HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Correcting a mutation causing SMAJ in patient myoblasts with CRISPR-Cas9 genome editing

Volter Lukander

Master's Thesis

Master's Degree Programme in Genetics and Molecular Biosciences

Faculty of Biological and Environmental Sciences

University of Helsinki

2022

Abstract

Tiedekunta – Fakultet – Facult	Koulutusohjelma – Utbildningsprogram – Degree Programme
Faculty of Biology and Environmental Sciences	Genetics and Molecular Biosciences MSc.

Tekijä – Författare – Author
Volter Lukander

Työn nimi – Arbetets titel – Title
Correcting a mutation causing SMAJ in patient myoblasts with CRISPR-Cas9 genome editing

Oppiaine/Opintosuunta – Läroämne/Studieinriktning – Subject/Study track
Genetics and Molecular Biosciences, specialization in Genetics and Genomics

Työn laji – Arbetets art – Level	Aika – Datum – Month and year	Sivumäärä – Sidoantal – Number of pages
Master's thesis	May 2022	54

Tiivistelmä – Referat – Abstract

Spinal muscular atrophy of Jokela type (SMAJ) is an autosomal dominant motor-neuron disease caused by a missense mutation c.197G>T, p.G66V in the gene CHCHD10. Coiled-coil-helix-coiled-coil-helix domain-containing protein 10 (CHCHD10) is a nuclear-encoded mitochondrial protein located in the intermembrane space (IMS) of mitochondria with an unknown exact function and disease-causing mechanism. In this project, the overarching aim was to correct a heterozygous SMAJ-causing mutation in patient myoblast cells with CRISPR-Cas9 genome editing. The goal was to create a genetically identical, isogenic, cell line to study only the effects of the mutation on cellular phenotype *in vitro*. Human myoblast cells isolated from patient biopsies provide the most pertinent experimental model to study neuromuscular atrophy-associated mutations in their natural genomic environment.

More specific aims included genome editing optimization with myoblast cells, since it is not as widely conducted as with some other cell types, such as iPSCs. CRISPR-Cas9 ribonucleoprotein (RNP) complex and associated donor template were used to induce homology-directed repair (HDR) in the genome of patient-derived myoblast cells and correct the mutation. After optimization of electroporation conditions for myoblast cells, guide RNAs were designed and transfected into patient myoblasts. Clonal cell lines were made by utilizing techniques such as fluorescence adjusted cell sorting (FACS) and manual colony picking. The success and precision of genome editing were analyzed by Sanger sequencing, comparing the performance of the different guide RNAs with restriction enzyme analysis and Synthego ICE CRISPR web tool, and screening regions of potential off-target genome editing.

A genome-edited myoblast cell line with the CHCHD10 c.197G>T mutation corrected, was successfully generated to provide an isogenic control for the patient myoblast cell line. Optimization of myoblast electroporation was successful and conditions used proved to be effective. Clonal cell line creation proved to be challenging with myoblast cells and work is still needed to improve the viability of single-cell clones after FACS. Nevertheless, the advances taken here regarding myoblast genome editing with CRISPR-Cas9 offer a fertile avenue for future research of myoblasts genome manipulation, myogenic disorders, and the role of CHCHD10 in skeletal muscle and SMAJ. Comparing the CHCHD10 protein level and mRNA expression between patient cells, corrected myoblasts, and differentiated myotubes is an area of future research. Future work also includes measuring the mitochondrial integrated stress response in both cell lines and co-culturing myotubes and iPSC derived motor neurons to study the effects of p.G66V on neuromuscular junction (NMJ) formation.

Avainsanat – Nyckelord – Keywords

CRISPR-Cas9, genome editing, disease modeling, CHCHD10, SMAJ, mitochondrial disease, myoblast

Ohjaaja tai ohjaajat – Handledare – Supervisor or supervisors

Henna Tyynismaa PhD, Associate Professor
Sandra Harjuhaahto, MSc
Tiina Rasila, PhD

Säilytyspaikka – Förvaringsställe – Where deposited
HELDA - Helsingin yliopiston digitaalinen arkisto / HELDA - Helsingfors universitets digitala publikationsarkiv /HELDA - Digital Repository of the University of Helsinki

Muita tietoja – Övriga uppgifter – Additional information

Tiedekunta – Fakultet – Facult	Koulutusohjelma – Utbildningsprogram – Degree Programme
Bio- ja ympäristötieteellinen tiedekunta	Genetiikan ja molekulaaristen biotieteiden maisteriohjelma

Tekijä – Författare – Author

Volter Lukander

Työn nimi – Arbetets titel – Title
SMAJ-tautia aiheuttavan mutaation korjaus potilaan lihaksen kantasoluissa CRISPR-Cas9 geenieditointitekniikalla

Oppiaine/Opintosuunta – Läroämne/Studieinriktning – Subject/Study track
Genetiikka ja genomiikka

Työn laji – Arbetets art – Level	Aika – Datum – Month and year	Sivumäärä – Sidoantal – Number of pages
Maisterin tutkielma	Toukokuu 2022	54

Tiivistelmä – Referat – Abstract

Vuonna 2011 Suomesta löydettiin uusi perinnöllinen sairaus, jonka todettiin olevan lähtöisin Pohjois-Karjalan alueelta ja siten osa suomalaista tautiperimää. Taudin nimi on Jokela tyypin spinaalinen lihasatrofia (Spinal Muscular Atrophy of Jokela Type = SMAJ) ja sitä sairastaa yhteensä arviolta 200–400 potilasta. SMAJ on autosomaalisesti vallitseva, alempien liikehermosolujen rappeumasairaus, jonka aiheuttaa pistemutaatio c.197G>T geenissä *CHCHD10* ja yhden aminohapon vaihtuminen p.G66V vastaavassa proteiinissa. CHCHD10 on saman nimisestä geenistä tuotettu proteiini, jota esiintyy mitokondrioiden ulko- ja sisäkalvon välissä. Sen tarkkaa toimintaa ei tunneta, eikä miten mutaatio siihen vaikuttaa. Taudin aiheuttavan mekanismin selvittäminen on kuitenkin olennaisen tärkeää, jotta mahdollisia hoitoja ja diagnostisia testejä on mahdollista kehittää. Tutkielman tavoite on korjata tämä mutaatio heterotsygootin SMAJ-potilaan myoblasteissa (lihassolun esiaste) käyttämällä CRISPR-Cas9 geenieditointi teknologiaa. Korjaamalla mutaatio, voidaan luoda solulinja, joka on muuten täysin identtinen potilaan solujen kanssa, mutta mutaatio geenissä CHCHD10 on korjattu vastaamaan normaalia. Vertailemalla tätä solulinjaa potilaan soluihin, voidaan selvittää ainoastaan mutaatiosta johtuvat erot soluissa.

Tutkielmassa pyritään selvittämään taudin aiheuttavan mutaation vaikutuksia ihmisen lihassoluissa. Myoblasti-solujen CRISPR-Cas9 geenieditointi ei ole yleistä, sillä usein vastaavan kaltaisia tutkimuksia tehdään indusoiduilla kantasoluilla, joista kohdekudoksen soluja on helppo erilaistaa. Nyt käytettävissä oli kuitenkin potilaan tutkimuskäyttöön luovuttamia lihaksen kantasoluja ja oli mielenkiintoista tutkia mitokondriaalisen CHCHD10 proteiinin aiheuttamaa tautia mitokondriorikkaissa lihassoluissa. CRISPR-Cas9 ribonukleproteiinikompleksia (RNP) ja sitä vastaavaa korjaustemplaattia käytettiin mutaation korjaamisessa. RNP-kompleksi transfektoitiin soluihin elektroporaation avulla, jota ennen elektroporaatio-olosuhteet optimoitiin lihassoluille edullisiksi. Elektroporoitujen solulinjojen geenieditoinnin onnistuminen arvioitiin sekä restriktioentsyymianalyysin, että Synthego ICE CRISPR internettyökalun avulla. Klonaaliset solulinjat luotiin fluoresenssiavusteisella solun lajittelu teknologialla (FACS) ja manuaalisesti poimimalla kolonioita solu maljoilta. Kloonien genotyypit selvitettiin Sanger sekvensoimalla ja arvioitiin off-target geenieditoinnin varalta.

Yksi korjattu solulinja saatiin valmistettua ja geenieditointiprosessin optimisaatio myoblasti-soluille onnistui. Tuotetun isogeenisen solulinjan avulla voidaan tulevaisuudessa tutkia CHCHD10:n proteiini- ja mRNA-tasojen eroja verrattuna potilassolulinjaan, ja näin saada arvokasta informaatiota sairauden vaikutuksista lihassoluissa. Tulevaisuuden tehtäviin lukeutuu myös mutaation vaikutuksien hermolihasliitokseen tutkiminen indusoiduista kantasoluista erikoistettujen hermosolujen avulla. Yhteissolukulttuurit mahdollistavat editoitujen myoblastien, sekä hermosolujen kasvattamisen ja hermolihasliitoksen tutkimisen *in vitro*. Hermolihasliitos on tärkeä yhdistäjä alempien liikehermosolujen ja luustolihasten välissä ja mutaation vaikutuksen tutkiminen voisi valaista SMAJ-tautimekanismia entisestään.

Avainsanat – Nyckelord – Keywords
CRISPR-Cas9, geenieditointi, CHCHD10, SMAJ, mitokondriaalinen sairaus, myoblasti, kantasolu, hermorappeumasairaus, spinaalinen lihasatrofia

Ohjaaja tai ohjaajat – Handledare – Supervisor or supervisors
Henna Tyynismaa PhD, Vastuuprofessori
Sandra Harjuhaahto, MSc
Tiina Rasila, PhD

Säilytyspaikka – Förvaringsställe – Where deposited
HELDA - Helsingin yliopiston digitaalinen arkisto / HELDA - Helsingfors universitets digitala publikationsarkiv /HELDA - Digital Repository of the University of Helsinki

Muita tietoja – Övriga uppgifter – Additional information

Table of Contents

Abbreviations

ALS	Amyotrophic lateral sclerosis
Cas	CRISPR associated protein
CHCHD10	Coiled-coil-helix-coiled-coil-helix domain-containing protein 10
CHCHD2	Coiled-coil-helix-coiled-coil-helix domain-containing protein 2
CMT	Charcot-Marie-Tooth disease
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CRISPR-Cas9	Clustered Regularly Interspaced Short Palindromic Repeats – CRISPR associated protein 9
crRNA	CRISPR RNA
dsODN	Double-stranded donor oligonucleotide
FACS	Fluorescent activated cell sorting
GFP	Green fluorescent protein
gRNA	guide RNA
HDR	Homology-directed repair
HR	Homologous recombination
IMM	Inner mitochondrial membrane
IMS	Mitochondrial intermembrane space
iPSC	Induced pluripotent stem cell
MICOS	Mitochondrial contact site and cristae organizing system
mtISR	Mitochondrial integrated stress response
NHEJ	Nonhomologous end-joining
NMJ	Neuromuscular junction
PAM	Protospacer adjacent motif
RFP	Red fluorescent protein
RNP	Ribonucleoprotein
sgRNA	Single guide RNA
SMA	Spinal muscular atrophy
SMAJ	Spinal muscular atrophy of Jokela type
ssODN	single-stranded donor oligonucleotide
tracrRNA	Trans-activating CRISPR RNA

1. Introduction

1.1 Spinal muscular atrophy of Jokela type

Spinal muscular atrophy of Jokela type (SMAJ, MIM#615048) is an autosomal dominant late-onset motor neuron disease, caused by a missense mutation in the gene *CHCHD10*. The mutation is a single base-pair substitution c.197G>T in the second exon of the gene, which changes the amino acid glycine to valine in the 66th position of the resulting protein (p.G66V). The disease was originally known as late-onset spinal motor neuronopathy (LOSMoN) and was characterized in 2011 by Jokela et al. in two families from Northeastern Finland. In 2015, the cause of the disease was located in *CHCHD10* in the locus 22q11.23, and the disease-causing c.197G>T mutation was found in all 55 studied patients from 17 different families (Penttilä et al., 2015). The discovery was surprising, since CHCHD10 is a mitochondrial protein, and no previous findings had indicated significant mitochondrial dysfunction in SMAJ patients. All documented families with SMAJ had identical disease-associated haplotypes, implying a founder effect in the Finnish population (Penttilä et al., 2014). This makes SMAJ the newest member of Finnish heritage diseases, which is a group of around 40 hereditary diseases significantly more common in Finland than in other parts of the world (Norio, 2003).

SMAJ is a moderately benign autosomal dominant type of spinal muscular atrophy (SMA). It is a late-onset disease, so most of the symptoms occur later in life, at the age of 30 to 40. The symptoms include cramps, fasciculations, lowered or absent tendon reflexes, elevated levels of creatine kinase, and hand tremors. The disease is relatively slow progressive and leads to weakness and muscular atrophy at a later age. Patients have been reported to maintain walking ability after decades of diagnosis and they typically retain normal life expectancy. The prevalence of SMAJ in Finland is 2-4 cases in 100 000, making it one of the most prevalent Finnish heritage diseases (Jokela *et al.*, 2011; Penttilä *et al.*, 2015). Early-onset spinal muscular atrophy (SMA) is typically caused by a loss of function (LOF) mutation in the *SMN1* gene, but at least 17 less common mutations have also been associated with SMA. The genetic background of spinal muscular atrophies is heterogeneous and can be caused by several different mutations in different genes. SMA genes have a high clinical overlap with other neuromuscular diseases such as hereditary spastic paraplegia, amyotrophic lateral sclerosis (ALS), inherited motor neuropathies, and Charcot–Marie–Tooth disease (CMT) (Peeters et al., 2014; Rossor et al., 2012). *CHCHD10* is the first SMA-causing gene, which encodes for a mitochondrial protein (Penttilä *et al.*, 2015).

1.2 CHCHD10 and CHCHD2

Coiled-coil-helix-coiled-coil-helix domain-containing protein 10 (CHCHD10) is a nuclear-encoded mitochondrial protein situated in locus 22q11.23 (Bannwarth et al., 2014). Its exact function is currently unknown, but it is known to form homodimers with itself and heterodimers with homologous protein CHCHD2 (Huang et al., 2018). CHCHD10 and CHCHD2 belong to the mitochondrial coiled-coil-helix-coiled-coil-helix (CHCH) domain protein family characterized by twin CX9C motifs with two cysteine residues separated by nine amino acids and they share a 54% similarity in amino acid sequence (Imai et al., 2019). By 2017, there were nine members identified in the CHCHD protein family (Zhou et al., 2017). Mutations in *CHCHD10* have been linked to motor neuron phenotypes such as ALS, CMT, and SMAJ, whereas mutation in CHCHD2 has been known to cause Parkinson's disease (Auranen et al., 2015; Bannwarth *et al.*, 2014; Funayama et al., 2015). CHCHD2 and CHCHD10 originated from a common ancestor through a gene duplication event predating human speciation and both proteins are conserved in all metazoans (Cavallaro, 2010; Longen et al., 2009). Both genes are expressed in the nucleus and the proteins are imported to the intermembrane space of mitochondria (IMS) via the mitochondrial intermembrane space import and assembly (MIA) pathway by the protein MIA40 (Lehmer et al., 2018).

CHCHD10 functions in the IMS of mitochondria and is enriched at the cristae junctions. It has been speculated to be involved in the maintenance of cristae morphology and mitochondrial DNA (mtDNA) stability. Mutations in the *CHCHD10* are known to cause defects in mitochondrial dynamics and cristae formation (Bannwarth *et al.*, 2014). CHCHD10 is highly expressed in tissues with high quantities of mitochondria such as skeletal muscle, heart, and liver, and is present in all tissues, although in very low amounts in some of them (Ajroud-Driss et al., 2015). Because CHCHD10 was found enriched at cristae junctions and its neurodegeneration-related mutations caused cristae abnormalities, it was previously thought to be part of the mitochondrial contact site and cristae organizing system (MICOS) complex, which is crucial for mitochondrial membrane architecture and cristae organization (Genin et al., 2016). A recent study challenges this view by arguing that CHCHD10 mutations associated with neurological diseases or the deletion of CHCHD2 and CHCHD10 resulted in cristae abnormalities due to increased optic atrophy 1 (OPA1) processing, while a single knockout of either of the proteins did not induce such effect (Liu et al., 2020). CHCHD10 and CHCHD2 proteins are partially overlapping in function and can

compensate for each other when knocked down (Harjuhaahto et al., 2020; Huang *et al.*, 2018). This explains the results of many studies where single knockouts/downs of either protein did not result in significant effects, but double knockouts did (Anderson et al., 2019; Liu *et al.*, 2020; Ruan et al., 2022).

Recently, CHCHD10 and CHDCHD2 have been shown to interact with metalloendopeptidase OMA1 and suppress its enzyme activity, which restrains the initiation of the mitochondrial integrated stress response (mtISR) and suppresses the processing of OPA1 involved in the mitochondrial fusion of the inner mitochondrial membrane (IMM) (Ruan *et al.*, 2022). The suppressing role of CHCHD proteins to mtISR has been also shown before (Harjuhaahto *et al.*, 2020). OMA1 is a mitochondrial protease involved in the processing of OPA1 after entering mitochondria (Song et al., 2007). OPA1, on the other hand, is a dynamin-like GTPase, which mediates the fusion of IMM and plays a role in cristae remodeling (Mishra and Chan, 2014). The failure of fusing mitochondria properly leads to faulty mitochondrial dynamics and mitochondrial dysfunction, which is linked to neurodegenerative diseases (Malpass, 2013). Mitochondrial dysfunction triggers the integrated stress response and it can be concluded that CHCHD10 acts as a mtISR suppressor, regulating stress responses in mammalian cells (Ruan *et al.*, 2022). This has a significant pathological impact on CHCHD10/CHCHD2-linked neuromuscular and neurodegenerative diseases.

