HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Creating CRISPR-Cas9 genome edited iPSC lines to model a patient-specific mutation in mitochondrial disease

Nelli Jalkanen

Master's Thesis

Master's Degree Programme in Translational Medicine

Faculty of Medicine

University of Helsinki

2020

Abstract

Tiedekunta – Fakultet – Faculty	Koulutusohjelma – Utbildingsprogram – Degree Programme
Faculty of Medicine	Translational Medicine MSc.

Tekijä – Författare – Author

Nelli Jalkanen

Työn nimi – Arbetets titel – Title

Creating CRISPR-Cas9 genome edited iPSC lines to model a patient-specific mutation in mitochondrial disease

Oppiaine/Opintosuunta – Läroämne/Studieinriktning – Subject/Study track

Translational Medicine, specialisation in Neuroscience and Psychobiology

Työn laji – Arbetets art – Level	Aika – Datum – Month and year	Sivumäärä – Sidoantal – Number of pages
Master's thesis	October 2020	74

Tiivistelmä – Referat – Abstract

Mitochondrial aminoacyl tRNA-synthetases (mt-aaRS) catalyse the charging of tRNAs with their cognate amino acids in mitochondria. Mutations in mt-aaRS cause tissue-specific mitochondrial diseases, especially affecting tissues with high energy expenditure like the nervous system, heart, and kidneys. However, disease mechanisms for the heterogeneous group of diseases have not yet been fully elucidated. Harnessing CRISPR-Cas9 genome editing in induced pluripotent stem cells (iPSC) provides an opportunity to model mt-aaRS mutations *in vitro* and investigate the effects of individual mutations on cellular phenotype.

SARS2 encodes mitochondrial seryl tRNA-synthetase, and its c.1347 G>A mutation causes severe childhood-onset progressive spastic paresis. Here, CRISPR-Cas9 ribonucleoprotein (RNP) complex and associated donor template were used to induce homology directed repair (HDR) the genome of iPSC and knock-in the patient mutation. Guide RNAs were designed and tested for efficiency before electroporation into wild type iPSC. Clonal cell lines were made by low-density seeding and manual colony picking. The expression of pluripotency markers was measured by RT-qPCR. RT-qPCR and Western blot measured SARS2 mRNA expression and protein level respectively. The success and precision of genome editing were analysed by Sanger sequencing, comparing the performance of the different guide RNAs, and screening regions of potential off-target genome editing.

Two genome-edited iPSC lines with the SARS2 c.1347 G>A mutation were successfully generated to model the patient mutation. The iPSC lines expressed pluripotency markers and contained no off-target genome editing and modelled the patient's decrease in SARS2 protein level and mRNA expression. More evidence of differentiation ability is needed before differentiation into the affected cell type (motor neurons) and further disease modelling. The efficiency of CRISPR-Cas9 for genome editing, especially harnessing HDR in iPSC, is an area of future research.

Avainsanat – Nyckelord – Keywords

CRISPR-Cas9, genome editing, iPSC, disease modelling, mitochondrial aminoacyl tRNA-synthetase, seryl tRNA-synthetase, mitochondrial disease

Ohjaaja tai ohjaajat – Handledare – Supervisor or supervisors

Henna Tyynismaa PhD, Associate Professor

Tiina Rasila PhD

Säilytyspaikka – Förvaringställe – Where deposited

Helsingin yliopiston kirjasto, Helsingfors universitets bibliotek, Helsinki University Library

Muita tietoja – Övriga uppgifter – Additional information

Table of Contents

Abbreviations

aaRS	Aminoacyl tRNA-synthetase
ANOVA	Analysis of variance
AMP	Adenosine monophosphate
ATP	Adenosine triphosphate
bp	Base pair
BSCC	Biomedicum Stem Cell Center
c-Myc	MYC proto-oncogene
CNS	Central nervous system
COX I	Cytochrome C oxidase (complex IV) subunit 1
COX II	Cytochrome C oxidase (complex IV) subunit 2
CRISPRa	CRISPR-Cas transcriptional activation
CRISPR-Cas	Clustered Regularly Interspaced Short Palindromic Repeats – CRISPR associated protein
CRISPR-Cas9	Clustered Regularly Interspaced Short Palindromic Repeats – CRISPR associated protein 9
crRNA	CRISPR-RNA
DSB	Double strand break
dsDNA	Double-stranded DNA
E. coli	Escherichia coli
EDTA	Ethylenediaminetetraacetic acid
ETC	Electron transport chain
FACS	Fluorescence activated cell sorting
$FADH_2$	Flavin adenine dinucleotide
FBS	Foetal bovine serum
FIMM	Institute for Molecular Medicine Finland
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GFP	Green fluorescent protein
gRNA	Guide RNA
HDR	Homology directed repair
IFN-γ	Interferon gamma
INDEL	Insertion/deletion
iPSC	Induced pluripotent stem cell

Klf4	Krüppel-like factor 4
Lin28a	Lin-28 homolog A
mt-aaRS	Mitochondrial aminoacyl tRNA-synthetase
mt-DNA	Mitochondrial DNA
mTOR	Mammalian target of rapamycin
MTS	Mitochondrial targeting sequence
NADH	Nicotinamide adenine dinucleotide
Nanog	Homeobox protein Nanog
NHEJ	Non-homologous end joining
nt	Nucleotide
Oct4	Octamer-binding transcription factor 4
OXPHOS	Oxidative phosphorylation
PAM	Protospacer adjacent motif
PBS	Phosphate buffered saline
PP_i	Pyrophosphate
P/S	Penicillin/streptomycin
PVDF	Polyvinylidene difluoride
p53	Cellular tumour antigen p53
RC	Respiratory chain
RNAi	RNA interference
RNP	Ribonucleoprotein
ROCK	p160 Rho-associated coiled-coil kinase
ROS	Reactive oxygen species
RT	Room temperature
RT-qPCR	Real-time quantitative polymerase chain reaction
rRNA	Ribosomal RNA
SARS2	Mitochondrial seryl tRNA-synthetase gene
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
sgRNA	Single guide RNA
Sox2	Sex determining region Y-box 2
S. pyogenes	Streptococcus pyogenes
spCas9	Streptococcus pyogenes CRISPR associated protein 9

SSEA3	Stage specific embryonic antigen 3
SSEA4	Stage specific embryonic antigen 4
ssODN	Single-stranded donor oligonucleotide
TALEN	Transcription activator-like effector nucleases
TBE	Tris/borate/EDTA
tracrRNA	Trans-activating CRISPR RNA
tRNA	Transfer RNA
TRA1-60	TRA-1-60 antibody
TRA1-81	TRA-1-81 antibody
UPR	Unfolded protein response
UPR-mt	Unfolded protein response-mitochondrial
ZFN	Zinc finger nuclease

1. Introduction

1.1 Mitochondria

Mitochondria (singular mitochondrion) are membrane-bound organelles found in the cytoplasm of most eukaryotic cells. They can be tubular, round, or irregular shaped due to developmental perturbations. Mitochondria contain a double membrane of a permeable outer membrane and a highly folded inner membrane with an intermembrane space between them. The space enclosed by the inner membrane is called the matrix. Mitochondria have their unique transcriptional and translational machineries, and a unique circular genome (mitochondrial DNA, mtDNA) in the matrix containing 37 genes, 13 of which encode for components of the respiratory chain (RC) along with 2 rRNAs and 22 tRNAs (Anderson et al., 1981). The number of mitochondria in a cell depends on the cell-type specific metabolic needs, with most cells containing hundreds. Mitochondria are constantly undergoing fission and fusion to replicate. Their high replication rate makes mtDNA susceptible to a higher mutation rate than nuclear DNA. Damage to mitochondrial networks is especially problematic for long-lived post-mitotic cells like neurons, cardiomyocytes, renal tubular cells, and pancreatic islet cells (reviewed in Yaffe, 1999; Van der Bliek, Sedensky & Morgan, 2017).

The primary function of mitochondria is to produce energy for cellular and mechanical functions in the form of adenosine triphosphate (ATP) from sugars, fats, and other chemicals with the help of molecular oxygen. The major source of ATP is through oxidative phosphorylation (OXPHOS), where ATP is synthesised in the inner mitochondrial membrane through the electron transport chain (ETC), which is the transfer of electrons from NADH or $FADH_2$ to O_2 by a series of electron carriers. OXPHOS complexes I-IV form the ETC, and complex V is ATP synthase. Electrons are transported through acceptor proteins with serially higher reduction ability at the ETC complexes by a series of redox reactions and finally passed to O_2. The transport of electrons creates a proton gradient across the inner mitochondrial membrane, which creates a pH gradient and transmembrane electrical force, resulting in proton-motive force and finally phosphoryl transfer potential. The flow of protons back to the matrix is used by ATP synthase to phosphorylate ADP to create ATP (reviewed in Chan, 2006). As an additional by-product of OXPHOS, reactive oxygen species (ROS) are formed when an electron bypasses any of the ETC complexes, and oxygen is incompletely reduced. In normal amounts, ROS partake in cell signalling. In addition to ATP production, mitochondria play an important role for example in calcium storage and homeostasis for cell signalling activities, ROS signalling, iron-sulphur cluster and haem synthesis, and regulating cell growth and apoptosis (Craven et al., 2017; Ylikallio & Suomalainen, 2012).

1.2 Mitochondrial tRNA-synthetases (mt-aaRS) in health and mitochondrial disease

Humans have two distinct sets of essential nuclear-encoded aminoacyl tRNA-synthetases (aaRSs), the cytosolic- and mitochondrial aaRS. The genes for mitochondrial aaRS, or mt-aaRS are denoted with a 2 at the end of aaRS name, e.g. 'SARS2'. Their canonical role is as enzymes in the attachment of an amino acid into their isoaccepting cognate tRNAs. Each of the 19 aaRS and 19 mt-aaRS for each of the 20 amino acids act in a two-step reaction, first activating a single amino acid (aa) via an adenylate intermediate, and second covalently linking it to one of the ribose hydroxyls at the 3' end of a set of tRNA isoacceptors:

$ATP + aa \rightleftarrows aa - AMP + PPi$ (inorganic pyrophosphate) (1)

$Aa - AMP + tRNA^{AA} \text{ à } aa - tRNA^{AA} + AMP$ (2)

Once the tRNA is charged, a ribosome can transfer the amino acid from the tRNA into a position a growing peptide according to the anticodon of the tRNA. All mt-aaRSs are encoded by nuclear genes and synthesised in the cytosol, from where transportation to mitochondria occurs with the help of mitochondrial targeting sequence (MTS), which is cleaved off during entry to mitochondria (Figure 1). Mitochondrial aminoacyl tRNA synthetases can be categorised into two classes I and II based on their domain architecture and tRNA binding modes. Class I synthetases contain a characteristic Rossmann fold catalytic domain and are mostly monomeric. Class II synthetases have anti-parallel beta-sheet folds flanked by alpha helices, are mostly dimeric or multimeric, and contain at least three conserved domains (reviewed in Sissler, González-Serrano & Westhof, 2017).

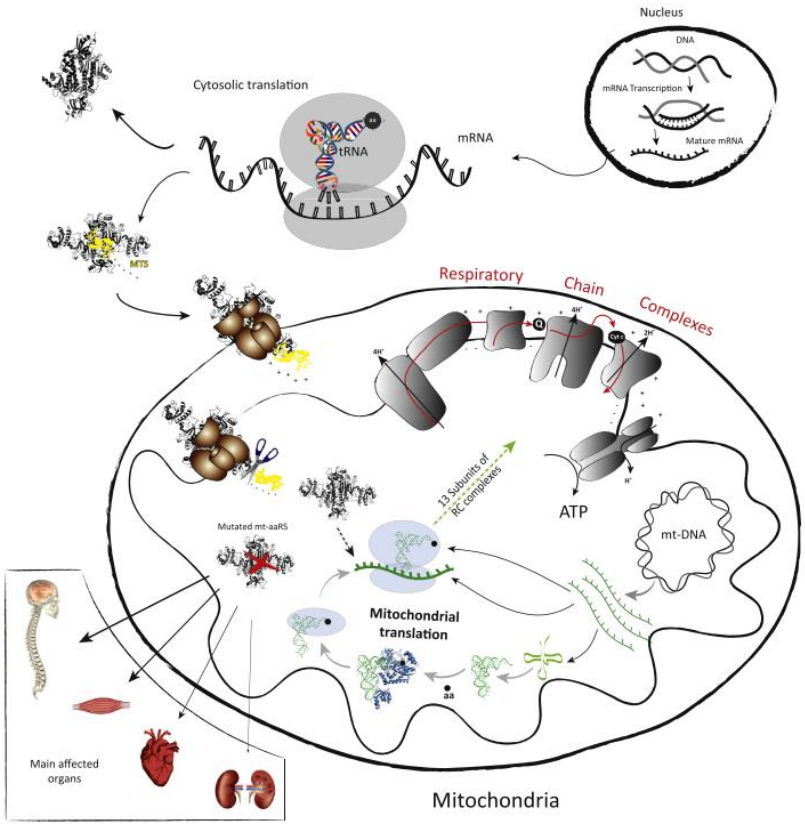


Figure 1. The localisation and canonical functions of mt-aaRS in mitochondrial translation, and the main organ systems affected by pathogenic mutations. Human mt-aaRS are nuclear encoded, synthesised in the cytosol, and delivered into the mitochondria by mitochondrial targeting sequence (MTS), which is cleaved upon entry to mitochondria. aaRS participate in mitochondrial translation. Tissue types and organs requiring high ATP amounts, such as nervous, muscular, cardiovascular, and urinary systems, are more severely affected by mutations in mt-aaRS genes (Figure source: Sissler, González-Serrano & Westhof, 2017).

Mitochondrial diseases refer to a genetically and clinically heterogenous group of disorders caused by OXPHOS dysfunction, that altogether make up the most common inherited metabolic diseases with a prevalence of 1:5000 (Diodato, Ghezzi & Tiranti, 2014). Inherited mitochondrial diseases can arise from mutations nuclear DNA or from the maternally inherited mtDNA. They can arise sporadically, or follow autosomal, recessive, maternal, or X-linked forms of inheritance (Suomalainen and Battersby, 2017). Mitochondrial diseases can manifest in all cells of the body except mature red blood cells, and most commonly in tissues with high energy expenditure (Craven et al., 2016). They can occur through defects in mitochondrial gene expression, nuclear-mitochondrial crosstalk, mtDNA maintenance, or mitochondrial stress responses, among other reasons (Suomalainen & Battersby, 2017). The expression of mtDNA in addition to the nuclear encoded DNA partaking in OXPHOS are crucial for OXPHOS assembly and function and maintaining overall cellular homeostasis. Defects in OXPHOS cause mitochondrial dysfunction and energy production impairment (Anderson et al.,

1981). In addition to ATP deficiencies, characteristics of mitochondrial disorders include systemic drop in pH caused by increased plasma lactate levels. OXPHOS dysfunction can lead to an elevated NADH/NAD ratio due to the inability of the citric acid cycle to feed forward NADH to the electron transport chain. Pyruvate levels are also increased by inhibition of the citric acid cycle. The elevated NADH/NAD ratio and increased pyruvate levels to shift the NADH-dependent lactate dehydrogenase towards production of lactate (Ylikallio & Suomalainen, 2012). Additionally, phenotype-specific signs can include ragged red fibres in muscle, exercise intolerance, or neurological problems, which are found in biochemical and histopathologic and exercise tolerance examinations in the clinic (Ognjenović & Simonović, 2017).

Several mitochondrial diseases involving mitochondrial protein synthesis and quality control deficiencies are caused by mutations in all of the nuclear encoded mt-aaRS. Defects in mt-aaRS involve either homozygous or compound heterozygous mutations, resulting in autosomal recessive disorders presenting as isolated clinical signs or syndromes, sometimes with overlapping clinical presentations (González-Serrano, Chihade & Sissler, 2019; Ognjenović & Simonović, 2017). Mutations in the synthetases cause depletion of charged tRNAs on ribosomes, resulting in translation interruptions causing OXPHOS assembly defects (Suomalainen & Battersby, 2017). Although all 19 mt-aaRS have the same biochemical function of charging tRNAs with their cognate amino acids (aminoacylation), mutations in different mt-aaRS lead to very different disease phenotypes of varying severity in different tissue types, presenting in various stages of life (Nunnari & Suomalainen, 2012) In addition to mt-aaRS mutations, mt-tRNA mutations exist causing an even wider variety of phenotypes (González-Serrano, Chihade & Sissler, 2019). These variables make them difficult to study (Ylikallio & Suomalainen, 2012), and no cures exist with treatment centred around relieving symptoms and treating complications (Suomalainen, 2011; Gorman et al., 2016). Mutations in mt-aaRS can cause tissue-specific or multi-organ syndromes almost exclusively affecting the central and peripheral nervous systems, and other organ systems with high mitochondria number and high ATP demand like musculoskeletal, cardiovascular, and urinary (Sissler, González-Serrano & Westhof, 2017; Vafai & Mootha, 2012).

Even mutations in one mt-aaRS gene can cause disparate phenotypes. For example, homozygous mutations in the gene *SARS2* coding for mitochondrial seryl-tRNA synthetase leading to amino acid changes p.D390G or p.R402H in the aminoacylation domain of the protein cause HUPRA syndrome (hyperuricemia, pulmonary hypertension, renal failure in infancy, and alkalosis) (Belostotsky et al., 2011; Rivera et al., 2013), but a splicing variant c.1347 G>A in *SARS2* causes spastic paresis (Linnankivi et al., 2016). Different alanyl-tRNA synthetase mutations have been linked to cardiomyopathy, or leukoencephalopathy, among others (Sissler, González-Serrano & Westhof, 2017; Götz et al., 2011). Pathogenic mutations in these enzymes can be studied in their downstream consequences, such as mitochondrial respiratory chain deficiencies, OXPHOS deficiency, or defective

mitochondrial protein synthesis. Disease mechanisms are also studied through consequences of mutations on protein architecture and resulting defects in functional domains (Moulinier et al., 2017).

Since many of the mt-aaRS mutations are not comparable, i.e. they affect the synthetases differently, variation in tissue- and organ specificity may be dependent on which an amino acid residue is affected and to what extent aminoacylation is compromised as a result (Diodato, Ghezzi & Tiranti, 2014). Some mt-aaRS contain an editing domain in addition to the canonical aminoacylation core domains. The role of the editing domain is to prevent the mischarging of a tRNA by deacylating any wrong amino acids either pre- or post- transfer (Schimmel & Schmidt, 1995; Hilander et al., 2017). Not all mt-aaRS contain an editing domain, for example *SARS2*, so mutations in other domains of the synthetase may be the cause of disease phenotypes. tRNA-charging defects, for example due to unstable catalytic function, have been suggested as the cause for disturbed mt-aaRS enzyme activity (Belostotsky et al., 2011). Reasons for variable age-of-onset are not known, but systemic involvement and more severe brain abnormalities are more prevalent in early-onset cases, and as reported in *AARS2* leukoencephalopathy (Fine et al., 2019).

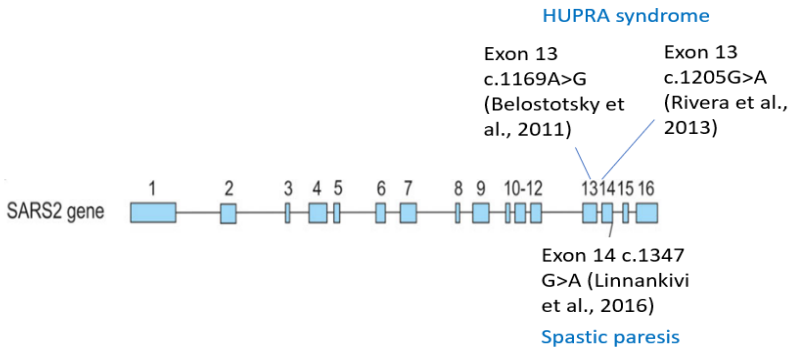
Mechanisms behind phenotype variability have been suggested. While non-canonical functions of mt-aaRS are poorly known, it is suggested that non-canonical, non-translational functions (e.g. immune response, inflammation, mTOR, IFN-γ and p53 signalling, neuronal differentiation) may play a role in phenotype heterogeneity of mt-aaRS mutations (Guo & Schimmel, 2013; Fine et al., 2019). Given the canonical function of mt-aaRSs to bind amino acids to their cognate tRNA, it can be expected that pathogenic mutations impair this pathway for example by reducing tRNA charging. Differences in mitochondrial chaperone activities between tissue types can affect the stability of some mutants (Konovalova & Tyynismaa, 2013). In neurons, the unique cell morphology and protein quality check may make them vulnerable to malfunction when parts of the protein quality check machinery are faulty (González-Serrano, Chihade & Sissler, 2019). Progressive phenotypes could be explained by faulty protein import, protein folding, or accumulation of misfolded and unfolded proteins that lead to endoplasmic reticulum (ER) stress and abnormal activation of unfolded protein response (UPR) and mitochondrial unfolded protein response (UPR^{mt}) (Rolland et al., 2019; Wang & Kaufman, 2016). Although reasons for heterogeneity of phenotypes and molecular mechanisms of mt-aaRS mutations have been hypothesised and even partly explained, disease modelling of the pathogenic genotypes and the functional analyses of their downstream consequences is required (González-Serrano, Chihade & Sissler, 2019).

1.3 Mitochondrial seryl-tRNA synthetase

SARS2 is a protein coding gene on chromosome 19 encoding for mitochondrial seryl-tRNA synthetase that contains 16 exons and spans 15.1 kb (structure shown in Figure 2 A). It belongs to the class II aminoacyl tRNA synthetase family, that are mostly dimeric or multimeric with an anti-parallel beta-sheet flanked by alpha helices. The aminoacyl group is coupled to the 3'-hydroxyl site of the tRNA.

The product is a 518 amino-acid protein that weighs 58 kDa (Figure 2 B). Specifically, it provides serine aminoacylation to two mitochondrial tRNAs, tRNA-ser (AGY) and tRNA-ser (UCN) by catalysing the esterification of seryl to both compatible tRNAs to form an aminoacyl-tRNA. tRNA-ser (AGY) is a small mitochondrial tRNA, completely lacking the D-domain (Steinberg et al., 1994).

A.



B.

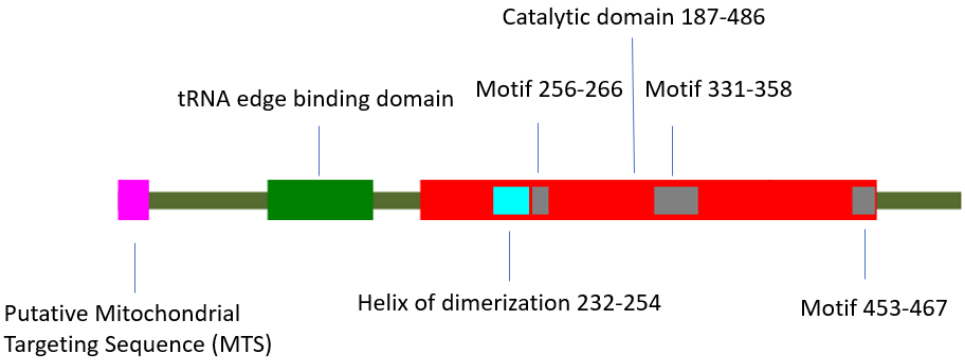


Figure 2. A. The *SARS2* gene encoding mitochondrial seryl-tRNA synthetase, containing 16 exons (Linnankivi et al., 2016), including the mutations that have been identified and associated phenotypes. **B.** The mitochondrial seryl-tRNA synthetase protein. Structures from MiSynPat.org (Moulinier et al., 2017).

