Genetic causes and stem-cell-based therapeutic strategies in neuromuscular diseases

Genetic causes and stem-cell-based therapeutic strategies in neuromuscular diseases

Le Guo

The studies presented in this thesis were performed within the MHeNs (School for Mental Health and Neuroscience) at Maastricht University.

Cover design by: Guillem Contestí Rubert / Antoni Vallbona Garcia
Layout by: Le Guo
Printed by: Gildeprint
ISBN: 978-94-6419-246-9

Genetic causes and stem-cell-based therapeutic strategies in neuromuscular diseases

Dissertation

to obtain the degree of Doctor at Maastricht University,
on the authority of the Rector Magnificus Prof.dr. Rianne M. Letschert
in accordance with the decision of the Board of Deans,
to be defended in public on
Wednesday, 23rd of June 2021 at 13:00 hours

By

Le Guo

Supervisor:
Prof. dr. H.J.M. (Bert) Smeets

Co-supervisors:
Dr. F.H.J. (Florence) van Tienen
Dr. M. (Mike) Gerards

Assessment committee:
Prof. dr. W.M. (Harry) Steinbusch (Chairman)
Dr. J.G.J. (Janneke) Hoeijmakers
Prof. dr. M. (Maurilio) Sampaolesi (KU Leuven, Belgium)
Prof. dr. G. (Geneviève) Gourdon (INSERM, France)

Contents

Chapter 1

General Introduction and Thesis outline

1. Introduction

1.1 Neuromuscular disease

Neuromuscular disease refers to a group of heterogeneous disorders that impair the function of motor nerves, muscles or neuromuscular junctions. The symptoms of neuromuscular disease vary according to the condition of patients and may be mild, moderate or life threatening. Some of the muscle symptoms include: muscular weakness, muscle wastage, muscular cramps, muscle spasticity, which later causes joint or skeletal deformities, muscle pain, breathing and swallowing difficulties. Meta-analysis of peer reviewed literature from 1990-2014 showed the prevalence rates for most neuromuscular diseases ranged from 1 to 10 per 100, 000/yr (1). Generally, neuromuscular disorders can be classified into four main groups, including motor neuron diseases, neuropathies, neuromuscular junction disorders and myopathies. Causes of neuromuscular disorders range from genetic mutations, injuries, viral infections, dietary deficiencies to certain drugs and poisons. To date, 1042 genetic neuromuscular diseases resulting from defects in 587 different genes have been identified in humans (2). Diagnosis can be performed by testing nerve conduction, electromyography, biochemistry, muscle biopsies and genes. Current treatment mainly focuses on palliative care and supportive management, although there are adequate therapies available for some types of neuromuscular diseases, like myasthenia gravis. And several novel therapeutic approaches are reaching the clinical trial stage (3).

Genetic neuromuscular diseases are usually complex and heterogeneous. A given gene or mutation may be involved in several different clinical manifestations, which is referred to phenotypic or clinical heterogeneity. Conversely, a given clinical entity may be produced by a defect in several alternative genes, which is named genetic heterogeneity. In neuromuscular diseases both kinds of heterogeneity occur, which increases the difficulty to identify the underlying genetic cause. Moreover, it has become increasingly clear that not all genetic neuromuscular diseases are monogenic and defects in multiple genes are detected, explaining the complex clinical manifestations.

1.1.1 Myotonic dystrophy: neuromuscular diseases with a single causative gene

Myotonic dystrophy type 1 (DM1; OMIM #160900) is an autosomal dominant

multisystem disease. Clinical symptoms include progressive skeletal muscle weakness, myotonia, cardiac defects, smooth muscle dysfunction and abnormalities of the eye, endocrine system, and central nervous system (4). DM1 is common with an estimated prevalence of 1 in 8000 worldwide (5). DM1 is always caused by an expansion of a CTG triplet in the 3' untranslated region of the dystrophia myotonica-protein kinase (DMPK) gene. The expanded repeat in the DMPK gene is transcribed into a toxic CUG repeat in the mRNA, which can be visualized as nuclear foci. The expanded transcripts sequester RNA-binding proteins, such as the muscle blind-like 1 (MBNL1) protein, resulting in aberrant splicing of a large number of downstream genes. These aberrant splicing events have been proposed to contribute to the multisystem clinical presentation of DM1 (6). The CTG repeat in healthy individuals is below 38 repeat units, whereas DM1 patients have between 50 and several thousand repeat units. Expanded repeats are mosaics. The age at onset and the severity of symptoms positively correlate with the repeat length (7). Besides, as the CTG triplet is unstable and inclines to expand when passing on to the next generation, anticipation is observed in successive generations demonstrated by an earlier age at onset, even congenitally, with increasingly severe symptoms in the younger generations. Based on the severity of the clinical features, DM1 patients can be divided into three major groups: congenital (almost exclusively transmitted by the mother), classical (adult-onset) and mild (asymptomatic) type (Table 1) (4, 8).

Table 1: Overview of DM1 classification

Subtypes	Symptoms	CTG length	Age of onset
Congenital	Infantile hypotonia, respiratory deficits, mental retardation, classic signs in adults	>1000	Birth to 10 years
Classical	Muscle weakness, myotonia, cataracts, cardiac defects	~100-~1000	10-30 years
Mild	Mild myotonia, premature cataracts	50-~150	>20 years

DM1 is initially diagnosed by physical examination and family history. Classical DM1 patients have a characteristic facial appearance of wasting and weakness of the jaw and neck muscles (9). All DM1 patients, reported so far, carry an expanded CTG repeat in the DMPK gene and therefore the genetic test has become the prime test to confirm the diagnosis, when clinically DM1 is suspected, thereby replacing other clinical and laboratory investigations, such as electromyography and muscle

biopsies. In terms of genetic diagnosis, conventional PCR with fluorescently-labeled primers flanking the CTG repeat region followed by fragment-length (GeneScan) analysis is performed to identify repeat size below 150. Larger repeat sizes (over 150 repeats) can be detected by electrophoresis and Southern blotting of the PCR product, followed by hybridization with a fluorescent oligonucleotide probe, detecting the repeat expansion. This gives a rough indication of the repeat size. However, in some cases false-negative results can be obtained due either the extreme size of the repeat expansion (>1000 repeats) or other PCR issues (5). Therefore, triplet repeat-primed PCR (TP-PCR) can be performed to detect any expanded CTG repeat (10). Although TP-PCR cannot determine the exact number of CTG repeat, it is a fast and cost-effective method to identify the presence of expanding CTG repeats and establishing the diagnosis DM1.

1.1.2 Mitochondrial disease: neuromuscular disease with multiple gene mutations

Mitochondria are essential intracellular organelles predominantly responsible for oxidative phosphorylation (OXPHOS), resulting in the production of ATP, a direct energy source for cells. The ATP production relies on five multi-subunit polypeptide complexes [I – V] located within the inner mitochondrial membrane. Except for complex II, which is entirely encoded by the nuclear genome, the other four complexes are encoded by both the nuclear and mitochondrial genomes. The human mitochondrial genome (mtDNA) is a 16.6 kb double-stranded DNA encoding 13 respiratory chain proteins, 22 tRNAs and 2 rRNAs, which is required for mtDNA encoded protein synthesis. As the mtDNA is present in multiple copies within cells, mtDNA variants can be either homoplasmic (all mtDNA copies carry this variant) or heteroplasmic (part of mtDNA copies carry the variant). In case of heteroplasmic, pathogenic mtDNA variants, a biochemical OXPHOS defect and subsequent clinical manifestations usually occur when the mtDNA mutation exceeds a tissue-specific threshold. The vast majority of proteins in mitochondrial function (>1500) are encoded by the nuclear genome and imported into the mitochondria. These are proteins involved in the respiratory chain as structural subunits or assembly factors, enzymes involved in mtDNA replication, fusion and fission proteins or proteins more indirectly connected to mitochondrial function.

Mitochondrial disease is a group of genetically and clinically heterogeneous diseases. The genetic defect can be either in the mtDNA or nuclear genome. Patients with mitochondrial disease show a diverse clinical spectrum, ranging from ataxia, epilepsy, myopathy and sensorineural deafness (neurological), optic atrophy, ophthalmoplegia and ptosis (ophthalmological), cardiomyopathy (cardiac) to diabetes (endocrinal) (11). Mitochondrial diseases can be specific, syndromic or involve common manifestations, like deafness, diabetes and/or cardiomyopathy, not necessarily primary mitochondrial diseases. It can be a multisystemic manifestation or only affect a single organ, like in Leber Hereditary Optic Neuropathy (LHON). The genetic heterogeneity of mitochondrial disease can be best illustrated by Leigh syndrome, which can be caused mutations in more than 75 genes, either located in the mtDNA or in the nuclear genome (12). Vice versa, a single mutation, like the A3243G $tRNA^{Leu(UUR)}$ mutation in mtDNA, may cause various phenotypes, including maternally inherited diabetes and deafness (MIDD), mitochondrial encephalomyopathy, lactic acidosis and stroke like episodes (MELAS), cardiomyopathy, renal dysfunction or chronic progressive external ophthalmoplegia (CPEO) (13).

Diagnosis of mitochondrial disease used to follow a phenotype to genotype approach. For patients suspect of mitochondrial disease their family histories need to be checked. Maternal inheritance in a patient suggests mtDNA defects as the cause of mitochondrial disease. However, as nDNA mutations can cause mitochondrial disease, the inheritance mode of mitochondrial disease may also be autosomal recessive, autosomal dominant or X-linked. Besides, de novo mutations in both mtDNA and nDNA can cause mitochondrial disease as well. Recognition of classic syndromes such as MIDD, MELAS and CPEO helps the diagnosis of mitochondrial disease. Multi-organ involvement of symptoms suggests mitochondrial disease, although there are exceptions such as Leber hereditary optic neuropathy (LHON), in which only the eye is involved. Moreover, clinical or laboratory examinations such as lactic acidemia, characteristic MRI findings and abnormal immunohistochemistry staining in muscle biopsy also contribute clues for confirming the diagnosis of mitochondrial disease (11, 14). Biochemical tests such as OXPHOS enzyme activity measurement are important to validate novel genes and pathogenic mutations. Several classification systems have been established, including and weighing all clinical and laboratory parameters (for example the Nijmegen clinical criteria for mitochondrial disease (15) or the Newcastle Mitochondrial disease adult scale – NMDAS (16)).

As often there is no straight line from phenotype to genotype and as novel mitochondrial genes are still being idenitified, diagnosis of mitochondrial disease is increasingly becoming a genotype to phenotype approach. This is possible due to the rapid development and reduced cost of massive parallel sequencing. Generally, patients suspect of mitochondrial disease are first screened for mutations in mtDNA, if excluded, continued to be analyzed by whole exome sequencing (WES) or whole genome sequencing (WGS). In a recent study, a two-step NGS-based approach revealed nearly 70% of disease-causing mutations in a cohort of 117 patients with the phenotype suggesting a genetic mitochondrial disease or a genetic neuromuscular disease, with a mitochondrial cause as one of the options, and 30% of the disease-causing mutations were identified in novel genes, previously not known to be mitochondrial (17). Compared to conventional Sanger sequencing, which only solved 11% (18), NGS approach is a huge leap forward both in success rate and turn-around time of weeks instead of years.

1.2 Advances in gene and cell therapy

As no effective cure is currently available for neuromuscular disease, many researchers turn their attention to develop novel gene and cell therapies, especially by gene editing and stem cell therapy. Gene therapy is the introduction, removal or alteration in genetic material, including DNA or RNA, into specific target cells or tissues of a patient for the purpose of curing or altering symptoms. Typically, genetic material is transferred into the target cell by a vector, such as adeno-associated virus (AAV) or lentivirus. Once inside the cell, a copy of exogenous gene will help produce functioning proteins despite the presence of a mutated gene, thus improving the health of the patient. Based on the delivery methods and conditions of the target cells, gene therapy can be divided into *in vivo* approach, which targets cells directly in the patients, and *ex vivo* approach, which modifies somatic stem or progenitor cells in culture, followed by transplantation back to the patient. Cell therapy is the transfer of beneficial cells into a patient with the goal of curing a disease. One promising approach is the use of autologous cells through gene-modified cell therapy, similar to *ex vivo* approach, which collects the target cells from the patient, then either introduces an exogenous wild-type gene or corrects an endogenous mutated gene, and eventually administer these autologous modified cells to the patient.

1.2.1 Gene editing

Besides adding a functioning gene copy as in the traditional gene therapy, gene editing can remove or correct the mutant gene using engineered nucleases, such as zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and Cas9. An overview of the three engineered nucleases is shown in Table 2 (19).

Table 2: Comparison of different engineered nucleases

Characteristic	ZFNs	TALENs	CRISPR/Cas9
Specificity determinant	Zinc-finger proteins	Transcription activator-like effectors	CRISPR RNA of guide RNA
Nucleases	*Fok*I	*Fok*I	Cas9
Restriction in target site	G-rich	Start with T and end with A	PAM
Specificity	Low	Medium	High
Ease of engineering	Difficult	Moderate	Easy
Ease of multiplexing	Low	Moderate	High
Off-target effects	Moderate	Low	Variable
Cytotoxicity	Moderate	Moderate	Low
Cost	High	Moderate	Low

The common basis for all gene editing approaches relies on endonucleases to generate double-strand breaks (DSBs). The DSB is then repaired by cellular machinery either via nonhomologous end joining (NHEJ) pathway, which leads to imprecise insertion/deletion (indel) mutations, or via homology-directed repair (HDR) pathway, which requires an exogenous DNA template and can generate a precise modification at the target locus (20). The phase of the cell cycle determines the repair pathway choice between NHEJ and HDR. NHEJ directly ligates broken DNA ends throughout the cell cycle, whereas HDR is restricted to the S and G2 phases when the sister chromatid is available as a repair template (21). Hence, engineered nucleases can effectively generate NHEJ-mediated mutations in most cell types, whereas HDR-mediated editing generally does not occur in post-mitotic cells, such as myofibers and neurons.

- Zinc finger nucleases (ZFNs): ZFNs rely on DNA-binding motif of Cys2-His2 zinc finger domain for DNA target recognition (22, 23). Up to 18 bp of DNA can be recognized by arrays of zinc finger domains (24). By fusing zinc fingers to other nucleases, like FokI, ZFNs acquire the power of programmable genome engineering. However, major drawbacks of ZFNs include the difficulties of designing linkage between each zinc finger unit and targeting any desired sequence, partly due to interference between adjoining zinc fingers.
- Transcription activator-like effector nuclease (TALEN): By switching from zinc finger domains to transcription activator-like effectors (TALEs) from Xanthomonas bacteria, an advanced version of genome engineering nuclease (TALENs) is generated. TALENs can recognize a single nucleotide based on a basic unit of a repeat variable di-residue and be assembled into arrays to target specific DNA sequences (25, 26). As no linkage of repeat di-residues is required and less interactions between each domain, TALENs are easy to design and have broader choices of target regions. However, cloning and maintaining highly repetitive TALEs is challenging for TALENs.
- CRISPR/Cas9: Adapted from a bacterial immune system against bacteriophage, CRISPR/Cas9 system is simple to design and use (27). The DNA-targeting principle is based on Watson-Crick RNA-DNA pairing instead of protein-DNA interactions, as in ZFNs and TALENs. A single guide RNA (gRNA) is required to direct Cas9 to the specific DNA target next to a protospacer adjacent motif (PAM) sequence. Depending on the origin of Cas9, the PAM sequence varies, with 5'-NGG-3' for spCas9 and 5'-NNGRRT-3' for SaCas9. The combination of different kinds of Cas9 and gRNA offers the CRISPR/Cas9 system advantages of highly adaptable and accessible, promoting diverse application for this novel genome-editing tool. During the past decade, numerous studies have utilized the CRISPR/Cas9 system to modify specific sequences in the genome or correct monogenic diseases in cultured cells (28), animal models (29) and multiple postnatal organs, such as brain (30), liver (29) and muscle (31).

Patients suffered from neuromuscular diseases usually have limited choices of treatment. However, the development of CRISPR/Cas9 system hold promising future for curing neuromuscular diseases, such as DM1 and mitochondrial disease. Several groups used a CTG expansion deletion strategy in human DM1 myoblasts, fibroblasts, induced pluripotent stem cells (iPSCs), iPSC-derived myogenic cells, iPSC-derived neural stem cells, HEK293 cells and a transgenic DMSXL mouse model, and reported that targeted CTG expansion deletion led to complete

disappearance of ribonuclear foci, restoration of the localization of MBNL proteins and correction of alternative splicing patterns (32-35). However, some studies reported the deletion strategy resulted in undesired inversion of CTG repeats in edited DM1 myogenic clones (32) and large deletion in HEK293 cells (36), presumably due to the imperfect NHEJ repair of DSBs induced by the CRISPR/Cas9 system. Another study group designed an HDR-based polyadenylation signal insertion strategy to induce termination of premature (CUG)n expansion transcripts (37). Their results showed elimination of toxic RNA transcripts and reversal of aberrant alternative splicing in iPSC, neural stem cells, cardiomyocytes and skeletal muscle myofibers. Still, this insertion strategy has risks of detrimental effects of the remaining CTG expansion and DM1 anti-sense transcripts (38, 39). Alternatively, transcriptional inhibition by dCas9/gRNA (40) and expanded (CUG)n RNA degradation by RNA-targeting dCas9 (41) were shown to rescue the molecular and cellular phenotypes in DM1 cell lines, while the duration remains to be tested.

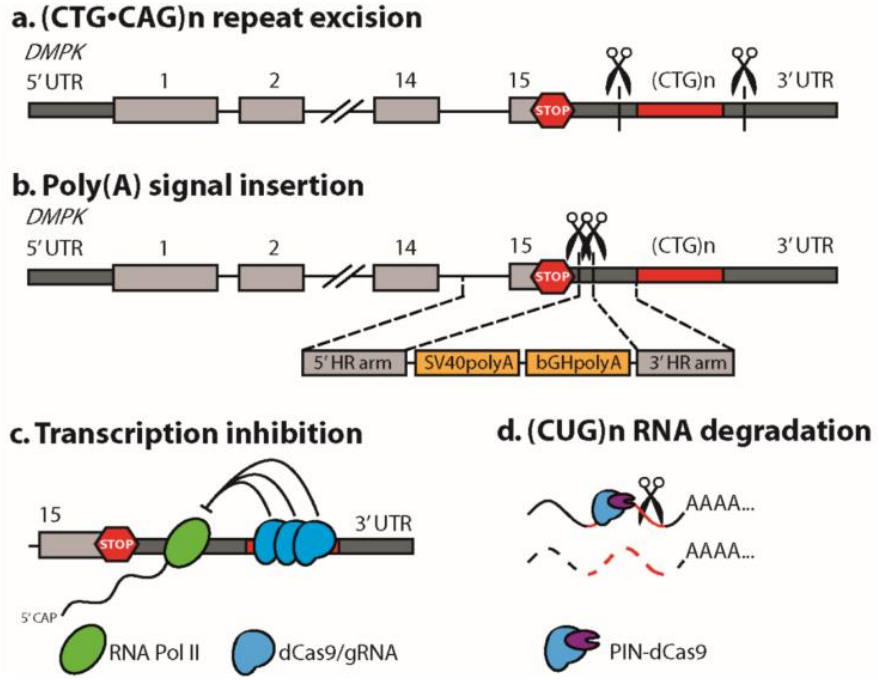


Figure 1: Four gene editing strategies for DM1. Gene editing strategies through (CTG•CAG)n repeat excision (a), poly(A) signal insertion (b), transcription inhibition by dCas9 to the expanded (CTG)n repeat (c), or (CUG)n RNA degradation by RNA-targeting dCas9 (d) (cited from (42)).

1.2.2 Stem cell therapy

Stem cell therapy is based on the transplantation of large numbers of live stem cells to regenerate the damaged tissue or organ of the patient. In neuromuscular disease, one of the important therapeutic targets is muscle, thus the candidate stem cells for treating neuromuscular disease should have myogenic capability. The source of therapeutic cells can be healthy, immune-compatible donors (allogeneic), or genetically corrected patient's own cells (autologous). Several cell types have been found and tested in animal models and clinical trials, and some of them hold great potential in developing treatments for neuromuscular disease, like Duchenne muscular dystrophy (DMD).

- Myoblasts are mononuclear muscle progenitor cell which can either proliferate or undergo terminal differentiation (43). Intramuscular transplanted healthy myoblasts resulted in fusion with dystrophic muscle fibres and formation of new dystrophin-positive fibres in mdx mouse models, with a partial or complete rescue of the phenotype (44, 45). However, clinical trials in DMD patients received no improvements in muscle strength (46), which is most likely due to the immune-rejection, poor survival and limited migration of injected myoblasts (47, 48). Subsequent studies applied immune suppression, more cells with multiple injections, resulted in 26% myofibers containing wild-type dystrophin
- Satellite cells are the major contributor to skeletal muscle regeneration, defined by their peripheral position beneath the basal lamina of myofibers (49). Upon injury in healthy adults, satellite cells leave their quiescent status, start to differentiate and eventually fuse to form new fibres (50, 51). Because of their self-renewal ability, satellite cells are more preferable for stem cell therapy than myoblasts. Besides, mouse muscles transplanted with satellite cells showed better engraftment than the muscles transplanted with myoblasts (52). However, human satellite cells have comparable issues such as limited migration, reduced proliferation potency after intra-muscular injection and lack of systemic delivery (53).
- Mesoangioblasts or pericytes are vessels-associated multipotent cells that are destined to differentiate into mesodermal cell lineages, including skeletal muscle cells. Mesoangioblasts were shown to ameliorate muscle function in dystrophic dogs (54) and mice (55, 56). A Phase I/II clinical trial with intra-arterial infusions of donor mesoangioblasts in pediatric DMD patients proved the feasibility and relative safety of the method, but there was no/minimal clinical

benefit (57). Although not designed for efficacy, reasons behind the lack of efficacy may be not reaching the optimal dose and immunosuppression applied. Further improvements include switching to autologous mesoangioblasts and starting the injection at an earlier age.

Criteria for optimal stem cells for treating neuromuscular diseases include being abundantly available from tissues, expandable *in vitro* without losing proliferation potency, easy homing to skeletal muscle through systemic delivery and efficient differentiation into skeletal muscle *in vivo*. Pericytes or mesoangioblasts fulfill all these criteria, which makes them an ideal candidate cell type to develop an effective stem cell therapy for neuromuscular disease.

1.3 Aims and outline of this thesis

Developing a therapy for neuromuscular disorders, first requires identifying the genetic defect. This can be straightforward as in myotonic dystrophy type 1, where the same CTG-repeat expansion causes the disease in every patient or it can be a challenge as in mitochondrial disorders, due to the large clinically (various clinical phenotypes) and genetically (over 1500 genes involved). However, whole exome sequencing brings the power of solving the latter issue, enabling us to identify the underlying genetic cause and to characterize the underlying pathophysiological process. Still, despite knowing the genetic cause in myotonic dystrophy for over 27 years, no cure is available for the vast majority of patients. Also, in mitochondrial disease, only few examples exist in which the genetic defect directly reveals a successful treatment. But there is light at the end of the tunnel. CRISPR/Cas technology offers us the opportunity to manipulate the genome to develop gene-based therapy and to better understand the disease mechanism. The work describes in this thesis starts with identifying the genetic defect in mitochondrial diseases and moves towards *ex vivo* correction of the genetic defects in mesoangioblasts of DM1 patients, preparing them for autologous therapy. The aims of this thesis are:

1) **To identify and functionally characterize gene defects in mitochondrial disease**. Patients with a possible mitochondrial disease will be screened by next-generation sequencing for mutations in the mitochondrial DNA and by whole exome sequencing for mutations in nuclear genes. Variants will be filtered by bioinformatics approaches and, in case of novel genes/variants, functionally validated in cell lines by complementation studies.

2) **To develop a generic autologous stem cell therapies to treat muscle pathology in the neuromuscular disorders DM1**. Mesoangioblasts, which are myogenic stem cells, seem to be particularly suited for clinical application. After correcting the gene defect with the CRISPR/Cas9 gene editing technology, patient-derived mesoangioblasts can be used as a potential autologous gene and stem cell treatment for regeneration of healthy muscle.

Chapter 2-3 describe the identification and characterization of the genetic defects in 2 families with mitochondrial diseases using whole exome sequencing. **Chapter 2** reports a boy of Leigh syndrome, with complex I and IV deficiency, in whom we identified two heterozygous pathogenic mutations in the *SLIRP* gene. This protein has been previously reported to be involved in mitochondrial OXPHOS function. Transduction of patient-derived fibroblasts with wild-type *SLIRP* but not mutant-type *SLIRP* not only restored mitochondrial encoded mRNA transcripts expression, it also recovered mitochondrial mass and enzymatic activity in complex I and IV. These results reveal *SLIRP* mutations as a novel cause of Leigh syndrome.

Chapter 3 describes gene defects identified in an Indian patient from a consanguineous family. The patient presented with progressive external ophthalmoplegia (PEO) and multiple mtDNA deletions, suggestive of mitochondrial disease. For this patient, whole exome sequencing revealed a homozygous deletion in *C1QBP* gene, which was responsible for the mtDNA multiple deletions and the genetic cause of PEO.

Chapter 4 describes our in-depth characterization of mesoangioblasts in 30 DM1 patients, including clinical characteristics, CTG repeat size, mtDNA copy number and mitochondrial function. The purpose is to characterize the pathophysiology in mesoangioblasts, identify possible treatment targets and define the read-out parameters to monitor the correction strategy (chapter 5).

Chapter 5 explores a novel CRISPR/Cas9-based approach for removing the pathological effect of the DM1 repeat expansion by abolishing a normal splice acceptor site, forcing the use of an alternative splice acceptor, which leads to skipping of the expanded CTG-repeat from the mRNA. This approach resulted in a sufficiently high total editing efficiency and an increased expression of the alternatively spliced product without the expanded CTG-repeat in six edited patient-derived cell lines. Moreover, on/off-target effects were evaluated by next-generation sequencing platform.

Chapter 6, the general discussion, provides an overview of gene editing therapies and on-going clinical trials and compares our gene editing approach in this broader context.

Chapter 7, the impact paragraph, describes the impact of the research presented in this thesis in getting more genetic diagnoses quicker and cheaper, and in developing gene/stem cell therapy in neuromuscular diseases.

Chapter 8 gives a summary of the entire thesis.

References

1. Deenen JCW, Horlings CGC, Verschuuren JJGM, Verbeek ALM, van Engelen BGM. The Epidemiology of Neuromuscular Disorders: A Comprehensive Overview of the Literature. Journal of Neuromuscular Diseases. 2015;2(1):73-85.
2. Benarroch L, Bonne G, Rivier F, Hamroun D. The 2020 version of the gene table of neuromuscular disorders (nuclear genome). Neuromuscular Disorders. 2019;29(12):980-1018.
3. Jungbluth H, Treves S, Zorzato F, Sarkozy A, Ochala J, Sewry C, et al. Congenital myopathies: disorders of excitation–contraction coupling and muscle contraction. Nature Reviews Neurology. 2018;14(3):151-67.
4. Meola G, Cardani R. Myotonic dystrophies: An update on clinical aspects, genetic, pathology, and molecular pathomechanisms. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease. 2015;1852(4):594-606.
5. Udd B, Krahe R. The myotonic dystrophies: molecular, clinical, and therapeutic challenges. The Lancet Neurology. 2012;11(10):891-905.
6. Lee Johanna E, Cooper Thomas A. Pathogenic mechanisms of myotonic dystrophy. Biochemical Society Transactions. 2009;37(6):1281-6.
7. Tsilfidis C, MacKenzie AE, Mettler G, Barceló J, Korneluk RG. Correlation between CTG trinucleotide repeat length and frequency of severe congenital myotonic dystrophy. Nature Genetics. 1992;1(3):192-5.
8. Bird TD. Myotonic Dystrophy Type 1. [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2020. [updated 2019 Oct 3. Available from: https://www.ncbi.nlm.nih.gov/books/NBK1165/.
9. National Human Genome Research Institute. About Myotonic Dystrophy [Available from: https://www.genome.gov/Genetic-Disorders/Myotonic-Dystrophy.
10. Warner JP, Barron LH, Goudie D, Kelly K, Dow D, Fitzpatrick DR, et al. A general method for the detection of large CAG repeat expansions by fluorescent PCR. Journal of Medical Genetics. 1996;33(12):1022-6.
11. Schapira AHV. Mitochondrial diseases. The Lancet. 2012;379(9828):1825-34.
12. Lake NJ, Compton AG, Rahman S, Thorburn DR. Leigh syndrome: One disorder, more than 75 monogenic causes. Ann Neurol. 2016;79(2):190-203.
13. Janssen GM, Hensbergen PJ, van Bussel FJ, Balog CI, Maassen JA, Deelder AM, et al. The A3243G tRNALeu(UUR) mutation induces mitochondrial dysfunction and variable disease expression without dominant negative acting translational defects in complex IV subunits at UUR codons. Hum Mol Genet. 2007;16(20):2472-81.
14. Ng YS, Turnbull DM. Mitochondrial disease: genetics and management. J Neurol. 2016;263(1):179-91.
15. Morava E, van den Heuvel L, Hol F, de Vries MC, Hogeveen M, Rodenburg RJ, et al. Mitochondrial disease criteria. Diagnostic applications in children. 2006;67(10):1823-6.