1.3 Myoblast cells and the role of skeletal muscle in SMAJ

Myoblast cells are myogenic stem cells that arise from mesodermal progenitor cells and differentiate into myotubes by fusing together. After myotube formation, they maturate into myofibers that build up a muscle, and muscle innervation occurs through the newly assembled neuromuscular junction (NMJ). Myofibers are elongated, multinuclear functional units of skeletal muscle (Hall and Sanes, 1993). Cultured myoblasts are a useful tool for researching cell biology and disease processes in the context of neuromuscular diseases such as SMAJ. In this project, myoblast lines used were obtained by muscle biopsy from SMAJ and non-SMAJ patients and immortalized as previously described (Mamchaoui et al., 2011). Immortalized cell lines rather than primary cells were used as they are easier to maintain and produce more consistent experimental results, although they may not fully represent *in vivo* conditions (Obinata, 2007).

SMAJ is known to be caused by the CHCHD10 p.G66V/+ mutation, but other than that, the exact disease mechanism of SMAJ is still unknown. Under normal conditions, CHCHD10 is highly expressed in

mitochondria-rich tissues such as skeletal muscle, which makes investigating the role of the mutation in muscle cells interesting (Martherus et al., 2010). SMAJ is a neurogenerative disease affecting lower motor neurons, cells that control voluntary muscles of the body via NMJ. The G66V/+ mutation impairs the function of CHCHD10 which likely causes OXPHOS deficiency, loss of cristae junctions, and destabilization of internal membrane structure within mitochondria at the muscle end of NMJ, therefore implying that muscle CHCHD10 is required for normal neurotransmission between motoneuron and skeletal muscle fibers (Genin et al., 2019). CHCHD10 malfunction in muscle cells can be the cause of NMJ impairment and the degradation of motor neurons (Xiao et al., 2020). This suggests a key role for muscle in motor neuron disease associated with CHCHD10 mutations, such as SMAJ.

In this study, the object was to create isogenic cell lines by correcting the disease mutation in patient myoblast cells with CRISPR-Cas9 genome editing to study the effects of the CHCHD10 mutation in muscle and further in NMJ formation. Despite the minute invasiveness of induced pluripotent stem cells (iPSC) to patients, myoblast cells were chosen for the project since the target was to elucidate CHCHD10 function in muscle cells. Further, it was convenient to use existing myoblast cultures, rather than using myoblasts derived from iPSCs. Human myoblasts isolated from patient biopsies, despite being immortalized, provide the most pertinent experimental model to assess the role of the mutation in its natural genomic environment.

1.4 Isogenic cell lines

Isogenic cell lines are essentially genetically identical. Genetically matched (isogenic) cell lines are only distinguished by the change induced within the targeted locus. Such cell lines create a highly controlled environment in which any phenotypic variation is more attributable to that specific change (Avior et al., 2016). Therefore, the creation of isogenic cell lines is an optimal strategy for studying monogenic disorders such as SMAJ. The creation of isogenic cell lines can take one of two routes: induction or correction of relevant mutations. Induction of a supposedly significant disease-causing gene variant verifies whether this single genetic change can generate the disease phenotype by removing it from the affected individual's genetic background. However, if no influence on the molecular or cellular phenotype of interest is observed, it is impossible to determine if this allele contributes to the disease. Correction includes the repairing of a pre-existing, suspected causal mutation in a cell line obtained from a disease patient. This determines if a genetic variation contributes to the disease phenotype, but not

whether it is sufficient to produce the disease. However, if the ability of the mutation to cause the disease is already established, it does not matter. It also has the significant advantage of expressing whatever cellular or molecular phenotype is causing the disease in the patient-derived cell line, and so reversion of this phenotype in the altered line can serve as feedback (Bassett, 2017).

Recent advances in genome engineering have dramatically improved the ability to induce such specified modifications into the genomes of human cells. In this project, patient-derived myoblasts carrying the disease mutation are corrected with the CRISPR-Cas9 gene editing system (**Figure 1**). For the obtaining of isogenic cell lines, the lines must originate from a single cell with the desired engineered characteristics to make sure the cells are identical clones. This means, that the edited cell culture must be isolated into single cells. A common method for achieving this is fluorescent activated cell sorting (FACS), which sorts a heterogenous cell population one cell at a time depending on its fluorescence properties and particular refraction of light (Bonner et al., 1972). Genome edited cells containing a fluorescence tag such as a Green fluorescent protein (GFP) -marker, are separated from the cells that do not have the target characteristics (Cormack et al., 1996). FACS can also be utilized to sort cell populations into single cells without a fluorescence signal. An alternative method for isolating single cells from heterogeneous cell populations is colony picking, where cells are sorted manually for *in vitro* disease modeling purposes, by picking colonies originating from individual cells from cell culture dishes. The advantage of manual colony picking is its accuracy and potential higher survival rate of isolated single cells (Bruntraeger et al., 2019).

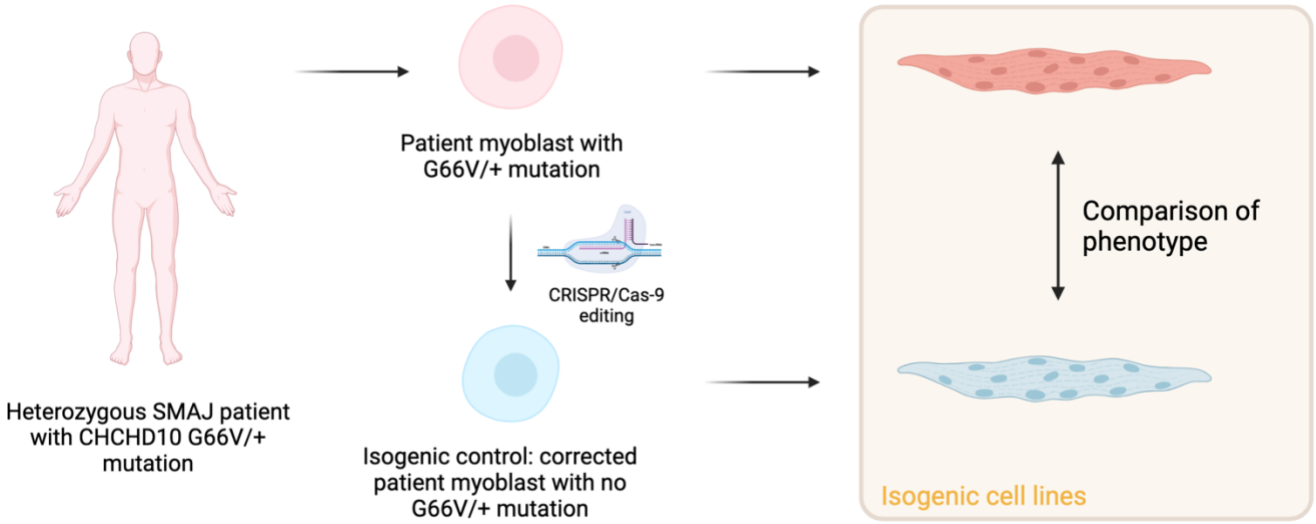

Figure 1. Isogenic myoblast cell lines. In this project, myoblast cells were derived from an SMAJ patient. The heterozygous mutation p.G66V/+ is corrected using the CRISPR-Cas9 genome editing technology, thus resulting in isogenic pairings of cell lines that reveal the real impact of the engineered alteration on cellular phenotype. Molecular and cellular phenotypes can be compared in the resulting myoblasts and differentiated myotubes. (Figure created with BioRender.com).

1.5 CRISPR

1.5.1 From a Microbial Repeated Sequence to Genome Editing Technology

Since its discovery almost 30 years ago, CRISPR has become the most widely used gene editing tool and is now used by thousands of research groups (and home laboratories) all over the world. CRISPR is an adaptive immune system used by prokaryotes to defend against invading viruses by recording and targeting their DNA sequences. This system uses CRISPR-associated proteins (Cas) to cut viral DNA and include snippets of it into the host genome so that the next time the same virus invades the microbe, it can transcribe small RNA molecules from these snippets and guide the Cas-protein cutting machinery to disable the invading viral DNA. The included snippets are called spacers in the CRISPR array of the microbe. These spacers are regularly interspaced by short palindromic DNA repeats, which the system obtains its name: CRISPR-Cas stands for "clustered regularly interspaced short palindromic repeats-CRISPR associated systems" (Jansen et al., 2002).

CRISPR was originally discovered in the archaeal microbe *Haloferax mediterranei* from a hypersaline environment by Francesco Mojica when investigating how growth media affect how restriction enzymes cut the microbe's genome (1993). After this, similar repeat sequences were also found from other closely related microbes and Mojica encountered a paper reporting the same structures in eubacteria *Escherichia coli* (Ishino et al., 1987). A paper reporting this new class of repeats in microbes was published in 1995 (Mojica et al.). CRISPR was found to be a prokaryotic adaptive immune system after spacer sequences were discovered to align with viral DNA, although immunity was initially thought to be conferred by RNA interference, which eventually proved to be wrong (Bolotin et al., 2005; Mojica et al., 2005). Experimental proof that CRISPR confers immunity was provided later by Barrangou, Fremaux et al. They also characterized the roles of two Cas proteins: Cas7 and Cas9. Cas7 was found to be needed to gain resistance but not to retain it, Cas9, on the other hand, was earlier found to include HNH and RuvC nuclease motifs and therefore suspected to be involved in cutting genetic material (Bolotin, Quinquis et

al. 2005). Cas9 was needed to retain resistance to invading bacteriophages, therefore being the active component of the bacterial immune system (Barrangou et al., 2007).

In 2008, Cas9 was proven to target DNA instead of RNA, as earlier hypothesized (Marraffini and Sontheimer, 2008). Further light was shed on how CRISPR systems work when it was demonstrated that Cas9 cuts DNA at specific points defined by the crRNA (CRISPR RNA, encoded from the spacer sequences) upstream of a protospacer adjacent motif (PAM) sequence in the viral genome (Garneau et al., 2010). When examining the RNA profile of the bacterial species *Streptococcus pyogenes*, Emmanuel Charpentier discovered trans-activating CRISPR RNA (tracrRNA), which was found to be essential for processing crRNA (Deltcheva et al., 2011). Later, Charpentier joined forces with Jennifer Doudna and elaborated the function of tracrRNA to include a complex formation with crRNA and Cas9 protein to enable DNA cleaving. This effort also led to the discovery that Cas9 could cut purified DNA *in vitro*, that it could be programmed with custom-designed crRNAs, and that the two nuclease domains, HNH and RuvC, cut opposite strands: complementary and non-complementary, respectively. In addition, they demonstrated that crRNA and tracrRNA could be fused into one single guide RNA (sgRNA) for easier programmability (Jinek et al., 2012). In the paper, Emmanuel Charpentier and Jennifer Doudna acknowledged for the first time the possibility to use the CRISPR technology for genome editing and earned the Nobel Prize in Chemistry for their work in 2020.

The last step before CRISPR could take over the research world was to make it work in mammalian cells. This was achieved in 2013 by three independent research groups. They established that CRISPR-Cas9 genome engineering could be performed in human and mouse cells when single guide RNAs (sgRNAs) were fused with 3' hairpin loops and by adding nuclear localization signals to the CRISPR-Cas9 complex (Cong et al., 2013; Jinek et al., 2013; Mali et al., 2013). After the publication of these papers, research on CRISPR began to rise steeply and has not slowed down since. The CRISPR revolution brought with it the interest in potential applications in human therapeutics, commercial agriculture, and countless other fields, as well as the ethical concerns of human germline editing, designer babies, and the future of humanity.

1.5.2 Types of CRISPR systems and bacterial immunity mechanisms

The CRISPR adaptive immune system is widely distributed in the prokaryotic life forms and is present in 42% of bacteria and 85% of archaea (Koonin and Makarova, 2019). CRISPR immunity response

comprises three main phases (Nussenzweig and Marraffini, 2020). The first phase is the naïve spacer acquisition or immunization phase, where a new spacer sequence is integrated into the host genome, which is created based on a protospacer sequence of invading mobile genetic element e.g., bacteriophage nucleic acid sequence (**Figure 2a**). In the second phase long, precursor CRISPR RNA (crRNA) is transcribed from the CRISPR array and further processed into small crRNAs which contain the sequences of the invaders. The crRNAs combine with Cas nucleases and other RNA molecules to recognize and destroy the invader nucleic acids. This phase is called the targeting phase (**Figure 2b**). Because CRISPR immunity is sequence-specific, organisms relying on CRISPR-based immunity systems have developed a third phase, known as the primed spacer acquisition phase (**Figure 2c**). This is necessary because invading bacteriophages can acquire a mutation in the sequence, which is originally recognized by the bacteria, thus preventing immune recognition. In this phase, new spacers are acquired during target destruction and integrated into the CRISPR array (Nussenzweig et al., 2019).

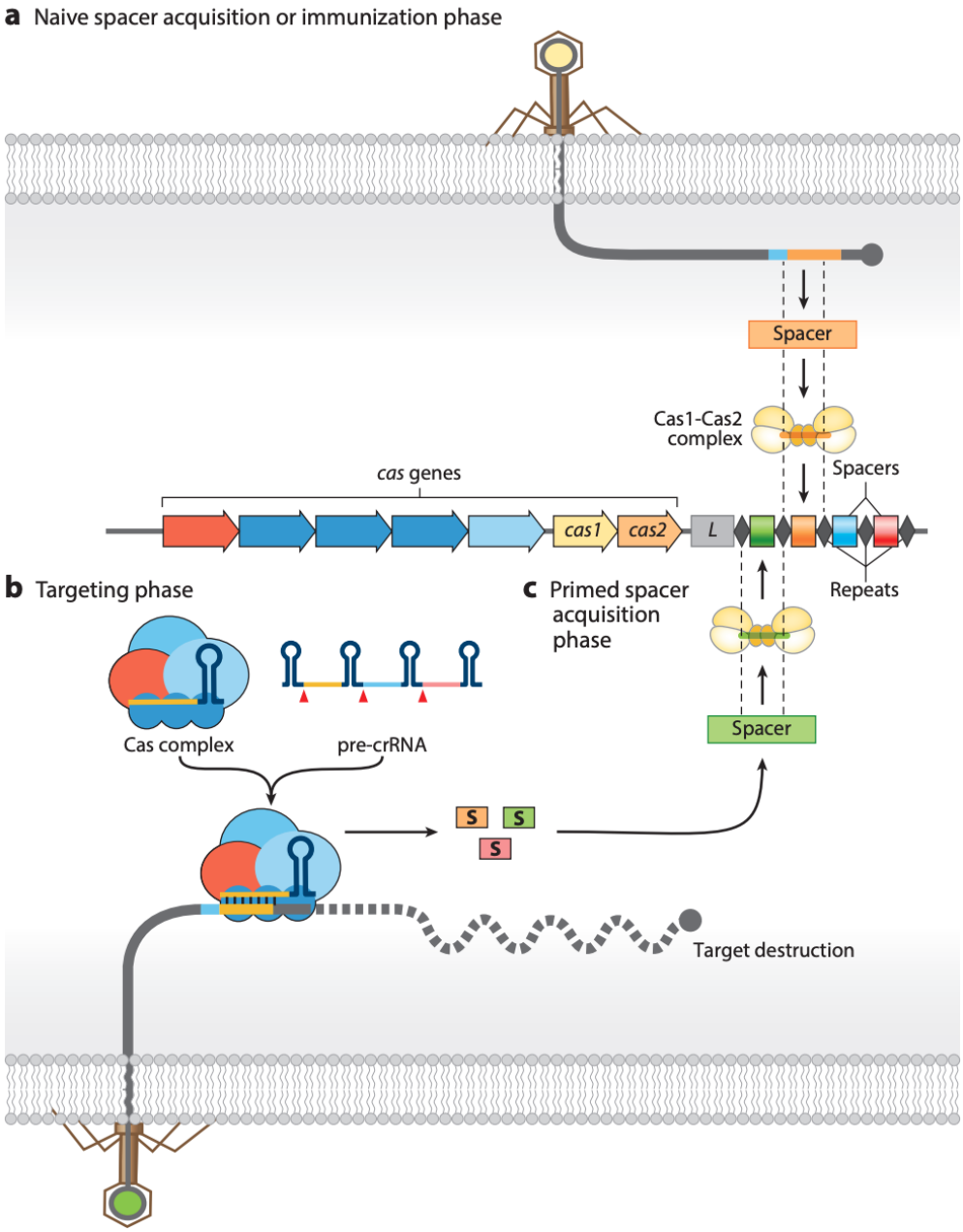

Figure 2. The three phases of CRISPR-Cas immune response. The CRISPR operon contains the CRISPR-associated (Cas) genes (red and blue arrows) including the Cas1-Cas2 spacer acquisition machinery (yellow and orange arrows), and the CRISPR array comprising palindromic repeat sequences (dark grey diamonds) and the spacer sequences (colored rectangles). **(a)** The first phase of the CRISPR adaptive immune system is the naïve spacer acquisition phase. In this phase, a spacer is determined (orange spacer) and integrated into the CRISPR array by the Cas1-Cas2 complex. **(b)** The targeting phase starts with the biogenesis of crRNAs from a long precursor CRISPR RNA (pre-crRNA), which is encoded by the newly immunized CRISPR array. The mature crRNAs merge with the Cas targeting complex to find and eliminate harmful nucleic acids in a sequence-specific way. **(c)** New substrates for spacer acquisition are created together with target destruction and inserted into the CRISPR array during

the primed spacer acquisition phase (green spacer). (Figure from (Nussenzweig and Marraffini, 2020), copyright acquired from Annual Reviews of Genetics).

The sequence and gene organization of CRISPR–Cas genomic loci are extremely diverse. Therefore, CRISPR systems are organized into two main classes, six types (I-IV) and 50 subtypes within them (Koonin and Makarova, 2019). Classes are formed based on the CRISPR system's use of either multi-subunit RNA-guided Cas complexes, i.e., effector complexes, or a single crRNA-Cas ribonucleoprotein complex. Multi-subunit effector complexes, Type I, III, and IV systems are classified as Class 1 systems, whereas single-subunit effector Type II, V, and VI systems are designated as Class 2 (Makarova et al., 2015). To avoid targeting themselves, type V and VI systems recognize the protospacer adjacent motif (PAM) sequence upstream of the protospacer sequence, while type II recognizes the PAM downstream of the protospacer (Hille and Charpentier, 2016). Type II is defined by its Cas9 protein and is the classic CRISPR type repurposed for genome engineering (Ishino et al., 2018).