1.4 *SARS2* patient genetics and clinical phenotype

Linnankivi, Neupane, Richter et al. in 2016, reported that a patient presenting with a severe childhood-onset spastic paresis had a homozygous single nucleotide variant c.1347 G>A in the last nucleotide of exon 14 of seryl tRNA-synthetase. The mutation was revealed by the MGZ Mitochondrial Diseases Multi Gene Panel (ID 87.01) containing 168 genes. Spastic paresis refers generally to an increased muscle tonus and tendon reflexes, and partial loss of voluntary movement. The nucleotide change causes a frameshift after exon 14 that completely changes the C-terminus amino acid sequence and results in a premature stop codon. The expected abnormal protein is 510 amino acids, compared to the normal, which is 518. The parents of the patient were heterozygous for the same mutation (Figure 3) and healthy. The altered C-terminus of the synthetase was shown to severely compromise tRNA charging and affect tRNA-ser (AGY) specifically. Western blot revealed a clear reduction in SARS2 protein level in patient fibroblasts to just 5% of the level in normal fibroblasts, but no changes in the cytochrome C oxidase subunit 1 and 2 (COX1 and COX2) proteins of respiratory complex IV encoded by mtDNA. Mitochondrial protein synthesis was normal (Linnankivi et al., 2016).

The patient's genotype and phenotype differ markedly from the previously reported homozygous *SARS2* missense mutations in exon 13 c.1169A>G resulting in p.D390G, and c.1205G>A resulting in p.R402H (Belostotsky et al., 2011; Rivera et al., 2013) causing HUPRA syndrome. HUPRA syndrome is a multisystem phenotype characterised by progressive renal failure in infancy leading to electrolyte imbalances, metabolic alkalosis, pulmonary hypertension, hypotonia, and delayed development (Belostotksy et al., 2011). While both mutations have destabilising effects on tRNA-ser (AGY) isoacceptors, the SARS2 patients with HUPRA syndrome were reported to have only 10-20% of the normal amount of tRNA-ser (AGY) in lymphocyte culture, while the spastic paresis patient had levels of tRNA-ser (AGY) reduced to 40% of normal fibroblasts. The differential level of stable tRNA-ser (AGY) and the resulting failure of SARS2 to aminoacylate tRNA-ser (AGY) leading to its degradation may be an explaining factor behind the large phenotype differences (Linnankivi et al., 2016).

The patient's development was normal until age 7-8 months, when the first signs of increased muscle tone in the legs were noticed. By age 5, spasticity progressed to involve upper extremities and oral motorics. While MRI at 1 year of age already showed a T2 signal increase in dentate nuclei, atrophy of cerebellar hemispheres and vermis were present at 5 years of age. Thalamic MR spectroscopy at 7 years of age showed elevated lactate levels. Muscle biopsy and respiratory chain enzymes were normal. Kidneys appeared normal in an abdominal ultrasound, but a liver biopsy showed mild steatosis. Blood-count, acid-base balance, electrolytes, plasma creatinine, cystatin C, uric acid, phosphate and parathormone levels were normal. Also, blood lactate, pyruvate, and urine organic and

amino acids were normal. Neuropsychological assessment at 8 years of age indicated cognitive developmental level of a 4.5-year-old. Since age 10 in 2016 the patient has had severe spastic tetraparesis.

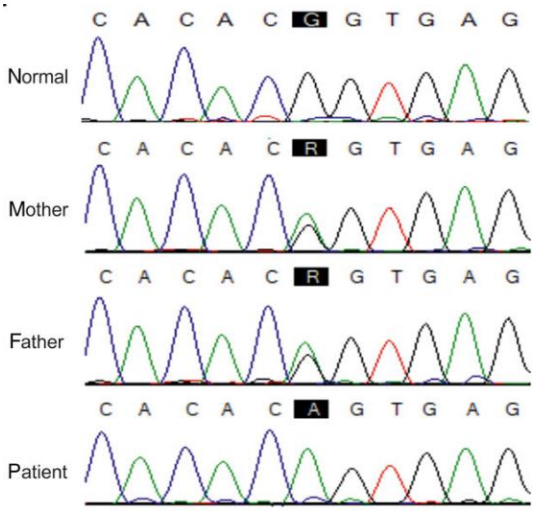


Figure 3. Sequencing traces of the patient's and parents' DNA compared against a normal healthy individual to show the splicing variant c.1347 G>A in *SARS2* (source: Linnankivi et al., 2016).

1.5 Induced pluripotent stem cells in disease modelling

Induced pluripotent stem cells (iPSC) are stem cells that have been reprogrammed from differentiated somatic cells back to a pluripotent stem cell state. iPSC resemble embryonic stem cells and can by definition differentiate into any cell type of the three embryonic germ layers (endoderm, mesoderm, ectoderm), and maintain the ability to proliferate indefinitely in culture. While embryonic stem cells originate from pre-gastrula embryos, limiting their use in research due to ethical concerns, iPSC overcome these issues. iPSC form into dense colonies with clear borders. The first reprogramming was reported in 2006, when Shinya Yamanaka's lab generated iPSC from embryonic and adult fibroblasts of mice by the activation of four genes encoding transcription factors, c-Myc, Oct3/4, Sox2 and Klf4 (Takahashi & Yamanaka, 2006). In 2007, the lab reprogrammed the first human iPSC using the same transcription factors (Takahashi et al., 2007). The applications of iPSC include disease modelling, regenerative medicine as autologous cell transplants without immune rejection concerns, and drug candidate and drug toxicity screening (Sterneckert, Reinhardt & Schöler, 2014; Wiegand & Banerjee, 2019; Lorenz et al., 2017). The use of isogenic iPSC in drug candidate screening is promising due to their possibility to reveal drug efficiency concerning the mutation of interest (Robinton & Daley, 2012; Barral & Kurian, 2016).

iPSC have been increasingly used since the 2010s for disease modelling of genetic diseases, especially ones with unknown molecular disease mechanisms. iPSC are an especially important model system for neuronal and cardiac phenotypes, where extraction of primary tissue is difficult and

harmful (Bellin et al., 2012). The standard of disease modelling is that a model is based on a known genetic lesion in a specific organism, represents the phenotypes present, and can model clinical manifestations and comorbidities (Doss & Sachinidis, 2019). Considering these validity parameters, iPSC disease models provide an advantage over patient primary cells because it is possible to create cell type-specific models that contain patient-specific genetic background without the need for invasive and many times impossible sample collection, or collection of large amounts of patient cells (Sterneckert, Reinhardt & Schöler, 2014; Bassett, 2017). To ensure their validity as stem cells, newly reprogrammed iPSC and cell lines created from them are subjected to quality control analyses. These include measuring pluripotency by the presence and quantity of surface markers TRA1-60, TRA1-81, SSEA3 and SSEA4, and/or transcription factors such as OCT4, SOX2, NANOG and LIN28A (O'Malley et al., 2013; Loh et al., 2006; Yu et al., 2007).

Differentiation methods that utilise knowledge of endogenous developmental pathways that lead to a certain cell type are being developed and used, especially for neurons, cardiomyocytes and pancreatic β-cells (Du et al., 2015; Yoshida & Yamanaka, 2017; Balboa et al., 2018). It also holds promise for studying complex phenotypes such as multi-system syndromes, in that they can establish whether the mutation of interest is responsible for the defects in different cell- or tissue types (Soldner et al., 2011; Yusa et al., 2011; Hockemeyer & Jaenisch, 2017). Such defects can be studied using genome engineering of a patient mutation in a parental cell line. In addition to knocking in a gene variant of interest in an isogenic cell line, the mutation in patient iPSC can be similarly corrected using genome editing. Patient mutations that have been corrected to be healthy should then behave like wild type cells (Sterneckert, Reinhardt & Schöler, 2014; Giacalone et al., 2018).

Other reprogramming methods are under development in order to improve pluripotency and thereby reliability, validity and efficiency of disease modelling pipelines. Human iPSC have most commonly been reprogrammed from fibroblasts or blood cells due to their availability and non-invasiveness of sampling. However, iPSC and cells differentiated from them may sometimes retain characteristics of cells of foetal origin (Hrvatin et al., 2014), and can contain epigenetic features of the cell of origin (Kim et al., 2010). However, more recently it has been found that epigenetic effects on cellular state are minimal, and isogenic iPSC from fibroblasts and blood cells have similar differentiation propensities (Kyttälä et al., 2016). In addition to retroviral delivery of the four key transcription factors in 2006, adenoviral gene delivery, microRNAs, RNA interference (RNAi), use of small molecules to replace the introduction of transcription factors into cells, and CRISPRa based endogenous promoter activation are other iPSC reprogramming methods that hold potential in improving iPSC generation pipelines (Zhou et al., 2019; Weltner et al., 2018).

1.6 The creation and significance of isogenic control clonal cell lines

One iPSC disease modelling approach in studying the effect of genotype on the disease phenotype is using isogenic disease models, which are selected or engineered to model the genetics specific to a patient- or patient population *in vitro*. Isogenic disease models are used as controls in analyses of genotype-phenotype relationships (Avior, Sagi & Benvenisty, 2016). Patient iPSC and primary cell populations may contain background genetic variation that may mask subtle phenotypes, in addition to the disease mutation of interest revealed by gene panels or Whole Exome/Whole Genome sequencing. Creating the mutation of interest in a healthy control cell line allows characterisation of the significance of the mutation of interest on a phenotype in a more precise manner (Bassett, 2017). Discovering the role and importance of genes and proteins in different cell types can be important for the characterisation of their basic functions and their roles in disease. Isogenic controls made using human specimen can model human lesions with greater specificity to human physiology than animal models can (Shi, Inoue, Wu & Yamanaka, 2017; Ben Jehuda, Shemer & Binah, 2018).

There are several different approaches to creating clonal cell lines. Clonal cell lines must be uniform and identical, originating from a single original cell that has the selected or engineered characteristics. To achieve this, after genome editing or selection, a cell population must be isolated into single cells. A common method for many cell types is cell sorting (FACS) and subsequent expansion sorted cells into clonal cell lines (Ran et al., 2013). In FACS, a heterogeneous cell population is sorted one cell at a time based on the fluorescent characteristics and specific light scattering of each cell. Cells with certain target characteristics, such as genome-edited cells containing a fluorescent label tagging the target region can be separated from those cells that do not have the characteristic by FACS (Bonner et al., 1972; Ran et al., 2013; Haupt et al., 2018). Immunomagnetic separation and microfluidic cell sorting are other precise methods for cell separation (Rahmanian et al., 2016). For the purpose of *in vitro* disease modelling with clonal cell lines, cell populations can also be separated into single cells by manual colony picking. An advantage of this cell sorting approach over FACS is the higher survival rate of electroporated iPSC (Bruntraeger et al., 2019).

1.7 CRISPR-Cas

1.7.1 Origins of CRISPR-Cas and developments in research

The discovery of clustered DNA repeats in prokaryotes occurred around the same time in three different parts of the world (Hermans et al., 1991; Ishino et al., 1987; Nakata et al., 1989). A 1987 study investigating the role of isozyme conversion of alkaline phosphatase in *Escherichia coli*, which sequenced a DNA fragment spanning the region of *iap* gene (isozyme of alkaline phosphatase), was the first study to recognise short DNA repeats in prokaryotes (Ishino et al., 1987). In 1993, similar DNA repeats were observed in the archaea *Haloferax mediterranei*, and since then in nearly all of

archaeal and about half of bacterial genomes (Ishino et al., 2018). Across the 1990s and early 2000s, Francisco Mojica's work further demonstrated that the spacer regions of the DNA repeats were similar to bacteriophages (Mojica et al., 1993; Mojica et al., 2005).

In 2001, the term CRISPR for Clustered Regularly Interspaced Palindromic Repeats was conjoined to refer to the DNA-repeats found in archaea and bacteria. Four conserved genes that are found adjacent to CRISPR regions were designated as *cas* genes (CRISPR-associated genes) 1-4. The CRISPR locus consists of three components, which are *cas* genes, a leader sequence, and a repeat-spacer array. In 2005, Bolotin's studies of *Streptococcus thermophilus* discovered that the CRISPR array contained novel cas genes, including a large cas gene with nuclease activity that later became known as Cas9. The study also revealed that the spacers that have homology to viral genes all shared a common sequence at the end, the PAM (protospacer adjacent motif) sequence (Bolotin et al., 2005).

The role of CRISPR-Cas in bacterial immunity for defence against viruses gathered more evidence in 2006 (Makarova et al., 2006). CRISPRs consist of DNA sequence repeats, and spacer sequences, which are remnants of foreign DNA of viruses that have previously invaded. In 2008, it was demonstrated that spacer sequences derived from phages or plasmid DNA are transcribed into CRISPR RNAs (crRNAs), which are small RNAs that guide Cas proteins to the target and interfere with the foreign genome (Brouns et al., 2008). The crRNA interferes with DNA also, not only RNA as previously hypothesised based on the view that CRISPR is parallel to RNAi interference in eukaryotes (Maraffini & Sontheimer, 2008). Garneau et al. (2010) demonstrated the distinguishing feature of Type II CRISPR systems at work, which is that double-strand breaks can be created three nucleotides upstream of the PAM mediated by Cas9 in conjunction with crRNA. In 2011, small RNA sequencing of *Streptococcus pyogenes* found that the CRISPR-Cas9 system contains another small RNA, trans-activating CRISPR RNA (tracrRNA), that forms a duplex with crRNA and guides Cas9 to the target location (Deltcheva et al., 2011). In 2013, the Zhang lab published the first CRISPR-Cas method to edit mouse and human genomes (Cong et al., 2013).

1.7.2 The CRISPR-Cas immunity mechanism and types of CRISPR-Cas

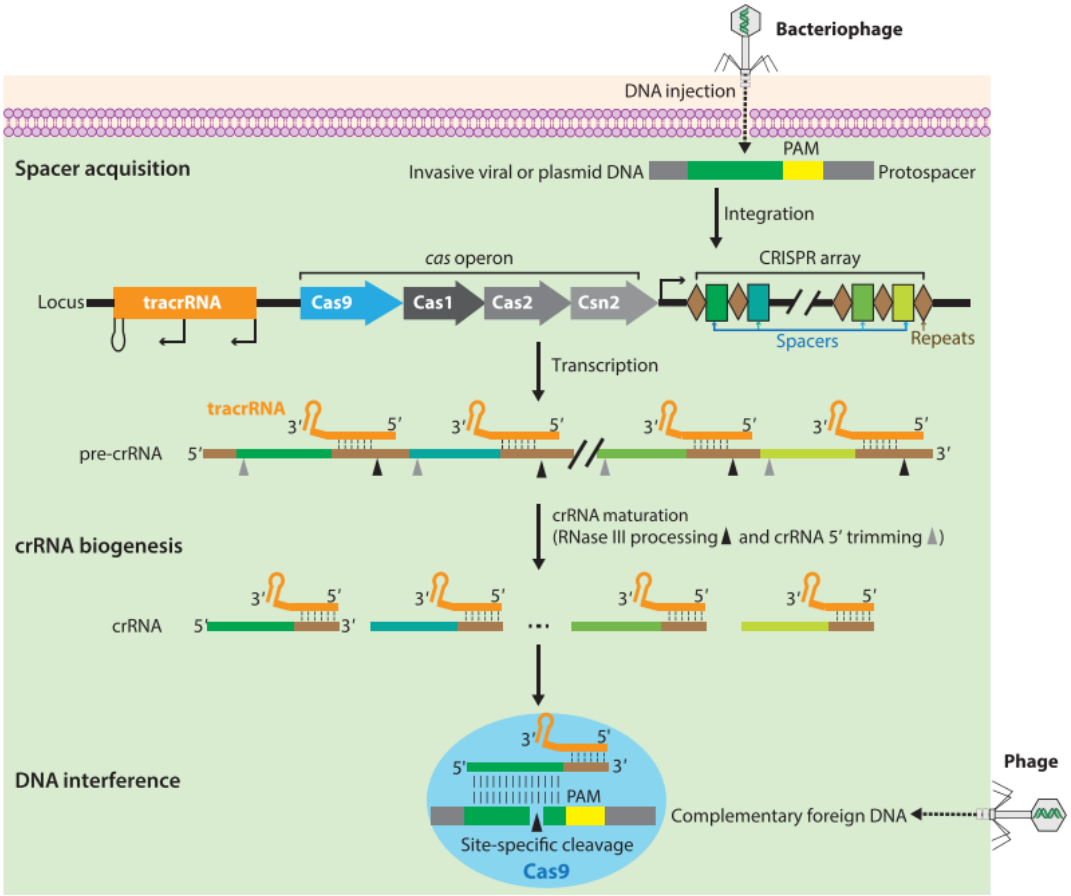
The CRISPR-Cas adaptive immunity system can be divided into three phases by which it recognises and cleaves foreign DNA or RNA: (i) Adaptation, (ii) crRNA biogenesis, and (iii) Target interference (Figure 4a). In the adaptation phase, a prokaryote comes into contact with an invading Mobile Genetic Element, from which a protospacer sequence is incorporated into the CRISPR array (collection of identical repeat sequences) as a new spacer. In the crRNA biogenesis phase, immunity is enabled when the CRISPR array is transcribed into a long precursor crRNA (pre-crRNA), which is further processed by Cas proteins into a mature crRNA containing the sequences of invaders. In the target interference phase, mature crRNAs activate the CRISPR mechanism and act as guides to target and

interfere with the cognate invading DNA or RNA, and the Cas protein cleaves them (Hille & Charpentier, 2016).

CRISPR-Cas systems can be divided into two main classes, which involve six types and several more subtypes, based on Cas genes, operons, and CRISPR loci (Barrangou, 2015). Class 1 CRISPR-Cas systems (types I, III and IV) have an effector Cas protein module consisting of a multi-protein complex. Class 2 systems (types II, V and VI) use only one effector protein for target interference. To avoid self-targeting, types I and V recognise the PAM sequence that is located upstream of the protospacer, and type II recognises the PAM downstream of the protospacer (Ishino et al., 2018; Hille & Charpentier, 2016). Type I consists of multiple Cas proteins, forming a chain with the Cas3 protein. Type II involves its characteristic protein Cas9. Cas1 and Cas2 are involved in adaptation in both main CRISPR-Cas types (Ishino et al., 2018).

Type II CRISPR-Cas systems, or CRISPR-Cas9, are characterised by the CRISPR-repeat spacer array, a *Cas9* gene as the first gene in an operon containing two to three cas adaptation modules (Cas1, Cas2, csn2, or Cas4), along with a tracrRNA (Charpentier et al., 2015). Cas9 has six domains, REC I, REC II, Bridge Helix, PAM-interacting (PID), and HNH and RuvC (Jinek et al., 2012). The PID recognises the PAM sequence in the invading genome. REC I is involved in binding the guide RNA to the target DNA. The PID binds to PAM, and initiates binding to target DNA. A defining feature of the type II (and type I) systems is the R-loop, which forms when the guide RNA segment of a crRNA invades a DNA double helix to form an RNA-DNA hybrid helix with the target DNA strand while displacing the non-target strand (Figure 4b) (Jinek et al., 2012). In CRISPR-Cas9, the Cas9 endonuclease together with a crRNA and a tracrRNA can form an R-loop (Jiang et al., 2016). The R-loop interaction occurs on the target strand by complementary pairing with the guide RNA, with a double-stranded DNA cleavage by Cas9 3 bp upstream of the PAM. The double strand cleavage occurs via the HNH and RuvC catalytic sites (Anders et al., 2014). In *Streptococcus pyogenes*, the PAM is 5'-NGG-3.' Each nuclease has a characteristic PAM (Jinek et al., 2012).

A.



B.

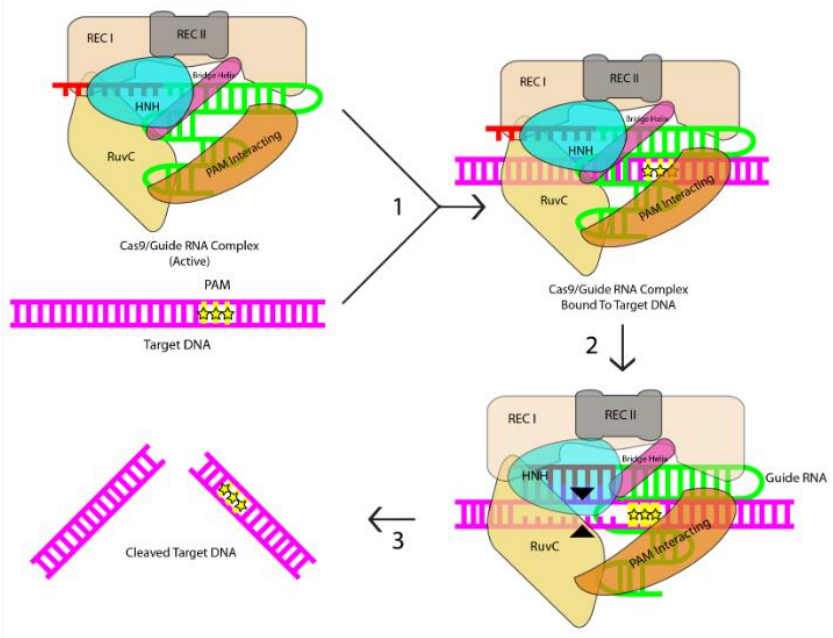


Figure 4. The CRISPR-Cas9 system. (A) CRISPR-Cas9 mediated DNA interference in bacterial adaptive immunity. The CRISPR-Cas adaptive immunity mechanism is usually described in three

steps: First in adaptation, a new spacer (dark green) from invading phage DNA is incorporated into the CRISPR array by acquisition machinery (Cas1, Cas2, Csn). The tracrRNA precedes the Cas operon, and encodes unique non-coding RNA homologous to the repeat sequences. After integration, the new spacer is transcribed together with all other spacers into a long precursor-crRNA (pre-crRNA) containing repeats (brown lines) and spacers (dark green, blue, light green, yellow lines). In crRNA biogenesis, the tracrRNA anneals to the pre-crRNA repeats for crRNA maturation by RNase III cleavage. The length of the crRNA is reduced to ~20 nt by trimming at the 5' end (grey arrows). During DNA interference, the crRNA-tracrRNA structure directs the Cas9 endonuclease to cleave the foreign DNA that contains the 20 nt crRNA complementary sequence preceding the PAM site (Image from Jiang & Doudna, 2017). **(B)** Target DNA binding and cleavage by the Cas9 endonuclease. (1) The crRNA-tracrRNA guided Cas9 scans the DNA for a suitable PAM. (2) When the crRNA sequence binds to the target DNA, (3) the RuvC and HNH domains of Cas9 cleave the DNA 3 bp upstream of the PAM (Image from Jinek et al., 2014).

1.7.3 CRISPR-Cas9 in genome editing and disease modelling

Genome editing (or gene editing) refers to the technologies used to target and add, remove, or alter parts of an organism's DNA (Fernandez et al., 2017). CRISPR-Cas9 has been found to be the most efficient and customisable for editing multiple genes at once and in a range of eukaryotic cells, compared to other previously developed genome editing methods. Different genome editing applications including CRISPR-Cas9 are used increasingly widely in disease modelling in cell-based and whole animal-based systems (Paquet et al., 2016; Jin et al., 2019; Platt et al., 2013). Genetic loci can be knocked-out or modified, or transcription regulated, to investigate their function in a particular biological pathway and phenotype (Hille & Charpentier, 2016; Paquet et al., 2016). CRISPR-Cas9 systems are also being researched as diagnostic tools for the diseases being modelled, antimicrobial treatments, and as autologous transplantation treatment for degenerative disorders, to target the genome and epigenome of cells in cancers (Martinez-Lage et al., 2018), and to repair mutations *ex-vivo* in somatic and induced pluripotent stem cells for the treatment of monogenic disorders (reviewed in Savic &Schwank, 2016). In 2020, the Nobel prize in Chemistry was awarded to Jennifer Doudna and Emmanuelle Charpentier for the development of the CRISPR-Cas9 method for genome editing.

