The cell medium was changed daily thereafter, until 70-80% confluence.

48 hours after the electroporation, the plates were returned to a humidified incubator set to 37°C and 5% CO_2.

3.8 Post-electroporation maintenance

Once the wells reached 70-80% confluence, the following actions were followed, as illustrated in Figure 7.

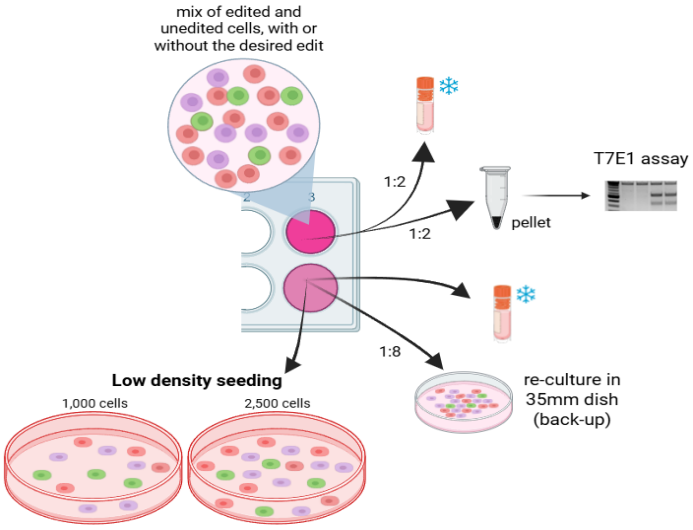

Figure 7. Experimental workflow following the recovery of the electroporated cells. One well was mainly used to collect the pellet for the T7 Endonuclease 1 assay, whereas the other one to seed single cells into 10cm dishes at a low density. The remaining cells were either cryopreserved or kept in the cell culture for back-up. Figure was created with Biorender.com.

From the well that was initially seeded with more cells (top well), half of the cells were cryopreserved to serve as a backup. The rest were collected as a pellet and stored at -80°C for the T7E1 assay, as described in Section 3.9. The bottom well (initially seeded with less cells) was used for single colony isolation and back-up.

To collect the cells from one confluent well of a 6-well plate, the cell medium was first removed, and the cells were washed once with 1 ml of sterile 1x DPBS at room temperature. The cells were incubated with 1 ml pre-warmed Accutase for up to 5 minutes at 37°C. Following incubation, the dissociation solution was carefully removed, and the cells were mechanically dissociated into a single-cell suspension in 1ml sterile 1X DPBS. To acquire the cell pellet from the single-cell suspension, 500 µl were collected in 1.5 ml microcentrifuge tubes and centrifuged at 300 x g, for 7 minutes at room temperature (Eppendorf, 5424 R centrifuge). The supernatant was removed, and the pellet was transferred immediately onto dry ice and stored at -80°C until use. To cryopreserve an aliquot of the cells from the suspension, the cells were centrifuged as described above, the remaining pellet was resuspended in 1 ml of freezing medium and cryopreserved as described in Section 3.5.3. To re-culture the cells in a 35mm

dish to serve as a back-up in case the seeding of single cells was not successful, 125 µl of the 1ml cell suspension (1:8 passage ratio) were plated into a 35mm Matrigel-coated dish containing 2 ml of E8 supplemented with 1:10 CloneR.

3.9 T7 Endonuclease I assay to assess editing outcomes

3.9.1 Overview of the experimental workflow and expected outcomes

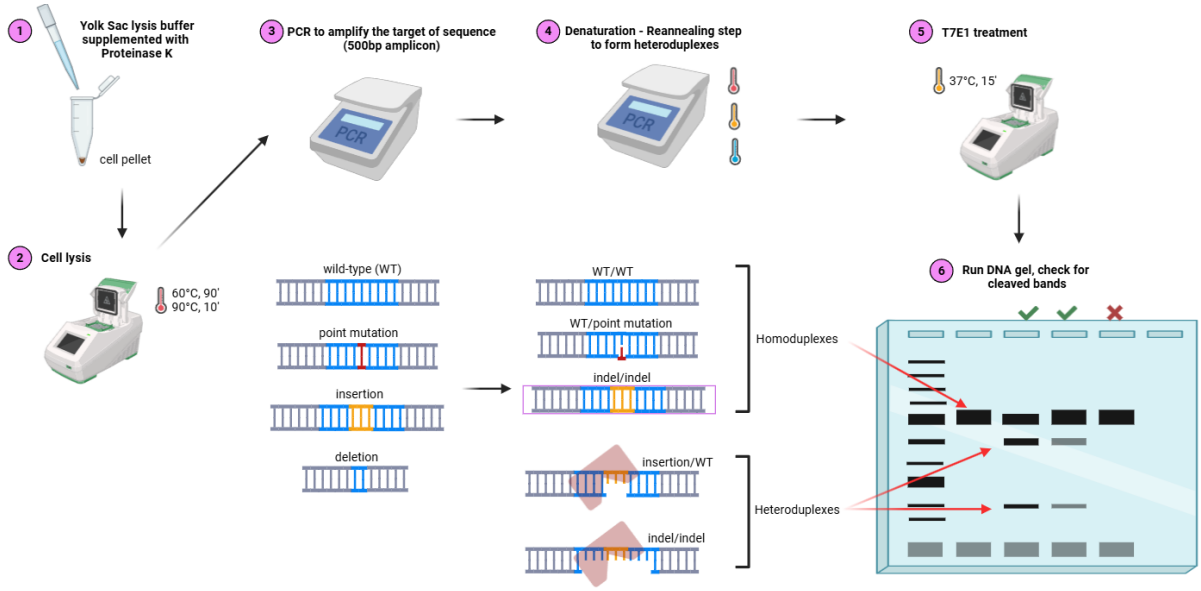

Figure 8. Experimental workflow of the T7 Endonuclease 1 assay. Each cell pellet contains a pool of unedited and edited cells with variable editing results. A PCR step selectively amplifies the target sequence. The subsequent heating-cooling step generates identically matched homoduplexes and mismatched heteroduplexes, which are cleaved by the T7 enzyme. The cleavage efficiency of each single guide RNA is then evaluated with gel electrophoresis. Figure created with Biorender.com.

The assay estimates the gene editing efficiency for each sgRNA and is mainly dependent on the edits occurred by the NHEJ pathway after the Cas9 DSB. The enzyme recognizes and cleaves mismatched heteroduplex DNA close to the mismatch site, usually leaving perfectly matched homoduplexes uncut (Sentmanat et al., 2018). Mismatches and small indels that distort the double helix are strongly recognized. However, single-base indels or point mutations in heteroduplexes that occur during the editing process are not efficiently detected by the enzyme (Sentmanat et al., 2018). This NHEJ-biased assay does not give evidence on whether the desired edit is present in the nucleofected cells, but it indicates sgRNA-Cas9 cleavage efficiency. Figure 8 schematically illustrates the experimental steps and expected results of the T7E1 assay.

3.9.2 Cell lysis of iPSCs

Pelleted cells from each 6-well plate (one nucleofection) and from two non-electroporated cell lines were used to assess the editing outcomes. The pelleted cells were resuspended in 1 mL of Yolk Sac lysis buffer (prepared by Institute of Biotechnology media kitchen, University of Helsinki) supplemented with 20 µl of 20mg/ml mM Proteinase K (Macherey-Nagel, cat. # 740506.30) and incubated at room temperature for 30-60 seconds. The lysate was collected in a sterile 2 ml sterile Eppendorf tube and incubated at 60°C for 90 minutes on a dry heat block (Eppendorf Thermomixer C, Thermo Fisher Scientific). To inactivate the activity of the Proteinase K, the lysate was incubated at 95°C for 10 minutes.

3.9.3 PCR to amplify the sequence of interest

To amplify the target sequence for each lysate, a 25 µl PCR mix was prepared according to Table 7, using the 500bp pair of primers (Table 5) and the lysate diluted 1:10 in nuclease-free water as DNA template. The samples were incubated in a PCR thermal cycler (Applied Biosystems, SimpliAmp Thermal Cycler) by executing the PCR cycling program described in Table 8.

Table 7. PCR mix up to 25 µl to generate a 500bp amplicon of the target sequence using the 500bp primers pair and the 1:10 diluted lysate as DNA template.

Reagents	Volume (µL)	Stock concentration	Working concentration
1:10 diluted lysate	5		
KAPA HiFi HotStart Polymerase Ready mix (Kapa Biosystems, cat. # 09420711001)	12.5	2X	1X
500bp forward primer (Sigma-Aldrich)	0.75	10µM	0.3µM
500bp reverse primer (Sigma-Aldrich)	0.75	10µM	0.3µM
Nuclease-free water (Macherey-Nagel, cat. # 15879077)	Up to 25 µl		

Table 8. PCR cycling program to amplify the 500bp target sequence from diluted cell lysates.

Step	Temperature (ºC)	Duration	Cycles
Initial denaturation	95	3 minutes	1
Denaturation	98	20 seconds	X 35
Annealing	**60**	15 seconds	
Extension	72	30 seconds	
Final extension	72	5 minutes	1

To confirm the success of the PCR, 5 µl of the PCR product was loaded on a 1% agarose gel, according to the instructions at Section 3.4.

3.9.4 Generation of heteroduplexes and incubation with T7 Endonuclease I

The 500bp amplicon from each lysate was subsequently divided into two duplicate samples (10 µl each), one to receive the enzyme and the other one serve as untreated control. To each sample, 2 µl of NE Buffer 2 (10X, New England Biolabs, cat. # B7002S) and 7 µl of nuclease-free water were added. To form the heteroduplexes, the samples were incubated in a thermocycler according to the program described in Table 9.

Table 9. Denaturation-annealing steps to generate heteroduplexes formed by mismatches in heterozygous DNA molecules.

Temperature (ºC)	Duration
95	10 minutes
95-85	Ramp -2 ºC / second
85-25	Ramp -0.3 ºC / second

Upon heating-cooling step, 1 µl of T7 Endonuclease 1 (New England Biolabs, cat. # MO302L) was added to the samples meant to receive the enzyme, and 1 µl of NE Buffer 2 was added to the control duplicate samples. The samples were then incubated in a PCR thermal cycler at 37 ºC for 15 minutes to activate the T7 enzyme, which cleaved the mismatched heteroduplexes. Then, 20 µl of each sample was loaded onto a 2.5% agarose gel (Section 3.4) to evaluate the editing efficiency.

3.10 Isolating and expanding single cell derived clones

3.10.1 Sub-culturing single cells from 6-well plates to 10cm plates

Once the bottom well of the 6-well plates with the electroporated iPSCs reached ~80% confluence (see Section 3.8), single cells were sub-cultured onto 10cm cell culture plates (Corning, cat. # 430167). This step is crucial for isolating and screening single cells that might have the desired edit among the cell population.

For each nucleofection (one 6-well plate), two 10cm plates were pre-coated with Synthemax according to the instructions at Section 3.5.2. The plates were incubated at 37ºC with 10 ml complete E8 medium supplemented with 1 mL of CloneR V2 prior to adding the cells. The cells from one well of the 6-well plate were washed once with 1 mL sterile 1X DPBS before adding 1 mL of pre-warmed Accutase. The plates were incubated at 37 °C/5% CO_2 incubator for up to 5 minutes. Following incubation, the cells were gently resuspended and collected in a 15-ml

falcon tube containing 3 mL of E8 medium to inactivate the Accutase. A cell pellet was formed after centrifuging at 210 x g (5804 R) for 4 minutes at room temperature and then resuspended in 1 mL of E8 medium. 9 µl of the suspension was mixed with 1 µl of Acridine Orange dye and counted using Luna FX7 automated cell counter (Logos Biosystems, LOG-L70001).

The live cell count was used to calculate the volume of suspension needed for the plating. 250,000 live cells were aliquoted in 1 ml of E8 medium. Following serial 1:10 dilutions in E8 medium to minimize pipetting errors, 1,000 and 2,500 single cells were seeded to each of the Synthemax-coated 10cm dishes respectively. The dishes were then incubated at 37°C, 5% CO_2 incubator. The remaining cell suspension was pelleted and cryopreserved as described at Section 3.8. The culture medium was replaced every two days thereafter without CloneR, until single-cell colonies grew to up to 1 mM in diameter.

3.10.2 Preparation of cell culture media and 96-well plates for colony picking

For each set of 96 colonies to be picked from the two 10cm dishes (one nucleofection), 20 mL of pre-warmed cell culture medium was prepared by adding 2 ml CloneR, 20 µl of 10mM ROCK inhibitor, and 40 µl of 50mg/ml Primocin (InvivoGen, cat. # ant-pm) in 18 ml of complete E8 medium. The addition of the apoptotic inhibitors and the antibiotic aims to prevent increased cell death and minimize risk of contamination during the procedure of colony picking. Similarly for each set of 10cm dishes seeded with 1,000 and 2,500 single cells, two replicate flat-bottom 96-well plates (Corning, cat. # 3596) were pre-coated: one with 50 µl of Matrigel solution per well for freezing, and one with 50 µl of Synthemax solution per well for DNA lysis and genotyping.

3.10.3 Picking single-cell-derived colonies and expanding into replicate 96-well plates

Once single-cell-derived colonies reached approximately 1mM in diameter, individual, round-shaped colonies were picked and sub-cultured into replicate 96-well plates for clonal expansion. 10µM of ROCK inhibitor were added to each 10cm plate 4 hours prior to colony picking to minimize cell death.

Under a stereomicroscope inside the biosafety cabinet, round colonies of intermediate size and bright edges were picked from the plate using a P200 by gently scraping the colony and collecting the cell fragments in 50 µl of media. Each colony was transferred to an empty well of a U-bottom 96-well plate (SARSTEDT AG&Co.KG, cat. # 83.3925). After the collection of 96 colonies from the 10cm dishes, 150 µl of the prepared medium were added to each well using a multichannel pipette. The cells were triturated 5-7 times in the U-bottom 96-well plate to break each colony into multiple small clumps and to increase the chances that the cells will

be evenly divided into the replicate wells. 100 µl of the cell suspension per well was then transferred to corresponding rows of the pre-warmed Matrigel and Synthemax-coated flat-bottom 96-well plates. The plates were incubated at 37°C, 5% CO_2 humidified incubator. The culture medium was replaced with regular E8 medium after 48 hours and changed every two days until the wells reached 80% confluency.

3.10.4 Freezing iPSCs from Matrigel-coated 96-well plates

Once a few wells of the Matrigel-coated 96-well plates reached 80% confluency, the culture medium was removed, and the cells were washed once with 100 µl of sterile 1X DPBS per well. 30 µl of pre-warmed Accutase was added to each well and the plate was incubated at 37°C for up to 5 minutes. To inactivate the activity of the chelating reagent, 175 µl of the freezing medium (KSR supplemented with 10% v/v DMSO) was added to each well, column by column, and the cells were immediately triturated 5-10 times to obtain a single-cell suspension. 100 µl of mineral oil (EmbryoMax Light Mineral Oil Embryonic Stem cell qualified, Merck, cat. # 4005820) was added carefully on top of the freezing medium per well to avoid mixing with the freezing medium. The plates were then firmly sealed with parafilm and transferred immediately to dry ice and then -80°C for long-term storage.

3.10.5 Lysing iPSCs from Synthemax-coated 96-well plates

Similarly, once a few wells of the Synthemax-coated 96-well plates reached 80% confluency, the culture medium was removed, and the cells were washed once with 100 µl of sterile 1X DPBS per well. Each well was incubated with 100 µL of Yolk Sac lysis buffer supplemented with 2 µl of 20mg/mL Proteinase K at 37°C for 5 minutes. The bottom surface of each well was then scraped using a multichannel pipette and cells were triturated a few times before transferring the lysate into a 96-well non-skirted PCR plate (Azenta Life Sciences, cat. # 4TI-0750). The PCR plate was subsequently incubated in a thermal cycler for 90 minutes at 60°C. The following incubation step at 95°C for 10 minutes ensured the inactivation of the proteinase. The lysates were then stored at -20°C for long-term storage until use for genotyping.

3.11 Genotyping

Three rounds of PCR were required to amplify the target sequence in addition to introducing sequencing adaptors with barcodes. First, for each well of the 96-well plate, 10 µl of PCR master mix was prepared according to Table 10, using the 500bp forward and reverse primers described at Table 5. The DNA lysates from the lysis step were diluted 1:10 with nuclease-free water and used as the DNA template. The 96-well PCR plate was then incubated following the PCR thermocycler program described in Table 8. Following PCR, 3 µl of the PCR products from

a few wells were loaded onto a 1% agarose gel according to the instructions at the Section 3.4 to validate the successful amplification of the 500bp target region. A second round of PCR was prepared to obtain a 200bp long amplicon by using the 200bp forward and reverse primers of the *INTS6* sequence (Table 5), 0.5 µl of the 500bp PCR product as the DNA template, and following the PCR cycling program as in Table 11. Similarly to the first PCR round, gel electrophoresis was performed to confirm the success of the PCR. Subsequently, the 200bp PCR product was diluted 1:100 in sterile nuclease-free water for the next PCR round.

Table 10. PCR mix to a final volume of 10 µl to amplify the *INTS6* target sequence.
The volumes correspond to the PCR mix per well of a 96-well PCR plate.

Reagents	Volume (µl)	Stock concentration	Working concentration
DNA template	0.5 - 2		
KAPA HiFi HotStart Polymerase Ready mix (Kapa Biosystems, cat. # 09420711001)	5	2X	1X
Forward primer (Sigma-Aldrich)	0.3	10µM	0.3µM
Reverse primer (Sigma-Aldrich)	0.3	10µM	0.3µM
Nuclease-free water	Up to 10 µl		

Table 11. Touchdown PCR cycling program to amplify the 200bp target sequence.

Step	Temperature (ºC)	Duration	Cycles
Initial denaturation	95	3 minutes	1
Denaturation	98	20 seconds	X 10
Annealing	**70**	15 seconds	
Extension	72	30 seconds	
Denaturation	98	20 seconds	X 25
Annealing	**60**	15 seconds	
Extension	72	30 seconds	
Final extension	72	5 minutes	1

The last round of PCR (indexing PCR) introduced unique combinations of amplicons per sample by using barcoded MiSeq primers (Supplementary Table 3, Supplementary Table 4). 10 µl of PCR mix were prepared according to Table 12, using 0.5 µl of the 1:100 diluted 200bp amplicon from the second PCR round as the DNA template. Each well of one 96-well PCR plate received the same i5 primer and a different i7 primer to distinguish between the samples from one plate. Each 96-well PCR plate received a different i5 primer to differentiate the samples from different plates. The same PCR cycling program as in Table 8 was used for this PCR

round, with the annealing temperature being 70°C instead of 60°C. To check whether the PCR was successful, gel electrophoresis was performed as in previous sections.

Table 12. Indexing PCR mix to generate barcoded 200bp amplicons for MiSeq analysis.

Reagents	Volume (µl)	Stock concentration	Working concentration
1:100 diluted 200bp PCR product (DNA template)	0.5		
KAPA HiFi HotStart Polymerase Ready mix (Kapa Biosystems, cat. # 09420711001)	5	2X	1X
i7 primer (unique per well of a 96-well plate)	0.6	5 µM	0.3 µM
i5 primer (unique per 96-well plate)	0.3	10 µM	0.3 µM
Nuclease-free water (Macherey-Nagel, cat. # 15879077)	Up to 10 µl		

3.12 Analysis of MiSeq results using CRISPResso

2 µl of the final PCR barcoded amplicons from each sample were pooled and mixed in a 1.5ml Eppendorf tube. The samples were submitted for paired-end 150 nucleotide sequencing on an Illumina MiSeq platform, with a minimum of 1000 reads per amplicon. The Illumina MiSeq sequencing services were provided by the Biomedicum Functional Genomics Unit at the Helsinki Institute of Life Science and Biocenter Finland at the University of Helsinki.