The class 2 type II bacterial CRISPR-Cas9 locus contains, from upstream to downstream, the sequence encoding for tracrRNA, the *Cas9* gene first in the *cas* operon, and the CRISPR array. Upon infection, the Cas9 endonuclease combines with the crRNA and the tracrRNA, and the complex cuts the invading genetic material at a location determined by a short motif (PAM) juxtaposed to the complementary region in the target DNA, as well as base-pairing complementarity between the crRNA and the target protospacer sequence. Cas9 cleaves the DNA three base pairs upstream of the PAM sequence. The Cas9 protein consists of two regions: the NUC (nuclease) lobe and the REC (recognition) lobe (**Figure 3**). The REC lobe is responsible for identifying the DNA sequence and the NUC lobe contains the endonuclease domains HNH and RuvC and the PAM interacting domain (PI) which binds to the PAM sequence of the invading DNA. Each Cas nuclease has a characteristic PAM sequence that they recognize. With Cas9, it is usually in the form of 5'-NGG-3' (Jinek *et al.*, 2012).

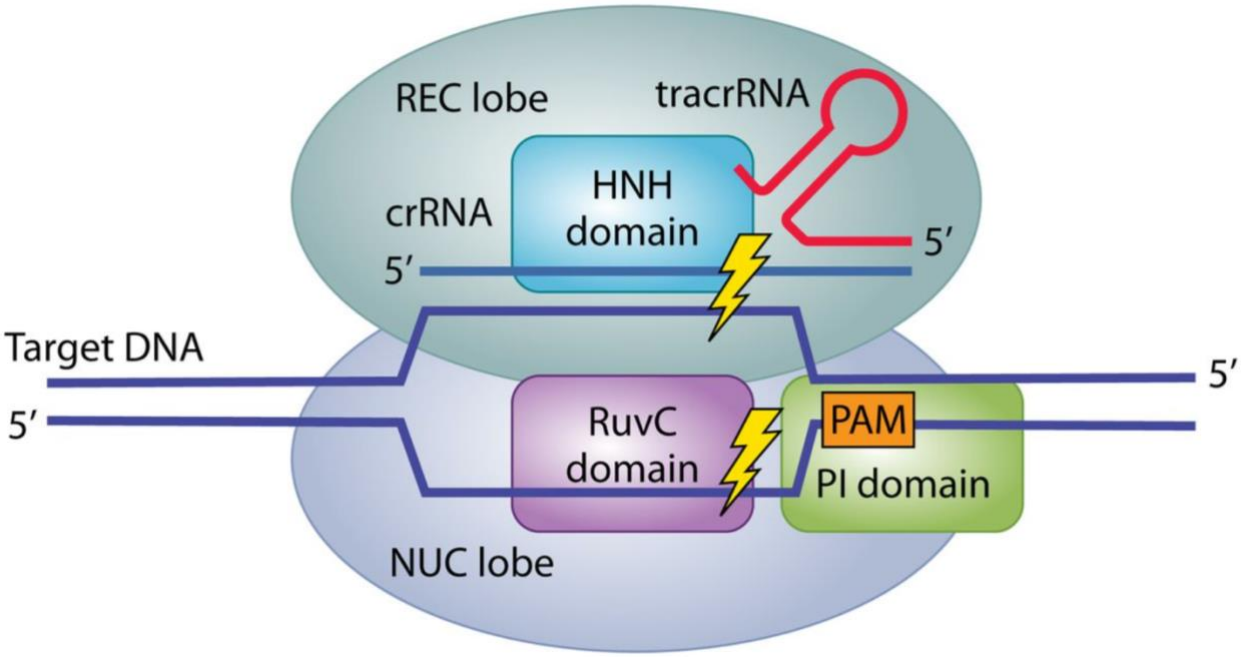

Figure 3. The CRISPR-Cas9 complex. The Cas9 protein contains two lobes: the REC lobe and the NUC lobe. The REC lobe recognizes the nucleic acid sequence. The NUC lobe contains the two endonuclease domains: HNH and RuvC, and the PAM interacting (PI) domain which binds to the PAM sequence and initiates binding to the target DNA. The two endonuclease domains cleave the DNA strand, producing a duplex with crRNA and the other DNA strand, respectively, causing a double-strand break in the target DNA. (Figure from (Ishino *et al.*, 2018), copyright acquired from ASM Journal of Bacteriology).

1.5.3 Genome engineering with CRISPR-Cas9

Genome editing, also known as genome engineering or gene editing, is a sort of genetic engineering that involves inserting, deleting, modifying, or replacing DNA in a living organism's genome (Bak et al., 2018). Due to its easy programmability and single sub-unit effector protein Cas9, CRISPR-Cas9 has become the most widely used tool for genome engineering in biomedical research. It is used to study the role of specific genes and demonstrate potential treatment strategies in a variety of genetic and infectious disease models. Like in this project, CRISPR-Cas9 can be used to treat monogenic illnesses by repairing mutations *ex vivo* in somatic or induced pluripotent stem cells (Savić and Schwank, 2016).

CRISPR-Cas9 can function as a programmable restriction enzyme. The discovery that crRNA and tracrRNA can be fused together to act as a single guide RNA (sgRNA) made the system very easy to use and requires only the designing of a single RNA molecule to program the gene editing machinery to

work in a wide range of different organisms, including plants, animals, and microbes (Jinek *et al.*, 2012). The sgRNA needs to match the desired DNA targeting site adjacent to PAM. The NGG PAM motif appears once per 8 bp on average in the genome, making it possible to target nearly any gene of interest (Wang et al., 2016). Also, a variety of Cas9s with modified PAM sequences have been created as a result of engineering the existing Cas9s, thereby increasing the targetable regions within the mammalian genome (Kleinstiver et al., 2015). Before CRISPR, genome editing required the designing and production of specific proteins for every genomic site to be edited. Such tools included zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) (Bogdanove and Voytas, 2011; Gaj et al., 2013). They were the first-generation tools of genome editing and were the first tools used to modify the DNA of mouse and human pluripotent stem cells (Urnov et al., 2010). These mechanisms were relatively labor-intensive and time-consuming. With CRISPR, genome editing advanced to a new era. The easy programmability of CRISPR with sgRNAs and the wide range of different CRISPR types and Cas proteins quickly made it the leading technique of gene editing in the research world and allowed for the development of a variety of applications.

Once in the target site, the Cas9 generates a double-stranded break (DSB), which initiates DNA repair by endogenous cellular mechanisms such as nonhomologous end joining (NHEJ) or homology-directed repair (HDR) (Rudin et al., 1989). NHEJ induces random insertion and deletion changes (INDELs) at the DSB site, which might result in gene knockout by shifting the target gene's reading frame or changing a key region of the encoded protein. HDR uses a homologous repair template to fix the DSB. The most common form of HDR in human cells is homologous recombination (HR), where sister chromatid is used as a repair template. This mechanism can be exploited by providing an exogenous repair template, in the form of single-stranded donor oligonucleotide (ssODN), facilitating the introduction of precise gene modifications (**Figure 4**) (Lieber, 2010). The restriction of HDR is that it can occur only in the G2/S cell cycle phase, when sister chromatids are accessible to serve as repair templates after DNA replication is completed, whereas NHEJ can occur at any phase of the cell cycle, making it the more general, but less precise repair mechanism (Heyer et al., 2010). In addition, the cell type, design of the repair template, and target location in the genome all play a role in DSB repair via the HDR pathway (Ran et al., 2013). An exception to this is CRISPR base editing, which utilizes inactive Cas9 endonuclease (dead Cas9, dCas9) or Cas9 nickase and does not require DSB to generate a single nucleotide change in the DNA (Eid et al., 2018; Shen et al., 2014).

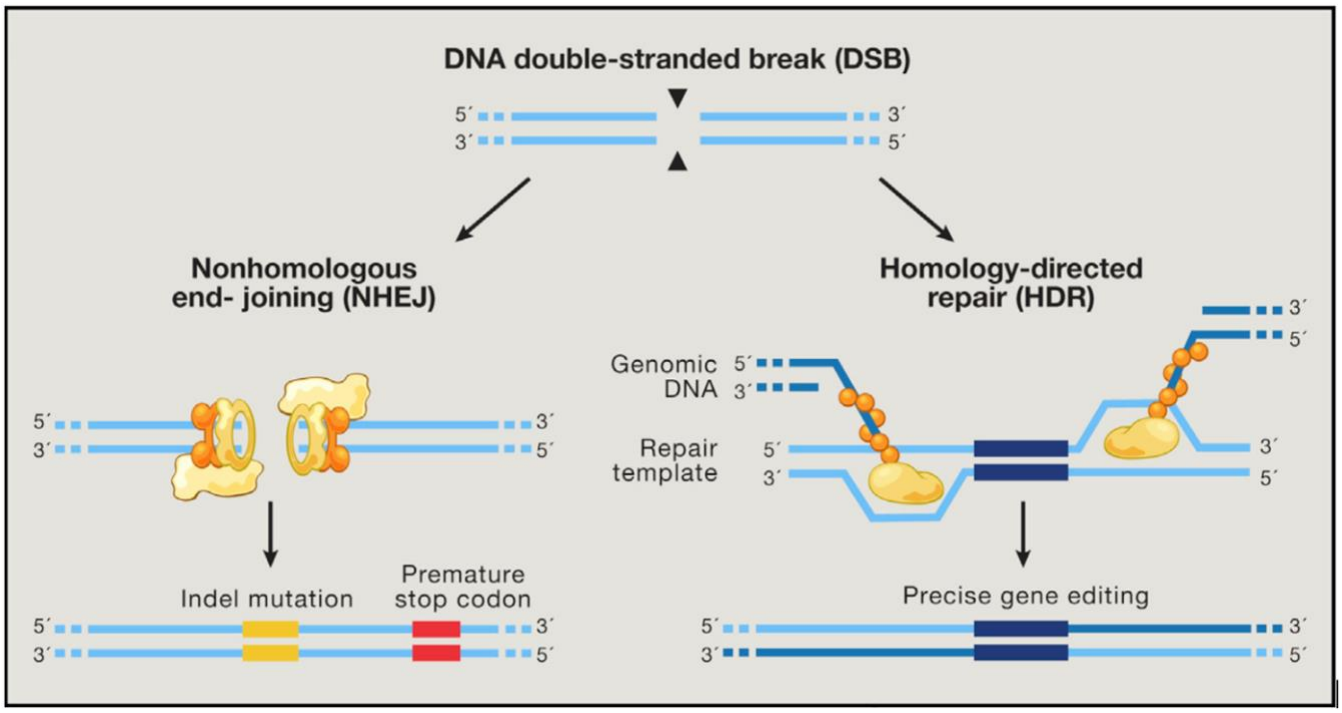

Figure 4. Intrinsic cellular DNA repair machinery. Nonhomologous end-joining (NHEJ) and homology-directed repair (HDR) are the most common methods for repairing DNA double-strand breaks (DSBs) in living cells. NHEJ frequently leads to insertions/deletions (INDELs), which can cause frameshifts or premature stop codons which is why it is preferable for making knockout mutations. Homology-directed repair (HDR), usually in the form of homologous recombination (HR) is the preferred repair mechanism for precise genome editing. (Figure from (Hsu et al., 2014), copyright acquired from CellPress).

A most efficient way of getting the CRISPR-Cas9 editing machinery into cells is assembling it into a ribonucleoprotein (RNP) complex. The RNP complex produces fewer off-targets and degrades faster in cells than Cas9-plasmids (Lin et al., 2014). The Cas9-guide RNA ribonucleoprotein contains the sgRNA and the Cas9 endonuclease. Transfection is the process of delivering nucleic acids into eukaryotic cells. Like in this project, nucleofection is often used as the transfection method to get the RNP complex into the cell *in vitro*. Nucleofection is an electroporation-based transfection method. Electroporation uses electrical shocks of certain voltages to widen the pores of the cell membrane to allow the transfected material to enter the cell cytoplasm (Recillas-Targa, 2006). To enable knock-in editing of the genome via the HR-pathway, a template must be provided to the cell with the RNP-complex, that contains the wanted sequence to be knocked in. A recommended HDR template for single-nucleotide substitutions planned in this project is a single-stranded donor oligonucleotide (ssODN). It is recommended over double-stranded donor oligonucleotide (dsODN) because of lower toxicity and lower frequency of

random integration into the genome (Li et al., 2019). The ssODN is complementary to the target sequence to be edited, which enables the HR to use it as a template for repair after the DSB has occurred. In this project, the ssODN is a 101 nucleotides long strand where the homology arms of either side of the point mutation are 50 nucleotides long.

In addition to the target site, sgRNA can inadvertently lead the RNP complex to an imperfectly matching "off-target" site in the genome. Because of this, guide RNAs need to be designed to match as few sequences in the genome as possible, preferably matching to unique sequences, where the PAM site is no further than 10 nucleotides from the mutation site (Fu et al., 2013). To maximize the chance of HR-pathway exploitation in the gene editing process, small molecules can be used to act as HDR enhancers. These can work either by arresting the cell cycle to the optimal G2/S phases for the HDR, when replication of DNA has already occurred or by inhibiting the NHEJ pathway. Inhibition of the NHEJ pathway works by inhibiting the molecules involved in the DSB repair process (Heyer *et al.*, 2010; Shrivastav et al., 2008).

1.6 Aims of the study

The primary aim of this study was to correct a heterozygous mutation c.197G>T in the gene *CHCHD10* in SMAJ patient myoblasts. Correcting the mutation creates a genetically matched (isogenic) healthy cell line, that can be used to model the disease mechanisms of the patient *in vitro* by comparing the phenotype to the diseased cells.

The more specific aims were:

1. To correct the heterozygous c.197G>T (p.G66V) SMAJ-causing mutation in the gene *CHCHD10*, with CRISPR-Cas9 genome editing technology.

2. To optimize electroporation and cell sorting conditions for CRISPR-Cas9 genome editing with myoblast cells.

3. To assess the effectiveness and precision of genome editing in the electroporated cell populations, as well as the quality of editing in the corrected myoblast cell line with off-target analysis.

2. Materials & Methods

2.1 Myoblast cell culture

Myoblast lines used in this study were obtained by muscle biopsies from SMAJ and non-SMAJ patients. Patients gave written informed consent, and the study was approved by HUS coordinating ethics committee for projects "Lihastautien geenidiagnostiikka - NEMU-SEQFIN ja MYO-SEQ projektit" (192/13/03/00/2015). All cell lines used in the laboratory are coded to enable the anonymity of patient and control donors.

Myoblast cells were cultured on a 100 mm cell culture dish (Thermo Scientific, #130182) in myoblast cell culture medium, 500 ml containing: 420 ml Ham's F-10 Nutrient Mix, GlutaMAX™ Supplement (Thermo Scientific, #41550021), 250 mg Bovine Serum Albumin (Sigma, #A4503), 250 mg fetuin (Sigma, #F2379), 90 mg insulin (Sigma, #I5500), 5 ml dexamethasone (Sigma, # D1756) and 500 µl Epidermal Growth Factor (EGF) (Sigma, #E5036). Cells were allowed to grow until they reach 85 – 95 % confluency, after which they were passaged by incubating for 3 minutes with 1x Trypsin (Thermo Scientific, #R001100). Cultures were maintained in an incubator at conditions +37 °C, 5 % CO2, and 95 % O2.

The cells were harvested by scraping on ice with 1 ml +4 °C PBS. Collected cell suspensions were centrifuged for 10 minutes at 1500 g at +4 °C. The supernatant of the centrifuged samples was aspirated, and cell pellets were stored at -80 °C

2.2 The design of crRNA

Two crRNAs were used in this project to target the mutation site c.197G>T in *CHCHD10* exon 2, referred to as guide#1 and guide#4. The crRNAs were 20 nt long and had *S. pyogenes* Cas9 (SpCas9) PAM site of the form NGG. Guide#4 was previously designed to knock-in a heterozygous c.197G>T mutation in a healthy wild-type cell line. It was tested along with other crRNAs and proved to be the most effective for inducing mutations and therefore chosen also for this project. Guide#4 targeted the reverse genomic strand and was non-allele specific whereas guide#1 targeted the forward genomic strand and was allele-specific, meaning that it only recognized the allele with the heterozygous mutation. Cut distances from the mutation site were 11 nt and 9 nt for guide#4 and guide1 respectively (**Table 1**). Both cRNAs were

designed based on the *CHCHD10* Ensembl gene ENSG00000250479 located in Chromosome 22: 23,765,834-23,767,972 reverse strand (GRCh38.p13, Homo Sapiens).

CRISPOR version 4.97 ((Concordet and Haeussler, 2018); http://crispor.tefor.net/) and Benchling (https://benchling.com/ crispr) were used for designing and evaluating the specificity and efficiency of the crRNAs. The specificity score of CRISPOR's CFD (cutting frequency determination) ranges from 0 to 100, with 100 being the most precise and linked with fewer off-target impacts. Guide#1 had a CFD specificity score of 93, whereas guide#4 had a score of 95. The specificity of the crRNAs was high enough (CFD >60) that they were chosen for CRISPR-Cas9 genome editing. Benchling graded on-target and off-target performance on a scale of 0-100, with 100 being the best. On-target scores indicate the likelihood of Cas9 binding to the crRNA-targeted area. The off-target score is the converse likelihood of Cas9 endonuclease binding off-target, therefore a higher score means Cas9 is less likely to bind elsewhere in the genome (Doench et al., 2016). The crRNA oligonucleotides were ordered from Integrated DNA Technologies (IDT) (**Table 1**).