Genome editing methods used for disease modelling and therapy research prior to CRISPR-Cas9 include zinc-finger nucleases (ZFN) and transcription activator-like effector nucleases (TALEN). Beginning from early 2000s, ZFN were the first generation of genome editing tools applied to gene correction, addition and disruption in plant and mammalian cells including mouse and human pluripotent stem cells (Urnov et al., 2010). The ZFN structure consists of a zinc finger protein (ZFP) array of 3-6 Cys2-His2 zinc-finger domains fused in tandem, with each domain binding to 3 nucleotides, and a Fok1 endonuclease domain fused with an intervening linker to the ZFP (Kim et al.,

1996). TALENs also work through a FokI endonuclease fused to a DNA binding domain derived from TALE proteins originally found in *Xanthomonas sp*. Each domain interacts with a single target nucleotide and allows genome editing in any location in the genome through the nuclease creating a DSB that gets corrected through NHEJ. However, the Fok1 endonuclease works in a dimer and is cumbersome to design to target a very specific region (binds 3 nucleotides in ZFN) and to transfect to some cell types (Boch et al., 2009). Although ZFN and especially TALEN setups have been improved for life science and disease modelling , the advantage of CRISPR-Cas9 over other genome editing methods is its efficiency due to the robustness and relatively small size of the Cas9 endonuclease, speed of assembly, customisability and affordability of guide RNAs for a range of mammalian cells, especially difficult to transfect cells such as iPSC (Ding et al., 2013; Ran et al., 2013, Hockemeyer & Jaenisch, 2017; Giacalone et al., 2018).

CRISPR-Cas is used as a genome editing technology repurposed and adapted from the naturally occurring genome editing system that archaea and bacteria use for adaptive immunity. CRISPR-Cas9 genome editing harnesses the use of a synthetic sgRNA, usually consisting of a crRNA and tracrRNA in a duplex, and Cas9 endonuclease. sgRNAs are non-coding RNA sequences with helicase- and nuclease active components, that target the complementary target DNA locus. Allele-specific and non-allele specific sgRNAs can be synthesised by polymerase chain reaction (PCR) (Balboa et al., 2015), or be cloned into plasmids that are transfected into cells. The PCR-synthesised sgRNA transcriptional cassette consists of a U6-promoter that controls sgRNA expression in transfected cells, and crRNA-tracrRNA complex consisting of a 20 bp gRNA and a 19 bp matching tracrRNA (Balboa et al., 2015).

There are two main pathways by which DSBs are repaired: non-homologous end joining (NHEJ) and homology directed repair (HDR). NHEJ occurs more commonly in eukaryotic cells when repairing a DSB created by a Cas endonuclease. NHEJ usually results in insertions/deletions (INDELs), which can result in frameshifts or premature stop codons, preferable for making knockout mutations (Hille & Charpentier, 2016). HDR, the most common form being homologous recombination, can be harnessed to repair mutations or insert precise mutations with a homologous DNA sequence (Lieber, 2010). The homologous duplex template can be naturally occurring, or synthetic to create large modifications, such as a knock-in mutation, or to insert reporter genes (Ran et al., 2013; Cong et al., 2013). HDR usually only occurs in the G2/S cell cycle phase, while NHEJ is not dependent on cell cycle phase (Heyer et al., 2010). DSB repair through the HDR pathway also depends on cell type, the design of the repair template, and target location in the genome (Saleh-Gohari, 2004; Ran et al., 2013).

One method to deliver CRISPR-Cas components, including CRISPR-Cas9, into cells *in vitro* is by using a ribonucleoprotein (RNP) complex (Lin et al., 2014). An RNP complex consists of a targeting sgRNA in complex with a Cas9 protein. The sgRNA consists of a unique short ~20 nucleotide target-

specific CRISPR-RNA (crRNA) that corresponds to the target genomic DNA region on either strand, and a trans-activating RNA (tracrRNA) that is the stem-loop structure bound by the Cas9 nuclease (Figure 5). The genomic DNA region corresponding to the crRNA must be chosen so that the 3' end has a protospacer adjacent motif (PAM) sequence (5'-NGG-3', where N is any nucleotide base), which is a 2-6 nucleotide sequence downstream of the DNA sequence targeted by the crRNA. The crRNA-tracrRNA duplex directs the Cas9 protein to a site upstream of the PAM sequence, and Cas9, activated by binding of the crRNA, makes a DSB 3-4 bp upstream of the PAM (Jinek et al., 2014). In addition to the RNP complex, a repair template can be added in a plasmid, PCR-product or chemically synthesised oligonucleotide to insert a new sequence with any wanted mutation. Repair templates can be single- (single-stranded donor oligonucleotide, ssODN) or double stranded (double stranded DNA, dsDNA). An ssODN, recommended for single-nucleotide substitutions used in this project, is a 101-nucleotide long strand containing homology arms of ideally 75-100 nt (in this project 50 nt) corresponding to the sequence immediately flanking the mutation site at 5' and 3' ends. An ssODN can either be complementary to the 5'-3' or 3'-5' strand (Ran et al., 2013). Major advantages of ssODN over dsDNA templates are their lower toxicity and lower frequency of random integration (Li et al., 2017).

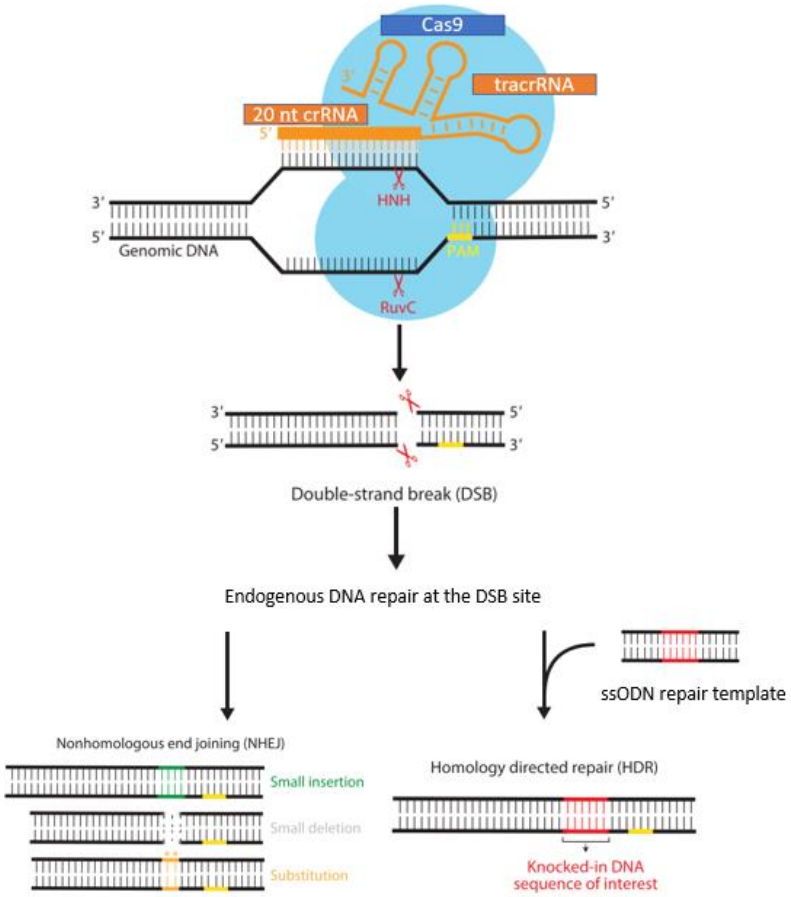


Figure 5. Schematic of the CRISPR-Cas9 ribonucleoprotein (RNP) system for gene editing through the NHEJ and HDR pathways. The crRNA contains a unique 20 nt sequence complementary to the target sequence in the genome. The crRNA-tracrRNA duplex guides the Cas9 endonuclease to the target region to create a double-strand break 3-4 nucleotides upstream of the PAM, resulting in a DSB. A DSB can be repaired by NHEJ to make insertions, deletions or substitutions, or through HDR with a homologous repair template to make a precise desired modification (Figure adapted from Jiang & Doudna, 2017).

Electroporation is the most common and efficient RNP nucleofection method to date for introducing substrates into the cytoplasm and nucleus of a cell, especially for difficult to transfect cells like iPSC (Ihry et al., 2018). Electroporation creates pores in the lipid bilayer of the cell membrane, increasing permeability for the RNP complex to enter the cytoplasm. Since the RNP complex degrades relatively quickly and acts on the cells only for 24 hours post-electroporation, so mosaicism and off-target cleavage and mutagenesis by the spCas9 protein can be minimised (Kim et al., 2014; Bruntraeger et al., 2019). In addition to cutting at the site targeted by the crRNA, Cas9 can cut imperfectly matched sequences elsewhere in the genome, called “off-target” insertions and deletions (Fu et al., 2013). Off-targets can be reduced by designing truncated <20 nt gRNAs with low similarity to sequences elsewhere in the genome (Fu et al., 2014). The PAM sequence should be as close as possible to the

mutation and not be present elsewhere in the genome in high quantities. An ssODN as close as possible to the DSB site is most efficient for making single-nucleotide changes (Bruntraeger et al., 2019). To maximise chances of HDR-pathway exploitation by the ssODN, HDR enhancing small molecules can be added into transfected cell populations. HDR enhancers work by inhibiting NHEJ (e.g. with DNA-PKc inhibitors and other inhibitors for enzymes involved in DSB processing) (Shrivastav et al., 2008), or by enhancing HDR, by compounds arresting cells in G2/S phase when HDR is more likely to occur. Naturally, cells in the proliferative phase are more favourable to HDR (Heyer et al., 2010).

1.8 Aims

The overarching aim of this project was to create the patient mutation c.1347 G>A in *SARS2* into an established healthy iPSC line. The edited iPSC lines are then applicable for differentiation into the affected cell type for further discovery of disease mechanisms.

The more specific aims were:

1. To create iPSC lines using CRISPR-Cas9 genome editing for disease modelling of a patient-specific *SARS2* splicing variant c.1347 G>A that causes severe childhood-onset progressive spastic paresis.

2. To analyse the efficiency and precision of the genome-editing and quality of the genome-edited iPSC lines, and thereby validity of the models to be used in studies of disease mechanisms.

3. To analyse the effect of the *SARS2* splicing variant c.1347 G>A on mRNA expression and protein level in the genome-edited iPSC lines.

2. Materials & Methods

2.1 Cell culture

2.1.1 HEK293 cells

HEK293 cells were cultured on a 100 mm cell culture dish (Thermo Scientific, #130182) in DMEM (Lonza, #BE12-614F)- based cell culture medium containing 10% Foetal Bovine Serum (FBS) (Merck, #TMS-013-B), 1% penicillin-streptomycin (Lonza, #09-757F), and 1% L-Glutamine (Thermo Scientific, #25030149). Cells were passaged using 1x Trypsin (Thermo Scientific, #R001100) and stored in an incubator at conditions +37 °C, 5 % CO_2 and 95 % O_2.

2.1.2 iPSC

Human iPSC line HEL24.3 was previously reprogrammed from a healthy control's neonatal foreskin fibroblasts by Sendai virus technology (Trokovic, Weltner & Otonkoski, 2015) at Biomedicum Stem Cell Center (University of Helsinki, Finland). *SARS2* patient cell line HEL204.6 was reprogrammed from the patient's skin fibroblasts into pluripotent stem cells also at the Biomedicum Stem Cell Center.

The patient and control iPSC (for use as healthy control and as parental for CRISPR-Cas9 electroporation) were cultured on 10 µg/mL Vitronectin XF (STEMCELL Technologies, #07180) in PBS (Lonza, #17-516F) coated 35 mm cell culture plates (Thermo Scientific). Cell culture medium was TeSR™-E8™ basal medium (#05990) supplemented with TeSR™-E8™ supplement (STEMCELL Technologies) and 1:500 Primocin antibiotic (Invivogen, #ant-pm-05). Post-thawing and pre- and post-electroporation, the cell culture medium was also supplemented with 1:100 RevitaCell Supplement (Thermo Scientific, #A2644501) that contains p160-Rho-associated coiled-coil kinase (ROCK)-inhibitor for 24 hours to regulate apoptosis and aid cell survival. Cells were passaged to multiple dishes upon reaching 80-90% confluency, using 1 mL of Gentle Cell Dissociation Reagent (STEMCELL Technologies, #07190) per 35 mm dish. iPSC were stored in an incubator at conditions +37 °C, 5 % CO_2 and 95 % O_2.

2.1.3 Fibroblasts

Fibroblasts were cultured on a 100 mm cell culture dish (Thermo Scientific, #130182), in DMEM (Lonza)- based cell culture medium containing 10% Foetal Bovine Serum (FBS) (Merck, #TMS-013-B), 1% penicillin-streptomycin (Lonza, #09-757F), 1% L-Glutamine (Thermo Scientific, #25030149), and 0.2% Uridine. Fibroblasts were stored in an incubator at conditions +37 °C, 5 % CO_2 and 95 % O_2.

2.2 crRNA design

Four crRNAs to target *SARS2* exon 14 in iPSC were designed using the web tools https://benchling.com/crispr and https:/crispr.mit.edu based on the SARS2 gene Ensembl gene ENSG00000104835, location Chromosome 19: 38,915,266-38,930,763 (GRCh38.p13, Homo sapiens), between positions 33769-33849. The crRNAs were 20 nt long, with SpCas9 PAM NGG. crRNA1 and crRNA2 were designed to target the forward genomic strand and crRNA3 and crRNA4 targeted the reverse genomic strand. Cut distances as close as possible to the c.1347 G>A mutation was 5 nt in crRNA1 (in the PAM region), 0 nt for crRNA 2 and 3, and 1 nt for crRNA 4 (Table 1). The crRNAs tested in HEK293 cells are called crRNA 1-4, and of these the two crRNAs (1 and 4) chosen for iPSC electroporation are called crRNA 1 and 2 respectively.

Specificity and efficiency of the crRNAs was simulated by CRISPOR version 4.97 (Concordet & Haeussler, 2018; http://crispor.tefor.net/) and Benchling. CRISPOR's CFD (cutting frequency determination) specificity score ranges from 0-100, 100 being most specific associated with fewer off-target effects. It correlates better with the total off-target cleavage fraction of a guide and has fewer false-positive off-targets and is more sensitive to off-targets with a single-mismatch from the guide RNA than the MIT-score, which the CFD is modelled after to measure uniqueness of a guide RNA in the target reference genome (Tycko et al., 2019). CFD specificity scores were 93 for crRNA1 and 82 for crRNA2. The specificity of the crRNAs were sufficient enough (CFD of >60) to be chosen for the CRISPR-Cas9 editing of iPSC. Benchling provided on-target and off-target scores from 0-100, 100 being better. On-target scores represent the probability of Cas9 binding to the region targeted by the crRNA. The off-target score is the inverse probability of off-target binding by Cas9 endonuclease, so a higher score indicates a lower chance of Cas9 binding elsewhere in the genome (Doench et al., 2016). crRNA oligonucleotides were supplied by Integrated DNA Technologies (IDT) (Table 1).

crRNA	**Sequence (5' → 3')**	**Orientation**	**PAM**	**Cut distance to mutation (nt)**	**CFD specificity score (CRISPOR)**	**On-target/off-target scores (Benchling)**
1	GGAGCTGCAGTTTGCCCACA	Forward	cgg	5	93	49.3/38.7
2	TGCAGTTTGCCCACACGgtg	Forward	agg	0	93	53.2/45.5

3	gctgtgcgggcctcacCGTG	Reverse	tgg	0	89	48.4/41.1
4	ctgtgcgggcctcacCGTGT	Reverse	ggg	1	82	49.7/45.6

Table 1. Characteristics of the crRNA sequences designed to target *SARS2* exon 14

2.3 gRNA production

Transcriptional units for gRNA expression were prepared by PCR-amplification and concatenated using Golden Gate assembly as described in Balboa et al., 2015. The PCR-amplification generated short products that include the 5' tailed U6 promoter, the gRNA sequence, and the tailed terminator. Four *SARS2* non-allele specific gRNA sequences were produced (Table 4). All PCR amplifications were done using the Bio-Rad S1000^{TM} Thermal Cycler.

The 5' tailed U6 promoter was generated by digesting plasmid PX335-dCas9-VP192-PGK puro (gift from Timo Otonkoski) with restriction enzymes Bbsl and Pvul (Thermo Scientific, #ER1011 and #ER0621). 5 ng of the product was run on a 1% w/v agarose gel (in TBE) for 1 hour, and the 1500 bp band cut from the gel under UV light and purified using the Nucleospin Tissue XS kit (Macherey-Nagel # 740901.50). 20 ng of the purified U6 template was used for a PCR reaction, along with 4 µL 5X HF buffer, 1.6 µL of 2.5 mM dNTP's, 1 µL each of both forward and reverse primers, 0.2 µL Phusion enzyme (Thermo Scientific # F530S), filled up to 20 µL with nuclease-free H_2O (Table 2). The PCR product was purified using the PCR and Gel Cleanup kit (Macherey Nagel #740609.50) and digested with Bbsl overnight to eliminate any remaining plasmid. The product was run on a 1% agarose gel, and the the 266 bp band cut from the gel under UV light and purified through the Nucleospin Tissue XS kit (Macherey Nagel). SYBR® Safe DNA gel stain was used as the nucleic acid visualisation stain in the agarose gel (Thermo Scientific, #S33102).

Number of Cycles	**Temperature (˚C)**	**Time**
1 x	98	3 min
35 x	98 60 72	10 s 30 s 9 s
1 x	72	8 min

Table 2. PCR for U6-promoter

The previously digested U6 promoter (20 ng) and tailed terminator were then PCR amplified separately. For the PCR reactions of both products, 20 µL of HF buffer, 8 µL of 2.5 mM dNTPs, 0.5 µL of 100 µM Forward and Reverse primers, 1 µL of Phusion enzyme (Thermo Scientific), and 2 µL of U6 or TT template, and filled up to 100 µL with MilliQ water (Table 3). The PCR products were loaded on 1% agarose gel (BioNordika, #BN-50004) in TBE, with negative water controls included, 266 bp U6 band and 103 bp tailed terminator band were cut under UV light and purified using the PCR and Gel Cleanup kit (Macherey Nagel).

The final gRNA expression cassette including the U6 promoter and tailed terminator was assembled by a PCR reaction containing 5 ng of both the U6 promoter and tailed terminator, 20 µL 5x HF-

buffer, 8 µL 2.5 mM dNTPs, 2 µL 1 µM gRNA, 0.5 µL of 100 µM both Forward and Reverse primers, 1 µL Phusion enzyme (Thermo Scientific), and filled to 20 µL with nuclease-free H_2O and run with the cycling program (Table 2). 5 µL of the PCR products were loaded on a 1% agarose gel in TBE and checked for a 455 bp band. The rest of the PCR product was then purified using the PCR and Gel Cleanup kit (Macherey-Nagel).

Number of Cycles	**Temperature (°C)**	**Time**
1 x	98	3 min
35 x	98 52 72	10 s 30 s 12 s
1 x	72	8 min

Table 3: PCR to create tailed U6 and tailed terminator

5' tailed U6 promoter (266 bp)	Template	~1500 bp DNA template digested from plasmid
	Fw-primer	GTAAAACGACGGCCAGTGAGGGCCTATTTCCCATGATTC
	Rv-primer	GGTGTTTCGTCCTTTCCAC
	Final sequence	GTAAAACGACGGCCAGTGAGGGCCTATTTCCCATGAT TCCTTCATATTTGCATATACGATACAAGGCTGTTAGAG AGATAATTGGAATTAATTTGACTGTAAACACAAAGAT ATTAGTACAAAATACGTGACGTAGAAAGTAATAATTT CTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAAT GGACTATCATATGCTTACCGTAACTTGAAAGTATTTCG ATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAAC ACC
Tailed terminator (103 bp)	Template	AAAAAAAGCACCGACTCGGTGCCATTTTTCAAGTTGATAACGGACTA GCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC
	Fw-primer	GTTTTAGAGCTAGAAATAGCAAG
	Rv-primer	AGGAAACAGCTATGACCATGAAAAAAAGCACCGACTCGGTGCCAC
	Final sequence	<u>GTTTTAGAGCTAGAAATAGCAAG</u>TTAAAATAAGGCTA GTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGT GCTTTTTTTCATGGTCATAGCTGTTTCCT

Final gRNA PCR-product with U6 promoter and tailed terminator (455 bp)	Template	5’ tailed U6 promoter + gRNA oligo + tailed terminator
	Fw-primer	actgaattcggatcctcgagcgtctcaccctgtaaaacgacggccagt
	Rv-primer	catgcggccgcgtcgacagatctcgtctcacatgaggaaacagctatgaccatg
	Final sequence	ACTGAATTCGGATCCTCGAGCGTCTCACCCTGTAAAAC GACGGCCAGTGAGGGCCTATTTCCCATGATTCCTTCAT ATTTGCATATACGATACAAGGCTGTTAGAGAGATAAT TGGAATTAATTTGACTGTAAACACAAAGATATTAGTA CAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGT AGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTAT CATATGCTTACCGTAACTTGAAAGTATTTCGATTTCTT GGCTTTATATATCTTGTGGAAAGGACGAAACACCGNN NNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATA GCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAA AAAGTGGCACCGAGTCGGTGCTTTTTTTCATGGTCATA GCTGTTTCCTCATGTGAGACGAGATCTGTCGACGCGGC CGCATG

Table 4. Sequences in non-allele specific sgRNA PCR-product synthesis. The underlined portion is the gRNA.

2.4 gRNA testing in HEK293 cells

The four assembled and concatenated gRNA sequences were transfected into HEK293 cells to determine transfection efficiency. A 48-well plate of 90% confluent HEK293 cells was chosen for the jetPRIME (Polyplus) protocol. Three transfections were done for each gRNA, and alongside a GFP-Cas9 only control and a negative control without the jetPRIME reagent.

0.1 µg of sgRNA, 50 µL of jetPRIME buffer, 0.5 µg of GFP-Cas9-Plasmid were mixed and vortexed. 1 µL of jetPRIME reagent was added, vortexed for 10 seconds, spun down, and incubated at room temperature for 20 minutes. 50 µL of the transfection mix was added per well of the 48-well plate. The plate was rocked gently side to side and placed in an incubator in conditions +37 °C, 5 % CO_2 and 95 % O_2. The transfection medium was replaced with TeSR-E8 medium after 4 hours and returned to the incubator.

After 48 hours, GFP-positivity was checked on the EVOS® F fluorescence microscope (Life Technologies) for an estimate of transfection efficiency. Cells were collected by first washing with 400 µL PBS (Lonza) and adding 400 µL 1x Trypsin-EDTA in PBS, incubating at 37 °C for 3 minutes.

Cells were collected into 1 mL DMEM/F-12 (Gibco #11320033), centrifuged at 200 g for 3 minutes, supernatant aspirated, and pellet stored at -80 °C.

2.5 T7 endonuclease assay

The T7 assay (NE Biolabs #M0302S) is based on the formation of homo- and heteroduplexes between PCR-amplified genomic wild type and mutant DNA fragments. Mutant, in this case edited DNA, contain mismatches that are cleaved by the T7 enzyme and can be visualised on an agarose gel. A cleaved T7-result suggests sufficient activity of the non-allele specific gRNA, or gRNA at the target site.

Genomic DNA was first extracted by adding 15 µL QuickExtract DNA Extraction Solution (Lucigen #QE09050) into approximately 10^4 cells from a 96-well plate well, vortexing for 15 seconds and incubating at 65 ˚C for 6 minutes, vortexing for 15 seconds and incubating at 98 ˚C for 2 minutes. Genomic DNA extraction from cells expanded on 48-well and 35 mm cell culture plates was done according to instructions from the NucleoSpin Tissue Mini Kit (Macherey Nagel, #740952.50). PCR-amplification was set up with 12.5 µL of AmpliTaq Gold 360 Master Mix (Thermo Scientific # 4398881), 0.5 µL of 10 µm forward and reverse primers, 100 µg DNA template filled to a 25 µL reaction with nuclease-free H_2O (Table 5). The AmpliTaq Gold positive kit control (Thermo Scientific, #N8080241) was run in parallel. 5 µL of the PCR-amplified product was run on a 3% agarose gel at 110V for 1 h 30 min to detect a single band of correct size at 206 bp.

Number of Cycles	**Temperature (˚C)**	**Time**
1 x	95	10 min
40 x	95 60 72	30 s 30 s 30 s
1 x	72	7 min
1 x	4	Hold

Table 5. T7-assay thermocycling program

To proceed with the cleavage assay, heteroduplex formation was set up with 10 µL PCR-product, 2 µL NE-Buffer (NE Biolabs), filled to a 19.5 µL reaction with nuclease-free H_2O. Negative enzyme controls were run in parallel. The samples were denatured and annealed by thermal cycling at 95 °C and lowered by 2˚C per second from 95 to 85 ˚C, and -0.1˚C from 85 to 25 ˚C (Table 6).