To analyse the sequencing data from multiple 96-well plates at the simultaneously, the CRISPResso2 v2.3.2 software pipeline was run in HDR mode using the Snakemake module (Clement et al., 2019). CRISPResso2 (http://crispresso2.pinellolab.org/) is an updated version of the original CRISPResso tool developed by Pinello et al. (2016) with advanced implementations for analysis and comparison of multiple genome editing experiments (Clement et al., 2019). A Phred quality threshold of 25 was applied to filter low-quality bases. The FASTQ files generated from amplicon sequencing, and both the reference and expected HDR sequences for each CRISPR edit were provided as an input. The CRISPResso2 workflow then aligned the sequencing reads to the input sequences and generated output files of quantified editing outcomes (including insertions, deletions, and substitutions), and calculated the percentage of HDR, NHEJ, and unedited reads for each sample. The results were used to identify and prioritize clones with successful HDR events for downstream validation. The pipeline was run using a custom implementation of the CRISPResso2-Snakemake. All analysis scripts and configuration files used are publicly available at https://github.com/Kilpinen-group/crispressoWorkflow2/.

4 Results

4.1 Optimization of the sgRNA design methodology

4.1.1 The initial sgRNA design approach provides good predictive scores to edit variants in 6 out of 8 genes associated with SCZ at exome-wide significance.

For each variant, I designed sgRNAs for two different edits: one to introduce the pathogenic allele into wild-type genetic background, and another to revert it back to the wild-type genotype. The objective was to precisely modify the genome with these two edits per variant to create isogenic pairs for disease modelling. To maximize HDR efficiency, the *in silico* designed sgRNAs were first filtered based on their proximity to the target site. Then, a plot-based scoring framework was applied to visually rank the top 2-3 sgRNAs with the highest efficiency and specificity scores per gene. For some genes (*HERC1*, *SETD1A*, *SP4*) more than one variant were identified in the SUPER cohort (Table 1). In that case, the sgRNAs were plotted and assessed together because the functional effect of the edit is observed on the gene level. Figure 9 depicts all sgRNAs designed per gene to knock-in the pathogenic variants, while Figure 10 presents all sgRNAs designed for rescue experiments.

The methodology revealed that for most genes, strong candidate sgRNAs could be designed for both edits. The variants in *CACNA1G* and *XPO7* genes had the highest potential for successful editing with at least four high-quality sgRNAs each (Figure 9, Figure 10). *RB1CC1* and *TRIO* also had two strong candidate sgRNAs each. *HERC1* and *SETD1A* had 7 and 4 variants respectively, enabling the design of multiple efficient sgRNAs for each gene. This has the advantage of using distinct pathogenic variants to target the same gene, enhancing its utility for disease modelling.

In contrast, sgRNA design for *CUL1* and *SP4* did not yield optimal efficiency scores, mainly due to the lack of 'NGG' PAM sequences nearby and the very low efficiency scores of the few designed sgRNAs (Figure 9, Figure 10). The requirement for a PAM sequence close to the edit site is a constraint for sgRNA design. Therefore, alternative sgRNA design strategies should be explored for variants predicted to have low editing efficiency under this approach.

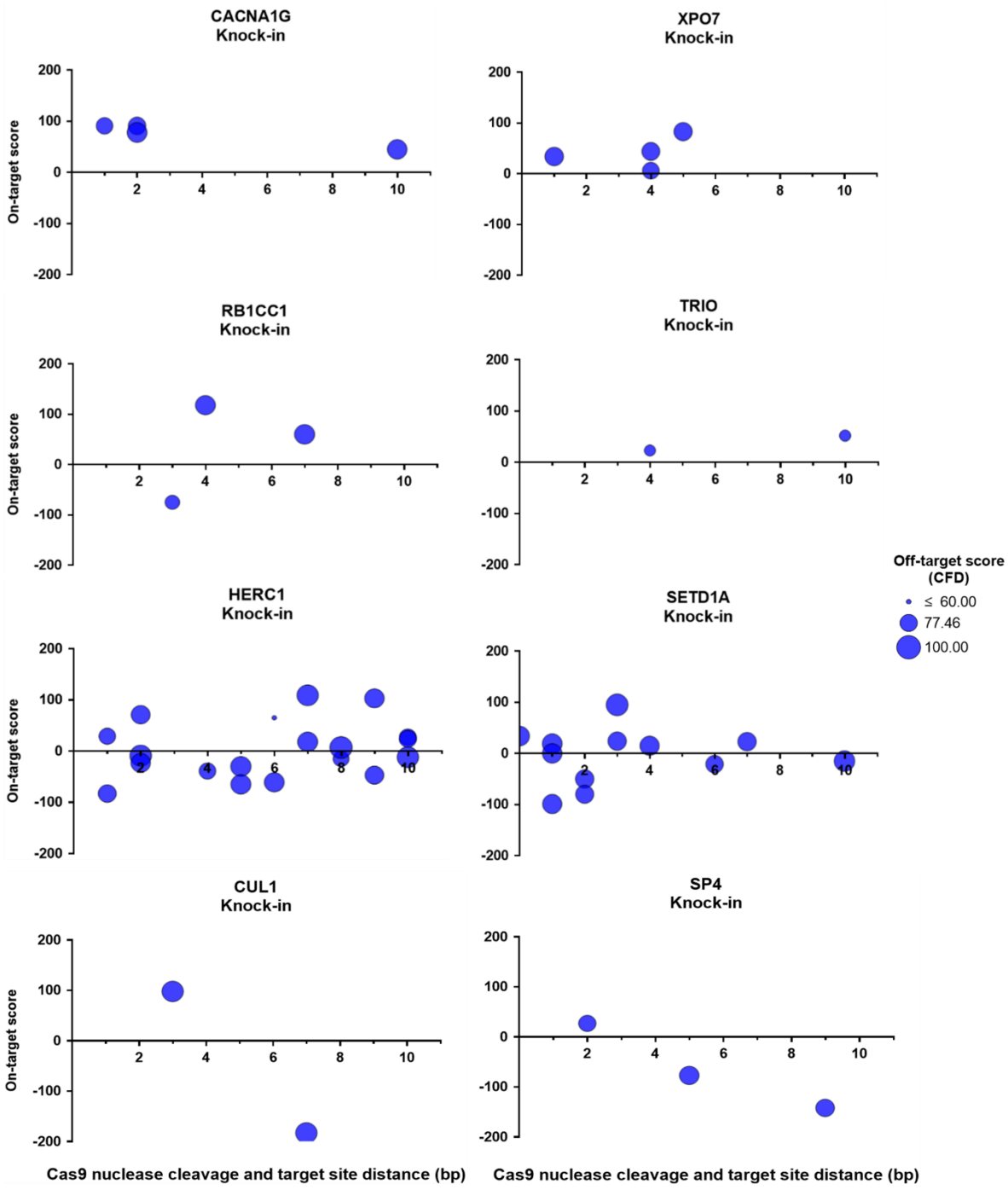

Figure 9. Bubble plots displaying individual single guide RNA (sgRNA) scores for variants of the same genes based on their efficiency, specificity, and proximity to the target site. The goal is to introduce each pathogenic variant in a wild-type genetic background. Each dot represents one *in silico* designed sgRNA. X-axis depicts the distance between the Cas9 cleavage and the edit site. The y-axis denotes the on-target efficiency score, calculated using the Doench RuleSet3 algorithm. The bubble size corresponds to the cutting frequency determination (CFD) specificity score. The sgRNAs with high on-target and off-target scores were prioritized. Not enough optimal candidate sgRNAs could be designed for variants in *CUL1* and *SP4* genes. Plots created using Prism GraphPad version 10.4.1.

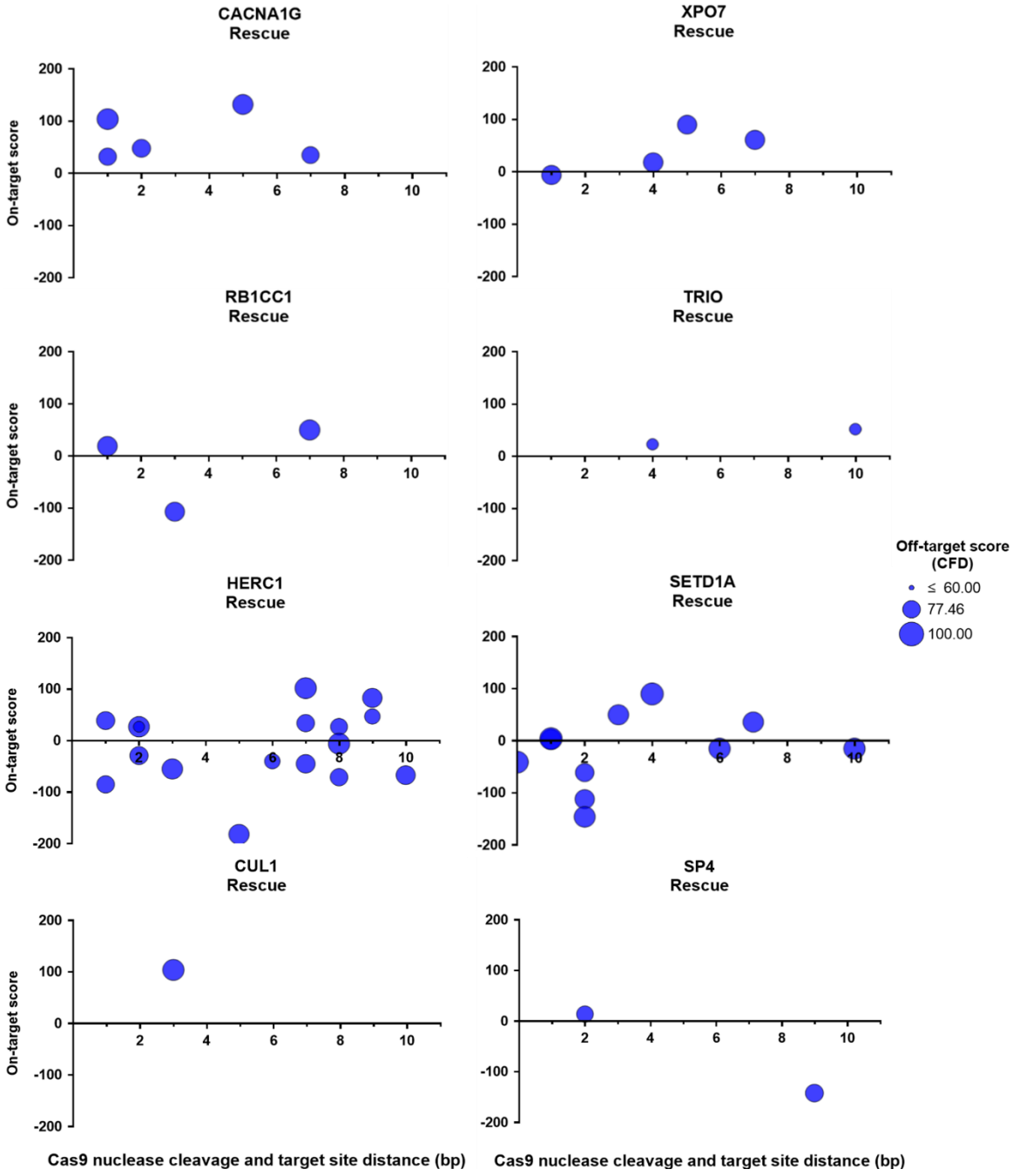

Figure 10. Bubble plots displaying individual single guide RNA (sgRNA) scores for variants of the same genes based on their efficiency, specificity, and proximity to the target site. The goal is to correct the pathogenic variant back to the wild-type genotype. Each dot represents one *in silico* designed sgRNA. X-axis depicts the distance between the Cas9 nuclease cleavage and the edit site. The y-axis denotes the on-target efficiency score, calculated using the Doench RuleSet3 algorithm. The bubble size corresponds to the cutting frequency determination (CFD) specificity score. The sgRNAs with high on-target and off-target scores were prioritized. Not enough optimal candidate sgRNAs could be designed for variants in *CUL1* and *SP4* genes. Plots created using Prism GraphPad version 10.4.1.

4.1.2 CRISPR Design using an engineered alternative of Cas9 predicts higher editing efficiency and specificity for variants in *SP4* and *CUL1* genes.

The variants in *CUL1* and *SP4* genes have low potential of successful HDR when using the wild-type Cas9 as the nuclease to load the sgRNAs. However, this limitation was addressed by designing sgRNAs using a more versatile PAM sequence. The 'NGN' PAM sequence ('N' denotes any nucleotide) is recognized by SpG, an engineered Cas9 variant, and provides wider flexibility to design sgRNAs. Employing this approach, more efficient sgRNAs were designed for both genes, improving the possibility for successful HDR edits (Figure 11).

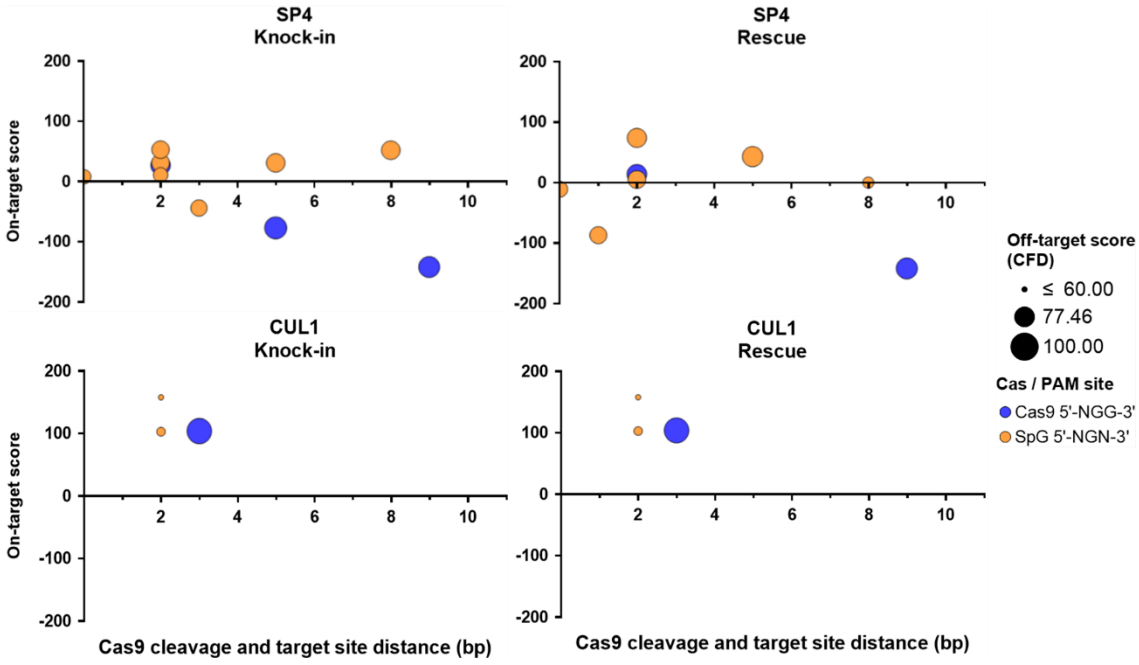

Figure 11. Bubble plots displaying individual single guide RNA (sgRNA) scores for variants of the same genes based on their efficiency, specificity, and proximity to the target site. The goal is either to introduce each pathogenic variant in a wild-type genetic background (left) or rescue the allele back to wild type (right). Each dot represents one *in silico* designed sgRNA. X-axis depicts the distance between the Cas9 cleavage and the edit site. The y-axis denotes the on-target efficiency score, calculated using the Doench RuleSet3 algorithm. The bubble size corresponds to the cutting frequency determination (CFD) specificity score. Blue colour denotes the initial sgRNA designed using the Cas9 'NGG' protospacer adjacent motif (PAM) sequence, whereas the orange colour represents the sgRNAs designed using the broader 'NGN' PAM recognized by the SpG nuclease. The sgRNAs with high on-target and off-target scores were prioritized. Plots created using Prism GraphPad version 10.4.1.

4.2 Evaluation of the CRISPR/Cas9-edited human iPSCs

4.2.1 The cell recovery patterns after electroporation indicate a CRISPR-editing effect on cell viability and morphology.

In addition to these rare variants in SCZ, a novel deleterious variant in the *INTS6* gene has recently been discovered in ID cases in Finland, showing a strong association with ID. Suggestive evidence for the association of *INTS6* with SCZ have also been found in the schizophrenia exome meta-analysis (SCHEMA) consortium, where rare *INTS6* variants have been recently identified in SCZ cases. Consequently, to investigate the causality of *INTS6* in the onset of SCZ and ID*, in vitro* cellular models of disease where CRISPR/Cas9 editing produces isogenic human iPSC-derived neuronal cell types are required.

To generate isogenic pairs of CRISPR/Cas9-edited human iPSC lines, iPSCs were cultured from two donors of the same family: one patient with mild ID carrying a heterozygous *INTS6* mutation (HEL246.2) and one wild-type unaffected individual (HEL252.2). Both iPSC lines were cultured for several passages before the electroporation to ensure cellular stability. These cell lines were very delicate and exhibited increased cell death during the early passages after thawing. To support post-splitting survival, the cells were supplemented with CloneR, a ROCK inhibitor which leads to colony formation with irregular borders. The patient-derived cell line (HEL246.2) grew faster than the wild-type HEL252.2, but the electroporation was performed when both cell lines reached a stable growth rate.

For both lines, electroporation had a considerable impact on iPSC morphology compared to the non-electroporated controls. Two days post-electroporation, irregular, jagged colony edges were seen in all wells, a morphological feature of cellular stress. The same observation was made for the routinely passaged iPSCs (Figure 12, Figure 13). However, by day four, the regularly maintained cells had recovered and formed smooth, well-rounded colonies, whereas the cells electroporated with the Cas9-sgRNA complex maintained the irregular, elongated colony edges (Figure 12, Figure 13). The ‘spikey’ morphology was a result of both stress response and treatment with CloneR but diminished once these factors were removed and the cells started growing together in compact colonies. Like the routinely cultured iPSCs, the Cas9-only electroporated control (no sgRNA) for HEL246.2 was also observed to form compact colonies with smooth borders.

Both the electroporation and the CRISPR/Cas9 editing affected the observed cell viability. In both cell lines, the electroporated cells with the Cas9 RNP complex were less confluent compared to the non-electroporated controls. This likely reflects the effects of both electroporation-induced stress and CRISPR gene editing. Interestingly, the decrease in confluency of the Cas9-only control in HEL246.2 was more moderate (Figure 12). This suggests that a change in genotype because of the CRISPR editing may led to increased cell death in the Cas9 RNP electroporated cells, pointing towards CRISPR editing activity.