Table 1. Characteristics of the crRNA sequences designed to target *CHCHD10* exon 2

crRNA	**Sequence (5' → 3')**	**Orientation**	**PAM**	**Cut distance to mutation (nt)**	**CFD specificity score (CRISPOR)**	**On-target/off-target scores (Benchling)**
guide#1	GTCATGGTCA GCGCCCTGAC	Forward	CGG	9	93	58.8/42.9
guide#4	CCCGCTGAAG GCTCCGGTCA	Reverse	GGG	11	95	47.6/45.6

2.3 The design of ssODN repair template

Two 101 nucleotides long repair templates were designed with Benchling (https://benchling.com/crispr) to be utilized in combination with the two crRNAs to correct the mutation c.197G>T. Thymine was replaced with guanine (adenine with cytosine with the sense-oriented template), and 50 nucleotides of wild-type DNA were inserted as homology arms on both sides of the mutation site. To avoid the ssODN

sequence being cleaved by Cas9 before or after HDR, one nucleotide in the PAM-sequence was changed in guide#4 ssODN so, that the corresponding amino acid would not change. The guide#1 ssODN did not include a PAM site change, but a one nucleotide change near the PAM site. PAM site change in ssODN is done to prevent the cleavage of Cas9 after the desired edit. If the PAM site cannot be changed without an amino acid change, the change should be done as near the PAM site as possible to prevent any undesired cleaving. Because guide#1 was in forward orientation, the ssODN for it was designed to be in sense orientation, and ssODN for guide#4 was designed to be antisense because guide#4 was in reverse orientation. Both repair templates (**Table 2**) were ordered from IDT.

Table 2. ssODN template sequences used in myoblast CRISPR-Cas9 RNP complex electroporations. The mutation sites are indicated in bold and PAM sequence changes are underlined.

ssODN	**Sequence**	**Orientation**
ssODN for guide#1	5’ -G CTG GGA GGG CTC CGA GCT CCC GCT GAA GGC TCC GGT GAG GGC GCT G**C**C CAT GAC GTG TCC CAC AGC CGA GCC CAC GGC TAC CCC TGC GGC CGT GGT C- 3’	Sense
ssODN for guide#4	5’ -G ACC ACG GCC GCA GGG GTA GCC GTG GGC TCG GCT GTG GGA CAC GTC ATG G**G**C AGC GCA CTG ACC GGA GCC TTC AGC GGG AGC TCG GAG CCC TCC CAG C- 3’	Antisense

2.4 Amplification of genomic DNA

Genomic DNA of the patient and the electroporated cell pools were extracted using a NucleoSpin® Tissue DNA extraction kit (Macherey-Nagel #740952). The concentrations of extracted genomic DNA were measured using DeNovix® DS-11 FX Spectrophotometer/Fluorometer and diluted to appropriate

final concentrations. The DNA was amplified using 12.5 µl of MyTaq HS Red Mix (Bioline #BIO-25047), 1 µl of 10 µM forward and reverse primers (**Table 3**), and a 100 µg DNA template in a 25 µl reaction with nuclease-free water. (**Table 4**). All PCR amplifications were done using the Bio-Rad S1000 Thermal Cycler. A single 427 bp band was produced on a 1 % agarose gel run at 140V for 45 minutes and imaged with a Bio-Rad ChemiDoc XRS+.

Table 3. Primers for *CHCHD10* exon 2 amplification

Primer	**Sequence (5' → 3')**
CHCHD10 exon 2 Forward	TTAACCCTGCTTCCTCCCAC
CHCHD10 exon 2 Reverse	GGAAGCCTGCCTCTAAGTGA

Table 4. Thermocycling protocol for DNA amplification

Number of cycles	**Temperature (°C)**	**Time**
1	95	1 min
34	95 60 72	15 s 15 s 30 s
1	72	5 min
	12	Hold

2.5 Genome editing of the myoblast cell lines

2.5.1 Optimization of electroporation

To optimize electroporation conditions for myoblast cells, eight different conditions were tried, described in **Table 5**. Before electroporation, cells were let to grow to 75% confluence in 100 mm cell culture dishes (Thermo Scientific, #130180).

Table 5. Optimization of electroporation conditions for myoblast cells.

Number of cells	**Conditions**			
1.00*E+05 cells per well	1700 V, 3 x 20 ms	1650 V, 3 x 10 ms	1450 V, 2 x 20 ms	1100 V, 2 x 20 ms
2.00*E+05 cells	1700 V,	1650 V,	1450 V,	1100 V,

per well	3 x 20 ms	3 x 10 ms	2 x 20 ms	2 x 20 ms

Cells were prepared by washing twice with 1 ml 5% FBS in PBS and detached with 1X Trypsin (Thermo Scientific, #R001100) -EDTA (Thermo Scientific, #15575020) (TE) in PBS for 3 minutes at 37 °C. Cells were gently pipetted into a 50 ml Falcon tube containing 4 ml 5% FBS-PBS, centrifuged for 5 minutes at 250 g, and resuspended in 1 ml 5% FBS-PBS. The number of living cells was counted using Countess Automated Cell Counter (Thermo Scientific), and appropriate number of cells was placed in a 1.5 ml Eppendorf tube, centrifuged at 200 g for 3 minutes, and resuspended in 300 µl Neon electroporation R buffer (Invitrogen). The cells were then combined with a plasmid containing a GFP marker to test the electroporation efficiency. At a time, 100 ml of the mixture was pipetted into a cuvette containing 3 ml electroporation buffer E2 (Invitrogen) and electroporated with all conditions using the Neon Electroporator (Invitrogen). The cells were plated to a 24-well cell culture dish (Thermo Scientific) according to **Table 5**. Duplicates were prepared for all conditions in both cell concentrations.

To assess the transfection effectiveness in electroporated cells, the cell culture medium was replenished 24 hours after electroporation and observed using a Life Technologies EVOS® FL-electro fluorescence microscope for the presence of GFP-tagged cells. The fluorescent signal was weak after 24 hours, so the media was changed and left to incubate for another 24 hours. After 48 hours, the signal was still weak in all conditions but based on visual inspection, decided to use the condition 1650 V, 3 x 10 ms with a fewer number of cells (100 000) for subsequent electroporation.

2.5.2 RNP electroporation

The crRNA-tracrRNA duplex (100 µM), or sgRNA, was prepared by incubating 2.5 µl of 200 µM solution of the synthesized crRNA of interest (IDT, AltR CRISPR-Cas9 crRNA) + 2.5 µM of ATTO-550 labeled tracrRNA (IDT, #1075925) for 5 minutes in 95 °C, and allowed to cool to room temperature. The RNP complex was prepared by incubating 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) at room temperature for 20 minutes. This was done with both crRNAs (guide#1 & guide#4).

After the optimization of electroporation conditions, the primary transfection of this project was done with 100 mm cell culture plates with 75 % confluency. The cells were transfected separately with both

crRNAs, guide#1, and guide#4, and RNP complex was used instead of GFP-plasmid. The cell preparation was performed as described earlier. Cells were resuspended in 400 µl R-buffer and divided into two (200 µl each) for electroporation with both gRNAs. The 200 µl of cells were combined with 6 µl RNP complex, 30 µM HDR enhancer (IDT, #1075915), and 4 µl of ssODN template. The HDR enhancer used was an NHEJ inhibitor to boost HDR in the electroporated cells. Two electroporations were made for both crRNAs: 100 µl of the mixture was taken into a cuvette with 3ml electroporation buffer E2 and electroporated with 1650V for 3 x 10 ms with the Neon Electroporator and plated to 60 mm cell culture dishes (Thermo Scientific) with myoblast media. One dish with each crRNA was incubated at 32 °C for 24h and the other at 37 °C to see which condition would produce better HDR efficiency. Cells were observed and imaged with a fluorescence microscope 18 hours after electroporation after which the media was changed.

2.5.3 Fluorescence assisted cell sorting

First fluorescence assisted cell sorting (FACS) was performed 24 hours after electroporation with BD Influx Cell Sorter (BD Biosciences). Both electroporated cell pools were allowed to grow to 80% confluency in 100 mm cell culture dishes and the cells were counted with Countess Automated Cell Counter. FACS buffer was prepared in advance consisting of 3815 µl 1X Hank's Balanced Salt Solution (HBSS) (Thermo Scientific #14025092), 10 µl 0.5 M EDTA, 125 µl 1M HEPES (Thermo Scientific #15630080) and 500 µl 100% inactivated Foetal Bovine Serum (FBS) (Life Technologies #10270106). Cells were washed with 5% FBS in PBS to dispose of dead cells, treated with 1x TE in PBS, and incubated for 3 minutes at 37 °C to detach cells. Cells were agitated with a pipet to make them single cells, transferred to a 50 ml polystyrene Falcon tube containing 5 volumes of 5 % FBS-PBS, and centrifuged at 750 rpm for 7 minutes. The pellet was resuspended into 250 µl FACS buffer.

Four cell populations (guide#1 32 °C, guide#1 37 °C, guide#4 32 °C and guide#4 37 °C) were sorted based on viability and their ATTO 550 positivity to three 96-well plates each. To distinguish between dead and live cells, the machine uses size estimation (dead cells are smaller than live cells), along with a SYTOX™ Blue Dead Cell Stain (Thermo Scientific # S34857) which stains dead cells. The cell sorter conditions were standard: fluidic pressure was 45 psi and the nozzle used was an 85 µm flow stream nozzle. After FACS, the plates were centrifuged at 70 g for 3 minutes, incubated for 72 hours at 37 °C, and the media was changed every other day.

Subsequent FACS sortings were conducted later, with different conditions, for guide#4 37 °C only. For the 2nd FACS, the cells were sorted into 96-well plates, so that for the first plate, the machine did not distinguish between dead and live cells but did so for the other two in the same fashion as in 1st FACS. This was done to see if the myoblast cells were sensitive to sorting and would be more viable with no sorting treatment except separated into single cells. 3rd FACS was done only by sorting single cells. No distinction between dead and live cells was made. The viability of cells was measured with Countess Automated Cell Counter (Thermo Scientific) and the cells were sorted into four 96-well plates. The sorting conditions used were also gentler, with 22 psi fluidic pressure, 100 µm flow stream nozzle, and 10 psi, 140 µm for 2nd FACS and 3rd FACS respectively. Otherwise, the protocol was the same as described earlier.

2.5.4 Subcloning

Before subcloning, electroporated myoblast cell pools were enlarged to 80 percent confluency in 35 mm cell culture dish. The cell culture medium was aspirated, 1 ml PBS (Lonza #17-516F) washed once, and 1 ml Accutase (Thermo Scientific) was added to each plate before incubating at +37 °C for 6 minutes to dissociate cells. Cells were collected into a 15 ml Falcon tube containing 5 ml prewarmed myoblast medium, centrifuged at 250 g for 3 minutes, and resuspended in 1 ml myoblast media. The number of living cells was counted using a Countess Automated Cell Counter (Thermo Scientific), and 500-1000 cells were plated into 100 mm cell culture dish (Thermo Scientific, #130182) containing 10 ml of myoblast medium. Dishes were gently agitated to uniformly distribute cells before being left undisturbed in an incubator for 48 hours (Bruntraeger *et al.*, 2019) Cell culture media were replaced 48 hours after subcloning with 10 ml myoblast media and changed every 48 hours for 8 days until colony harvesting.

2.5.5 Colony picking

Single cell-derived myoblast colonies distinctively separate from other colonies were manually picked from the cell culture dish 8 days after low-density seeding of the cells in subcloning. The colony picking process was performed on two cell populations: guide#1 32 °C and guide#4 37 °C. First, the medium was aspirated, the dish washed once with 5 ml PBS, and 8 ml myoblast medium added. Individual colonies were scraped and collected, using a dissection microscope, in a P20 pipette set at 10 µl and put into a U-bottom 96-well plate (Corning, #0058) with 100 µl myoblast medium per well. To avoid cross-

contamination, the tip was changed between each colony. To create duplicate plates of colonies, cells were triturated 6 times and 50 µl transferred to two separate 96-well cell culture plates (Thermo Scientific).

2.6 Sanger sequencing

The services of The Institute for Molecular Medicine Finland's (FIMM) SeqLab sequencing facility were utilized to perform primary genotyping on all expanded clonal cell lines. The service included product cleaning, sequencing with the ABI3730xl DNA Analyzer, and base-calling using Sequencing Analysis 5.2. To examine the effect of genome editing, chromatogram data were evaluated with SnapGene 6.0 software (Insightful Science; available at snapgene.com) and compared to a reference sequence.

First, DNA was extracted from colonies from 96-well cell culture plates by aspirating the media and adding 15 µl of QuickExtract DNA Extraction Solution 1.0 (Lucigen #QE09050). Cells were scraped from the wells and transferred to PCR tubes, vortexed for 15 seconds, centrifuged down, and incubated at 65 °C for 6 minutes. After this the solution was vortexed again for 15 seconds, centrifuged, and incubated at 98 °C for 2 minutes. The extracted genomic DNA was amplified as described in section 2.4. The PCR products were purified for sequencing by mixing 2 µl ExoSAP-IT PCR Product Cleanup kit (Applied Biosystems, #78200.2) into 5 µl PCR product, incubating at 37 °C for 15 minutes to activate the enzyme, and then incubating at 80 °C for 15 minutes to inactivate the enzyme. The Ready-To-Run sequencing reaction consisted of 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 6**), and 2 µl purified PCR-product as the template. Before being sent to FIMM, the samples were run through a thermocycler program (**Table 6**) on a Bio-Rad S1000TM Thermal Cycler.

Table 6. Thermocycling protocol for BigDye 3.1 sequencing

Number of cycles	**Temperature (°C)**	**Time**
1	96	1 min
29	96 55 60	10 s 10 s 4 min

	10	Hold until delivery to SeqLab

2.7 Restriction endonuclease assay

Restriction endonuclease assay was started by extracting DNA from the myoblast cells. Five cell populations, guide#1 37 °C, guide#1 32 °C, guide#4 37 °C, unedited SMAJ patient, and unedited non-SMAJ patient, were harvested according to section 2.1, and genomic DNA was extracted using NucleoSpin® Tissue DNA extraction kit (Macherey-Nagel #740952). The concentrations were measured with DeNovix® DS-11 FX Spectrophotometer/Fluorometer and diluted to a final concentration of 100 ng/µl. The DNA was amplified with PCR as described in section 2.4 (**Table 4**). The PCR product was then purified using PCR and Gel Cleanup kit (Macherey Nagel #740609.50) and concentrations were measured. For the digestion, 400 ng of DNA was used in a total volume of 20 µl, with 2 µl 10X CutSmart Buffer (New England BioLabs #B7204), 0.5 µl TseI restriction enzyme (New England BioLabs #R0591L), and nuclease-free water. Negative controls with no enzyme were also included. The reaction was incubated at 65 °C for one hour. The digested DNA was run in 3 % agarose gel at 140 V for 2.5 hours and imaged using Bio-Rad ChemiDoc XRS+. The band intensity and DNA cleaving efficiency were quantified using ImageLab 6.0.

2.8 Analysis of knock-in efficiency

The sequenced cell pools were also analyzed with the Synthego ICE (Inference of CRISPR Edits) web tool (https://ice.synthego.com/#/, (Hsiau et al., 2019)). The edited pools with both guides were analyzed and the tool compares the sequences to the unedited patient sequence to determine the efficiency of gene editing. By comparing the edited and control traces, the Indel Percentage determines the editing efficiency (percentage of the pool with non-wild type sequence). Potential editing outcomes are presented and fitted to the observed data using linear regression. When the ICE linear regression is computed during the generation of the ICE Score, the Pearson correlation coefficient (r) is also computed and reported. This is depicted as the ICE tool's Model Fit (R2) Score. The higher the R2 score, the more confident you can be with the ICE score. The Knockout Score denotes the fraction of cells with a frameshift or a 21+ bp indel and the Knock-in Score estimates the percentage of sequences that have the desired knock-in edit. The Synthego ICE analysis tool can detect changes in the rates of HDR and INDELs between the

two crRNAs used in the myoblast electroporation, hence determining the efficiency of the differently designed crRNAs.

2.9 Off-target genome editing

CRISPOR Batch Gene Targeting Assistant 4.97 was used to test off-target genome editing for guide#4. The algorithm compared the crRNA of interest to a reference genome and identified exonic, intronic, and intergenic sites with similar sequences to the crRNA's target regions, as well as PCR primers for these regions. Cutting frequency determination (CFD) scores were used to rank the off-target loci. CFD scores range from 0 to100 and indicate how specific the crRNA is to the target, with 100 indicating the strongest specificity and 0 indicating a weak association due to mismatches between the crRNA and the target. CFD scores are then converted into weighted scores ranging from 0-1, with weak interactions (<20) being lowered to 0 and strong interactions (>80) being transformed to 100 (Doench et al., 2016.) The seven top off-target sites according to their CFD scores were analyzed. PCR primers were designed based on CRISPOR suggestions and ordered from Merck. UCSC *in silico* PCR web tool (https://genome.ucsc.edu/cgi-bin/hgPcr Genome assembly: Dec. 2013 GRCh38/hg38) was used to determine the lengths of the amplicons. Off-targets were sequenced in FIMM SeqLab as described in section 2.6. The sequences were compared to a reference genome with NCBI Basic Local Alignment Tool, Nucleotide Collection.

3. Results

Two crRNAs were chosen for transfection to correct the heterozygous *CHCHD10* mutation c.197G>T for SMAJ disease modeling in patient-derived myoblast cells. The project workflow (**Figure 5**) included two separate electroporations of myoblast cells with RNP complexes with corresponding crRNAs (guide#1 and guide#4) and ssODNs after optimization of the electroporation conditions. Best cell lines were selected according to gene modification efficiency and estimated HDR rates. Cell pools were sorted using FACS and manual colony picking, and corrected clones were isolated based on genotyping and sequence analysis. The possible off-target Cas9 cleavage sites were screened in the corrected cell line to assess genome editing accuracy and quality. The individual results will be presented in subsequent sections in more detail.