16. Schaefer AM, Phoenix C, Elson JL, McFarland R, Chinnery PF, Turnbull DM. Mitochondrial disease in adults: A scale to monitor progression and treatment. Neurology. 2006;66(12):1932-4.
17. Theunissen TEJ, Nguyen M, Kamps R, Hendrickx AT, Sallevelt S, Gottschalk RWH, et al. Whole Exome Sequencing Is the Preferred Strategy to Identify the Genetic Defect in Patients With a Probable or Possible Mitochondrial Cause. Front Genet. 2018;9:400.
18. Neveling K, Feenstra I, Gilissen C, Hoefsloot LH, Kamsteeg EJ, Mensenkamp AR, et al. A post-hoc comparison of the utility of sanger sequencing and exome sequencing for the diagnosis of heterogeneous diseases. Hum Mutat. 2013;34(12):1721-6.
19. Im W, Moon J, Kim M. Applications of CRISPR/Cas9 for Gene Editing in Hereditary Movement Disorders. J Mov Disord. 2016;9(3):136-43.
20. Mao Z, Bozzella M, Seluanov A, Gorbunova V. DNA repair by nonhomologous end joining and homologous recombination during cell cycle in human cells. Cell Cycle. 2008;7(18):2902-6.
21. Symington LS, Gautier J. Double-Strand Break End Resection and Repair Pathway Choice. Annual Review of Genetics. 2011;45(1):247-71.
22. Gersbach CA, Gaj T, Barbas CF, 3rd. Synthetic zinc finger proteins: the advent of targeted gene regulation and genome modification technologies. Acc Chem Res. 2014;47(8):2309-18.
23. Wolfe SA, Nekludova L, Pabo CO. DNA Recognition by Cys2His2 Zinc Finger Proteins. Annual Review of Biophysics and Biomolecular Structure. 2000;29(1):183-212.
24. Liu Q, Segal DJ, Ghiara JB, Barbas CF. Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. Proceedings of the National Academy of Sciences. 1997;94(11):5525.
25. Moscou MJ, Bogdanove AJ. A Simple Cipher Governs DNA Recognition by TAL Effectors. Science. 2009;326(5959):1501.
26. Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A, et al. Targeting DNA Double-Strand Breaks with TAL Effector Nucleases. Genetics. 2010;186(2):757.
27. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, et al. CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. Science. 2007;315(5819):1709.
28. Cao J, Wu L, Zhang SM, Lu M, Cheung WK, Cai W, et al. An easy and efficient inducible CRISPR/Cas9 platform with improved specificity for multiple gene targeting. Nucleic Acids Res. 2016;44(19):e149.
29. Singh K, Evens H, Nair N, Rincón MY, Sarcar S, Samara-Kuko E, et al. Efficient In Vivo Liver-Directed Gene Editing Using CRISPR/Cas9. Molecular Therapy. 2018;26(5):1241-54.
30. Swiech L, Heidenreich M, Banerjee A, Habib N, Li Y, Trombetta J, et al. In vivo interrogation of gene function in the mammalian brain using CRISPR-Cas9. Nature

biotechnology. 2015;33(1):102-6.
31. Tabebordbar M, Zhu K, Cheng JK, Chew WL, Widrick JJ, Yan WX, et al. In vivo gene editing in dystrophic mouse muscle and muscle stem cells. Science. 2016;351(6271):407-11.
32. Provenzano C, Cappella M, Valaperta R, Cardani R, Meola G, Martelli F, et al. CRISPR/Cas9-Mediated Deletion of CTG Expansions Recovers Normal Phenotype in Myogenic Cells Derived from Myotonic Dystrophy 1 Patients. Mol Ther Nucleic Acids. 2017;9:337-48.
33. van Agtmaal EL, Andre LM, Willemse M, Cumming SA, van Kessel IDG, van den Broek W, et al. CRISPR/Cas9-Induced (CTGCAG)n Repeat Instability in the Myotonic Dystrophy Type 1 Locus: Implications for Therapeutic Genome Editing. Mol Ther. 2017;25(1):24-43.
34. Dastidar S, Ardui S, Singh K, Majumdar D, Nair N, Fu Y, et al. Efficient CRISPR/Cas9-mediated editing of trinucleotide repeat expansion in myotonic dystrophy patient-derived iPS and myogenic cells. Nucleic Acids Res. 2018;46(16):8275-98.
35. Lo Scrudato M, Poulard K, Sourd C, Tomé S, Klein AF, Corre G, et al. Genome Editing of Expanded CTG Repeats within the Human DMPK Gene Reduces Nuclear RNA Foci in the Muscle of DM1 Mice. Molecular Therapy. 2019;27(8):1372-88.
36. Ikeda M, Taniguchi-Ikeda M, Kato T, Shinkai Y, Tanaka S, Hagiwara H, et al. Unexpected Mutations by CRISPR-Cas9 CTG Repeat Excision in Myotonic Dystrophy and Use of CRISPR Interference as an Alternative Approach. Mol Ther Methods Clin Dev. 2020;18:131-44.
37. Wang Y, Hao L, Wang H, Santostefano K, Thapa A, Cleary J, et al. Therapeutic Genome Editing for Myotonic Dystrophy Type 1 Using CRISPR/Cas9. Mol Ther. 2018;26(11):2617-30.
38. Zu T, Gibbens B, Doty NS, Gomes-Pereira M, Huguet A, Stone MD, et al. Non-ATG–initiated translation directed by microsatellite expansions. Proceedings of the National Academy of Sciences. 2011;108(1):260-5.
39. Gudde A, van Heeringen SJ, de Oude AI, van Kessel IDG, Estabrook J, Wang ET, et al. Antisense transcription of the myotonic dystrophy locus yields low-abundant RNAs with and without (CAG)n repeat. RNA Biol. 2017;14(10):1374-88.
40. Pinto BS, Saxena T, Oliveira R, Mendez-Gomez HR, Cleary JD, Denes LT, et al. Impeding Transcription of Expanded Microsatellite Repeats by Deactivated Cas9. Mol Cell. 2017;68(3):479-90 e5.
41. Batra R, Nelles DA, Pirie E, Blue SM, Marina RJ, Wang H, et al. Elimination of Toxic Microsatellite Repeat Expansion RNA by RNA-Targeting Cas9. Cell. 2017;170(5):899-912 e10.
42. Raaijmakers RHL, Ripken L, Ausems CRM, Wansink DG. CRISPR/Cas Applications in Myotonic Dystrophy: Expanding Opportunities. Int J Mol Sci. 2019;20(15).
43. Verhaart IEC, Aartsma-Rus A. Therapeutic developments for Duchenne muscular dystrophy. Nat Rev Neurol. 2019;15(7):373-86.

44. Partridge TA, Morgan J, Coulton G, Hoffman E, Kunkel L. Conversion of mdx myofibres from dystrophin-negative to-positive by injection of normal myoblasts. Nature. 1989;337(6203):176-9.
45. Morgan JE, Hoffman EP, Partridge TA. Normal myogenic cells from newborn mice restore normal histology to degenerating muscles of the mdx mouse. Journal of Cell Biology. 1990;111(6):2437-49.
46. Mendell JR, Kissel JT, Amato AA, King W, Signore L, Prior TW, et al. Myoblast Transfer in the Treatment of Duchenne's Muscular Dystrophy. New England Journal of Medicine. 1995;333(13):832-8.
47. Vilquin J. Myoblast transplantation: clinical trials and perspectives. Mini-review. Acta myologica: myopathies and cardiomyopathies: official journal of the Mediterranean Society of Myology. 2005;24(2):119-27.
48. Qu Z, Balkir L, van Deutekom JCT, Robbins PD, Pruchnic R, Huard J. Development of Approaches to Improve Cell Survival in Myoblast Transfer Therapy. Journal of Cell Biology. 1998;142(5):1257-67.
49. Mauro A. Satellite cell of skeletal muscle fibers. The Journal of Biophysical and Biochemical Cytology. 1961;9(2):493-5.
50. Collins CA, Olsen I, Zammit PS, Heslop L, Petrie A, Partridge TA, et al. Stem cell function, self-renewal, and behavioral heterogeneity of cells from the adult muscle satellite cell niche. Cell. 2005;122(2):289-301.
51. Dhawan J, Rando TA. Stem cells in postnatal myogenesis: molecular mechanisms of satellite cell quiescence, activation and replenishment. Trends Cell Biol. 2005;15(12):666-73.
52. Sacco A, Doyonnas R, Kraft P, Vitorovic S, Blau HM. Self-renewal and expansion of single transplanted muscle stem cells. Nature. 2008;456(7221):502-6.
53. Sun C, Serra C, Lee G, Wagner KR. Stem cell-based therapies for Duchenne muscular dystrophy. Exp Neurol. 2020;323:113086.
54. Sampaolesi M, Blot S, D'Antona G, Granger N, Tonlorenzi R, Innocenzi A, et al. Mesoangioblast stem cells ameliorate muscle function in dystrophic dogs. Nature. 2006;444(7119):574-9.
55. Sampaolesi M, Torrente Y, Innocenzi A, Tonlorenzi R, D'Antona G, Pellegrino MA, et al. Cell Therapy of α-Sarcoglycan Null Dystrophic Mice Through Intra-Arterial Delivery of Mesoangioblasts. Science. 2003;301(5632):487-92.
56. Iyer PS, Mavoungou LO, Ronzoni F, Zemla J, Schmid-Siegert E, Antonini S, et al. Autologous Cell Therapy Approach for Duchenne Muscular Dystrophy using PiggyBac Transposons and Mesoangioblasts. Molecular Therapy. 2018;26(4):1093-108.
57. Cossu G, Previtali SC, Napolitano S, Cicalese MP, Tedesco FS, Nicastro F, et al. Intra-arterial transplantation of HLA-matched donor mesoangioblasts in Duchenne muscular dystrophy. EMBO Mol Med. 2015;7(12):1513-28.

Chapter 2

SLIRP variants as a novel cause of mitochondrial encephalomyopathy with complex I and complex IV deficiency

Le Guo, Bob P.H. Engelen, Irenaeus F.M. de Coo, Maaike Vreeburg, Suzanne C.E.H. Sallevelt, Debby M.E.I. Hellebrekers, Edwin H. Jacobs, Farah Sadeghi-Niaraki, Florence H.J. van Tienen, Hubert J.M. Smeets*, Mike Gerards*

* Both senior author

Submitted: Major revision European Journal of Human Genetics

Abstract

In a Dutch boy of non-consanguineous parents, with a clinical diagnosis of mitochondrial encephalomyopathy and complex I and IV deficiency, whole exome sequencing (WES) revealed compound heterozygosity for two variants in the *SLIRP* gene, being a deletion in exon 3 and a splicing variant in intron 1. RT-PCR analysis revealed three *SLIRP* transcripts, adding up in total to only 63% of normal transcript levels in controls. The largest transcript, present in 18%, showed retention of a 106bp fragment of intron 1, leading to a frameshift and premature stop codon. The two smaller fragments contained either the deletion (74%) or wild-type *SLIRP* (8%). SRA stem-loop-interacting RNA-binding protein (SLIRP) is predominantly localized in mitochondria and forms a complex with LRPPRC, which is involved in regulation of mitochondrial DNA (mtDNA) gene expression. Indeed, in fibroblasts of the patient a reduced RNA level of the mtDNA genes *MT-ND1*, *MT-ND6* and *MT-CO1* was observed. Lentiviral transduction of wild-type *SLIRP* cDNA in the patient fibroblasts increased the expression of the *MT-ND1*, *MT-ND6*, *MT-CO1* genes (2.5 to 7.2-fold), whereas transduction with the mutants did not. The decreased complex I and IV enzyme activities seemed to be the result of a reduced mitochondrial mass, reflected by a fourfold decrease of citrate synthase (CS) compared to total protein in patient fibroblasts. Transduction with wild-type *SLIRP* cDNA led to a 2.4-fold increase of CS compared to total protein and an increase in complex I and IV activities. The results of the complementation assay confirmed the causality of the *SLIRP* variants and revealed the mechanism, leading to the mitochondrial encephalomyopathy in this patient. In conclusion, we report the first variants in *SLIRP* gene as a novel cause of mitochondrial encephalomyopathy with combined complex I and IV deficiency.

Introduction

Mitochondrial encephalomyopathy (ME) is a genetically and clinically heterogeneous group of neuromuscular disorders, characterized by oxidative phosphorylation (OXPHOS) defects resulting in decreased ATP production (1). The usually multi-systemic clinical manifestation of ME can be highly variable, ranging from chronic progressive external ophthalmoplegia, generalized myopathy, to central nervous system involvement and endocrine dysfunction (2). Highly dependent on oxidative metabolism, brain and skeletal muscle are often affected. ME is caused by variants in either the mtDNA or in a nuclear gene involved in OXPHOS (3). However, the variants and genes, identified so far, only explain part of the cases. Whole exome sequencing (WES) provides the power to identify the genetic defects of ME in undiagnosed patients. Recent examples of novel ME genes, resolved by WES, are *SLC25A42* (encoding an inner mitochondrial membrane protein that imports Coenzyme A into the mitochondrial matrix) (4), *VARS2* (encoding the mitochondrial valyl tRNA-synthetase that engages in mitochondrial protein synthesis) (5), and *FBXL4* (encoding F-box and leucine-rich repeat 4 protein that controls mtDNA homeostasis and maintenance) (6). These examples illustrate the heterogeneity in genetic causes and affected pathways in ME, which pose a challenge to an accurate molecular diagnosis of ME.

In a Dutch boy with an ME, presenting as congenital hypotonia, diffuse cerebral atrophy and hypomyelination and a complex I and IV deficiency, WES revealed compound heterozygosity for two candidate variants in the *SLIRP* gene. The pathogenicity was confirmed by a complementation assay. We showed for the first time that pathogenic variants in the *SLIRP* gene can cause ME, by which we broadened the genetic heterogeneity and further clarified underlying mechanisms of ME.

Materials and methods

Patient

The patient was the second child of non-consanguineous Dutch parents, born at 41 5/7 weeks after an uncomplicated pregnancy (birth weight 4.40 kg, >97th percentile) with Apgar scores being 1/5/10. Delayed progression of the delivery necessitated a cesarean section. The boy was floppy from birth onwards and showed bouts of opisthotonic posturing. He slowly developed some motor skills, with at best reaching with his fingers to his mouth, but lost these abilities gradually around 5 months of age. After 6 months of age he developed an inspiratory stridor, a nystagmus was noted and he was difficult to console. The arms and legs were weak and not able to move against gravity (MRC 2). Reflexes could not be elicited. Tremors were seen in the muscles from head and shoulder-girdle. Because of this severe psychomotor developmental delay, extensive clinical and laboratory investigations were undertaken.

Brain MRI at 9 months of age revealed delayed white matter maturation in particular in the frontal and parietal regions and symmetrical in the center semi-ovale, though some maturing was seen in the occipital region. Apart from mild diffuse cerebral atrophy, no structural abnormalities were observed. Spectroscopy at 14 months of age did not show signs of cell loss nor a lactic acid peak in the white matter and basal ganglia voxel analyzed. Electroneurography showed a lack of sensory nerve action potential and low compound motor action potential, which demonstrated an axonal or a mixed neuropathy. The visual evoked response via both eyes showed severe delay, which fits the defective myelination of the visual tract. Light and electron microscopic analysis found no signs of hypomyelination on a sural nerve biopsy, and no abnormalities in conjunctival and rectal biopsies. The boy further deteriorated and died at 18 months of age from general muscle weakness with respiratory insufficiency and pneumonia.

Morphology studies on biopsy material from quadriceps muscle showed a denervation pattern. No ragged red, COX negative fibres or signs of increased histochemical staining of succinate dehydrogenase were found. Biochemical measurements of the OXPHOS complexes in quadriceps muscle biopsy and in cultured skin fibroblasts showed a combined isolated Complex I (muscle 0.06 (29%) and fibroblast 0.09 (45%) normalized to CS; muscle 5.53 and fibroblast 2.41 µmol NADH/min normalized to wet protein) and Complex IV (muscle 0.57 (48%) and fibroblast 0.31 (44%) normalized to CS; muscle 50.26 and fibroblast 8.62 µmol

NADH/min normalized to wet protein) activity deficiency. This indicated a mitochondrial defect, with a remark that the quality of the muscle was considered poor and the enzyme measurement could not be repeated due to the death of the patient. Metabolic screening of blood and cerebrospinal fluid (CSF) showed a slight increase of lactate in blood (3.1 mmol/l; normal < 2.3 mmol/l) and in CSF (3.8 mmol/l; normal < 2.8 mmol/l). Amino acid, organic acid, purines and pyrimidine metabolism and sialotransferrine patterns were normal. Genetic analysis did not show chromosomal aberrations. Sequencing of the mtDNA (7) and a panel of 412 nuclear mitochondrial genes (8) did not reveal the genetic cause.

Whole Exome Sequencing

In order to identify the genetic cause in the patient, WES was performed using 1 µg of DNA for fragmentation and exons, including untranslated regions, were captured with the Agilent SureSelect Human All Exon v4 plus UTR's (Agilent Technologies, USA) kit for exome enrichment. Sequencing was performed on an Illumina HiSeq 2000 (Illumina), followed by an in-house annotation Python/R script that matched variants to the RefGene (refGene_131114), GenCode v19 (genCode_v19_030215), and dbSNP144 hg19 tracks from the UCSC genome browser as previously described (8). Exome data of the patient were filtered for non-synonymous homozygous or compound heterozygous variants due to an autosomal recessive inheritance, with allele frequency below 0.01 in ExAC, dbSNP144 and gnomAD, conservation by PhyloP (>1.5), and protein damage by SIFT (<0.05) and Polyphen2 (>0.85). Possible splice site variants were analyzed with SpliceSiteFinder-like and NNSPLICE. Candidate variants were validated by Sanger sequencing using an ABI 3730 sequencer (Applied Biosystems) after PCR amplification with primers listed in Supplementary Table S1. The segregation of variants was determined by Sanger sequencing the DNA from both parents. SLIRP transcript NM_031210.5/ENST00000557342.6 was used for variant nomenclature and exon numbering. Identified variants were submitted to the Leiden Open Variation Database (individual: 00320233, DB-ID: SLIRP-000001 and SLIRP-000002, https://databases.lovd.nl/shared/individuals/00320233).

RNA extraction, cDNA synthesis and qRT-PCR

Dermal fibroblasts derived from the index patient and healthy controls were cultured in complete medium containing Dulbecco's modified Eagle Medium, high glucose (DMEM, all reagents were from Thermo Fisher Scientific unless stated)

supplemented with 10% fetal bovine serum (FBS, Bodinco) and 1% penicillin-streptomycin in a 37°C, 5% CO_2 humid incubator. Total RNA was isolated from consecutive-passage cultured fibroblasts with TRIzol (Invitrogen) and converted into cDNA using qScript cDNA Synthesis Kit (Quantabio) according to the manufacturer's protocol. cDNA was diluted 25-fold and used in downstream qPCR reactions. qRT-PCR analysis for the candidate *SLIRP* gene, *LRPPRC* gene and *MT-CYB*, *MT-ND1*, *MT-ND6* and *MT-CO1* was performed using a LightCycler 480 (Roche) with SensiMix SYBR Hi-ROX kit (Bioline). The delta-delta Ct method was used for relative quantification and the housekeeping gene *hTBP* was used for normalization. Reactions were run in biological and technical triplicate. Primers are listed in Supplementary Table S2. Statistical significance was calculated using a two-sample t-test, one-tailed, equal variance. A 95% confidence interval and alpha-level 0.05 was applied.

TA cloning

In order to identify the effect of the splicing variant on *SLIRP* mRNA, PCR products of the patient cDNA were amplified with primers listed in Supplementary Table S1, and cloned into pCR2.1 vector according to the protocol provided by the manufacturer. Positive clones were picked from LB agar plate containing 100 µg/ml ampicillin and validated by colony PCR and Sanger sequencing with specific primers (same as for amplifying cDNA).

Complementation assay

Feline immunodeficiency virus-based lentivirus transduction was performed according to the protocol of the manufacturer (GeneCopoeia). *SLIRP* cDNA sequences were cloned into a lentiviral expression vector (pReceiver-Lv21). The correct insertion of wild-type, insertion-type and deletion-type *SLIRP* cDNA was validated by Sanger sequencing. For feline immunodeficiency virus production, HEK293T cells were plated in DMEM supplemented with 10% heat-inactivated fetal bovine serum. When the cells reached 70% confluence, the medium was replaced with DMEM supplemented with 5% fetal bovine serum and the HEK293T cells were co-transfected using Endofectin Lenti reagent (GeneCopoeia) with the packaging plasmids and each type of the *SLIRP* cDNA expression vectors in a 1:1 ratio. The supernatants containing each type of viral particles were collected 48 hours post

transfection, centrifuged at 500 x g for 5 min and filtered through a 0.45 µm filter (Corning). Patient fibroblasts (70% confluence) were infected by replacing the normal culture medium with medium containing lentiviral particles and 8 µg/ml polybrene (Sigma-Aldrich). The medium containing lentiviral particles was replaced by fresh complete medium the next day. Stably transduced fibroblasts were selected with 200 µg/ml G418 (Sigma-Aldrich) for 2 weeks, starting 48 hours post transduction, followed by culturing in 5 mM D-(+)-galactose (Sigma-Aldrich) supplemented DMEM (no glucose) medium for 48 hours. qRT-PCR analysis for the candidate *SLIRP* gene and *MT-CYB*, *MT-ND1*, *MT-ND6* and *MT-CO1* was performed as described above using a LightCycler 480 (Roche) with SensiMix SYBR Hi-ROX kit (Bioline). Complex I, IV and citrate synthase activities and protein levels were measured following a clinical diagnostic procedure as described before (9).

Results

WES revealed two probably pathogenic variants in the SLIRP gene

After filtering the WES data by allele frequency, conservation and predicted pathogenicity, three candidate genes remained with two possibly disease-causing variants: *TMPRSS9* (c.1253C>T, c.1519G>A, NM_182973), *NPC1* (homozygous c.2131-5_2131-4del, NM_000271) and *USP9Y* (homozygous c.3826A>G, NM_004654). Of these, *TMPRSS9* encodes a membrane-bound type II serine polyprotease involved in cancer progression. *NPC1* is linked to Niemann-Pick disease, type C1, which is mainly caused by cholesterol trafficking defects. *USP9Y* is associated with male infertility. As the patient suffered from a ME with a deficiency of OXPHOS complex I and IV activities, the variants in these genes could not explain the phenotype. Thus, we re-analyzed the data and included genes with one possible disease-causing variant but with a clear relation to OXPHOS. This led to the identification of a heterozygous deletion variant causing a frameshift in exon 3 (NM_031210.5:c.248_252del; NP_112487.1:p.(Ile83Argfs*10)). Further evaluation of the *SLIRP* sequences revealed another heterozygous *SLIRP* variant (NC_000014.8:g.78177003A>G; NM_031210.5:c.98-178A>G), deep in intron 1, which was predicted to cause alternative splicing by introducing a novel splice donor site (SpliceSiteFinder-like: 79.6%, NNSPICE: 70.0%). Both *SLIRP* variants were confirmed in the patient by Sanger sequencing and both parents were carrier of one of the variants (Figure 1).

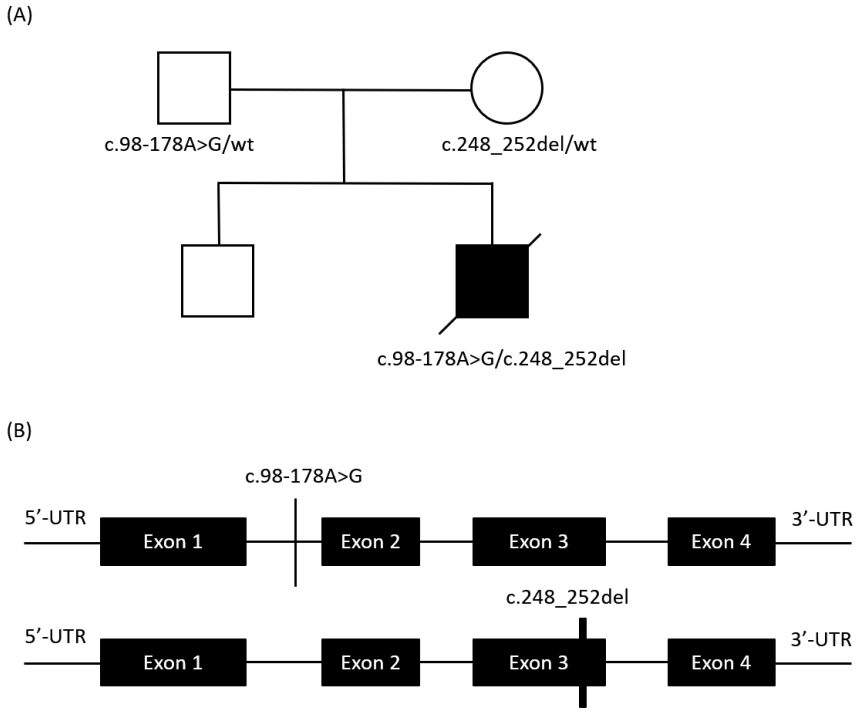
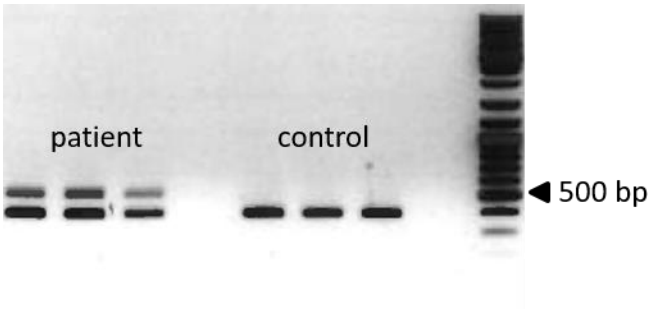


Figure 1: Schematic view of *SLIRP* gene and positions of variants. (A) Pedigree of the index family. The index patient was diagnosed with mitochondrial encephalomyopathy as depicted by the filled black symbol. (B) Schematic diagram showing the position of two *SLIRP* variants. Exons and introns are not drawn to scale. UTR = untranslated region.

Effect of SLIRP variants at the RNA level

RT-PCR was performed using primers at the start and end of the *SLIRP* cDNA, to investigate the effect of the predicted alternative splicing due to the c.98-178A>G variant. In the control cDNA, only one PCR product of the expected size was observed, whereas two products were detected in the patient cDNA (Figure 2A). TA cloning was performed on the PCR products to investigate the sequence of each product separately. Colony PCR results revealed three different fragments. The largest fragment showed retention of a 106 bp fragment of intron 1 at the boundary of exon 1 and exon 2 in the cDNA, leading to a frameshift and premature stop codon (p.(Ser33Argfs*9), Figure 2B), resulting from the c.98-178A>G variant. The two smaller fragments consisted of *SLIRP* cDNA with the c.245_249del deletion or the wild-type *SLIRP* cDNA sequence.