Number of Cycles	Temperature (˚C)	Time
1 x	95	5 min
	95-85	-2˚C/s
	85-25	-0.1˚C/s
	4	Hold

Table 6: T7-assay denature and anneal thermocycling program

The annealed products are digested with the T7 endonuclease I. 1 µL of T7 endonuclease (NE Biolabs), or H_2O for negative enzyme controls, was added to the 19.5 µL of the annealed product. The samples were resuspended and incubated at 37 ˚C for 15 minutes and inactivated with 1 µL 0.5 M EDTA (Thermo Scientific #15575020). The samples were analysed on 2% agarose gel at 100 V for 1 h 30 min. Imaging was done with Bio-Rad ChemiDoc XRS+, and band intensity and percentage of DNA cleaved was quantified using ImageLab 6.0.

Gene modification efficiency *f* (percentage of DNA cleaved in the pool of cells) was calculated with the following equation:

$f = 100x\ [1 - (1 - fraction\ cleaved)^{\frac{1}{2}}]$, where the fraction cleaved is the sum of cleaved bands.

2.6 ssODN repair template design

Two repair templates to be used in conjunction with the two crRNAs to create the knock-in mutation c.1347 G>A were designed to be 101 nucleotides long (Benchling, https://benchling.com/crispr). Guanine was changed to adenine, and 50 nucleotides of the wild type DNA was added to both sides of the mutation site as homology arms. The patient mutation is in the last nucleotide of exon 14, so the ssODN contains 15 nucleotides of intron. In ssODN 2, one guanine in the PAM-sequence (NGG) was changed to another nucleotide to become NCG, to avoid the ssODN sequence being cleaved by Cas9 before or after HDR. ssODN 1 was designed to be in sense orientation since crRNA 1 was in forward orientation, and ssODN 2 designed to be antisense since crRNA 2 was reverse. A silent mutation and restriction site were not designed in order to avoid additional and possibly deleterious mutations in the disease model. Repair templates shown in Table 7 were ordered from IDT.

ssODN	Sequence	Orientation
ssoDN 1	5’-CTC CAC ATC ATG TTC CAG ACC GAG GCT GGG GAG CTG CAG TTT GCC CAC AC**A** GTG AGG CCC GCA CAG CCT CCT GCC CGC GTG CCC TCG CCC GCA GCC TCT GC-3’	Sense
ssoDN 2	5’-GCA GAG GCT GCG GGC GAG GGC ACG CGG GCA GGA GGC TGT GCG GGC CTC AC**T** GTG TGA GCA AAC TGC AGC TCC CCA GCC TCG GTC TGG AAC ATG ATG TGG AG-3’	Antisense

Table 7. ssODN repair template sequences used in iPSC CRISPR-Cas9 RNP complex electroporations. ssODN 1 is used in conjunction with crRNA 1 and ssODN 2 with crRNA 2. The knock-in mutations are indicated in bold.

2.7 Creating the genome edited iPSC lines

2.7.1 Electroporation

Cells were passaged with 0.75X TrypLESelect cell dissociation enzyme (Thermo Scientific #12563029) in PBS (Lonza) two days before electroporation to accustom the proliferative cells to a single-cell state. Confluence on the day of electroporation was 70% in four 35mm cell culture plates (Thermo Scientific). This translates to about two million cells per electroporation as counted by Countess Automated Cell Counter (Thermo Scientific). Three hours before electroporation, fresh cell culture medium containing RevitaCell supplement was added to enhance survival throughout the process.

First, the AltR® CRISPR-Cas9 crRNA and 20 nmol Alt-R® CRISPR-Cas9 tracrRNA, ATTO™ 550 (IDT, #1075925) were individually resuspended in a nuclease-free buffer (IDT) to a final concentration of 200 µM in a dark Eppendorf tube. The crRNA/tracrRNA duplex was formed by resuspending the crRNA and tracrRNA together to a final concentration of 100 µM, incubated at 95 °C for 5 minutes and allowed to cool down to room temperature. Second, the ribonucleoprotein (RNP) complex was formed by combining in a 1.5 mL Eppendorf tube 3 µL of the crRNA/tracrRNA duplex, 3 µL of 10 µG recombinant Alt-R® S.p. HiFi Cas9 Nuclease V3 (IDT, #1081061), 2 µg ssODN (IDT), and 3 µL of Neon electroporation R buffer (Invitrogen, #MPK10096). The complex was incubated at room temperature for 20 minutes.

iPSC were washed once with 1 mL 5% FBS-PBS and dissociated using 1 mL 0.75X TrypLE Select (Thermo Scientific) at 37 °C for 3 minutes. Cells were pipetted carefully into a 50 mL polystyrene Falcon tube containing 4 mL 5% FBS-PBS and centrifuged at 250 g for 5 minutes. The pellet was

resuspended into 1 mL 5% FBS-PBS. The number of live cells was counted with Countess Automated Cell Counter (Thermo Scientific), and two million cells were taken into a 1.5 mL Eppendorf tube, centrifuged at 200 g for 3 minutes and resuspended in 300 µL Neon electroporation R buffer (Invitrogen). Then, 9 µL RNP complex, 6 µL electroporation enhancer (IDT, #1075915), 6 µL ssODN and 6 µL pCXLE-mp53DD plasmid (AddGene, #41859) were mixed with the cells. With the Neon pipette, 100 µL of the mixture was taken into in a cuvette with 3mL electroporation buffer E2 (Invitrogen) and electroporated with two pulses of 1100V for 20 ms with the Neon Electroporator (Invitrogen).

Electroporated samples were plated immediately onto pre-warmed 35 mm cell culture plates as described earlier, one supplemented with Alt-R HDR enhancer (IDT, #1075915), one with 20 µM, and one without HDR Enhancer. Cell culture medium was replaced 24 hours from the electroporation and imaged with a fluorescence microscope for the presence of ATTO™ 550 labelled cells to determine transfection efficiency in attached cells. Each pool showing a positive fluorescence result was passaged 1:3, one for T7 assay and sequencing, one for subcloning, and one for cryopreservation as backup. For freezing the electroporated pools, cells were detached and resuspended in 1 mL of freezing medium containing DMEM/F-12 (Gibco) with 20% FBS (Gibco, #10270106) and 10% DMSO (Merck, #102952). They were incubated at -80 °C overnight and transferred to -140 °C for long-term storage.

2.7.2 Subcloning

Electroporated iPSC pools were expanded to 80% confluency before subcloning. Cell culture medium was aspirated, washed once with 1 mL PBS (Lonza), 1 mL of Accutase (Thermo Scientific) was added to each dish, and incubated at +37 °C for 6 minutes to dissociate cells. Cells were collected into a 50 mL Falcon tube containing 4 mL TeSR™-E8™ media supplemented with 1X CloneR (STEMCELL Technologies, #05888), centrifuged at 200 g for 5 minutes, and resuspended in 1 mL TeSR™-E8™ media supplemented with 1X CloneR. The number of live cells was counted with Countess Automated Cell Counter (Thermo Scientific), and the volume of triturated medium containing 500-1000 cells were plated into 10 cm tissue-culture dishes coated with Vitronectin XF at a concentration of 5 µg/cm^2 containing 10 mL with TeSR™-E8™ media supplemented with 1X CloneR. Dishes were gently rocked to disperse cells evenly and placed in an incubator undisturbed for 48 hours (Bruntraeger et al., 2019).

48 hours after subcloning, cell culture medium was replaced with 10 mL of Essential 8 Flex medium supplemented with E8 Flex medium supplement (Gibco, #A2858501) and changed every 48 hours for 8 days until colony picking.

2.7.3 Colony picking

Single-cell derived iPSC colonies of approximately 1-2 mm diameter, uniform morphology and clear edges (indicative of colonies originating from one cell) were manually picked from the cell culture dish 8 days after low-density seeding of the cells in subcloning. First, medium was aspirated, the dish washed once with 5 mL PBS (Lonza), and 5 mL Gentle Cell Dissociation Reagent (STEMCELL Technologies, #07174) added and incubated at 37 °C for 2 min 30 s. Dissociation reagent was aspirated carefully without disturbing the colonies, and 10 mL of TeSR™-E8™ medium added. Using a dissection microscope, individual colonies were scraped and collected in a P20 pipette set at 10 µL and transferred into a U-bottom 96-well plate (Corning, #0058) containing 100 µL TeSR™-E8™ media supplemented with RevitaCell Supplement (Thermo Scientific) per well. A fresh tip was used for each colony to prevent cross-contamination. Cells were triturated 6 times and 50 µL transferred to one 96-well cell culture plate (Thermo Scientific) coated with Matrigel (Corning), and 50 µL to another 96-well cell culture plate coated with Vitronectin XF (STEMCELL Technologies) to create duplicate plates of colonies.

2.7.4 Cryopreservation of clonal cell lines

Five days post-picking at 60-90% confluency, the colonies from the Matrigel-coated plates were passaged 1:2 into two Matrigel-coated 96-well plates (Thermo Scientific). Once 60-90% confluent, spent medium was removed, colonies washed once with 100 µL PBS (Lonza), 25 µL of pre-warmed Accutase (Life Technologies) added per well and incubated at 37°C for 6 minutes. 75 µL of Knock-Out Serum Replacement (Life Technologies, #10828128) was added per well to quench the dissociation reaction. The colonies were triturated 6 times, and 100 µL of Knock-Out Serum Replacement containing 20% DMSO added. The solution was triturated 4 times, and the contents transferred into a 96-well 0.5 mL 2D matrix plate (Thermo Scientific, #3725). 100 µL of sterile EmbryoMax® light mineral oil (Millipore, #ES-005-C) was added on top of the freezing medium to prevent evaporation. The matrix plates were incubated at -80°C overnight and transferred to liquid nitrogen for long term storage.

2.7.5 Thawing

Individual vials from the matrix plates were removed from liquid nitrogen and thawed rapidly at room temperature. Mineral oil was removed with a pipette, and cells suspended in freeze medium were transferred into 1 mL of pre-warmed TeSR-E8 medium and centrifuged at 300 x g for 3 minutes. Pellets were resuspended in 400 µL TeSR-E8 medium supplemented with CloneR, transferred into individual 48-well plate wells coated with Vitronectin XF at 2 µg/cm^2 and cultured overnight at 37°C.

After a 400 µL medium change the following day, cells were expanded to 35 mm cell culture plates for secondary genotyping and further analyses.

2.8 Sanger sequencing

Primary genotyping of all expanded clonal cell lines was done using the SeqLab sequencing service at the Institute for Molecular Medicine Finland (FIMM). The service included product-cleanup, sequencing with ABI3730xl DNA Analyzer, and base-calling with Sequencing Analysis 5.2. Chromatogram data were interpreted with the Sequencher 5.2.4 software (Gene Codes Corporation), and FASTA-file data compared to a reference genome with NCBI BLAST search tool to evaluate effects of genome editing.

First, genomic DNA was PCR-amplified with 12.5 µL of AmpliTaq Gold 360 Master Mix (Thermo Scientific), 0.5 µL of 10 µm forward and reverse primers (Table 8), 100 µg DNA template filled to a 25 µL reaction with nuclease-free H_2O (Table 4). The production of a single 206 bp band was checked on 2% agarose gel run at 140V for 45 minutes and imaging with Bio-Rad ChemiDoc XRS+. The PCR-products were purified for sequencing by adding 2 µL Exo-SAP-IT reagent (Thermo Scientific, #78200.200.UL) into 5 µL PCR-product, incubating at 37°C for 15 minutes to activate the enzyme, and incubating at 80°C for 15 minutes to inactivate the enzyme. The Ready-To-Run sequencing reaction included 5.2 µL MilliQ water, 1.9 µL 5X BigDye dilution buffer (Life Technologies, #4336697), 0.25 µL BigDye RR Mix reaction enzyme (Life Technologies), 0.65 µL 5 µm sequencing primer (Table 9), and 2 µL of purified PCR-product as the template. The samples were run in a thermocycler program (Table 8) on Bio-Rad S1000™ Thermal Cycler before delivery to FIMM.

Number of cycles	**Temperature (°C)**	**Time**
1 x	96	1 min
24 x	96 59 60	10 s 5 s 4 min
	10	Hold until delivery to SeqLab

Table 8. BigDye 3.1 sequencing protocol thermocycling

Primer	**Sequence (5’ → 3’)**
SARS2 exon 14 Forward	CCCATCCCTGATCCGCTC
SARS2 exon 14 Reverse	CAAGGTGGGTGGGTCTAGG

Table 9. Sequencing primers for *SARS2* exon 14

2.9 TOPO-Cloning

Clone KI 1 from gRNA 2 was TOPO-cloned to determine the sequence of individual strands at the mutation due to unclear genotype at the mutation site. First, the high-fidelity PCR reaction was set up with 2 µL DNA template, 0.3 µL of 10 µM forward and reverse primers, 5 µL KAPA HiFi HotStart ReadyMix (2X) (KapaBiosystems, #KR0370), and filled to 10 µL with nuclease-free H_2O. Thermocycling was according to the program in Table 10.

Number of cycles	**Temperature (°C)**	**Time**
1 x	95	3 min
35 x	98 65 72	20 s 15 s 30 s
1 x	72	5 min

Table 10. KAPA HiFi HotStart ReadyMix PCR cycling program

The reaction is set up according to the Zero Blunt® TOPO® PCR Cloning Kit instructions (Thermo Scientific, #450245). 4 µL of PCR product, 1 µL of salt solution, and 1 µL of pCR™-Blunt-TOPO vector were combined. The mixture was incubated for 5 minutes and RT and then transferred to ice.

The recombinant vector was then transformed into competent *E. coli.* 2 µL of recombinant vector was added to 50 µL of DH5α competent *E. coli* cells and incubated on ice for 30 minutes. The transformation was done with giving the mixture a heat-shock at 42°C for 30 seconds, then immediately transferred to ice for 2 minutes. 250 µL of RT S.O.C. medium was added to the mixture and incubated horizontally at 37°C for 1 hour on a shaker at 200 rpm. 250 µL of the transformation mixture was spread on a prewarmed LB plates containing 50 µg/ml Kanamycin and incubated at 37°C overnight. 18 colonies were culture overnight in LB medium containing 50 µg/ml Kanamycin. Plasmid DNA for sequencing was isolated using GeneJET Plasmid Miniprep Kit (Thermo Scientific, #00827352).

The Ready-To-Run sequencing reaction for plasmid DNA included 6.35 µL nuclease-free H_2O, 1.5 µL 5X BigDye dilution buffer (Life Technologies), 1 µL 2.5X BigDye RR mix (Life Technologies), 0.65 µL of 10 µM primer and 0.5 µL plasmid DNA. Thermocycling was according to the program in Table 11.

Number of cycles	Temperature (°C)	Time
1 x	96	1 min
24 x	96 50 60	10 s 5 s 4 min
	10	Hold until delivery to SeqLab

Table 11. BigDye 3.1 sequencing protocol thermocycling for plasmid DNA

2.10 Analysis and sequencing of off-target genome editing

Off-target genome editing for crRNA 2 was evaluated with CRISPOR Batch Gene Targeting Assistant 4.97. The program aligned the crRNA of interest to a reference genome and reported exonic, intronic, and intergenic regions with a similar sequence as targeted by the crRNA, and PCR primers for these regions. Off-target loci were ranked according to cutting frequency determination (CFD) scores. CFD measures the specificity of the crRNA to the target with scores ranging from 0-100 with 100 suggesting the strongest specificity to the crRNA and 0 suggesting a weak interaction due to mismatches between the crRNA and the target. CFD scores are further transformed so that weak interactions <20 are reduced to 0 and strong scores >80 are transformed to 100, after which they are turned into weighted scores ranging from 0-1. The CFD is based on a large dataset of cleavages in cells by infection with a lentiviral library containing thousands of guides targeting the CD33 gene for all possible nucleotide mismatches and 1 bp INDELs in all PAMs (Doench et al., 2016; Primers for regions with CFD the highest CFD scores, in this case above 0.2 and exonic regions with CFD scores above 0.14 were simulated using UCSC *in silico* PCR (Genome assembly: Dec. 2013 GRCh38/hg38). The primers that produced a ~600 bp amplicon at the correct locus were ordered from Merck, and regions sequenced using the FIMM service described in section 2.7. Sequences were compared to a reference genome using NCBI Basic Local Alignment Tool, Nucleotide Collection.

2.11 Analyses of HDR efficiency

The presence of INDELs and efficiency of HDR-mediated knock-ins in the electroporated cell pools was done using an online bioinformatics tool ICE CRISPR analysis tool version 2 from Synthego (https://ice.synthego.com/#/ , Hsiau et al., 2018). This tool was used to analyse Sanger sequencing files of the electroporated cell pools to find INDELs resulting from the use of each of the crRNAs. The inputs include the crRNA sequence excluding the PAM, the donor sequence with homology arms, and Sanger sequencing files of a reference wild type and the test sample. The output parameters are the PAM sequence, INDEL percentage, KO-score, KI-score, model fit (R^2). The INDEL percentage is the percentage of cells in the pool with a non-wild type sequence, showing overall editing efficiency. The KO-score (knockout score) represents the proportion of cells in the sequence

that have frameshifts or >21 bp INDEL. The KI-score (knock-in score) is the proportion of cells with the desired knock-in mutation. Potential editing outcomes are fitted to observed data by linear regression, and model fit shows reliability of the ICE output parameters. Whereas the T7 test determined genome editing efficiency by the percentage of INDELs present in electroporated cell populations, the ICE CRISPR analysis tool could determine any differences in the rates of HDR and INDELs between the two crRNAs used in iPSC electroporation and thereby the efficiencies of the differently designed crRNAs.

2.12 Cell collection for RT-qPCR and Western blotting

80% confluent iPSC and fibroblasts on 35 mm cell culture dishes were first washed with 1 ml of RT PBS to remove cell debris. 1 ml of +4 ºC PBS was added, and cells were detached carefully with a cell scraper (Thermo Scientific) and placed on ice. The cell suspensions were centrifuged at 1000 g for 10 minutes in +4 ºC, supernatant aspirated, and pellets stored -80 ºC.

2.13 RT-qPCR

RNA was extracted from iPSC and fibroblasts using the Nucleospin RNA kit (Macherey-Nagel, #740955.250), and 500 ng cDNA synthesised with the Maxima First Strand cDNA Synthesis Kit (Thermo Scientific, #K1641). 20 µL PCR reactions contained 10 µL SybrGreen (Thermo Scientific, #K0222), 3 µL MQ water, 1 µL of 10 µM forward primer, 1 µL of 10 µM reverse primer (Table 12), and 5 µL (5 ng) of template. Three technical replicates of each sample and negative (water) template controls were included. Human fibroblasts, a differentiated cell type, were used as a control against pluripotency. The transcription factors Sox2, Oct4 and Nanog were chosen as pluripotency markers due to their role in maintaining pluripotency, and their validity and widespread use in literature.

PCR amplifications were carried out with a CFX Touch Real Time PCR Detection system (BioRad) for 7 minutes at 95 °C followed by 40 cycles of 95 °C for 10 seconds, 60 °C for 30 seconds and 72 °C for 20 seconds. The program ended with a cooling step of 10 seconds at 40 °C. A standard curve for each primer pair with mixed serial cDNA dilutions of 10 ng/ µL to 0.156 ng/ µL was included as an efficiency control. Cycle threshold (Ct)- values were automatically calculated by Bio-Rad CFX Maestro (v 4.1.2433.1219) for each sample and mean Ct values were calculated for three technical replicates of each cell line.

mRNA expression values were normalised to the GAPDH internal control to correct for any differences in RNA quantity, quality, and cDNA synthesis efficiency. Changes in relative mRNA expression were calculated using the ΔΔCt ($2^{-\Delta\Delta CT}$) method (Livak & Schmittgen 2001). Graphs were plotted using GraphPad Prism 8.4.1 (GraphPad Software) with data presented as mean ± standard

deviation (SD). The significance of differences between means was tested using a one-way ANOVA and Tukey post-hoc test for multiple comparisons, with $p<0.05$ being significant.

Transcript	**Forward primer sequence (5' → 3')**	**Reverse primer sequence (5' → 3')**
Sox2	GCCCTGCAGTACAACTCCAT	TGCCCTGCTGCGAGTAGGA
Nanog	CTCAGCCTCCAGCAGATGC	TAGATTTCATTCTCTGGTTCTGG
Oct4	TTGGGCTCGAGAAGGATGTG	TCCTCTCGTTGTGCATAGTCG
Gapdh	cgctctctgctcctcctgtt	ccatggtgtctgagcgatgt
Sars2	TTGGTCCTACGTTCCTCCCA	GAGAAAACTGGCTTGTCTCCG

Table 12. Primers used for RT-qPCR of pluripotency markers and Sars2 in iPSC and fibroblasts

2.14 Western blot

2.14.1 Protein extraction

iPSC and fibroblast cell pellets were lysed with 40 µL RIPA-buffer (Cell Signaling, #9806) supplemented with 10 µL/ml of Halt™ protease inhibitor (Thermo Scientific, #10137963). Cell lysates were incubated on ice for 10 minutes, centrifuged at 12000 x g for 10 minutes at 4°C, and supernatant collected. Protein concentration was quantified with the Pierce™ BCA Protein Assay Kit (Thermo Scientific, # 23225), using Bovine Serum Albumin (Biowest, #P6154) samples as a standard, with concentrations ranging from 0.25 mg/ml to 10 mg/ml.

2.14.2 SDS-PAGE electrophoresis and blotting

10 µg of protein sample was added to 1:3 of the total volume 4 x Laemmli loading buffer (BioRad, # 1610747) containing 10 % β-mercaptoethanol (BioRad, #1610710) filled to 30 µL with MQ water. Samples were loaded on a 10 % stain free 31 gel Mini-PROTEAN-TGX (BioRad # 456-8033) gel, with 1 X running buffer (0.3 % Tris, 1.44 % Glycine, 0.06 % SDS) (Thermo Scientific) with Precision Plus Protein™ Dual Color Standards (BioRad, #1610374) as the size ladder. Samples were first run for 10 minutes at 110 V and then at 125 V for 1 hour 30 minutes.

Proteins from SDS-PAGE were transferred onto a 0.45 µM nitrocellulose membrane (Bio Rad, #1620234) by semi-dry blotting using the Trans-Blot Turbo transfer system (Bio Rad). The membranes were blocked in 5 % milk (Valio) in TBS–Tween 20 (0.1 %). Immunoblotting was done with the indicated primary antibodies (Table 13) and the corresponding secondary antibodies anti-rabbit (Jackson Immunoresearch, #111-335-14) and anti-mouse (Jackson Immunoresearch, #111-035-146).

Epitope	Marker of	Host species	Cat. Number	Supplier	Size (kDA)	Dilution (in 5% BSA/TBS-T)
SARS2	Mitochondrial seryl-tRNA synthetase	Rabbit	PA5-31473	Thermo Scientific	58	1:2500
β-actin	Cytoskeleton	Mouse	sc-47778	Santa-Cruz	45	1:1000

Table 13. Primary antibodies used in Western blotting of iPSC

Signal detection was done using enhanced chemiluminescent substrate enhanced luminal & peroxide solution (BioRad) and Molecular Imager ChemiDoc XRS+ with ImageLab software (BioRad). ImageLab Software 6.0.1 (BioRad) was used to quantify the bands. Relative band intensities were normalized to loading control β-actin.

3. Results

The project workflow (Figure 6) involved creating genome-edited iPSC lines to model a patient-specific splicing variant c.1347 G>A in *SARS2* causing early-onset progressive spastic paresis. First, non-allele specific sgRNAs were tested for gene modification efficiency in HEK293 cells. Based on their cleavage ability, two crRNAs for CRISPR-Cas9 RNP complexes were chosen and corresponding ssODNs designed to make a knock-in mutation in a healthy iPSC line HEL 24.3. Cell lines were created from electroporated cell populations with the best gene modification efficiency and estimated HDR rates. The editing and HDR efficiencies of the two crRNAs were compared between cell populations and cell lines, and potential off-target Cas9 cleavage loci screened in the two genome-edited disease model cell lines to test the precision and quality of the genome editing. Pluripotency markers were measured for iPSC quality control. SARS2 expression at the gene and protein levels were measured to compare the genome-edited mutant to the patient and wild type iPSC lines, and thereby the ability of the genome-edited lines to recapitulate the effects of the patient mutation. The individual results will be discussed next.