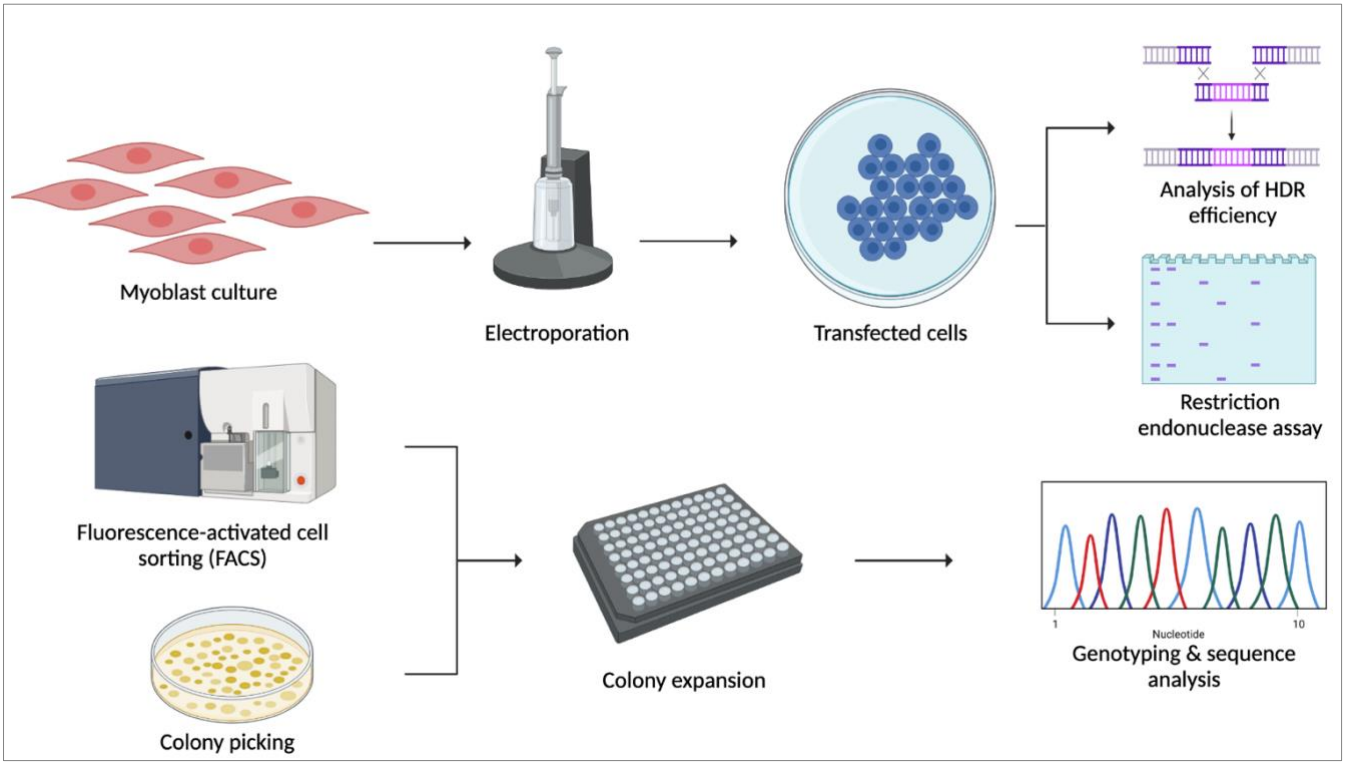

Figure 5. Schematic outline of the project workflow. Patient-derived myoblast cultures were electroporated with RNP complexes, using two different crRNAs separately, allele-specific guide#1 and non-allele-specific guide#4. Transfected cell pools were analyzed using the Synthego ICE web tool (https://ice.synthego.com/#/, (Hsiau *et al.*, 2019)) for HDR efficiency and restriction endonuclease assay for Cas9 cleaving efficiency. The pools were then sorted with FACS and manual colony picking, expanded, and sequenced for genotyping. (Figure created with BioRender.com).

3.1 Myoblast electroporation

3.1.1 Electroporation condition for primary transfection selected based on GFP-intensity

Electroporation conditions were optimized for myoblast cells by using four different conditions for two different cell concentrations. The electroporation conditions used normally for iPSC in the thesis research group where this thesis was conducted are 1100 V for 2 x 20 ms, and therefore it acted as a starting point and the voltage was increased up to 1700 V (Harjuhaahto *et al.*, 2020). The two cell concentration tried were 100 000 and 200 000 cells per electroporation. The optimization electroporation was performed with a plasmid containing a GFP marker and did not include an RNP complex, to see which condition would yield the best permeability for the cells to accept the transfected nucleic acids into the cytoplasm. However, the transfected GFP-plasmid is larger than the RNP complex designed for the actual experiment and therefore should have more difficulties entering the cell. Cells were imaged with a fluorescent microscope 24 hours after electroporation (**Figure 6**). By visual inspection, the GFP intensity was strongest in the condition where 200 000 cells were electroporated with 1650 V for 3 x 10 ms, and this condition was selected for the subsequent electroporation of the RNP complex.

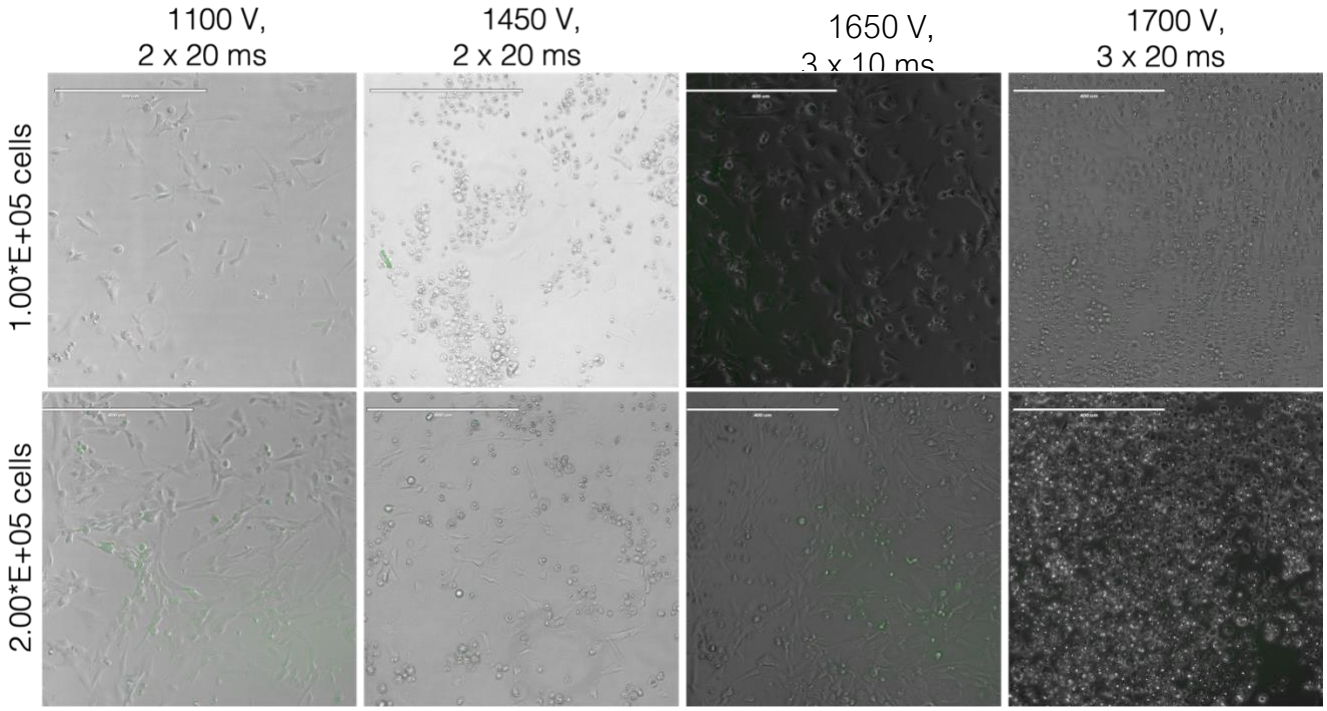

Figure 6. Optimization of electroporation conditions for myoblast cells. Electroporated myoblast cells with different voltages, varying numbers, and lengths of pulses, at two different cell concentrations. The condition selected for the subsequent RNP complex electroporation was 1650V 3 x 10 ms with 200 000 cells per reaction. The myoblast cells did not survive the higher 1700 V condition, which can be seen as dead, rounded cells. Images are taken 24 hours post electroporation with Life Technologies EVOS® FL-electro fluorescence microscope, using the GFP channel. The scale bars are 400 µm.

3.1.2 RNP complex was transfected successfully

The RNP complex with the ATTO550 labeled tracrRNA was electroporated into myoblast cells as described in section 2.4.1. The patient cells containing the heterozygous *CHCHD10* c.197G>T mutation were electroporated with two gRNAs individually: the allele-specific guide#1 and the non-allele-specific guide#4. After electroporation, both cell populations were split and incubated for 24 hours at two different temperatures: 32 °C and 37 °C. The lower temperature is rather extreme for human cells, but cold shock has been shown to increase the frequency of HDR gene editing in human iPSCs, and thus was also tried here for myoblast cells (Guo et al., 2018). After 24 hours, cell populations were imaged with a fluorescent microscope using the RFP channel. ATTO550 signal was detected from all cells, indicating that the RNP complex had entered the cells in all cell populations very well (**Figure 7**). All cell populations were viable after electroporation.

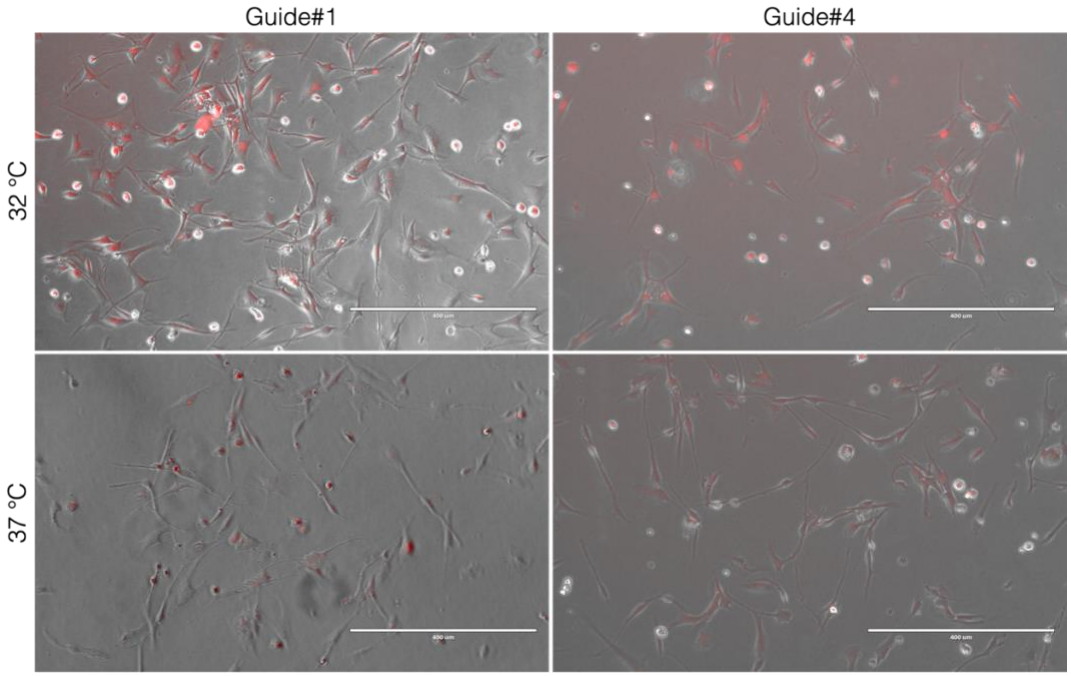

Figure 7. RFP images of successfully electroporated myoblast cells. Transfection was successful with both sgRNAs and temperature conditions. All cell populations were viable after electroporation. The images were taken 24 hours after electroporation with Life Technologies EVOS® FL-electro fluorescence microscope, using the RFP channel. The scale bars are 400 μm.

3.1.3 Guide#4 37 °C population had the best gene modification efficiency

Gene modification efficiency of the electroporated cell populations was estimated with a restriction endonuclease assay. The c.197G>T mutation deletes a TseI restriction endonuclease cleaving site. If edited successfully, the heterozygous c.197G>T mutation is corrected and TseI cleaving site restored. When the wild-type non-SMAJ patient is digested with TseI, the 427 bp long amplified genomic DNA sequence is cleaved at five different sites, resulting in six DNA snippets of varying sizes: 122 bp, 106 bp, 69 bp, 53 bp, 52 bp, and 25 bp. The c.197G>T mutation deletes one TseI cleaving site, thus resulting in five snippets of DNA: 175 bp, 122 bp, 53 bp, 52 bp, and 25 bp. The cleaving site between the 106 bp and 69 bp snippets has been removed and a 175 bp snippet is produced. The SMAJ patient studied here is heterozygous for the mutation and thus it only removes the TseI cleaving site from one allele. This results in seven different bands in the restriction digestion, as only half of the 175 bp snippet is cleaved into 106 bp and 69 bp snippets. Gene modification efficiency is determined by comparing the relative band intensities of 106 bp + 69 bp bands to the intensity of the 175 bp band. The proportion of the combined 106 bp + 69 bp band intensity should increase when the amount of correctly gene-edited cells increases in the cell populations. The analysis shows that guide#4 37 °C cell population has the highest amount of successfully gene edited cells (**Figure 8**). Restriction analysis was performed with three cell populations and the ones with the highest gene modification efficiencies (guide#4 37 °C and guide#1 32 °C) were chosen for subsequent subcloning.

A. **B.**

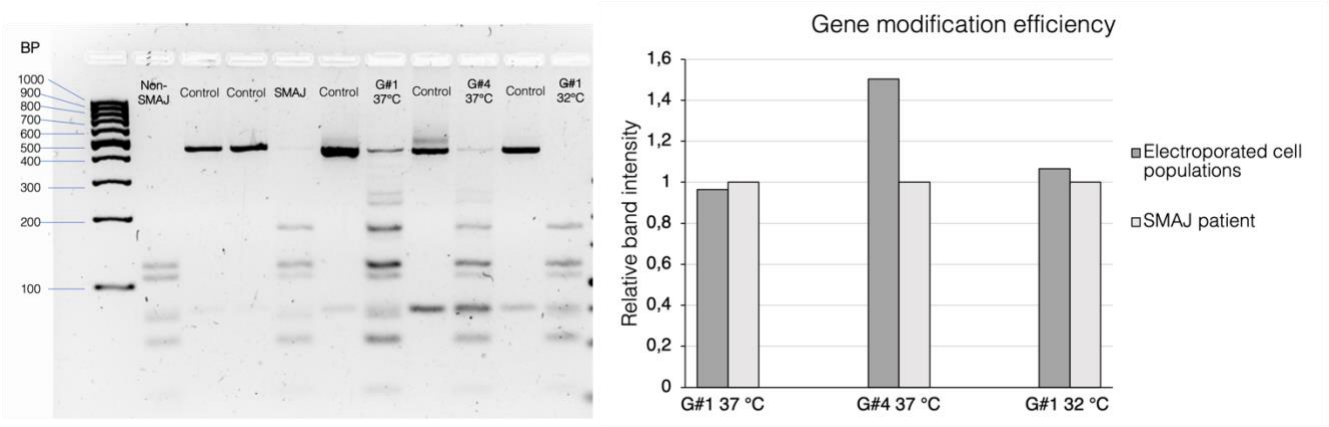

Figure 8. Gene modification efficiencies estimated with TseI digestion. (A) Restriction endonuclease digestion of electroporated cell populations, non-SMAJ patient, and SMAJ patient produced bands with varying relative intensities on 3% agarose gel. Negative controls had only PCR products with no enzyme. Non-SMAJ patient acted as an experimental control. **(B)** Gene modification efficiencies were estimated by comparing the relative band intensities of 106 bp + 69 bp bands to the intensity of the 175 bp band after normalization to the SMAJ patient. The analysis shows that guide#4 37 °C cell population (G#4 37 °C) has the highest gene modification efficiency.

3.1.4 Knock-in efficiency in the electroporated cell populations

The HDR efficiency and the integration of the ssODN template to the electroporated cell populations were analyzed using the Synthego ICE CRISPR bioinformatics web tool (Hsiau *et al.*, 2019). Only the cell population guide#4 37 °C was managed to analyze with the software since guide#4 32 °C was bacterially contaminated when optimizing the FACS conditions after electroporation (see section 3.3.1) and the allele-specific guide#1 was not able to be analyzed. The allele specificity made the HDR efficiency in guide#1 cell populations impossible to analyze with currently available web tools, which require the gRNA sequence to be found in the control (SMAJ patient in this case) sequence. Since the patient was heterozygous for the mutation, the patient showed "K" in the sequence (**Figure 9**) and the web tools did not allow the gRNA sequence to include a nucleotide "K". For this reason, only the cell population guide#4 37 °C was analyzed, which had indel percentage of 28 % and knock-in percentage (cells with the desired KI-edition) of 20 %. The model fit (R^2) score was 0.99.

A visual examination of the chromatograms was conducted. In the chromatograms of the cell population sequences, the gRNAs and the corresponding PAM sites are shown. Both guide#1 cell populations show

minimal variation in the nucleotides after the cut site and in the nucleotide G in position 160/161, which should change to C if the edition would be successful. Guide#4 37 °C population shows a lot of cleaving activity and variance in nucleotides after the cut site resulting from NHEJ-based DSB repair. Also, the cytosine in position 157 in the middle of the PAM site (depicted in blue) shows that it has been modified in some cells to adenine according to the intended ssODN integration (**Figure 9**). Based on these results, guide#4 pool was the cell pool with the most effective genome editing, ssODN integration, and HDR pathway exploitation.