(A)



(B)

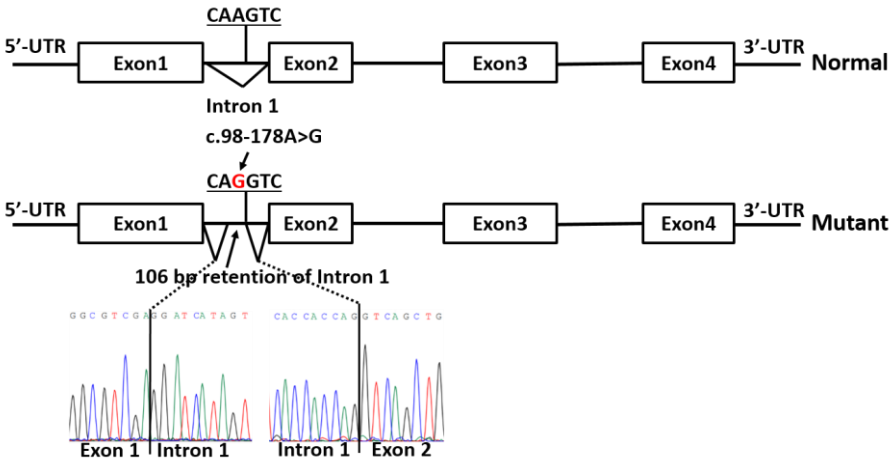


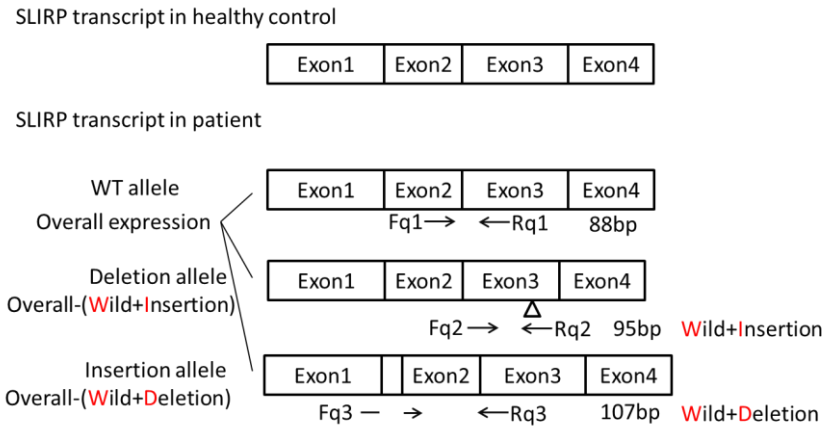
Figure 2: The effects of the c.98-178A>G variant in *SLIRP* intron 1 on RNA splicing. (A) PCR product of entire *SLIRP* cDNA. The control showed the expected wild-type size, whereas the patient had an additionally larger transcript product (n=3). (B) Sequence analysis after cloning patient *SLIRP* cDNA revealed a 106 bp retention from intron 1 due to the c.98-178A>G variant in *SLIRP* gene. Exons and introns are not drawn to scale.

Reduced expression wild-type SLIRP gene in fibroblasts of the patient

To quantify the abundance of each *SLIRP* transcript, qRT-PCR was performed using different primer pairs designed to capture overall *SLIRP* expression and each transcript type (wild-type, deletion-type and insertion-type) separately in the patient (Figure 3A). A forward primer (Fq1) located in exon 2 and a reverse primer (Rq1) in exon 3 were used to measure overall *SLIRP* expression, which was decreased to 63% in the patient compared to controls (Figure 3B). A forward primer (Fq2) located

in exon 3 and a reverse primer (Rq2) encompassing the 5 bp deletion in exon 3 were used to measure wild-type and insertion-type (106 bp intron 1 retention) *SLIRP* mRNA, which was 16% compared to control *SLIRP* expression (Figure 3B). A forward primer (Fq3) spanning the exon 1-exon 2 boundary and a reverse primer (Rq3) in front of the deletion in exon 3 were used to determine the level of wild-type and deletion-type transcripts, which was 52% compared to total *SLIRP* expression in controls (Figure 3B). From these data we calculated that of all *SLIRP* transcripts in the patient, 74% is deletion-type, 18% is insertion type and only a small minority (8%) is wild-type. As the *SLIRP* expression is overall decreased to 63%, this means that only 5% of wild-type *SLIRP* RNA is present in the patient fibroblasts compared to the control.

(A)



(B)

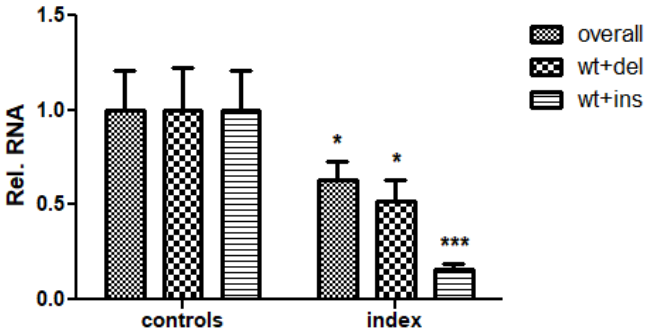


Figure 3: *SLIRP* mRNA expression levels in patient and control fibroblasts. (A) Locations of specific primer pairs for capturing overall (Fq1, Rq1), wild-type and insertion-type (Fq2, Rq2), wild-type and deletion-type (Fq3, Rq3) and wild-type (Fq3, Rq2) *SLIRP* transcripts. The deletion variant position is denoted as a hollow triangle in the deletion allele and the splicing variant-induced retention of intron 1 is labelled in the insertion allele. (B) Total *SLIRP* mRNA expression (denoted as overall) was reduced to 63% (*, p=2.8E-3) in patient fibroblasts compared to control fibroblasts. Wild-type plus deletion-type (denoted as wt+ins) was 52% (*, p=8.0E-4) compared to controls, while wild-type plus insertion-type (denoted as wt+del) was 16% (***, p=5.7E-6) compared to control *SLIRP* expression. The patient fibroblasts only had 5% wild-type transcript. *hTBP* was used as endogenous control. Data are represented as means ± SD (n=4). Error bars indicate SD. Student's *t* test was used for statistics. *p < 0.05, **p < 0.01, ***p < 0.001.

Wild-type SLIRP transduction rescues mtDNA gene expression and OXPHOS enzyme activity

Recent studies have shown that SLIRP facilitates the correct association of mt-mRNAs with the mitoribosome, thus playing an important role in mt-mRNA homeostasis (10). To evaluate the effect of the two *SLIRP* variants on the levels of mt-mRNA transcripts, qRT-PCR was performed in fibroblasts of the patient. We found a general decrease of mt-mRNA transcripts, including *MT-CO1*, *MT-ND1* and *MT-CYB* (Figure 4A, Supplementary Figure S1). To investigate whether reduced mRNA level of SLIRP has an impact on the mRNA levels of LRPPRC, qRT-PCR was performed. The level of LRPPRC mRNA transcripts was decreased to 76% in the patient fibroblasts, although this was not statistically significant (Figure 4A).

To clarify the pathogenicity of the two *SLIRP* variants, a complementation assay was performed. Fibroblasts from the patient were stably transduced with lentiviral clones containing wild-type *SLIRP*, *SLIRP* containing c.245_249del variant, and *SLIRP* containing the 106 bp intron 1 retention due to c.98-178A>G variant. At the mRNA level, *MT-CO1* (encoding a complex IV subunit) mRNA expression in wild-type transduced patient fibroblasts showed a 7.2-fold increase (***, p<0.001) compared to non-transduced patient fibroblasts (Figure 4B). Neither the deletion-type *SLIRP* nor insertion-type *SLIRP* rescued *MT-CO1* expression (Figure 4B), which demonstrated the rescue of *MT-CO1* mRNA expression by the wild-type, but not by the mutants. *MT-ND1* and *MT-ND6* (encoding complex I subunits) mRNA expression also showed a similar pattern upon transduction of the different *SLIRP* types (Figure 4B), being a 2.5- and 4.2-fold increase in wild-type transduced patient fibroblasts, respectively. These results indicate that both insertion and deletion type *SLIRP* transcripts resulted in overall decrease of mtDNA encoded gene expression.

The enzyme activity of CS and OXPHOS complex I and IV was measured in the patient fibroblasts before and after complementation. After transfection with wild-type *SLIRP* cDNA, the CS versus protein ratio increased to 64% of control values. Related to the protein level, complex I activity in patient fibroblasts was fully restored by wild-type *SLIRP*, and complex IV partly (Figure 4C). Taken together, our results showed that the wild-type *SLIRP* cDNA transduction in the patient fibroblasts can rescue mt-mRNA transcripts expression, mitochondrial mass as well as OXPHOS enzyme activity, thus providing the evidence of the pathogenicity of the two *SLIRP* variants.

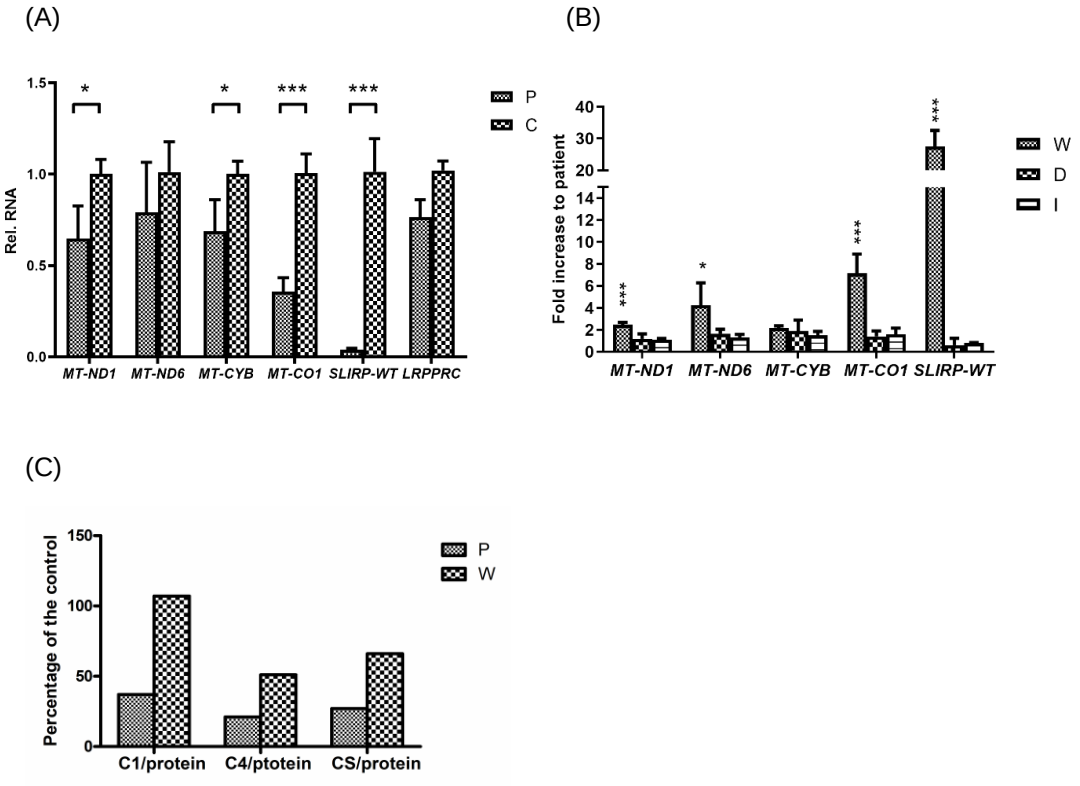


Figure 4: Complementation assay of SLIRP in patient fibroblasts. (A) Mitochondrial transcript steady-state levels assessed by qPT-PCR in patient fibroblasts (denoted as P) and controls (denoted as C). *hTBP* was used as endogenous control. Data are represented as means ± SD (n=3). Error bars indicate SD. Student's *t* test was used for statistics. *p < 0.05, **p < 0.01, ***p < 0.001. (B) Transduction of patient fibroblasts with wild-type *SLIRP* recovers mitochondrial gene expression in galactose medium culture. Patient fibroblasts were stably transduced with lentiviral clones containing wild-type *SLIRP* (denoted as W), *SLIRP* containing c.245_249del (denoted as D) and *SLIRP* containing 106 bp intron 1 retention (denoted as I). *hTBP* was used as endogenous control. Relative RNA expressions of W, D and I were normalized to P. (C) Transduction of patient fibroblasts with wild-type *SLIRP* restores complex I (C1), IV activity (C4) and citrate synthase (CS), when normalized to the total protein level.

Discussion

In this study, two *SLIRP* c.248_252del and c.98-178A>G variants were identified by WES in a non-consanguineous Dutch patient as a novel cause of ME. The *SLIRP* gene was missed in the first analysis due to the presence of only one heterozygous deletion variant in the exonic and flanking intronic regions, whereas the mode of inheritance was autosomal recessive. After filtering the exome data, none of the genes with two variants could explain the ME. Therefore, the WES data were reanalyzed, selecting for genes with one possible disease-causing variant but a clear relation to OXPHOS. *SLIRP* was such a candidate. Initially, the gene was described as an RNA-binding protein that interacts with the STR7 substructure of steroid receptor RNA activator (SRA) and inhibits nuclear receptor transactivation (11). SLIRP harbors two ubiquitination sites (Lys36, Lys88) within an RNA recognition motif (RRM) domain (Ala21-Val91), which is essential for correct association of mt-mRNAs with the mitoribosome (10, 12-15). Reanalysis of the WES data revealed a deep-intronic variant in *SLIRP* intron 1 with a possible effect on splicing. Relative quantification of *SLIRP* transcripts level showed that total *SLIRP* mRNA expression was reduced to 63% in patient fibroblasts compared to control fibroblasts. Subsequent mutant-type and wild-type *SLIRP* transcripts analysis showed the majority of *SLIRP* transcripts in the patient was deletion-type (74%), followed by insertion-type (18%). The reduced level of total *SLIRP* transcripts and the low level of insertion-type *SLIRP* transcripts suggested nonsense-mediated mRNA decay due to the c.98-178A>G variant in intron 1, as the splicing variant created a premature stop codon early in the *SLIRP* transcripts. Only a small fraction of *SLIRP* transcripts was wild-type (8%), which, after corrected for the overall decrease in expression to 63%, indicated that only 5% wild-type SLIRP was present compared to the control. The other two transcripts would not yield a functional protein.

In human fibroblasts, SLIRP was demonstrated to be an OXPHOS regulator to maintain the steady-state levels of mitochondrial transcripts involved in OXPHOS (16). In *SLIRP* knockout mice and MEFs, SLIRP was also shown to be involved in presenting mature mRNAs to the mitoribosome to fine-tune the rate of mitochondrial protein synthesis (10). This paper reported in the KO mice a drastic decrease in the steady-state levels of mt-mRNA transcripts, consistent with our findings of the decrease of the steady-state level of *MT-ND1*, *MT-ND6* and *MT-CO1* RNA in fibroblasts of the patient. The steady-state level of *MT-ND6* RNA from the patient fibroblasts only slightly decreased to 79%. This may be due to the relatively small size and short half-life of the *MT-ND6* transcripts compared to the rest of the tested

mt-mRNA transcripts, making the *MT-ND6* transcripts less affected by the *SLIRP* variants (17, 18). SLIRP physically interacts with LRPPRC and protects the latter from degradation by mitochondrial matrix proteases (10). SLIRP and LRPPRC form a ribonucleoprotein complex that promotes the post-transcriptional expression of mt-mRNA transcripts and both proteins are co-stabilized within the complex (16, 17). In our patient, no variants were discovered in *LRPPRC*, whereas the levels of *LRPPRC* mRNA transcripts decreased to 76% compared to control fibroblasts (though not statistically significant). This could be explained by the fact that the predominant deletion variant p.(Ile83Argfs*10) is located downstream to the key amino acids (His59, Arg60, Glu80, Asn81 and His82) for LRPPRC-SLIRP interaction, thus causing less disruption to the LRPPRC-SLIRP binding interface (19). Besides, low levels of wild-type *SLIRP* mRNA transcripts (8%) were present in fibroblasts of the patient, consistent with the role of SLIRP in maintaining normal levels of *LRPPRC* mRNA transcripts (10).

A complementation assay can provide definite evidence for variants being pathogenic or benign (20, 21). In our patient, a decrease in mtDNA-encoded transcripts, a decreased CS/protein ratio and a combined complex I and IV deficiency were detected in muscle and fibroblasts. The data suggested that the OXPHOS deficiency resulted from a decreased number of mitochondria, rather than a specific complex defect. Results of the complementation assay showed that the levels of *MT-CO1*, *MT-ND1* and *MT-ND6* were restored in patient fibroblasts after wild-type *SLIRP* transduction, whereas this did not happen when transduced with the deletion-type *SLIRP* or insertion-type *SLIRP*. Transduction with the wild-type SLIRP also showed an increase in mitochondrial mass per cell (increased CS/protein ratio) and an increase in complex I and IV activity, when normalized for the amount of protein. Based on these results, we concluded that the two *SLIRP* variants caused the complex I and IV deficiency in the fibroblasts of the patient by a decrease in mitochondrial mass. Considering the frequency, conservation and complementation study results, the two *SLIRP* variants are classified as likely pathogenic (class 4) according to ACMG guidelines (22).

WES is a powerful tool to identify pathogenic variants in mitochondrial diseases (8). Still, the method is not perfect and due to technical issues or the location of variants outside the captured regions, pathogenic variants will be missed. A recent study reported an additional 10% (5 of 48) increase of diagnostic rate by RNA sequencing in molecularly undiagnosed mitochondriopathy patients following a regular WES procedure (23). The study not only provided an example of intronic loss-of-function

variants with pathological relevance, it also showed the necessity to re-prioritized heterozygous rare variants. Our study confirmed the importance to reanalyze non-exonic variants. Imbalanced allele expression of *SLIRP* was discovered by analyzing RNA from fibroblasts of patient, indicating that RNA sequencing would have uncovered SLIRP as a candidate for ME in this family. RNA sequencing covers intronic and regulatory regions, allelic imbalances and splicing variants, under the condition that the genes are expressed in the cells or tissue studied. It is therefore an important complementary tool to identify rare, disease-causing variants in mitochondrial diseases.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

We would kindly acknowledge the family participating in this study. LG was supported by funding from China Scholarship Council and IdC by NeMO foundation, Grant/Award Number: 19_P02. Part of this work has been made possible with the support of the Dutch Province of Limburg (MG).

References

1. DiMauro S, Schon EA, Carelli V, Hirano M. The clinical maze of mitochondrial neurology. Nat Rev Neurol. 2013;9(8):429-44.
2. Lindner A, Hofmann E, Naumann M, Becker G, Reichmann H. Clinical, morphological, biochemical, and neuroradiological features of mitochondrial encephalomyopathies. Presentation of 19 patients. Molecular and cellular biochemistry. 1997;174(1-2):297-303.
3. DiMauro S. Mitochondrial encephalomyopathies—Fifty years on. The Robert Wartenberg Lecture. 2013;81(3):281-91.
4. Almannai M, Alasmari A, Alqasmi A, Faqeih E, Al Mutairi F, Alotaibi M, et al. Expanding the phenotype of SLC25A42-associated mitochondrial encephalomyopathy. Clin Genet. 2018;93(5):1097-102.
5. Bruni F, Di Meo I, Bellacchio E, Webb BD, McFarland R, Chrzanowska-Lightowlers ZMA, et al. Clinical, biochemical, and genetic features associated with VARS2-related mitochondrial disease. Human mutation. 2018;39(4):563-78.
6. Gai X, Ghezzi D, Johnson MA, Biagosch CA, Shamseldin HE, Haack TB, et al. Mutations in FBXL4, encoding a mitochondrial protein, cause early-onset mitochondrial encephalomyopathy. Am J Hum Genet. 2013;93(3):482-95.
7. Spruijt L, Smeets HJ, Hendrickx A, Bettink-Remeijer MW, Maat-Kievit A, Schoonderwoerd KC, et al. A MELAS-Associated ND1 Mutation Causing Leber Hereditary Optic Neuropathy and Spastic Dystonia. Archives of Neurology. 2007;64(6):890-3.
8. Theunissen TEJ, Nguyen M, Kamps R, Hendrickx AT, Sallevelt S, Gottschalk RWH, et al. Whole Exome Sequencing Is the Preferred Strategy to Identify the Genetic Defect in Patients With a Probable or Possible Mitochondrial Cause. Front Genet. 2018;9:400.
9. Sgobbo P, Pacelli C, Grattagliano I, Villani G, Cocco T. Carvedilol inhibits mitochondrial complex I and induces resistance to H2O2 -mediated oxidative insult in H9C2 myocardial cells. Biochim Biophys Acta. 2007;1767(3):222-32.
10. Lagouge M, Mourier A, Lee HJ, Spahr H, Wai T, Kukat C, et al. SLIRP Regulates the Rate of Mitochondrial Protein Synthesis and Protects LRPPRC from Degradation. PLoS Genet. 2015;11(8):e1005423.
11. Hatchell EC, Colley SM, Beveridge DJ, Epis MR, Stuart LM, Giles KM, et al. SLIRP, a small SRA binding protein, is a nuclear receptor corepressor. Mol Cell. 2006;22(5):657-68.
12. Matsumoto M, Hatakeyama S, Oyamada K, Oda Y, Nishimura T, Nakayama KI. Large-scale analysis of the human ubiquitin-related proteome. Proteomics. 2005;5(16):4145-51.
13. Wagner SA, Beli P, Weinert BT, Nielsen ML, Cox J, Mann M, et al. A proteome-wide, quantitative survey of in vivo ubiquitylation sites reveals widespread regulatory roles.

Molecular & Cellular Proteomics. 2011;10(10).

14. Beltrao P, Albanèse V, Kenner LR, Swaney DL, Burlingame A, Villén J, et al. Systematic functional prioritization of protein posttranslational modifications. Cell. 2012;150(2):413-25.
15. Stes E, Laga M, Walton A, Samyn N, Timmerman E, De Smet I, et al. A COFRADIC protocol to study protein ubiquitination. Journal of proteome research. 2014;13(6):3107-13.
16. Baughman JM, Nilsson R, Gohil VM, Arlow DH, Gauhar Z, Mootha VK. A computational screen for regulators of oxidative phosphorylation implicates SLIRP in mitochondrial RNA homeostasis. PLoS Genet. 2009;5(8):e1000590.
17. Chujo T, Ohira T, Sakaguchi Y, Goshima N, Nomura N, Nagao A, et al. LRPPRC/SLIRP suppresses PNPase-mediated mRNA decay and promotes polyadenylation in human mitochondria. Nucleic Acids Res. 2012;40(16):8033-47.
18. Bruni F, Proctor-Kent Y, Lightowlers RN, Chrzanowska-Lightowlers ZM. Messenger RNA delivery to mitoribosomes - hints from a bacterial toxin. FEBS J. 2020.
19. Spahr H, Rozanska A, Li X, Atanassov I, Lightowlers RN, Chrzanowska-Lightowlers ZM, et al. SLIRP stabilizes LRPPRC via an RRM-PPR protein interface. Nucleic Acids Res. 2016;44(14):6868-82.
20. Starita LM, Ahituv N, Dunham MJ, Kitzman JO, Roth FP, Seelig G, et al. Variant Interpretation: Functional Assays to the Rescue. Am J Hum Genet. 2017;101(3):315-25.
21. MacArthur DG, Manolio TA, Dimmock DP, Rehm HL, Shendure J, Abecasis GR, et al. Guidelines for investigating causality of sequence variants in human disease. Nature. 2014;508(7497):469-76.
22. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17:405.
23. Kremer LS, Bader DM, Mertes C, Kopajtich R, Pichler G, Iuso A, et al. Genetic diagnosis of Mendelian disorders via RNA sequencing. Nature Communications. 2017;8(1):15824.

Supplementary material

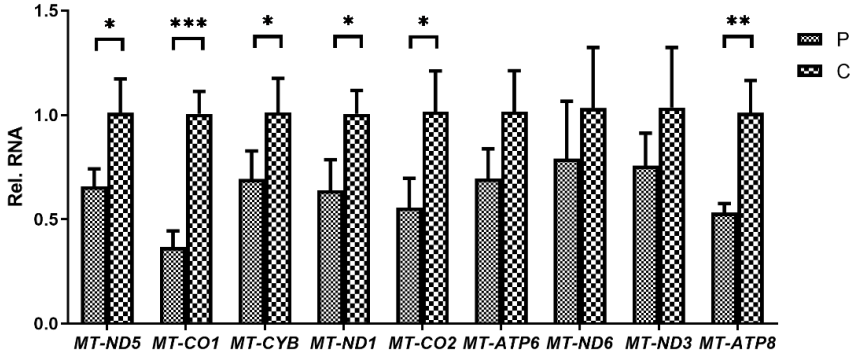
Supplementary Table S1. Primers for Sanger sequencing of the candidate genes

Primer name	Sequence (5'-3')	Candidate variants
NPC1_F	AGTGCACGTGAGAGTGATGT	NPC1 (c.2131-5_2131-4del, NM_000271)
NPC1_R	GGGACATGTTCAGGTAGCCA	
USP9Y_F	CCAAACTGCAGTGCCAAAGT	USP9Y (c.3826A>G, NM_004654)
USP9Y_R	TCTGCTTGGAACGGTAAGGT	
TMPRSS9_F1	CTTGTGTGGAGAGAGGTCCT	TMPRSS9 (c.1253C>T, NM_182973)
TMPRSS9_R1	CTTAGGAACGGTGACCCAGT	
TMPRSS9_F2	GGCAAACTACGGACACCTTG	TMPRSS9 (c.1519G>A, NM_182973)
TMPRSS9_R2	GGTTAGGCCCTGTTTGTCCT	

Supplementary Table S2. Primers for qRT-PCR

Primer name	Sequence (5'-3')	Purpose
Fq1	AACACTTTGCACAGTTCGGC	Overall *SLIRP* transcripts
Rq1	TGAACCCAACCCAAACCTCT	
Fq2	AGAGGTTTGGGTTGGGTTC	Wild-type + insertion-type
Rq2	CTGGACCTTTACTCCATCTATAATAT	*SLIRP* transcripts
Fq3	GCGTCGAGTCAGCTGAAAG	Wild-type + deletion-type
Rq3	TGAACCCAACCCAAACCTCT	*SLIRP* transcripts
ND5_F	GCCTTCTCCACTTCAAGTCAA	*MT-ND5* transcripts expression
ND5_R	TGGGTACAGATGTGCAGGAAT	
CO1_F	GGCCTGACTGGCATTGTATT	*MT-CO1* transcripts expression
CO1_R	TGGCGTAGGTTTGGTCTAGG	
CYB_F	TATCCGCCATCCCATACATT	*MT-CYB* transcripts expression
CYB_R	GGTGATTCCTAGGGGGTTGT	
ND1_F	CCCTAAAACCCGCCACATCT	*MT-ND1* transcripts expression
ND1_R	AGCGATGGTGAGAGCTAAGGTC	
CO2_F	CGACTACGGCGGACTAATCT	*MT-CO2* transcripts expression
CO2_R	CGGGAATTGCATCTGTTTTT	
ATP6_F	ACCCGCCGCAGTACTGATCAT	*MT-ATP6* transcripts expression
ATP6_R	AGGAGGTTAGTTGTGGCAAT	
ND6_F	TGATTGTTAGCGGTGTGGTC	*MT-ND6* transcripts expression
ND6_R	CCACAGCACCAATCCTACCT	
ND3_F	CGGCTTCGACCCTATATCCC	*MT-ND3* transcripts expression
ND3_R	GTAGGGCTCATGGTAGGGGT	
ATP8_F	GTATGGCCCACCATAATTACCC	*MT-ATP8* transcripts expression
ATP8_R	GGCAATGAATGAAGCGAACAG	
TBP_F	CACGAACCACGGCACTGATT	*hTBP* transcripts expression
TBP_R	TTTTCTTGCTGCCAGTCTGGAC	

Supplementary Figure S1. Mitochondrial mRNA expression levels in patient and control fibroblasts. Mitochondrial transcript steady-state levels assessed by qPT-PCR in patient fibroblasts (P) and controls (C). *hTBP* was used as endogenous control. Data are represented as means ± SD (n=4). Error bars indicate SD. Student's *t* test was used for statistics. *p < 0.05, **p < 0.01, ***p < 0.001.



Chapter 3

Whole exome sequencing reveals a homozygous *C1QBP* deletion as the cause of progressive external ophthalmoplegia and multiple mtDNA deletions

Le Guo, Periyasamy Govindaraj, Mariëlle Kievit, Irenaeus F.M. de Coo, Mike Gerards, Debby M.E.I. Hellebrekers, Fons P.M. Stassen, Narayanappa Gayathri, Arun B Taly, Bindu Parayil Sankaran*, Hubert J.M. Smeets*

* Both senior author

Submitted: Major revision Neuromuscular Disorders

Abstract

Whole exome sequencing (WES) is a powerful tool to identify novel mutations and genes in genetically heterogeneous mitochondrial diseases. Here we combined two bioinformatic tools, GENESIS and WeGET, for WES data analysis, and identified a homozygous deletion in *C1QBP* gene in a patient of Indian descent, presenting with progressive external ophthalmoplegia (PEO) and multiple deletions in mitochondrial DNA (mtDNA). *C1QBP* gene encodes complementary 1 Q subcomponent-binding protein (C1QBP; also known as p32), which is predominantly localized in the mitochondrial matrix and involved in mitochondrial homeostasis. Biallelic mutations in the *C1QBP* gene have been reported to cause mitochondrial cardiomyopathy and/or PEO, with variable age of onset in only a few patients yet. Compared to the previously reported phenotype, our patient showed only isolated late-onset PEO without overt cardiac involvement. Phenotype genotype correlation of these limited data suggests that the location of the mutation is important for the clinical manifestation, as amino acid changes associated with early-onset cardiomyopathy are localized in important structural domains whereas variants found in PEO only patients are localized on the coiled-coil region. This paper also highlights the need to consider *C1QBP* mutations in PEO patients, regardless of cardiac phenotype.