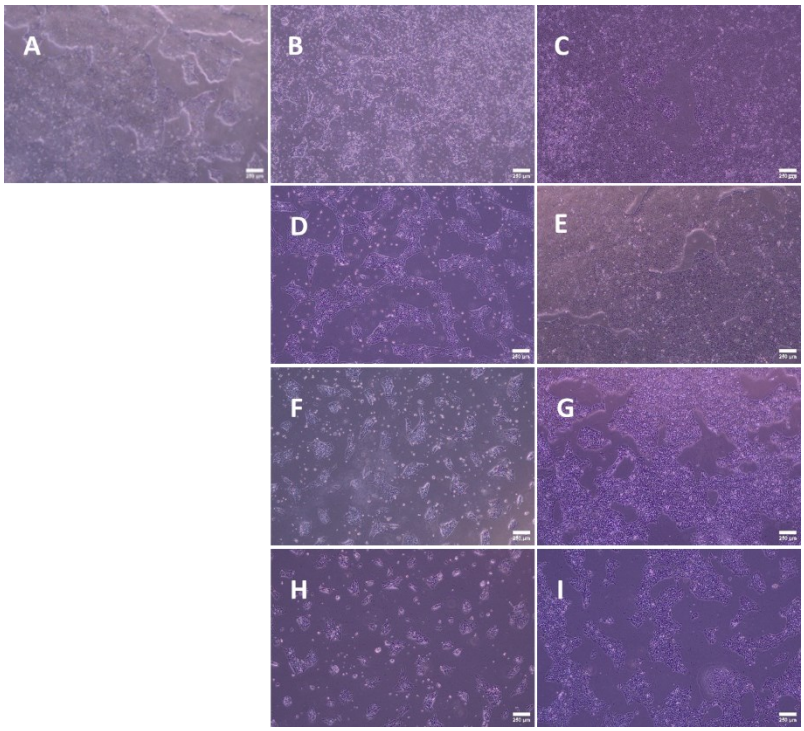

Figure 12. The effect of electroporation and targeted genome editing on the colony morphology and recovery patterns of the patient cell line (HEL246.2).
A: One day before electroporation.
B, C: Maintained regularly at day 2 and day 4 after 1:4 passage.
D, E: Day 2 and day 4 after electroporation with Cas9, without single guide RNA (sgRNA).
F, G: Day 2 and day 4 after electroporation with Cas9 and reverse sgRNA.
H, I: Day 2 and day 4 after electroporation with Cas9 and forward sgRNA.
All images were taken with Olympus CKX53 microscope at 4x magnification. Scale bar = 250µm.

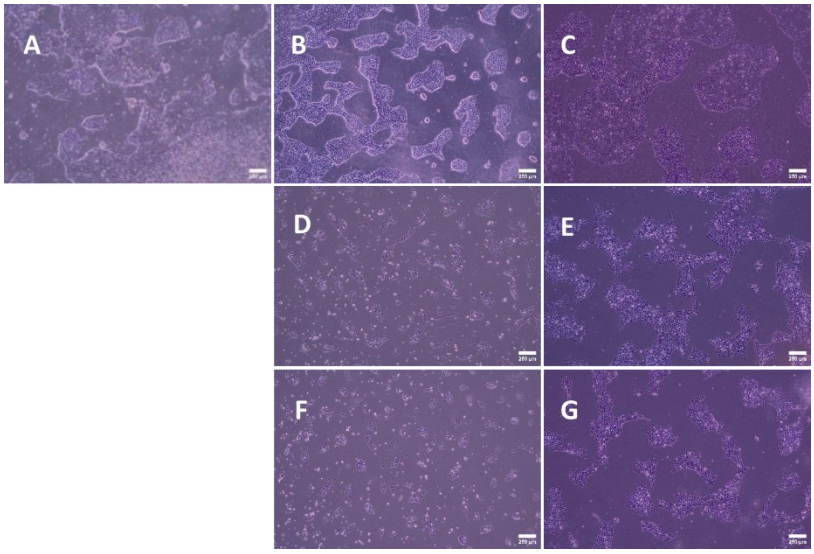

Figure 13. The effect of electroporation and targeted genome editing on the colony morphology and recovery patterns of the wild-type cell line (HEL252.2).
A: One day before electroporation.
B, C: Maintained regularly at day 2 and day 4 after 1:4 passage.
D, E: Day 2 and day 4 after electroporation with Cas9 and reverse single guide RNA (sgRNA).
F, G: Day 2 and day 4 after electroporation with Cas9 and forward sgRNA.
All images were taken with Olympus CKX53 microscope at 4x magnification. Scale bar = 250µm.

4.2.2 No morphological differences were observed between the cell lines during single cell derived colony expansion

After the low density seeding to isolate and expand single cell derived colonies, the iPSCs were kept in culture for several days supplemented with CloneR to promote cell survival. During the colony expansion phase, many neighbouring colonies appeared to have merged, forming larger colonies with typically irregular shapes, which were not selected for colony picking. Single cell derived colonies typically have a rounded shape and bright, smooth borders. The shape, size, and morphology of the colonies did not differ between the cell lines and enough colonies for characteristics indicative for single cell origin were identified on the 10cm plates (Figure 14).

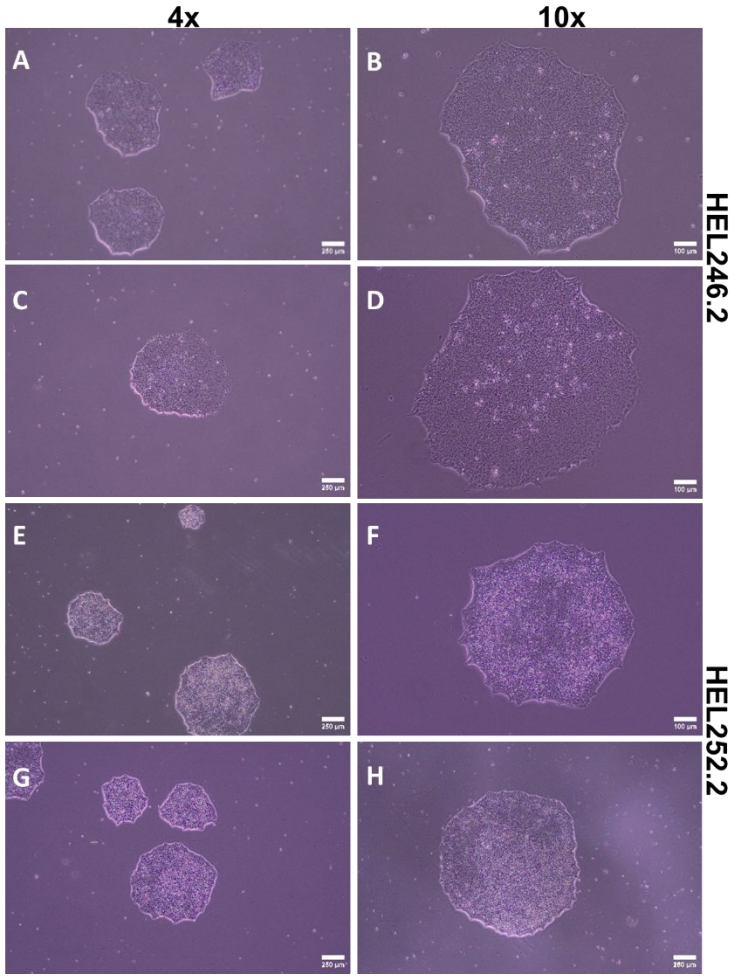

Figure 14. The ideal morphology of single cell derived colonies 7 days after low-density seeding. Circular colonies with clear, bright, and smooth edges are good indicators of single cell origin. No morphological differences were observed between the cell lines.
A, B: HEL246.2 electroporated with the reverse single guide RNA (sgRNA).
C, D: HEL246.2 electroporated with the forward sgRNA.
E, F: HEL252.2 electroporated with the reverse sgRNA.
G, H: HEL252.2 electroporated with the forward sgRNA.
All cells were cultured under the same conditions. All images were captured with Olympus CKX53 microscope. Images of the left column were taken at 4x magnification (scale bars = 250µm), whereas images of the right column were taken at 10x magnification (scale bars = 100µm).

4.2.3 T7 Endonuclease 1 assay indicates low to modest editing efficiency but provides inconclusive results.

T7E1 assay was used to assess the editing efficiency for each sgRNA. Cleavage by the enzyme was detected in all electroporated samples, indicating the presence of heteroduplexes and suggesting Cas9-mediated cleavage at the target locus (Figure 15). The homozygous wild-type cell line HEL252.2 yielded the expected cleavage bands, although very faint, indicating low to modest cleavage efficiency. However, this conclusion cannot be made for the heterozygous HEL246.2 because heteroduplexes are expected to form regardless of the editing. Also, rescue edits would ideally produce wild-type homozygotes, but the HDR efficiency had to be ~100% in that case. Similarly, the HEL246.2 Cas9 control and the HEL246.2 unedited control showed this faint cleavage pattern (Figure 15). However, faint bands were also observed at the HEL252.2 unedited control, which raises concerns about the reliability and specificity of the T7E1 assay due to potential background cleavage or endogenous mismatches. As expected, all replicate T7-untreated samples exhibited no cleavage.

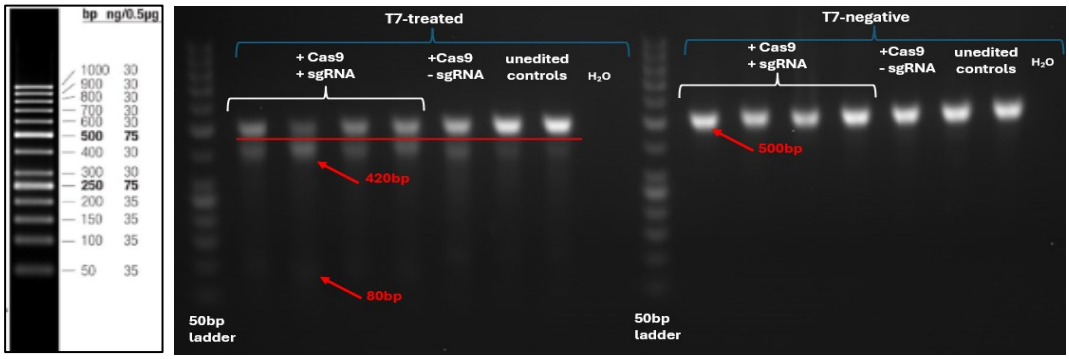

Figure 15. Agarose gel electrophoresis of the T7 Endonuclease 1 assay to Cas9-mediated cleavage. The expected band size of the homoduplexes is ~500 base pairs (bp). Heteroduplexes form mismatches upon denaturation-annealing and are cleaved by the enzyme, which yields ~420 bp and ~80 bp fragments on the gel. Cleavage bands were formed in all electroporated samples, but also in both the heterozygous Cas9-only control and unedited patient cell line, which hints for moderate CRISPR activity but also raises concerns about the assay's specificity. Left: All samples received T7 treatment. Right: Untreated controls. Samples by order for each group: HEL252.2_sg01_rv, HEL252.2_sg02_fw, HEL246.2_sg03_rv, HEL246.2_sg04_fw, HEL246.2_Cas9, HEL252.2 unedited, HEL246.2 unedited, water only control. Left picture: band sizes of the 50 bp DNA ladder (from ThermoFisher Scientific)

4.2.4 Three rounds of PCR result in the successful amplification of the barcoded sequences for sequencing by MiSeq.

Nested PCR successfully amplified the 200bp sequence around the *INTS6* target site. The first round of PCR amplified the broader target region using the 500bp primers, while the second round used the 200bp primers with partial sequencing adapters to further enrich the desired sequence within the first PCR products. A touchdown PCR program was needed to enhance

the sensitivity and ensure the selective amplification of the 200bp amplicons over the 500bp products (Figure 16). Then, indexing PCR was carried out to ligate the Illumina adapters and introduce the well-specific sequencing barcodes for MiSeq. After the third round of PCR, the expected size of the MiSeq amplicons carrying the sequencing adapters and barcodes was confirmed (Figure 17).

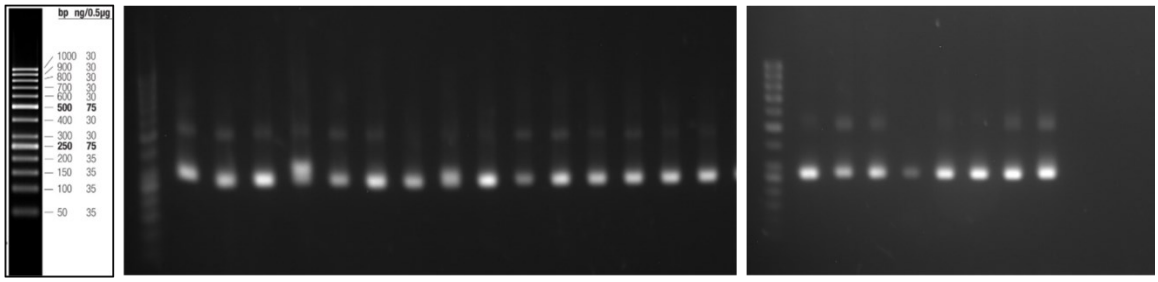

Figure 16. Agarose gel electrophoresis of PCR products after the second round of PCR, using the 200bp primers to amplify the *INTS6* target sequence. The expected size of the amplicons is ~265 base pairs (bp). Left gel: GeneRuler 50 bp DNA ladder, then PCR products from 5 random wells from each 96-well plate: HEL252.2_sg02_fw, HEL246.2_sg03_rv, HEL246.2_sg04_fw. Right gel: GeneRuler 50bp DNA ladder, 6 random wells from HEL252.2_sg01_rv, then two unedited positive controls (HEL252.2, HEL246.2), and two negative controls (water, no DNA template). Left image: band sizes of the 50 bp DNA ladder (acquired from ThermoFisher Scientific).

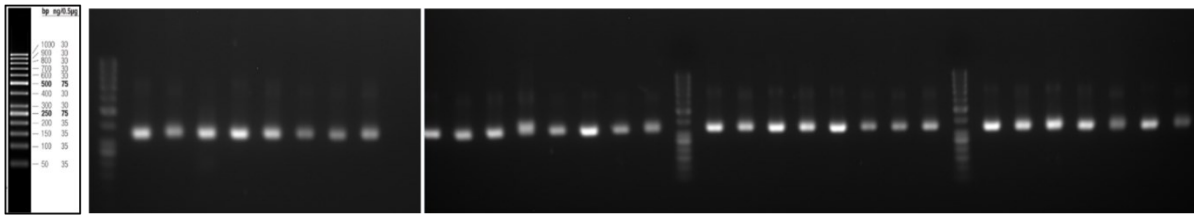

Figure 17. Agarose gel electrophoresis confirmed the successful amplification of the final PCR product with the sequencing barcodes before delivering for sequencing analysis. The expected amplicon size is ~340 base pairs (bp). Left gel: GeneRuler 50 bp DNA ladder, then PCR products from six random wells from HEL252.2_sg01_rv, followed by two positive unedited controls and two negative controls. Right gel: Eight random wells from each HEL252.2_sg02_fw, HEL246.2_sg03_rv, HEL246.2_sg04_fw in order, separated by GeneRuler 50bp DNA ladder. Left image: band sizes of the 50bp DNA ladder (ThermoFisher Scientific).

4.2.5 MiSeq data analysis results reveal successful HDR edits for one sgRNA

The sequencing reads for each sample were aligned with the given input sequences to determine the editing outcome. The forward sgRNAs were highly ineffective and they generated no perfect HDR events for both cell lines (Figure 18, Figure 19, Table 13). However, modified reference alleles were identified in both cases which means that the Cas9 indeed cleaved the target site, but NHEJ edits were predominant over HDR edits (Figure 18).

On the other hand, the reverse sgRNAs were more effective, as expected from their higher on-target prediction scores (Table 13). For the knock-in experiment, the reverse sgRNA was very

efficient as 51% of the clones exhibited HDR homozygous edits, which means that both alleles were knocked-out (Figure 18). Despite this, most of these colonies were growing normally in their wells, indicating that knocking out the *INTS6* gene was not lethal for the iPSCs *in vitro*. Surprisingly, only 4 clones carried HDR heterozygous edits (HDR = 40-60%) and just one of these was defined as successfully edited heterozygote with unmodified reference allele (Unmodified Reference = 40-60%) (Figure 19, Table 13). Nevertheless, since the introduced allele is the loss-of-function variant, successful heterozygotes can also be found from NHEJ events, provided that the mutations disrupt the gene function.

Table 13. Summary of the predicted and observed editing outcomes.

Sample ID	Cas9 cut-edit distance	sgRNA on-target score CRISPOR/ Benchling	Heterozygous edits (40-60% HDR)	Homozygous edits (>60% HDR)	Heterozygous edits (40-60% HDR) with unmodified reference / isogenic correction
HEL252.2_sg01_RV	4 bp	86 / 67.3	4 (4.2%)	49 (51%)	1 (1.0 %)
HEL252.2_sg02_FW	1 bp	6 / 53.3	0	0	0
HEL246.2_sg03_RV	4 bp	90 / 69.3	61 (63.5%)	6 (6.25%)	49 (51%)
HEL246.2_sg04_FW	1 bp	22 / 55.7	0	0	0

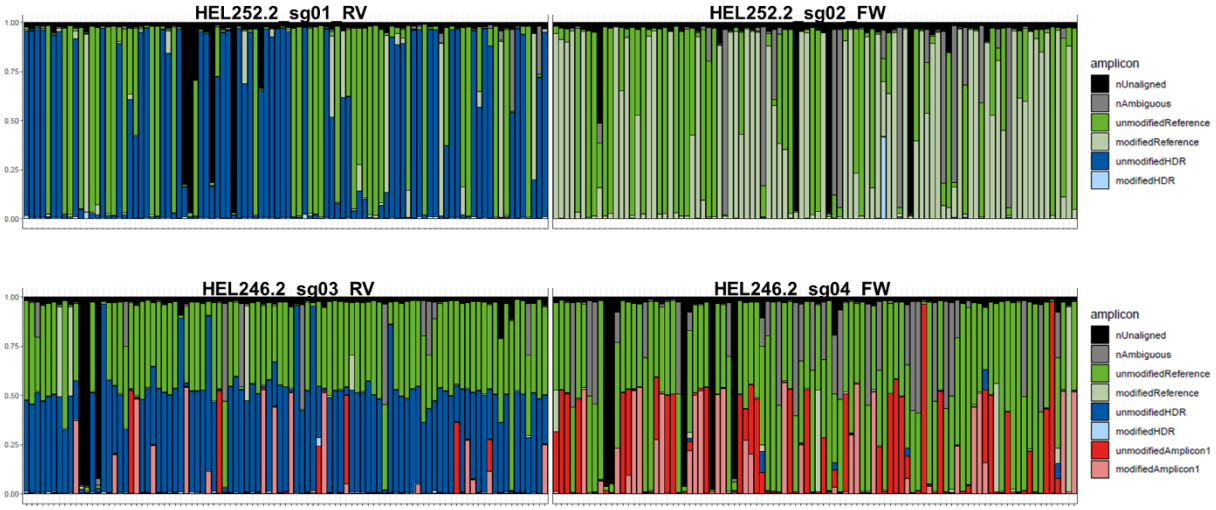

Figure 18. **Proportional stacked bar plot of sequencing reads per sample**. The "Reference" sequence reflects the wild-type allele, whereas the "Amplicon1" is the variant allele detected in the patient cell line in heterozygosity. The "HDR" sequences represent the designed homology-directed repair (HDR) template DNA which is unique for each single guide RNA (sgRNA). "Modified" symbolizes the presence of mutation(s) other than the expected. "Unmodified" denotes the unchanged input sequence. Each column refers to one clone. Y-axis depicts the proportion (0-1) of the read counts per sample. The sample IDs (x-axis) have been removed for simplicity. No HDR events were observed with the forward sgRNAs, although the presence of modified sequences indicates introduction of insertions and deletions by the non-homologous end-joining (NHEJ) repair mechanism. HEL252.2_sg01_RV was very efficient introducing homozygous knock-ins, while HEL246.2_sg03_RV produced isogenic corrections in half of the clones.

Notably, the reverse sgRNA for the isogenic rescue (HEL246.2) was very successful in generating heterozygous edits with unmodified wild-type allele (Table 13, Figure 18). The

clones with 40-60% unmodified HDR alleles and 40-60% unmodified reference (wild type) alleles were classified as successfully edited heterozygotes. 51% of the clones were rescued, while 6 additional clones exhibited homozygous HDR edits (>60% HDR, Table 13, Figure 19). The remarkable difference in the number of homozygous edits between the cell lines also pinpoints the allele-specificity of the sgRNAs for the rescue experiment where the sgRNAs were designed to target only the variant allele and not the wild-type sequence.

Figure 19. Bar plot depicting the percentage of reads mapped to the repair template sequences per sample. Each column refers to one individual clone. The sample names (x-axis) have been removed for simplicity. The percentage of homologous-directed repair (HDR) edits determines the success of the integration of the DNA repair template in the target genomic site. The forward single guide RNAs (sgRNAs) were very ineffective for either cell lines. In contrast, the reverse sgRNAs were very effective. Notably, more than half of the clones in the knock-in experiment had homozygous edits, while only 4 had heterozygous edits (40-60% HDR).