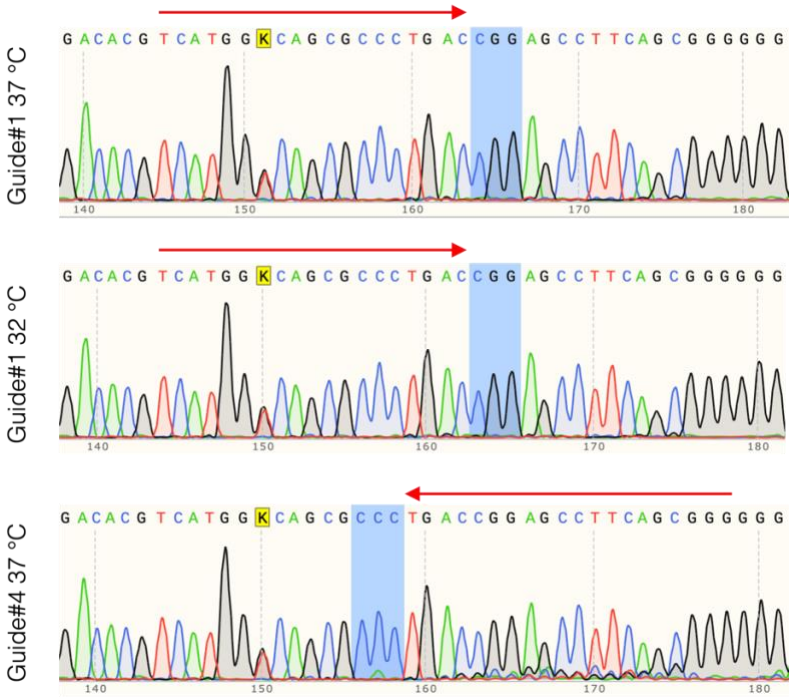

Figure 9. Chromatogram sequences of electroporated cell populations. Guide#1 32 °C and 37 °C shows the allele-specific guide#1 (red arrow) in a forward direction and the corresponding PAM site CGG (blue box). Guide#4 37 °C population shows the reverse non-allele specific guide#4 with the corresponding PAM site GGG (here CCC). Guide#4 sequence shows more editing activity, seen by nucleotide variability, than guide#1 sequences. Chromatogram images are taken with SnapGene 6.0 software (Insightful Science; available at snapgene.com).

3.2 Creation of clonal cell lines

3.2.1 Both FACS and manual colony picking produced viable clones

1st FACS. Cell populations were sorted into single cells 24 hours after electroporation to produce clonal

cell lines (**Figure 11**). All four cell populations (guide#1 32 °C & 37 °C, guide#4 32 °C & 37 °C) were sorted into three 96-well plates each with fluidic pressure of 45 psi and 85 µm flow stream nozzle according to section 2.5.4. The remaining cells were plated in 60 mm cell culture plates. Guide#4 32 °C cell population was bacterially contaminated during cell sorting and had to be discarded. In addition, the incubator cabinet was exhausted of CO2 for a couple of hours post FACS and resulted in no cells surviving in the sorted 96-well plates.

Colony picking. Because of the poor FACS results, cell pools were sorted by manually picking colonies from cell culture dishes. Myoblast cells do not form colonies of uniform morphology and clear edges like iPSCs. Instead, groups of cells were present in the cell culture dishes (**Figure 10**), and only colonies clearly separate from others were picked, starting from the center of the colony to avoid cross-contamination. Colony picking was tried in addition to FACS as it can be gentler, and clones can start expanding from many cells instead of just a single cell, like in FACS. The downside is that manual colony picking is considerably more labor-intensive than automatic cell sorting. Two cell populations were selected for colony picking based on HDR and restriction endonuclease analyses: guide#1 32 °C and guide#4 37 °C. Colonies were picked eight days after subcloning approximately 500-1000 cells to cell culture dishes and expanding into colonies. The number of colonies picked was 30 and 37 from guide#1 and guide#4 populations, respectively (summarized in **Table 7**). Out of the total 67 colonies picked, 36 produced viable clones (53.73 %) that were genotyped, but none of the clones contained the expected *CHCHD10* c.197G>T correction.

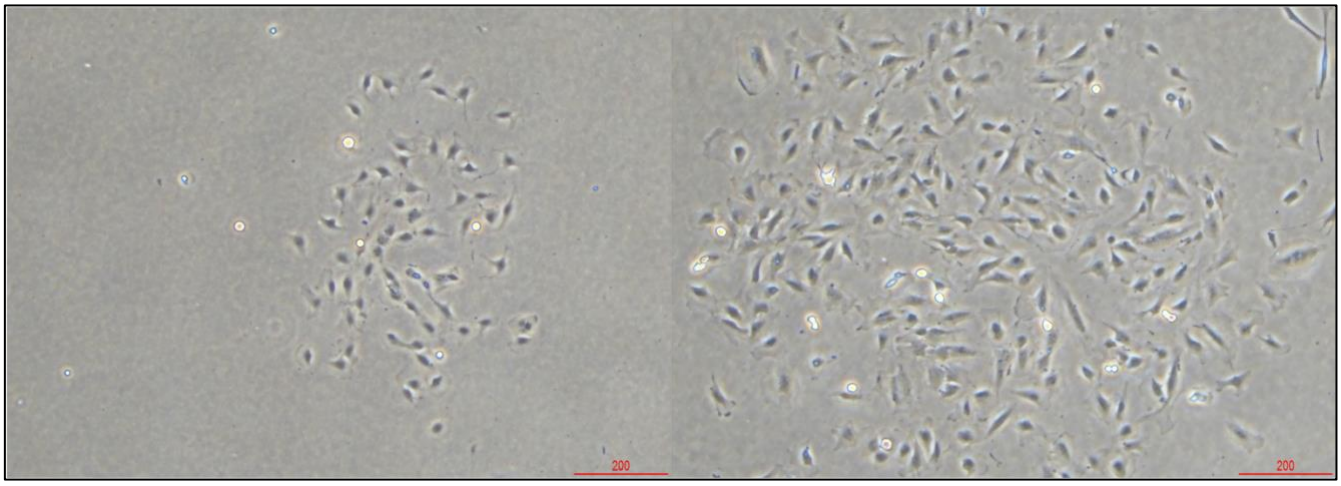

Figure 10. Light microscopy image of myoblast colonies 8 days after subcloning. Myoblast colonies do not form uniform colonies with clear edges like iPSCs. Only colonies that were clearly separated from

other surrounding colonies, were selected for picking. Colonies were picked starting from the center of the colony to avoid cross-contamination. The scale bar is 200 µm.

2nd FACS. Second FACS was performed to increase the number of viable clones. Based on the results obtained from HDR efficiency analysis and overall better cleaving (and because guide#4 32 °C was lost due to contamination), the sorting was conducted only with guide#4 37 °C cell population. The FACS conditions were altered from the 1st FACS to be gentler to the cells and sorting was done based not on ATTO550 since the fluorescence had already faded (**Figure 11**), but on size, according to section 2.5.4. Dead cells were distinguished from live cells based on SYTOX™ Blue Dead Cell Stain and their size (dead cells are smaller), and single cells were sorted into three 96-well plates. The conditions used for this sorting were 22 psi of fluidic pressure and 100 µm flow stream nozzle tip. Out of total 288 sorted cells, 28 produced viable clones (9.72 %) and one clone contained the looked-for gene editing (**Table 7**).

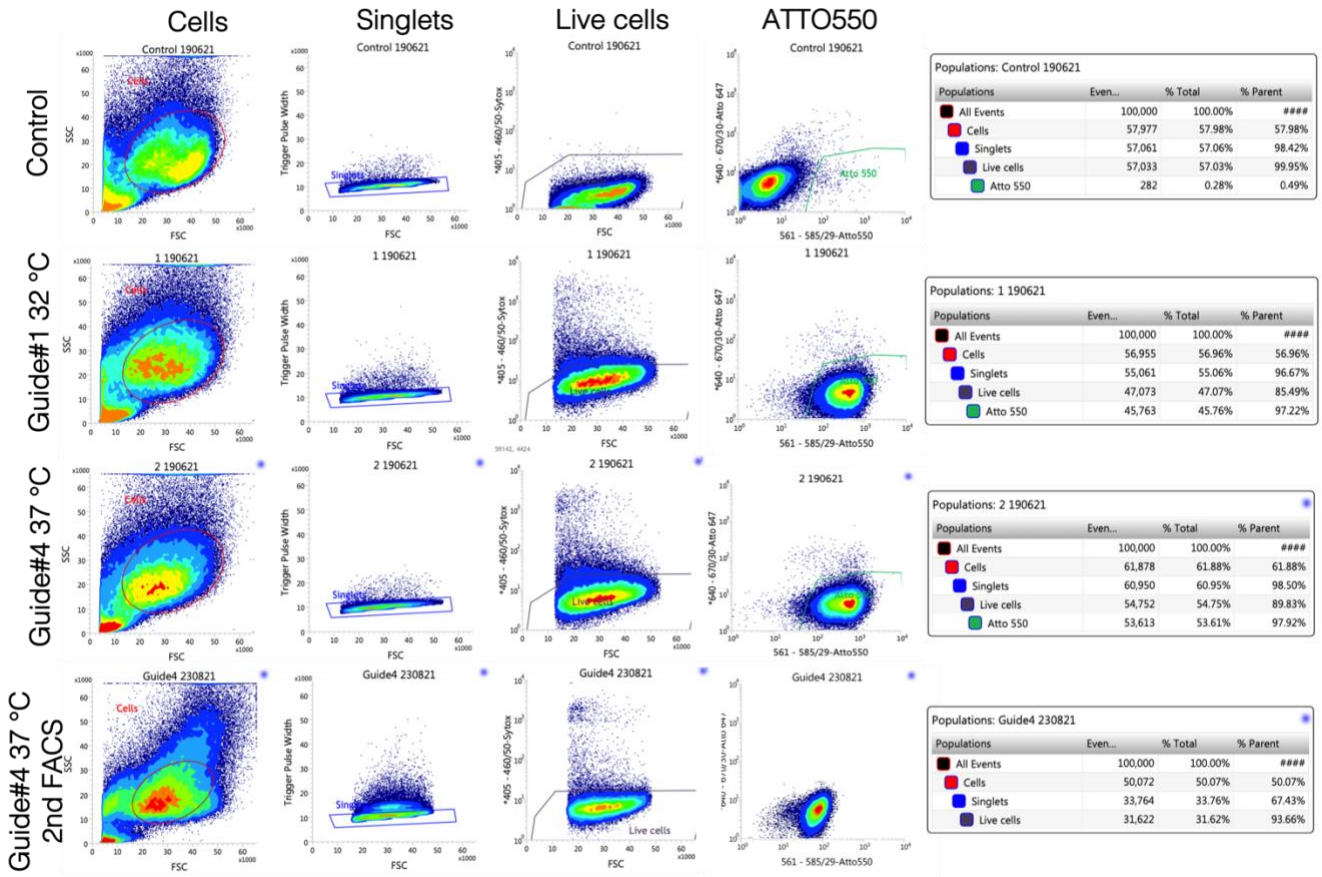

Figure 11. FACS data from 1st and 2nd cell sortings. Here are depicted the cell sorting data from control, guide#1 32 °C and guide#4 37 °C cell populations. First three rows are from the 1st cell sorting 24 hours after electroporation and the last row is from 2nd FACS with only guide#4 37 °C. Control cells

were non electroporated patient cells and do not have the ATTO550 marker in them. No ATTO550 signal was detected from guide#4 37 °C in the 2nd FACS, since it had been already faded, and sorting was done based on live cells. The data tables show the proportion of single cells, live cells and ATTO550 positive cells from all events (graphics courtesy of Biomedicum Flow Cytometry Unit).

3rd FACS. To further increase the number of clones, one additional FACS was conducted with guide#4 32 °C cell population. Conditions were made even gentler by lowering the pressure to 10 psi and increasing the nozzle size to 140 µm. Again, the cells were sorted based on their size, since the fluorescence had faded. Single cells were sorted into four 96-well plates totaling 384 sorted cells. Out of these 20 viable clones were produced until the experiment was needed to bring to an end (**Table 7**). None of the clones produced in time contained successful correction of the mutation.

Table 7. Comparison of sorting methods for creating clonal myoblast cell lines. Here are presented the number of clones sorted from each sorting method, the number of viable clones produced, the proportion of clones that were viable out of the total clones sorted and the amount of successfully gene edited clones.

Sorting method	**Clones sorted**	**Viable clones**	**The proportion of viable clones**	**Successfully gene edited clones**
Colony picking	67	36	36/67 (53.73 %)	0
FACS #2	288	28	28/288 (9.72 %)	1
FACS #3	384	20*	20/384 (5.21 %)*	0*

*The number of viable clones would almost certainly have been higher if allowed more time for the experiment. The proportion is not an accurate representation and cannot be compared to the other sorting methods.

3.2.2 Clonal myoblast cell line with corrected *CHCHD10* c.197G>T mutation was created

Out of 84 clones in total, 71 clones edited with guide#4 and 13 clones with guide#1, were sequenced and one clone was found with the heterozygous *CHCHD10* c.197G>T mutation corrected, that matched the normal sequence and contained the PAM site edit. The corrected clone is shown in **Figure 12** along with the sequences of a non-SMAJ patient having no disease mutation, an SMAJ patient with the heterozygous mutation, and an edited clone with a cellular NHEJ repair mechanism employed to repair the DSB, which resulted in a frameshift mutation and created a knock-out cell line. The corrected clone went through a quality control analysis of off-target site sequencing to validate the quality of genome editing.

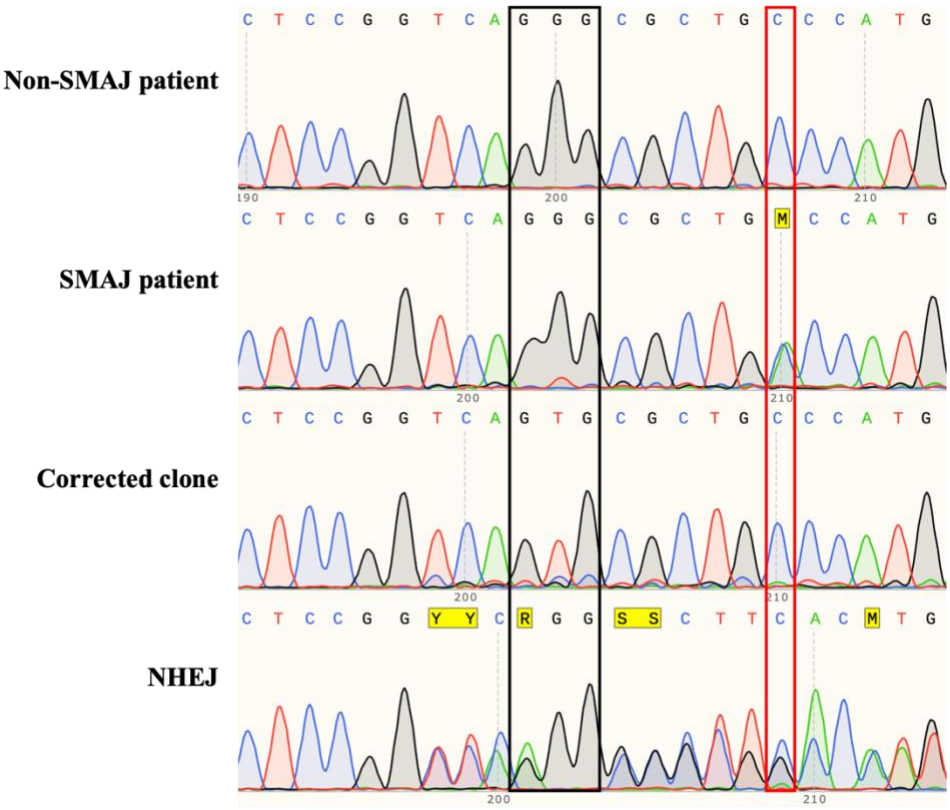

Figure 12. Reverse sequences of *CHCHD10* exon 2 show the disease mutation and targeted PAM site. Reverse sequences are shown here since the non-allele-specific guide#4 targeted the reverse strand. Non-SMAJ patient's sequence shows no editing in the PAM site (black rectangle) and no c.197G>T mutation (C>A in this case because it is a reverse strand). The heterozygous c.197G>T mutation is present in the SMAJ patient. The corrected clone shows the PAM site editing and the corrected mutation site (red rectangle). NHEJ sequence displays editing process where the cell has utilized NHEJ instead of HDR repair mechanism (knock-out cell line). Chromatograms are from SnapGene 6.0 software (Insightful Science; available at snapgene.com).

3.3 No off-target CRISPR-Cas9 cleavage was detected in the corrected myoblast cell line

The corrected clone was screened for potential Cas9 off-target cleaving. Off-target sites are similar loci by sequence as *CHCHD10* exon 2 the crRNA is designed to target, and therefore are susceptible to unintentional Cas9 cleavage. The corrected clonal cell line created with the RNP complex containing guide#4 underwent Sanger sequencing according to section 2.5 for quality control analysis. The CFD off-target scores for guide#4 (CCCGCTGAAGGCTCCGGTCA), 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) are summarized in **Table 8**. CRISPOR Batch Gene Targeting

Assistant found 91 off-target sites in total. Sanger sequencing and NCBI BLAST® revealed no off-target Cas9 cleavage in the locations, although one nucleotide was missing from the off-target 7 sequence, probably due to sequencing error (**Figure 13**). (Basic Local Alignment Tool, Nucleotide Collection, 12th April 2022).

Table 8. Off-target sequences of guide#4 targeted CRISPR-Cas9 genome editing. The top seven off-targets were chosen for analysis based on their CFD score. Mismatches between the guide and the off-target are indicated in bold in the off-target sequence.