Introduction

Progressive external ophthalmoplegia (PEO; OMIM phenotypic series PS157640) is a group of disorders characterized by progressive ptosis and decrease in ocular motility, which may occur in isolation (isolated PEO) or accompanied with other mitochondrial related symptoms (PEO plus) (1). PEO exhibits considerable genetic heterogeneity. Approximately half of the PEO patients are sporadic cases, presenting with a single large mtDNA deletion (1, 2). The other half are either maternally inherited or has autosomal dominant (AD), or autosomal recessive (AR) inheritance. Maternally inherited PEO is caused by point mutations in mtDNA, leading to either a multi-systemic phenotype or isolated PEO. AD PEO is caused by mutations in at least seven nuclear genes, including *POLG* (MIM# 174763), *POLG2* (MIM# 604983), *SLC25A4* (MIM# 103220), *TWNK* (MIM# 606075), *RRM2B* (MIM# 604712), *DNA2* (MIM# 601810), and *OPA1* (MIM# 605290). In contrast to AD PEO, AR PEO is less common and results from mutations in nuclear genes such as *TYMP* (MIM# 131222), *DGUOK* (MIM# 601465), *TK2* (MIM# 188250) and *RNASEH1* (MIM# 604123). These genes encode proteins involved in mtDNA replication and maintenance, and defects in these proteins are associated with the development of multiple deletions in mtDNA (3).

Whole exome sequencing (WES) has increasingly become a standard approach to identify causative variants in diseases with high genetic heterogeneity like PEO. However, bioinformatics analysis to discover causative variants remains challenging due to the vast amount of sequencing data. Two bioinformatic tools, GENESIS and WeGET, provide a solution to this issue. GENESIS is a next-generation sequencing data-sharing platform integrated with various external variants filtering tools (mainly based on frequency, pathogenicity, conservation) (4). Until now, over 160 studies have used or cited GENESIS, enabling the discovery of 72 novel pathogenic genes and/or expanding their genotype/phenotype relationship (5). WeGET is a computational tool to identify mammalian genes that strongly co-express with a human query gene set of interest (6). The main principle of WeGET is by assessing a query gene set's co-expression within approximately 1000 multi-tissue datasets using N100 statistic (7) and robust rank aggregation (8), and a high area under the rank-recall curve (AUC) score (≥ 0.75) indicates that the majority of the query genes are among the top-ranking possible disease-causing genes.

Here, we combined GENESIS and WeGET for WES data analysis in a patient with isolated PEO and multiple deletions in mtDNA. Our data lead to the discovery of a

novel homozygous deletion in the *C1QBP* gene, recently reported as a cause of PEO and/or cardiomyopathy (9, 10).

Methods

Clinical findings

This female patient (aged 33) was born to healthy consanguineous parents with no family history of neurological or systemic disease (Figure 1A). She was married and had two healthy children. She reported mild intermittent headache for two years, drooping of eyelid for 1.5 years and difficulty in swallowing for six months at her first visit. At follow up after three and a half years she reported improvement in headache but she had nasal twang and there was a mild progressive proximal upper and lower muscle weakness (MRC scale:4/5).

Apart from a mild high arch palate, her physical and systemic examination was normal. Her visual acuity and fundi were normal. She had bilateral symmetrical mild ptosis and ophthalmoparesis. Facial, neck flexor, finger extensors and hip flexor showed muscle weakness, whereas bulbar and other limb muscles showed normal strength. The sensory system was normal. All stretch reflexes were sluggish while ankle jerks were normal. The plantar response was flexor. There was no appreciable fatigue on repeated efforts.

Laboratory tests showed elevated creatine kinase (1032 U/L, normal<145 U/L), FGF21 (3351.26 pg/ml, normal<331 pg/ml) and GDF15 (5891.39 pg/ml, normal <1014 pg/ml), while low B12 level (101.8 pg/ml, normal>180 pg/ml) was observed in serum. Motor and sensory conduction study and repetitive nerve stimulation test from nasalis, orbicularis oculi, and abductor digiti minimi muscles were normal. Electro and echocardiographic studies were normal. Brainstem auditory evoked potentials and somatosensory evoked potentials were normal but there was a mild delay in P100 latency on the right side during the visual evoked potential study (rt. 113.3 ms, lt. 105.9 ms, normal<108 ms). MRI of the brain showed early mineralization of globus pallidum and dentate nuclei with normal MR spectroscopy. Biceps muscle biopsy revealed nearly 25% ragged-red (Figure 1B, MGT) and ragged-blue fibers (Figure 1B, SDH) and 50% COX-negative fibers (Figure 1B, COX-SDH). Electron microscopy showed aggregates of a large number of mitochondria of varying size, altered cristae, parking lot inclusions and unusual triangular crystalline inclusions (Figure 1B, EM). Complex IV deficiency was detected in muscle biopsies. Multiple mtDNA deletions in muscles were confirmed by long-range PCR

(Figure 1C). Pathogenic mtDNA and POLG mutations had been excluded by Sanger sequencing. Informed consent was obtained from the patient according to the Declaration of Helsinki.

Whole Exome Sequencing

Whole exome sequencing (WES) was performed using 1 µg of DNA on a HiSeq2000 platform (Illumina), with a 2 x 100 bp paired-end setting. Base calling, demultiplexing and variant calling occurred as reported before (11). Variants were mapped to Ensembl transcripts and RefSeq transcripts and confirmed via Alamut Visual v2.15.0 (Interactive Biosoftware). Candidate variants were validated by Sanger sequencing (ABI 3730, Applied Biosystems).

GENESIS analysis

Two different set of filters were performed on WES data in GENESIS analysis. The first set of filters (denoted as 'Inheritance Filter') was predefined in GENESIS as 'autosomal recessive, strict' mode. Briefly, the 'Inheritance Filter' is based on data quality, variant type and frequency, evolutionary conservation and recessive inheritance (Supplementary table S1). By contrast, the second set of filters (denoted as 'Mitochondrial Filter') selected nuclear genes associated with mitochondria or mitochondrial diseases from MSeqDR (12) and MitoCarta (13), prior to applying the predefined 'autosomal recessive, strict' filters in GENESIS.

WeGET analysis

WeGET analysis requires custom-made and precomputed gene lists as query genes. Custom-made gene lists were generated from literature reported PEO, mtDNA maintenance, mitochondrial myopathy associated genes (1, 3, 14) and MitoCarta (13). Precomputed gene lists from GO-terms 'Mitochondrial organization'; KEGG pathway 'DNA replication'; and REACTOME pathways 'Mitochondrial biogenesis' were directly accessed from WeGET database (6). Each gene list was cross-validated in WeGET with the AUC ≥ 0.75, and genes ranking outside the 500 highest co-expressed position were removed. Then the trimmed gene lists were loaded into the GENESIS with a less stringent set of filters (denoted as 'WeGET Filter', Supplementary table S2).

Manual Checking

Candidate variants after GENESIS and WeGET analysis were manually checked by the following criteria: (A) Minor allele frequency (MAF) from public population databases including gnomAD (version v.2.1.1 controls, n=141,456) (15), 1000 Genome (1000G, Phase 3, n=2,504) (16), Exome Variant Server (EVS, version ESP6500, n= 6,503) (17) and GENESIS (20201205.ver, n=10,813) (4); (B) Predicted pathogenicity by prediction tools including CADD Phred (cutoff>20) (18), SIFT (cutoff<0.05) (19), Provean (cutoff<-2.5) (20); (C) Known pathogenic variants in ClinVar (21) and Leiden Open Variation Database (LOVD) (22) were checked for variant assessment.

Results

Whole exome sequencing data analysis

The presence of multiple mtDNA deletions in muscle biopsies of the patient and consanguineous parents suggested an autosomal recessive inheritance of the candidate gene with likely homozygous variants (Figure 1A). The 'Inheritance Filter', 'Mitochondrial Filter' and 'WeGET Filter' of WES data resulted in a total of 11 homozygous variants, all of which are inherited in an autosomal recessive mode (Supplementary table S3). Subsequent manual checking on frequency, allele count and predicted pathogenicity further narrowed down the candidate variants to be a homozygous deletion in the *C1QBP* gene. The homozygous deletion (NM_001212.3: c.611_613del) in *C1QBP* (Figure 1D) was not present in 1000 Genomes, ClinVar or LOVD, with only one heterozygote from non-Finnish European in gnomAD (MAF=0.000004). Although this variant is an in-frame deletion, it would lead to the loss of a F204 amino acid residue (NP_001203.1: p.(Phe204del)), which is highly conserved across different species (Figure 1E). Moreover, this deletion variant was predicted to be pathogenic according to SIFT, Provean and CADD (summarized in Table 1).

Table 1: Predicted consequence of *C1QBP* variants.

Item	This patient-Allele 1+2
Mutation, cDNA (NM_001212.3)	c.611_613delTCT
Protein (NP_001203.1)	p.Phe204del
Mutation, genomic (hg19)	chr17:5336699_5336701delAGA
Exon position of the mutation	5
gnomAD frequency	0.000003976 (1/251480)
1000 genome frequency	Not available
EVS frequency	Not available
GENESIS	1 homozygous allele
SIFT prediction	damaging
SIFT confidence score	0.894
Provean prediction	Deleterious
Provean score (cutoff<-2.50)	-12.61
CADD Phred score (cutoff>20)	22.70000076

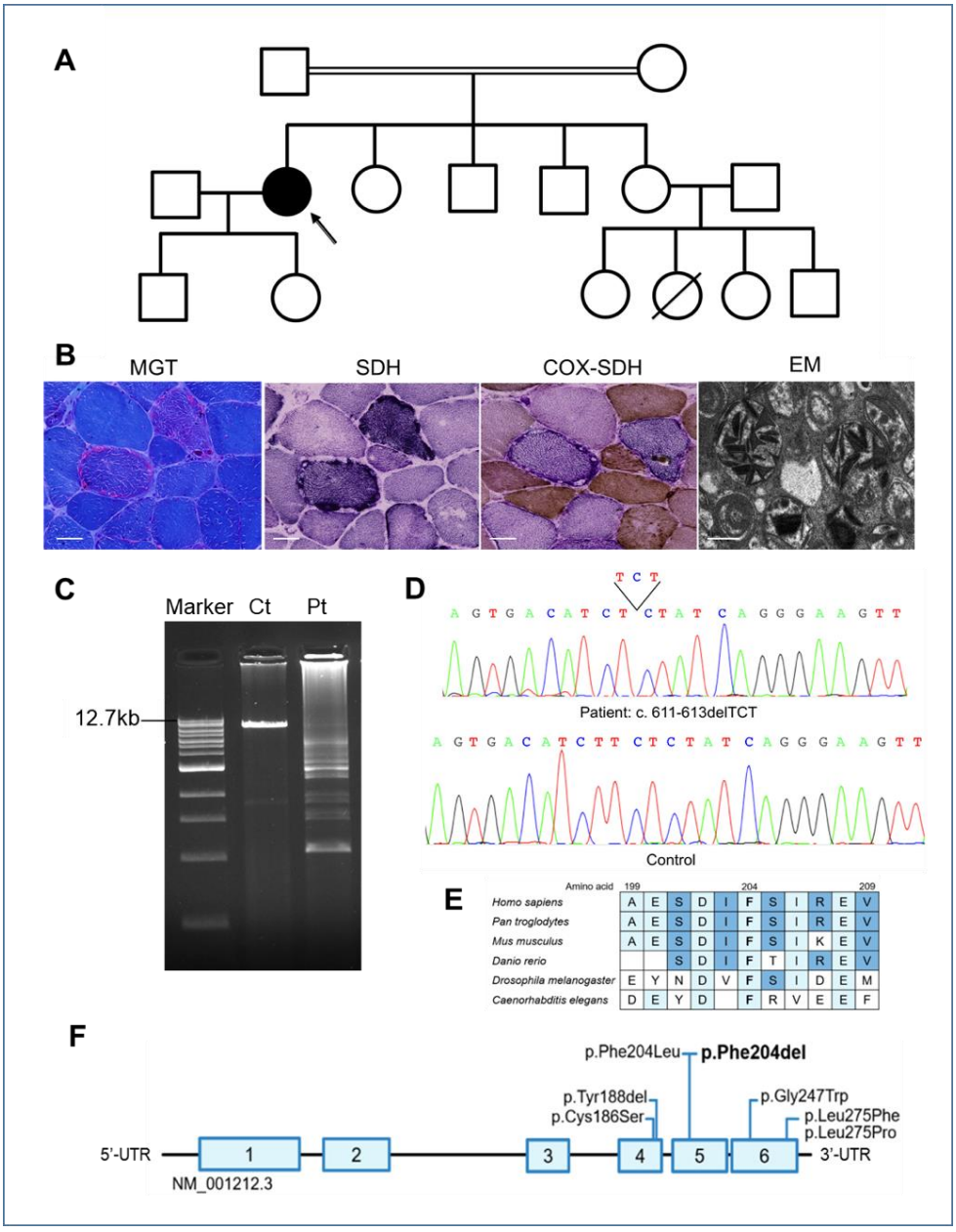


Figure 1: WES data analysis and genetic findings in the patient. (A) Pedigree of the patient family. The patient was depicted by the filled black symbol. (B) Transversely cut skeletal muscle tissue shows ragged red fibre (MGT) (× 200), ragged blue fibre (SDH), COX negative fibre (COX-SDH) (× 200) and electron microscopy (EM) shows abnormal mitochondria in size, altered cristae and parking lot inclusions. Scale bars, 50 µm (MGT, SDH, COX-SDH) and 1 µm (EM). (C) Long-range PCR of mtDNA showing multiple deletions in the patient (Pt) muscle biopsies, which are absent in the control (Ct). (D)

Electropherogram of the *C1QBP* regions containing the homozygous NM_001212.4:c.611_613del deletion in the patient. (E) Amino acid conservation study. The amino acid F204 (in bold) is conserved down to *C. elegans*, which is absent in the patient due to the NM_001212.3: c.611_613del variant. (F) Schematic view of *C1QBP* gene and all pathogenic variants identified to date. The variant in bold indicates the position of the amino acid changes present in the patient from this study.

Discussion

In this paper, we combined GENESIS and WeGET to identify candidate variants or genes causing PEO and multiple mtDNA deletions in an Indian patient. The consanguineous healthy parents of this patient indicated that the disease-causing gene follows an autosomal recessive inheritance. After applying three different filtering strategies and manual confirmation, rare and evolutionary conserved variants with high sequencing quality were prioritized, leading us to the discovery of a likely pathogenic variant in *C1QBP* gene. Considering the conservation and prediction results, as well as the diagnosis of PEO, the classification of this novel variant was changed to likely pathogenic (class 4) according to ACMG criteria (23).

C1QBP encodes the complement component 1Q binding protein (C1QBP), also known as p32. C1QBP is a ubiquitous protein mainly localized in mitochondrial matrix and has fundamental roles in mitochondrial ribosome biogenesis (24). Cellular models have been developed to study the role of C1QBP. Knockdown of C1QBP in HeLa cells shifted the normal mitochondrial network towards a more fragmented network, whereas overexpression of C1QBP showed a more fibrillar mitochondrial network (25). Reduced levels of C1QBP in cardiac cells, neurons and dendritic cells resulted in not only a defective mitochondrial function, but also a disturbed maturation, differentiation or signaling (26-28).

Biallelic *C1QBP* mutations were reported to cause mitochondrial cardiomyopathy and (or) PEO, with variable age of onset and severity of clinical presentations (summarized in Table 2). Till date, six pathogenic *C1QBP* variants have been reported from eight patients (Figure 1C). *C1QBP* mutations can cause a wide clinical phenotype, ranging from infantile lactic acidosis to childhood (cardio)myopathy and adult-onset PEO. In a previous study, four patients with biallelic *C1QBP* mutations were characterized with cardiomyopathy while lacking the involvement of the central nervous system (9). In that small cohort, a boy (deceased at 18 days) had two compound heterozygous variants (c.557G>C, p.Cys186Ser and c.612C>G, p.Phe204Leu) in *C1QBP*, of which the latter affects the same codon, which is deleted in our patient. In contrast to this patient who presented in the neonatal period, our

patient had a late-onset PEO with no overt cardiac symptoms. This illustrates the clinical heterogeneity of C1QBP defects. A recent study also reported two patients presenting with only a typical PEO phenotype (10), which is similar to the phenotype in our patient. In that recent study, an adult female (54 years) with late-onset PEO and multiple deletions in muscles had a homozygous variant (c.612C>G, p.Phe204Leu), which is the same variant found in compound heterozygosity with another missense variant (c.557G>C, p.Cys186Ser) in the boy previously reported. Such variable phenotypes could be explained by different position of the pathogenic *C1QBP* variants in the C1QBP protein. The amino acid changes associated with early-onset cardiomyopathy (e.g., p.Cys186Ser) are all localized in important structural domains (e.g., beta strand of the protein) whereas variants found in PEO patients (e.g., p.Phe204Leu) are localized on the coiled-coil region (10).

Multiple mtDNA deletions is typically observed in PEO patients. To explain the multiple mtDNA deletions in PEO patients harboring *C1QBP* mutations, one hypothesis is via interaction between C1QBP with RNase H1. RNase H1 is required to remove the replication template after one round of mtDNA replication, thus initiating a new replication round. Defects in RNase H1 are associated with late onset PEO with multiple mtDNA deletions (29-32). C1QBP was demonstrated to increase the cleavage activity of RNase H1 (29, 33), thus *C1QBP* mutations may affect the normal function of RNase H1. However, the detailed mechanism between *C1QBP* mutations and mtDNA deletions is still unclear and needs further investigation.

In conclusion, we identified a homozygous deletion in *C1QBP* from an Indian patient of consanguineous parents, presenting with adult-onset PEO and mtDNA multiple deletions. Despite the original paper associated *C1QBP* pathogenic variants with mitochondrial cardiomyopathy with variable age of onset, our data indicate that *C1QBP* mutations have to be considered in patients with isolated PEO and PEO plus phenotype.

Web resources

gnomAD	http://gnomad.broadinstitute.org/
1000 genome	https://www.internationalgenome.org/
EVS	https://evs.gs.washington.edu/EVS/
SIFT	https://sift.bii.a-star.edu.sg/
Provean	http://provean.jcvi.org/genome_submit_2.php?species=human
CADD Phred	https://cadd.gs.washington.edu/

Table 2: Genotype and phenotype comparison of patients carrying *C1QBP* variants.

Item	Patient 1	Patient 2	Patient 3	Patient 4
Literature	Wang et al., 2020(34)	Wang et al., 2020(34)	Marchet et al., 2020(10)	Marchet et al., 2020(10)
Gender	Male	Male	Female	Female
Age	14 years	9 years	54 years	65 years
Origin	Chinese descent	Chinese descent	Italian descent	Italian descent
Consanguinity	Unrelated	Unrelated	Second grade	Third grade
C1QBP variant (NM001212.3)	c.[823C>T];[823C>T]	c.[823C>T];[823C>T]	c.[612C>G];[612C>G]	c.[562_564del];[562_564del]
C1QBP variant (NP001203.1)	p.[Leu275Phe];[Leu275Phe]	p.[Leu275Phe];[Leu275Phe]	p.[(Phe204Leu); (Phe204Leu)]	p.[(Tyr188del);(Tyr188del)]
Onset	Childhood (1.5 years)	Childhood (2 years)	Adulthood (28 years)	Adulthood (45 years)
Eye	PEO, ptosis	Unapparent ptosis	PEO, bilateral ptosis	PEO, bilateral ptosis
Muscle	Exercise intolerance with fatigue	Exercise intolerance with fatigue	Mitochondrial myopathy	Mitochondrial myopathy, exercise intolerance
Heart	Cardiomyopathy	Cardiomyopathy	Normal	Normal
mtDNA	Normal	Normal	Multiple deletions in muscle	Multiple deletions in muscle
OXPHOS defects	Not available	Not available	Partially reduced Complex I, III, and IV in muscle	Normal in muscle
Histochemical analysis in muscle	Not available	Not available	RRF, RBF, and COX-negative/SDH-positive fibers	RRF, RBF, and COX-negative/SDH-positive fibers
Other presentations	Slack skins, upturned nose, increased blood lactate and ammonia	Increased blood lactate and ammonia	Muscle cramps and waddling gait, cushingoid appearance, hirsutism, neck drop, severe proximo-distal hyposthenia, severe dysphagia, and rhinolalia	Depression, reduced strength of neck flexors, neck tremor, proximal arms and legs hyposthenia, swallowing dysfunction

Patient 5	Patient 6	Patient 7	Patient 8	Patient 9
Feichtinger et al., 2017(9)	Feichtinger et al., 2017(9)	Feichtinger et al., 2017(9)	Feichtinger et al., 2017(9)	This study
Male	Female	Male	Male	Female
18 days (deceased)	4 days (deceased)	22 years	70 years (deceased)	33 years at biopsy
British descent	Japanese descent	Austrian descent	Italian descent	Indian descent
Unrelated	Unrelated	Unrelated	Unrelated	Third grade
c.[557G>C];[612C>G]	c.[739G>T];[c.824T>C]	c.[823C>T];[823C>T]	c.[562_564del];[562_564del]	c.[611_613del];[611_613del]
p.[Cys186Ser];[Phe204Leu]	p.[Gly247Trp];[Leu275Pro]	p.[Leu275Phe];[Leu275Phe]	p.[(Tyr188del);(Tyr188del)]	p.[(Phe204del);(Phe204del)]
Childhood (4 days)	Infantile (birth)	Childhood (5 years)	Adulthood (57 years)	Adulthood (33 years)
Not available	Not available	PEO, ptosis, astigmatism, amblyopia,	PEO, ptosis	PEO, bilateral ptosis
Not available	Not available	Myopathy, exercise intolerance	Myopathy, exercise intolerance, weakness	Mitochondrial myopathy, weakness in proximal limbs
Cardiomyopathy	Cardiomyopathy	Cardiomyopathy	Cardiomyopathy	Normal electroencephalography and echocardiogram
Increased copy number in muscle	Increased copy number in liver	Multiple deletions in muscle	Multiple deletions in muscle	Multiple deletions
Reduced Complex I, III, and IV in muscle	Reduced Complex I, II, III, and IV in liver	Reduced Complex I, III, and IV, increased CS in muscle	Reduced Complex I, III, and IV, increased CS in muscle	Reduced Complex IV in muscle
RRF, COX-deficient fibers	Not available	RRF, COX-deficient fibers	Isolated COX-deficient fibers	RRF, RBF, and COX-negative fibers

Abbreviations: RRF, ragged red fibers; RBF, ragged blue fibers; SDH, succinate dehydrogenase; COX, cytochrome C oxidase; CS, citrate synthase.

Acknowledgements
We would kindly acknowledge the family who is participating in this study.

Funding
LG was supported by funding from China Scholarship Council (Grant Number: 201606310036). Part of this work has been made possible with the support of the Dutch Province of Limburg (MG). BPS is supported by a clinical research fellowship from Mito-Foundation Australia. IdC is supported by Stichting NeMO (Grant Number: N9_2020).

Conflict of interest
The authors declare no conflict of interest.

References

1. McClelland C, Manousakis G, Lee MS. Progressive External Ophthalmoplegia. Current Neurology and Neuroscience Reports. 2016;16(6):53.
2. Moraes CT, DiMauro S, Zeviani M, Lombes A, Shanske S, Miranda AF, et al. Mitochondrial DNA Deletions in Progressive External Ophthalmoplegia and Kearns-Sayre Syndrome. New England Journal of Medicine. 1989;320(20):1293-9.
3. Viscomi C, Zeviani M. MtDNA-maintenance defects: syndromes and genes. J Inherit Metab Dis. 2017;40(4):587-99.
4. Gonzalez M, Falk MJ, Gai X, Postrel R, Schüle R, Zuchner S. Innovative genomic collaboration using the GENESIS (GEM.app) platform. Hum Mutat. 2015;36(10):950-6.
5. The Genesis Project [Internet]. [cited 2020]. Available from: https://www.tgp-foundation.org/.
6. Szklarczyk R, Megchelenbrink W, Cizek P, Ledent M, Velemans G, Szklarczyk D, et al. WeGET: predicting new genes for molecular systems by weighted co-expression. Nucleic acids research. 2015;44(D1):D567-D73.
7. Mootha VK, Lepage P, Miller K, Bunkenborg J, Reich M, Hjerrild M, et al. Identification of a gene causing human cytochrome c oxidase deficiency by integrative genomics. Proceedings of the National Academy of Sciences. 2003;100(2):605-10.
8. Kolde R, Laur S, Adler P, Vilo J. Robust rank aggregation for gene list integration and meta-analysis. Bioinformatics (Oxford, England). 2012;28(4):573-80.
9. Feichtinger RG, Olahova M, Kishita Y, Garone C, Kremer LS, Yagi M, et al. Biallelic C1QBP Mutations Cause Severe Neonatal-, Childhood-, or Later-Onset Cardiomyopathy Associated with Combined Respiratory-Chain Deficiencies. Am J Hum Genet. 2017;101(4):525-38.
10. Marchet S, Legati A, Nasca A, Di Meo I, Spagnolo M, Zanetti N, et al. Homozygous mutations in C1QBP as cause of progressive external ophthalmoplegia (PEO) and mitochondrial myopathy with multiple mtDNA deletions. Hum Mutat. 2020.
11. Nguyen M, Boesten I, Hellebrekers DMEI, Vanoevelen J, Kamps R, de Koning B, et al. Pathogenic CWF19L1 variants as a novel cause of autosomal recessive cerebellar ataxia and atrophy. Eur J Hum Genet. 2016;24(4):619-22.
12. Shen L, Diroma MA, Gonzalez M, Navarro-Gomez D, Leipzig J, Lott MT, et al. MSeqDR: A Centralized Knowledge Repository and Bioinformatics Web Resource to Facilitate Genomic Investigations in Mitochondrial Disease. Hum Mutat. 2016;37(6):540-8.
13. Calvo SE, Clauser KR, Mootha VK. MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. Nucleic acids research. 2015;44(D1):D1251-D7.
14. Amberger JS, Bocchini CA, Scott AF, Hamosh A. OMIM.org: leveraging knowledge across phenotype-gene relationships. Nucleic acids research. 2019;47(D1):1038-43.
15. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. Variation

across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. bioRxiv. 2019:531210.

16. Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, et al. A map of human genome variation from population-scale sequencing. Nature. 2010;467(7319):1061-73.
17. Tran KT, Le VS, Vu CD, Nguyen LT. A novel de novo variant of LAMA2 contributes to merosin deficient congenital muscular dystrophy type 1A: Case report. Biomed Rep. 2020;12(2):46-50.
18. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic acids research. 2018;47(D1):D886-D94.
19. Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic acids research. 2012;40(Web Server issue):W452-7.
20. Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics (Oxford, England). 2015;31(16):2745-7.
21. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic acids research. 2018;46(D1):D1062-d7.
22. Fokkema IFAC, Taschner PEM, Schaafsma GCP, Celli J, Laros JFJ, den Dunnen JT. LOVD v.2.0: the next generation in gene variant databases. Hum Mutat. 2011;32(5):557-63.
23. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17:405.
24. Saha SK, Kim KE, Islam SMR, Cho SG, Gil M. Systematic Multiomics Analysis of Alterations in C1QBP mRNA Expression and Relevance for Clinical Outcomes in Cancers. J Clin Med. 2019;8(4).
25. Hu M, Crawford SA, Henstridge DC, Ng IH, Boey EJ, Xu Y, et al. p32 protein levels are integral to mitochondrial and endoplasmic reticulum morphology, cell metabolism and survival. The Biochemical journal. 2013;453(3):381-91.
26. Gotoh K, Morisaki T, Setoyama D, Sasaki K, Yagi M, Igami K, et al. Mitochondrial p32/C1qbp Is a Critical Regulator of Dendritic Cell Metabolism and Maturation. Cell Rep. 2018;25(7):1800-15.e4.
27. Saito T, Uchiumi T, Yagi M, Amamoto R, Setoyama D, Matsushima Y, et al. Cardiomyocyte-specific loss of mitochondrial p32/C1qbp causes cardiomyopathy and activates stress responses. Cardiovascular research. 2017;113(10):1173-85.
28. Yagi M, Uchiumi T, Sagata N, Setoyama D, Amamoto R, Matsushima Y, et al. Neural-specific deletion of mitochondrial p32/C1qbp leads to leukoencephalopathy due to

undifferentiated oligodendrocyte and axon degeneration. Scientific reports. 2017;7(1):15131-.