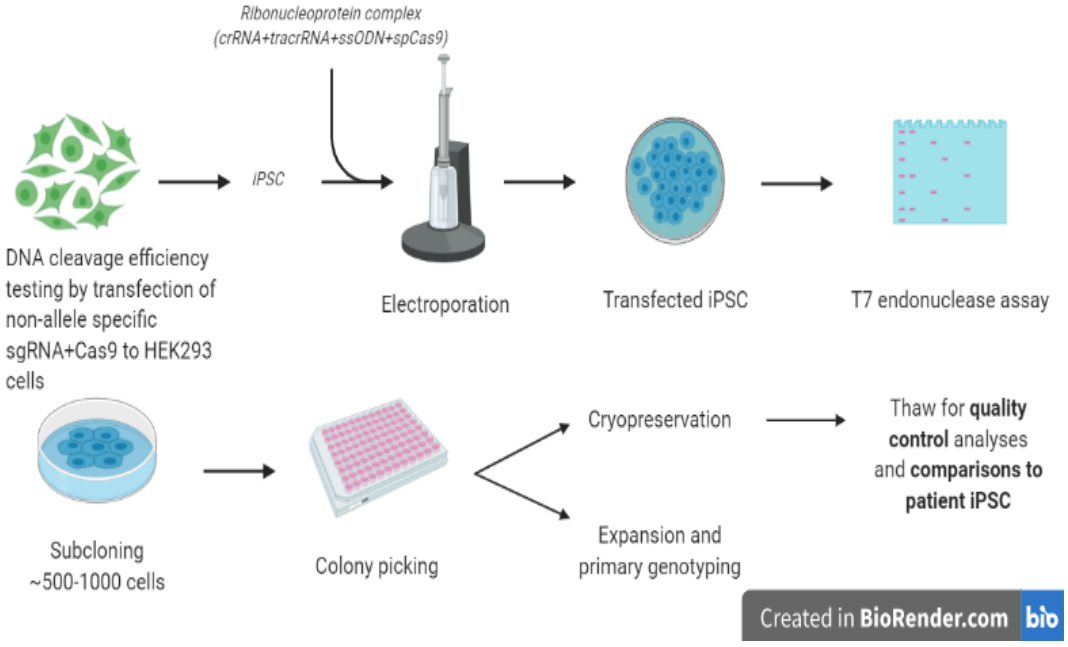


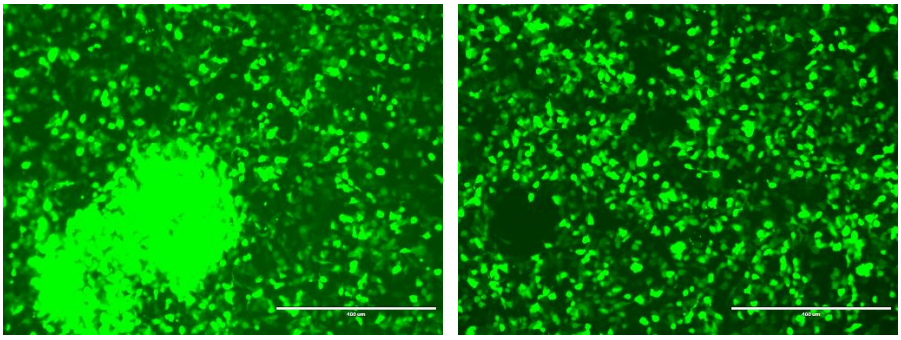
Figure 6. Workflow of the creation of genome-edited cell lines in this thesis. The method of clonal cell line creation used in this project is adapted from the method reported by Bruntraeger, Byrne, Long and Bassett (2019). The method is based on the delivery of a ribonucleoprotein complex (RNP complex) and single-stranded donor oligonucleotide (ssODN) repair template delivery into iPSC by electroporation to create clonal cell lines. The electroporated cell pools were screened for genome editing efficiency using a T7 Endonuclease Assay (NE Biolabs) that recognises and cleaves the insertions/deletions (INDELs) in the DNA made by Cas9. Approximately 500-1000 cells from cell

pools with the highest genome editing efficiency (DNA cleavage) were seeded on a 10 cm tissue culture dish at a low density to create clonal cell populations. Colonies were then manually picked and expanded, and duplicate plates of clones were made for cryopreservation and primary genotyping.The clones with the desired edit for the *SARS2* c.1347 G>A variant were then thawed, expanded, re-genotyped and quality controlled.

3.1 Two out of four crRNA were chosen for iPSC electroporation based on gene modification efficiency in HEK293 cells

Four non-allele specific sgRNAs were designed to test transfection efficiency in HEK293 cells with the jetPRIME transfection protocol. Their gene modification efficiencies were tested using the T7-endonuclease assay. crRNA 1 and 4 had gene modification efficiencies of approximately 1.17% and 15.4% as estimated by ImageLab software band intensity analysis respectively. crRNA 2 and 3 had gene modification efficiencies of less than 1%. crRNA 1 and 4 had the highest GFP-positivity suggesting transfection efficiency and successful entry of gRNA into nuclei. crRNA 1 and 4 also produced the highest T7 endonuclease assay gene modification efficiency scores, suggesting Cas9 cleavage (Figures 7 A 1 and 2.). crRNA 1 and 4 were chosen for electroporation into iPSC based on their higher efficiencies of DNA cleavage in HEK293 cells as compared to minimal DNA cleavage in the other two crRNAs (efficiencies summarised in Figure 7 B and Table 14). In iPSC electroporation, crRNA1 is referred to as crRNA1 and crRNA4 as crRNA2.

A 1. **A 2.**



B

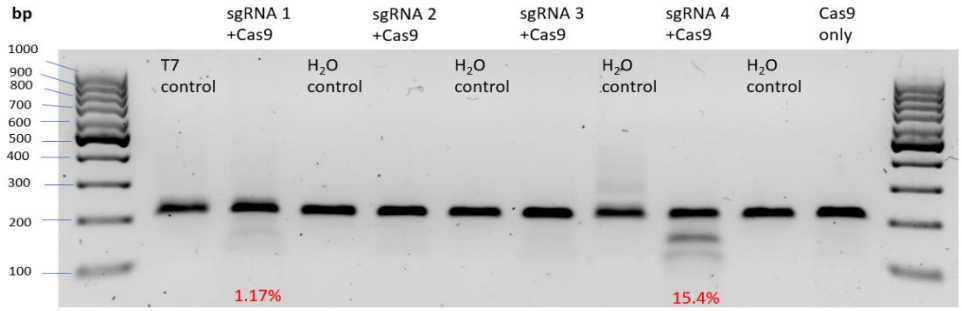


Figure 7. T7 results for sgRNA 1-4 in HEK293 cells. **(A)** Successful transfection with sgRNA 1 **(1)** and 4 **(2)** respectively showing GFP-intensity as detected by Life Technologies EVOS® FL-electro fluorescence microscope 24 hours after transfection. The scale of the ruler is 400 µm. (**B)** Gene modification efficiency is shown by the intensity of cleaved bands and the percentage of DNA cleaved on a 2% agarose gel, as estimated by ImageLab software band intensity analysis. The two sgRNAs with the highest T7 cleavage results (1 and 4) were chosen for iPSC electroporation. The lanes in between the tested sgRNAs 1, 2 , 3 and 4 represent negative H_2O controls.

sgRNA	Gene modification efficiency	Sequence
1	1.17%	GGAGCTGCAGTTTGCCCACA
2	0.73%	TGCAGTTTGCCCACACGgtg
3	0.71%	gctgtgcgggcctcacCGTG
4	15.4%	ctgtgcgggcctcacCGTGT

Table 14: Summary of T7 gene modification efficiencies of non-allele specific sgRNAs transfected to HEK293 cells

3.2 iPSC electroporation

3.2.1 Gene modification efficiencies in the electroporated cell populations

Three separate electroporations were done for each of the two crRNAs to ensure the development of least one viable cell pool for each crRNA (crRNA 1 and 2). One pool of each was supplemented with 10 µM Alt-R HDR enhancer (IDT, #1075915), one with 20 µM, and one without the supplement, to enhance HDR while also enhancing cell survival through electroporation. Since the HDR enhancer is diluted in 3 mM DMSO and is toxic to cells, concentration titration of the HDR enhancer was done to maximise chances of survival in the electroporated cell pools, and fresh cell culture medium without HDR enhancer was changed 24 hours post-electroporation. The HDR enhancer also acted as a NHEJ inhibitor to boost HDR. The presence or concentration of HDR enhancer did not affect cell survival, as observed by confluency 24 hours post electroporation. Additionally, a p53 dominant negative plasmid was electroporated along with the CRISPR-Cas9 RNP complex to increase the survival of iPSC with DSBs and thereby increase genome editing efficiency.

The three electroporated pools for each of the two crRNAs were each passaged 1:3 following expansion to 80% confluence post-electroporation. Genomic DNA from cell pools was PCR-amplified, and the T7-endonuclease assay tested gene modification efficiency. The electroporated pools with the highest gene modification efficiencies chosen for subcloning were 8.45% and 13% for crRNA 1 and 2 respectively (Table 15). These cell pools with the highest efficiencies chosen for cloning and subcloning were the ones supplemented with HDR enhancer (Figure 8).

A. B.

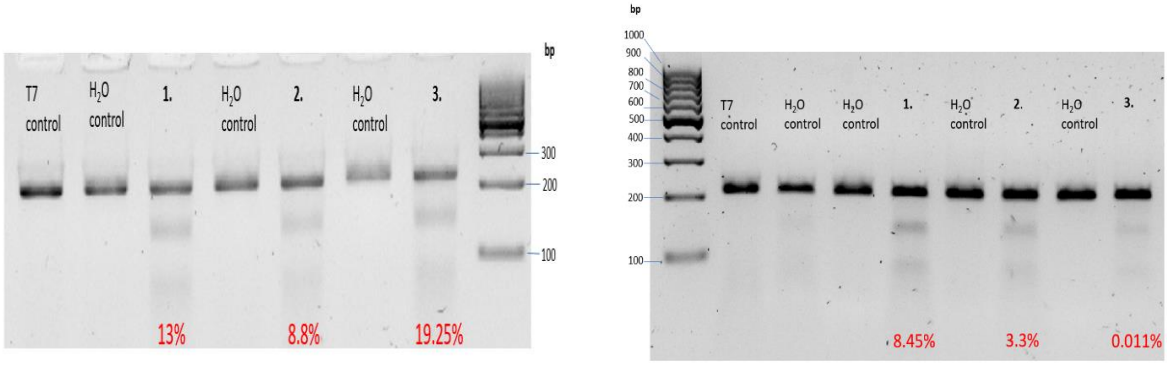


Figure 8. T7 results in electroporated iPSC pools. (A) Gene modification efficiency is shown by the intensity of cleaved bands and the percentage of DNA cleaved on a 2% agarose gel. 1. refers to iPSC pool electroporated with crRNA1 supplemented with 10 µM Alt-R HDR-enhancer (IDT), 2.

refers to one supplemented with 20 µM HDR-enhancer, and 3. the cell pool not supplemented with HDR-enhancer. **(B)** Results interpreted the same manner for crRNA2.

crRNA	**T7 gene modification efficiency**	**Sequence (5' → 3')**	**Orientation**	**PAM**
1 +10 µM HDR enhancer	8.45%	GGAGCTGCAGTTTGCCCACA	Forward	cgg
2 +10 µM HDR enhancer	13%	ctgtgcgggcctcacCGTGT	Reverse	ggg

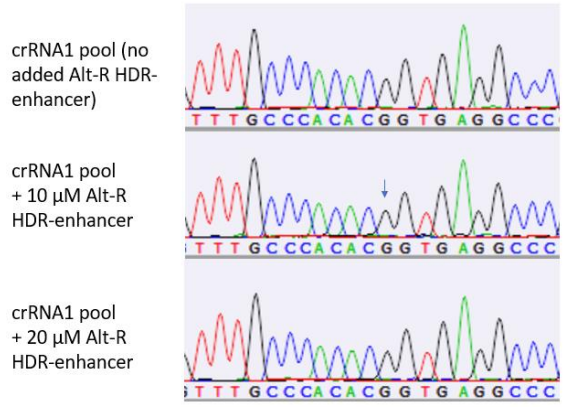
Table 15. Summary of electroporated pools chosen for subcloning. Gene modification efficiencies in electroporated iPSC pools chosen for subcloning based on T7-assay. crRNA 1 supplemented with 10 µM HDR enhancer and crRNA 2 pool supplemented with 10 µM HDR enhancer were chosen for subcloning considering their T7 gene modification scores. They were chosen for their higher gene modification efficiencies and the highest probability of containing cells with the knock-in mutation.

3.2.2 ssODN integration and HDR efficiency in the electroporated pools

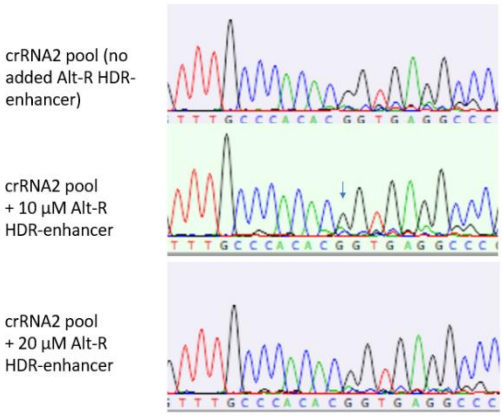
Overall HDR efficiency in the three electroporated cell pools for each crRNA was estimated using a bioinformatic tool ICE CRISPR analysis tool version 2 from Synthego (Hsiau et al., 2018). Increasing the concentration of Alt-R HDR enhancer did not produce higher HDR efficiency. Overall, crRNA 2 yielded higher percentages of indels and ssODN integration efficiencies than crRNA 1. Especially the crRNA 2 pool that was supplemented with 10 µM of the commercial Alt-R HDR enhancer (IDT) had the highest indel percentage (27%) and KI-score (10% of cells with the desired KI). The highest concentration of HDR enhancer used (20 µM) did not yield better results for crRNA 2 but improved the KI-score slightly for crRNA 1 raising KI-score to 1, whereas it was 0 with lower concentrations (summarised in Figure 9). crRNA 2 was designed as a reverse sequence, with the cut site to the mutation being just 1 nt.

A.

1.



2.



B.

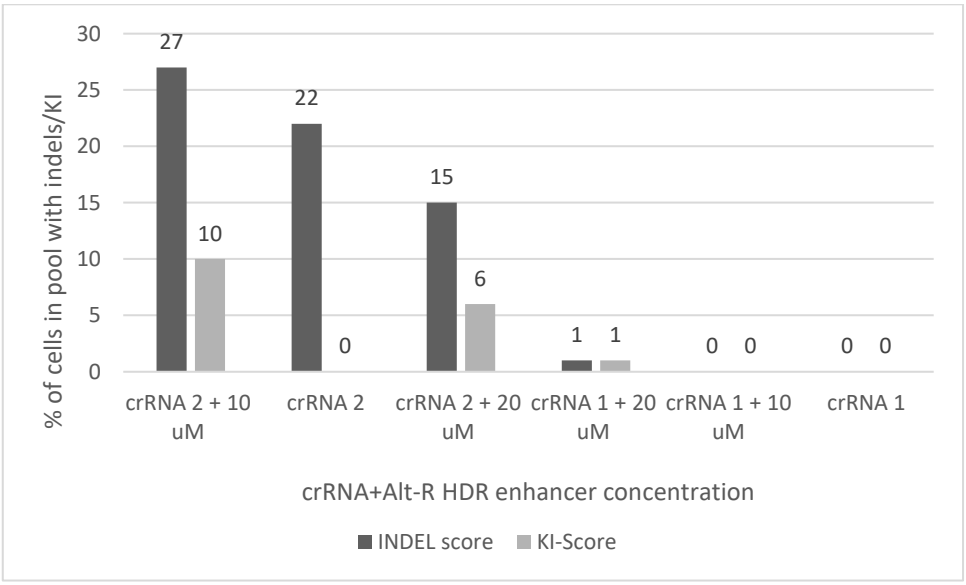


Figure 9. Sequencing of electroporated iPSC pools and efficiency of NHEJ and HDR. (A) Chromatograms of the electroporated pools from **(1)** crRNA1 and **(2)** crRNA2 used as input for the ICE CRISPR analysis tool. The blue arrows show the intended mutation site. **(B)** ICE CRISPR analysis tool output results of INDEL percentage and KI-score shows estimates of INDEL percentage and proportion, and ssODN integration in the electroporated iPSC pools. Electroporated pool crRNA2 supplemented with 10 µM Alt-R HDR enhancer (IDT) showed the highest INDEL percentage (27%) and knock-in score (10%). crRNA 1 pool with 10 µM Alt-R HDR enhancer had an indel score of 1 and KI score of 1, while the other two crRNA 1 pools had 0 for both scores.

3.3 Clonal cell line creation

3.3.1 Rates of NHEJ and HDR in the clonal cell lines

For the cloning step of creating the clonal cell lines, 1-2 mm diameter colonies of the appropriate morphology (Figure 10) were picked eight days post subcloning. 72 colonies were picked for crRNA1, and 96 colonies for crRNA2. Out of these, 64 colonies of crRNA1 and 85 colonies of crRNA2 survived for expansion and cryopreservation. Genomic DNA from expanded cells was PCR-amplified and sequenced. 48.44% of crRNA1 clones and 67.06% of crRNA2 clonal cell lines contained DNA cleavage in SARS2 exon 14. 1.56% and 1.18% of clones for crRNA 1 and 2 respectively were heterozygotes for the mutation (summarised in Table 16).

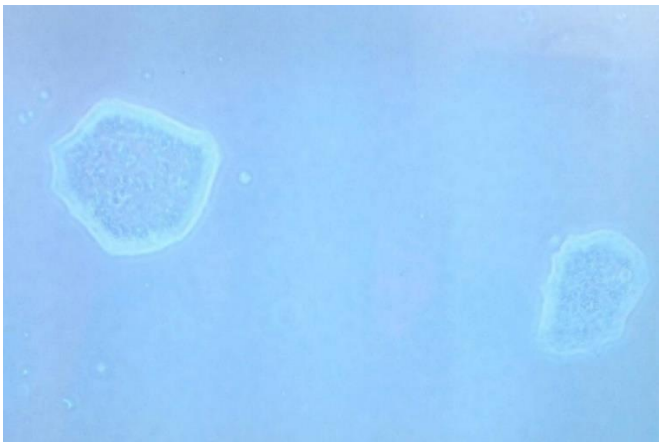


Figure 10. iPSC colonies 8 days post subcloning. Light microscope image showing a colony on the left with ideal morphology for colony picking to create clonal cell lines (clear, round edges, uniform in shape and texture, 1-2 mm in diameter) suggesting origin from a single cell. The colony on the right is uneven in shape, and therefore not suitable for colony picking.

crRNA	Wild type (no editing)	NHEJ	NHEJ clones with PAM change	Homozygote HDR	Heterozygotes
1	32/64 (50%)	31/64 (48.44%)	0	0	1/64 (1.56%)
2	25/81(29.41%)	57/85 (67.06%)	0	2/85(2.35 %)	1/85 (1.18%)

Table 16. Genome editing- and HDR efficiencies in CRISPR-Cas9 edited clonal iPSC lines created from crRNA 1 and 2. Presented as percentages of total number of clonal cell lines expanded from picked colonies. 'Wild type' cell lines contained no editing despite undergoing electroporation, NHEJ clones refer to clones that underwent double-strand break correction by NHEJ, NHEJ clones with PAM change to clones that underwent NHEJ along with the PAM change correspondent to the crRNA, HDR clones refer to clones that underwent ssODN integration via the HDR pathway of

double-stranded break repair, and heterozygotes refer to clones containing the c.1347 G>A mutation in one allele only.

3.3.2 Two *SARS2* c.1347 G>A knock-in clonal iPSC lines were created

2 out of 85 clonal cell lines created with crRNA 2 contained the c.1347 G>A mutation indicating both cleavage by Cas9 and ssODN integration through the HDR pathway (summarised in Table 15). These two clonal iPSC lines are named KI 1 and KI 2. The KI 1 and KI 2 cell lines contain the same disease mutation as the patient cell line HEL 204.6, as shown by their DNA sequences. TOPO-cloning of the DNA fragments and re-sequencing of an expanded cell population confirmed homozygous knock-in mutations. These two knock-in cell lines were used for subsequent quality control analyses, alongside two wild type (WT) lines that underwent electroporation, the parental cell line HEL 24.3, patient iPSC line HEL 204.6, and one cell line that underwent DSB correction by NHEJ in one allele resulting in a frameshift (Figure 11).

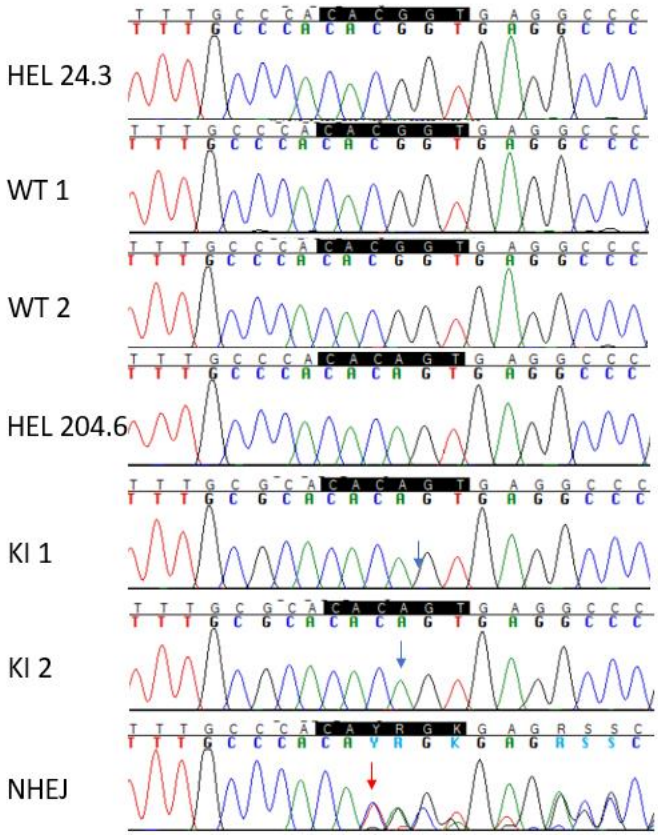
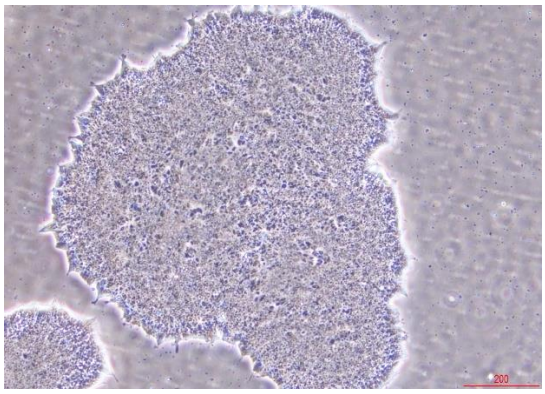


Figure 11. *SARS2* Sanger sequencing results. Comparison of chromatograms between the parental cell line HEL 24.3, wild type (WT) (electroporated but not genome edited), patient cell line HEL 204.6, the two edited iPSC lines demonstrates the successful editing of c.1347 G>A mutation by HDR in the edited KI 1 and 2 lines, indicated by blue arrows. DSB correction by NHEJ is indicated by the red arrow.