5 Discussion

5.1 sgRNA design principles and observed editing outcomes

Designing the CRISPR components greatly defines the success of precise genome editing. This depends on two events in the cellular level (Pacesa et al., 2024). At first, the Cas9 endonuclease must bind specifically to the target sequence guided by the sgRNA and introduce a DSB near the desired edit site. This process relies on the activity and specificity of the Cas9-sgRNA complex, determined by the *in silico* design of the sgRNA. Secondly, the editing relies on DNA repair mechanisms, and the prevalence of the HDR machinery is required to introduce a specific modification. The sgRNA activity, the position of the Cas9 cleavage, and the design of the exogenous repair template determine the HDR rate. These results demonstrated that with variant-specific adjustments, such as selecting an alternative nuclease with broader PAM requirements, efficient candidate sgRNAs could be designed for each genetic variant, underscoring the versatility of this genome editing tool.

Online design and score prediction tools were utilized to select the best candidate sgRNAs and DNA repair templates for each editing attempt. Though, it is necessary to acknowledge that these sgRNA on-target and off-target scores only provide an estimate of the sgRNA activity *in vitro*, and they do not always reflect experimental outcomes. These tools are particularly useful in identifying and ruling out highly ineffective sgRNAs, but beyond this the correlation between observed and expected editing outcomes has been reported to be weak (Bruntraeger et al., 2019; Haeussler et al., 2016). Therefore, it is essential to select and test at least two candidate sgRNAs per variant to maximize the likelihood of successful HDR (Bruntraeger et al., 2019).

As predicted by their on-target scores, the reverse sgRNAs were highly efficient and resulted in high ratio of bi-allelic HDR edits and only one successful mono-allelic edit in the knock-in experiment *in vitro*. The aim was to introduce the pathogenic *INTS6* variant in wild-type iPSCs to compare natural heterozygous patient-derived iPSCs to heterozygous iPSCs with wild-type genetic background. Modelling both heterozygous and knock-out iPSCs could also provide invaluable information about the gene's function, but in this experiment the generation of heterozygous edits was not as successful as the homozygous knock-in. The reverse sgRNA had high efficiency, leading to bi-allelic editing. A less efficient sgRNA could have been more suitable in generating heterozygous edits.

The approach of selecting at least two sgRNAs per edit was further supported by the fact that only the reverse sgRNAs successfully generated HDR edits for each cell line in the *INTS6*

experiment, whereas the forward sgRNAs failed to introduce any HDR edits. Although the failed sgRNAs were predicted to induce a DSB just 1 bp away from the target site and had medium efficiency scores, no HDR events occurred. The presence of modified reference sequences indicated that NHEJ-mediated edits were predominant over HDR, confirming that the sgRNAs effectively induced a DSB in the target site. Thus, the reason why the repair template did not integrate in the genome when using these sgRNAs may not be solely attributed to the sgRNA design. These findings highlight that other factors, such as chromatin accessibility, design of the HDR template, and experimental variability also influence the genome editing outcome and should be considered when interpreting CRISPR-editing results.

Moreover, it is worth noting that the allele-specific sgRNA design yielded allele-specific edits for both sgRNAs. Like the reverse sgRNAs, the forward sgRNAs used for each cell line differed only at a single nucleotide because the rescue experiment targeted the variant allele, with the sgRNA sequences overlapping the edit site (Table 3, Figure 5). In fact, the MiSeq results confirmed that both sgRNAs in the isogenic experiment mainly edited the target allele and not the wild-type allele. This highlights the specificity of the sgRNAs and supports the principle that mismatches in the seed region of the sgRNA can effectively prevent the Cas9 from cleaving the target sequence (Jinek et al., 2012; Hsu et al., 2013).

5.2 The cause of the low HDR rates for the forward sgRNAs remains unknown

In addition to the sgRNA design, genome editing success is defined by a wide range of experimental and design parameters. In this project, the delivery method of choice was the Cas9 RNP complex using synthetic sgRNAs via electroporation, a method known to be more effective, specific, and less stressful for iPSCs compared to transient plasmid transfection (Skarnes et al., 2019; Pacesa et al., 2024). Along with the Cas9 RNP, 100-nt long symmetric ssODNs were used as repair templates for single-base edits. Symmetric ssODNs are more efficient compared to dsODNs or asymmetric repair templates and provide higher recombination activity to facilitate higher HDR rates (Liang et al., 2017). To favour HDR over NHEJ after a genomic DNA lesion, a commercial HDR enhancer was added to the cell culture medium during electroporation and cold shock treatment was applied to the electroporated cells. These conditions have been previously shown to enhance editing via the HDR pathway and set the groundwork for an efficient workflow (Skarnes et al., 2019; Bruntraeger et al., 2019).

MiSeq analysis showed that the reverse sgRNA was very efficient in generating HDR edits in both cell lines, with over half of the clones being edited with the DNA repair template. In contrast, the NHEJ edits were more frequently observed with the forward sgRNA, hinting that

the sgRNA design or the delivery method was not the issue for failed HDR. HDR enhancer was included in all conditions, and given the high efficiency of the reverse sgRNAs, it is unlikely that it contributed to the poor performance of the forward sgRNAs.

All nucleofections were performed on the same day and all experimental procedures were carried out using the same reagents and under identical conditions. Therefore, the probability that a defective Cas9 nuclease or another reagent impacted the reduced HDR efficiency of the forward sgRNAs is low. Furthermore, since similar results were observed in both cell lines, a cell line specific effect can be ruled out too. In addition, the sgRNAs and ssODN HDR templates were resuspended on the same day using the same resuspension buffer and the subsequent DNA measurements confirmed their optimal DNA concentrations, giving little chances that resuspension issues were the root of cause.

Since NHEJ edits were present in both cell lines, the design of the ssODN HDR templates then rises as a potential reason for the inefficient HDR using the forward sgRNAs. The designed antisense oligos carried PAM blocking mutations to prevent re-cutting of the Cas9. Studies have shown that antisense symmetric oligos are more effective in iPSCs for single-base modifications (Okamoto et al., 2019), while others support that the orientation of the donor template does not impact HDR rates when the cut site is within 5 bp from the edit locus (Bruntraeger et al., 2019). Given these parameters, the ssODN HDR templates for the forward sgRNAs were optimally designed, suggesting that they are unlikely to be responsible for low HDR efficiency. To explore whether the ssODNs were the issue, further experiments using the same sgRNAs with a combination of ssODN templates of varying lengths and orientations should be carried out.

5.3 Limitations and suggestions for improvement

5.3.1 Restricted use of enough biological replicates and experimental conditions

The scope of this thesis was to optimize sgRNA design to edit specific variants associated with SCZ and ID cases and subsequently generate CRISPR-edited iPSCs to model a novel *INTS6* variant. Although this study fulfilled its aims with very promising results, some critical considerations need to be emphasized.

Most importantly, the laborious experimental workflow and the defined time period of the thesis did not allow the inclusion of more biological replicates. In this study, one biological sample (iPSC line) was used for each edit because the biggest workload was the clonal expansion, isolation, and genotyping to identify successfully edited iPSCs. To account for the

genetic variability between the donors, clonal variability within the same iPSC clone, and different off-target edits after the CRISPR editing, at least three independent biological replicates should be used and analysed (Bassett, 2017).

To further save on workload and material costs, the restriction of experimental conditions and sample size was necessary. The Cas9-only electroporated control was only included in the patient-derived cell line (HEL246.2), since HEL246.2 reached a stable growth rate faster than HEL252.2. When interpreting the T7E1 assay results, a Cas9-only control for HEL252.2 could provide more insights on the assay's specificity, and whether Cas9 alone influenced cell toxicity and DNA damage by randomly introducing DSBs without a sgRNA in place. The HEL246.2 is a natural heterozygote, so heteroduplexes would have been formed even with no editing, unlike the wild-type homozygote HEL252.2. However, none of these controls would have been used for clonal expansion and genotyping. Moreover, testing each sgRNA with more ssODN HDR templates could potentially yield more successful edits in case the design of the exogenous template was the limiting factor for HDR.

5.3.2 Experimental hurdles and room for improvements

The MiSeq analysis results showing the observed edits and HDR rates per clone were the main readout of this established experimental workflow. Though, earlier readouts, such as the T7E1 assay or comments on cell morphology and viability, did not provide conclusive evidence on whether the CRISPR editing and HDR edits were successful. Also, the differences in editing efficiencies between the sgRNAs were not captured by the T7E1 assay. The cleavage bands present in all electroporated conditions indicated low to moderate editing efficiency primarily by detecting indels caused by the NHEJ pathway, but no quantitative conclusion could be drawn. False positive bands were also detected in the unedited heterozygote samples due to endogenous mismatches, while homozygous biallelic mutations cannot be captured. Despite suggesting Cas9 activity, T7E1 assay was not a reliable method to assess the editing outcomes and estimate the percentage of HDR events in single-base modifications because of its low specificity and inaccurate detection.

Implementing additional detection methods at an earlier stage of the workflow could significantly save on material expenses and workload, especially when certain sgRNAs might be inefficient. For instance, bulk DNA Sanger sequencing from a pool of edited iPSCs and subsequent analysis could reveal whether the desired HDR edits were present in the mix of electroporated cells and hint on which sgRNAs and experimental conditions to prioritize. Nevertheless, it is more laborious than the T7E1 assay and would not substitute the typical experimental pipeline. A different strategy could be the implementation of a PCR-based

restriction enzyme recognition site assay, which provides higher accuracy for single-base modifications and short indels than the T7E1 assay in case a restriction site is available (Kim et al., 2014).

Another consideration lies on the lengthy colony isolation and screening process. The single-cell suspension prior to the low-density seeding procedure ensures the homogeneous distribution of single cells on the 10cm plate for the expansion of clonal colonies. During colony picking, only round-shaped, bright-edged colonies were selected for subcloning, which are indicators of a single-cell derived colony (Bruntraeger et al., 2019). However, the single-cell origin of each colony could not be guaranteed because colonies could have merged during culturing. Therefore, this possibility must be considered when interpreting sequencing results, as they may reflect a mixture of genotypes rather than a single clone. Re-sequencing of selected clones is essential to confirm the expected genotype. One modification to the protocol could be single-cell seeding at even lower densities, albeit this comes with lower chances of getting enough colonies to pick. An alternative approach is to employ fluorescence activated cell sorting to isolate iPSCs of interest based on the fluorescent signal transmitted by a tagged fluorescent marker (Caillaud et al., 2022). This efficient and robust method also comes with certain limitations though, including high risk of contamination and increased cell death (Caillaud et al., 2022).

Regarding the CRISPR experimental design, a few improvements to the standard protocol can be proposed. The inclusion of NHEJ inhibitors in addition to HDR enhancers could shift the editing towards more efficient HDR. To further increase HDR rates, an alternative method that covalently binds the ssODN molecule to the Cas9 RNP complex for spatiotemporal co-localisation of the repair template to the DSB has been proposed (Aird et al., 2018). Selecting the appropriate Cas9 system is also an important consideration when designing a CRISPR experiment. Whether an “NGG” PAM sequence is found close to the target site, the wild-type Cas9 nuclease should be preferred. Otherwise, an alternative nuclease with different PAM site recognition can be employed, such as Cas12 in AT-rich regions, or engineered Cas9 variants. On the same note, the ‘paired nicking strategy’ using a nickase with two different sgRNAs targeting opponent strands is demonstrated to be highly efficient and more specific than the wild-type Cas9 (Schubert et al., 2021). Moreover, novel CRISPR-based approaches for precise genome editing have emerged. Prime editors and base editors are new gene editing technologies that resemble the CRISPR/Cas9 mechanism, but do not require Cas9-mediated DSB and overcome the limitations of HDR (Pacesa et al., 2024).

5.3.3 Limitations of iPSCs

The value of iPSCs in human disease modelling is profound. iPSC technology is valuable in investigating neuropsychiatric polygenic disorders and provide high scalability and high-throughput applications. Nevertheless, significant challenges and technical limitations must be considered when using iPSCs for disease modelling. 2D iPSC cultures fail to recapitulate and mimic the complex *iv vivo* 3D tissue microenvironment and its spatiotemporal dynamics, which restricts the reproducibility and clinical relevance of the results (Bassett, 2017; Silva-Pedrosa et al., 2023). So far, cognitive deficits in neuropsychiatric disorders can only be studied in animal models, whereas *in vitro* models can only be used to assess the molecular and cellular disease phenotype (Cerneckis et al., 2024).

The inherent genetic variability and diverse differentiation potential of iPSCs poses an additional challenge (Silva-Pedrosa et al., 2023). Reprogramming somatic cells towards a pluripotent stem cell fate and their subsequent maintenance often results in the accumulation of genetic mutations, changes in epigenetic memory, and disruptions in X-inactivation process (Ho et al., 2015). These events can diversify gene expression and DNA methylation patterns among iPSC clones of the same donor and impact their phenotypic features and quality (Ho et al., 2015; Silva-Pedrosa et al., 2023). Also, donor-specific genetic and epigenetic information impacts their differentiation potential and robustness of experimental outcomes in disease mechanism studies (Ho et al., 2015; Kurishev et al., 2023). The differentiated cells might not functionally reflect the desired cell type since the differentiation protocols are not always efficient (Christian et al., 2020; Silva-Pedrosa et al., 2023). Therefore, careful validation and characterisation of the iPSCs and iPSC-derived cells is necessary. However, comparing isogenic patient-derived with wild-type cell cultures helps establish causality between the studied disease genotype and its underlying biological and molecular mechanisms.

The MiSeq results clearly disproved the initial hypothesis that the *INTS6* gene is lethal *in vitro*, at least at the iPSC stage, because homozygous knock-in and NHEJ edits that result in a knock-out genotype were identified in the colonies. This means that the increased cell death of the electroporated cells compared to the Cas9-only control could be attributed to other cytotoxic effects of the CRISPR/Cas9 editing, such as potential devastating off-targets that may led to chromosomal rearrangements and excessive DNA damage. Unlike heterozygous patients, the *INTS6*$^{-/-}$ knock-out genotype is not viable in the human population, probably because of the devastating effects of the gene's complete absence. However, the artificially generated *INTS6*$^{-/-}$ iPSCs *in vitro* were not lethal for the cells in their undifferentiated state. This highlights the fact that iPSCs and conventional iPSC-derived 2D *in vitro* models do not fully resemble the in

vivo physiological environment (Bassett, 2017). Therefore, the natural causal effect of genetic variants may be invisible *in vitro* for complex phenotypes where cell-cell interactions and epistatic effects are limited.

5.4 Future prospects

The desired HDR edits were generated for both iPSC lines in this experiment. The following step is the identification of these clones in the replicate 96-well plates and their re-introduction in the cell culture for expansion and re-sequencing to confirm their genotype (Figure 20). If the desired edited clones cannot be validated or retrieved, recovering more clones from the cryopreserved back-up samples and re-screening will be necessary. Otherwise, repeating the CRISPR protocol with different parameters, such as more efficient sgRNAs, ssODN HDR templates, or nucleases, could potentially lead to successfully edited clones. After confirming the genotype of the successfully CRISPR-edited iPSCs, further characterization is needed (Figure 20). Karyotyping and off-target screening will check the DNA integrity and genomic stability of the clones, while immunocytochemistry with stem cell markers can be employed to ensure that the iPSCs remain in pluripotent state (Figure 20). Moreover, increasing the number of biological samples to edit is also crucial to achieve statistical power for downstream applications.

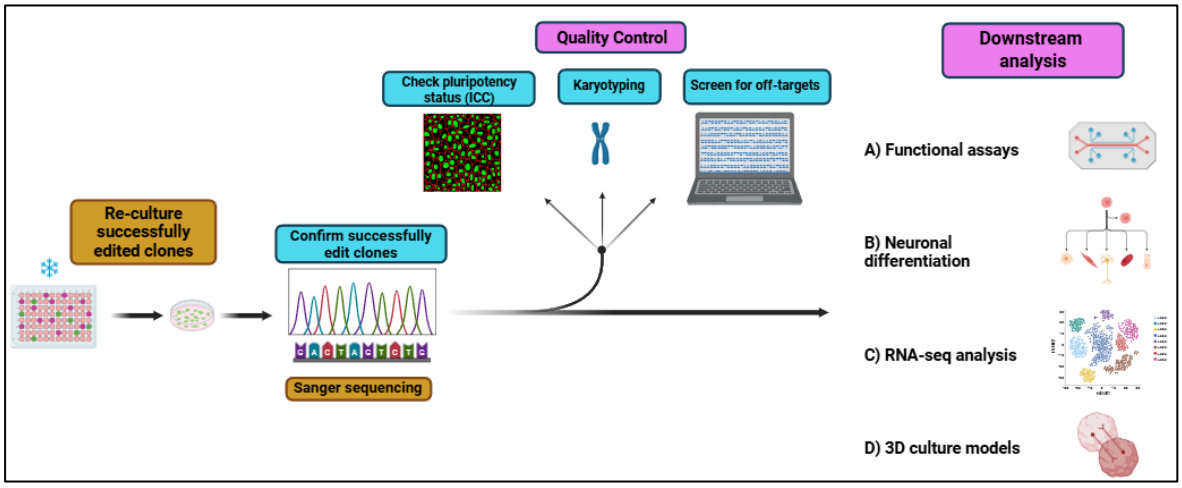

Figure 20. Future prospects upon identifying successfully edited induced pluripotent stem cells (iPSCs). The successfully edited clones will be re-introduced in the cell culture for expansion and re-sequencing. A quality control step is then necessary to confirm their pluripotency status and validate their normal karyotype and absence of devastating off-target edits. Differentiation protocols, as well as 2D and 3D cellular models can be then employed for disease modelling.

In polygenic disorders like SCZ, the cumulative effect of multiple variants is responsible for the disease's progression, making it challenging to investigate the causality of individual variants alone. Thus, in addition to the *INTS6* mutation, more rare genetic variants can be used for

CRISPR experiments and downstream analyses to model the pathogenesis of SCZ and ID and to generate additional data to further optimize the CRISPR editing design and experimental protocols. For instance, the variants in *CACNA1G* and *XPO7* genes used in this project hold the greatest potential for successful editing given the predictive scores of the available sgRNAs.

The crucial next step after generating characterized isogenic pairs of CRISPR-edited iPSCs with the intended modifications is their application in complex cellular systems for disease modelling. Functional analysis of these iPSC-derived isogenic pairs helps determine whether the studied genotype is sufficient and necessary to cause the pathogenic phenotype (Christian et al., 2020; Bassett, 2017). *INTS6* is a key modulator of the Integrator complex, and its disruption can result in improper processing of snRNA transcripts, and transcriptional dysregulation (Oegema et al., 2017; Welsh & Gardini, 2023). Neuronal differentiation and subsequent RNA-seq analysis of specific isogenic neuronal cell types would then provide valuable information about the transcriptomic effects of the gene (Figure 20). For variants in genes with distinct biological functions, such as ion transport (*CACNA1G*) or neuronal migration and growth (*TRIO*), different 2D and 3D cellular models and functional assays can also be employed based on the expected pathogenic phenotype to investigate biological processes of psychiatric disorders (Figure 20).

Traditional 2D cultures are highly accessible and easy to use in producing and manipulating individual neuronal cell types to study neurodegeneration and cell-cell interactions (Silva-Pedrosa et al., 2023). Though, the altered cellular behaviour and limited reproducibility compared to *iv vivo* conditions have led to the implementation of 3D cellular models, such as organoids, assembloids, and spheroids (Christian et al., 2020; Silva-Pedrosa et al., 2023). These *in vitro* systems represent closer the molecular dynamics and tissue architecture of the brain. Different isogenic CRISPR-edited iPSC-derived neuronal cell types can be integrated into such 3D models and complement the observations made in animal models and 2D cultures about the causality of specific genetic variants on the disease phenotype (Silva-Pedrosa et al., 2023).