Name	**Locus description (CRISPOR)**	**CFD off-target score**	**Mismatches**	**Forward primer (5' → 3')**	**Reverse primer (5' → 3')**	**Off-target sequence**	**Amplicon length (bp)**
Off-target 1	intergenic: RNU7-74P-RBM47/RP11-588L15.2	0.8667	2	GAACCCAGGCAGGGCTATG	AAGGGGAAGAAGGCACCTTG	CC**A**GCTGAAGGCTCC**A**GTCA	103
Off-target 2	intergenic: RNU6-546P-AC093375.1	0.3467	4	TGTCTCCAGGGCATTTCAGA	ATGCTGGAACTTTTGGCTGG	CC**A**GCTGAA**A**GCTCC**A**G**C**CA	340
Off-target 3	intergenic: HBQ1-LA16c-OS12.2	0.3025	4	AAGGATGCCACCCCAGATAG	TTGTACCACCTTTGCCTCCT	CCC**A**CT**C**A**G**GGCT**T**CGGTCA	202
Off-target 4	intergenic:AC093702.1-AC009236.2	0.2632	4	CCACCAACATCCAAGTTACCA	TGAGCACAGGAGTTTGAGGT	CC**A**GCTGAAG**T**C**C**CC**A**GTCA	238
Off-target 5	intergenic: EXTL3-EXTL3-AS1	0.2326	4	ATGGGAGGAATCCTGAGAGG	CTGGGGTTGCATCTGGGAAA	**AC****T**GCTG**T**AGGCC**C**CGGTCA	113

Off-target 6	intron: TAF4	0.2095	4	CTCGGGGAAGAGCTCTACAG	TCAGAGAGCAGTGGACATGG	CC**A**GCTGAAGG**AC**CCGG**C**CA	204
Off-target 7	intergenic: RP11-211G23.2-RP11-211G23.1	0.2066	4	GGAGCATGCACACCTAAGGT	GCTGTAGAGTCAGACAGCCC	CCCGC**C****A**AAGGCTCC**A**GT C**C**	140

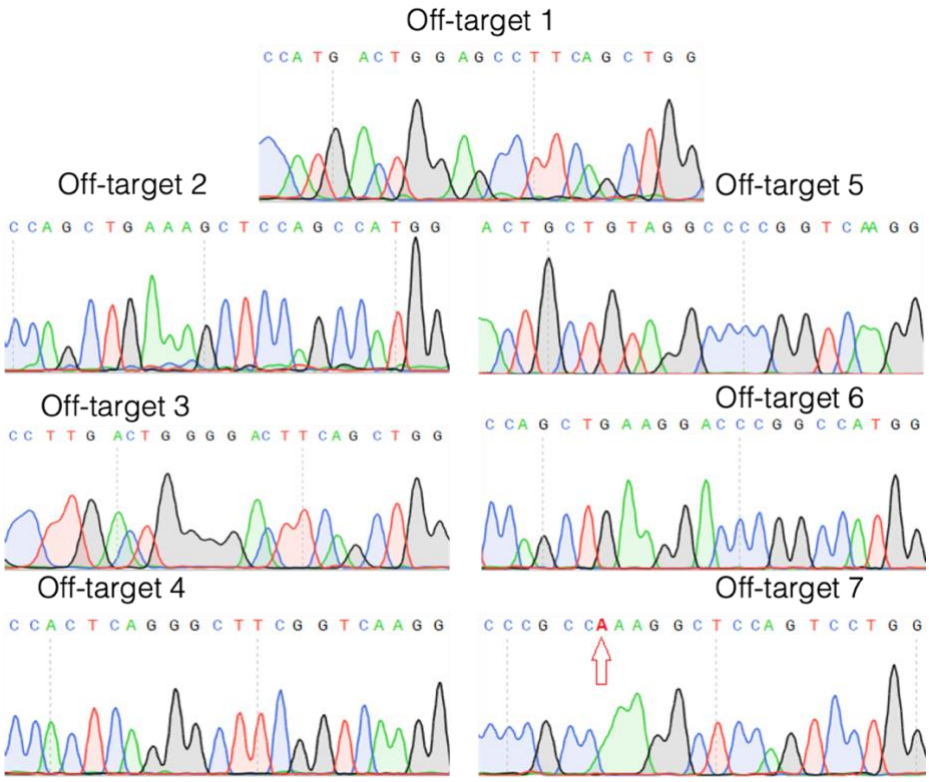

Figure 13. Off-target sequences of the successfully edited clone with guide#4 crRNA. No Cas9 cleavage was detected in the off-target sites. There was one adenine missing from the off-target 7 sequence (red arrow), but otherwise, the sequence was identical to the reference. The one missing nucleotide is probably due to a sequencing error. All other off-target sequences matched to reference sequence (NCBI BLAST).

4. Discussion

The central aim of this project was to correct an SMAJ causing mutation c.197G>T in the gene *CHCHD10* in heterozygous patient myoblast cells with CRISPR-Cas9 genome editing. The goal was to create isogenic myoblast cell lines that could be used to study the role of skeletal muscle in the disease *in vitro*. SMAJ is an autosomal dominant lower motor neuron disease with currently unknown exact disease mechanism. Lower motor neurons are cells that control the voluntary muscles of the body via the neuromuscular junction (NMJ). There is evidence, that CHCHD10 expression in skeletal muscles is required for normal neurotransmission between the motor neurons and the muscle fiber (Genin *et al.*, 2019). Therefore, the impairment of the G66V/+ mutation to the function of CHCHD10 in skeletal muscle plays a key role in motor neuron diseases associated with CHCHD10, such as SMAJ. For this reason, existing patient-derived myoblast cell cultures were used in this project for the creation of isogenic cell lines to enable the modeling of the disease in muscle cells.

Genome editing with CRISPR-Cas9 is not as widely conducted in myoblast cells as in iPSCs or some other types of human stem cells, such as fibroblasts or HEK cells. Consequently, the conditions for the electroporation of myoblast cells were optimized with GFP-containing plasmids, before two different CRISPR-Cas9 ribonucleoprotein complexes and associated ssODN templates were transfected to correct the heterozygous disease mutation. The performance of the two crRNAs was investigated and based on this, cell populations were selected for clonal cell line creation. FACS and manual colony picking conditions were optimized for myoblast cell sorting. One successfully genome edited clone was created by one of the RNP complexes and was partially validated with off-target screening.

4.1 RNP complex design and performance

Two different crRNAs were used in this project: allele-specific guide#1, which targeted the forward genomic strand, and non-allele-specific guide#4, which targeted the reverse genomic strand. Guide#4 was designed earlier for knock-in a heterozygous c.197 G>T mutation in a healthy wild-type cell line, had proved to be effective and was selected for this project to correct the mutation in a patient cell line. Guide#1 was designed for the purposes of this project to target the chromosome with the disease mutation to increase the specificity of the crRNA and to reduce off-target cleaving. The cut site from the targeted knock-in was 9 bp in guide#1 and 11 bp in guide#4. It has been reported, that the cut distance correlates

with gRNA efficiency and that cutting efficiency should decrease dramatically when the distance to the cut site surpasses 10 bp (Ran *et al.*, 2013). This would suggest that guide#1 would perform better than guide#4 but the opposite was observed here. It was evident, that the overall genome editing efficiency was higher in guide#4 crRNA than in guide#1. Overall, more clones showed Cas9 cleaving action in guide#4 edited cell populations and the one c.197G>T corrected clone was produced with guide#4. The allele specificity of guide#1 could have reduced the cleaving efficiency overall, since it only targets the disease allele, but should not affect the HDR knock-in efficiency of the correction. Guide#4 was designed in a reverse direction, but the orientation of the crRNA should be irrelevant since Cas9 can detect a PAM sequence in either sense or antisense direction (Wang *et al.*, 2016)

However, genome editing depends on more than just crRNA design. The choice of a suitable site, CRISPR-Cas9 delivery method, and the design of the selected delivery method all have a role in the success and accuracy of genome editing. The CRISPR-Cas9 delivery method of choice for this project was the RNP complex which is less stressful to stem cells and more efficient than plasmid transfection (Kim et al., 2014). The crRNA in the RNP complex cleaves DNA instantly, and the concentration of the homologous recombination template is highest immediately after electroporation but dramatically decreases over 24 hours, minimizing off-target effects. Plasmid expression is a much slower process and risks the integration of the plasmid into the genome. Plasmid integrated CRISPR-Cas9 complexes can also be cytotoxic in some cell types (Kim *et al.*, 2014). Along with the RNP complex, an ssODN template was delivered to cells. SsODNs are preferred templates over dsODNs for inducing single-nucleotide knock-ins, due to efficient repair without compromising DNA stability (Li *et al.*, 2019). Additionally, ssDNA has higher recombinogenic tendencies and is not integrated randomly into the genome as often as dsDNA templates (Bassett, 2017; Bruntraeger *et al.*, 2019). The design of the ssODN template can be modified to favor a certain DSB repair pathway: the HDR pathway can best be exploited by symmetric, rather than asymmetric ssODNs (Bollen et al., 2018). Moreover, long ssODN templates are generally better performing knock-in templates (Li *et al.*, 2019). In this project, ssODNs containing 50 nt long homology arms were used.

The performance of the crRNAs used in this project was analyzed with a restriction enzyme assay and with the Synthego ICE CRISPR web tool. A TseI restriction enzyme was used to digest the genome edited cell populations. The test showed that editing efficiency is minimal in both guide#1 cell populations, although slightly higher in guide#1 32 °C. Contrastingly, the guide#4 cell population

exhibited great efficiency. TseI enzyme was used in place of the standard T7 assay for genome editing mutation detection because T7 cannot distinguish between heterozygous and wild-type samples and therefore was not suitable for analyzing the correction of the heterozygous disease mutation. ICE CRISPR analysis was only able to be conducted with the non-allele specific guide#4 cell populations since the allele specificity of guide#1 caused problems in the analysis. Guide#4, however, presented a significant 20 % HDR score and 28 % INDEL score. The chromatograms of the edited cell population sequences were inspected visually and again based on this; guide#4 had the most Cas9 cleaving action.

4.2 Genome editing and clonal cell line creation with myoblast cells

Electroporation was used as a transfection method for the delivery of the CRISPR-Cas9 RNP complex into myoblast cells. The electroporation was conducted with Neon Transfection System (Life Technologies), which is the most commonly used electroporator in the field, along with Lonza's Nucleofector (Bassett, 2017; Ran *et al.*, 2013; Wang *et al.*, 2016). Electroporation conditions were optimized for myoblast cells. Two different cell concentrations were tried along with four different voltages and pulse lengths. In the research group where this thesis was conducted, the condition of 1100V for two pulses of 20 ms has been previously used (Harjuhaahto *et al.*, 2020). The Neon Transfection System has been used with a diverse range of conditions with different cell types. Previously, it was discovered that one pulse of 1300 V and 30 ms was the optimal condition for plasmid electroporation into PC12 cells and three pulses of 1650 V for 10 ms have been used for iPSC electroporation (Covello et al., 2014; Tidball et al., 2018).

Taking these results into consideration, myoblast electroporation was optimized with four conditions: 1100V 2x20 ms, 1450V 2x20 ms, 1650 3x10 ms, and 1700V 3x20 ms. The optimization process included electroporating a GFP containing plasmid into the cells to assess the electroporation efficiency with a fluorescence microscope. All conditions produced reasonable results, except the highest 1700 voltage, which killed the cells. Based on GFP intensity, the 1650V 3x10ms condition with the higher concentrations of cells proved to be the most effective with myoblast cells and was selected for the subsequent RNP complex transfection. Cell density is a major variable in electroporation procedures, with higher concentrations generally producing better results (Covello *et al.*, 2014). The CRISPR-Cas9 RNP complex with ATTO550 fluorescent marker tagged tracrRNA was electroporated into myoblast cells with 1650V 3x10ms condition. Based on visual inspection of fluorescence, the complex entered the

cells almost at 100 % efficiency. The RNP complex having a better transfection efficiency than the GFP-plasmid could be explained by plasmid transfection often being inefficient and stressful to cells (Kim *et al.*, 2014).

Electroporations with both crRNAs were supplemented with HDR supporting reagent. The reagent used acted as an NHEJ inhibitor. HDR can be supported by both NHEJ inhibition and HDR enhancement, although HDR enhancement is generally preferred due to possible cytotoxic side effects of NHEJ inhibition. No cytotoxic effects were detected here, and all transfected cell populations were viable. NHEJ is vastly preferred over HDR in most cell types, with HDR occurring only in the late-S, G2 stages of the cell cycle when DNA replication is complete and sister chromatids are available to act as repair templates. HDR is the least error-prone DSB repair approach, although implementing it in experimental settings is complex, and designs are frequently poor in terms of editing efficiency (Shrivastav *et al.*, 2008). The percentage of NHEJ in DSB repairs might be as high as 80-90 % in experimental setups (Bassett, 2017). In addition to small molecule HDR enhancers, cold shock has also been demonstrated to shift DSB repair to favor HDR, resulting in HDR in 70% of unselected human iPSC when combined (Guo *et al.*, 2018; Skarnes et al., 2019). After transfection, cell populations were incubated in two different temperature conditions: 32 °C cold shock treatment and 37 °C standard condition. Guide #4 32 °C cell population was lost due to bacterial contamination from the FACS facility, but guide#1 32 °C showed enhanced gene modification compared to guide#1 37 °C cell population with restriction enzyme analysis, although HDR efficiency could not be analyzed with ICE CRISPR. This suggests that cold treatment could prove a sufficient method to enhance HDR even in myoblast cells. Cold shock treatment seems harsh given that 37 °C is normally considered optimal for human cells, but the treatment did not seem to affect the cell viability negatively.

Single cell sorting for clonal cell line production proved to be a challenge with myoblast cells. A day after electroporation, cells were FACS sorted based on their ATTO550 fluorescence to filter out cells that had not accepted the RNP complex inside. The conditions were standard (85 μm nozzle and 45 psi pressure) used at the Biomedicum Flow Cytometry Unit (University of Helsinki, Finland) with BDInflux Flow Cytometer. No viable clones were produced with this sorting. The question remains if this was due to harsh FACS conditions or the fact that the cells were depleted of CO2 for a few hours after sorting due to reservoir exhaustion. Furthermore, guide#4 32 °C cell population caught a bacterial infection from the FACS facility and was not able to be used at later cell sortings. Manual colony picking required the

subcloning of myoblast cells. Myoblast cells do not form colonies of uniform morphology and clear edges like iPSCs do, and therefore, to avoid cross-contamination, only very distinctive colonies were picked starting from the middle, so no surrounding drifting cells from other colonies would be picked. Colony picking proved to produce clones effectively since over 50 % of clones were viable. The major disadvantages of manual cell sorting are its leisureliness resulting in a low number of clones overall and its lower reliability compared to automatic cell sorting for creating clonal cell lines (Bruntraeger *et al.*, 2019).

Additional FACS sorting was conducted with guide#4 37 °C edited cell population to produce more viable clones for sequencing. The conditions were altered from the first FACS to be gentler for the cells, so the nozzle size was increased to 100 µm and pressure was lowered to 22 psi. The general rule of thumb is that the nozzle size should be about 4-5 times larger than the size of the cells being interrogated. Based on measurements from microscope images, the myoblasts used in this study vary in size from 20-200 µm in length and about 10-20 µm in diameter. The lengthiness of human myoblast cells may affect cell sorting effectiveness. Cells were sorted based on the added SYTOX™ Blue Dead Cell Stain, which allowed the cytometer to distinguish between dead and live cells. The cytometer was also adjusted based on data gathered from the first cell sorting to crop the cell population to include as many viable cells as possible. The sorting produced around the same number of viable clones with minimal effort compared to manual colony picking, with the difference that only 9.72 % of the clones were viable compared to 53.73 % from colony picking. The c.197G>T corrected clone was successfully produced from this cell sorting. To increase the number of clones, even more, a third FACS was conducted with conditions adjusted to be even more gentle to the cells: 140 µm nozzle and 10 psi of pressure, while other factors remained the same. It appeared that clones started growing faster compared to preceding sortings. The full potential of this sorting could not be documented since the experiment ran out of time, although the general impression was that the larger the nozzle size, the better it works for myoblast cells.

The pursued gene editing that produced the corrected mutation site and PAM site change was present in 1 of the 84 (1.19 %) sequenced clonal cell lines, of which 71 were produced with guide#4 37 °C and 13 with guide#1 32 °C. The successfully corrected clonal cell line, with the desired c.197G>T correction and the PAM site edit generated by guide#4, exhibited no distinct off-target indels in the seven most likely off-target regions that were analyzed. The crRNA can tolerate up to five mismatches in the target sequence and many of these off-target sites are mutagenized with frequencies comparable to, or even

higher than those observed at the intended on-target site (Fu *et al.*, 2013). Understanding how to reduce off-target effects is crucial for accurate disease modeling and representation of acquired results. As discussed before, the use of ssODN over dsODN as an HDR template, and guiding the cell to favor HDR as a DSB repair pathway reduces random integration of the template to the genome and therefore undesired off-target gene editing (Bruntraeger *et al.*, 2019). Moreover, the use of RNP complex over plasmids as a delivery system for CRISPR-Cas9 also reduces random integration (Kim *et al.*, 2014).

Due to the lack of genomic cleavage and therefore minimal off-target effects, base editors are an excellent choice for generating single nucleotide alterations. Base editors are confined to specific base replacements and are better suited for creating heterozygous mutations rather than homozygous mutations, which would have made them an option to use also in this project (Eid *et al.*, 2018). Off-target gene editing might alter other parts of the genome thus resulting in differences in the cell lines considered to be isogenic. For example, differences in CHCHD10 mRNA expression levels might be the result of off-target editing and would not be reliable. Therefore, it is crucial for disease modeling, that off-target sites are screened for. Given the error-prone nature of CRISPR-Cas9, karyotyping and extensive sequencing should be performed to properly validate the quality of the corrected cell line obtained in this project and hence the validity of the disease model.

4.3 Conclusions and future considerations

In this project, important steps towards isogenic control myoblast-based disease modeling were taken with the optimization of CRISPR-Cas9 genome editing of human myoblast cells and subsequent augmentation of cell sorting methods for clonal cell line creation. The corrected isogenic clone obtained here can be employed in a wide range of downstream analyses and experiments to illuminate the function of CHCHD10 in skeletal muscle and its role in SMAJ. To elucidate the disease mechanism of SMAJ in myoblast cells, future work is required. The project originally included the characterization of mRNA expression and protein levels of CHCHD10 in both the patient and the corrected isogenic cell line to study only the effects of the mutation. However, the optimization of genome editing of the myoblast cell lines turned out to provide enough challenge for the scope of this thesis. Myoblast differentiation into myotubes and the inspection of myogenic regulatory factors (MRFs), as well as CHCHD10 levels in patient and healthy cells before and after differentiation, would provide exciting information about the disease. Similarly, the investigation of mitochondrial respiration in both the patient and healthy cell lines

would clarify the role of CHCHD10 in mtISR regulation. The more far-reaching project would be to correct the disease mutation in human iPSC and differentiate them into neurons to produce co-cultures of isogenic motor neurons and skeletal muscle cells to study the effect of the mutation on NMJ formation and morphology.