29. Reyes A, Melchionda L, Nasca A, Carrara F, Lamantea E, Zanolini A, et al. RNASEH1 Mutations Impair mtDNA Replication and Cause Adult-Onset Mitochondrial Encephalomyopathy. The American Journal of Human Genetics. 2015;97(1):186-93.
30. Bugiardini E, Poole OV, Manole A, Pittman AM, Horga A, Hargreaves I, et al. Clinicopathologic and molecular spectrum of RNASEH1-related mitochondrial disease. Neurology Genetics. 2017;3(3):e149.
31. Sachdev A, Fratter C, McMullan TFW. Novel mutation in the RNASEH1 gene in a chronic progressive external ophthalmoplegia patient. Canadian Journal of Ophthalmology. 2018;53(5):e203-e5.
32. Carreño-Gago L, Blázquez-Bermejo C, Díaz-Manera J, Cámara Y, Gallardo E, Martí R, et al. Identification and Characterization of New RNASEH1 Mutations Associated With PEO Syndrome and Multiple Mitochondrial DNA Deletions. Frontiers in Genetics. 2019;10(576).
33. Wu H, Sun H, Liang X, Lima WF, Crooke ST. Human RNase H1 is associated with protein P32 and is involved in mitochondrial pre-rRNA processing. PLoS One. 2013;8(8):e71006.
34. Wang J, Li H, Sun M, Yang Y, Yang Q, Liu B, et al. Early Onset of Combined Oxidative Phosphorylation Deficiency in Two Chinese Brothers Caused by a Homozygous (Leu275Phe) Mutation in the C1QBP Gene. Front Pediatr. 2020;8:583047.

Supplementary materials

Supplementary table S1. Detailed settings in 'Inheritance Filter'.

'Inheritance Filter'	
Sequence quality	Depth filter > 12 Genotype quality filter > 75 Quality filter > 50
Variant type	Variant type: SNV, insertions, deletions Loss-of-function = transcript ablation, splice acceptor, splice donor, stop gained, frameshift, stop lost, start lost Nonsynonymous = transcript amplification, inframe insertion, inframe deletion, missense, splice region, incomplete terminal codon, stop retained, protein altering, coding sequence Show Loss-of-Function variants irrespective of conservation: No
Frequency*	ExAC and gnomAD MAF < 0.01 GENESIS heterozygous allele count* < 11 GENESIS homozygous allele count* < 11
Evolutionary conservation	PhastCons (100 vertebrates) > 0.5 phyloP (100 vertebrates) > 1.5
Inheritance	Both recessive patterns Hide results not supported by affected individuals

*Updated in January 2020

Abbreviations: ExAC, Exome Aggregation Consortium; gnomAD, genome Aggregation Database; MAF, Minor Allele Frequency; SNV, Single Nucleotide Variant.

Supplementary table S2. Detailed settings in 'WeGET filter'.

'WeGET Filter'	
Sequence quality	Depth filter > 8 Genotype quality filter > 50 Quality filter > 35
Variant type	Variant type: SNV, insertions, deletions Intron Loss-of-function = transcript ablation, splice acceptor, splice donor, stop gained, frameshift, stop lost, start lost Nonsynonymous = transcript amplification, inframe insertion, inframe deletion, missense, splice region, incomplete terminal codon, stop retained, protein altering, coding sequence Synonymous Show Loss-of-Function variants irrespective of conservation: No
Frequency*	ExAC and gnomAD MAF < 0.01 GENESIS heterozygous allele count* < 11 GENESIS homozygous allele count* < 11
Evolutionary conservation	PhastCons (100 vertebrates) > 0.5 phyloP (100 vertebrates) > 1.0
Inheritance	Both recessive patterns Hide results not supported by affected individuals

*Updated in January 2020

Abbreviations: ExAC, Exome Aggregation Consortium; gnomAD, genome Aggregation Database; MAF, Minor Allele Frequency; SNV, Single Nucleotide Variant.

Supplementary table S3. Candidate variants after 'Inheritance Filter', 'Mitochondrial Filter' and 'WeGET Filter' screening of WES data.

Nucleotide change	Amino acid change	g.(hg19)	Gene(s)	Ontology
NM_021962.4:c.1507T>C	NP_068781.2:p.(Tyr503His)	chr17:959329A>G	ABR	Missense
NM_001257999.2:c.1462C>T	NP_001244928.1:p.(Arg488Cys)	chr17:4098309G>A	ANKFY1	Missense
NM_001212.3:c.611_613del	NP_001203.1:p.(Phe204del)	chr17:5336699_5336701delAGA	C1QBP	Inframe deletion
NM_005687.4:c.986A>G	NP_005678.3:p.(Lys329Arg)	chr2:223489175T>C	FARSB	Missense
NM_212482.2:c.6923T>C	NP_997647.1:p.(Val2308Ala)	chr2:216232681A>G	FN1	Missense
NM_001136265.1:c.1457C>T	NP_001129737.1:p.(Pro486Leu)	chr1:19235152G>A	IFFO2	Missense
NM_004843.3:c.1205C>T	NP_004834.1:p.(Ala402Val)	chr19:14159856C>T	IL27RA	Missense
NM_015074.3:c.2107T>C	NP_055889.2:p.(Trp703Arg)	chr1:10381802T>C	KIF1B	Missense
NM_181537.3:c.847-1G>A	NP_853515.2:p.?	chr17:38935880C>T	KRT27	Splice acceptor
NM_004870.3:c.652G>T	NP_004861.2:p.(Val218Leu)	chr17:7490777G>T	MPDU1	Missense
NM_032409.2:c.1183G>A	NP_115785.1:p.(Gly395Ser)	chr1:20975057G>A	PINK1	Missense

ClinVar/LOVD	gnomAD allele count	gnomAD allele frequency	phastcons100way	phylop100way	SIFT	Provean	CADD_score
Not reported	122	0.0004955	1	6.8649997	Tolerated	Deleterious	17.85
Not reported	32	0.0001154	1	2.615	Damaging	Deleterious	21.1
Not reported	1	0.000004	1	7.651	Damaging	Deleterious	22.7
Not reported	110	0.0003971	1	4.8109998	Tolerated	Tolerated	22.5
Not reported	300	0.0012184	0.9789999	9.3430004	Damaging	Tolerated	13.39
Unknow effect	27	0.00015	1	6.6319999	Damaging	Deleterious	27.3
Likely benign	149	0.000538	0.939	1.9819999	Damaging	Tolerated	8.098
Benign	728	0.0026265	1	7.6449999	Damaging	Deleterious	29.8
Unknow effect	1775	0.0064077	1	2.9360001	Not available	Not available	19.97
Benign	438	0.0017797	1	2.184	Tolerated	Tolerated	22.1
Not reported	2	0.0000081	1	6.1209998	Tolerated	Deleterious	22.2

Chapter 4

Characterization of cellular and mitochondrial parameters in mesoangioblasts of myotonic dystrophy type 1 patients

Le Guo‡, Ruby Zelissen‡, Erika Timmer, Wendy Nguyen, Isis Joosten, Carla Gorissen, Karin Faber, Florence van Tienen*, Hubert Smeets*

‡ Both first author, * Both senior author

In preparation.

Chapter 5

CRISPR/Cas9 corrected mesoangioblasts as potential therapeutic strategy for myotonic dystrophy type 1

Le Guo, Erika Timmer, Isis Joosten, Rick Kamps, Patrick Lindsey, Florian Caiment, Carla Gorissen, Karin Faber, Florence van Tienen*, Bert Smeets*

* Both senior author

In preparation

Chapter 6

General Discussion

In this general discussion, I will fully focus on CRISPR/Cas9 technology, as it is one of the most promising developments in the scientific field.

1. Introduction

Clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) protein system provides researchers a simple, easy and efficient genome editing tool, which is widely adopted and greatly facilitates genetic engineering and translational studies. The CRISPR/Cas systems can be divided into two classes, six types and 33 subtypes, depending on the construction of the effector modules and the characteristic of *cas* genes and Cas9 proteins (1). Among them, type I, III and IV CRISPR systems (belong to class 1) consist of multiple Cas protein subunits, while type II (e.g. Cas9), type V (e.g. Cas12a, also known as Cpf1) and type VI CRISPR systems (belong to class 2) only require a single multidomain protein as the effector (1), as illustrated in Figure 1.

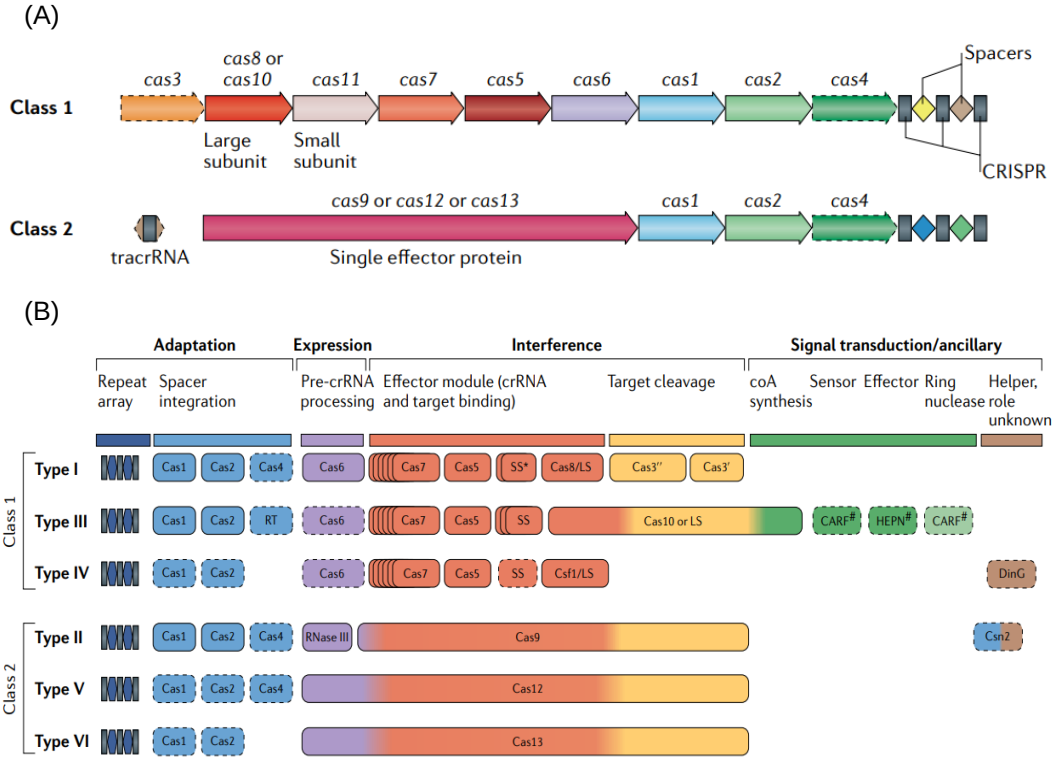


Figure 1. Classification of the CRISPR/Cas systems and their modular components (cited from (1)). (A) Generic compositions of the class 1 and class 2 CRISPR/Cas system. (B) Functional modules of CRISPR/Cas systems.

The most commonly used genome editing tool CRISPR/Cas9 belongs to the type II CRISPR system, and has originated from the prokaryotic adaptive immune system of *Streptococcus pyogenes* (SpCas9) against invading viruses and plasmids (2). CRISPR/Cas9 is an RNA-guided site-specific DNA cleavage system, composing of a non-specific Cas9 endonuclease protein, a programmable CRISPR RNA (crRNA) which directs Cas9 to cleave DNA and a fixed transactivating crRNA (tracrRNA) which hybridizes with crRNA to facilitate RNA-guided targeting of Cas9 (2, 3) (Figure 2). The tracrRNA:crRNA duplex can also be engineered as a single chimeric RNA (single guide RNA, sgRNA) that directs Cas9-mediated DNA cleavage (4). At the target DNA sequences, CRISPR/Cas9-mediated cleavage requires a 20nt DNA protospacer sequence complementary to the crRNA and a protospacer adjacent motif (PAM, typically 5'-NGG for SpCas9) which is an essential short sequence motif adjacent to the crRNA-targeted sequences. After binding to the target site, the crRNA matching DNA strand and the opposite strand is cleaved by the HNH nuclease domain and the RuvC nuclease domain of Cas9, respectively, generating a double-strand break (DBS) at the third base pairs upstream the PAM (5, 6). Following the generation of DSBs, subsequent cellular DNA repair process will start by error-prone nonhomologous end joining (NHEJ) pathway, which causes imprecise insertion/deletion (indel) mutations, or by homology-directed repair (HDR) pathway, which requires an exogenous DNA template (donor DNA) and generates a precise correction at the target locus (7-9). As NHEJ is functioning during the whole cell cycle whereas HDR is most active in the S phase, NHEJ is more efficient than HDR (10). In human proliferating cells, when offering an exogenous DNA template, NHEJ repairs about ¾ of DSBs, and HDR repairs the remaining ¼ (11).

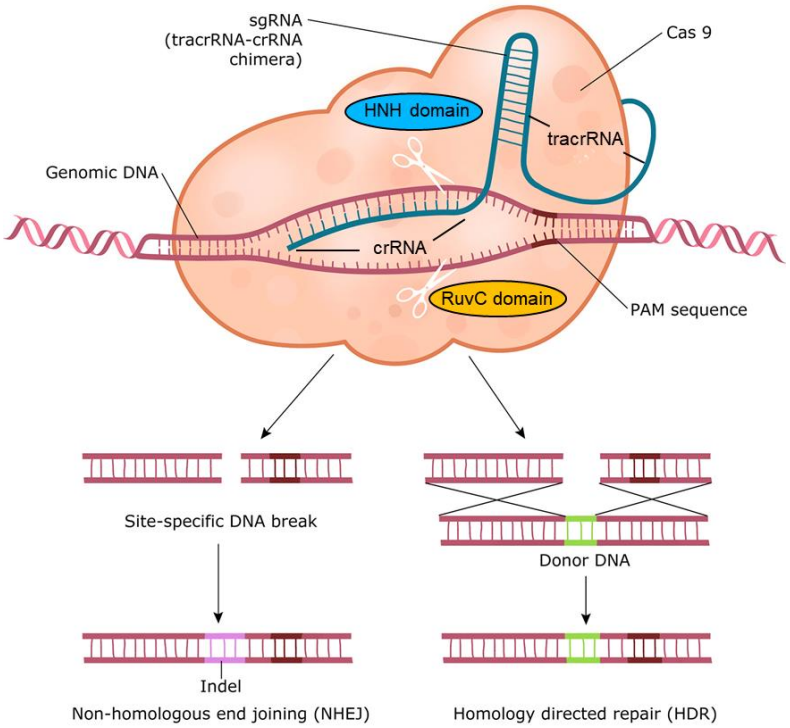


Figure 2. CRISPR/Cas9-mediated genome editing and two cellular repair pathways following the generation of a double-strand break (adapted from (12)).

CRISPR/Cas9 is different from other earlier-developed genome editing tools like zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs). First, CRISPR/Cas9 relies on Watson-Crick base pairing between guide RNA and target DNA, rather than protein-DNA interactions, as in ZFNs and TALENs, for target recognition. Second, CRISPR/Cas9 uses Cas9 protein, instead of Fok I used in ZFNs and TALENs for target DNA cleavage. Third, CRISPR/Cas9 only requires programmable guide RNAs and universal Cas proteins to generate sequence-specific cleavages, which is easy, quick and cheap. In contrast, ZFNs and TALENs need new combinations of protein recognition motifs each time, with context-dependent crosstalk between adjacent modules when assembled, which makes the designing and validating process labor-intensive, time-consuming and expensive (13, 14). Besides, CRISPR/Cas9 harbors facile multiplexing capability which allows multiple gRNAs targeting different DNA targets simultaneously, whereas ZFNs and TALENs suffer from costly and complicated assembly processes for multiplexing.

Moreover, CRISPR/Cas9 system is more efficient and specific, while less toxic to cells, compared to ZFNs and TALENs (Table 1). Due to these advantages over ZFNs and TALENs, CRISPR/Cas9 system becomes the mainstream tool of genome editing.

Table 1. A comparison of genome editing tools

Nuclease	Engineering	Efficiencies	Specificity	Toxicity	Clinical development
ZFN	difficult	high	low	medium	phase 1/2
TALEN	easy	medium	medium	medium	phase 1
Cas9/sgRNA	very easy	high	high	low	preclinical

2. An overview of different CRISPR/Cas9 platforms

2.1 Wild-type Cas9

Wild-type Cas9 (WT Cas9) causes DSBs at target sites, triggering the cellular machinery to repair the damage by one of at least two different pathways present in nearly all cell types and organisms: NHEJ and HDR (7-9). As mentioned above, NHEJ and HDR can have different editing outcomes for various applications. NHEJ is an error-prone repair, which can induce small indels of various lengths, leading to premature stops if targeting the coding exons or alternative splicing if targeting the splicing elements. Eventually, a knockout cell line or an animal model can be produced in this way. Besides, targeted deletion can also be achieved by applying two Cas9/gRNA complexes flanking the desired deletion region. In contrast, HDR-directed repair is precise and can insert specific point mutations or desired sequences by homologous recombination of the target locus with external DNA repair templates. Consequently, a knock-in or precisely corrected cell line or animal model can be created. The initial low efficiency of HDR activity can be improved by suppressing NHEJ key molecules like KU70, KU80 and DNA ligase IV (15), or by cell cycle synchronization and timed delivery of CRISPR/Cas9 components (16).

Besides SpCas9 derived from *Streptococcus pyogenes*, novel versions of Cas9 proteins are being continuously identified in various species of bacteria and found to be functional in mammalian cells, such as SaCas9 from *Staphylococcus aureus* (17, 18), NmCas9 from *Neisseria meningitides* (19), St1Cas9 from *Streptococcus thermophiles* (20, 21). Each CRISPR system can recognize a unique PAM for

targeted genome editing, thus broadening the choices of DNA targets. Due to the smaller coding sequences of SaCas9 proteins than other Cas9 proteins, the SaCas9 system enables larger cargo capability when packaged into plasmid vectors.

2.2 High-fidelity Cas9

WT Cas9-mediated genome editing can cause serious off-target cleavage at some loci targeting endogenous human genes, like *VEGFA* and *EMX1* (22). To tackle this issue, mutant Cas9 has been created with one or multiple specific variants, aiming at decreasing non-specific interactions between the Cas9 protein and the target site. Multiple mutations of DNA-interacting residues of the Cas9 protein have been engineered (on purpose or by evolution) to reduce the energetics of target DNA recognition and cleavage, resulting in several high-fidelity Cas9 variants, including spCas9-HF1 (23), eSpCas9 (1.1) (24), HypaCas9 (25), xCas9 3.7 (26), HiFi Cas9 (27) and Sniper-Cas9 (28).

spCas9-HF1 (carrying p.N497A, p.R661A, p.Q695A and p.Q926A) was demonstrated to retain high on-target activities comparable to WT Cas9 and reduced off-target activities to levels generally undetectable by GUIDE-seq and targeted deep sequencing, even in some targets with atypical homopolymeric or repetitive sequences (23). Besides, further engineered SpCas9-HF1 variants, such as SpCas9-HF2 (with an additional p.D1135E substitution) and SpCas9-HF4 (with an additional p.Y450A substitution), were shown to further reduce some residual off-target sites that susceptible to the SpCas9-HF1 (23). eSpCas9 (1.1) (carrying p.K848A, p.K1003A and p.R1060A) was engineered in a structure-guided manner, and was shown to reduce off-target effects in human cells by targeted deep sequencing and unbiased whole-genome off-target analysis, and to maintain robust on-target cleavage (24). HypaCas9 (carrying p.R692A, p.M694A, p.Q695A and p.H698A) was designed based on an observation that a non-catalytic domain within Cas9, REC3, recognizes target complementarity and governs the HNH nuclease to regulate overall catalytic competence (25). Experimental data demonstrated high genome-wide specificity without compromising on-target activity in human cells (25). HiFi Cas9 (carrying p.R691A) was identified through an unbiased bacterial selection of Cas9 mutants, which could maintain on-target editing efficiency while reducing off-target editing when delivered in ribonucleoproteins (RNPs) (27). When compared to HiFi Cas9, the high-fidelity Cas9 variants like spCas9-HF1, eSpCas9(1.1) and HypaCas9 exhibited reduced on-target editing at multiple tested sites when used in

an RNP format (27). xCas9 3.7 (carrying p.A262T, p.R324L, p.S409I, p.E480K, p.E543D, p.M694I and p.E1219V) was developed using phage-assisted continuous evolution with expanded PAM sequences (5′-NG-3′, 5′-GAA-3′ and 5′-GAT-3′) in mammalian cells (26). In addition to the broadened PAM compatibility, xCas9 3.7 was shown to have the minimal genome-wide off-target effect of tested sites with all PAMs, which can be explained by the conformational rearrangement in REC2 and REC3 domains (26). Sniper-Cas9 (carrying p.F539S, p.M763I and p.K890N) is also a high-fidelity Cas9 variant, deriving from *E.coli* following a directed evolution approach, and was shown to have high specificities without sacrificing on-target activities in human cells (28). Moreover, Sniper-Cas9 was shown to have WT-level on-target activities with extended or truncated gRNAs with further reduced off-target activities and work well in the RNP format to allow DNA-free genome editing (28).

2.3 Cas9 nickase

A Cas9 nickase is generated by mutating active cleavage sites of the WT Cas9, being either the HNH (e.g., H840A) or the RuvC domain (e.g., D10A), leading to the conversion of the WT Cas9 into a nicking nuclease Cas9 H840A or Cas9 D10A, respectively, with single-stranded DNA cleavage activity (2, 20). In contrast to the WT Cas9 nuclease which is guided by a gRNA to mediate a DSB at a particular target locus, the Cas9 D10A nickase can be led by a pair of appropriately spaced and oriented gRNAs to simultaneously introduce single-stranded nicks on both strands of the target DNA. To facilitate efficient double nicking in Cas9 D10A nickase, the pair of gRNAs must be designed with 5′ overhangs generated upon nicking, and the gRNA pairs must also have an optimal gap of approximately 20 bp (29).

The double-nicking Cas9 has shown to reduce off-target activity by 50- to 1500-fold compared to the WT Cas9 in cell lines and enable gene knockout in mouse zygotes without sacrificing editing efficiency (29). The individual nicks are predominantly repaired by the high-fidelity base excision repair pathway (30), whereas nicking of both DNA strands by a pair of Cas9 nickases leads to site-specific DSBs and NHEJ, resulting in indels formation with similar efficiency to that of the WT Cas9 (29). As single-stranded nicks are repaired without the indels formation, DSBs would only occur if both gRNAs are located at target sequences with proper space. Thus, the double nicking approach effectively doubles the number of bases needed for specific targets recognition and significantly increases the specificity of genome editing.

2.4 dCas9-Fok I

A catalytically inactive version of Cas9 or dead Cas9 (dCas9) and Fok I constitute the dCas9-Fok I system. The idea is adapted from ZFNs and TALENs in which the protein recognition motif domain is fused to the Fok I nuclease domain. The dCas9 maintains its ability to bind to target DNA sites defined by the gRNA, but DNA cleavage by dCas9-Fok I requires two gRNAs designed at around 15 or 25 base pairs apart to guide two dCas9-Fok I monomers to bind target sites simultaneously. The two adjacent dCas9-Fok I monomers subsequently assemble into a catalytically active Fok I nuclease dimer, thus triggering target DNA cleavage. Compared to Cas9 nickase, the dCas9-Fok I has a more stringent spacer, gRNA orientation and pairing requirements (31).

The benefit of dCas9-Fok I lies in the increased specificity as it requires two adjacent gRNAs for target recognition, which makes it ideal for target loci with highly similar sequences, pseudogenes or homologous gene families. In human cells, dCas9-Fok I was reported to have over 140-fold higher specificity than WT Cas9 and at least fourfold higher than that of paired nickases without compromised on-target cleavage efficiency (31). Target sites were calculated to occur on average every 34bp in the human genome, suggesting the versatility of dCas9-Fok I (31).

2.5 CRISPR activation and interference (CRISPRa/i)

Instead of fusing with endonuclease Fok I, the dCas9 can also fuse with transcriptional activator (e.g., VP64) or repressor domains (e.g., KRAB), which together act as CRISPR activation (CRISPRa) or CRISPR interference (CRISPRi), respectively, resulting in targeted gene regulation without introducing DSBs (32-35).

CRISPRa was shown to increase the expression of endogenous genes in human and mouse cells (36-38). Initial low expression activation efficiency can be improved by multiplexing gRNAs targeted to the same promotor, possibly due to the synergistic functioning (36-38). Besides, higher activation can be achieved by adopting VPR, SAM or SunTag activator system (39-42). The better performance of these CRISPRa systems relies on the multiplexing gRNAs, and many factors like the selected gRNAs, locus and the cell line, contribute to the difference in efficiency (35, 42).

CRISPRi was demonstrated to block transcriptional elongation, RNA polymerase binding or transcription factor binding, with no detectable off-target effects, confirmed

by both whole-genome sequencing and RNA-seq (32). As the innate nature of multiplexing of CRISPR/Cas9, CRISPRi can also repress multiple genes simultaneously. Besides, CRISPRi knockdown is inducible and reversible, making it a suitable platform to interrogate endogenous regulatory networks (32). CRISPRi has many advantages over RNA interference (RNAi). Eukaryotes lack CRISPR/Cas systems, thus avoiding endogenous pathway competition as in RNAi. Moreover, CRISPR/Cas9 can permanently change the genetic code or regulate gene expression at the (post)transcriptional level, whereas RNAi often generates complete gene knockouts and suffers from substantial off-target effects.

2.6 Base editing

Contrary to the classic CRISPR/Cas system, base editing does not rely on inducing and repairing DSBs nor homologous templates to achieve single base or base pair alteration in cells, which circumvents the issue of inaccurate repairing by NHEJ or HDR (43, 44). Base editing increases the efficiency of targeted genome modification and reduces the off-target events (43, 44). So far, two types of base editing have been developed, being DNA base editors and RNA base editors. Generally, DNA base editors consist of a guide RNA, a dCas9 nickase, an engineered deaminase as a base-modification enzyme and a short linker. And RNA base editors are similar to DNA base editors, except replacing the base-modification enzyme to the RNA-specific deaminase (45).

A recently engineered cytosine base editor YE1-BE3-FNLS with mutations in the predicted DNA-binding site can significantly decrease the off-target effects while maintaining high on-target editing efficiency (46). The cytosine base editors BE3, a fusion of APOBEC1, uracil glycosylase inhibitor, and Cas9-D10A nickase mutant, can convert C-G into A-T base pair, which is a widely used DNA base editor (47). The adenine base editor with tRNA specific adenosine deaminase and a Cas9 nickase can convert A-T to C-G, inactivating genes by converting four codons into STOP codons with fewer off-target mutations than Cas9 (48). Also, BE4 and SaBE4 were engineered to increase the efficiency of C-G to T-A conversion by approximately 50%, while halving the frequency of undesired by-products compared to BE3 (49). Moreover, fusing BE3, BE4, SaBE3, or SaBE4 to Gam, a bacteriophage DSBs-binding protein, further reduces indel formation to below 1.5% in most cases and improves product purity (49). Till now, multiplex base editing has been successfully performed in larger animal models, such as cynomolgus monkeys,

sheep and pig, providing the fundamental data before translational application (50-52). Still, established base editors can only make C-T or A-G base alteration, not gene knockout or knockin, thus limiting the choice of engineering the target DNA. Besides, base editors are too large to be packaged into AAV vectors, which hinders the systemic delivery for *in vivo* application.

2.7 Prime editing

Prime editing has been added recently to the toolbox of genome editing, which can directly write new genetic information into a targeted site, thus achieving all possible base conversions (53). The composition of the prime editing system includes an engineered Cas9 nickase fused to a reverse-transcriptase and a prime editing guide RNA called pegRNA including the target site and desired edit site (53). Prime editing can induce small indels and all 12 possible base-to-base conversions at target sites without DSBs or donor templates (53). Till now, prime editing has generated more than 175 edits in human cells and mouse cortical neurons, and specificity analysis showed that the prime editing had fewer off-targets than HDR without compromising the editing efficiency (53). However, as prime editing is still in its infancy, the editing efficiency only reached 20% to 50% in HEK293T and 7.1% in mouse cortical neurons (53). The large size of the engineered Cas9 makes it difficult to be delivered into cells, limiting the further application at this moment (54).

2.8 CRISPR-associated transposon

Unlike the canonical genome engineering based on the NHEJ and HDR, the transposon-encoded CRISPR/Cas system can efficiently insert DNA fragments in both mitotic and nonmitotic cells, with great potentials of clinical applications. A programmable CRISPR-associated transposase, composed of bacterial Tn7-like transposons and type V-K CRISPR effector, can catalyze RNA-guided integration of mobile genetic elements into the genome (55). Another transposon-encoded CRISPR/Cas system named INTEGRATE can induce the site-specific DNA integration via Tn7-like transposon and TniQ (56).

2.9 Conclusion

CRISPR/Cas9 is a game changer in the field of genetic editing, the reason why the Noble prize in Chemistry in 2020 was awarded to Emmanuelle Charpentier and Jennifer A. Doudna, who discovered this tool and developed the technology for application in eukaryotes. Using CRISPR/Cas9 researchers can change the DNA of animals, plants and microorganisms with extremely high precision. This technology has had a revolutionary impact on the life sciences, is contributing to new cancer therapies and may make the dream of curing inherited diseases come true.

3. Delivery

3.1 Delivery format

The key to successful CRISPR-mediated genome editing is the presence of a functional Cas9:gRNA ribonucleoprotein (RNP) complex in the cell nucleus. To achieve that goal, CRISPR/Cas9 can be delivered in three different formats: (1) a plasmid DNA (pDNA) encoding both the Cas9 enzyme and the sgRNA; (2) the Cas9 mRNA and sgRNA; (3) the Cas9:sgRNA RNP. Each of these delivery formats has advantages and disadvantages in overall efficacy as well as unique challenges.

The CRISPR/Cas9 pDNA is generally cheap and stable, but it has disadvantages such as the risk of host genome integration, plasmid immunogenicity and delayed effect (17, 57). The Cas9 mRNA and sgRNA enable transient delivery, but it suffers from the risk of degradation due to the innate instability of mRNA (58, 59). The RNPs delivery exhibits quick effect, short duration and low off-target effect. Compared to other delivery formats, the RNP delivery by electroporation showed higher editing efficacy, decreased off-target effects and less cellular toxicity in human stem cells and primary cell lines (60, 61). However, the RNP components can be costly to produce or purchase.

3.2 Delivery method

Multiple delivery methods have been explored to transport the CRISPR/Cas9 components into various types of cells, and these methods can be generally classified into viral and non-viral.

3.2.1 Viral vectors

Viral delivery utilizes a viral vector to encapsulate CRISPR/Cas9 components in DNA or RNA form to facilitate efficient delivery. The viral vector can be lentivirus, adenovirus, and adeno-associated virus (AAV). Characteristics of the three types of viral vectors are summarized in table 2 (62, 63).

Table 2: Characteristics of the three types of viral vectors for genome editing

Type	Genome	Payload	Expression duration	Immuno-genicity	Application	Tropism
Lentivirus	ssRNA	<8 kb	Transient	Low	*In vitro* and *ex vivo*	Not applicable
Adenovirus	dsDNA	5~37 kb	Long term (weeks to months)	High	*In vitro*, *ex vivo* and *in vivo*	Liver, neuro-muscular system
AAV	ssDNA	<5 kb	Long term (years)	Low	*In vitro*, *ex vivo* and *in vivo*	Broad

Lentiviral vectors are a subclass of retroviral vectors characterized by reverse transcription-mediated replication (64). Conventional lentiviral vectors harbor HIV-1 element, which enables permanent genetic modification of target cells by the semirandom chromosomal insertion of the transported cargo (65). Such semirandom chromosomal insertion would increase the risk of off-target events, thus raising the safety concerns in the genome editing. Therefore, integrase-defective lentiviral vectors were developed by mutating key positions in the HIV-1 *pol* region without interfering the viral transduction process (64). These integrase-defective lentiviral vectors were shown to be valuable vehicles for nucleic acid delivery in various human cell types like HEK293T (66) and iPSC cells (67).

Adenoviral vectors can deliver the viral double-stranded DNA rather than incorporate it into the host cell genome, making them suitable for applications requiring transient protein expression. Adenoviral vectors successfully delivered the Cas9 and a gRNA targeting PCSK9 into the liver of humanized mice, resulting in decreased plasma PCSK9 levels and cholesterol levels with undetectable off-target mutagenesis (68). Adenoviral transduction of Cas9/gRNA constructs was shown to delete the mutated exon 23 of *DMD*, leading to the partial recovery of *DMD* expression and improvement of the skeletal muscles in mdx mice (69). However, serious inflammatory response from early practical trials of adenoviral vectors raises safety concerns, which requires further investigation (70).

AAV vector is one of the leading delivery platforms for *in vivo* gene therapy (71). AAV is capable of delivering its single-stranded DNA cargo to various tissues and cell types, both dividing and non-dividing cells, and only causes mildly immunogenic responses in a dose-dependent manner (72). The vector genome largely remains stable as episome inside host cells without integrating into the host genome, which ensures its safety and durable therapeutic efficacy (72). However, limited cargo size (~4.7 kb) is the main drawback of AAVs, which could not package the spCas9, sgRNAs expression cassettes and donor template in one plasmid. Recently, a dual AAV system separating the Cas9 and sgRNAs into two vectors was efficiently delivered into a DMD mouse model, rendering functional recovery of skeletal muscle and heart in the treated mice (73). This dual AAV system successfully reduced the viral dose and preferential depletion of the sgRNA AAV genome, thus providing a solution for AAV-mediated systemic delivery of CRISPR/Cas9 components (73).

3.2.2 Non-viral delivery

Non-viral delivery can deliver all the three CRISPR/Cas9 formats via physical approaches (for example, microinjection and electroporation), chemical approaches (for example, cationic lipid-mediated polymers and gold-based nanoparticles) and physicochemical approaches (for example, iTOP). Compared to viral vectors, non-viral delivery systems are less toxic and immunogenic.

Microinjection is a direct method for delivery of CRISPR components into cells by penetrating the cell membrane with glass microcapillaries. This method has successfully generated a wide variety of knockout and transgenic animals, such as in mice (74), sheep (75) and pigs (76) by injecting zygotes/oocytes. Although microinjection is very efficient to deliver the CRISPR components and allows precise control of the delivery amount into single cells, it has drawbacks like low-throughput, labor-intensive and difficult for *in vivo* application (77).

Electroporation exerts a strong electric field to create transient holes in cell membranes, allowing nucleic acids and proteins to enter the cell. This technology has been successfully applied to deliver the three CRISPR formats *in vitro* and *ex vivo* for genome editing. For example, human B-cells were electroporated with CRISPR/Cas9 RNP to induce the production of therapeutic proteins (78). Mouse zygotes were electroporated with CRISPR/Cas RNP to target tyrosinase gene for precise sequence modification (79). However, the cellular toxicity is the main concern (62).

Cationic lipid-based delivery vectors are commercially available and widely used in transfecting various nucleic acids (80). The positively charged liposomes mediates the interaction of the nucleic acid cargo and the cell membrane, allowing for fusion of the cationic liposome/nucleic acid transfection complex with the negatively charged cell membrane. Cationic lipid vectors have been shown to efficiently deliver Cas9:sgRNA complexes into cultured human U2OS cells and hair cells from the mouse inner ear, resulting in up to 80% and 20% genome modification, respectively (81). Furthermore, the cationic lipid-based Cas9:sgRNA complexes transfection tolerates the use of 10% serum in culturing media and have higher specificity compared to DNA transfection (81).

Gold-based nanoparticles (AuNPs) are compatible with CRISPR/Cas9 components and have been demonstrated to deliver CRISPR/Cas9 components into human cells *in vitro* and animal models *in vivo* (82, 83). Of note, AuNPs complexed with cationic endosomal disruptive polymer were shown to efficiently deliver CRISPR/Cas9 RNP and donor DNA into a DMD mouse model, inducing HDR-mediated correction of the causative mutation, with minimal off-target events (83).

Induced transduction by osmocytosis and propanebetaine (iTOP) enables efficient transduction of native proteins and other macromolecules, such as small RNAs, into a wide choice of primary cells, without the help of any transduction peptide (84). The underlying mechanism is through an active uptake process in which NaCl-mediated hypertonicity, in combination with propanebetaine as a transduction compound, induces micropinocytosis and intracellular release of the transported macromolecules (84). Cas9 proteins and gRNAs were successfully co-transduced into KBM7 and H1 human embryonic stem cells by iTOP in a dTomato reporter system targeting AAVS1, reaching up to 56% and 26% positive cells, respectively (84).

3.3 Conclusion

CRIPSR/Cas9 systems can be delivered in three formats (pDNA, RNA and RNP) with a variety of delivery methods (viral and non-viral). Critical in choosing which method to use is the balance between delivery efficiency, which is linked to correction efficiency, and the viability of the cells after delivery. This is especially an issue if adult stem cells are *ex vivo* corrected in order to use them for treatment. Apart from correction efficiency and viability, also safety has to be considered as these stem cells are transplanted into human patients.

4. Off-target activity

As CRISPR/Cas9-mediated genome editing results in permanent modifications within the genome, the targeting specificity of Cas9 nucleases is of significant importance. CRISPR/Cas9 may cause more than half of off-target activity *in vitro*, leading to undesired mutations and cytotoxicity (85, 86). However, recent studies have shown that off-target events are barely detected in large animal models like sheep and goat (87, 88). Cas9 target recognition is dependent on the DNA-RNA interaction via Watson-Crick base pairing. Mismatch tolerance between the gRNA and its target DNA directly causes off-target events. Previous studies suggested that the first 8-12 PAM-proximal sequences, or the seed sequences in gRNAs determine the specificity of spCas9 (2, 3). However, recent studies demonstrated that all bases within the gRNA contribute to overall specificity (22, 89-91). Particularly, the specificity of CRISPR/Cas9 is decided by multiple factors including: (i) position of mismatches: the 8–14 bp on the 3′end of the guide sequence, also known as the seed sequence, is less tolerant of mismatches than the 5′bases, thus determining the specificity; (92) (ii) quantity of mismatches: Cas9 was reported to tolerate up to five mismatches within the target site; (iii) guide sequence context: some guide sequences are less tolerant of mismatches than others; and (iv) concentration of Cas9 and gRNA: off-target cleavage is highly sensitive to the transfected amounts, as well as relative ratios of Cas9 and gRNA. Finally, species-specific genomic sequence backgrounds and delivery methods also have an impact on the off-target activity.

4.1 Detection methods of off-target

Typically, off-target sites can be computationally predicted by searching for genomic sequences similar to the target locus, and various *in silico* tools are available, such as CCTop (93), Cas-OFFinder (94) and COSMID (95), leading to a list of the most likely off-target sites. Subsequent analysis by in-gel T7 endonuclease I (detection limit 2%~5% mutation frequency) or Surveyor nuclease assay of top-ranked off-target sites can provide a general number for assessing specificity. Although these methods are quick and easy to offer a general profile of off-target activities, biased results can be generated due to the incomplete predicted off-target candidate list.

To avoid biased results, next-generation sequencing platforms like whole genome sequencing can be integrated into thoroughly screening for off-target sites. Whole-genome-based detection approaches like genome-wide unbiased identification of DSBs enabled by sequencing (GUIDE-seq), *in vitro* Cas9-digested whole-genome sequencing (Digenome-seq) and circularization for *in vitro* reporting of cleavage effects by sequencing (CIRCLE-seq) have been proposed to assess an unbiased specificity of Cas9 cleavage, with CIRCLE-seq having the best performance among the three approaches detecting genome-wide off-target sites *in vitro* (96-98). More detection methods have been reported recently, including target-enriched GUIDE-seq (TEG-Seq), iGUIDE-seq (an improved GUIDE-seq), and DIG-seq (based on Digenome-seq), which provide more choices to accurately detect off-targets in genome editing system (99-101).

Off-target activity can cause tumorigenesis if oncogenes are unexpectedly hit, thus it is important to assess the consequence of off-target events for clinical and translational research. Tumorigenesis can be examined *in vitro* or *in vivo* (102, 103). *In vitro*, CRISPR/Cas9 edited cells can be evaluated by soft agar colony formation assay, with cervical cancer cells forming colony as the positive control. *In vivo*, CRISPR/Cas9 edited cells can be transplanted to nude mice or rat to examine the possibility of teratoma formation, with pluripotent stem cells forming teratoma as the positive control.

4.2 Solutions to reduce off-target

Several plausible strategies, already partly discussed above, have been reported to reduce off-target activity:

(1) Using a double-nicking Cas9 system. The double-nicking Cas9 increases the overall number of bases that are specifically recognized in the target DNA, and off-target nick sites are precisely repaired by the high-fidelity base excision repair pathway. This double-nicking strategy was shown to improve the specificity by up to 1,500-fold compared to the WT Cas9 (29, 30);

(2) Using an RNA-guided dCas9-Fok I system. Two dCas9-Fok I monomers are required to bind target sites simultaneously with a defined distance. Since the dCas9-Fok I system relies on RNA-guided Fok I dimerization, two adjacent off-target binding events and subsequent cleavage are less likely to occur than a single off-target cleavage;

(3) Using the base editing approach. Base editing directly converts a target base or base pair into another without inducing DSBs, therefore it circumvents the inaccurate repairing of DSBs (43, 44);

(4) Designing gRNAs properly. *In silico* predicted gRNAs with low off-target effects should be preferred. gRNAs truncated by 2 or 3 nt at 5' end were reported to significantly increase the specificity of Cas9, possibly due to greater mismatch sensitivity (104). Experimental data showed that the high-GC content of gRNA can increase off-targets by stabilizing the RNA-DNA hybridization at non-specific sites (105). Besides, modified gRNAs like bridged nucleic acids or locked nucleic acids at specific locations of crRNAs also successfully reduced the off-target DNA cleavage (106). These design strategies of gRNAs can be combined with the double-nicking or other strategies to further reduce the off-target activity;

(5) Using high-fidelity Cas9 variants. These Cas9 variants were engineered to weaken the DNA binding activity between the Cas9 protein and the target DNA, leading to improved specificity (24, 25, 28);

(6) Using RNP complexes to deliver CRISPR/Cas9 components. Compared to plasmid-based delivery, RNP initially presents with a high concentration. However, RNP has a shorter turnover time within cells due to the degradation by endogenous proteases and RNases, which can reduce the chance of off-target risks (107).

4.3 Conclusion

Multiple factors contribute to off-target events after CRISPR/Cas9-mediated genome editing. Enzymatic digestion, unbiased whole-genome-based sequencing and tumorigenic assay can be used to evaluate the off-target events. A careful design and selection of CRISPR/Cas9 platforms can help minimize the off-target events.

5. CRISPR-based gene editing applications

To date, CRISPR/Cas9 has been widely used to achieve efficient and accurate genome editing in a number of species and cell types, including human cell lines, mouse, rat, rabbit, zebrafish, fruit fly, silkworm, yeast, bacteria and common crops (6). Different scientific purposes have been addressed by CRISPR/Cas9 approaches to facilitate a variety of basic and translational research questions, including gene therapy, diagnosis, disease modelling, functional elements screening and live cell labelling (108) (Figure 3).

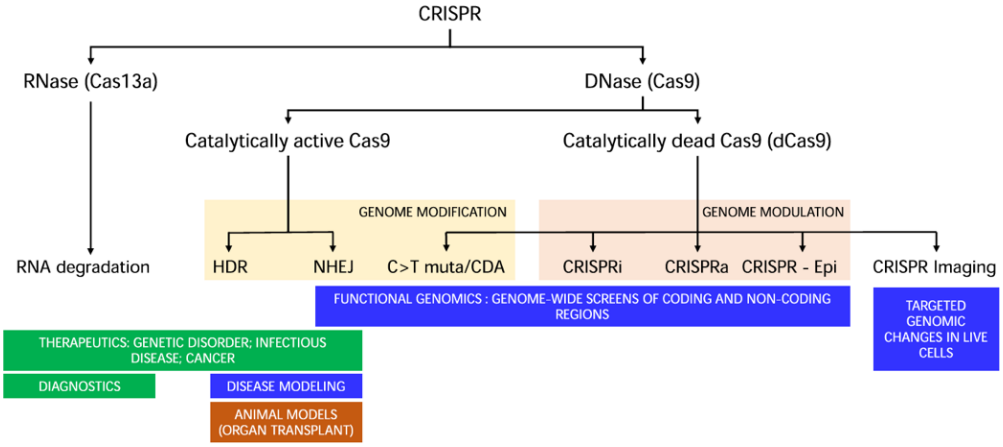


Figure 3. Applications of CRISPR-based gene editing

5.1 Therapy in genetic disease and cancer

5.1.1 Gene therapy

Gene therapy is either adding a functional gene copy (gene replacement) or correcting an existing mutated copy (gene correction). Adding a normal copy has been successfully done for spinal muscular atrophy (SMA) and developed into an FDA-approved gene therapy called Zolgensma. However, gene replacement is ineffective for diseases due to gain-of-function mutations (like DM1). In this part of the discussion, I will focus on gene correction strategies rather than gene replacement.

In 2017, Luxturna, the first human gene therapy drug for an inherited retinal dystrophy, together with Kymriah and Yescarta, two cell-based gene therapies for acute lymphoblastic leukemia, are approved by FDA and entered the US market. Since then, nearly twenty gene therapy products have been approved and over two thousand gene therapy clinical trials have been reported worldwide (109).

Among these various gene therapy products and clinical trials, they can be categorized into two approaches, *in vivo* or *ex vivo*. *In vivo* approach aims to introduce genome editing reagents into patients systemically or locally, usually via genetically engineered viruses to directly manipulate cells in the body. In *ex vivo* approach, genome editing components are introduced into isolated human cells to achieve the desired genetic modification. After validation and expansion, the genetically modified cells are infused into patients to have a therapeutic effect.

5.1.2 CRISPR clinical trials – *in vivo* approach

Some human cell types are highly differentiated and belong to post-mitotic cells (for example, neurons and myotubes) which are difficult to isolate and expand. As these cells are functional and manipulatable *in vivo*, *in vivo* genome editing is more suitable to modify these cell types relevant to a range of human diseases. However, the safety concerns of using virus for systemic delivery and of irreparable off-target effects remain to be one of the major obstacles of *in vivo* CRISPR therapy. A summary of *in vivo* CRISPR/Cas9-based clinical trials are listed in Table 3.

Table 3 CRISPR clinical trials adopting an *in vivo* approach

Product name	Developer/ sponsor	Structure and mechanism	Therapeutic indication and target tissue	Clinical trials ID
AGN-151587	Allergan	An AAV5 vector contains two gRNAs and Cas9 to correct IVS26 mutation in CEP290 gene	Leber Congenital Amaurosis 10 (LCA10)	NCT03872479
ND	First Affiliated Hospital, Sun Yat-Sen University	CRISPR/Cas9 plasmids in gel targeting HPV E6/E7	HPV-related cervical intraepithelial neoplasia	NCT03057912

ND, not disclosed; Data are from www.clinicaltrials.gov (accessed on March, 2021).

5.1.3 CRISPR clinical trials – *ex vivo* approach

The robustness of introducing CRISPR-based genome editing components into cell cultures, the possibility of confirming edited cells before infusion into patients and the limited human immune responses are major advantages of the *ex vivo* approach, promoting a number of CRISPR clinical trials targeting various diseases (Table 4). Current *ex vivo* CRISPR clinical trials mainly focus on cancer, solid tumor and HIV infection.

Table 4 CRISPR clinical trials adopting an *ex vivo* approach

Product name	Developer/ sponsor	Structure and mechanism	Therapeutic indication and target tissue	Clinical trials ID
Cyclophosphamide (PD-1 Knockout T Cells)	Sichuan University	Ex vivo gene manipulation of T cells	Metastatic non-small cell lung cancer	NCT02793856
PD-1 Knockout T Cells	Yang Yang	Ex vivo gene manipulation of peripheral blood lymphocytes by PD1 gene knockout	Advanced-stage EBV associated malignancies	NCT03044743
PD-1 Knockout T Cells	Hangzhou Cancer Hospital	Ex vivo gene manipulation of T cells	Esophageal cancer	NCT03081715
CTX001	Vertex Pharma-ceuticals Incorporated	Autologous CD34+ hematopoietic stem and progenitor cells manipulated at enhancer position of the BCL11A gene leads to increase in fetal hemoglobin	Hematologic diseases	NCT03655678 NCT03745287
CTX110	CRISPR Therapeutics	Allogeneic CRISPR/Cas9 gene-edited CAR-T cell therapy targeting CD19	B-cell malignancy	NCT04035434
CTX120	CRISPR Therapeutics	Allogeneic CRISPR/Cas9 gene-edited CAR-T cells	Multiple myeloma	NCT04244656

		targeting B-cell maturation antigen		
CCR5 gene modification	Affiliated Hospital to Academy of Military Medical Sciences	CD34+ hematopoietic stem/progenitor cells from donor are manipulated with CRISPR/Cas9 aiming at CCR5 gene deletion from cell surface	HIV-1-infection	NCT03164135
UCART019	Chinese PLA General Hospital	CRISPR/Cas9 mediated CAR-T cells targeting CD19 in individuals with relapsed/ refractory CD19+ leukemia and lymphoma	leukemia and lymphoma	NCT03166878
Anti-Mesothelin CAR-T cells	Chinese PLA General Hospital	Knocking out PD-1 and TCR genes with CRISPR/Cas in CAR-T cells	Adult solid tumor	NCT03545815
Universal Dual Specificity CD19 and CD20 or CD22 CAR-T Cells	Chinese PLA General Hospital	Using CRISPR/Cas9 to target CD19 and CD20 or CD22 in CAR-T cells for treatment of relapsed or refractory leukemia and lymphoma	Leukemia, Lymphoma	NCT03398967

Data are from www.clinicaltrials.gov (accessed on March, 2021).

5.2 Diagnosis

CRISPR/Cas9 technologies can be applied in the field of disease diagnosis, and several CRISPR-based diagnostic platforms have been developed to detect certain types of nucleic acids.

Specific high-sensitivity enzymatic reporter unlocking (SHERLOCK), a CRISPR-based diagnostic platform, combines recombinase polymerase amplification (110) with Cas13 or Cas12 for specific recognition of desired DNA or RNA sequences, allowing multiplexed, portable and sensitive detection of DNA or RNA viruses (111). The detection limit of the Cas13-SHERLOCK was reported to be 2 aM and 20 aM of a synthetic RNA and DNA dilution, respectively.

DNA endonuclease-targeted CRISPR trans reporter (DETECTR), a Cas12-based assay, enables rapid (<40 min), easy and accurate detection of COVID-19 (112). DETECTR is based on simultaneous reverse transcription and isothermal amplification using loop-mediated amplification (113) for RNA extracted from swabs, followed by Cas12 detection of predefined coronavirus sequences and cleavage of a FAM-biotin reporter molecule. When tested in 36 COVID-19 infected patients and 42 non-COVID-19 patients, DETECTR could reach 95% positive predictive agreement and 100% negative predictive agreement.

Combinatorial arrayed reactions for multiplexed evaluation of nucleic acids (CARMEN), a platform for scalable, multiplexed pathogen detection with Cas13, which was shown to simultaneously differentiate all 169 human-associated viruses and comprehensively subtype influenza A strains and identify dozens of HIV drug-resistance mutations (114).

5.3 Disease modeling

CRISPR/Cas9-mediated genome editing enables the rapid generation of cellular and animal models, thus providing a suitable testing platform for novel therapeutic drugs and treatment. CRISPR/Cas-mediated genome editing has been used to model the specific causative variant for monogenetic disease harboring a particular mutation, such as Duchenne muscular dystrophy (115) and cystic fibrosis (116). The multiplexing capability of Cas9 enables targeting several different gene loci simultaneously. Therefore, CRISPR/Cas-mediated genome editing has also been used to model polygenic diseases harboring multiple mutations, such as diabetes (117), cardiovascular disease (118) and Parkinson’s disease (119).

A major advantage of CRISPR-based disease modeling is that CRISPR-generated models can reproduce genetic mutations found in patients, instead of relying on disease models that only phenocopy a particular disorder. For example, specific mutations can be introduced or corrected in isogenic adult stem cells or iPSCs by CRISPR/Cas9 (120, 121). Besides, CRISPR-based editing could also be used to establish the causal roles of specific genes or genetic variations. Genome-wide association studies (GWAS) have identified large numbers of haplotypes that show strong association with certain disease risks (122). However, it is often difficult to determine which of several genes or genetic variants are responsible for the phenotype. By contrast, CRISPR/Cas9 models could help to study the effect of each individual variant, at least if the usually small effect of such a variant can be measured and if the effect is not dependent on other variants, or test the effect of each individual gene on an isogenic background by editing stem cells and differentiating them into cell types of interest.

5.4 Functional elements screening

Functional elements screening requires genome perturbation, which can be achieved by RNAi. However, the RNAi approach has issues of partial knockdown, extensive off-target effects and restrictive targets selection in transcribed genes (123).

The efficiency and multiplexing capacity of the CRISPR/Cas9 system makes it suitable for genome-wide functional screening of genomic elements. Cas9-mediated pooled sgRNAs screenings have been shown to have increased sensitivity as well as consistency (124). Besides, these sgRNAs could be designed to target nearly any DNA sequence, including enhancers, promotors and other regulatory elements.

The common approach of CRISPR/Cas9-based screenings for functional elements is via lentiviral transduction of sgRNAs targeting genes of interest, thus perturbing multiple genomic elements simultaneously (125-127). Conjugating dCas9 to different effector domains could also facilitate genomic screens. For example, dCas9 fused to epigenetic modifiers has been used to study the effects of methylation (128), certain chromatin-protein interaction (129) or disease mechanisms (130).

5.5 Live cell imaging

Studying the interactions of specific genes and chromatin dynamics would require a robust method to visualize DNA in living cells. Traditional techniques for labeling DNA, such as fluorescence in situ hybridization (FISH), require sample fixation and are therefore not suitable for live cell imaging. The characteristic of specific binding between dCas9/gRNA and target genomic loci serves the purpose of live cell imaging, which can be achieved by fluorescently labeled dCas9 or gRNA (131). Novel CRISPR-based genome labeling systems have been developed and tested (132). For example, a biomolecular fluorescence complementation (BIFC)-dCas9/gRNA strategy was shown to have high signal-to-noise ratios without non-specific foci (133). Advances in novel Cas9 proteins or modified sgRNAs would create multi-color and multi-locus capabilities, thus enhancing the power of CRISPR-based imaging for studying complex chromosomal architecture and nuclear organization.

5.6 Other organisms

Conventional crop and farm animal breeding mainly rely on screening beneficial genetic variation from spontaneous mutations and recombination following hybridization, which is often time-consuming and labor-intensive. Recent developments of CRISPR technology enables targeted and precise genetic engineering of plants and animals, thus accelerating the transition towards precision breeding for traits improvement and disease resistance.

In food crops, the main focus lies on the improvement of crop yields, quality and stress resistance, which can be achieved by CRISPR-mediated knockout of genes regulating undesirable traits (134). For example, knocking out *Gn1a*, *DEP1* and *GS3* in rice led to expanded grain number, dense erect panicles and larger grain size (135); disrupting the waxy gene *Wx1* in maize resulted in high amylopectin content with improved digestibility that holds the potential to be commercialized (136); and destroying the *MLO* allele generated powdery mildew-resistant wheat and tomato (137). As knockout of essential genes can cause seedling lethality, CRISPR-mediated gene regulation acts as a promising solution without disrupting these important genes. For example, the promoters of genes related to quantitative traits such as fruit size, inflorescence branching and plant architecture were mutated by creating a serial of variation for tomato breeding, which efficiently altered gene

expression levels while avoiding the integration of foreign DNA (138). Many beneficial traits in crops can also be generated by base editing. For example, the base editing of acetolactate synthase has successfully created herbicide-resistant wheat, watermelon and rice, which increased the productivity of agricultural systems by controlling weeds and preserving soil (139-142).