5.5 Concluding remarks

The primary objective of this thesis was to explore the utility of the CRISPR/Cas9 gene editing tool to study the cause-and-effect relationship of genetic variants for disease modelling. It proposed an optimized design methodology and experimental workflow to generate isogenic pairs of CRISPR-edited human iPSCs. This is particularly relevant for disease modelling in neuropsychiatric disorders with a strong genetic component, such as SCZ, where mature brain cell types are post-mitotic and largely inaccessible.

18 coding variants in 8 genes associated with SCZ at exome-wide significance were used as an example to design sgRNAs and DNA repair templates for isogenic editing. It was evidenced that the CRISPR design methodology could be optimized for each variant, and that specific parameters could be tweaked to improve the *in silico* design of sgRNAs in some cases. Subsequently, a novel variant in the *INTS6* gene was modelled *in vitro* using wild-type and patient-derived iPSCs. Two sgRNAs were employed for each cell line aiming to introduce the variant allele in wild-type genetic background and to revert the genotype in the patient-derived cells. In the patient-derived cell line, only one Cas9-sgRNA complex successfully generated heterozygous corrections. For the knock-in experiment, the reverse sgRNA was highly efficient, producing bi-allelic edits in more than half of the clones, along with just one successful HDR heterozygous clone. In both cases, the forward Cas9-sgRNA complex failed to produce HDR edits, although indels were generated via the NHEJ pathway.

These findings suggest that although the CRISPR/Cas9 technology can be effective in human iPSCs, careful *in silico* design and experimental considerations are essential. Upon optimization of the workflow and further validation steps, the isogenic CRISPR-edited iPSCs can then be used for a wide range of downstream applications to understand the molecular mechanisms underlying complex genetic disorders, such as SCZ and ID. Combining CRISPR/Cas9 with iPSC technology and 3D cellular systems has brought tremendous possibilities in personalised medicine and therapeutics, getting us one step closer to treating genetic diseases. During the past decade, the rapidly evolving CRISPR technology has also been expanded beyond human disease modelling, bringing revolutionary advancements in biotechnology, drug discovery, epigenetics, and agriculture (Pacesa et al., 2024). It is therefore evident that this technology holds great promise for a range of applications and more innovations await to be discovered.

6 Acknowledgements

First and foremost, I would like to deeply thank my supervisors, Group Leader Olli Pietiläinen and PhD researcher Nelli Jalkanen for welcoming me in the group, and their constant and dedicated guidance and support throughout the thesis project. Nelli's supervision was instrumental, kind, and inspiring at all parts of the project, from guiding me in the cell culture and always being present, to assisting me during thesis writing and in 'debugging' the sequencing analysis issues, as well as creating a very supportive environment to work with. The primary guidance and insightful conversations with Olli helped me tremendously to identify and decode my thesis scope, in addition to the general advice and great mentorship.

My sincere thanks to all Pietiläinen group members for creating a pleasant, inclusive and warm working atmosphere, as well as to all Kilpinen group members for making me feel like home in the lab. Special thanks to Riina Lampela, Reyhane Eghtedarian, Adithi Sundaresh, Robin Forsen, and my GMB friend, Marc Carillo Perez, for their generous and professional advice on the CRISPR experiments and for creating a well-organized and clean workspace in the lab. Many thanks to Dimitri Meistermann for his valuable input in the MiSeq analysis pipeline.

I would also like to thank the iPS Core Facility of Biomedicum Stem Cell Center for providing the human iPSC lines I used in this project. In addition, my sincere appreciation to the Biomedicum Functional Genomics Unit at the Helsinki Institute of Life Science and Biocenter Finland at the University of Helsinki for their quick and reliable sequencing services.

Moreover, I wouldn't like to forget the tutors and responsible teachers of my study programme for their continuous and prompt support when needed and their open-door policy. Finally, huge thanks to my family, and all my friends in Finland and beyond for truly influencing and standing by me in both rough and pleasant times during my studies in Helsinki.

References

Aird, E. J., Lovendahl, K. N., St Martin, A., Harris, R. S., & Gordon, W. R. (2018). Increasing Cas9-mediated homology-directed repair efficiency through covalent tethering of DNA repair template. *Communications Biology*, *1*(1), 54. https://doi.org/10.1038/s42003-018-0054-2

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A., & Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, *315*(5819), 1709–1712. https://doi.org/10.1126/science.1138140

Bassett A. R. (2017). Editing the genome of hiPSC with CRISPR/Cas9: disease models. *Mammalian Genome*, *28*(7-8), 348–364. https://doi.org/10.1007/s00335-017-9684-9

Bolotin, A., Quinquis, B., Sorokin, A., & Ehrlich, S. D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*, *151*(8), 2551–2561. https://doi.org/10.1099/mic.0.28048-0

Brouns, S. J. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J. H., Snijders, A. P. L., Dickman, M. J., Makarova, K. S., Koonin, E. V., & Van Der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, *321*(5891), 960–964. https://doi.org/10.1126/science.1159689

Bruntraeger, M., Byrne, M., Long, K., & Bassett, A. R. (2019). Editing the Genome of Human Induced Pluripotent Stem Cells Using CRISPR/Cas9 Ribonucleoprotein Complexes. *Methods in molecular biology*, *1961*, 153–183. https://doi.org/10.1007/978-1-4939-9170-9_11

Caillaud, A., Lévêque, A., Thédrez, A., Girardeau, A., Canac, R., Bray, L., Baudic, M., Barc, J., Gaborit, N., Lamirault, G., Gardie, B., Idriss, S., Rimbert, A., May, C. L., Cariou, B., & Si-Tayeb, K. (2022). FACS-assisted CRISPR-Cas9 genome editing of human induced pluripotent stem cells. *STAR Protocols*, *3*(4), 101680. https://doi.org/10.1016/j.xpro.2022.101680

Cerneckis, J., Cai, H., & Shi, Y. (2024). Induced pluripotent stem cells (iPSCs): molecular mechanisms of induction and applications. *Signal Transduction and Targeted Therapy*, *9*(1). https://doi.org/10.1038/s41392-024-01809-0

Chang, H. H. Y., Pannunzio, N. R., Adachi, N., & Lieber, M. R. (2017). Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nature Reviews Molecular Cell Biology*, *18*(8), 495–506. https://doi.org/10.1038/nrm.2017.48

Cho, S. W., Kim, S., Kim, J. M., & Kim, J. S. (2013). Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nature biotechnology*, *31*(3), 230–232. https://doi.org/10.1038/nbt.2507

Christian, K. M., Song, H., & Ming, G. (2020). Using Two- and Three-Dimensional Human IPSC Culture Systems to model psychiatric disorders. *Advances in Neurobiology*, 237–257. https://doi.org/10.1007/978-3-030-45493-7_9

Clement, K., Rees, H., Canver, M. C., Gehrke, J. M., Farouni, R., Hsu, J. Y., Cole, M. A., Liu, D. R., Joung, J. K., Bauer, D. E., & Pinello, L. (2019). CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nature Biotechnology*, *37*(3), 224–226. https://doi.org/10.1038/s41587-019-0032-3

Concordet, J., & Haeussler, M. (2018). CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Research*, *46*(W1), W242–W245. https://doi.org/10.1093/nar/gky354

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., & Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science*, *339*(6121), 819–823. https://doi.org/10.1126/science.1231143

Cui, Y., Xu, J., Cheng, M., Liao, X., & Peng, S. (2018). Review of CRISPR/Cas9 sgRNA Design Tools. *Interdisciplinary sciences, computational life sciences*, *10*(2), 455–465. https://doi.org/10.1007/s12539-018-0298-z

Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., Eckert, M. R., Vogel, J., & Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, *471*(7340), 602–607. https://doi.org/10.1038/nature09886

Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., Virgin, H. W., Listgarten, J., & Root, D. E. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature biotechnology*, *34*(2), 184–191. https://doi.org/10.1038/nbt.3437

Doench, J. G., Hartenian, E., Graham, D. B., Tothova, Z., Hegde, M., Smith, I., Sullender, M., Ebert, B. L., Xavier, R. J., & Root, D. E. (2014). Rational design of highly active sgRNAs for CRISPR-Cas9–mediated gene inactivation. *Nature Biotechnology*, *32*(12), 1262–1267. https://doi.org/10.1038/nbt.3026

Elton, T. S., Hossain, M. I., Carvajal-Moreno, J., Wang, X., Skaggs, D. J., & Yalowich, J. C. (2022). Maximizing the efficacy of CRISPR/Cas homology-directed repair gene targeting. *IntechOpen eBooks*. https://doi.org/10.5772/intechopen.109051

Fu, Y., Foden, J. A., Khayter, C., Maeder, M. L., Reyon, D., Joung, J. K., & Sander, J. D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature Biotechnology*, *31*(9), 822–826. https://doi.org/10.1038/nbt.2623

Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M., & Joung, J. K. (2014). Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nature Biotechnology*, *32*(3), 279–284. https://doi.org/10.1038/nbt.2808

Genovese, G., Fromer, M., Stahl, E. A., Ruderfer, D. M., Chambert, K., Landén, M., Moran, J. L., Purcell, S. M., Sklar, P., Sullivan, P. F., Hultman, C. M., & McCarroll, S. A. (2016). Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. *Nature neuroscience*, *19*(11), 1433–1441. https://doi.org/10.1038/nn.4402

Guo, Q., Mintier, G., Ma-Edmonds, M., Storton, D., Wang, X., Xiao, X., Kienzle, B., Zhao, D., & Feder, J. N. (2018). 'Cold shock' increases the frequency of homology directed repair gene editing in induced pluripotent stem cells. Scientific Reports, 8(1), 2080. https://doi.org/10.1038/s41598-018-20358-5

Haeussler, M., Schönig, K., Eckert, H., Eschstruth, A., Mianné, J., Renaud, J., Schneider-Maunoury, S., Shkumatava, A., Teboul, L., Kent, J., Joly, J., & Concordet, J. (2016). Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biology*, *17*(1), 148. https://doi.org/10.1186/s13059-016-1012-2

Ho, S. M., Topol, A., & Brennand, K. J. (2015). From "directed differentiation" to "neuronal induction": modeling neuropsychiatric disease. *Biomarker insights*, *10*(Suppl 1), 31–41. https://doi.org/10.4137/BMI.S20066

Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., Li, Y., Fine, E. J., Wu, X., Shalem, O., Cradick, T. J., Marraffini, L. A., Bao, G., & Zhang, F. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology*, *31*(9), 827–832. https://doi.org/10.1038/nbt.2647

Ihry, R. J., Worringer, K. A., Salick, M. R., Frias, E., Ho, D., Theriault, K., Kommineni, S., Chen, J., Sondey, M., Ye, C., Randhawa, R., Kulkarni, T., Yang, Z., McAllister, G., Russ, C., Reece-Hoyes, J., Forrester, W., Hoffman, G. R., Dolmetsch, R., & Kaykas, A. (2018). p53 inhibits CRISPR–Cas9 engineering in human pluripotent stem cells. *Nature Medicine*, *24*(7), 939–946. https://doi.org/10.1038/s41591-018-0050-6

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., & Nakata, A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *Journal of bacteriology*, *169*(12), 5429–5433. https://doi.org/10.1128/jb.169.12.5429-5433.1987

Jansen, R., Van Embden, J. D. A., Gaastra, W., & Schouls, L. M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology*, *43*(6), 1565–1575. https://doi.org/10.1046/j.1365-2958.2002.02839.x

Jauhar, S., Johnstone, M., & McKenna, P. J. (2022). Schizophrenia. *Lancet (London, England)*, *399*(10323), 473–486. https://doi.org/10.1016/S0140-6736(21)01730-X

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable Dual-RNA–Guided DNA endonuclease in adaptive bacterial immunity. *Science*, *337*(6096), 816–821. https://doi.org/10.1126/science.1225829

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., & Doudna, J. (2013). RNA-programmed genome editing in human cells. *eLife*, *2*. https://doi.org/10.7554/elife.00471

Kim, J. M., Kim, D., Kim, S., & Kim, J. (2014). Genotyping with CRISPR-Cas-derived RNA-guided endonucleases. *Nature Communications*, *5*(1), 3157. https://doi.org/10.1038/ncomms4157

Kleinstiver, B. P., Prew, M. S., Tsai, S. Q., Topkar, V. V., Nguyen, N. T., Zheng, Z., Gonzales, A. P. W., Li, Z., Peterson, R. T., Yeh, J. J., Aryee, M. J., & Joung, J. K. (2015). Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature*, *523*(7561), 481–485. https://doi.org/10.1038/nature14592

Kurishev, A. O., Karpov, D. S., Nadolinskaia, N. I., Goncharenko, A. V., & Golimbet, V. E. (2023). CRISPR/Cas-based approaches to study schizophrenia and other neurodevelopmental disorders. *International journal of molecular sciences*, *24*(1), 241. https://doi.org/10.3390/ijms24010241

Kurki, M. I., Saarentaus, E., Pietiläinen, O., Gormley, P., Lal, D., Kerminen, S., Torniainen-Holm, M., Hämäläinen, E., Rahikkala, E., Keski-Filppula, R., Rauhala, M., Korpi-Heikkilä, S., Komulainen–Ebrahim, J., Helander, H., Vieira, P., Männikkö, M., Peltonen, M., Havulinna, A. S., Salomaa, V., . . . Palotie, A. (2019). Contribution of rare and common variants to intellectual disability in a sub-isolate of Northern Finland. *Nature Communications*, *10*(1), 410. https://doi.org/10.1038/s41467-018-08262-y

Lane, D. P. (1992). p53, guardian of the genome. *Nature*, *358*(6381), 15–16. https://doi.org/10.1038/358015a0

Liang, X., Potter, J., Kumar, S., Ravinder, N., & Chesnut, J. D. (2017). Enhanced CRISPR/Cas9-mediated precise genome editing by improved design and delivery of gRNA, Cas9 nuclease, and donor DNA. *Journal of biotechnology*, *241*, 136–146. https://doi.org/10.1016/j.jbiotec.2016.11.011

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E., & Church, G. M. (2013). RNA-guided human genome engineering via Cas9. *Science*, *339*(6121), 823–826. https://doi.org/10.1126/science.1232033

Mao, Z., Bozzella, M., Seluanov, A., & Gorbunova, V. (2008). Comparison of nonhomologous end joining and homologous recombination in human cells. *DNA Repair*, *7*(10), 1765–1771. https://doi.org/10.1016/j.dnarep.2008.06.018

Marshall, C. R., Howrigan, D. P., Merico, D., Thiruvahindrapuram, B., Wu, W., Greer, D. S., Antaki, D., Shetty, A., Holmans, P. A., Pinto, D., Gujral, M., Brandler, W. M., Malhotra, D., Wang, Z., Fajarado, K. V. F., Maile, M. S., Ripke, S., Agartz, I., Albus, M., Alexander, M., … CNV and Schizophrenia Working Groups of the Psychiatric Genomics Consortium (2017). Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nature genetics*, *49*(1), 27–35. https://doi.org/10.1038/ng.3725

Maurissen, T. L., & Woltjen, K. (2020). Synergistic gene editing in human iPS cells via cell cycle and DNA repair modulation. Nature Communications, 11(1), 2876. https://doi.org/10.1038/s41467-020-16643-5

Mitchell, K. J. (2012). What is complex about complex disorders? *Genome Biology*, *13*(1), 237. https://doi.org/10.1186/gb-2012-13-1-237

Mojica, F. J., Díez-Villaseñor, C., Soria, E., & Juez, G. (2000). Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Molecular microbiology*, *36*(1), 244–246. https://doi.org/10.1046/j.1365-2958.2000.01838.x

Mojica, F. J., Díez-Villaseñor, C., García-Martínez, J., & Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of molecular evolution*, *60*(2), 174–182. https://doi.org/10.1007/s00239-004-0046-3

Moreno-Mateos, M. A., Vejnar, C. E., Beaudoin, J. D., Fernandez, J. P., Mis, E. K., Khokha, M. K., & Giraldez, A. J. (2015). CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nature methods*, *12*(10), 982–988. https://doi.org/10.1038/nmeth.3543

Murray, R. M., & Lewis, S. W. (1988). Is schizophrenia a neurodevelopmental disorder?. *British medical journal (Clinical research ed.)*, *296*(6614), 63. https://doi.org/10.1136/bmj.296.6614.63

Nakamura, T., & Takata, A. (2023). The molecular pathology of schizophrenia: an overview of existing knowledge and new directions for future research. *Molecular psychiatry*, *28*(5), 1868–1889. https://doi.org/10.1038/s41380-023-02005-2

Niemi, M. E. K., Martin, H. C., Rice, D. L., Gallone, G., Gordon, S., Kelemen, M., McAloney, K., McRae, J., Radford, E. J., Yu, S., Gecz, J., Martin, N. G., Wright, C. F., Fitzpatrick, D. R., Firth, H. V., Hurles, M. E., & Barrett, J. C. (2018). Common genetic variants contribute to risk of rare severe neurodevelopmental disorders. *Nature*, *562*(7726), 268–271. https://doi.org/10.1038/s41586-018-0566-4

Oegema, R., Baillat, D., Schot, R., van Unen, L. M., Brooks, A., Kia, S. K., Hoogeboom, A. J. M., Xia, Z., Li, W., Cesaroni, M., Lequin, M. H., van Slegtenhorst, M., Dobyns, W. B., de Coo, I. F. M., Verheijen, F. W., Kremer, A., van der Spek, P. J., Heijsman, D., Wagner, E. J., Fornerod, M., … Mancini, G. M. S. (2017). Human mutations in integrator complex subunits link transcriptome integrity to brain development. *PLoS Genetics*, *13*(8), e1006923. https://doi.org/10.1371/journal.pgen.1006923

Okamoto, S., Amaishi, Y., Maki, I., Enoki, T., & Mineno, J. (2019). Highly efficient genome editing for single-base substitutions using optimized ssODNs with Cas9-RNPs. *Scientific Reports*, *9*(1), 4811. https://doi.org/10.1038/s41598-019-41121-4

Owen, M. J., & O'Donovan, M. C. (2017). Schizophrenia and the neurodevelopmental continuum: evidence from genomics. *World psychiatry*, *16*(3), 227–235. https://doi.org/10.1002/wps.20440

Owen, M. J., O'Donovan, M. C., Thapar, A., & Craddock, N. (2011). Neurodevelopmental hypothesis of schizophrenia. *The British journal of psychiatry*, *198*(3), 173–175. https://doi.org/10.1192/bjp.bp.110.084384

Pacesa, M., Pelea, O., & Jinek, M. (2024b). Past, present, and future of CRISPR genome editing technologies. *Cell*, *187*(5), 1076–1100. https://doi.org/10.1016/j.cell.2024.01.042

Panda, A., Suvakov, M., Mariani, J., Drucker, K. L., Park, Y., Jang, Y., Kollmeyer, T. M., Sarkar, G., Bae, T., Kim, J. J., Yoon, W. H., Jenkins, R. B., Vaccarino, F. M., & Abyzov, A. (2023). Clonally Selected Lines After CRISPR-Cas Editing Are Not Isogenic. *The CRISPR journal*, *6*(2), 176–182. https://doi.org/10.1089/crispr.2022.0050

Pinello, L., Canver, M. C., Hoban, M. D., Orkin, S. H., Kohn, D. B., Bauer, D. E., & Yuan, G. (2016). Analyzing CRISPR genome-editing experiments with CRISPResso. Nature Biotechnology, 34(7), 695–697. https://doi.org/10.1038/nbt.3583

Rees, E., Kendall, K., Pardiñas, A. F., Legge, S. E., Pocklington, A., Escott-Price, V., MacCabe, J. H., Collier, D. A., Holmans, P., O'Donovan, M. C., Owen, M. J., Walters, J. T. R., & Kirov, G. (2016). Analysis of intellectual disability copy number variants for association with schizophrenia. *JAMA psychiatry*, *73*(9), 963-969. https://doi.org/10.1001/jamapsychiatry.2016.1831

Sapranauskas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., & Siksnys, V. (2011). The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. *Nucleic acids research*, *39*(21), 9275–9282. https://doi.org/10.1093/nar/gkr606

Ripke, S., Neale, B. M., Corvin, A., Walters, J. T. R., Farh, K., Holmans, P. A., Lee, P., Bulik-Sullivan, B., Collier, D. A., Huang, H., Pers, T. H., Agartz, I., Agerbo, E., Albus, M., Alexander, M., Amin, F., Bacanu, S. A., Begemann, M., Belliveau, R. A., . . . O'Donovan, M. C. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, *511*(7510), 421–427. https://doi.org/10.1038/nature13595

Schubert, M. S., Thommandru, B., Woodley, J., Turk, R., Yan, S., Kurgan, G., McNeill, M. S., & Rettig, G. R. (2021). Optimized design parameters for CRISPR Cas9 and Cas12a homology-directed repair. *Scientific reports*, *11*(1), 19482. https://doi.org/10.1038/s41598-021-98965-y

Sentmanat, M. F., Peters, S. T., Florian, C. P., Connelly, J. P., & Pruett-Miller, S. M. (2018). A survey of Validation Strategies for CRISPR-CAS9 editing. *Scientific Reports*, *8*(1), 888. https://doi.org/10.1038/s41598-018-19441-8

Silva-Pedrosa, R., Salgado, A. J., & Ferreira, P. E. (2023). Revolutionizing disease modeling: the emergence of organoids in cellular systems. *Cells*, *12*(6), 930. https://doi.org/10.3390/cells12060930

Singh, T., Kurki, M. I., Curtis, D., Purcell, S. M., Crooks, L., McRae, J., Suvisaari, J., Chheda, H., Blackwood, D., Breen, G., Pietiläinen, O., Gerety, S. S., Ayub, M., Blyth, M., Cole, T., Collier, D., Coomber, E. L., Craddock, N., Daly, M. J., Danesh, J., … Barrett, J. C. (2016). Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. *Nature neuroscience*, *19*(4), 571–577. https://doi.org/10.1038/nn.4267

Singh, T., Poterba, T., Curtis, D., Akil, H., Eissa, M. A., Barchas, J. D., Bass, N., Bigdeli, T. B., Breen, G., Bromet, E. J., Buckley, P. F., Bunney, W. E., Bybjerg-Grauholm, J., Byerley, W. F., Chapman, S. B., Chen, W. J., Churchhouse, C., Craddock, N., Cusick, C. M., . . . Daly, M. J.