The creation of isogenic cell lines with CRISPR-Cas9 genome editing enables an important platform to investigate disease genes and their functions *in vivo*. Human myoblast cells isolated from patient biopsies provide the most pertinent experimental model to study neuromuscular atrophy-associated mutations in their natural genomic environment. In this project, one patient-specific genome-edited control myoblast cell line was successfully and precisely generated and partly validated as an isogenic disease model. The remaining challenges of using CRISPR-Cas9 for myoblast-based disease modeling include low genome editing efficiency in myoblast cells and low viability of clonal cell lines after sorting into single cells. Nevertheless, the advances taken here regarding myoblast genome editing with CRISPR-Cas9 offer a fertile avenue for future research of myoblasts genome manipulation, myogenic disorders, and the role of CHCHD10 in skeletal muscle and SMAJ.

5. Acknowledgements

I would first like to thank my supervisors Associate professor Henna Tyynismaa, Sandra Harjuhaahto MSc, and Tiina Rasila PhD for the guidance and support throughout the thesis project. Thank you, Henna, for the opportunity to work in your great research group. Thank you, Riitta Lehtinen and Jana Pennonen for your guidance in the laboratory and technical support. Thanks also to Julius Rönkkö, who helped me whenever needed. I would also like to thank the Institute for Molecular Medicine Finland (FIMM) for sequencing all the clonal cell lines and the Biomedicum Flow Cytometry Unit for sorting the edited cell populations. Lastly, I want to thank the whole Tyynismaa group for general warm atmosphere and supportive work environment.

6. References

Ajroud-Driss, S., Fecto, F., Ajroud, K., Lalani, I., Calvo, S.E., Mootha, V.K., Deng, H.-X., Siddique, N., Tahmoush, A.J., and Heiman-Patterson, T.D. (2015). Mutation in the novel nuclear-encoded mitochondrial protein CHCHD10 in a family with autosomal dominant mitochondrial myopathy. Neurogenetics *16*, 1-9.

Anderson, C.J., Bredvik, K., Burstein, S.R., Davis, C., Meadows, S.M., Dash, J., Case, L., Milner, T.A., Kawamata, H., and Zuberi, A. (2019). ALS/FTD mutant CHCHD10 mice reveal a tissue-specific toxic gain-of-function and mitochondrial stress response. Acta neuropathologica *138*, 103-121.

Auranen, M., Ylikallio, E., Shcherbii, M., Paetau, A., Kiuru-Enari, S., Toppila, J.P., and Tyynismaa, H. (2015). CHCHD10 variant p.(Gly66Val) causes axonal Charcot-Marie-Tooth disease. Neurology Genetics *1*.

Avior, Y., Sagi, I., and Benvenisty, N. (2016). Pluripotent stem cells in disease modelling and drug discovery. Nature reviews Molecular cell biology *17*, 170-182.

Bak, R.O., Gomez-Ospina, N., and Porteus, M.H. (2018). Gene editing on center stage. Trends in Genetics *34*, 600-611.

Bannwarth, S., Ait-El-Mkadem, S., Chaussenot, A., Genin, E.C., Lacas-Gervais, S., Fragaki, K., Berg-Alonso, L., Kageyama, Y., Serre, V., and Moore, D.G. (2014). A mitochondrial origin for frontotemporal dementia and amyotrophic lateral sclerosis through CHCHD10 involvement. Brain *137*, 2329-2345.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science *315*, 1709-1712.

Bassett, A.R. (2017). Editing the genome of hiPSC with CRISPR/Cas9: disease models. Mammalian Genome *28*, 348-364.

Bogdanove, A.J., and Voytas, D.F. (2011). TAL effectors: customizable proteins for DNA targeting. Science *333*, 1843-1846.

Bollen, Y., Post, J., Koo, B.-K., and Snippert, H.J. (2018). How to create state-of-the-art genetic model systems: strategies for optimal CRISPR-mediated genome editing. Nucleic acids research *46*, 6435-6454.

Bolotin, A., Quinquis, B., Sorokin, A., and Ehrlich, S.D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology *151*, 2551-2561.

Bonner, W., Hulett, H., Sweet, R., and Herzenberg, L. (1972). Fluorescence activated cell sorting. Review of Scientific Instruments *43*, 404-409.

Bruntraeger, M., Byrne, M., Long, K., and Bassett, A.R. (2019). Editing the genome of human induced pluripotent stem cells using CRISPR/Cas9 ribonucleoprotein complexes. In CRISPR Gene Editing, (Humana Press, New York, NY), pp. 153-183.

Cavallaro, G. (2010). Genome-wide analysis of eukaryotic twin CX9C proteins. Molecular BioSystems *6*, 2459. 10.1039/c0mb00058b.

Concordet, J.-P., and Haeussler, M. (2018). CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. Nucleic acids research *46*, W242-W245.

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

Cormack, B.P., Valdivia, R.H., and Falkow, S. (1996). FACS-optimized mutants of the green fluorescent protein (GFP). Gene *173*, 33-38.

Covello, G., Siva, K., Adami, V., and Denti, M.A. (2014). An electroporation protocol for efficient DNA transfection in PC12 cells. Cytotechnology *66*, 543-553.

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

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

Eid, A., Alshareef, S., and Mahfouz, M.M. (2018). CRISPR base editors: genome editing without double-stranded breaks. Biochemical Journal *475*, 1955-1964.

Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K., and Sander, J.D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nature biotechnology *31*, 822-826.

Funayama, M., Ohe, K., Amo, T., Furuya, N., Yamaguchi, J., Saiki, S., Li, Y., Ogaki, K., Ando, M., and Yoshino, H. (2015). CHCHD2 mutations in autosomal dominant late-onset Parkinson's disease: a genome-wide linkage and sequencing study. The Lancet Neurology *14*, 274-282.

Gaj, T., Gersbach, C.A., and Barbas III, C.F. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends in biotechnology *31*, 397-405.

Garneau, J.E., Dupuis, M.-È., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A.H., and Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature *468*, 67-71.

Genin, E.C., Madji Hounoum, B., Bannwarth, S., Fragaki, K., Lacas-Gervais, S., Mauri-Crouzet, A., Lespinasse, F., Neveu, J., Ropert, B., and Augé, G. (2019). Mitochondrial defect in muscle precedes neuromuscular junction degeneration and motor neuron death in CHCHD10S59L/+ mouse. Acta neuropathologica *138*, 123-145.

Genin, E.C., Plutino, M., Bannwarth, S., Villa, E., Cisneros-Barroso, E., Roy, M., Ortega-Vila, B., Fragaki, K., Lespinasse, F., Pinero-Martos, E., et al. (2016). CHCHD 10 mutations promote loss of mitochondrial cristae junctions with impaired mitochondrial genome maintenance and inhibition of apoptosis. EMBO Molecular Medicine *8*, 58-72. 10.15252/emmm.201505496.

Guo, Q., Mintier, G., Ma-Edmonds, M., Storton, D., Wang, X., Xiao, X., Kienzle, B., Zhao, D., and Feder, J.N. (2018). 'Cold shock'increases the frequency of homology directed repair gene editing in induced pluripotent stem cells. Scientific reports *8*, 1-11.

Hall, Z.W., and Sanes, J.R. (1993). Synaptic structure and development: the neuromuscular junction. Cell *72*, 99-121.

Harjuhaahto, S., Rasila, T.S., Molchanova, S.M., Woldegebriel, R., Kvist, J., Konovalova, S., Sainio, M.T., Pennonen, J., Torregrosa-Muñumer, R., and Ibrahim, H. (2020). ALS and Parkinson's disease genes CHCHD10 and CHCHD2 modify synaptic transcriptomes in human iPSC-derived motor neurons. Neurobiology of disease *141*, 104940.

Heyer, W.-D., Ehmsen, K.T., and Liu, J. (2010). Regulation of homologous recombination in eukaryotes. Annual review of genetics *44*, 113-139.

Hille, F., and Charpentier, E. (2016). CRISPR-Cas: biology, mechanisms and relevance. Philosophical transactions of the royal society B: biological sciences *371*, 20150496.

Hsiau, T., Conant, D., Rossi, N., Maures, T., Waite, K., Yang, J., Joshi, S., Kelso, R., Holden, K., and Enzmann, B.L. (2019). Inference of CRISPR edits from Sanger trace data. BioRxiv, 251082.

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

Huang, X., Wu, B.P., Nguyen, D., Liu, Y.-T., Marani, M., Hench, J., Bénit, P., Kozjak-Pavlovic, V., Rustin, P., and Frank, S. (2018). CHCHD2 accumulates in distressed mitochondria and facilitates oligomerization of CHCHD10. Human molecular genetics *27*, 3881-3900.

Imai, Y., Meng, H., Shiba-Fukushima, K., and Hattori, N. (2019). Twin CHCH proteins, CHCHD2, and CHCHD10: key molecules of Parkinson's disease, amyotrophic lateral sclerosis, and frontotemporal dementia. International journal of molecular sciences *20*, 908.

Ishino, Y., Krupovic, M., and Forterre, P. (2018). History of CRISPR-Cas from encounter with a mysterious repeated sequence to genome editing technology. Journal of bacteriology *200*, e00580-00517.

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., and 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*, 5429-5433.

Jansen, R., Embden, J.D.v., Gaastra, W., and Schouls, L.M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. Molecular microbiology *43*, 1565-1575.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. science *337*, 816-821.

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., and Doudna, J. (2013). RNA-programmed genome editing in human cells. elife *2*, e00471.

Jokela, M., Penttilä, S., Huovinen, S., Hackman, P., Saukkonen, A.M., Toivanen, J., and Udd, B. (2011). Late-onset lower motor neuronopathy: a new autosomal dominant disorder. Neurology *77*, 334-340.

Kim, S., Kim, D., Cho, S.W., Kim, J., and Kim, J.-S. (2014). Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. Genome research *24*, 1012-1019.

Kleinstiver, B.P., Prew, M.S., Tsai, S.Q., Topkar, V.V., Nguyen, N.T., Zheng, Z., Gonzales, A.P., Li, Z., Peterson, R.T., and Yeh, J.-R.J. (2015). Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature *523*, 481-485.

Koonin, E.V., and Makarova, K.S. (2019). Origins and evolution of CRISPR-Cas systems. Philosophical Transactions of the Royal Society B *374*, 20180087.

Lehmer, C., Schludi, M.H., Ransom, L., Greiling, J., Junghänel, M., Exner, N., Riemenschneider, H., van der Zee, J., Van Broeckhoven, C., and Weydt, P. (2018). A novel CHCHD10 mutation implicates a Mia40-dependent mitochondrial import deficit in ALS. EMBO molecular medicine *10*, e8558.

Li, H., Beckman, K.A., Pessino, V., Huang, B., Weissman, J.S., and Leonetti, M.D. (2019). Design and specificity of long ssDNA donors for CRISPR-based knock-in. BioRxiv, 178905.

Lieber, M.R. (2010). The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. Annual review of biochemistry *79*, 181-211.

Lin, S., Staahl, B.T., Alla, R.K., and Doudna, J.A. (2014). Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. elife *3*, e04766.

Liu, Y.-T., Huang, X., Nguyen, D., Shammas, M.K., Wu, B.P., Dombi, E., Springer, D.A., Poulton, J., Sekine, S., and Narendra, D.P. (2020). Loss of CHCHD2 and CHCHD10 activates OMA1 peptidase to disrupt mitochondrial cristae phenocopying patient mutations. Human molecular genetics *29*, 1547-1567.

Longen, S., Bien, M., Bihlmaier, K., Kloeppel, C., Kauff, F., Hammermeister, M., Westermann, B., Herrmann, J.M., and Riemer, J. (2009). Systematic analysis of the twin Cx9C protein family. Journal of molecular biology *393*, 356-368.

Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J., Charpentier, E., and Haft, D.H. (2015). An updated evolutionary classification of CRISPR–Cas systems. Nature Reviews Microbiology *13*, 722-736.

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

Malpass, K. (2013). Defective mitochondrial dynamics in the hot seat—a therapeutic target common to many neurological disorders? Nature Reviews Neurology *9*, 417-417.

Mamchaoui, K., Trollet, C., Bigot, A., Negroni, E., Chaouch, S., Wolff, A., Kandalla, P.K., Marie, S., Di Santo, J., and St Guily, J.L. (2011). Immortalized pathological human myoblasts: towards a universal tool for the study of neuromuscular disorders. Skeletal muscle *1*, 1-11.

Marraffini, L.A., and Sontheimer, E.J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. science *322*, 1843-1845.

Martherus, R.S., Sluiter, W., Timmer, E.D., VanHerle, S.J., Smeets, H.J., and Ayoubi, T.A. (2010). Functional annotation of heart enriched mitochondrial genes GBAS and CHCHD10 through guilt by association. Biochemical and biophysical research communications *402*, 203-208.

Mishra, P., and Chan, D.C. (2014). Mitochondrial dynamics and inheritance during cell division, development and disease. Nature reviews Molecular cell biology *15*, 634-646.

Mojica, F., Ferrer, C., Juez, G., and Rodríguez-Valera, F. (1995). Long stretches of short tandem repeats are present in the largest replicons of the Archaea Haloferax mediterranei and Haloferax volcanii and could be involved in replicon partitioning. Molecular microbiology *17*, 85-93.

Mojica, F.J., Díez-Villaseñor, C., García-Martínez, J., and Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. Journal of molecular evolution *60*, 174-182.

Mojica, F.J., Juez, G., and Rodriguez-Valera, F. (1993). Transcription at different salinities of Haloferax mediterranei sequences adjacent to partially modified PstI sites. Molecular microbiology *9*, 613-621.

Norio, R. (2003). Finnish disease heritage I. Human genetics *112*, 441-456.

Nussenzweig, P.M., and Marraffini, L.A. (2020). Molecular mechanisms of CRISPR-Cas immunity in bacteria. Annual Review of Genetics *54*, 93-120.

Nussenzweig, P.M., McGinn, J., and Marraffini, L.A. (2019). Cas9 cleavage of viral genomes primes the acquisition of new immunological memories. Cell host & microbe *26*, 515-526. e516.

Obinata, M. (2007). The immortalized cell lines with differentiation potentials: their establishment and possible application. Cancer science *98*, 275-283.

Peeters, K., Chamova, T., and Jordanova, A. (2014). Clinical and genetic diversity of SMN1-negative proximal spinal muscular atrophies. Brain *137*, 2879-2896.

Penttilä, S., Jokela, M., Bouquin, H., Saukkonen, A.M., Toivanen, J., and Udd, B. (2015). Late onset spinal motor neuronopathy is caused by mutation in CHCHD 10. Annals of Neurology *77*, 163-172.

Penttilä, S., Jokela, M., Huovinen, S., Saukkonen, A.M., Toivanen, J., Lindberg, C., Baumann, P., and Udd, B. (2014). Late-onset spinal motor neuronopathy–A common form of dominant SMA. Neuromuscular Disorders *24*, 259-268.

Ran, F., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nature protocols *8*, 2281-2308.

Recillas-Targa, F. (2006). Multiple strategies for gene transfer, expression, knockdown, and chromatin influence in mammalian cell lines and transgenic animals. Molecular biotechnology *34*, 337-354.

Rossor, A.M., Kalmar, B., Greensmith, L., and Reilly, M.M. (2012). The distal hereditary motor neuropathies. Journal of Neurology, Neurosurgery & Psychiatry *83*, 6-14.

Ruan, Y., Hu, J., Che, Y., Liu, Y., Luo, Z., Cheng, J., Han, Q., He, H., and Zhou, Q. (2022). CHCHD2 and CHCHD10 regulate mitochondrial dynamics and integrated stress response. Cell death & disease *13*, 1-12.

Rudin, N., Sugarman, E., and Haber, J.E. (1989). Genetic and physical analysis of double-strand break repair and recombination in Saccharomyces cerevisiae. Genetics *122*, 519-534.

Savić, N., and Schwank, G. (2016). Advances in therapeutic CRISPR/Cas9 genome editing. Translational Research *168*, 15-21.

Shen, B., Zhang, W., Zhang, J., Zhou, J., Wang, J., Chen, L., Wang, L., Hodgkins, A., Iyer, V., and Huang, X. (2014). Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. Nature methods *11*, 399-402.

Shrivastav, M., De Haro, L.P., and Nickoloff, J.A. (2008). Regulation of DNA double-strand break repair pathway choice. Cell research *18*, 134-147.

Skarnes, W.C., Pellegrino, E., and McDonough, J.A. (2019). Improving homology-directed repair efficiency in human stem cells. Methods *164*, 18-28.

Song, Z., Chen, H., Fiket, M., Alexander, C., and Chan, D.C. (2007). OPA1 processing controls mitochondrial fusion and is regulated by mRNA splicing, membrane potential, and Yme1L. The Journal of cell biology *178*, 749-755.

Tidball, A.M., Swaminathan, P., Dang, L.T., and Parent, J.M. (2018). Generating loss-of-function iPSC lines with combined CRISPR indel formation and reprogramming from human fibroblasts. Bio-protocol *8*.

Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S., and Gregory, P.D. (2010). Genome editing with engineered zinc finger nucleases. Nature Reviews Genetics *11*, 636-646.

Wang, H., La Russa, M., and Qi, L.S. (2016). CRISPR/Cas9 in genome editing and beyond. Annual review of biochemistry *85*, 227-264.

Xiao, Y., Zhang, J., Shu, X., Bai, L., Xu, W., Wang, A., Chen, A., Tu, W.-Y., Wang, J., and Zhang, K. (2020). Loss of mitochondrial protein CHCHD10 in skeletal muscle causes neuromuscular junction impairment. Human molecular genetics *29*, 1784-1796.

Zhou, Z.-D., Saw, W.-T., and Tan, E.-K. (2017). Mitochondrial CHCHD-containing proteins: physiologic functions and link with neurodegenerative diseases. Molecular neurobiology *54*, 5534-5546.