In livestock, disease resistance is one of the most popular applications for CRISPR. For example, CRISPR-edited pigs were shown to be protected from African swine fever, which killed over 90% of infected domestic pigs (143). Economically significant traits can also be added to the livestock by CRISPR/Cas9. For example, CRISPR-mediated knockout of the *MSTN* gene in sheep embryos resulted in the desired muscle hypertrophy without detectable off-target effects, and the live mutant sheep were healthy (144). Valuable add-on products can be produced by CRISPR/Cas9 editing as well. For example, CRISPR-mediated knockout of beta-lactoglobulin goats could produce milk without allergen and edited cows could produce human serum albumin in milk (145).

5.7 Conclusion

CRISPR-mediated genome editing can be applied in diverse scientific fields, including gene therapy, diagnosis, disease modelling, functional elements screening and live cell labelling, which can address a variety of basic and translational research questions. As it is more precise than gene therapy, where a functional copy of gene is being added, it holds a much better promise for precision medicine in the near future.

6. Ethical and regulatory issues

The ethical and regulatory issues regarding CRISPR-mediated genome editing can be generally divided in non-human and human aspects.

For CRISPR-based editing in non-human species, ecological risks and impacts are the major concerns. For example, gene modified crops by CRISPR (considered as genetically modified organisms in Europe, but not in the United States and other countries), enabling traits like drought resistance and fast growth may become invasive weeds and may pose an unexpected risk to other species. Besides, the

CRISPR technologies provide possibilities to produce bioweapons by engineering infectious pathogens that infect humans or crops.

For CRISPR-based editing in human, ethical issues largely depend on the treated cells, being somatic cells and germline cells. Gene editing in somatic cells and most pluripotent stem cells (PSCs), are intended to treat patients themselves or study the cellular pathological mechanisms, and if used for treatment, these cells are not transmitted to offspring. So, ethical issues are limited when somatic gene editing became safe and cheap, and it is a generally accepted strategy for disease treatment. In contrast, gene editing in germline cells can be passed on to future generations and has generated intense ethical debates and discussions. At this moment editing the human germline is forbidden in most countries and the discussion is going on whether it should be allowed for research purposes.

Ethical discussions on human germline genome editing lie in basic research, pre-clinical research and future clinical settings, though there may be an overlap between these types.

- In the basic research area, some debates and questions of genome editing in human germline include: is it safe to specifically correct germ cells or embryos? What is the identity of abnormally fertilized/corrected embryos? Is a later 14-day limit regarding embryo research be possible? What measures can be taken to ensure the autonomy and protect the welfare of oocyte donors? Will correction of germ cells/embryos be a slippery slope (146).
- Pre-clinical research aims at evaluating the effectiveness and safety of future clinical germline genome editing. Although the value of CRISPR technologies in biomedical and agricultural application is substantial as aforementioned, certain risks are still present. Current risks of genome editing mainly involve incomplete editing (mosaicism) (147), off-target events (148) and unknown long-term effect (due to long life-cycle) or undesired applications.
- In future clinical settings, germline genome editing has similar health concerns and medical risks as in the pre-clinical research. And whether a step towards clinical application of germline genome editing will be taken depends on the results of basic and pre-clinical research (safety), as well as the ethical, social and legal discussions.

In terms of regulations and guidelines, different countries have developed distinct policies or laws towards human germline gene editing. These ethical guidelines vary widely around the world, such as ‘restrictive’ in the United States, ‘legal prohibition’ in the UK, ‘prohibition by guidelines’ in Japan, China, Ireland and India, and ‘ambiguous’ in Russia, Argentina, South Africa, Chile, Slovakia, Peru, Colombia, Iceland and Greece (92). By contrast, the modification of human embryo beyond 14 days of its development is ethically prohibited in many countries (149). Also, germline genome editing is prohibited and highly discourages in International law including the UNESCO Declaration on the Human Genome and Human Rights as well as the Council of Europe Convention on Human Rights and Biomedicine (150).

The CRISPR babies experiment by the Chinese researcher He Jiankui was a bad example for germline editing. Although the purpose of trying to make children of HIV-positive spouses free from HIV infection is unquestioned, he illegally forged approval of ethics committee and performed the experiment without knowing whether his CRISPR method would work. Besides, the two edited CRISPR babies could not avoid the mosaic problem, which would lead to the failure of the immune protection against HIV. After the report of gene-edited babies, a moratorium was called to establish an international framework (151). The CRISPR babies event sets back the development of therapeutic genome editing research and reflects the urgent need for regulations and guidelines regarding the human germline genome editing. Several international organizations of health-care professionals try to create consensus by formulating guidelines and statements.

7. Future outlook

The discovery of CRISPR/Cas9 systems has remodeled the modern biotechnology. Diverse CRISPR platforms have been developed and applied in cultured cells, animals and plants, accelerating the pace of both basic and translational research, and enabling therapeutic and agricultural breakthroughs.

The CRISPR technology is developing rapidly, especially in the field of stem cell and gene therapy. Multiple CRISPR-mediated clinical trials are ongoing or beginning soon, and the results of these trials will guide the application for *in vivo* and *ex vivo* stem cell editing in patients. Although certain challenges of CRISPR technology such as innate immune response in human, undesired editing by off-target effects are still present, advancement of the CRISPR technology may provide new solutions.

Current CRISPR technology has some limitations and potential pitfalls, such as DNA damage, immunotoxicity, PAM requirement, off-target effects, delivery methods and the incidence of HDR. Further investigations are needed to judge the eventual impact of CRISPR/Cas9 adequately. One important direction of future CRISPR research will be how to use CRISPR technology wisely and safely, including improving editing efficiency and reducing off-target effects, exploiting novel delivery strategies and monitoring long-term effect. Scientists and researchers must comply with the current laws and regulations to design preclinical genome editing trials, establish proper criteria for patient recruitment and management, thus ensuring legal and safe experiments. As CRISPR technology is in its infancy, comprehensive investigation and in-depth research are required to better understand this novel technology. Deliberative and global discussions are also needed to reconsider the existing ethical guidelines and regulations to develop more international criteria and laws for CRISPR-mediated genome editing.

References

1. Makarova KS, Wolf YI, Iranzo J, Shmakov SA, Alkhnbashi OS, Brouns SJJ, et al. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. Nature Reviews Microbiology. 2020;18(2):67-83.
2. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science. 2012;337(6096):816-21.
3. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex Genome Engineering Using CRISPR/Cas Systems. Science. 2013;339(6121):819-23.
4. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, et al. RNA-Guided Human Genome Engineering via Cas9. Science. 2013;339(6121):823-6.
5. Sternberg SH, LaFrance B, Kaplan M, Doudna JA. Conformational control of DNA target cleavage by CRISPR-Cas9. Nature. 2015;527(7576):110-3.
6. Sander JD, Joung JK. CRISPR-Cas systems for editing, regulating and targeting genomes. Nat Biotechnol. 2014;32(4):347-55.
7. Haber JE. Partners and pathways: repairing a double-strand break. Trends in Genetics. 2000;16(6):259-64.
8. Lieber MR. The mechanism of human nonhomologous DNA end joining. J Biol Chem. 2008;283(1):1-5.
9. Thompson LH, Schild D. Homologous recombinational repair of DNA ensures mammalian chromosome stability. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis. 2001;477(1):131-53.
10. Mao Z, Bozzella M, Seluanov A, Gorbunova V. DNA repair by nonhomologous end joining and homologous recombination during cell cycle in human cells. Cell Cycle. 2008;7(18):2902-6.
11. Mao Z, Bozzella M, Seluanov A, Gorbunova V. Comparison of nonhomologous end joining and homologous recombination in human cells. DNA Repair (Amst). 2008;7(10):1765-71.
12. Savić N, Schwank G. Advances in therapeutic CRISPR/Cas9 genome editing. Translational Research. 2016;168:15-21.
13. Maeder ML, Thibodeau-Beganny S, Osiak A, Wright DA, Anthony RM, Eichtinger M, et al. Rapid “open-source” engineering of customized zinc-finger nucleases for highly efficient gene modification. Molecular cell. 2008;31(2):294-301.
14. Juillerat A, Dubois G, Valton J, Thomas S, Stella S, Maréchal A, et al. Comprehensive analysis of the specificity of transcription activator-like effector nucleases. Nucleic acids research. 2014;42(8):5390-402.
15. Chu VT, Weber T, Wefers B, Wurst W, Sander S, Rajewsky K, et al. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. Nat Biotechnol. 2015;33(5):543-8.
16. Lin S, Staahl BT, Alla RK, Doudna JA. Enhanced homology-directed human genome

engineering by controlled timing of CRISPR/Cas9 delivery. Elife. 2014;3:e04766.
17. Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS, Kriz AJ, et al. In vivo genome editing using Staphylococcus aureus Cas9. Nature. 2015;520(7546):186-91.
18. Kleinstiver BP, Prew MS, Tsai SQ, Nguyen NT, Topkar VV, Zheng Z, et al. Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. Nat Biotechnol. 2015;33(12):1293-8.
19. Hou Z, Zhang Y, Propson NE, Howden SE, Chu L-F, Sontheimer EJ, et al. Efficient genome engineering in human pluripotent stem cells using Cas9 from *Neisseria meningitidis*. Proceedings of the National Academy of Sciences. 2013;110(39):15644-9.
20. Gasiunas G, Barrangou R, Horvath P, Siksnys V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc Natl Acad Sci U S A. 2012;109(39):E2579-86.
21. Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell. 2015;163(3):759-71.
22. Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nature biotechnology. 2013;31(9):822-6.
23. Kleinstiver BP, Pattanayak V, Prew MS, Tsai SQ, Nguyen NT, Zheng Z, et al. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. Nature. 2016;529(7587):490-5.
24. Slaymaker IM, Gao L, Zetsche B, Scott DA, Yan WX, Zhang F. Rationally engineered Cas9 nucleases with improved specificity. Science. 2016;351(6268):84-8.
25. Chen JS, Dagdas YS, Kleinstiver BP, Welch MM, Sousa AA, Harrington LB, et al. Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. Nature. 2017;550(7676):407-10.
26. Hu JH, Miller SM, Geurts MH, Tang W, Chen L, Sun N, et al. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. Nature. 2018;556(7699):57-63.
27. Vakulskas CA, Dever DP, Rettig GR, Turk R, Jacobi AM, Collingwood MA, et al. A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells. Nat Med. 2018;24(8):1216-24.
28. Lee JK, Jeong E, Lee J, Jung M, Shin E, Kim YH, et al. Directed evolution of CRISPR-Cas9 to increase its specificity. Nat Commun. 2018;9(1):3048.
29. Ran FA, Hsu PD, Lin CY, Gootenberg JS, Konermann S, Trevino AE, et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell. 2013;154(6):1380-9.
30. Dianov GL, Hubscher U. Mammalian base excision repair: the forgotten archangel. Nucleic Acids Res. 2013;41(6):3483-90.

31. Guilinger JP, Thompson DB, Liu DR. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. Nat Biotechnol. 2014;32(6):577-82.
32. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell. 2013;152(5):1173-83.
33. Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell. 2013;154(2):442-51.
34. Bikard D, Jiang W, Samai P, Hochschild A, Zhang F, Marraffini LA. Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. Nucleic Acids Res. 2013;41(15):7429-37.
35. Gilbert LA, Horlbeck MA, Adamson B, Villalta JE, Chen Y, Whitehead EH, et al. Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell. 2014;159(3):647-61.
36. Maeder ML, Linder SJ, Cascio VM, Fu Y, Ho QH, Joung JK. CRISPR RNA-guided activation of endogenous human genes. Nat Methods. 2013;10(10):977-9.
37. Perez-Pinera P, Kocak DD, Vockley CM, Adler AF, Kabadi AM, Polstein LR, et al. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. Nat Methods. 2013;10(10):973-6.
38. Cheng AW, Wang H, Yang H, Shi L, Katz Y, Theunissen TW, et al. Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. Cell Res. 2013;23(10):1163-71.
39. Chavez A, Scheiman J, Vora S, Pruitt BW, Tuttle M, E PRI, et al. Highly efficient Cas9-mediated transcriptional programming. Nat Methods. 2015;12(4):326-8.
40. Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, et al. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature. 2015;517(7536):583-8.
41. Zalatan JG, Lee ME, Almeida R, Gilbert LA, Whitehead EH, La Russa M, et al. Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. Cell. 2015;160(1-2):339-50.
42. Chavez A, Tuttle M, Pruitt BW, Ewen-Campen B, Chari R, Ter-Ovanesyan D, et al. Comparison of Cas9 activators in multiple species. Nat Methods. 2016;13(7):563-7.
43. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature. 2016;533(7603):420-4.
44. Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, Bryson DI, et al. Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. Nature. 2017;551(7681):464-71.
45. Rees HA, Liu DR. Base editing: precision chemistry on the genome and transcriptome of living cells. Nature Reviews Genetics. 2018;19(12):770-88.

46. Zuo E, Sun Y, Yuan T, He B, Zhou C, Ying W, et al. A rationally engineered cytosine base editor retains high on-target activity while reducing both DNA and RNA off-target effects. Nat Methods. 2020.
47. Grunewald J, Zhou R, Garcia SP, Iyer S, Lareau CA, Aryee MJ, et al. Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. Nature. 2019;569(7756):433-7.
48. Liang P, Xie X, Zhi S, Sun H, Zhang X, Chen Y, et al. Genome-wide profiling of adenine base editor specificity by EndoV-seq. Nat Commun. 2019;10(1):67.
49. Komor AC, Zhao KT, Packer MS, Gaudelli NM, Waterbury AL, Koblan LW, et al. Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. Science Advances. 2017;3(8):eaao4774.
50. Zhang W, Aida T, Del Rosario RCH, Wilde JJ, Ding C, Zhang X, et al. Multiplex precise base editing in cynomolgus monkeys. Nat Commun. 2020;11(1):2325.
51. Zhou S, Cai B, He C, Wang Y, Ding Q, Liu J, et al. Programmable Base Editing of the Sheep Genome Revealed No Genome-Wide Off-Target Mutations. Front Genet. 2019;10:215.
52. Xie J, Ge W, Li N, Liu Q, Chen F, Yang X, et al. Efficient base editing for multiple genes and loci in pigs using base editors. Nat Commun. 2019;10(1):2852.
53. Anzalone AV, Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM, et al. Search-and-replace genome editing without double-strand breaks or donor DNA. Nature. 2019;576(7785):149-57.
54. Yan J, Cirincione A, Adamson B. Prime Editing: Precision Genome Editing by Reverse Transcription. Mol Cell. 2020;77(2):210-2.
55. Strecker J, Ladha A, Gardner Z, Schmid-Burgk JL, Makarova KS, Koonin EV, et al. RNA-guided DNA insertion with CRISPR-associated transposases. Science. 2019;365(6448):48-53.
56. Klompe SE, Vo PLH, Halpin-Healy TS, Sternberg SH. Transposon-encoded CRISPR-Cas systems direct RNA-guided DNA integration. Nature. 2019;571(7764):219-25.
57. Hughes TS, Langer SJ, Virtanen SI, Chavez RA, Watkins LR, Milligan ED, et al. Immunogenicity of intrathecal plasmid gene delivery: cytokine release and effects on transgene expression. The Journal of Gene Medicine. 2009;11(9):782-90.
58. Shen B, Zhang W, Zhang J, Zhou J, Wang J, Chen L, et al. Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. Nat Methods. 2014;11(4):399-402.
59. Chang N, Sun C, Gao L, Zhu D, Xu X, Zhu X, et al. Genome editing with RNA-guided Cas9 nuclease in zebrafish embryos. Cell Res. 2013;23(4):465-72.
60. Wagenblast E, Azkanaz M, Smith SA, Shakib L, McLeod JL, Krivdova G, et al. Functional profiling of single CRISPR/Cas9-edited human long-term hematopoietic stem cells. Nat Commun. 2019;10(1):4730.

61. Xu X, Gao D, Wang P, Chen J, Ruan J, Xu J, et al. Efficient homology-directed gene editing by CRISPR/Cas9 in human stem and primary cells using tube electroporation. Sci Rep. 2018;8(1):11649.
62. Yin H, Kauffman KJ, Anderson DG. Delivery technologies for genome editing. Nat Rev Drug Discov. 2017;16(6):387-99.
63. Chen X, Goncalves MA. Engineered Viruses as Genome Editing Devices. Mol Ther. 2016;24(3):447-57.
64. Wanisch K, Yanez-Munoz RJ. Integration-deficient lentiviral vectors: a slow coming of age. Mol Ther. 2009;17(8):1316-32.
65. Matrai J, Chuah MK, VandenDriessche T. Recent advances in lentiviral vector development and applications. Mol Ther. 2010;18(3):477-90.
66. Ortinski PI, O'Donovan B, Dong X, Kantor B. Integrase-Deficient Lentiviral Vector as an All-in-One Platform for Highly Efficient CRISPR/Cas9-Mediated Gene Editing. Mol Ther Methods Clin Dev. 2017;5:153-64.
67. Olgasi C, Talmon M, Merlin S, Cucci A, Richaud-Patin Y, Ranaldo G, et al. Patient-Specific iPSC-Derived Endothelial Cells Provide Long-Term Phenotypic Correction of Hemophilia A. Stem Cell Reports. 2018;11(6):1391-406.
68. Ding Q, Strong A, Patel KM, Ng SL, Gosis BS, Regan SN, et al. Permanent alteration of PCSK9 with in vivo CRISPR-Cas9 genome editing. Circ Res. 2014;115(5):488-92.
69. Xu L, Park KH, Zhao L, Xu J, El Refaey M, Gao Y, et al. CRISPR-mediated Genome Editing Restores Dystrophin Expression and Function in mdx Mice. Mol Ther. 2016;24(3):564-9.
70. Raper SE, Chirmule N, Lee FS, Wivel NA, Bagg A, Gao GP, et al. Fatal systemic inflammatory response syndrome in a ornithine transcarbamylase deficient patient following adenoviral gene transfer. Mol Genet Metab. 2003;80(1-2):148-58.
71. Wang D, Tai PWL, Gao G. Adeno-associated virus vector as a platform for gene therapy delivery. Nat Rev Drug Discov. 2019;18(5):358-78.
72. Wang D, Zhang F, Gao G. CRISPR-Based Therapeutic Genome Editing: Strategies and In Vivo Delivery by AAV Vectors. Cell. 2020;181(1):136-50.
73. Zhang Y, Li H, Min Y-L, Sanchez-Ortiz E, Huang J, Mireault AA, et al. Enhanced CRISPR-Cas9 correction of Duchenne muscular dystrophy in mice by a self-complementary AAV delivery system. Science advances. 2020;6(8):eaay6812.
74. Raveux A, Vandormael-Pournin S, Cohen-Tannoudji M. Optimization of the production of knock-in alleles by CRISPR/Cas9 microinjection into the mouse zygote. Sci Rep. 2017;7:42661.
75. Crispo M, Mulet AP, Tesson L, Barrera N, Cuadro F, dos Santos-Neto PC, et al. Efficient Generation of Myostatin Knock-Out Sheep Using CRISPR/Cas9 Technology and Microinjection into Zygotes. PLoS One. 2015;10(8):e0136690.
76. Hai T, Teng F, Guo R, Li W, Zhou Q. One-step generation of knockout pigs by zygote injection of CRISPR/Cas system. Cell Res. 2014;24(3):372-5.
77. Xu W. Microinjection and micromanipulation: a historical perspective. Microinjection:

Springer; 2019. p. 1-16.
78. Hung KL, Meitlis I, Hale M, Chen CY, Singh S, Jackson SW, et al. Engineering Protein-Secreting Plasma Cells by Homology-Directed Repair in Primary Human B Cells. Mol Ther. 2018;26(2):456-67.
79. Chen S, Lee B, Lee AY, Modzelewski AJ, He L. Highly Efficient Mouse Genome Editing by CRISPR Ribonucleoprotein Electroporation of Zygotes. J Biol Chem. 2016;291(28):14457-67.
80. Yin H, Kanasty RL, Eltoukhy AA, Vegas AJ, Dorkin JR, Anderson DG. Non-viral vectors for gene-based therapy. Nat Rev Genet. 2014;15(8):541-55.
81. Zuris JA, Thompson DB, Shu Y, Guilinger JP, Bessen JL, Hu JH, et al. Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. Nat Biotechnol. 2015;33(1):73-80.
82. Mout R, Ray M, Yesilbag Tonga G, Lee YW, Tay T, Sasaki K, et al. Direct Cytosolic Delivery of CRISPR/Cas9-Ribonucleoprotein for Efficient Gene Editing. ACS Nano. 2017;11(3):2452-8.
83. Lee K, Conboy M, Park HM, Jiang F, Kim HJ, Dewitt MA, et al. Nanoparticle delivery of Cas9 ribonucleoprotein and donor DNA in vivo induces homology-directed DNA repair. Nat Biomed Eng. 2017;1:889-901.
84. D'Astolfo DS, Pagliero RJ, Pras A, Karthaus WR, Clevers H, Prasad V, et al. Efficient intracellular delivery of native proteins. Cell. 2015;161(3):674-90.
85. Zhang XH, Tee LY, Wang XG, Huang QS, Yang SH. Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. Mol Ther Nucleic Acids. 2015;4:e264.
86. Wang DC, Wang X. Off-target genome editing: A new discipline of gene science and a new class of medicine. Cell Biol Toxicol. 2019;35(3):179-83.
87. Wang X, Liu J, Niu Y, Li Y, Zhou S, Li C, et al. Low incidence of SNVs and indels in trio genomes of Cas9-mediated multiplex edited sheep. BMC Genomics. 2018;19(1):397.
88. Li C, Zhou S, Li Y, Li G, Ding Y, Li L, et al. Trio-Based Deep Sequencing Reveals a Low Incidence of Off-Target Mutations in the Offspring of Genetically Edited Goats. Front Genet. 2018;9:449.
89. Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. Nature biotechnology. 2013;31(9):827.
90. Mali P, Aach J, Stranges PB, Esvelt KM, Moosburner M, Kosuri S, et al. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nature biotechnology. 2013;31(9):833.
91. Pattanayak V, Lin S, Guilinger JP, Ma E, Doudna JA, Liu DR. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. Nature biotechnology. 2013;31(9):839.
92. Xiao-Jie GEL, Li-Juan J, Ishii T. Germ line genome editing in clinics: the approaches, objectives and global society. Briefings in Functional Genomics. 2015;16(1):46-56.

93. Stemmer M, Thumberger T, del Sol Keyer M, Wittbrodt J, Mateo JL. CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. PLOS ONE. 2015;10(4):e0124633.
94. Bae S, Park J, Kim J-S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. Bioinformatics. 2014;30(10):1473-5.
95. Cradick TJ, Qiu P, Lee CM, Fine EJ, Bao G. COSMID: a web-based tool for identifying and validating CRISPR/Cas off-target sites. Molecular Therapy-Nucleic Acids. 2014;3:e214.
96. Tsai SQ, Zheng Z, Nguyen NT, Liebers M, Topkar VV, Thapar V, et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nat Biotechnol. 2015;33(2):187-97.
97. Kim D, Bae S, Park J, Kim E, Kim S, Yu HR, et al. Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. Nat Methods. 2015;12(3):237-43, 1 p following 43.
98. Tsai SQ, Nguyen NT, Malagon-Lopez J, Topkar VV, Aryee MJ, Joung JK. CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets. Nat Methods. 2017;14(6):607-14.
99. Tang P-Z, Ding B, Peng L, Mozhayskiy V, Potter J, Chesnut JD. TEG-seq: an ion torrent-adapted NGS workflow for in cellulo mapping of CRISPR specificity. BioTechniques. 2018;65(5):259-67.
100. Nobles CL, Reddy S, Salas-McKee J, Liu X, June CH, Melenhorst JJ, et al. iGUIDE: an improved pipeline for analyzing CRISPR cleavage specificity. Genome Biol. 2019;20(1):14.
101. Kim D, Kim JS. DIG-seq: a genome-wide CRISPR off-target profiling method using chromatin DNA. Genome Res. 2018;28(12):1894-900.
102. Sato Y, Bando H, Di Piazza M, Gowing G, Herberts C, Jackman S, et al. Tumorigenicity assessment of cell therapy products: The need for global consensus and points to consider. Cytotherapy. 2019;21(11):1095-111.
103. Ito E, Miyagawa S, Takeda M, Kawamura A, Harada A, Iseoka H, et al. Tumorigenicity assay essential for facilitating safety studies of hiPSC-derived cardiomyocytes for clinical application. Scientific Reports. 2019;9(1):1881.
104. Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nat Biotechnol. 2014;32(3):279-84.
105. Fu Y, Reyon D, Keith Joung J. Chapter Two - Targeted Genome Editing in Human Cells Using CRISPR/Cas Nucleases and Truncated Guide RNAs. In: Doudna JA, Sontheimer EJ, editors. Methods in Enzymology. 546: Academic Press; 2014. p. 21-45.
106. Cromwell CR, Sung K, Park J, Krysler AR, Jovel J, Kim SK, et al. Incorporation of bridged nucleic acids into CRISPR RNAs improves Cas9 endonuclease specificity. Nat Commun. 2018;9(1):1448.

107. Vakulskas CA, Behlke MA. Evaluation and Reduction of CRISPR Off-Target Cleavage Events. Nucleic Acid Ther. 2019;29(4):167-74.
108. Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. Cell. 2014;157(6):1262-78.
109. Shahryari A, Saghaeian Jazi M, Mohammadi S, Razavi Nikoo H, Nazari Z, Hosseini ES, et al. Development and Clinical Translation of Approved Gene Therapy Products for Genetic Disorders. Front Genet. 2019;10:868.
110. Piepenburg O, Williams CH, Stemple DL, Armes NA. DNA detection using recombination proteins. PLoS Biol. 2006;4(7):e204.
111. Kellner MJ, Koob JG, Gootenberg JS, Abudayyeh OO, Zhang F. SHERLOCK: nucleic acid detection with CRISPR nucleases. Nat Protoc. 2019;14(10):2986-3012.
112. Broughton JP, Deng X, Yu G, Fasching CL, Servellita V, Singh J, et al. CRISPR-Cas12-based detection of SARS-CoV-2. Nat Biotechnol. 2020.
113. Notomi T, Okayama H, Masubuchi H, Yonekawa T, Watanabe K, Amino N, et al. Loop-mediated isothermal amplification of DNA. Nucleic acids research. 2000;28(12):e63-e.
114. Ackerman CM, Myhrvold C, Thakku SG, Freije CA, Metsky HC, Yang DK, et al. Massively multiplexed nucleic acid detection with Cas13. Nature. 2020.
115. Shimo T, Hosoki K, Nakatsuji Y, Yokota T, Obika S. A novel human muscle cell model of Duchenne muscular dystrophy created by CRISPR/Cas9 and evaluation of antisense-mediated exon skipping. Journal of Human Genetics. 2018;63(3):365-75.
116. Fan Z, Perisse IV, Cotton CU, Regouski M, Meng Q, Domb C, et al. A sheep model of cystic fibrosis generated by CRISPR/Cas9 disruption of the CFTR gene. JCI insight. 2018;3(19).
117. Roh J-i, Lee J, Park SU, Kang Y-S, Lee J, Oh A-R, et al. CRISPR-Cas9-mediated generation of obese and diabetic mouse models. Experimental animals. 2018:17-0123.
118. Jarrett KE, Lee C, De Giorgi M, Hurley A, Gillard BK, Doerfler AM, et al. Somatic editing of Ldlr with adeno-associated viral-CRISPR is an efficient tool for atherosclerosis research. Arteriosclerosis, thrombosis, and vascular biology. 2018;38(9):1997-2006.
119. Yang W, Li S, Li X-J. A CRISPR monkey model unravels a unique function of PINK1 in primate brains. Molecular neurodegeneration. 2019;14(1):1-4.
120. Schwank G, Koo BK, Sasselli V, Dekkers JF, Heo I, Demircan T, et al. Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. Cell Stem Cell. 2013;13(6):653-8.
121. Li T, Pires C, Nielsen TT, Waldemar G, Hjermind LE, Nielsen JE, et al. Generation of induced pluripotent stem cells (iPSCs) from an Alzheimer's disease patient carrying an A79V mutation in PSEN1. Stem Cell Res. 2016;16(2):229-32.
122. Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. Benefits and limitations of genome-wide association studies. Nature Reviews Genetics. 2019;20(8):467-84.

123. Stojic L, Lun AT L, Mangei J, Mascalchi P, Quarantotti V, Barr AR, et al. Specificity of RNAi, LNA and CRISPRi as loss-of-function methods in transcriptional analysis. Nucleic Acids Research. 2018;46(12):5950-66.
124. Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen TS, et al. Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. Science. 2014;343(6166):84-7.
125. Koike-Yusa H, Li Y, Tan EP, Velasco-Herrera MDC, Yusa K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. Nature Biotechnology. 2014;32(3):267-73.
126. Zhou Y, Zhu S, Cai C, Yuan P, Li C, Huang Y, et al. High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. Nature. 2014;509(7501):487-91.
127. Evers B, Jastrzebski K, Heijmans JPM, Grernrum W, Beijersbergen RL, Bernards R. CRISPR knockout screening outperforms shRNA and CRISPRi in identifying essential genes. Nature Biotechnology. 2016;34(6):631-3.
128. Galonska C, Charlton J, Mattei AL, Donaghey J, Clement K, Gu H, et al. Genome-wide tracking of dCas9-methyltransferase footprints. Nature Communications. 2018;9(1):597.
129. Liu X, Zhang Y, Chen Y, Li M, Zhou F, Li K, et al. In Situ Capture of Chromatin Interactions by Biotinylated dCas9. Cell. 2017;170(5):1028-43.e19.
130. Baeumler TA, Ahmed AA, Fulga TA. Engineering Synthetic Signaling Pathways with Programmable dCas9-Based Chimeric Receptors. Cell Reports. 2017;20(11):2639-53.
131. Chen B, Gilbert LA, Cimini BA, Schnitzbauer J, Zhang W, Li G-W, et al. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell. 2013;155(7):1479-91.
132. Wu X, Mao S, Ying Y, Krueger CJ, Chen AK. Progress and Challenges for Live-cell Imaging of Genomic Loci Using CRISPR-based Platforms. Genomics, Proteomics & Bioinformatics. 2019;17(2):119-28.
133. Hong Y, Lu G, Duan J, Liu W, Zhang Y. Comparison and optimization of CRISPR/dCas9/gRNA genome-labeling systems for live cell imaging. Genome Biology. 2018;19(1):39.
134. Zhang Y, Massel K, Godwin ID, Gao C. Applications and potential of genome editing in crop improvement. Genome biology. 2018;19(1):210.
135. Li M, Li X, Zhou Z, Wu P, Fang M, Pan X, et al. Reassessment of the four yield-related genes Gn1a, DEP1, GS3, and IPA1 in rice using a CRISPR/Cas9 system. Frontiers in plant science. 2016;7:377.
136. Waltz E. CRISPR-edited crops free to enter market, skip regulation. Nature Publishing Group; 2016.
137. Nekrasov V, Wang C, Win J, Lanz C, Weigel D, Kamoun S. Rapid generation of a transgene-free powdery mildew resistant tomato by genome deletion. Scientific

reports. 2017;7(1):1-6.
138. Rodríguez-Leal D, Lemmon ZH, Man J, Bartlett ME, Lippman ZB. Engineering quantitative trait variation for crop improvement by genome editing. Cell. 2017;171(2):470-80. e8.
139. Zong Y, Song Q, Li C, Jin S, Zhang D, Wang Y, et al. Efficient C-to-T base editing in plants using a fusion of nCas9 and human APOBEC3A. Nature Biotechnology. 2018;36(10):950-3.
140. Zhang R, Liu J, Chai Z, Chen S, Bai Y, Zong Y, et al. Generation of herbicide tolerance traits and a new selectable marker in wheat using base editing. Nature plants. 2019;5(5):480-5.
141. Tian S, Jiang L, Cui X, Zhang J, Guo S, Li M, et al. Engineering herbicide-resistant watermelon variety through CRISPR/Cas9-mediated base-editing. Plant cell reports. 2018;37(9):1353-6.
142. Li C, Zong Y, Wang Y, Jin S, Zhang D, Song Q, et al. Expanded base editing in rice and wheat using a Cas9-adenosine deaminase fusion. Genome biology. 2018;19(1):59.
143. Whitworth KM, Rowland RR, Ewen CL, Trible BR, Kerrigan MA, Cino-Ozuna AG, et al. Gene-edited pigs are protected from porcine reproductive and respiratory syndrome virus. Nat Biotechnol. 2016;34(1):20-2.
144. Wang X, Niu Y, Zhou J, Yu H, Kou Q, Lei A, et al. Multiplex gene editing via CRISPR/Cas9 exhibits desirable muscle hypertrophy without detectable off-target effects in sheep. Sci Rep. 2016;6:32271.
145. Zhou W, Wan Y, Guo R, Deng M, Deng K, Wang Z, et al. Generation of beta-lactoglobulin knock-out goats using CRISPR/Cas9. PLoS One. 2017;12(10):e0186056.
146. Brokowski C, Adli M. CRISPR Ethics: Moral Considerations for Applications of a Powerful Tool. J Mol Biol. 2019;431(1):88-101.
147. Tu Z, Yang W, Yan S, Yin A, Gao J, Liu X, et al. Promoting Cas9 degradation reduces mosaic mutations in non-human primate embryos. Scientific Reports. 2017;7(1):42081.
148. Li D, Zhou H, Zeng X. Battling CRISPR-Cas9 off-target genome editing. Cell Biol Toxicol. 2019;35(5):403-6.
149. Hyun I, Wilkerson A, Johnston J. Embryology policy: Revisit the 14-day rule. Nature. 2016;533(7602):169-71.
150. Dickenson D, Darnovsky M. Did a permissive scientific culture encourage the 'CRISPR babies' experiment? Nature Biotechnology. 2019;37(4):355-7.
151. Lander ES, Baylis F, Zhang F, Charpentier E, Berg P, Bourgain C, et al. Adopt a moratorium on heritable genome editing. Nature. 2019;567(7747):165-8.

Chapter 7

Impact Paragraph

Neuromuscular disease is a large group of genetic disorders with complex symptoms and few treatment options, which places a heavy burden on society and health care. The work presented in this thesis will contribute immediately to improve and accelerate genetic diagnosis and will pave the way for future gene/stem cell therapy in these patients with the ultimate goal of precision diagnosis and personalized treatment.

WES in genetic diagnosis

Conventional genetic diagnosis of a patient in the past largely relied on sequential Sanger sequencing of known candidate genes inferred from the patient's phenotype, which is time-consuming, expensive and inefficient. On the contrary, WES enables massive parallel sequencing of the complete exome of a patient. The sequencing-by-synthesis platform used in this thesis, Illumina Hiseq 2000, is capable of efficiently sequencing up to 100 human exomes in a single run. WES was reported to solve 50% of a patient cohort with mitochondrial or mitochondrial associated disease, whereas the solving rate reduced to about 10% by Sanger sequencing in a heterogeneous patient cohort (1, 2). Moreover, with new sequencing platforms emerging, the average time and costs has rapidly declined over the last decade, making these ideal for identifying genetic defects in clinical diagnostics of heterogeneous mitochondrial disease (3). Next-generation sequencing has greatly accelerated the discovery of novel pathogenic genes in neuromuscular disease and facilitated the establishment of genotype-phenotype relationships (4). In 2013 alone, more than 180 novel disease-causing genes had been identified by WES with a broad variety of clinical manifestations (5). This can be exemplified by the identification for the first time of two heterozygous pathogenic mutations in an OXPHOS RNA homeostasis regulator *SLIRP* applying WES and functional validation. *SLIRP* turned out to be a novel gene, involved in mitochondrial encephalomyopathy and OXPHOS complex I and IV deficiency (chapter 2).

Variants detected by WES data should be analyzed with a comprehensive and structured approach to maximize the likelihood of discovering pathogenic variants. When no candidate gene is detected, matching the criteria and genetic model, a second-round examination should be performed to identify genes, in which for technical reasons or due to cut-off criteria variants have been missed (20-30% of all variants). As is shown in chapter 2, the initial WES data filtering only discovered one deletion variant in *SLIRP*. The second deep-intronic splicing variant was found by

reanalyzing the WES data in more detail based on the functional role of *SLIRP* in OXPHOS. Subsequent RNA analysis of *SLIRP* revealed an additional transcript in the patient compared to the control, thus making *SLIRP* the best candidate. Therefore, it is important to combine other approaches, like for example RNASeq as a general strategy, to prioritize the candidate list when analyzing WES data.

During WES data analysis, there might still remain a large number of variants after an initial step of allele frequencies screening in several public population databases (e.g., gnome, 1000G, EVS). Fortunately, a large variety of bioinformatic tools can be used to reduce and prioritize the candidate variants, such as cloud-based GENESIS and co-expression-based WeGET. In chapter 3, we applied these two *in silico* tools in analyzing WES data from a patient having progressive paralysis of the extraocular muscles and multiple deletions in mtDNA, and identified a homozygous deletion in *C1QBP*, encoding complement component 1 Q subcomponent-binding protein involved in mitochondrial homeostasis. GENESIS provided a comprehensive pathogenicity prediction score calculated from a set of prediction tools, which circumvented the conflicting results from each individual tool and made the prediction more reliable. WeGET evaluated the query gene set's co-expression within approximately 1000 multi-tissue datasets and ranked the variants by the co-expression level, which assessed the WES data from the co-expression perspective and complemented the other pathogenicity-based prediction methods.

Our data demonstrates that WES is a powerful tool to identify pathogenic mutations in neuromuscular disease, which not only has research values of understanding genetic causes and underlying pathophysiological mechanisms, but also helps clinical geneticists and physicians acquire accurate molecular diagnosis of patients, thus offering better health care to the patients and preventing disease transmission by prenatal and preimplantation genetic diagnosis.

Gene/stem cell therapy in neuromuscular disease

In recent years, a number of gene/stem cell therapies advanced to human clinical trial stage, some of them had even been approved and entered into the market (6). As described in this thesis, CRISPR-based gene editing and stem cells like mesoangioblasts can be developed into potential treatments for treating the dystrophic muscles in neuromuscular disease like myotonic dystrophy type 1 (DM1). In chapter 5, we designed and tested a novel *ex vivo* CRISPR/Cas9 editing approach in autologous mesoangioblasts from DM1 patients. The CRISPR/Cas9 component

is delivered in the form of ribonucleoprotein by electroporation, which can reduce the risk of off-target events and immune response. The approach turned out to be specific and efficient in skipping the transcription of CTG expansion with limited effect on the viability of the mesoangioblast. The underlying methodology is also suitable to correct the genetic defect in mesoangioblasts from patients with other neuromuscular diseases. As all methods and tools have been established in compliance with GMP, we expect to transfer this correction strategy towards clinical trials. In the future, once the safety and efficacy of CRISPR-edited mesoangioblasts is fully established, production of therapeutic mesoangioblasts can be further commercialized into a cell-based medicinal product, providing the possibility of a spin-off company with pharmaceutical and economic value.

References

1. Theunissen TEJ, Nguyen M, Kamps R, Hendrickx AT, Sallevelt S, Gottschalk RWH, et al. Whole Exome Sequencing Is the Preferred Strategy to Identify the Genetic Defect in Patients With a Probable or Possible Mitochondrial Cause. Front Genet. 2018;9:400.
2. Neveling K, Feenstra I, Gilissen C, Hoefsloot LH, Kamsteeg E-J, Mensenkamp AR, et al. A Post-Hoc Comparison of the Utility of Sanger Sequencing and Exome Sequencing for the Diagnosis of Heterogeneous Diseases. Human Mutation. 2013;34(12):1721-6.
3. Schwarze K, Buchanan J, Taylor JC, Wordsworth S. Are whole-exome and whole-genome sequencing approaches cost-effective? A systematic review of the literature. Genetics in Medicine. 2018;20(10):1122-30.
4. Boycott KM, Vanstone MR, Bulman DE, MacKenzie AE. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. Nature Reviews Genetics. 2013;14(10):681-91.
5. Zhang X. Exome sequencing greatly expedites the progressive research of Mendelian diseases. Frontiers of medicine. 2014;8(1):42-57.
6. Shahryari A, Saghaeian Jazi M, Mohammadi S, Razavi Nikoo H, Nazari Z, Hosseini ES, et al. Development and Clinical Translation of Approved Gene Therapy Products for Genetic Disorders. Front Genet. 2019;10:868.

Chapter 8

Summary

Neuromuscular diseases are a broad group of heterogeneous disorders that impair normal function of neuromuscular systems, manifesting with a broad variety of symptoms. Muscle symptoms include muscular weakness, muscle wastage, muscular cramps, muscle spasticity, muscle pain, breathing and swallowing difficulties, with different onset and severity. Causes of neuromuscular disorders range from genetic mutations, injuries, viral infections, dietary deficiencies to certain drugs and poisons. Genetic neuromuscular diseases are usually complex and heterogeneous, which can be caused by a defect (or defects) in a single gene (e.g., myotonic dystrophy type 1, DM1), or in multiple genes (e.g., mitochondrial diseases) (**chapter 1**). Such highly clinical and genetic heterogeneity creates hurdles to identify the genetic defect for many neuromuscular disorders. However, whole exome sequencing (WES) brings the power to solve this issue, and enables identifying the underlying genetic cause and characterizing the unknown pathophysiological process. Despite the successes in identifying the genetic cause successful treatments are lagging behind. However, many novel gene and cell-based therapies are being developed these days and some have successfully reversed the phenotype in animal models of neuromuscular diseases (like CRISPR/Cas9 mediated correction in DM1 mice and mesoangioblasts infusion in dystrophic dogs). So promising treatments for these so far incurable neuromuscular diseases are emerging (**chapter 1**).

Our central hypothesis was that **(**1) whole exome sequencing followed by functional validation would enable us to identify the genetic cause of neuromuscular diseases and confirm the pathogenicity of the identified variants, especially in novel genes; (2) CRISPR/Cas9-mediated correction of known genetic defects in patient-derived muscle stem cells would be a potential autologous stem cell therapy for treating the muscle pathology.

Our first aim was to identify and functionally characterize gene defects in mitochondrial diseases. In **chapter 2**, we report a Dutch boy of non-consanguineous parents, with a clinical diagnosis of mitochondrial encephalomyopathy and complex I and IV deficiency, where whole exome sequencing revealed compound heterozygosity for two variants in the *SLIRP* gene, being a deletion in exon 3 and a splicing variant in intron 1. RT-PCR analysis revealed three *SLIRP* transcripts, adding up in total to only 63% of normal transcript levels in controls. The largest transcript, present in 18%, showed retention of a 106bp fragment of intron 1, leading to a frameshift and premature stop codon. The two smaller fragments contained either the deletion (74%) or wild-type *SLIRP* (8%). SRA

stem-loop-interacting RNA-binding protein (SLIRP) is predominantly localized in mitochondria and forms a complex with LRPPRC, which is involved in regulation of mitochondrial DNA (mtDNA) gene expression. Indeed, in fibroblasts of the patient a reduced RNA level of the mtDNA genes *MT-ND1*, *MT-ND6* and *MT-CO1* was observed. Lentiviral transduction of wild-type *SLIRP* cDNA in the patient fibroblasts increased the expression of the *MT-ND1*, *MT-ND6*, *MT-CO1* genes (2.5 to 7.2-fold), whereas transduction with the mutants did not. The decreased complex I and IV enzyme activities seemed to be the result of a reduced mitochondrial mass, reflected by a fourfold decrease of citrate synthase (CS) compared to total protein in patient fibroblasts. Transduction with wild-type *SLIRP* cDNA led to a 2.4-fold increase of CS compared to total protein and an increase in complex I and IV activities. The results of the complementation assay confirmed the causality of the *SLIRP* variants and revealed the mechanism, leading to the mitochondrial encephalomyopathy in this patient. In conclusion, we report the first variants in *SLIRP* gene as a novel cause of mitochondrial encephalomyopathy with combined complex I and IV deficiency.

Next in **chapter 3**, whole exome sequencing, analyzed with GENESIS and WeGET as 2 filtering strategies, revealed a homozygous deletion in the *C1QBP* gene in a patient with progressive external ophthalmoplegia (PEO) and multiple mtDNA deletions. The gene encodes the mitochondria-located complementary 1 Q subcomponent-binding protein, involved in mitochondrial homeostasis. Interestingly, of the reported eight patients with *C1QBP* variants, four were diagnosed with cardiomyopathy, but without involvement of the central nervous system. Our patient showed only isolated late-onset PEO without overt cardiac involvement. So, biallelic mutations in *C1QBP* cause mitochondrial cardiomyopathy and/or PEO with variable age of onset. Available data suggest that early-onset cardiomyopathy variants localize in important structural domains and PEO-only variants in the coiled-coil region. Anyhow, *C1QBP* mutations should be considered in PEO patients, regardless of cardiac phenotype.

Our second aim was to develop a generic autologous stem cell therapy to treat muscle pathology in neuromuscular disorders. Mesoangioblasts, which are myogenic stem cells, seem to be particularly suited for clinical application. Benefits of using mesoangioblasts as a somatic cell therapy product include abundantly availability from tissues, easy expansion to large numbers *in vitro* without losing proliferation potency, homing to skeletal muscle through the blood stream, allowing systemic delivery and efficient differentiation into skeletal muscle *in vivo*. Furthermore, mesoangioblasts were shown to ameliorate muscle function in

dystrophic dogs and mice.

In **chapter 4**, we analyzed mesoangioblasts from 30 clinically well-characterized DM1 patients. Recent studies demonstrated mitochondrial dysfunction in DM1 fibroblasts and peripheral blood mononuclear cells (PBMCs). However, little is known about the pathophysiology and mitochondrial function in myogenic progenitor cells of DM1 patients. Therefore, in this study, we characterized cellular hallmarks, mitochondrial content and metabolism in DM1 mesoangioblasts. Our results revealed that DM1 mesoangioblasts displayed typical DM1 hallmarks, including expanded CTG repeat and downstream splicing defects. The expanded CTG repeat size in DM1 mesoangioblasts positively correlated to the degree of muscle impairment while negatively to the age of onset. No clear correlation was observed between expanded CTG repeat size, mitochondrial content and mitochondrial respiration, in contrast, most DM1 mesoangioblasts displayed normal metabolism and mitochondrial function. Our results implicate that variation in cellular DM1 phenotype can exist among different cell types and addressed the importance of cell type specific research in DM1 pathology, thereby focusing on the relation of DM1 hallmarks and stemness of cells.

In **chapter 5**, we explored a novel CRISPR/Cas9-mediated genome editing approach in mesoangioblasts derived from six DM1 patients. Our approach was to abolish the splice-acceptor site of intron 14 in mesoangioblasts derived from DM1 patients, using CRISPR/Cas9-based Non-Homologous End Joining (NHEJ), in order to induce alternative splicing and exclude the CTG-repeat from the 3' UTR of the DMPK-RNA. We achieved a total editing efficiency of 36-51%, based on next-generation sequencing of the on-target site, and an increased expression of the alternative exon 16 downstream of the CTG repeats in six edited DM1 mesoangioblasts. In 9-27%, the expanded allele was corrected, which showed that our strategy was able to remove the expanded repeat from mesoangioblasts in sufficient numbers for eventual treatment, although it was still insufficient to see reversal of the cellular phenotype in the mosaic cell population. Currently, purification strategies are being explored to extract the correctly edited mesoangioblasts from the overall pool of edited and non-edited mesoangioblasts. Notably, next-generation sequencing of the ten most-likely off-target sites of the single guide RNA (sgRNA) revealed almost no mutations at those sites. This study demonstrates that CRISPR/Cas9 ribonucleoprotein-mediated genome edited DM1 mesoangioblasts provide a novel therapeutic strategy for DM1.

Chapter 6 reviews the state-of-art CRISPR/Cas9 technology, as it is one of the most important novel technologies in science, rewarded with the Nobel prize for its inventors Emmanuelle Charpentier and Jennifer A. Doudna in 2020. We give an overview of different CRISPR/Cas9 platforms, delivery formats and methods. We summarize the factors contributing to off-target events and approaches to avoid these undesired off-target events. We discuss the applications of CRISPR/Cas9 in different scientific fields, with a focus on therapeutic applications, as well as current ethic and regulation issues and a glimpse of future development.

Appendices

Curriculum Vitae

List of publications

Acknowledgements

Curriculum Vitae

Le Guo was born on 7th of November, 1991, in Jingzhou, Hubei province, China. In 2013, he received his Bachelor's degree in Biotechnology and teaching certificate in Biology at Shanxi Normal University in Linfen, China. He won an excellent thesis award for his Bachelor thesis titled "Selection of Extraction Methods Suitable for SDS-PAGE Analysis of Arabidopsis thaliana Chloroplast Proteins" in 2013. After graduation, he continued his studies at State Key Laboratory of Molecular Vaccinology and Molecular Diagnostics, Xiamen University in Xiamen, China and obtained his Master in Biochemistry and Molecular Biology in 2016. With great study achievements, he got second-class scholarship in Xiamen University during his master period (2013-2016). His Master project was about "Joint Detection of Mutations in Hyperphenylalaninemia". In September 2016, Le started his PhD project under supervision of Pro. Dr. Smeets Hubert J.M. (promotor), Dr. Gerards Mike and Dr. van Tienen Florence (co-promotors) at the Department of Toxicogenomics, Maastricht University, accomplishing this thesis. The topic of his research was to identify novel genetic causes and to develop therapeutic strategies in neuromuscular diseases. During this period, he was supported by the Chinese Scholarship Council (2016.09-2020.09).

List of publications

Le Guo, Bob P.H. Engelen, Irenaeus F.M. de Coo, Maaike Vreeburg, Suzanne C.E.H. Sallevelt, Debby M.E.I. Hellebrekers, Ed H. Jacobs, Farah Sadeghi-Niaraki, Florence van Tienen, Mike Gerards* and Hubert J.M. Smeets*. Pathogenic *SLIRP* variants as a novel cause of mitochondrial encephalomyopathy with complex I and complex IV deficiency. Submitted to European Journal of Human Genetics (IF=3.657). Under major revision.

Le Guo, Periyasamy Govindaraj, Mariëlle Kievit, Irenaeus F.M. de Coo, Mike Gerards, Debby M.E.I. Hellebrekers, Alphons P.M. Stassen, Narayanappa Gayathri, Arun B Taly, Bindu Parayil Sankaran*, Hubert J.M. Smeets*. Homozygous C1QBP deletion in a patient with progressive external ophthalmoplegia and mtDNA deletions. Submitted to Neuromuscular Disorders (IF=3.115). Under major revision.

Le Guo#, Ruby Zelissen#, Erika Timmer, Wendy Nguyen, Isis Joosten, Carla Gorissen, Karin Faber, Florence van Tienen*, Hubert J.M. Smeets* Characterization of cellular and mitochondrial parameters in mesoangioblasts of myotonic dystrophy type 1 patients. Manuscript in preparation.

Le Guo, Erika Timmer, Isis Joosten, Rick Kamps, Patrick Lindsey, Florian Caiment, Carla Gorissen, Karin Faber, Florence van Tienen*, Hubert J.M. Smeets*. CRISPR/Cas9 corrected mesoangioblasts as potential therapeutic strategy for Myotonic Dystrophy Type 1. Manuscript in preparation.

Oral presentations

- September 2020, Department of Toxicogenomics, Maastricht University, lab presentation: Identification, characterization and treatment of nuclear gene mutations in neuromuscular diseases.
- January 2021, Department of Toxicogenomics, Maastricht University, lab presentation: CRISPR/Cas9-mediated correction of the Myotonic Dystrophy Type 1 (DM1) mutation in mesoangioblasts (MABs).

Poster presentations

- November 2018, Annual Research Day, School for Mental Health & Neuroscience, Hotel Van der Valk, Maastricht, poster presentation: Towards genetic correction of Myotonic Dystrophy Type 1 mutation in mesoangioblasts for autologous cell therapy.
- September 2019, EURON PhD Symposium, University of Luxembourg, poster presentation: Towards genetic correction of Myotonic Dystrophy Type 1 mutation in mesoangioblasts for autologous cell therapy.
- September 2020, Annual Research Day, School for Mental Health & Neuroscience, Maastricht University, poster presentation: Pathogenic *SLIRP* variants as a novel cause of mitochondrial encephalomyopathy with complex I and complex IV deficiency.

Teaching

- Supervision of Biomedical Master student, Junior Internship (2017)
- Supervision of Biomedical Bachelor student, Junior Internship (2018)
- Tutoring a Master course, Biomedical Challenges (2019/2020)
- Supervision of Biomedical Bachelor student, Junior Internship (2020/2021)

Acknowledgements

It feels like yesterday when I took my step out of the crowded Schiphol airport and into Maastricht train station. Time flies but the good memories last. It is a great experience to have my PhD study at Maastricht University, and I love the motto “leading in learning”. During my PhD study, it was so lucky to meet and work with so many nice people, and I would like to acknowledge all of you.

First, I would like to express deepest gratitude to my supervision team. **Bert**, my promotor, thank you for giving me such a great opportunity to carry out my PhD research in Clinical Genomics Lab. Your great sense of humor has definitely made my study and life interesting. Your kind support and constant guidance have helped me through many difficult times. Your wise advice and suggestions have motivated me towards the finishing line of my PhD. Also thank you for putting great efforts to revise this thesis and other manuscripts for publication.

Mike, my co-promotor, thank you so much for your generous support and constant supervision. Your positive and optimistic attitude has continuously encouraged me through the dark. Thank you for your immense help when experiments got stuck, for your broad knowledge and skills during our routine project discussions, all of which are of great value for me to learn new things and complete my PhD project.

My sincere gratitude to **Florence**, my co-promotor, who always have intelligent ideas and solutions to every problem. I would appreciate your timely supervision and useful suggestions. You are very careful and thoughtful in the experimental design and data analysis, which have set a really good example to me on how to become an outstanding researcher.

René, I am so grateful for your professional knowledge in neurology, for your contributive opinions towards neuromuscular diseases and your detailed explanation of complex neurological items, and I have learned a lot from you.

Debby, many thanks for your selfless help in obtaining clinical materials and information, searching for similar cases in the database and commenting on possible candidate variants.

Special thanks to **Rick**, for providing me with constructive advice and ample knowledge about next-generation sequencing. I also gratefully acknowledge the

tremendous help with finalizing the sequencing work during my last period of PhD.

Next, I would like to thank my lovely colleagues. **Ruby**, you are really a nice colleague and have a good personality. I am amazed that you are talented in so many aspects, like cooking, painting, weaving and gardening. I am pretty sure you can start a company from one of your hobbies. Thank you for your delicious food and the pleasant dinner with **Milena** and **Somaieh** at your house. Hope you and **Tom** have a happy life forever.

Milena, thank you for your kind help and patience whenever I ask you for extracting sequencing data from the server. I still remember that we joined the PhD day together and talked a lot about study and life. Hope you have a wonderful wedding with your boyfriend in the near future and good luck with finalizing your thesis.

Somaieh, it is a great pleasure to know you and thanks for your generous help with confocal imaging. Also, thanks for your kind suggestions and funny video clips. I still remember you are the one who led me to the lunch room asking me a favor, but turned out to be a surprise celebrating my wedding with other colleagues. Good acting. Hope your lovely kids stay healthy and grow up happily.

Toni, my dear friend and colleague, it is great to have you as my roommate. Thanks for the kind invitation to visit your hometown and meet your family. You are absolutely a skilled driver and thank you for showing me the beautiful scenery, also for the delicious Spanish food. It was really an unforgettable trip and I enjoyed it a lot. Hope you have a fruitful PhD life.

Sincere gratitude to **Erika**, **Bieke** and **Iris**, who are always cheerful and helpful during my stay in the lab. Thank you for all your care and concern about me, and I also really appreciate your ordering all sorts of primers and reagents for me. All the best in your work.

Many thanks to **Diana** and **Christa**, our lab secretary, for planning all the meeting and solving all the practical issues. I hope everything goes well with your work.

Furthermore, I am much grateful to past and present group members for your kind help, valuable collaboration and pleasant conversations during my entire PhD period, **Bob, Jo**, **Auke**, **Ivo**, **Tom**, **Minh**, **Evelien**, **Ellen**, **Barbie**, **Janine**, **Markos**, **Michiel**. Hope you have a bright future.

Thanks to my dear friends, **Shujin**, **Aomin**, **Ying**, **Yiyong**, **Qian, Lingling, Li-yen, Zhuojun**. I will miss you and let us keep in touch. I would also like to acknowledge **China Scholarship Council** for the financial support for my entire PhD study in Maastricht University.

Last but not the least, I would thank my wife **Qiannan** and my parents for your unlimited love and support, for your warm encouragement, for understanding me and taking care my little daughter **Chuxuan**. And I will always stay strong for you.