(2022). Rare coding variants in ten genes confer substantial risk for schizophrenia. *Nature*, *604*(7906), 509–516. https://doi.org/10.1038/s41586-022-04556-w

Singh, T., Walters, J. T. R., Johnstone, M., Curtis, D., Suvisaari, J., Torniainen, M., Rees, E., Iyegbe, C., Blackwood, D., McIntosh, A. M., Kirov, G., Geschwind, D., Murray, R. M., Di Forti, M., Bramon, E., Gandal, M., Hultman, C. M., Sklar, P., INTERVAL Study, UK10K Consortium, … Barrett, J. C. (2017). The contribution of rare variants to risk of schizophrenia in individuals with and without intellectual disability. *Nature genetics*, *49*(8), 1167–1173. https://doi.org/10.1038/ng.3903

Skarnes, W. C., Pellegrino, E., & McDonough, J. A. (2019). Improving homology-directed repair efficiency in human stem cells. *Methods (San Diego, Calif.)*, *164-165*, 18–28. https://doi.org/10.1016/j.ymeth.2019.06.016

Song, J., Yao, S., Kowalec, K., Lu, Y., Sariaslan, A., Szatkiewicz, J. P., Larsson, H., Lichtenstein, P., Hultman, C. M., & Sullivan, P. F. (2022). The impact of educational attainment, intelligence and intellectual disability on schizophrenia: a Swedish population-based register and genetic study. *Molecular Psychiatry*, *27*(5), 2439–2447.https://doi.org/10.1038/s41380-022-01500-2

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., & Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*, *131*(5), 861–872. https://doi.org/10.1016/j.cell.2007.11.019

Takahashi, K., & Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell*, *126*(4), 663–676. https://doi.org/10.1016/j.cell.2006.07.024

Trubetskoy, V., Pardiñas, A. F., Qi, T., Panagiotaropoulou, G., Awasthi, S., Bigdeli, T. B., Bryois, J., Chen, C., Dennison, C. A., Hall, L. S., Lam, M., Watanabe, K., Frei, O., Ge, T., Harwood, J. C., Koopmans, F., Magnusson, S., Richards, A. L., Sidorenko, J., . . . Van Os, J. (2022). Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature*, *604*(7906), 502–508. https://doi.org/10.1038/s41586-022-04434-5

Walton, R. T., Christie, K. A., Whittaker, M. N., & Kleinstiver, B. P. (2020). Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science*, *368*(6488), 290–296. https://doi.org/10.1126/science.aba8853

Weinberger D. R. (1987). Implications of normal brain development for the pathogenesis of schizophrenia. *Archives of general psychiatry*, *44*(7), 660–669. https://doi.org/10.1001/archpsyc.1987.01800190080012

Welsh, S. A., & Gardini, A. (2023). Genomic regulation of transcription and RNA processing by the multitasking Integrator complex. *Nature reviews. Molecular cell biology*, *24*(3), 204–220. https://doi.org/10.1038/s41580-022-00534-2

Xue, C., & Greene, E. C. (2021). DNA repair pathway choices in CRISPR-CAS9-Mediated Genome Editing. *Trends in Genetics*, *37*(7), 639–656. https://doi.org/10.1016/j.tig.2021.02.008

Appendices

Appendix 1: Sequences of all the in silico designed sgRNAs and repair templates to edit the selected SUPER variants.

Supplementary Table 1. Sequences and scores of all designed single-guide RNAs (sgRNAs) to edit the selected variants in the SUPER cohort. The SpCas9 nuclease is predicted to introduce a double-strand break within 10 base pairs from the target site for all designed sgRNAs, to maximize the likelihood of homology-directed repair (HDR). All sgRNAs were 20 nucleotides long. 5'-NGG-3' PAM sequences are recognized by the widely used Cas9 nuclease, while 5'-NGN-3' PAM sites (highlighted in bold) are recognized by the SpG nuclease, an engineered variant of the Cas9. 'N' denotes any nucleotide. For each variant, two isogenic edits were engineered: a knock-in edit to introduce the pathogenic allele in a wild-type cell line, and a rescue edit to reverse the variant back to the wild-type sequence. The cell line IDs corresponding to each variant can be cross-referenced in **Table 1**. Higher on-target scores (Benchling: 0-100, CRISPOR: -200 - +200) reflect higher cleavage efficiency, whereas higher off-target scores (0-100) indicate higher predicted sgRNA specificity. Per edit, the designed sgRNAs were sorted based on their proximity to the edit site and their on-target and off-target scores.

Gene Name	*Pathogenic variant*	*Edit type*	*PAM (5' -> 3')*	*sgRNA sequence (5' -> 3')*	*Edit – cut site distance (bp)*	*Benchling / CRISPOR on-target scores*	*Benchling / CRISPOR off-target score (CFD)*
CACNA1G	chr17:50624357:CAGG:C	Knock-in	AGG	GGACAAGACGGAGCCTGTGG	1	60.5 / 91	52 / 78
			GGG	GACAAGACGGAGCCTGTGGA	2	60.3 / 91	57.3 / 81
			TGG	AACGGACAAGACGGAGCCTG	2	58.9 / 78	77.2 / 89
			CGG	CTGGGAGTGAACGGACAAGA	10	54.9 / 45	60.2 / 88
		Rescue	TGG	GTGAACGGACAAGACGGAGG	2	56.8 / 104	80.7 / 91
			AGG	GGAGTGAACGGACAAGACGG	5	62.2 / 132	81.6 / 88
			GGG	ACGGACAAGACGGAGGTGGA	2	64.3 / 48	61 / 81
			AGG	AACGGACAAGACGGAGGTGG	1	51.8 / 32	80.7 / 91
			CGG	CTGGGAGTGAACGGACAAGA	7	54.9 / 35	60.2 / 78
CUL1	chr7:148790376:T:C	Knock-in	CGG	ACAGTGGCCGAAAATTGACG	3	73.3 / 98	85 / 93
			GGT	CAGTGGCCGAAAATTGACGT	2	64.3 / 96	43.4 / 86
			CGT	TTAGACAACTGATATAACCA	2	84.9 / 150	27.7 / 37

Gene Name	*Pathogenic variant*	*Edit type*	*PAM (5' -> 3')*	*sgRNA sequence (5' -> 3')*	*Edit – cut site distance (bp)*	*Benchling / CRISPOR on-target scores*	*Benchling / CRISPOR off-target score (CFD)*
		Rescue	CGG	ACAGTGGCCGAAAATTGACG	3	76.3 / 104	85 / 93
			GGT	CAGTGGCCGAAAATTGACGC	2	48.4 / 103	82.1 / 61
			CGT	TTAGACAACTGATATAACCG	2	95.5 / 158	54.5 / 44
HERC1	chr15:63746986:G:A	Knock-in	AGG	TAGCTACCAGCATTCTCGGG	7	69 / 109	86.9 / 93
			AGG	TACCAGCATTCTCGGGAGGC	2	52.9 / -23	77.3 / 86
			AGG	AGCATTCTCGGGAGGCAGGC	1	49.5 / -83	58.4 / 81
		Rescue	AGG	TACCAGCATTCTCGGGAGGT	2	62.9 / 27	81.1 / 91
			AGG	TAGCTACCAGCATTCTCGGG	7	69.3 / 102	86.9 / 93
			AGG	AGCATTCTCGGGAGGTAGGC	1	51.1 / -85	71.4 / 80
HERC1	chr15:63775279:A:C	Knock-in	TGG	AAAGTGCATAAAAGACACGC	9	64.4 / 103	76.3 / 86
		Rescue	TGG	AAAGTGCCTAAAAGACACGC	9	65.4 / 83	80.2 / 86
			AGG	TGTACTCCAGCGTGTCTTTT	5	20 / -182	72.6 / 89
HERC1	chr15:63655874:G:A	Knock-in	AGG	TTTCTACCCAGCTTTCTCCG	2	68.6 / 71	66.3 / 84
			GGG	GCAATTCCTCGGAGAAAGCT	7	57.6 / 18	68 / 88
			CGG	TGTTCTCTTCAGCAATTCCT	4	50.9 / -39	53.4 / 77
			TGG	AGCAATTCCTCGGAGAAAGC	6	44.1 / -61	70.9 / 87
		Rescue	AGG	TTTCTACCCAGCTTTCTCTG	2	67.8 / 27	44.5 / 65
			GGG	GCAATTCCTCAGAGAAAGCT	7	58.3 / 34	50.9 / 80
			TGG	AGCAATTCCTCAGAGAAAGC	6	45 / -40	49 / 74
HERC1	chr15:63713615:G:A	Knock-in	**AGA**	GTAGCTCGTAGCCGAGACCG	1	83.6 / 128	93.8 / --
			AGA	CGTGAAGTAGCTCGTAGCCG	6	66.8 / 63	93.3 / --
		Rescue	**AGA**	GTAGCTCGTAGCCGAGACTG	1	51.5 / 128	90.1 / 82

Gene Name	*Pathogenic variant*	*Edit type*	*PAM (5' -> 3')*	*sgRNA sequence (5' -> 3')*	*Edit – cut site distance (bp)*	*Benchling / CRISPOR on-target scores*	*Benchling / CRISPOR off-target score (CFD)*
			AGA	CGTGAAGTAGCTCGTAGCCG	6	70.2 / 63	90.3 / 91
			AGC	CTATCTCAGTCTCGGCTACG	8	63.5 / 70	92.6 / 93
HERC1	chr15:63656291:G:A	Knock-in	CGG	ACATTAATCGAACTAGCGTT	8	63.1 / 7	92.7 / 97
		Knock-in	TGG	GCTAGTTCGATTAATGTGCT	10	54 / -12	86.5 / 93
		Rescue	CGG	ACATTAATCAAACTAGCGTT	8	59.1 / -6	82.8 / 93
		Rescue	TGG	GCTAGTTTGATTAATGTGCT	10	53.1 / -67	62.9 / 85
HERC1	chr15:63693963:C:T	Knock-in	AGG	ACTGAGAAGAAAGATTTCAG	6	71.6 / 65	43.9 / 58
		Rescue	AGG	ACTGAGAAGAAAGATTTCAG	5	34.5 / 65	36.4 / --
			AGC	AAGATTTCAGAGATAATGTA	4	21.4 / 44	34.3 / --
HERC1	chr15:63637595:CTG:C	Knock-in	TGG	CCAGGCCAATGGCACAGTGT	1	48.3 / 29	59 / 77
			TGG	ACCTTTGTCATCCAGGCCAA	8	49.5 / 27	60 / 78
			TGG	TGGCACAGTGTTGGCTTGTG	10	48 / 24	63 / 79
			TGG	ACAAGCCAACACTGTGCCAT	5	45.5 / -30	66.4 / 89
			GGG	AATGGCACAGTGTTGGCTTG	8	40 / -15	63.2 / 75
		Rescue	TGG	ATCCAGGCCAATGGCAGTGT	1	41.4 / 39	57.8 / 82
			GGG	AATGGCAGTGTTGGCTTGTG	9	42.5 / 47	58.3 / 74
			TGG	ACCTTTGTCATCCAGGCCAA	8	48.9 / 27	60 / 78
			TGG	CCACAAGCCAACACTGCCAT	3	47.5 / -55	65.1 / 90
			TGG	CCAATGGCAGTGTTGGCTTG	7	43.9 / -45	65.2 / 84
RB1CC1	chr8:52657117:C:CTGAG	Knock-in	AGG	AGAGTTAGTATGCCTTGAGG	4	69.2 / 118	69.8 / 88
			AGG	GGGAGAGTTAGTATGCCTTG	7	80.2 / 60	78.9 / 89
			AGG	TATTTTGTAAAACCTCCTCA	3	64.8 / -75	55.1 / 72

Gene Name	*Pathogenic variant*	*Edit type*	*PAM (5' -> 3')*	*sgRNA sequence (5' -> 3')*	*Edit – cut site distance (bp)*	*Benchling / CRISPOR on-target scores*	*Benchling / CRISPOR off-target score (CFD)*
		Rescue	AGG	TTAGTATGCCTTGAGGACTC	1	62.5 / 19	75.6 / 86
			AGG	GGGAGAGTTAGTATGCCTTG	7	83 / 50	78.9 / 89
			AGG	TTGTAAAACCTGAGTCCTCA	3	61.5 / -107	55.7 / 85
SETD1A	chr16:30966112:C:CA	Knock-in	AGG	CAGAGGGGCATACTCACGGG	3	71.1 / 95	85.5 / 97
			GGG	GGGCAGAGGGGCATACTCAC	0	57.6 / 34	74.2 / 88
			CGG	AGGGCAGAGGGGCATACTCA	1	54.9 / 19	71.1 / 89
		Rescue	AGG	AGAGGGGCATACTCAACGGG	4	69.8 / 90	90.2 / 96
			CGG	GGGCAGAGGGGCATACTCAA	1	59.1 / 3	73.8 / 88
			GGG	GGCAGAGGGGCATACTCAAC	0	48.5 / -41	85.1 / 95
SETD1A	chr16:30961466:AG:A	Knock-in	AGG	CACCAGCACTCGGGGCGCCA	4	54.3 / 15	73 / 87
			CGG	CACTCGGGGCGCCAAGGAAA	1	51.4 / 0	70.6 / 88
			TGG	GAGGTTTTTGACCGTTTCCT	1	42.7 / -99	80.7 / 87
		Rescue	CGG	GCACTCGGGGCGCCAAGAAA	1	42.2 / 4	87.3 / 96
			TGG	GGAGGTTTTTGACCGTTTCT	2	32.6 / -146	81.1 / 91
SETD1A	chr16:30963433:G:T	Knock-in	GGG	TCATTCGTTGTTGTCCTGGA	2	52.9 / 24	79.8 / 84
			GGG	TCGTTGTTGTCCTGGAGGGA	6	58.7 / 23	57.4 / 84
			AGG	TTCGTTGTTGTCCTGGAGGG	5	50.6 / -21	58.4 / 81
			AGG	TTCATTCGTTGTTGTCCTGG	1	57.6 / -50	69.2 / 84
			TGG	TATTTCATTCGTTGTTGTCC	1	45.2 / -80	72.2 / 83
		Rescue	GGG	TCATTCGTTGTTGTACTGGA	2	57.2 / 50	84.5 / 89
			GGG	TCGTTGTTGTACTGGAGGGA	6	60.2 / 36	71.3 / 89
			AGG	TTCGTTGTTGTACTGGAGGG	5	56.3 / -15	72.1 / 92

Gene Name	*Pathogenic variant*	*Edit type*	*PAM (5' -> 3')*	*sgRNA sequence (5' -> 3')*	*Edit – cut site distance (bp)*	*Benchling / CRISPOR on-target scores*	*Benchling / CRISPOR off-target score (CFD)*
			AGG	TTCATTCGTTGTTGTACTGG	1	51.5 / -61	78.6 / 82
			TGG	TATTTCATTCGTTGTTGTAC	1	43.8 / -112	73.2 / 85
SETD1A	chr16:30966303:C:T	Knock-in	**AGA**	ATGAAGAGCATCATGCAGCG	1	60 / 65	70.6 / --
			TGC	GTTGAGGTCTCGCTGCATGA	5	51 / 97	77.5 / --
		Rescue	AGG	TCATGCAGTGAGACCTCAAC	8	67.2 / 59	62.4 / --
			CGC	TCTCCACCATCTTGCGGTTG	10	49.8 / -15	83.2 / 92
SP4	chr7:21429774:TAATC:T	Knock-in	**TGC**	AGTATAGCTTGATTATTACC	2	11.1 / 53	73.3 / -
			AGA	ATAGCTTGATTATTACCTGC	5	60.7 / 31	75.6 / -
			AGC	ATTTCTGCAGGTAATAATCA	2	6.4 / 11	67.6 / -
			TGA	TAGCAGCTGTGAGTATAGCT	8	29.2 / 52	75.9 / -
			AGG	CAAATTCAGCTCATTTCTGC	9	54 / -142	53.2 / 82
		Rescue	**TGC**	TGTGAGTATAGCTTATTACC	2	12.2 / 74	76.9 / -
			AGA	GAGTATAGCTTATTACCTGC	5	44.2 / 43	79.8 / -
			GGT	AAATTCAGCTCATTTCTGCA	8	34.5 / 0	63.1 / -
			AGC	GCTCATTTCTGCAGGTAATA	1	32.3 / -87	72.1 / -
			AGG	CAAATTCAGCTCATTTCTGC	9	54 / -142	53.2 / 82
SP4	chr7:21481973:G:T	Knock-in	TGG	ATCTGTCATATTGAAGGATG	2	42.2 / 27	55 / 78
			GGT	TCTGTCATATTGAAGGATGT	2	24.5 / 31	73.8 / -
			TGT	ATATCTGTCATATTGAAGGA	0	52.1 / 8	67.3 / -
			GGA	AGCATATCTGTCATATTGAA	3	44.3 / -44	70.8 / -
			AGG	CAGCATATCTGTCATATTGA	1	41.5 / -77	67.2 / 85
		Rescue	TGG	ATCTGTCATATTGAATGATG	1	45.9 / 14	57.4 / 78

Gene Name	*Pathogenic variant*	*Edit type*	*PAM (5' -> 3')*	*sgRNA sequence (5' -> 3')*	*Edit – cut site distance (bp)*	*Benchling / CRISPOR on-target scores*	*Benchling / CRISPOR off-target score (CFD)*
			GGT	TCTGTCATATTGAATGATGT	2	34.1 / 5	73.6 / -
			TGT	ATATCTGTCATATTGAATGA	0	34.3 / -11	69.7 / -
TRIO	chr5:14472658:G:GGTGTGTA	Knock-in	AGG	GGACCTTGGCTATGTGGTTG	4	61 / 23	64.9 / 82
			TGG	TGTGCGGGACCTTGGCTATG	10	64.7 / 52	78.8 / 83
		Rescue	AGG	GGACCTTGGCTATGTGGTTG	4	61 / 23	64.9 / 82
			TGG	TGTGCGGGACCTTGGCTATG	10	64.7 / 52	78.8 / 83
XPO7	chr8:21976513:G:A	Knock-in	GGG	TTACCTGATCTCCAGCTGGT	5	61.7 / 83	62 / 84
			TGG	TCTGTTACCTGATCTCCAGC	1	57.4 / 34	68.3 / 85
			TGG	GTTACCTGATCTCCAGCTGG	4	56.5 / 44	55.5 / 83
			TGG	CAGTGCAGATTCCCACCAGC	4	55.9 / 6	65.9 / 79
		Rescue	GGG	TTACCTGATCTCTAGCTGGT	5	61.7 / 90	71.4 / 84
			TGG	GTTACCTGATCTCTAGCTGG	7	57.6 / 61	71.6 / 85
			AGG	ATTCCCACCAGCTAGAGATC	4	55.1 / 18	69.4 / 85
			TGG	TCTGTTACCTGATCTCTAGC	1	57.2 / -6	71.3 / 85

Supplementary Table 2. Sequences of all single-strand DNA (ssDNA) repair templates for each single guide RNA (sgRNA) to edit the selected variants in the SUPER cohort.
For each isogenic edit, 100-nucleotide long ssDNA repair templates were designed for each of the top two or three ranked sgRNAs, according to the **Supplementary Table 1**. The intended edits are denoted in red letters, whereas the PAM site blocking / silent mutations are highlighted in orange. The asterisks at the first and last two nucleotides indicate the phosphorylated nucleotides.