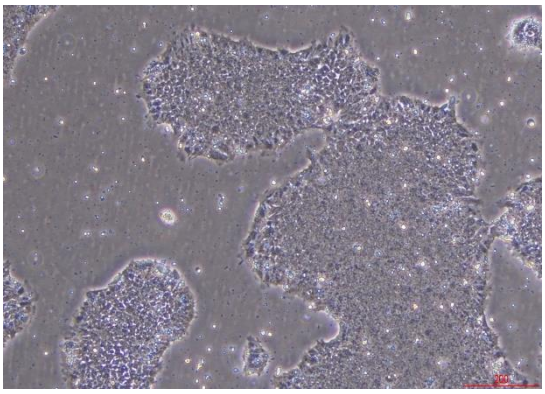
3.3.3 iPSC morphology of wild type, KI and patient cells

The two KI, patient, and wild type iPSC grew at a similar rate and had typical iPSC morphology. The colony in Figure 12. A is an example of an ideal iPSC morphology: round colonies, clear borders and even surface. Further analyses of cell morphology could be important in explaining disease mechanisms modelled by the genome-edited iPSC lines.

A



B



C

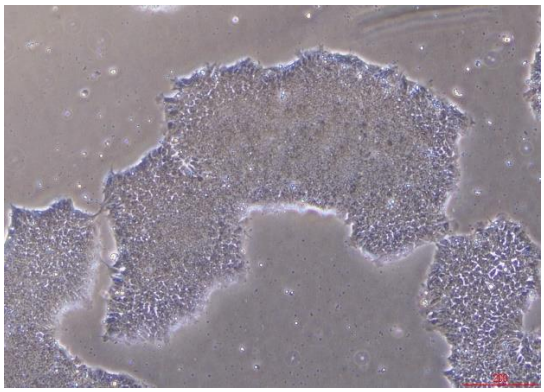


Figure 12. Light microscope images of iPSC colonies. (A) KI 1 passage 11 **(B)** KI 2 passage 11 **(C)** Parental HEL 24.3 passage 34. The ruler scale is 200 µm.

3.4 No off-target CRISPR-Cas9 cleavage was present in the two knock-in iPSC lines

Quality control analyses involved bioinformatic analyses of potential off-target genome editing by CRISPR-Cas9. Off-target regions were sequences similar to the complementary SARS2 region targeted by crRNA 2. Mismatches between crRNA-DNA pairing can be well tolerated at some similar loci and result in Cas9 cleavage. The two clonal cell lines KI 1 and KI 2 created with the RNP complex containing crRNA 2 underwent Sanger sequencing according to section 2.7 for the screening of off-target genome editing by SARS2 crRNA 2 (ctgtgcgggcctcacCGTGT). Table 17 summarises the

CFD off-target scores of the 7 most significant potential off-target loci for crRNA 2, and PCR primers from CRISPOR Batch Gene Targeting Assistant (http://crispor.tefor.net/) and amplicons from *in-silico* PCR (https://genome.ucsc.edu/cgi-bin/hgPcr). In total, CRISPOR Batch Gene Targeting Assistant estimated 69 off-target regions. No off-target Cas9 cleavage was detected in the regions (Figure 13) by Sanger sequencing and NCBI BLAST® (Basic Local Alignment Tool, Nucleotide Collection, 26th March 2020).

Locus description (CRISPOR)	**CFD off-target score**	**Forward primer (5' → 3')**	**Reverse primer (5' → 3')**	**Off-target sequence**	**Amplicon length (bp)**
intergenic:SPNS3-AC118754.4	0.40625	TATCTCTGTGTGCGTGTGCC	AGACACACAGGTTCAACCCCC	CTGTGTGTGCATCACCATGTTGG	596
intergenic:RP11-91J19.4-TDRP	0.363174603	ATGACCTCAGAACGTGGCAG	CACCACCACTTCGTCCATGAC	CCCACATGGTGAGTCCAGGACAG	537
intron:LHPP	0.233333333	ACACGAGTCTCACCATGCAG	GGTCTCAGCAGCACAGTCTTC	CCATGCGGGGCTCACCATGTGGG	533
intergenic:AC091150.1-ZBTB7C	0.225	GTGAGGATTTGCCAGGCTCT	AATTGAGGCCTGGGCTTCTCC	CCCACACGGTAAGGTCCCCAC	582
intron:ADARB2-AS1	0.202222222	TTACCCAGTGGCATGTGGTC	AGGTCCTCGCCTTCCCTAATC	CCAACACTGCGAGTCCCCCACAG	578
exon:BLK	0.077777778	CCATCTCTGCAGACCCTTGG	ACCGTATAGCCCTCAGGTGTC	CTGTGCGGGGGACACAGTGTGGG	553
exon:C15orf61/IQCH-AS1	0.070643642	AAGTACCACTGCTCCAAGGC	ACGAACTTGCCGGGAATTCAT	CCCACGCGGTGAAGCCCGCCCGG	561

Table 17. Analysis of CRISPR-Cas9 off-target effects. Top hits for potential off-target areas of SARS2 crRNA2 *(*ctgtgcgggcctcacCGTGT) and their CFD-scores, PCR primers for sequencing cell lines KI 1 and KI 2, and amplicon sequences.

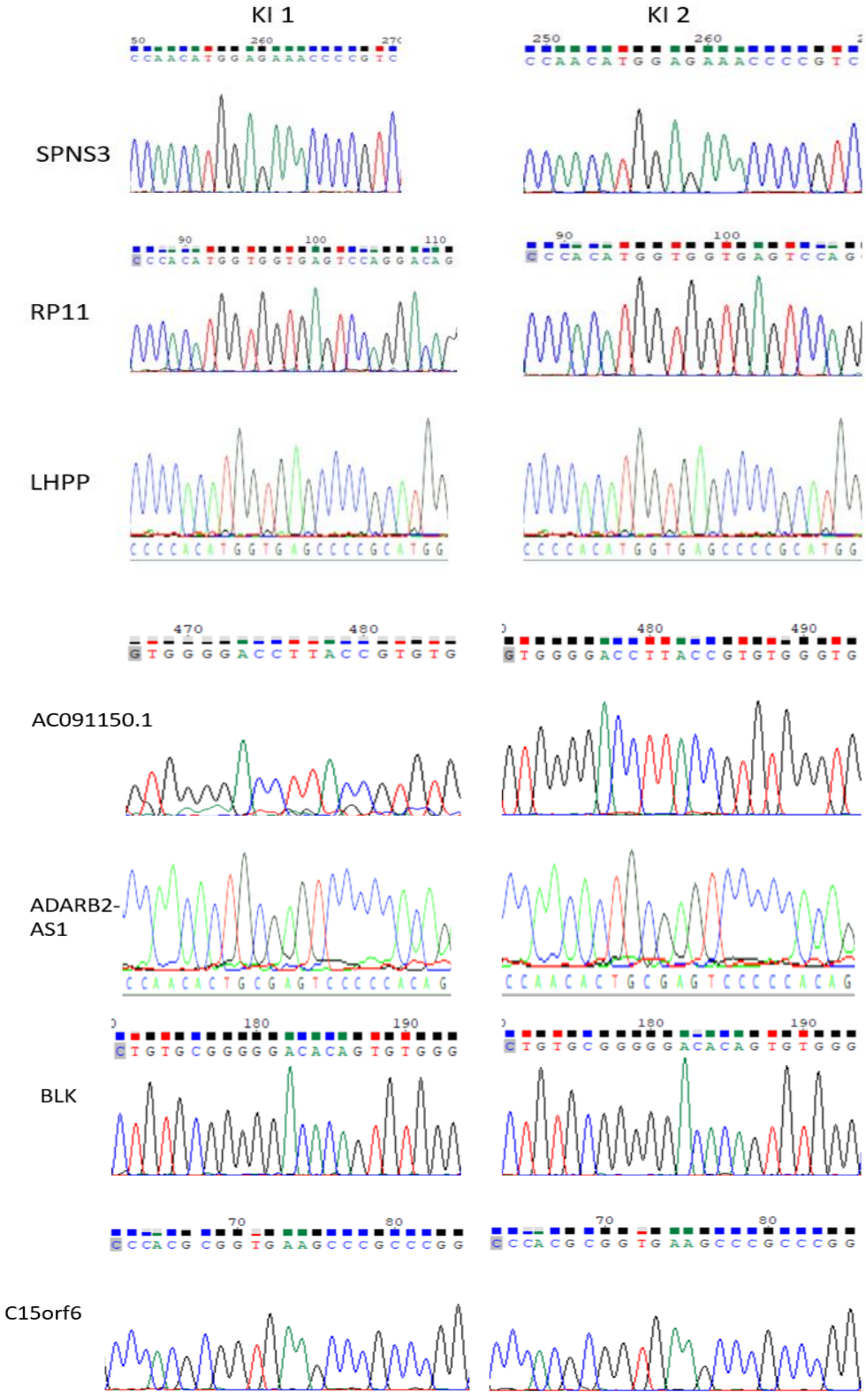


Figure 13. Sanger sequencing results of the off-target regions in genome-edited knock-in cell lines KI 1 and KI 2. No Cas9 cleavage or DSB correction by NHEJ was present as shown in the chromatograms for SPNS3, RP11, LHPP, AC091150.1, ADARB2-AS1, BLK and C15orf6. All the sequences corresponded to the reference sequence (NCBI BLAST®).

3.5 The genome-edited iPSC lines expressed pluripotency markers

iPSC need to remain in a pluripotent state to ensure their ability to be induced to differentiate into any cell type of the three germ layers (endoderm, mesoderm, ectoderm) for the purpose of disease modelling. The relative expressions of transcription factors Sox2, Oct4 and Nanog were significantly different ($P < 0.01$) between the iPSC lines and the wild type fibroblasts that were tested as a negative control for pluripotency (Figure 14). The expression of pluripotency markers in KI 1 and KI 2 demonstrates their pluripotency and thereby status as (induced) pluripotent stem cells.

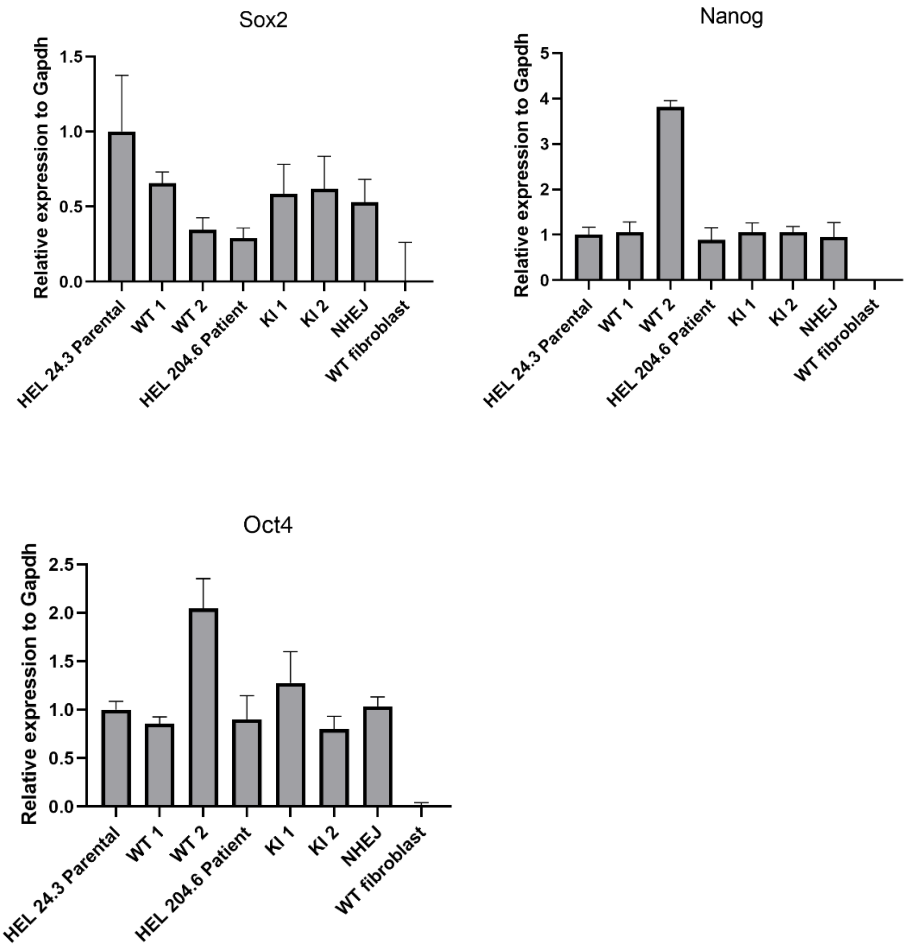
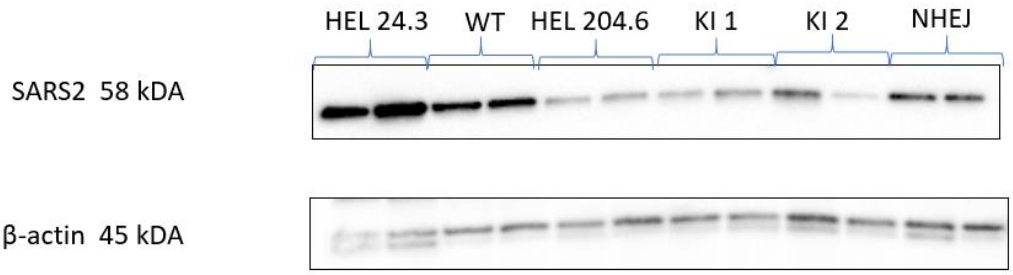


Figure 14. Relative expression of pluripotency markers in the compared iPSC lines. Relative expression of the pluripotency-related transcription factors Sox2, Oct4 and Nanog. Wild-type fibroblast as a negative control for pluripotency. Normalisation to GAPDH. Data shown as mean ± SD. One-way ANOVA ($P < 0.01$), n=3 technical replicates per cell line.

3.6 The genome-edited iPSC lines modelled the patient's SARS2 protein deficiency and mRNA expression

To test the comparability of the genome-edited knock-in iPSC lines to the patient and parental/wild type lines, SARS2 protein level was measured by Western blot, and relative Sars2 mRNA expression measured by RT-qPCR. SARS2 protein level was severely decreased in the patient and genome-edited iPSC, as shown by immunoblotting (Figure 15 A) and quantification of band intensity (Figure 15 B). The genome-edited iPSC lines modelled the patient's decrease in SARS2 protein level, with a decrease in intensity to 14% of normal compared to the parental in HEL 204.6, 15% in KI 1, 20% in KI 2, and 30% in the NHEJ clone. However, there was a marked difference in band intensity between the healthy controls (parental and wild type), which may be due to undiscovered off-target genome editing or other quality anomalies. Although KI 1, KI2 and patient iPSC showed a decreased Sars2 mRNA expression level, the isogenic lines KI 1 and KI 2 did not statistically significantly model the patient's significant decrease in Sars2 mRNA expression compared to parental and wild type. Again, there were statistically significant differences in Sars2 expression between the parental and other wild types (Figure 16).

A



B

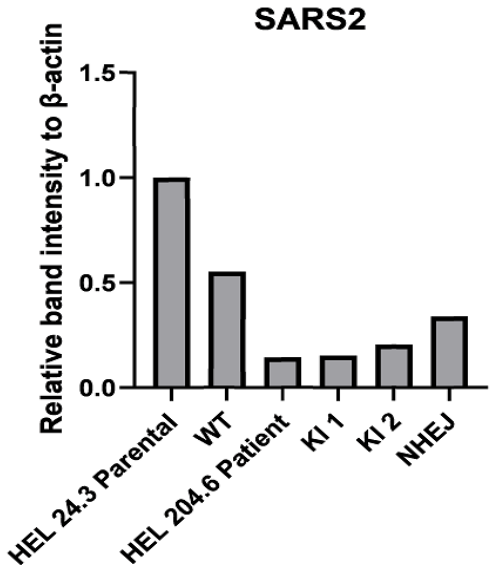


Figure 15. SARS2 Western blot results compared between control and patient cell lines. (A) Immunoblotting of SARS2 shows a marked decrease in the mitochondrial protein in the patient and knock-in mutants as compared to the healthy parental and wild type lines. The NHEJ clone shows a lesser decrease in protein than the KI mutation causing a frameshift. **(B)** The relative protein intensities normalised to loading control β-actin show a marked difference between the parental and wild type cell lines, and a decrease in intensity to 14% of parental in HEL 204.6, 15% in KI 1, 20% in KI 2, and 30% in the NHEJ clone (n=2 technical replicates per cell line).

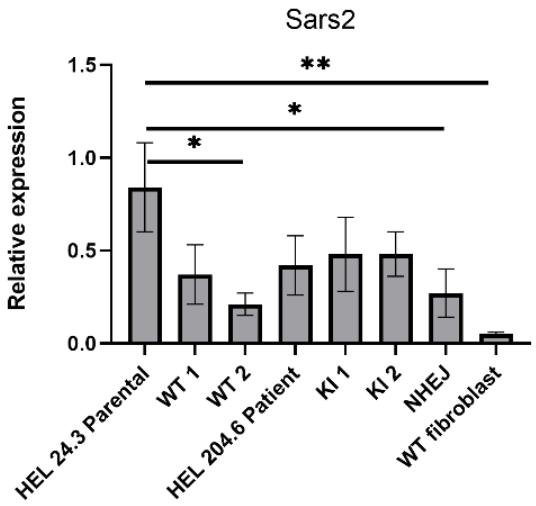


Figure 16. The relative mRNA expression of Sars2 was significantly different between the HEL 24.3 parental control and the NHEJ clone and fibroblast, but not between HEL 24.3 and the patient cell line HEL 204.6 or the two KI clones. Normalisation to GAPDH. Data shown as mean ± SD. One-way ANOVA *P < 0.05, **P < 0.01, n=3 technical replicates per cell line.

4. Discussion

The overarching aim of this project was to create genome-edited iPSC lines using CRISPR-Cas9 to model a patient-specific splicing variant c.1347 G>A in *SARS2* that causes severe childhood-onset progressive spastic paresis. All 19 mt-aaRS have been associated with disease-causing mutations causing a wide variety of phenotypes, presenting in disparate ways in different patients. Disease mechanisms for many are not known, and isogenic iPSC lines are a valuable tool for studying the molecular and cellular mechanisms of mt-aaRS related mitochondrial diseases.

Two different CRISPR-Cas9 ribonucleoprotein complexes and associated single-stranded donor oligonucleotides as repair templates were used to make a knock-in mutation in wild type iPSC. Two genome-edited iPSC cell lines that accurately model the patient genotype were created by one of the ribonucleoprotein complexes. The established iPSC line HEL24.3 was used as the parental cell line, as it is reported to express pluripotency markers Oct4, Sox2 and Nanog, and is able to differentiate into cells of all three germ layers by a teratoma assay, and has a normal karyotype (Trokovic et al., 2015). Additionally, the success and precision of the genome editing in the two genome edited disease models was established. The genome editing and HDR efficiencies by the two crRNAs used in the RNP complexes were investigated. The quality of the genome-edited iPSC lines in terms of pluripotency markers was measured, and thereby validity of the models to be used in differentiation to the affected cell type and studies of disease mechanisms. To preliminarily test the ability of the isogenic control iPSC lines to model the patient iPSC lines, SARS2 protein level and relative mRNA expression were measured.

4.1 The design and success of the CRISPR-Cas9 RNP complex

Out of two crRNAs designed to create genome edited knock-in iPSC lines, one successfully created two iPSC lines to model the patient mutation. The crRNA in the RNP complex immediately cleaves DNA, and the concentration of the homologous recombination template is highest immediately after electroporation, but rapidly declines over 24 hours reducing off-target effects (Kim et al., 2014). RNP complex delivery is less stressful to stem cells and more efficient than plasmid transfection (Kim et al., 2014). Many studies have shown that the most efficient and customisable for iPSC knock-in genome editing is CRISPR-Cas9 RNP with a repair template (reviewed in Bollen et al., 2018). Electroporation is the best transfection method for iPSC, which are difficult to transfect and edit compared to tumour cell lines or mouse embryonic stem cells (reviewed in Ihry et al., 2018). Electroporation has been reported to be most successful with Neon Electroporator (Life Technologies) that was used in this project, and Nucleofector (Lonza), which are the most commonly used electroporators (Li et al., 2015; Ran et al., 2013).

A wide range of monogenic, complex, chromosomal and epigenetic disorders occurring in various stages of life have been successfully modelled with human pluripotent stem cells (reviewed in Avior, Sagi & Benvenisty, 2016). The small phenotypic differences between cells differentiated from patient and control iPSC may reflect unpredictable genetic variation (differing in millions of SNPs in each genome) and non-genetic variation between individual iPSC lines, masking the disease phenotype. Experimental variation exists in cell differentiation and subsequent passaging. P53 mutations inhibit CRISPR-Cas9 editing and have toxic effects in human pluripotent stem cells. It has been reported that there is a reduced efficiency of precise editing in human pluripotent stem cells with a wild type p53 gene (Ihry et al., 2018). These may be confounding and limiting factors in the success of disease modelling. However, later passage iPSC have less epigenetic memory, which may be advantageous for modelling genomic mutations (Soldner & Jaenisch, 2012; Hockemeyer & Jaenisch, 2017).

The conditions of cell culture and genome editing are crucial to editing success. In this study, supplementing electroporated iPSC with 10 μm HDR enhancer produced the best rate of HDR and overall genome editing efficiency without affecting cell survival as seen by the number of attached cells 12 and 24 hours post-electroporation. The addition of a broad-spectrum antibiotic to avoid infections in cell culture did not affect cell survival or morphology at any stage of clonal cell line creation. iPSC were grown on a chemically defined vitronectin coating. Artificial cell culture conditions may affect the growth and characteristics of cells (Chen et al., 2011). iPSC of passage 32 were used for electroporation. Electroporation causes additional stress to iPSC, and stronger cells that divide quicker have an advantage in survival. A disadvantage of this is the potentially higher number of mutations that are harboured by higher passage cells. The addition of Rho-kinase (ROCK) inhibitor Y-27632 suppresses anoikis during disaggregation of human embryonic stem cell colonies and increases single-cell survival in stem cells (Watanabe et al., 2007). The addition of ROCK-inhibitor to cell culture medium before and after electroporation can contribute to iPSC survival in genome editing.

The rates of NHEJ and HDR in the electroporated iPSC pools were estimated based on Sanger sequencing input into the ICE CRISPR analysis software (Synthego) that estimates the proportion of each DNA repair mechanism. Based on these estimates, the occurrence of NHEJ in double-strand DNA break repair is higher than that of HDR, matching the results of previous studies (Kim et al., 2014; Bassett, 2017). The predicted crRNA success and the results from electroporation varied in several ways. On-target scores from the Benchling web tool, which represented the probability of Cas9 binding to the region targeted by the crRNA, suggested that crRNA1 and crRNA2 would have similar efficiency. However, crRNA 2 outperformed crRNA1 upon transfection to HEK293 cells and iPSC. CRISPOR's CFD (Cutting Frequency Determination) score ranked the uniqueness of a gRNA in the reference genome, considering the off-target cleavage fraction of the gRNA. Off-target regions should be as far away as possible from the PAM sequence to avoid direct interference with the

intended edit. CFD scores suggested that crRNA1 would be slightly more specific and efficient than crRNA2, which was not the case since the two knock-in cell lines were created with crRNA2. Additionally, after iPSC electroporation, the percentage of indels varied between T7 assay and estimates by the software ICE CRISPR analysis tool. This may be explained either by incomplete recognition of cleaved DNA by T7 endonuclease, or computational inaccuracies. However, the ICE CRISPR tool provided important information about the percentage of knock-ins, which was crucial for this project in choosing the right cell population for subcloning. A restriction enzyme could have been used to recognise the knocked-in point mutation more efficiently and reliably, but additional mutations next to the patient mutation were not made to avoid unwanted effects on the disease model. The use of a restriction enzyme would have eliminated the need for sequencing and indirect in-silico analysis. There are several possible reasons for the lower occurrence of HDR that will be discussed.

It was important to design multiple crRNAs, in this case four, and test the gene modification efficiencies on HEK293 cells. This was done to test crRNA efficiency under laboratory conditions, which may differ from computational efficiencies. Two of the four crRNAs produced favourable gene modification efficiencies in HEK293 cells based on GFP-positivity and T7-assay results. The laboratory results and crRNA on-target and off-target scores from two different *in silico* tools (Benchling, CRISPOR) were considered before iPSC electroporation.