Gene	Pathogenic variant	Edit type	sgRNA sequence (5' -> 3')	ssDNA repair template sequence (5' -> 3')	Orientation
CACNA1G	chr17:50624357:CAGG:C	Knock-in	GGACAAGACGGAGCCTGTGG	C*T*CCAGCTCCATTCTCTCCCCCCACCCCTCCCCC GCTTCCCTCCCTCCACCTCCGAGCTGTCCGTTCAC TCCCAGCCAGCAGATACCAGCTACATCCT*G*C	Sense
			GACAAGACGGAGCCTGTGGA	C*T*CCAGCTCCATTCTCTCCCCCCACCCCTCCCCC GCTTCCCTCCCTCCACCTCCGAGCTGTCCGTTCAC TCCCAGCCAGCAGATACCAGCTACATCCT*G*C	Sense
			AACGGACAAGACGGAGCCTG	C*T*CCAGCTCCATTCTCTCCCCCCACCCCTCCCCC GCTTCCCTCCCTCCACCTCCGAGCTGTCCGTTCAC TCCCAGCCAGCAGATACCAGCTACATCCT*G*C	Sense
		Rescue	GTGAACGGACAAGACGGAGG	T*C*CAGCTCCATTCTCTCCCCCCACCCCTCCCCCG CTTCCCTCCCTCCACAGGCTCCGTCTTGTCCGTTC ACTCCCAGCCAGCAGATACCAGCTACATC*C*T	Sense
			GGAGTGAACGGACAAGACGG	T*C*CAGCTCCATTCTCTCCCCCCACCCCTCCCCCG CTTCCCTCCCTCCACAGGCTCCGTCTTGTCCGTTC ACTCCCAGCCAGCAGATACCAGCTACATC*C*T	Sense
			ACGGACAAGACGGAGGTGGA	T*C*CAGCTCCATTCTCTCCCCCCACCCCTCCCCCG CTTCCCTAACTCCACAGGCTCCGTCTTGTCCGTTC ACTCCCAGCCAGCAGATACCAGCTACATC*C*T	Sense
CUL1	chr7:148790376:T:C	Knock-in	ACAGTGGCCGAAAATTGACG	A*G*CGATTCACAGCTTTCTACGCCAGCCGCCACA GTGGCCGAAAATTGACGCGCTTATATCAGTTGTCT AAAGGAGAATTGGTAACTAACTGCTTCAAA*A*A	Sense
			CAGTGGCCGAAAATTGACGT	A*G*CGATTCACAGCTTTCTACGCCAGCCGCCACA GTGGCCGAAAATTGACGCGCTTATATCAGTTGTCT AAAGGAGAATTGGTAACTAACTGCTTCAAA*A*A	Sense
			TTAGACAACTGATATAACCA	A*G*CGATTCACAGCTTTCTACGCCAGCCGCCACA GTGGCCGAAAATTGACGCGACTATATCAGTTGTCT AAAGGAGAATTGGTAACTAACTGCTTCAAA*A*A	Sense
		Rescue	ACAGTGGCCGAAAATTGACG	A*G*CGATTCACAGCTTTCTACGCCAGCCGCCACA GTGGCCGAAAACTTACTTGGTTATATCAGTTGTCTA AAGGAGAATTGGTAACTAACTGCTTCAAA*A*A	Sense
			CAGTGGCCGAAAATTGACGC	A*G*CGATTCACAGCTTTCTACGCCAGCCGCCACA GTGGCCGAAAACTTACGTGGTTATATCAGTTGTCT AAAGGAGAATTGGTAACTAACTGCTTCAAA*A*A	Sense

Gene	Pathogenic variant	Edit type	sgRNA sequence (5’ -> 3’)	ssDNA repair template sequence (5’ -> 3’)	Orientation
			TTAGACAACTGATATAACCG	A*G*CGATTCACAGCTTTCTACGCCAGCCGCCACA GTGGCCGAAAATTGACGTGGCTTTATCAGTTGTCT AAAGGAGAATTGGTAACTAACTGCTTCAAA*A*A	Sense
HERC1	chr15:63746986:G:A	Knock-in	TAGCTACCAGCATTCTCGGG	A*C*AGTTGAGTCCATCAGTCTGAAGAGCAAATTTC GAAGTGGACCTGCCTATCTCCCGAGAATGCTGGTA GCTACCCCTCCCGCAAGTGCAAGAGCAAG*G*T	Antisense
			TACCAGCATTCTCGGGAGGC	A*C*CTTGCTCTTGCACTTGCGGGAGGGGTAGCTA CCAGCATTCTCGGGAGGTAAGCAGGTCCACTTCG AAATTTGCTCTTCAGACTGATGGACTCAACT*G*T	Sense
		Rescue	TACCAGCATTCTCGGGAGGT	A*C*CTTGCTCTTGCACTTGCGGGAGGGGTAGCTA CCAGCATTCTCGGGAGGCAAGCAGGTCCACTTCG AAATTTGCTCTTCAGACTGATGGACTCAACT*G*T	Sense
			TAGCTACCAGCATTCTCGGG	A*C*AGTTGAGTCCATCAGTCTGAAGAGCAAATTTC GAAGTGGACCTGCCTGTCTCCCGAGAATGCTGGT AGCTACCCCTCCCGCAAGTGCAAGAGCAAG*G*T	Antisense
HERC1	chr15:63775279:A:C	Knock-in	AAAGTGCATAAAAGACACGC	G*C*TGCTGCTGCTTCACCTTGCCTTTGTCATGGTA TTTATTAGAAAGTGCCTAAAAGACACGCTGAAGTA CAAGCAGTCGTTTTCTAAGTGCCCCGGCA*A*A	Antisense
		Rescue	AAAGTGCCTAAAAGACACGC	G*C*TGCTGCTGCTTCACCTTGCCTTTGTCATGGTA TTTATTAGAAAGTGCATAAAAGACACGCTGAAGTA CAAGCAGTCGTTTTCTAAGTGCCCCGGCA*A*A	Antisense
HERC1	chr15:63655874:G:A	Knock-in	TTTCTACCCAGCTTTCTCCG	T*A*ACCACAGTTGATGACTCAATTCAGCGAAAGTT TCTACCCAGCTTTCTCTAAGGAATTGCTGAAGAGA ACAAGCTTGTGACCTCCCCAAACTTTGTT*G*T	Sense
			GCAATTCCTCGGAGAAAGCT	A*C*AACAAAGTTTGGGGAGGTCACAAGCTTGTTCT CTTCAGCAATTCCTCAGAGAAAAGAGGGTAGAAAC TTTCGCTGAATTGAGTCATCAACTGTGGT*T*A	Antisense
		Rescue	TTTCTACCCAGCTTTCTCTG	T*A*ACCACAGTTGATGACTCAATTCAGCGAAAGTT TCTACCCAGCTTCTTGCGAGGAATTGCTGAAGAGA ACAAGCTTGTGACCTCCCCAAACTTTGTT*G*T	Sense
			GCAATTCCTCAGAGAAAGCT	A*C*AACAAAGTTTGGGGAGGTCACAAGCTTGTTCT CTTCAGCAATTCCTCGGAGAAAAGAGGGTAGAAAC TTTCGCTGAATTGAGTCATCAACTGTGGT*T*A	Antisense
HERC1	chr15:63713615:G:A	Knock-in	GTAGCTCGTAGCCGAGACCG	A*A*ATCTTTCAGTGTTTCCTCTCAGCCCGTGAAGT AGCTCGTAGCCGAGATTAAGATAGAATGAACAGTG GGGCAGGGTCTGGGGCTCGAGCTGATGAT*C*C	Sense

Gene	Pathogenic variant	Edit type	sgRNA sequence (5’ -> 3’)	ssDNA repair template sequence (5’ -> 3’)	Orientation
			CGTGAAGTAGCTCGTAGCCG	G*G*ATCATCAGCTCGAGCCCCAGACCCTGCCCCA CTGTTCATTCTATCTCAGTCACGGGAACGAGCTAC TTCACGGGCTGAGAGGAAACACTGAAAGAT*T*T	Antisense
		Rescue	GTAGCTCGTAGCCGAGACTG	A*A*ATCTTTCAGTGTTTCCTCTCAGCCCGTGAAGT AGCTCGTTCCCGTGACCGAGATAGAATGAACAGTG GGGCAGGGTCTGGGGCTCGAGCTGATGAT*C*C	Sense
			CGTGAAGTAGCTCGTAGCCG	G*G*ATCATCAGCTCGAGCCCCAGACCCTGCCCCA CTGTTCATTCTATCTCGGTCACGAGAACGAGCTAC TTCACGGGCTGAGAGGAAACACTGAAAGAT*T*T	Antisense
			CTATCTCAGTCTCGGCTACG	G*G*ATCATCAGCTCGAGCCCCAGACCCTGCCCCA CTGTTCATTCTATCTCGGTCTCGGGATCGTGCTAC TTCACGGGCTGAGAGGAAACACTGAAAGAT*T*T	Antisense
HERC1	chr15:63656291:G:A	Knock-in	ACATTAATCGAACTAGCGTT	G*G*GCTGGTGGAGAGGCCAGCTCTCCCTGCTGCT GCCAAGCACATTAATCAAACTAGCGTTCGAATATC TGTTAGCCCCAGAGACTCAAGACCAGCAGC*C*A	Antisense
		Rescue	ACATTAATCAAACTAGCGTT	G*G*GCTGGTGGAGAGGCCAGCTCTCCCTGCTGCT GCCAAGCACATTAATCGAACTAGCGTTCGAATATC TGTTAGCCCCAGAGACTCAAGACCAGCAGC*C*A	Antisense
HERC1	chr15:63693963:C:T	Knock-in	ACTGAGAAGAAAGATTTCAG	A*G*AAAAAAAAGTTGACTCCAGTGGAGAAACTGAG AAGAAAGATTTCAGAGATAATGTAAGCCTCTTTCTC TGTCTTTACTTACAAGCATTTATTAAAA*C*C	Sense
		Rescue	ACTGAGAAGAAAGATTTCAG	A*G*AAAAAAAAGTTGACTCCAGTGGAGAAACTGAG AAGAAAGATTTTCGGGGTAATGTAAGCCTCTTTCT CTGTCTTTACTTACAAGCATTTATTAAAA*C*C	Sense
			AAGATTTCAGAGATAATGTA	A*G*AAAAAAAAGTTGACTCCAGTGGAGAAACTGAG AAGAAAGATTTCAGAGGTAATGTAAACCTCTTTCTC TGTCTTTACTTACAAGCATTTATTAAAA*C*C	Sense
HERC1	chr15:63637595:CTG:C	Knock-in	CCAGGCCAATGGCACAGTGT	C*T*AGGTCATTTGTGGTCAGAATTGTACCTTTGTC ATCCAGGCCAATGGCAGTGTTGGCTTGTGGGGAA GGAAGTTATGGCAGATTAGGACAAGGAAAT*T*C	Sense
			ACCTTTGTCATCCAGGCCAA	G*A*ATTTCCTTGTCCTAATCTGCCATAACTTCCTTC CCCACAAGCCAACACTGCCATTAGCCTGAATAACA AAGGTACAATTCTGACCACAAATGACCT*A*G	Antisense
			TGGCACAGTGTTGGCTTGTG	C*T*AGGTCATTTGTGGTCAGAATTGTACCTTTGTC ATCCAGGCCAATGGCAGTGTTGGCCTTTGGGGAA GGAAGTTATGGCAGATTAGGACAAGGAAAT*T*C	Sense

Gene	Pathogenic variant	Edit type	sgRNA sequence (5’ -> 3’)	ssDNA repair template sequence (5’ -> 3’)	Orientation
		Rescue	ATCCAGGCCAATGGCAGTGT	C*T*AGGTCATTTGTGGTCAGAATTGTACCTTTGTC ATCCAGGCCAATGGCACAGTGTTGGCTTGTGGGG AAGGAAGTTATGGCAGATTAGGACAAGGAAAT*T*C	Sense
			AATGGCAGTGTTGGCTTGTG	T*A*GGTCATTTGTGGTCAGAATTGTACCTTTGTCAT CCAGGCCAATGGCACAGTGTTGGCTTGTGGTGAA GGAAGTTATGGCAGATTAGGACAAGGAAA*T*T	Sense
			ACCTTTGTCATCCAGGCCAA	G*A*ATTTCCTTGTCCTAATCTGCCATAACTTCCTTC CCCACAAGCCAACACTGTGCCATTAGCCTGAATAA CAAAGGTACAATTCTGACCACAAATGACCT*A*G	Antisense
RB1CC1	chr8:52657117:C:CTGAG	Knock-in	AGAGTTAGTATGCCTTGAGG	T*G*AAAACAAAATTAAAAAATTGAAGGGAGAGTTA GTATGCCTTGAGGACTCAGGTTTTACAAAATAAAG ATAATGAATTTGCTTTGGTTAAACATGAA*A*A	Sense
			GGGAGAGTTAGTATGCCTTG	T*T*TTCATGTTTAACCAAAGCAAATTCATTATCTTTA TTTTGTAAAACCTGAGTCTTCCAAGCATACTAACTC TCCCTTCAATTTTTTAATTTTGTTTT*C*A	Antisense
		Rescue	TTAGTATGCCTTGAGGACTC	T*G*AAAACAAAATTAAAAAATTGAAGGGAGAGTTA GTATGCCTTGAGGAGGTTTTACAAAATAAAGATAAT GAATTTGCTTTGGTTAAACATGAA*A*A	Sense
			GGGAGAGTTAGTATGCCTTG	T*T*TTCATGTTTAACCAAAGCAAATTCATTATCTTTA TTTTGTAAAACCTCTTCCAAGCATACTAACTCTCCC TTCAATTTTTTAATTTTGTTTT*C*A	Antisense
SETD1A	chr16:30966112:C:CA	Knock-in	CAGAGGGGCATACTCACGGG	T*T*GCCGTATGCTCTATATGCACAGGGGCAGGAG GGCAGAGGGGCATACTCAACGGGAGGCCTACCAC CTGCCCATGCCAATGGCAGCCGAGCCCCTGC*C*C	Sense
			GGGCAGAGGGGCATACTCAC	T*T*GCCGTATGCTCTATATGCACAGGGGCAGGAG GGCAGAGGGGCATACTCAACGGGAGGCCTACCAC CTGCCCATGCCAATGGCAGCCGAGCCCCTGC*C*C	Sense
			AGGGCAGAGGGGCATACTCA	T*T*GCCGTATGCTCTATATGCACAGGGGCAGGAG GGCAGAGGGGCATACTCAACTGGAGGCCTACCAC CTGCCCATGCCAATGGCAGCCGAGCCCCTGC*C*C	Sense
		Rescue	AGAGGGGCATACTCAACGGG	T*T*GCCGTATGCTCTATATGCACAGGGGCAGGAG GGCAGAGGGGCATACTCACGGGAGGCCTACCACC TGCCCATGCCAATGGCAGCCGAGCCCCTGC*C*C	Sense
			GGGCAGAGGGGCATACTCAA	T*T*GCCGTATGCTCTATATGCACAGGGGCAGGAG GGCAGAGGGGCATACTCACTGGAGGCCTACCACC TGCCCATGCCAATGGCAGCCGAGCCCCTGC*C*C	Sense

Gene	Pathogenic variant	Edit type	sgRNA sequence (5’ -> 3’)	ssDNA repair template sequence (5’ -> 3’)	Orientation
SETD1A	chr16:30961466:AG:A	Knock-in	CACCAGCACTCGGGGCGCCA	A*G*CACCTGGGCCTGGCCCGTGTGCTCTTCACCA GCACTCGGGGCGCCAAGAAACGGTCAAAAACCTC CACCTTACCTCCGTCATGGGCAACATCATCC*A*T	Sense
			CACTCGGGGCGCCAAGGAAA	A*G*CACCTGGGCCTGGCCCGTGTGCTCTTCACCA GCACTCGGGGCGCCAAGAAAAGATCAAAAACCTC CACCTTACCTCCGTCATGGGCAACATCATCC*A*T	Sense
		Rescue	GCACTCGGGGCGCCAAGAAA	A*A*GCACCTGGGCCTGGCCCGTGTGCTCTTCACC AGCACTCGGGGCGCCAAGGAAACTGTCAAAAACC TCCACCTTACCTCCGTCATGGGCAACATCAT*C*C	Sense
			GGAGGTTTTTGACCGTTTCT	A*A*GCACCTGGGCCTGGCCCGTGTGCTCTTCACC AGCACTCGGGGCGCAAAGGAAACGGTCAAAAACC TCCACCTTACCTCCGTCATGGGCAACATCAT*C*C	Sense
SETD1A	chr16:30963433:G:T	Knock-in	TCATTCGTTGTTGTCCTGGA	G*G*TGTAGGAGCCATTGACAATTAGTTCATAGTAT TTCATTCGTTGTTGTACTGGAGGTAGGGAAATGGG AACCAATTGTCAGATGGAGATACTAACTC*C*T	Sense
			TCGTTGTTGTCCTGGAGGGA	A*G*GAGTTAGTATCTCCATCTGACAATTGGTTCCC ATTTACCTCCCTCCAGTACAACAACGAATGAAATAC TATGAACTAATTGTCAATGGCTCCTACA*C*C	Antisense
		Rescue	TCATTCGTTGTTGTACTGGA	G*G*TGTAGGAGCCATTGACAATTAGTTCATAGTAT TTCATTCGTTGTTGTCCTGGAGGTAGGGAAATGGG AACCAATTGTCAGATGGAGATACTAACTC*C*T	Sense
			TCGTTGTTGTACTGGAGGGA	A*G*GAGTTAGTATCTCCATCTGACAATTGGTTCCC ATTTACCTCCCTCCAGGACAACAACGAATGAAATA CTATGAACTAATTGTCAATGGCTCCTACA*C*C	Antisense
SETD1A	chr16:30966303:C:T	Knock-in	ATGAAGAGCATCATGCAGCG	T*G*GGCCGTGTGCTCGCCATGCTGGTCCAGGAGA TGAAGAGCATCATGCAGTAAGACCTCAACCGCAAG ATGGTGGAGAACGTGGCCTTCGGAGCCTTT*G*A	Sense
			GTTGAGGTCTCGCTGCATGA	T*G*GGCCGTGTGCTCGCCATGCTGGTCCAGGAGA TGAAGTCTATCATGCAGTGAGACCTCAACCGCAAG ATGGTGGAGAACGTGGCCTTCGGAGCCTTT*G*A	Sense
		Rescue	TCATGCAGTGAGACCTCAAC	T*G*GGCCGTGTGCTCGCCATGCTGGTCCAGGAGA TGAAGAGCATCATGCAGCGTGACCTTAATCGCAAG ATGGTGGAGAACGTGGCCTTCGGAGCCTTT*G*A	Sense
			TCTCCACCATCTTGCGGTTG	T*C*AAAGGCTCCGAAGGCCACGTTCTCCACCATCT TGCGGTTGAGATCTCGCTGCATGATGCTCTTCATC TCCTGGACCAGCATGGCGAGCACACGGCC*C*A	Sense