The success and precision of genome editing depend on choosing an appropriate site, crRNA and ssODN design, and Cas9 activity. The better HDR performing crRNA 2 was designed as a reverse sequence, with the cut site to the mutation being just 1 nt. This could suggest that an RNP complex with the reverse crRNA sequence with the cut site as close as possible to the mutation produced the best results in terms of NHEJ and HDR efficiency. Previous studies have found that a DSB occurring close to the intended mutation site correlates with gRNA efficiency (Ran et al., 2013). While in this project the two crRNAs were designed in different orientations, Cas9 can detect a PAM sequence in the sense or antisense directions, so crRNA orientation should be of trivial importance (Wang et al., 2016a). While a p53 dominant negative plasmid was included in the electroporations to lower chances of toxic effects of p53 on cells with DSBs, its use was not necessary. The protein from the plasmid is expressed too slowly compared to the RNP-mediated DNA cleavage, resulting in the risk of plasmid integration into the target genome. Plasmid transfection can also be cytotoxic for some cell types (Kim et al., 2014).

4.2 Efficiency and quality of genome editing in the two successfully created genome-edited knock-in iPSC lines

In this project, the use of CRISPR-Cas9 RNP complex for genome editing and manual colony picking method of cell line creation produced favourable genome editing results for the purpose of disease modelling. Two genome-edited cell lines to model the patient mutation c.1347 G>A in *SARS2* were

successfully created from the cell population electroporated with crRNA 2 supplemented with 10 µM Alt-R HDR enhancer. This amounts to 2.35% of all colonies that were expanded originating from that cell population. From both the crRNAs, only one heterozygous clone was created. The Sanger sequencing and TOPO-cloning results suggest the ssODN incorporated into the genome in the desired way for a homozygous *SARS2* c.1347 G>A mutation. Like in the electroporated pools, there was a higher percentage of DNA breaks repaired by NHEJ than by HDR. The overall genome editing efficiency was higher in crRNA 2 than crRNA 1. Out of the clonal cell lines created, only 50% were wild type for crRNA 1 and 29.41% were wild type for crRNA2. Although both NHEJ and HDR efficiencies were lower than in previous studies, two pluripotent isogenic iPSC lines with homozygous recombination without off-targets were successfully created from picking only 85 colonies. Saarimäki-Vire et al. (2017) reported 5.3% of iPSC clones having heterozygous or homozygous recombination and correction of the patient mutation with the repair template. Low toxicity of RNP complex electroporation and further ensuring cell survival of electroporated cells by manual colony picking over FACS possibly also contributed to the successful outcome in genome editing. While manual colony picking is less reliable than automatic cell sorting for creating clonal cell lines, careful sequencing of the clones KI 1 and KI 2 determined that they likely originate from a single cell (Bruntraeger et al., 2019).

In most cell types, NHEJ is massively favoured over HDR, with HDR only being likely in the late-S, G2 phases of the cell cycle when DNA replication is complete and sister chromatids are available to serve as repair templates. HDR is the least error-prone DSB repair strategy but harnessing it in experimental setups is difficult and designs are often low in editing efficiency (Shrivastav et al., 2008). Successful creation of knock-in mutants exploiting the HDR pathway is restricted to proliferative cells in a logarithmic growth phase without growing to full confluency. Mutagenesis by NHEJ varies between experimental setups but can be as high as 80-90% (Bassett, 2017). Cho et al. (2013) reported indel frequencies of up to 33% in human cells using CRISPR-Cas9. Genome editing efficiencies and especially HDR frequency tend to be lower in embryonic stem cells and iPSC than in cancer cell lines (Hockemeyer & Jaenisch, 2017). This phenomenon may also explain the lower rate of HDR compared to NHEJ rate in this project.

HDR can be supported by NHEJ inhibition or HDR enhancement. However, HDR enhancement is generally preferred due to cytotoxic side effects of NHEJ inhibition. Jayavaradhan et al. (2019) showed in K562 cells (a human immortalised myelogenous leukaemia cell line) that Cas9-DN1S fusion proteins (a dominant-negative mutant of 53BP1 bound to Cas9 nucleases) significantly blocked NHEJ events locally at Cas9 cut sites and improve HDR frequency Wang et al. (2017) reported that using a doxycycline-inducible Cas9 transgene on a PiggyBac transposon resulted in INDEL frequencies of 40-50% and HDR frequencies of 10-20%. Skarnes, Pellegrino & McDonough (2019) demonstrated that the added effects of a cold shock, small molecule HDR enhancer, and chemically

modified ssODN shift DNA break repair to favour HDR, resulting in HDR in 70% of unselected human iPSC. Combining cell cycle synchronisation in the G2/S phase with timed RNP delivery can increase HDR success by increasing nuclease activity in the HDR favourable phase of the cell cycle (Bollen et al., 2018). Using a paired nickase approach, Bothmer et al. (2017) observed a shift in repair outcome distribution and a specific signature of repair outcomes for two different overhang structures. D10A-induced dual-nicking results in a 50 overhang that more frequently engages HDR than N863A-induced 30 overhangs. Yu et al. (2019) reported that using a Cas9 ribonucleoprotein with 5'-modified double-stranded DNA donors with 50 bp homology arms lead up to a fivefold change in HDR rates at various locations in human stem cells. While the design of silent mutation outside the SNP of interest can decrease the chance of re-cutting by Cas9, the additional silent SNPs may affect protein translation efficiency (Bassett, 2017). The use of nickases to create single-strand breaks, modified Cas9 proteins, or design of silent mutations can all help lower off-target genome editing depending on the target locus and cell type. Most testing has been done on cancer cell lines (Hockemeyer & Jaenisch, 2017).

Rates of HDR to introduce disease-associated mutations can be increased by using an ssODN repair template over dsDNA due to somewhat different DNA repair mechanisms. ssODNs produce short conversion tracts preferentially through synthesis-dependent strand annealing, enabling efficient DNA repair without sacrificing genomic stability (Kan et al., 2017). Additionally, ssDNA has higher recombinogenic tendencies (Bassett, 2017). Zhang et al. (2017) found that the supplementation with both CCND1, a cyclin that functions in G1/S transition, and nocodazole, a G2/M phase synchronizer, doubles HDR efficiency to up to 30% in iPSCs. Results from other studies support the use of reversible chemical inhibitors to enhance HDR. Supplementation of stem cells and iPSC with ABT to arrest cells in G2 phase increases on-target genome editing and donor incorporation, compared to cells in G1 phase (Yang et al., 2016). Controlled timing by CRISPR-Cas9 delivery with nocodazole synchronization and high concentration of Cas9 RNP at 100 ρmol increased homology directed repair up to 31% without affecting cell morphology in culture (Lin et al., 2014). Yang et al. (2016) also showed that a 16-hour treatment with 1μg/ml Nocodazole and one hour in Nocodazole-free medium showed significant G2/M phase enrichment in 80% of cells, while 58.1% of untreated cells were in G1 phase. These results suggest that rates of HDR can be enhanced by timed delivery of CRISPR-Cas9 complex, and chemically arresting cells in G2 phase can enhance HDR without being significantly harmful at least for disease modelling purposes.

When considering *in vivo* mutation corrections that could be used to correct patient mutations such as the one in this project, precision and safety are even more crucial. Yusa et al. (2011) used ZFN with *piggyBac* transposition to make a point mutation (Glu342Lys) in the α_1-antitrypsin gene to rescue the α_1-antitrypsin deficiency in differentiated liver cells in vitro and in vivo. This demonstrates that genetic correction of human iPSC can generate clinically relevant cells for autologous transplantation.

Concerns about reliability of iPSC and their applicability for therapeutic use have arisen from reports about differences in global chromatin structure and gene expression profiles between human iPSC and embryonic stem cells. However, they show very few differences in terms of transcriptional programs and nucleosome structures (Guenther et al., 2010). Immune responses against Cas9 proteins are another hurdle for in-vivo use of CRISPR-Cas9 (Broeders et al., 2020).

Understanding how to avoid off-target DNA modification by CRISPR-Cas9 is important both for fidelity of disease modelling and for clinical applications of edited iPSC like autologous cell transplantation. In this project, no off-target indels were present in the likely off-target regions that were analysed. Off-target mutagenesis can disrupt endogenous gene function, producing artefacts that can be misleading in *in vitro* experiments, and harmful for clinical use (Frock et al., 2014). Undesired mutations in both alleles can also be present in correctly targeted clones created using the Cas9 nuclease, and Cas9-D10A single and double nickases (Merkle et al., 2015). The advantage of CRISPR-Cas9 RNP complexes is their lower rate of off-target mutations than viral-vector based CRISPR-Cas9 methods (Argani, 2019). Single and double mismatches are tolerated by Cas9 nucleases, and off-target sites can harbour up to five mismatches and many were mutagenized with frequencies comparable to (or higher than) those observed at the intended on-target site (Fu et al., 2013). Restricting the time period CRISPR-Cas9 is active, off-target mutagenesis can be lowered (Merkle et al., 2015). Some studies have found that reducing the amount of active Cas9 in the cell reduces off-target mutagenesis. However, the amount of Cas9 needs to be optimised so that on-target cleavage is ensured to a reasonable standard. To decrease off-targets, other studies have pointed to the use of Cas9 nickase mutants to create a pair of juxtaposed single-stranded DNA nicks, truncating the gRNA sequence at the 5' end, or using a pair of catalytically inactive Cas9 nucleases fused to a Fok1 nuclease domain (reviewed in Slaymaker et al., 2016). Based on high-throughput genome-wide translocation sequencing, Frock et al. (2015) report that the Cas9D10A paired nickase approach suppresses genome-wide off-target cleavage. Limitations in targetability due to restrictions by species-specific PAM sequence can be overcome by finding alternative orthologous CRISPR systems e.g. SaCas9 and Cpf1, in addition to engineering of Cas9 proteins (Bollen et al., 2018).

Off-targets can also be minimised by approaching HDR machinery. The composition of the donor template can be altered towards favouring a particular DSB repair pathway (Bollen et al., 2018). Non-specific integration of DNA is minimised by using an ssODN repair template over dsDNA (Bruntraeger et al., 2017). Symmetric ssODNs produce more favourable results than asymmetric ssODNs (reviewed in Bollen et al., 2018). Base editors are a good alternative for making single nucleotide substitutions, due to no genomic cleavage, eliminating the off-target effects. However, base editors are limited to certain base substitutions, and are more suited to creating heterozygous mutations than homozygous mutations (Rees & Liu, 2018). Since the goal of this project was to create a homozygous knock-in mutation, HDR-mediated genome editing remains the most feasible and

economical choice for this goal. Considering the error-prone nature of CRISPR-Cas9, karyotyping and comprehensive sequencing should be done to screen clones to fully establish the quality of the cell line and thereby the validity of the disease model.

Off-targets in exonic regions may be detrimental to disease modelling. The severity of off-target effects in the disease model would depend on the role of the off-target regions in cell health and the disease phenotype. The discrepant Sars2 mRNA expression level differences between the parental and other wild types could be due to other off-target genome editing not analysed, or technical errors, differences in RNA quality, or fluorescence. Similarly, discrepant results in pluripotency marker expression could be explained by the presence of differentiated cells in the collected iPSC population. However, all cell pellets collected contained no visible differentiated cells and were approximately 80% confluent on a 35 mm cell culture dish.

4.3 Questions to be answered using the genome-edited iPSC lines

All aminoacyl-tRNA synthetases and mitochondrial aminoacyl-tRNA synthetases have the same canonical role of charging the appropriate amino acid onto its tRNA (aminoacylation). The use of genome-edited and isogenic controls is crucial for studying the heterogeneous defects in mitochondrial aminoacyl tRNA-synthetase disorders. The use of isogenic controls and wild types provides a more precise approach to investigating and characterising disease phenotypes and mechanisms than the more commonly used method of transgenic overexpression, which cannot be used to accurately target recessive phenotypes (Bollen et al., 2018). Recent studies on different mt-aaRS based mitochondrial diseases suggest a categorisation of diseases based on tissue specificity of clinical manifestations: 1. Mainly affecting the CNS, 2. Affecting the CNS and another system, and 3. Affecting the CNS or another system, and 4. Affecting a system other than the CNS (reviewed in González-Serrano, Chihade & Sissler, 2019). Discovering the mechanisms behind mt-aaRS related disease phenotypes would help in understanding and categorisation of diseases and thereby discovering the functions of the individual mt-aaRS enzymes for targeting treatments.

The two iPSC lines created in this project modelled the patient's decrease in SARS2 protein level and mRNA expression. This preliminary disease mechanism analysis suggests that the c.1347 G>A mutation in *SARS2* causes a defect at the protein level leading to reduced SARS2 protein level. In the original patient report, Linnankivi et al. (2016) found in the patient fibroblasts a decrease in SARS2 protein down to 10% of normal. Further, differentiation of the patient iPSC and edited cell lines and their isogenic parental controls into motor neurons and characterisation of differential gene expression can reveal the neuron-specific effects of the mutation in protein and mRNA levels, development and morphology. More specific exploratory disease mechanism analyses include studying more specific structures like mitochondrial complexes and analyses of mitochondrial translation. The results could

reveal the importance and role of SARS2 in different cell types, which might be a clue to discovering disease mechanism pathways. In investigating the mechanism of the SARS2 defect, it is important to consider how crucial the SARS2 protein is for iPSC and motor neuron functions and morphology.

The expression of pluripotency markers Oct4, Sox2, and Nanog suggested that the genome-edited iPSC lines were indeed pluripotent stem cells. However, their validity as iPSC models and ability to be differentiated into any cell type in the three germ layers (endoderm, mesoderm, ectoderm) to investigate the mechanisms of the *SARS2* c.1347 G>A mutation in the childhood-onset progressive spastic paresis phenotype will need to be further established. This can be done by an in-vitro differentiation assay such as embryoid body differentiation where cells are guided to spontaneously differentiate, after which markers of each germ layer are analysed with immunocytochemistry. As a quality control step in addition to RT-qPCR of the pluripotency markers, immunocytochemistry should be used to see the localisation and intensity of pluripotency markers.

4.4 Conclusions and future perspectives

A generalised CRISPR-Cas9 platform for iPSC editing is an important method for both basic and biomedical research. A highly customisable, highly target-specific and quick to assemble CRISPR-Cas9 RNP complex can improve the speed and reliability of iPSC-based *in vitro* disease modelling in single-gene disorders of unknown etiology. Manipulating genes enables the characterisation of their functions. This could also enable increased precision for *in vivo* editing in animal models and modification of patient cells ex-vivo for autologous transplantation. iPSC as model systems provide many advantages over cancer cells, which have long been used as the default cell model system. Advantages of iPSC include patient specificity with relevant genotypic background, and customisability in that they can be edited and differentiated into disease-relevant cell types to elucidate cell type specific disease mechanisms. In this project, two patient-specific genome-edited control iPSC lines were successfully and precisely created and partly validated as disease models.

The remaining challenges of using CRISPR-Cas9 for iPSC-based disease modelling include low efficiency in iPSC, iPSC heterogeneity in morphology, and maintenance of pluripotency in culture. *In vitro* culture and genome editing by electroporation could alter genomic stability of the cells (Hockemeyer & Jaenisch, 2017). iPSC models can reveal information about cellular and intracellular events but are limited in their ability to model more complex phenotypes, for example ones that affect whole organs or organ systems. Additionally, cell culture on synthetic growth surfaces may affect disease modelling capabilities in known and unknown ways. To optimise CRISPR-Cas9 genome editing, more advanced *in silico* methods are being developed for designing and estimating efficiencies of crRNAs and ssODN/ssDNA homologous recombination templates with reduced off-targets. Other ways of improving gRNA fidelity include engineering gRNA stability and optimising

RNP complex chemistries. Considering iPSC culture conditions and CRISPR-Cas9 setup and optimisation can improve speed and accuracy of iPSC disease modelling and reporter cell line capabilities while maintaining iPSC survival rate and stemness.

5. Acknowledgements

I would like to thank my supervisors Associate Prof. Henna Tyynismaa and Post doc Tiina Rasila for their support and guidance throughout the thesis and for the chance to work on this project. Special thanks also goes to MSc. Rosa Woldegebriel for her advice on iPSC genome editing. Thanks to Biomedicum Stem Cell Center (BSCC) for the iPSC lines, and the Institute for Molecular Medicine Finland (FIMM) for sequencing all the clonal cell lines. I would also like to thank all the members of the Tyynismaa group, especially Riitta Lehtinen and Jana Pennonen for technical support.

6. References

Anders, C. Niewoehner O., Duerst, A. & Jinek, M. (2014). Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature*, 513, 569–573. doi:10.1038/nature13579

Anderson, S., Bankier, A., Barrell, B., de Bruijn, M., Coulson, A., & Drouin, J. et al. (1981). Sequence and organization of the human mitochondrial genome. *Nature*, *290*(5806), 457-465. doi: 10.1038/290457a0

Argani, H. (2019). Genome Engineering for Stem Cell Transplantation. *Experimental And Clinical Transplantation*, *17*(Suppl 1), 31-37. doi: 10.6002/ect.mesot2018.l34

Avior, Y., Sagi, I., & Benvenisty, N. (2016). Pluripotent stem cells in disease modelling and drug discovery. *Nature Reviews Molecular Cell Biology*, *17*(3), 170-182. doi: 10.1038/nrm.2015.27

Avior, Y., Sagi, I., & Benvenisty, N. (2016). Pluripotent stem cells in disease modelling and drug discovery. *Nature Reviews Molecular Cell Biology,* 17(3), 170-182. doi: 10.1038/nrm.2015.27

Balboa, D., Saarimäki-Vire, J., & Otonkoski, T. (2018). Concise Review: Human Pluripotent Stem Cells for the Modeling of Pancreatic β-Cell Pathology. *Stem Cells*, 37(1), 33-41. doi: 10.1002/stem.2913

Balboa, D., Weltner, J., Eurola, S., Trokovic, R., Wartiovaara, K., & Otonkoski, T. (2015). Conditionally stabilized dCas9 activator for controlling gene expression in human cell reprogramming and differentiation. *Stem Cell Reports*, *5*(3), 448-459. doi: 10.1016/j.stemcr.2015.08.001

Barral, S., & Kurian, M. (2016). Utility of Induced Pluripotent Stem Cells for the Study and Treatment of Genetic Diseases: Focus on Childhood Neurological Disorders. *Frontiers In Molecular Neuroscience*, 9. doi: 10.3389/fnmol.2016.00078

Barrangou, R. (2015). Diversity of CRISPR-Cas immune systems and molecular machines. *Genome Biology*, 16(1). doi: 10.1186/s13059-015-0816-9

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

Bellin, M., Marchetto, M., Gage, F., & Mummery, C. (2012). Induced pluripotent stem cells: the new patient?. *Nature Reviews Molecular Cell Biology*, *13*(11), 713-726. doi: 10.1038/nrm3448

Belostotsky, R., Ben-Shalom, E., Rinat, C., Becker-Cohen, R., Feinstein, S., Zeligson, S., Segel, R., Elpeleg, O., Nassar, S., Frishberg, Y. Mutations in the mitochondrial seryl-tRNA synthetase cause hyperuricemia, pulmonary hypertension, renal failure in infancy and alkalosis, HUPRA syndrome. Am. J. Hum. Genet. 88: 193-200, 2011. doi: 10.1016/j.ajhg.2010.12.010

Ben Jehuda, R., Shemer, Y., & Binah, O. (2018). Genome Editing in Induced Pluripotent Stem Cells using CRISPR/Cas9. *Stem Cell Reviews And Reports*, 14(3), 323-336. doi: 10.1007/s12015-018-9811-3

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., & Kay, S. et al. (2009). Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. *Science*, 326(5959), 1509-1512. doi: 10.1126/science.1178811

Bollen, Y., Post, J., Koo, B., & Snippert, H. (2018). How to create state-of-the-art genetic model systems: strategies for optimal CRISPR-mediated genome editing. *Nucleic Acids Research*, *46*(13), 6435-6454. doi: 10.1093/nar/gky571

Bolotin A, Quinquis B, Sorokin A, Ehrlich SD. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extra- chromosomal origin. *Microbiology*, 151, 2551–2561. doi:10 .1099/mic.0.28048-0.

Bonner, W., Hulett, H., Sweet, R., & Herzenberg, L. (1972). Fluorescence Activated Cell Sorting. *Review Of Scientific Instruments*, 43(3), 404-409. doi: 10.1063/1.1685647

Bothmer, A., Phadke, T., Barrera, L., Margulies, C., Lee, C., & Buquicchio, F. et al. (2017). Characterization of the interplay between DNA repair and CRISPR/Cas9-induced DNA lesions at an endogenous locus. *Nature Communications*, *8*(1). doi: 10.1038/ncomms13905

Broeders, M., Herrero-Hernandez, P., Ernst, M., van der Ploeg, A., & Pijnappel, W. (2020). Sharpening the Molecular Scissors: Advances in Gene-Editing Technology. *Iscience*, *23*(1), 100789. doi: 10.1016/j.isci.2019.100789

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

Bruntraeger, M., Byrne, M., Long, K., & Bassett, A. (2019). Editing the Genome of Human Induced Pluripotent Stem Cells Using CRISPR/Cas9 Ribonucleoprotein Complexes. *Methods In Molecular Biology*, 153-183. doi: 10.1007/978-1-4939-9170-9_11.

Chan, D. (2006). Mitochondria: Dynamic Organelles in Disease, Aging, and Development. *Cell*, *125*(7), 1241-1252. doi: 10.1016/j.cell.2006.06.010

Chen, G., Gulbranson, D., Hou, Z., Bolin, J., Ruotti, V., & Probasco, M. et al. (2011). Chemically defined conditions for human iPSC derivation and culture. *Nature Methods*, *8*(5), 424-429. doi: 10.1038/nmeth.1593

Charpentier, E., Richter, H., van der Oost, J., & White, M. (2015). Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity. *FEMS Microbiology Reviews*, 39(3), 428-441. doi: 10.1093/femsre/fuv023

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

Concordet, J., & Haeussler, M. (2018). CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Research*, 46(W1), W242-W245. doi: 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., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/ Cas systems. *Science*, 339, 819–823. https://doi.org/10.1126/science.1231143.

Craven, L., Alston, C., Taylor, R., & Turnbull, D. (2017). Recent Advances in Mitochondrial Disease. *Annual Review Of Genomics And Human Genetics*, *18*(1), 257-275. doi: 10.1146/annurev-genom-091416-035426

Deltcheva, E., Chylinski, K., Sharma, C., Gonzales, K., Chao, Y., & Pirzada, Z. et al. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, 471(7340), 602-607. doi: 10.1038/nature09886

Ding, Q., Regan, S., Xia, Y., Oostrom, L., Cowan, C., & Musunuru, K. (2013). Enhanced Efficiency of Human Pluripotent Stem Cell Genome Editing through Replacing TALENs with CRISPRs. *Cell Stem Cell*, 12(4), 393-394. doi: 10.1016/j.stem.2013.03.006

Diodato, D., Ghezzi, D., & Tiranti, V. (2014). The Mitochondrial Aminoacyl tRNA Synthetases: Genes and Syndromes. *International Journal Of Cell Biology*, *2014*, 1-11. doi: 10.1155/2014/787956

Doench, J., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E., & Donovan, K. et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology*, 34(2), 184-191. doi: 10.1038/nbt.3437

Doss, M., & Sachinidis, A. (2019). Current Challenges of iPSC-Based Disease Modeling and Therapeutic Implications. *Cells*, 8(5), 403. doi: 10.3390/cells8050403

Du, Z., Chen, H., Liu, H., Lu, J., Qian, K., & Huang, C. et al. (2015). Generation and expansion of highly pure motor neuron progenitors from human pluripotent stem cells. *Nature Communications*, 6(1). doi: 10.1038/ncomms7626

Fernández, A., Josa, S., & Montoliu, L. (2017). A history of genome editing in mammals. *Mammalian Genome*, 28(7-8), 237-246. doi: 10.1007/s00335-017-9699-2.

Fine, A., Nemeth, C., Kaufman, M., & Fatemi, A. (2019). Mitochondrial aminoacyl-tRNA synthetase disorders: an emerging group of developmental disorders of myelination. *Journal Of Neurodevelopmental Disorders*, 11(1). doi: 10.1186/s11689-019-9292-y

Frock, R., Hu, J., Meyers, R., Ho, Y., Kii, E., & Alt, F. (2014). Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nature Biotechnology*, *33*(2), 179-186. doi: 10.1038/nbt.3101

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

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

Garneau, J., Dupuis, M., Villion, M., Romero, D., Barrangou, R., & Boyaval, P. et al. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, *468*(7320), 67-71. doi: 10.1038/nature09523

Giacalone, J., Sharma, T., Burnight, E., Fingert, J., Mullins, R., Stone, E., & Tucker, B. (2018). CRISPR-Cas9-Based Genome Editing of Human Induced Pluripotent Stem Cells. *Current Protocols In Stem Cell Biology*, 5B.7.1-5B.7.22. doi: 10.1002/cpsc.46

González-Serrano, L., Chihade, J., & Sissler, M. (2019). When a common biological role does not imply common disease outcomes: Disparate pathology linked to human mitochondrial aminoacyl-tRNA synthetases. *Journal Of Biological Chemistry*, 294(14), 5309-5320. doi: 10.1074/jbc.rev118.002953

Gorman, G., Chinnery, P., DiMauro, S., Hirano, M., Koga, Y., & McFarland, R. et al. (2016). Mitochondrial diseases. *Nature Reviews Disease Primers*, *2*(1). doi: 10.1038/nrdp.2016.80

Götz, A., Tyynismaa, H., Euro, L., Ellonen, P., Hyötyläinen, T., & Ojala, T. et al. (2011). Exome Sequencing Identifies Mitochondrial Alanyl-tRNA Synthetase Mutations in Infantile Mitochondrial Cardiomyopathy. *The American Journal Of Human Genetics*, 88(5), 635-642. doi:10.1016/j.ajhg.2011.04.006

Guenther, M., Frampton, G., Soldner, F., Hockemeyer, D., Mitalipova, M., Jaenisch, R., & Young, R. (2010). Chromatin Structure and Gene Expression Programs of Human Embryonic and Induced Pluripotent Stem Cells. *Cell Stem Cell*, *7*(2), 249-257. doi: 10.1016/j.stem.2010.06.015

Guo, M., & Schimmel, P. (2013). Essential nontranslational functions of tRNA synthetases. *Nature Chemical Biology*, *9*(3), 145-153. doi: 10.1038/nchembio.1158

Haupt, A., Grancharova, T., Arakaki, J., Fuqua, M., Roberts, B., & Gunawardane, R. (2018). Endogenous Protein Tagging in Human Induced Pluripotent Stem Cells Using CRISPR/Cas9. *Journal Of Visualized Experiments*, (138). doi: 10.3791/58130

Hermans, P., van Soolingen, D., Bik, E., de Haas, P., Dale, J., & van Embden, J. (1991). Insertion element IS987 from Mycobacterium bovis BCG is located in a hot-spot integration region for insertion elements in Mycobacterium tuberculosis complex strains. *Infection And Immunity*, 59(8), 2695-2705. doi: 10.1128/iai.59.8.2695-2705.1991

Heyer, W., Ehmsen, K., & Liu, J. (2010). Regulation of Homologous Recombination in Eukaryotes. *Annual Review Of Genetics*, 44(1), 113-139. doi: 10.1146/annurev-genet-051710-150955

Hilander, T., Zhou, X., Konovalova, S., Zhang, F., Euro, L., & Chilov, D. et al. (2017). Editing activity for eliminating mischarged tRNAs is essential in mammalian mitochondria. *Nucleic Acids Research*, 46(2), 849-860. doi: 10.1093/nar/gkx1231

Hille, F., & Charpentier, E. (2016). CRISPR-Cas: biology, mechanisms and relevance. *Philosophical Transactions Of The Royal Society B: Biological Sciences*, 371(1707), 20150496. doi: 10.1098/rstb.2015.0496

Hockemeyer, D., & Jaenisch, R. (2016). Induced Pluripotent Stem Cells Meet Genome Editing. *Cell Stem Cell*, 18(5), 573-586. doi: 10.1016/j.stem.2016.04.013

Hrvatin, S., O'Donnell, C., Deng, F., Millman, J., Pagliuca, F., & DiIorio, P. et al. (2014). Differentiated human stem cells resemble fetal, not adult, β cells. *Proceedings Of The National Academy Of Sciences*, *111*(8), 3038-3043. doi: 10.1073/pnas.1400709111

Hsiau, T., Conant, D., Rossi, N., Maures, T., Waite, K., & Yang, J. et al. (2018). Inference of CRISPR Edits from Sanger Trace Data. *bioRxiv*, doi: 10.1101/251082

Ihry, R., Worringer, K., Salick, M., Frias, E., Ho, D., & Theriault, K. et al. (2018). p53 inhibits CRISPR–Cas9 engineering in human pluripotent stem cells. *Nature Medicine*, *24*(7), 939-946. doi: 10.1038/s41591-018-0050-6

Ishino, Y., Krupovic, M., & Forterre, P. (2018). History of CRISPR-Cas from Encounter with a Mysterious Repeated Sequence to Genome Editing Technology. *Journal Of Bacteriology*, *200*(7). doi: 10.1128/jb.00580-17

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. doi: 10.1128/jb.169.12.5429-5433.1987

Jayavaradhan, R., Pillis, D., Goodman, M., Zhang, F., Zhang, Y., Andreassen, P., & Malik, P. (2019). CRISPR-Cas9 fusion to dominant-negative 53BP1 enhances HDR and inhibits NHEJ specifically at Cas9 target sites. *Nature Communications*, *10*(1). doi: 10.1038/s41467-019-10735-7

Jiang, F., & Doudna, J. (2017). CRISPR–Cas9 Structures and Mechanisms. *Annual Review Of Biophysics*, 46(1), 505-529. doi: 10.1146/annurev-biophys-062215-010822

Jiang, F., Taylor, D., Chen, J., Kornfeld, J., Zhou, K., & Thompson, A. et al. (2016). Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science*, *351*(6275), 867-871. doi: 10.1126/science.aad8282

Jin, Y., Shen, Y., Su, X., Weintraub, N., & Tang, Y. (2019). CRISPR/Cas9 Technology in Restoring Dystrophin Expression in iPSC-Derived Muscle Progenitors. *Journal Of Visualized Experiments*, (151). doi: 10.3791/59432

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J., & Charpentier, E. (2012). A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*, *337*(6096), 816-821. doi: 10.1126/science.1225829

Jinek, M., Jiang, F., Taylor, D., Sternberg, S., Kaya, E., & Ma, E. et al. (2014). Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation. *Science*, *343*(6176), 1247997-1247997. doi: 10.1126/science.1247997

Kan, Y., Ruis, B., Takasugi, T., & Hendrickson, E. (2017). Mechanisms of precise genome editing using oligonucleotide donors. *Genome Research*, *27*(7), 1099-1111. doi: 10.1101/gr.214775.116

Kausar, S., Wang, F., & Cui, H. (2018). The Role of Mitochondria in Reactive Oxygen Species Generation and Its Implications for Neurodegenerative Diseases. *Cells*, *7*(12), 274. doi: 10.3390/cells7120274

Kim, K., Doi, A., Wen, B., Ng, K., Zhao, R., & Cahan, P. et al. (2010). Epigenetic memory in induced pluripotent stem cells. *Nature*, *467*(7313), 285-290. doi: 10.1038/nature09342

Kim, S., Kim, D., Cho, S., Kim, J., & Kim, J. (2014). Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Research*, *24*(6), 1012-1019. doi: 10.1101/gr.171322.113

Kim, Y., Cha, J., & Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proceedings Of The National Academy Of Sciences*, *93*(3), 1156-1160. doi: 10.1073/pnas.93.3.1156

Konovalova, S., & Tyynismaa, H. (2013). Mitochondrial aminoacyl-tRNA synthetases in human disease. *Molecular Genetics And Metabolism*, *108*(4), 206-211. doi: 10.1016/j.ymgme.2013.01.010

Kyttälä, A., Moraghebi, R., Valensisi, C., Kettunen, J., Andrus, C., & Pasumarthy, K. et al. (2016). Genetic Variability Overrides the Impact of Parental Cell Type and Determines iPSC Differentiation Potential. *Stem Cell Reports*, *6*(2), 200-212. doi: 10.1016/j.stemcr.2015.12.009

Li, H., Beckman, K., Pessino, V., Huang, B., Weissman, J., & Leonetti, M. (2017). Design and specificity of long ssDNA donors for CRISPR-based knock-in. bioRxiv, doi: 10.1101/178905

Li, H., Gee, P., Ishida, K., & Hotta, A. (2016). Efficient genomic correction methods in human iPS cells using CRISPR–Cas9 system. *Methods, 101*, 27-35. doi: 10.1016/j.ymeth.2015.10.015

Lieber, M. (2010). The Mechanism of Double-Strand DNA Break Repair by the Nonhomologous DNA End-Joining Pathway. *Annual Review Of Biochemistry, 79*(1), 181-211. doi: 10.1146/annurev.biochem.052308.093131

Lin, S., Staahl, B., Alla, R., & Doudna, J. (2014). Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife, 3*. doi: 10.7554/elife.04766

Lin, S., Staahl, B., Alla, R., & Doudna, J. (2014). Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife, 3*. doi: 10.7554/elife.04766

Linnankivi, T., Neupane, N., Richter, U., Isohanni, P., & Tyynismaa, H. (2016). Splicing Defect in Mitochondrial Seryl-tRNA Synthetase Gene Causes Progressive Spastic Paresis Instead of HUPRA Syndrome. *Human Mutation, 37*(9), 884-888. doi: 10.1002/humu.23021

Livak, K., & Schmittgen, T. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. *Methods, 25*(4), 402-408. doi: 10.1006/meth.2001.1262

Loh, Y., Wu, Q., Chew, J., Vega, V., Zhang, W., & Chen, X. et al. (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nature Genetics, 38*(4), 431-440. doi: 10.1038/ng1760

Lorenz, C., Lesimple, P., Bukowiecki, R., Zink, A., Inak, G., & Mlody, B. et al. (2017). Human iPSC-Derived Neural Progenitors Are an Effective Drug Discovery Model for Neurological mtDNA Disorders. *Cell Stem Cell, 20*(5), 659-674.e9. doi: 10.1016/j.stem.2016.12.013

Makarova, K., Grishin, N., Shabalina, S., Wolf, Y., & Koonin, E. (2006). *Biology Direct, 1*(1), 7. doi: 10.1186/1745-6150-1-7Marraffini L.A. & Sontheimer E.J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science 322: 1843–1845. https://doi.org/10.1126/science.1165771.

Martinez-Lage, M., Puig-Serra, P., Menendez, P., Torres-Ruiz, R., & Rodriguez-Perales, S. (2018). CRISPR/Cas9 for Cancer Therapy: Hopes and Challenges. *Biomedicines, 6*(4), 105. doi: 10.3390/biomedicines6040105

Merkle, F., Neuhausser, W., Santos, D., Valen, E., Gagnon, J., & Maas, K. et al. (2015). Efficient CRISPR-Cas9-Mediated Generation of Knockin Human Pluripotent Stem Cells Lacking Undesired Mutations at the Targeted Locus. *Cell Reports, 11*(6), 875-883. doi: 10.1016/j.celrep.2015.04.007

Mojica, F.J.M., Díez-Villaseñor, C., García-Martínez, J. & Soria, E. (2005). Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. J Mol Evol 60, 174–182.

Mojica, M.J., Juez, G., Rodríguez-Valera, F. (1993). Transcription at different salinities of Haloferax mediterranei sequences adjacent to partially modified PstI sites. *Molecular Microbiology, 9*, 613–621. Doi: 10.1111/j.1365 -2958.1993.tb01721.x.

Moulinier, L., Ripp, R., Castillo, G., Poch, O., & Sissler, M. (2017). MiSynPat: An integrated knowledge base linking clinical, genetic, and structural data for disease-causing mutations in human mitochondrial aminoacyl-tRNA synthetases. *Human Mutation, 38*(10), 1316-1324. doi: 10.1002/humu.23277

Nakata, A., Amemura, M., & Makino, K. (1989). Unusual nucleotide arrangement with repeated sequences in the Escherichia coli K-12 chromosome. *Journal Of Bacteriology, 171*(6), 3553-3556. doi: 10.1128/jb.171.6.3553-3556.1989

Nunnari, J., & Suomalainen, A. (2012). Mitochondria: In Sickness and in Health. *Cell*, *148*(6), 1145-1159. doi: 10.1016/j.cell.2012.02.035

O'Malley, J., Skylaki, S., Iwabuchi, K., Chantzoura, E., Ruetz, T., & Johnsson, A. et al. (2013). High-resolution analysis with novel cell-surface markers identifies routes to iPS cells. *Nature*, *499*(7456), 88-91. doi: 10.1038/nature12243

Ognjenović, J., & Simonović, M. (2017). Human aminoacyl-tRNA synthetases in diseases of the nervous system. *RNA Biology*, *15*(4-5), 623-634. doi: 10.1080/15476286.2017.1330245

Paquet, D., Kwart, D., Chen, A., Sproul, A., Jacob, S., & Teo, S. et al. (2016). Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature*, *533*(7601), 125-129. doi: 10.1038/nature17664

Platt, R., Chen, S., Zhou, Y., Yim, M., Swiech, L., & Kempton, H. et al. (2014). CRISPR-Cas9 Knockin Mice for Genome Editing and Cancer Modeling. *Cell*, *159*(2), 440-455. doi: 10.1016/j.cell.2014.09.014

Rahmanian, N., Bozorgmehr, M., Torabi, M., Akbari, A., & Zarnani, A. (2016). Cell separation: Potentials and pitfalls. *Preparative Biochemistry And Biotechnology*, *47*(1), 38-51. doi: 10.1080/10826068.2016.1163579

Ran, F., Hsu, P., Wright, J., Agarwala, V., Scott, D., & Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nature Protocols, 8(11), 2281-2308. doi: 10.1038/nprot.2013.143

Rees, H., & Liu, D. (2018). Base editing: precision chemistry on the genome and transcriptome of living cells. *Nature Reviews Genetics*, *19*(12), 770-788. doi: 10.1038/s41576-018-0059-1

Rivera, H., Martín-Hernández, E., Delmiro, A., García-Silva, M., Quijada-Fraile, P., & Muley, R. et al. (2013). A new mutation in the gene encoding mitochondrial seryl-tRNA synthetase as a cause of HUPRA syndrome. *BMC Nephrology*, *14*(1). doi: 10.1186/1471-2369-14-195

Robinton, D., & Daley, G. (2012). The promise of induced pluripotent stem cells in research and therapy. *Nature*, *481*(7381), 295-305. doi: 10.1038/nature10761

Rolland, S., Schneid, S., Schwarz, M., Rackles, E., Fischer, C., & Haeussler, S. et al. (2019). Compromised Mitochondrial Protein Import Acts as a Signal for UPRmt. *Cell Reports*, *28*(7), 1659-1669.e5. doi: 10.1016/j.celrep.2019.07.049

Saleh-Gohari, N. (2004). Conservative homologous recombination preferentially repairs DNA double-strand breaks in the S phase of the cell cycle in human cells. *Nucleic Acids Research*, *32*(12), 3683-3688. doi: 10.1093/nar/gkh703

Savić, N., & Schwank, G. (2016). Advances in therapeutic CRISPR/Cas9 genome editing. *Translational Research*, *168*, 15-21. doi: 10.1016/j.trsl.2015.09.008

Scheper, G., van der Klok, T., van Andel, R., van Berkel, C., Sissler, M., & Smet, J. et al. (2007). Mitochondrial aspartyl-tRNA synthetase deficiency causes leukoencephalopathy with brain stem and spinal cord involvement and lactate elevation. *Nature Genetics*, *39*(4), 534-539. doi: 10.1038/ng2013

Schimmel, P., & Schmidt, E. (1995). Making connections: RNA-dependent amino acid recognition. *Trends In Biochemical Sciences*, *20*(1), 1-2. doi: 10.1016/s0968-0004(00)88937-9

Shi, Y., Inoue, H., Wu, J., & Yamanaka, S. (2016). Induced pluripotent stem cell technology: a decade of progress. *Nature Reviews Drug Discovery*, *16*(2), 115-130. doi: 10.1038/nrd.2016.245

Shrivastav, M., De Haro, L., & Nickoloff, J. (2007). Regulation of DNA double-strand break repair pathway choice. *Cell Research*, *18*(1), 134-147. doi: 10.1038/cr.2007.111

Shrivastav, M., De Haro, L., & Nickoloff, J. (2007). Regulation of DNA double-strand break repair pathway choice. *Cell Research*, *18*(1), 134-147. doi: 10.1038/cr.2007.111

Sissler, M., González-Serrano, L., & Westhof, E. (2017). Recent Advances in Mitochondrial Aminoacyl-tRNA Synthetases and Disease. *Trends In Molecular Medicine*, *23*(8), 693-708. doi: 10.1016/j.molmed.2017.06.002

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

Slaymaker, I., Gao, L., Zetsche, B., Scott, D., Yan, W., & Zhang, F. (2015). Rationally engineered Cas9 nucleases with improved specificity. *Science*, *351*(6268), 84-88. doi: 10.1126/science.aad5227

Soldner, F., & Jaenisch, R. (2012). iPSC Disease Modeling. *Science*, *338*(6111), 1155-1156. doi: 10.1126/science.1227682

Soldner, F., Laganière, J., Cheng, A., Hockemeyer, D., Gao, Q., & Alagappan, R. et al. (2011). Generation of Isogenic Pluripotent Stem Cells Differing Exclusively at Two Early Onset Parkinson Point Mutations. *Cell*, *146*(2), 318-331. doi: 10.1016/j.cell.2011.06.019

Steinberg, S., Gautheret, D., & Cedergren, R. (1994). Fitting the structurally diverse animal mitochondrial tRNAsSer to common three-dimensional constraitns. *Journal Of Molecular Biology*, *236*(4), 982-989. doi: 10.1016/0022-2836(94)90004-3

Sterneckert, J., Reinhardt, P., & Schöler, H. (2014). Investigating human disease using stem cell models. *Nature Reviews Genetics*, *15*(9), 625-639. doi: 10.1038/nrg3764

Suomalainen, A. (2011). Therapy for mitochondrial disorders: Little proof, high research activity, some promise. *Seminars In Fetal And Neonatal Medicine*, *16*(4), 236-240. doi: 10.1016/j.siny.2011.05.003

Suomalainen, A., & Battersby, B. (2017). Mitochondrial diseases: the contribution of organelle stress responses to pathology. *Nature Reviews Molecular Cell Biology*, *19*(2), 77-92. doi: 10.1038/nrm.2017.66

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. doi: 10.1016/j.cell.2006.07.024

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. doi: 10.1016/j.cell.2007.11.019

Trokovic, R., Weltner, J., & Otonkoski, T. (2015). Generation of iPSC line HEL24.3 from human neonatal foreskin fibroblasts. *Stem Cell Research*, *15*(1), 266-268. doi: 10.1016/j.scr.2015.05.012

Tycko, J., Wainberg, M., Marinov, G., Ursu, O., Hess, G., & Ego, B. et al. (2019). Mitigation of off-target toxicity in CRISPR-Cas9 screens for essential non-coding elements. *Nature Communications*, *10*(1). doi: 10.1038/s41467-019-11955-7

Urnov, F., Rebar, E., Holmes, M., Zhang, H., & Gregory, P. (2010). Genome editing with engineered zinc finger nucleases. *Nature Reviews Genetics*, *11*(9), 636-646. doi: 10.1038/nrg2842

Vafai, S., & Mootha, V. (2012). Mitochondrial disorders as windows into an ancient organelle. *Nature*, *491*(7424), 374-383. doi: 10.1038/nature11707

van der Bliek, A., Sedensky, M., & Morgan, P. (2017). Cell Biology of the Mitochondrion. *Genetics*, *207*(3), 843-871. doi: 10.1534/genetics.117.300262

Wang, G., Yang, L., Grishin, D., Rios, X., Ye, L., & Hu, Y. et al. (2016b). Efficient, footprint-free human iPSC genome editing by consolidation of Cas9/CRISPR and piggyBac technologies. *Nature Protocols*, *12*(1), 88-103. doi: 10.1038/nprot.2016.152

Wang, M., & Kaufman, R. (2016). Protein misfolding in the endoplasmic reticulum as a conduit to human disease. *Nature*, *529*(7586), 326-335. doi: 10.1038/nature17041

Wang, Y., Liu, X., Ren, C., Zhong, G., Yang, L., Li, S., & Liang, Z. (2016a). Identification of genomic sites for CRISPR/Cas9-based genome editing in the Vitis vinifera genome. *BMC Plant Biology*, *16*(1). doi: 10.1186/s12870-016-0787-3

Watanabe, K., Ueno, M., Kamiya, D., Nishiyama, A., Matsumura, M., & Wataya, T. et al. (2007). A ROCK inhibitor permits survival of dissociated human embryonic stem cells. *Nature Biotechnology*, *25*(6), 681-686. doi: 10.1038/nbt1310

Weltner, J., Balboa, D., Katayama, S., Bespalov, M., Krjutškov, K., & Jouhilahti, E. et al. (2018). Human pluripotent reprogramming with CRISPR activators. *Nature Communications*, *9*(1). doi: 10.1038/s41467-018-05067-x

Wiegand, C., & Banerjee, I. (2019). Recent advances in the applications of iPSC technology. *Current Opinion In Biotechnology*, *60*, 250-258. doi: 10.1016/j.copbio.2019.05.011

Yaffe, M. (1999). Dynamic mitochondria. *Nature Cell Biology*, *1*(6), E149-E150. doi: 10.1038/14101

Yang, D., Scavuzzo, M., Chmielowiec, J., Sharp, R., Bajic, A., & Borowiak, M. (2016). Enrichment of G2/M cell cycle phase in human pluripotent stem cells enhances HDR-mediated gene repair with customizable endonucleases. *Scientific Reports*, *6*(1). doi: 10.1038/srep21264

Ylikallio, E., & Suomalainen, A. (2012). Mechanisms of mitochondrial diseases. *Annals Of Medicine*, *44*(1), 41-59. doi: 10.3109/07853890.2011.598547

Yoshida, Y., & Yamanaka, S. (2017). Induced Pluripotent Stem Cells 10 Years Later. *Circulation Research*, *120*(12), 1958-1968. doi: 10.1161/circresaha.117.311080

Yu, J., Vodyanik, M., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J., & Tian, S. et al. (2007). Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells. *Science*, *318*(5858), 1917-1920. doi: 10.1126/science.1151526

Yu, Y., Guo, Y., Tian, Q., Lan, Y., Yeh, H., & Zhang, M. et al. (2019). An efficient gene knock-in strategy using 5′-modified double-stranded DNA donors with short homology arms. *Nature Chemical Biology*, *16*(4), 387-390. doi: 10.1038/s41589-019-0432-1

Yusa, K., Rashid, S., Strick-Marchand, H., Varela, I., Liu, P., & Paschon, D. et al. (2011). Targeted gene correction of α1-antitrypsin deficiency in induced pluripotent stem cells. *Nature*, *478*(7369), 391-394. doi: 10.1038/nature10424

Yusa, K., Rashid, S., Strick-Marchand, H., Varela, I., Liu, P., & Paschon, D. et al. (2011). Targeted gene correction of α1-antitrypsin deficiency in induced pluripotent stem cells. *Nature*, *478*(7369), 391-394. doi: 10.1038/nature10424

Zhang, J., Li, X., Li, G., Chen, W., Arakaki, C., & Botimer, G. et al. (2017). Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage. *Genome Biology*, *18*(1). doi: 10.1186/s13059-017-1164-8

Zhou, J., & Sun, J. (2019). A Revolution in Reprogramming: Small Molecules. *Current Molecular Medicine*, *19*(2), 77-90. doi: 10.2174/1566524019666190325113945