Gene	Pathogenic variant	Edit type	sgRNA sequence (5’ -> 3’)	ssDNA repair template sequence (5’ -> 3’)	Orientation
SP4	chr7:21481973:G:T	Knock-in	ATCTGTCATATTGAAGGATG	G*A*GGCAGTAATGAACCAGGAAAAAAGAAGCAGC ATATCTGTCATATTGAGTAGTGTGGTAAAGTTTATG GCAAAACATCTCATTTACGAGCACATCTT*C*G	Sense
			TCTGTCATATTGAAGGATGT	G*A*GGCAGTAATGAACCAGGAAAAAAGAAGCAGC ATATCTGTCATATTGAATAGTGCGGTAAAGTTTATG GCAAAACATCTCATTTACGAGCACATCTT*C*G	Sense
		Rescue	ATCTGTCATATTGAATGATG	G*A*GGCAGTAATGAACCAGGAAAAAAGAAGCAGC ATATCTGTCATATTGAGGGGTGTGGTAAAGTTTAT GGCAAAACATCTCATTTACGAGCACATCTT*C*G	Sense
			TCTGTCATATTGAATGATGT	G*A*GGCAGTAATGAACCAGGAAAAAAGAAGCAGC ATATCTGTCATATTGAAGGGTGCGGTAAAGTTTAT GGCAAAACATCTCATTTACGAGCACATCTT*C*G	Sense
SP4	chr7:21429774:TAATC:T	Knock-in	AGTATAGCTTGATTATTACC	T*C*TCTACAGGATTTGCAGGGTCAAATTCAGCTCA TTTCTGCAGGTAATAAGCTATACTCACAGCTGCTAA CAGGACAGCTTCTGGGAATATTCTTGCT*C*A	Sense
			ATAGCTTGATTATTACCTGC	T*C*TCTACAGGATTTGCAGGGTCAAATTCAGCTCA TTTCTGCAGGTAATAAGCTATACTCACAGCTGCTAA CAGGACAGCTTCTGGGAATATTCTTGCT*C*A	Sense
		Rescue	TGTGAGTATAGCTTATTACC	T*C*TCTACAGGATTTGCAGGGTCAAATTCAGCTCA TTTCTGCAGGTAATAATCAAGCTATACTCACAGCTG CTAACAGGACAGCTTCTGGGAATATTCT*T*G	Sense
			GAGTATAGCTTATTACCTGC	T*C*TCTACAGGATTTGCAGGGTCAAATTCAGCTCA TTTCTGCAGGTAATAATCAAGCTATACTCACAGCTG CTAACAGGACAGCTTCTGGGAATATTCT*T*G	Sense
TRIO	chr5:14472658:G:GGTGTGTA	Knock-in	GGACCTTGGCTATGTGGTTG	T*G*GAGACAGAGCGTGACTATGTGCGGGACCTTG GCTATGTGGTTGAAGTGTGTAGTGTGTATTGCCAG AAATTTAGTATCTTCGTATCAGTTCCAAGA*G*T	Sense
			TGTGCGGGACCTTGGCTATG	A*C*TCTTGGAACTGATACGAAGATACTAAATTTCTG GCAATACACACTACACACCTCAACGACATAGCCAA GGTCCCGCACATAGTCACGCTCTGTCTC*C*A	Antisense
		Rescue	GGACCTTGGCTATGTGGTTG	T*A*GTGGAGACAGAGCGTGACTATGTGCGGGACC TTGGCTATGTCGTTGAAGTGTGTATTGCCAGAAAT TTAGTATCTTCGTATCAGTTCCAAGAGTTG*T*C	Sense
			TGTGCGGGACCTTGGCTATG	G*A*CAACTCTTGGAACTGATACGAAGATACTAAAT TTCTGGCAATACACACCTCAACGACATAGCCAAGG TCCCGCACATAGTCACGCTCTGTCTCCAC*T*A	Antisense

Gene	Pathogenic variant	Edit type	sgRNA sequence (5’ -> 3’)	ssDNA repair template sequence (5’ -> 3’)	Orientation
XPO7	chr8:21976513:G:A	Knock-in	TTACCTGATCTCCAGCTGGT	C*A*CTGATGAGTCCTCAGACGACCTGTGTACAGTG CAGATTCCCACATCCTAGAGATCAGGTAACAGAAC TTCCTCCACCTCAAAGGCTGTCTGTCACT*C*C	Sense
			TCTGTTACCTGATCTCCAGC	C*A*CTGATGAGTCCTCAGACGACCTGTGTACAGTG CAGATTCCCACAAGCTAGAGATCAGGTAACAGAAC TTCCTCCACCTCAAAGGCTGTCTGTCACT*C*C	Sense
			GTTACCTGATCTCCAGCTGG	C*A*CTGATGAGTCCTCAGACGACCTGTGTACAGTG CAGATTCCAACCAGCTAGAGATCAGGTAACAGAAC TTCCTCCACCTCAAAGGCTGTCTGTCACT*C*C	Sense
		Rescue	TTACCTGATCTCTAGCTGGT	C*A*CTGATGAGTCCTCAGACGACCTGTGTACAGTG CAGATTCCCACATCCTGGAGATCAGGTAACAGAAC TTCCTCCACCTCAAAGGCTGTCTGTCACT*C*C	Sense
			GTTACCTGATCTCTAGCTGG	C*A*CTGATGAGTCCTCAGACGACCTGTGTACAGTG CAGATTCCAACCAGCTGGAGATCAGGTAACAGAAC TTCCTCCACCTCAAAGGCTGTCTGTCACT*C*C	Sense

Appendix 2: i5 and i7 primer sequences

Supplementary Table 3. Sequences of the i5 primers
i5 primers D501 – D504 were used for the electroporated cells with the guide RNAs HEL252_sg01_rv, HEL252_sg02_fw, HEL246_sg03_rv, and HEL246_sg04_fw respectively. Each unique barcode is highlighted in bold letters.

i5 Primer_ID	Sequence (5' -> 3')
i5_D501	AATGATACGGCGACCACCGAGATCTACAC**TATAGCCT**ACACTCTTTCCCTACAC GACGCTCTTCCGAT*C*T
i5_D502	AATGATACGGCGACCACCGAGATCTACAC**ATAGAGGC**ACACTCTTTCCCTACAC GACGCTCTTCCGAT*C*T
i5_D503	AATGATACGGCGACCACCGAGATCTACAC**CCTATCCT**ACACTCTTTCCCTACAC GACGCTCTTCCGAT*C*T
i5_D504	AATGATACGGCGACCACCGAGATCTACAC**GGCTCTGA**ACACTCTTTCCCTACAC GACGCTCTTCCGAT*C*T

Supplementary Table 4. Sequences of the i7 primers.
Each well of a 96-well plate received a unique i7 barcode, highlighted in bold letters within the sequence.

Row	Column	i7 Primer_ID	Sequence (5' - 3')
A	1	iPCRtag1	CAAGCAGAAGACGGCATACGAGAT**AACGTGAT**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
A	2	iPCRtag2	CAAGCAGAAGACGGCATACGAGAT**AAACATCG**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
A	3	iPCRtag3	CAAGCAGAAGACGGCATACGAGAT**ATGCCTAA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
A	4	iPCRtag4	CAAGCAGAAGACGGCATACGAGAT**AGTGGTCA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
A	5	iPCRtag5	CAAGCAGAAGACGGCATACGAGAT**ACCACTGT**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
A	6	iPCRtag6	CAAGCAGAAGACGGCATACGAGAT**ACATTGGC**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
A	7	iPCRtag7	CAAGCAGAAGACGGCATACGAGAT**CAGATCTG**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
A	8	iPCRtag8	CAAGCAGAAGACGGCATACGAGAT**CATCAAGT**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
A	9	iPCRtag9	CAAGCAGAAGACGGCATACGAGAT**CGCTGATC**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
A	10	iPCRtag10	CAAGCAGAAGACGGCATACGAGAT**ACAAGCTA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
A	11	iPCRtag11	CAAGCAGAAGACGGCATACGAGAT**CTGTAGCC**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
A	12	iPCRtag12	CAAGCAGAAGACGGCATACGAGAT**AGTACAAG**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
B	1	iPCRtag13	CAAGCAGAAGACGGCATACGAGAT**AACAACCA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC

Row	Column	i7 Primer_ID	Sequence (5' - 3')
B	2	iPCRtag14	CAAGCAGAAGACGGCATACGAGAT**AACCGAGA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
B	3	iPCRtag15	CAAGCAGAAGACGGCATACGAGAT**AACGCTTA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
B	4	iPCRtag16	CAAGCAGAAGACGGCATACGAGAT**AAGACGGA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
B	5	iPCRtag17	CAAGCAGAAGACGGCATACGAGAT**AAGGTACA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
B	6	iPCRtag18	CAAGCAGAAGACGGCATACGAGAT**ACACAGAA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
B	7	iPCRtag19	CAAGCAGAAGACGGCATACGAGAT**ACAGCAGA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
B	8	iPCRtag20	CAAGCAGAAGACGGCATACGAGAT**ACCTCCAA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
B	9	iPCRtag21	CAAGCAGAAGACGGCATACGAGAT**ACGCTCGA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
B	10	iPCRtag22	CAAGCAGAAGACGGCATACGAGAT**ACGTATCA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
B	11	iPCRtag23	CAAGCAGAAGACGGCATACGAGAT**ACTATGCA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
B	12	iPCRtag24	CAAGCAGAAGACGGCATACGAGAT**AGAGTCAA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
C	1	iPCRtag25	CAAGCAGAAGACGGCATACGAGAT**AGATCGCA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
C	2	iPCRtag26	CAAGCAGAAGACGGCATACGAGAT**AGCAGGAA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
C	3	iPCRtag27	CAAGCAGAAGACGGCATACGAGAT**AGTCACTA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
C	4	iPCRtag28	CAAGCAGAAGACGGCATACGAGAT**ATCCTGTA**GAGATCGGTC TCGGCATTCCTGCTGAACCGCTCTTCCGATC
C	5	iPCRtag29	CAAGCAGAAGACGGCATACGAGAT**ATTGAGGA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
C	6	iPCRtag30	CAAGCAGAAGACGGCATACGAGAT**CAACCACA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
C	7	iPCRtag31	CAAGCAGAAGACGGCATACGAGAT**CAAGACTA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
C	8	iPCRtag32	CAAGCAGAAGACGGCATACGAGAT**CAATGGAA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
C	9	iPCRtag33	CAAGCAGAAGACGGCATACGAGAT**CACTTCGA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
C	10	iPCRtag34	CAAGCAGAAGACGGCATACGAGAT**CAGCGTTA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
C	11	iPCRtag35	CAAGCAGAAGACGGCATACGAGAT**CATACCAA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
C	12	iPCRtag36	CAAGCAGAAGACGGCATACGAGAT**CCAGTTCA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC

Row	Column	i7 Primer_ID	Sequence (5' - 3')
D	1	iPCRtag37	CAAGCAGAAGACGGCATACGAGAT**CCGAAGTA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
D	2	iPCRtag38	CAAGCAGAAGACGGCATACGAGAT**CCGTGAGA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
D	3	iPCRtag39	CAAGCAGAAGACGGCATACGAGAT**CCTCCTGA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
D	4	iPCRtag40	CAAGCAGAAGACGGCATACGAGAT**CGAACTTA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
D	5	iPCRtag41	CAAGCAGAAGACGGCATACGAGAT**CGACTGGA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
D	6	iPCRtag42	CAAGCAGAAGACGGCATACGAGAT**CGCATACA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
D	7	iPCRtag43	CAAGCAGAAGACGGCATACGAGAT**CTCAATGA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
D	8	iPCRtag44	CAAGCAGAAGACGGCATACGAGAT**CTGAGCCA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
D	9	iPCRtag45	CAAGCAGAAGACGGCATACGAGAT**CTGGCATA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
D	10	iPCRtag46	CAAGCAGAAGACGGCATACGAGAT**GAATCTGA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
D	11	iPCRtag47	CAAGCAGAAGACGGCATACGAGAT**GACTAGTA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
D	12	iPCRtag48	CAAGCAGAAGACGGCATACGAGAT**GAGCTGAA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
E	1	iPCRtag49	CAAGCAGAAGACGGCATACGAGAT**GATAGACA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
E	2	iPCRtag50	CAAGCAGAAGACGGCATACGAGAT**GCCACATA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
E	3	iPCRtag51	CAAGCAGAAGACGGCATACGAGAT**GCGAGTAA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
E	4	iPCRtag52	CAAGCAGAAGACGGCATACGAGAT**GCTAACGA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
E	5	iPCRtag53	CAAGCAGAAGACGGCATACGAGAT**GCTCGGTA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
E	6	iPCRtag54	CAAGCAGAAGACGGCATACGAGAT**GGAGAACA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
E	7	iPCRtag55	CAAGCAGAAGACGGCATACGAGAT**GGTGCGAA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
E	8	iPCRtag56	CAAGCAGAAGACGGCATACGAGAT**GTACGCAA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
E	9	iPCRtag57	CAAGCAGAAGACGGCATACGAGAT**GTCGTAGA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
E	10	iPCRtag58	CAAGCAGAAGACGGCATACGAGAT**GTCTGTCA**GAGATCGGTC TCGGCATTCCTGCTGAACCGCTCTTCCGATC
E	11	iPCRtag59	CAAGCAGAAGACGGCATACGAGAT**GTGTTCTA**GAGATCGGTC TCGGCATTCCTGCTGAACCGCTCTTCCGATC

Row	Column	i7 Primer_ID	Sequence (5' - 3')
E	12	iPCRtag60	CAAGCAGAAGACGGCATACGAGAT**TAGGATGA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
F	1	iPCRtag61	CAAGCAGAAGACGGCATACGAGAT**TATCAGCA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
F	2	iPCRtag62	CAAGCAGAAGACGGCATACGAGAT**TCCGTCTA**GAGATCGGTC TCGGCATTCCTGCTGAACCGCTCTTCCGATC
F	3	iPCRtag63	CAAGCAGAAGACGGCATACGAGAT**TCTTCACA**GAGATCGGTC TCGGCATTCCTGCTGAACCGCTCTTCCGATC
F	4	iPCRtag64	CAAGCAGAAGACGGCATACGAGAT**TGAAGAGA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
F	5	iPCRtag65	CAAGCAGAAGACGGCATACGAGAT**TGGAACAA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
F	6	iPCRtag66	CAAGCAGAAGACGGCATACGAGAT**TGGCTTCA**GAGATCGGTC TCGGCATTCCTGCTGAACCGCTCTTCCGATC
F	7	iPCRtag67	CAAGCAGAAGACGGCATACGAGAT**TGGTGGTA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
F	8	iPCRtag68	CAAGCAGAAGACGGCATACGAGAT**TTCACGCA**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
F	9	iPCRtag69	CAAGCAGAAGACGGCATACGAGAT**AACTCACC**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
F	10	iPCRtag70	CAAGCAGAAGACGGCATACGAGAT**AAGAGATC**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
F	11	iPCRtag71	CAAGCAGAAGACGGCATACGAGAT**AAGGACAC**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
F	12	iPCRtag72	CAAGCAGAAGACGGCATACGAGAT**AATCCGTC**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
G	1	iPCRtag73	CAAGCAGAAGACGGCATACGAGAT**AATGTTGC**GAGATCGGTC TCGGCATTCCTGCTGAACCGCTCTTCCGATC
G	2	iPCRtag74	CAAGCAGAAGACGGCATACGAGAT**ACACGACC**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
G	3	iPCRtag75	CAAGCAGAAGACGGCATACGAGAT**ACAGATTC**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
G	4	iPCRtag76	CAAGCAGAAGACGGCATACGAGAT**AGATGTAC**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
G	5	iPCRtag77	CAAGCAGAAGACGGCATACGAGAT**AGCACCTC**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
G	6	iPCRtag78	CAAGCAGAAGACGGCATACGAGAT**AGCCATGC**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
G	7	iPCRtag79	CAAGCAGAAGACGGCATACGAGAT**AGGCTAAC**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
G	8	iPCRtag80	CAAGCAGAAGACGGCATACGAGAT**ATAGCGAC**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
G	9	iPCRtag81	CAAGCAGAAGACGGCATACGAGAT**ATCATTCC**GAGATCGG TCTCGGCATTCCTGCTGAACCGCTCTTCCGATC
G	10	iPCRtag82	CAAGCAGAAGACGGCATACGAGAT**ATTGGCTC**GAGATCGGTC TCGGCATTCCTGCTGAACCGCTCTTCCGATC

Row	Column	i7 Primer_ID	Sequence (5' - 3')
G	11	iPCRtag83	CAAGCAGAAGACGGCATACGAGAT**CAAGGAGC**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
G	12	iPCRtag84	CAAGCAGAAGACGGCATACGAGAT**CACCTTAC**GAGATCGGTC TCGGCATTCCTGCTGAACCGCTCTTCCGATC
H	1	iPCRtag85	CAAGCAGAAGACGGCATACGAGAT**CCATCCTC**GAGATCGGTC TCGGCATTCCTGCTGAACCGCTCTTCCGATC
H	2	iPCRtag86	CAAGCAGAAGACGGCATACGAGAT**CCGACAAC**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
H	3	iPCRtag87	CAAGCAGAAGACGGCATACGAGAT**CCTAATCC**GAGATCGGTC TCGGCATTCCTGCTGAACCGCTCTTCCGATC
H	4	iPCRtag88	CAAGCAGAAGACGGCATACGAGAT**CCTCTATC**GAGATCGGTC TCGGCATTCCTGCTGAACCGCTCTTCCGATC
H	5	iPCRtag89	CAAGCAGAAGACGGCATACGAGAT**CGACACAC**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
H	6	iPCRtag90	CAAGCAGAAGACGGCATACGAGAT**CGGATTGC**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
H	7	iPCRtag91	CAAGCAGAAGACGGCATACGAGAT**CTAAGGTC**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
H	8	iPCRtag92	CAAGCAGAAGACGGCATACGAGAT**GAACAGGC**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
H	9	iPCRtag93	CAAGCAGAAGACGGCATACGAGAT**GACAGTGC**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
H	10	iPCRtag94	CAAGCAGAAGACGGCATACGAGAT**GAGTTAGC**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
H	11	iPCRtag95	CAAGCAGAAGACGGCATACGAGAT**GATGAATC**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC
H	12	iPCRtag96	CAAGCAGAAGACGGCATACGAGAT**GCCAAGAC**GAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGATC