Dissertation

Identification of two novel disease-causing mutations in pleiotropic genes resulting in unexpected phenotypes due to loss-of-function escape mechanisms

submitted by

Lukas KAUFMANN, BSc MSc

for the Academic Degree of

Doctor of Medical Science (Dr. scient. med.)

at the

Medical University of Graz

Diagnostic & Research Institute of Human Genetics

under the Supervision of

Assoz. Prof. Priv.-Doz. Mag. Dr.rer.nat.

Christian WINDPASSINGER

2024

STATUTORY DECLARATION

I hereby declare that this dissertation is my own original work and that I have fully acknowledged by name all of those individuals and organizations that have contributed to the research for this dissertation. Due acknowledgement has been made in the text to all other material used. Throughout this dissertation and in all related publications I followed the "Guidelines of the Medical University of Graz on Good Scientific Practice".

Lukas Kaufmann, May 2024

DISCLOSURES

Part of this thesis has been published in:

Kaufmann L, Pilic J, Auinger L, Mayer AL, Blatterer J, Semmler-Bruckner J, Abbas S, Rehman K, Ayaz M, Graier WF, Malli R, Petek E, Wagner K, Al Kaissi A, Khan MA, Windpassinger C. Analysis of a non-lethal biallelic frameshift mutation in *ZMPSTE24* reveals utilization of alternative translation initiation codons. Clin Genet. 2023 Oct;104(4):491-496. doi: 10.1111/cge.14381. Epub 2023 Jun 4. PMID: 37270786.

and

Schaflinger E*, Blatterer J*, Khan AS, Kaufmann L, Auinger L, Tatrai B, Abbasi SW, Zeeshan Ali M, Abbasi AA, Al Kaissi A, Petek E, Wagner K, Khan MA, Windpassinger C. An exceptional biallelic N-terminal frame shift mutation in *ZMPSTE24* leads to non-lethal progeria due to possible utilization of a downstream alternative start codon. Gene. 2022 Jul 30;833:146582. doi: 10.1016/j.gene.2022.146582. Epub 2022 May 18. PMID: 35597529.
*These authors contributed equally

Some of the results in this dissertation were also part of the following Master's thesis:

"Generation of ZMPSTE24 expression clones for the characterization of a potential start gain mutation on a cellular level" submitted by Anna-Lena Mayer in August 2021 at the Medical University Graz, Austria, supervised by Assoz. Prof. Priv.-Doz. Mag. Dr.rer.nat. Christian Windpassinger

The functional analysis of the detected *ZMPSTE24* mutation, as well as the phenotypic and genetic characterization of family 2 presented in this dissertation and published by Kaufmann et al., 2023, is linked to family 1 reported in the work of Schaflinger & Blatterer et al., 2022. The genetic analysis and phenotypic characterization of family 1 were also included in the results section of this dissertation in order to provide all essential details for comprehensive understanding of the work. I contributed to the work of Schaflinger & Blatterer et al., 2022, in the in-depth analysis and by creating Figure 10A and Figure 12.

All the co-authors have agreed to the inclusion of their published data in the dissertation and permission to reproduce illustrations and figures from own or third-party publications has been granted.

This work was only possible with the contribution of following people:

Project leader and supervision: **Christian Windpassinger[1,2]**

Supervision: **Klaus Wagner[1], Dietmar Enko[3]**

Sample recruitment *ZMPSTE24*-studies: **Safdar Abbas[4], Ansar Ahmad Abbasi[5], Sumra Wajid Abbasi[6], Muhammad Ayaz[7], Aiman Saeed Khan[4], Muzammil Ahmad Khan[4], Khurram Rehman[8], Muhammad Zeeshan Ali[4]**

Genetic analysis *ZMPSTE24*-studies: **Jasmin Blatterer[1], Erich Schaflinger[1], Johann Semmler-Bruckner[2], Benjamin Tatrai[1]**

Localization experiments *ZMPSTE24*-studies: **Lisa Auinger[1,9], Anna-Lena Mayer[1], Johannes Pilic[10]**

Clinical evaluation *ZMPSTE24*-studies: **Ali Al Kaissi[11]**

Critical review of *ZMPSTE24* article manuscript (Kaufmann et al., 2023): **Wolfgang F. Graier[10,12], Roland Malli[10,12], Erwin Petek[1]**

Genetic analysis *MN1*-study: **Christine Beichler[1], Ingrid Janisch[1]**

Clinical evaluation *MN1*-study: **Bence Csapó[13]**

[1]Diagnostic and Research Institute of Human Genetics, Diagnostic and Research Center for Molecular Biomedicine, Medical University of Graz, Graz, Austria

[2]Neurogenetics Laboratory, Department of Neurology, Medical University of Graz, Graz, Austria

[3]Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University of Graz, Graz, Austria

[4]Gomal Centre of Biochemistry and Biotechnology, Gomal University, Dera Ismail Khan, Pakistan

[5]Department of Zoology, Mirpur University of Science and Technology, Mirpur, AJK, Pakistan

[6]NUMS Department of Biological Sciences, National University of Medical Sciences, Rawalpindi, Punjab, Pakistan

[7]Department of Biological Sciences, Gomal University, Dera Ismail Khan, Pakistan

[8]Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Gomal University, Dera Ismail Khan, Pakistan

[9]Division of Haematology, Medical University of Graz, Graz, Austria

[10]Gottfried Schatz Research Center, Department of Molecular Biology and Biochemistry, Medical University of Graz, Graz, Austria

[11]Pediatric Orthopedic Department, Speising Hospital, Vienna, Austria

[12]BioTechMed Graz, Mozartgasse 12/II, Graz, Austria

[13]Division of Obstetrics, Department of Obstetrics and Gynaecology, Medical University of Graz, Graz, Austria

ACKNOWLEDGEMENTS

I would like to thank all those who have supported me professionally and personally during my time as a doctoral student at the Diagnostic & Research Institute of Human Genetics.

First of all, I would like to thank my supervisor **Prof. Christian Windpassinger** for giving me the opportunity to work on my doctoral studies as a part of his research group "complex genome research". Thank you for sharing your expertise with me. I am truly grateful for your advice, support, and motivation and for teaching me both for work and for life. I hope I was able to teach you something on the tennis court in return and I look forward to further shared projects to come.

I would also like to thank the other members of the dissertation committee, **Prof. Klaus Wagner** and **Dr. Dietmar Enko**, as well as **Prof. Michael Speicher**, who unfortunately left us far too early, for their support and advice.

I would like to especially thank my favorite lab partners **Anna-Lena Mayer** and **Lisa Auinger**. It was a pleasure to work with you. Many thanks also to **Jasmin Blatterer**, who helped me get started and kept our lab organized.

Furthermore, I would like to express my appreciation to **Prof. Muzammil Ahmad Khan** of Gomal University in Pakistan and his team for the collaboration that made the *ZMPSTE24* study possible. I would also like to thank **Johannes Pilic**, **Prof. Roland Malli** and **Prof. Wolfang Graier** for their expertise and the opportunity to analyze our samples with the confocal laser scanning microscope.

Special thanks go to **Ingrid Janisch**, **Christine Beichler** and all **my colleagues**, whose contribution made the *MN1* study possible.

A heartfelt thank you to all the patients who participated in our studies and to their families. I greatly appreciate your valuable contribution to our research.

Lukas Kaufmann received funding from the Medical University of Graz through the Doctoral School “Lifestyle-Related Diseases” and has been awarded a dissertation scholarship from the Medical University of Graz.

Finally, I would like to thank my **family** (mom, Robsi, Jana, dad, Sylvia, grandma, grandpa, Rudi and Sabine), my **friends** and **Leni** from the bottom of my heart for their support in every aspect of my life. You have to know: I love you and I am very happy to have you in my life. A very special thank you goes to my **son**, who motivated me to finish this work with his upcoming birth.

TABLE OF CONTENTS

ABBREVIATIONS

°C	Degree Celsius
aa	Amino acids
ACMG	American College of Medical Genetics and Genomics
AMP	Association for Molecular Pathology
array CGH	Array-based comparative genomic hybridization
bp	base pairs
CAAX motif	Sequence of cysteine, aliphatic amino acid, aliphatic amino acid and any amino acid
cDNA	Complementary DNA
CDS	Coding sequence
CEBALID syndrome	Craniofacial defects, dysmorphic ears, structural brain abnormalities, expressive language delay, and impaired intellectual development syndrome [alternative name of MCTT syndrome]
ClinGen	Clinical Genome Resource
CNV	Copy number variation
CPMER	Cytoplasmic mesoderm regulator
CSIM motif	Sequence of cysteine, serine, isoleucine and methionine
DECIPHER	DatabasE of genomiC varIation and Phenotype in Humans using Ensembl Resources
EJC	Exon junction complex
EMBL-EBI	European Bioinformatics Institute of the European Molecular Biology Laboratory
ER	Endoplasmic reticulum
EST	Expressed sequence tag
F1	Studied family 1
F2	Studied family 2
FDA	U.S. Food and Drug Administration

FISH	Fluorescence in situ hybridization
GARD	Genetic and Rare Diseases
GFP	Green fluorescent protein
GOF	Gain-of-function
HEXXH motif	Sequence of histidine, glutamic acid, any amino acid, any amino acid and histidine
HGPS	Hutchinson-Gilford progeria syndrome
HI	Haploinsufficiency
HPO	Human Phenotype Ontology
IRDiRC	International Rare Diseases Research Consortium
kb	Kilobases; equivalent to 1,000 base pairs
LMNA	Lamin A/C
lncRNA	Long non-coding RNA
LOF	Loss-of-function
MADB	Mandibuloacral dysplasia with type B lipodystrophy
MANE	Matched Annotation from NCBI and EMBL-EBI
Mb	Megabases; equivalent to 1,000,000 base pairs
MCTT syndrome	MN1 C-terminal truncation syndrome [alternative name of CEBALID syndrome]
MLPA	Multiplex ligation-dependent probe amplification
MME	Matchmaker Exchange
MN1	MN1 proto-oncogene, transcriptional regulator
mRNA	Messenger RNA
ms	millisecond
NCBI	National Center for Biotechnology Information
NGS	Next-generation sequencing
NMD	Nonsense-mediated mRNA decay
nt	Nucleotides

OMIM	Online Mendelian Inheritance in Man
ORF	Open reading frame
PCR	Polymerase chain reaction
PTC	Premature termination codon
qRT-PCR	quantitative real-time PCR
RACE	Rapid amplification of cDNA ends
RD	Restrictive dermopathy
RFLP	Restriction fragment length polymorphism
RFP	Red fluorescent protein
ROH	Runs of Homozygosity
SNV	Single nucleotide variant
SOP	Standard operating procedure
SV	Structural variation
TIS	Translation initiation site
Tm	Melting temperature
UCSC	University of California, Santa Cruz
V	Volt
VUS	Variant of uncertain significance
WES	Whole exome sequencing
WGS	Whole genome sequencing
WT	Wild-type
ZMPSTE24	Zinc metallopeptidase STE24

LIST OF FIGURES

LIST OF TABLES

ZUSAMMENFASSUNG

Die Entdeckung neuer Genotyp-Phänotyp-Assoziationen bei seltenen Erkrankungen sowie die Identifizierung der zugrundeliegenden Pathomechanismen stellen entscheidende Schritte auf dem Weg zu einem diagnostischen Erfolg dar und können zudem als Ausgangspunkt zur Entwicklung therapeutischer Maßnahmen dienen. Die Feststellung, ob und welche Auswirkungen zuvor unbekannte genetische Varianten auf den Phänotyp haben, stellt eine große Herausforderung bei der Varianteninterpretation bzw. der Diagnosestellung dar. Eine wichtige Rolle dabei spielen die computergestützte Varianten-Annotation, sowie Software-Tools zur Vorhersage möglicher Auswirkungen neuer Varianten.
Im Zuge dieser Dissertation konnten wir zwei zuvor unbekannte, krankheitsverursachende Mutationen identifizieren und charakterisieren. Beide Mutationen weisen eine bemerkenswerte Gemeinsamkeit auf: Die vorhergesagten Auswirkungen der Varianten stimmten nicht mit dem beobachteten Phänotyp der betroffenen Personen überein. Beide Varianten werden anhand der Annotation und in silico Vorhersage mit einem vollständigen Funktionsverlust (engl. Loss-of-function bzw. LOF) des betroffenen Gens assoziiert. Die Varianten betreffen zudem jeweils ein Gen mit bekannter Pleiotropie. Wir konnten zeigen, dass der vorhergesagte Funktionsverlust beider Varianten durch alternative Genexpressionsmechanismen umgangen wird.

In der ersten Studie dieser Dissertation wurde die zuvor unbekannte, homozygote, Frameshift-Variante c.28_29insA, p.(Leu10Tyrfs*37) in *ZMPSTE24* bei betroffenen Personen aus zwei pakistanischen Familien nachgewiesen. Anhand der in silico Vorhersagen, wird die entdeckte Variante mit einem vollständigen Funktionsverlust des *ZMPSTE24*-Gens assoziiert. Ein vollständiger, biallelischer Funktionsverlust dieses Gens ist ursächlich für die Restriktive Dermopathie (RD), welche bei betroffenen Säuglingen meist bereits innerhalb der ersten Lebenswoche zum Tod führt. Die betroffenen Personen der untersuchten Familien wiesen jedoch den progeroiden Phänotyp Mandibuloakrale Dysplasie mit Typ B Lipodystrophie (MADB) auf. Diese Erkrankung wird durch biallelische *ZMPSTE24*-Mutationen verursacht, die zu einer stark reduzierten Enzymaktivität des ZMPSTE24 Proteins führen. Anschließend durchgeführte Expressions- und Lokalisationsexperimente zeigten, dass der vorhergesagte vollständige Funktionsverlust des mutierten Gens durch die Nutzung von zwei N-terminalen, alternativen Translations-Initiationsstellen umgangen wird, was mit dem beobachteten Phänotyp der betroffenen Personen übereinstimmt. Bemerkenswert ist, dass eines dieser alternativen Startcodons an der Mutationsstelle der Frameshift-Variante neu entstanden ist.

Diese Konstellation wurde unseres Wissens nach noch nicht als LOF-Umgehungsmechanismus in der Literatur beschrieben.

In der zweiten Studie dieser Dissertation wurde eine bisher nicht beschriebene, heterozygote 22q12.1-Deletion bei einem Fetus mit schweren zerebralen Fehlbildungen nachgewiesen. Die Deletion betrifft das *MN1*-Gen, einschließlich des gesamten letzten Exons 2, sowie das lncRNA Gen *CPMER*, einschließlich des gesamten ersten Exons. Der aufgrund der Deletion erwartete Funktionsverlust eines *MN1*-Allels wäre mit überwiegend unspezifischen Phänotypauffälligkeiten, ohne Vorkommen von zerebralen Fehlbildungen, sowie einer unvollständigen Penetranz assoziiert, was auch in diesem Fall zu einer Genotyp-Phänotyp-Diskrepanz führt. Die durchgeführten RNA-Analysen deuten darauf hin, dass der erwartete LOF-Effekt durch die Expression von *MN1-CPMER*-Fusionstranskripten umgangen wird. Die Expression der C-terminal verkürzten MN1-Proteine, die anhand der nachgewiesenen RNA-Sequenzen vorhergesagt wurden, wird mit dem MN1-C-terminalen Trunkierungssyndrom (MCTT Syndrom) assoziiert. Bei diesem Syndrom werden unter anderem strukturelle Gehirnanomalien beschrieben. Eine Genfusion eines protein-kodierenden Gens mit einem nicht-kodierenden Gen in der Keimbahn als LOF-Umgehungsmechanismus in Verbindung mit einer Krankheitsassoziation, wurde unseres Wissens nach ebenfalls noch nicht in der Literatur beschrieben.

Mit dieser Arbeit wollen wir das Bewusstsein für die Möglichkeit alternativer Genexpressionsmechanismen bei der Varianteninterpretation schärfen und die Notwendigkeit aufzeigen, (Standard-) Annotationspipelines in diesem Zusammenhang zu verbessern. Darüber hinaus wollen wir die Wichtigkeit von zusätzlichen in-depth Analysen und weiterführenden Experimenten zur Feststellung der Auswirkungen von Kandidatenvarianten hervorheben.

ABSTRACT

The discovery of novel genotype-phenotype associations in rare diseases as well as the identification of the underlying disease mechanisms are crucial steps on the path to diagnostic success and can also serve as a starting point for the development of therapeutic interventions. A major challenge in providing a genetic diagnosis is determining whether and what effects previously unknown variants may have. Thereby, variant annotation and computer-based tools to predict possible effects of novel variants play an important role in the variant interpretation process.
In this work, we successfully discovered and analyzed two previously unknown, disease-causing mutations. Both mutations share a remarkable feature: the predicted effects of the variants did not match the observed phenotype of affected individuals. The variants each affect a gene with known pleiotropy and both variants are predicted to result in a complete loss of gene function. The predicted loss-of-function (LOF) effect is circumvented by alternative gene expression mechanisms in both variants.

In the first study presented, we identified the novel homozygous frameshift variant c.28_29insA, p.(Leu10Tyrfs*37) in *ZMPSTE24* in affected individuals of two Pakistani families. The detected mutation is predicted to result in a complete loss-of-function of the *ZMPSTE24* gene, associating the variant with lethal restrictive dermopathy (RD). However, affected individuals displayed the progeroid phenotype mandibuloacral dysplasia with type B lipodystrophy (MADB), which is associated with biallelic *ZMPSTE24* mutations preserving residual enzymatic activity of the protein. Expression and localization experiments revealed utilization of two N-terminal alternative translation initiation sites to bypass complete loss-of-function, consistent with the observed phenotype in affected individuals. One of these alternative start codons is newly formed at the insertion site, which to our knowledge has not yet been described as a LOF escape mechanism in the literature.

In the second study presented, we detected a previously unreported, heterozygous 22q12.1 deletion in a fetus with severe cerebral malformations. The deletion affects the *MN1* gene, including the entire last exon 2 and the lncRNA gene CPMER, including the entire first exon. The expected loss-of-function of the *MN1* gene is associated with mainly non-specific phenotype abnormalities, without brain anomalies, and incomplete penetrance, which also leads to a genotype-phenotype discrepancy in this case. RNA analyses indicate that the expected LOF effect is circumvented by the expression of *MN1-CPMER* fusion transcripts.

Expression of the C-terminally truncated MN1 proteins predicted based on the RNA sequences, is associated with MN1 C-terminal truncation (MCTT) syndrome, a very rare condition characterized by distinct brain malformations. To our knowledge, a disease-causing gene fusion of a protein-coding gene and a non-coding gene in the germline leading to LOF escape has also not yet been reported.

With this work, we want to raise awareness for the possibility of alternative gene expression mechanisms in the variant interpretation process and highlight the need to improve (standard) annotation pipelines in this context. Furthermore, we want to highlight the importance of additional in-depth analyses and further experiments to determine the effects of candidate variants.

1. Introduction

1.1. Rare genetic diseases

Even though rare diseases are individually “rare”, they affect a significant portion of the general population overall. Exact data on rare diseases vary greatly in the medical literature and even the definition differs by locality.

In the European Union, a disease is defined as “rare disease” if no more than 1 in 2,000 individuals (≤ 0.05%) are affected (1). Whereas in the United States, a rare disease is one that affects fewer than 200,000 people in the U.S. (2), which corresponds to less than 0.06% currently (3). In Japan, a rare disease is defined by less than 50,000 affected people in the country (4), which equals less than 0.04% at present (5). Richter et al. found 296 definitions for a rare disease from 32 jurisdictions and the global average prevalence threshold across all organizations within the jurisdictions was 40 cases/100,000 people, or 0.04% (4). Disorders that are even rarer are often termed “ultra-rare diseases” in the literature. There is no international definition, but in the United Kingdom ultra-rare diseases describe conditions with a prevalence of <1 case per 50,000 people (< 0.002%) (6, 7).

Also the exact number of different rare diseases identified so far varies, as there is no consensual standardization. One issue is the lack of consistency in defining distinct disease entities, as well as in reporting their frequencies in different countries or demographic groups. Another challenge lies in the shortcomings of the current terminologies and their evaluation (8). There are currently multiple databases for disease definitions available, such as OMIM (9), Orphanet (10), GARD (11), HPO (12), MedGen (13) etc. In order to harmonize disease definitions across the world, Vasilevsky et al. developed the Mondo Disease Ontology (Mondo). Mondo provides a logic-based structure for unifying multiple disease resources by combining information from the major knowledge sources on rare diseases (14). Mondo therefore offers perhaps the most comprehensive curated count, with a current total of 10,443 rare diseases (15). The website Orphadata, which extracts datasets from Orphanet, provides a similar number of 10,839 rare diseases (16).

There is also some variability in the medical literature when it comes to the cumulative number of people affected by rare diseases worldwide. Ferreira calculated a cumulative prevalence of 6,221/100,000 people, or 6.2% of the general population (17). Data in the European Union suggest a slightly higher cumulative prevalence of up to 8% of the European population (18, 19).

It is commonly stated that about 80% of all rare diseases have a genetic cause (17, 18). However, it is unclear what data this estimate is based on. Data in Orphadata shows that the proportion of rare diseases with a confirmed genetic cause is currently around 39% (16, 17). But novel gene-disease associations are constantly being described. Since 2014, approximately 300 new Mendelian phenotypes are added to OMIM every year, most of which represent a novel gene-disease association (20, 21). This indicates that the actual number of rare diseases with a genetic etiology is much higher than the above-mentioned 39% of currently confirmed associations would suggest (17, 22). To date, of the 38,268 human genes in total (20,433 protein-coding and 17,835 non-coding genes) (23), a disease association has been confirmed for 3,945 genes (16).

1.2. Discovery of novel gene-disease associations in mendelian diseases

Diseases that are caused by mutations in a single gene are known as Mendelian or monogenic diseases (24). In 1956 Vernon Martin Ingram discovered that a specific chemical alteration in hemoglobin is the cause of sickle cell disease due to the analysis of the protein (25). Based on the “genetic timeline” provided by the U.S. National Human Genome Research Institute, this was the first confirmed disease association of a single amino acid exchange in humans (26). The first DNA-based description of a specific disease-causing genetic alteration was the discovery of trisomy 21 as the cause of Down syndrome three years later, in 1959 (26, 27). It took another 24 years until the *HTT* gene was the first disease-associated gene to be mapped to a human chromosome, in 1983 (26, 28). Since then, our understanding of genetic variations and mechanisms in rare diseases has increased dramatically (29). This progress goes hand in hand with the enormous technological developments in the field of genetics over the last 67 years. The first method developed to examine genetic material in humans was conventional karyotyping in 1956, which was used to determine the number of chromosomes (30, 31). Other technological milestones that followed include: Southern blot (1975) (32), Sanger sequencing (1977) (33), FISH (1980) (34), PCR (1983) (26, 35), RFLP (1985) (36), chromosomal microarray (1995) (37), MLPA (2002) (38), Massive parallel sequencing/Next-generation DNA sequencing (emerged since 1993 (39-43) and commercially available since 2005 (44)) and RNA sequencing (2008) (45, 46), as well as the continuous development in the field of multiomic approaches and novel computational methods (29).

Due to their cumulative frequency, rare diseases contribute to a large number of morbidity, mortality and health care costs worldwide (47, 48). For an affected individual, a rare disease is often accompanied with a diagnostic odyssey and only symptom-based treatment. The

comprehensive diagnostic evaluation they have to undergo includes several specialist consultations, imaging examinations, invasive investigations and other laboratory and genetic tests. It is reported that about half of the patients with a rare disease never receive a diagnosis. Not receiving a definitive diagnosis also leads to prognostic uncertainty and represents a psychological burden for the patients (49-51). A rapid and accurate diagnosis may therefore eliminate the need for further diagnostic investigations, facilitate appropriate access to health care resources, enable participation in patient support networks and/or research studies, improve disease management and may even result in effective and targeted treatments (49, 51, 52). In his work, Ferreira found that an FDA-approved treatment exists for only around 2.4% of rare diseases (17). In the case of rare inherited diseases, providing the correct genetic diagnosis can have additional positive effects: it can impacting the health of relatives and families can be informed about the recurrence-risk, allowing reproductive decisions to be made (52).

The identification of novel gene-disease associations includes both, first-time disease associations of certain genes (novel disease gene discoveries) and associations of already known disease genes with other, distinct diseases (detection of new disease mechanisms) (29). The discovery of genes and mutations associated with rare diseases as well as the identification of the underlying disease mechanisms are the first steps on the path to diagnostic success and can also serve as a starting point for the development of therapeutic interventions (49, 53). Research focusing on new genotype-phenotype correlations is therefore of great importance.

1.2.1. Traditional strategies

In the past, the identification of Mendelian disease genes was based on the search for candidate genes and subsequent Sanger sequencing of these genes. Two different strategies were applied, which differed in the screening for candidate genes. The first strategy is called functional cloning, where candidate genes are selected based on the function of the gene or protein (54). More specifically, candidate genes are selected because the predicted protein function may be related to the disease etiology, or because they resemble genes associated with similar diseases (53). The second strategy is called positional cloning, in which candidate genes are identified based on its location in the genome (55). The great advantage of positional cloning is that it does not require any prior biological or medical knowledge and can be applied in an unbiased fashion (53). Important and classic genetic mapping approaches used were linkage analysis (56) and homozygosity mapping (57). If the identified disease locus and the number of genes in question are too large for the selection of candidate genes after genetic

mapping, biological and medical information can additionally be incorporated to minimize the number of candidate genes. Also the combination of data from multiple affected individuals is useful for narrowing the disease locus. When dealing with unrelated, but phenotypically similar patients, precise clinical characterization of the disease is important, as different genetic alterations could be causative in these individuals. Once candidate genes had been determined, they were analyzed for the presence of potential disease-causing mutations using Sanger sequencing. Segregation analyses, as well as functional studies in cellular and animal models, usually followed to confirm pathogenicity. This process of identifying disease genes is laborious, but had led to the characterization of approximately 1000 monogenic diseases by the year 2000, including Huntington's disease and cystic fibrosis (53, 58).
In general, for polygenic, sporadic and/or genetically heterogeneous diseases in small cohorts or individual patients, it can be very difficult or even impossible to identify disease genes using traditional strategies.

1.2.2. Modern strategies

The development of new genomic methods has led to an increase in the discovery of novel gene-disease associations in Mendelian diseases (59). For example, the invention of microarray-based comparative genomic hybridization (array CGH) has enabled to detect copy number variations (CNVs) at submicroscopic level. The change in gene dosage could then be linked to disease phenotypes, which represents another approach for discovering novel disease genes and/or disease mechanisms (60). The advent of next-generation sequencing (NGS) approaches in the same decade then dramatically accelerated the discovery of new gene-disease associations. Furthermore, the application of NGS-based approaches rapidly expanded our knowledge of genome function and changed the clinical practice of genetic medicine (59). With whole exome sequencing (WES), it became possible to sequence the entire protein-coding regions of the genome fast and in parallel. Whole genome sequencing (WGS) offers the possibility of sequencing the entire genome, which also enables the detection of variants in non-protein-coding regions. When using NGS-based approaches, a prior search for candidate genes was no longer necessary compared to traditional strategies. As a result, the challenge of discovering novel disease genes has shifted from the identification to the interpretation phase. Millions of variants are detected per genome, but only one or two of them may explain the Mendelian disease under investigation. Therefore, many different options are available for variant filtering and prioritization, which can be adapted to the respective individual case (53).

Recent improvements in bioinformatics and sequencing technologies have made it possible to identify almost the full range of causal genetic variation, including single-nucleotide variants, small insertions or deletions, repeat expansions, structural variations (SVs) and mitochondrial DNA mutations, with a single approach (29, 61, 62). This has contributed to a better understanding of more complex concepts in genetic diseases such as risk factors, penetrance and expressivity. Implementation of NGS-based approaches also led to major progress in the delineation of genetic diseases. “Phenotype-driven” delineation has been replaced by “genotype-driven” delineation (59).

If short-read exome or whole genome sequencing is not revealing, additional approaches can be used in combination to identify novel disease-causing mutations (51). These approaches include, among others: long-read technology (63-65), usage of a pan-genome reference (66, 67), case matching (51, 59), proteomics (68), transcriptomics (69, 70), metabolomics (71, 72), methylation profiling (73) and functional studies (74-77).

In their study Bamshad et al. showed that by 2017, NGS-based approaches had been used for around 36% (1,268/3,549) of all known gene discoveries in Mendelian diseases. However, since 2013, the majority of reported novel gene discoveries have been made using NGS-based approaches compared to traditional approaches. In 2017, the rate of NGS-based discoveries reached 87%. The impact of NGS-based approaches also affects the numbers of discoveries per year, which was 173 on average between 2002 and 2009 and increased to 256 discoveries per year on average between 2010 and 2017 (59). The pace of novel gene discoveries in Mendelian diseases per year appears to have stabilized at approximately 250, presumably reflecting the time required to build international cohorts, conduct functional experiments and publish results (52, 78). Bamshad et al. estimated that at a minimum, about 6,100–14,400 Mendelian diseases remain to be discovered (59).

1.2.3. Current challenges

The current challenges in identifying novel gene-disease associations are manifold and affect several areas (51, 52, 59, 79). On average, a genetic cause is found in only about 30% of patients undergoing clinical exome or genome sequencing (80, 81). In many of the other cases, the causative variant would be present in their sequencing data, but remains unrecognized because it has not yet been described as pathogenic. Instead, the variation is very likely one of millions of variants of uncertain significance (VUS), based on current knowledge, within the sequencing data (82, 83). This massive amount of VUS detected per person by genome sequencing represents one of the major challenges of recent times. Variants of uncertain significance can be located in both protein-coding and non-coding regions. For a protein-

coding VUS, clinical significance may be uncertain because the effect on the protein is unclear or because the affected gene is currently not linked to disease. Further evaluation of variants of uncertain significance can only be performed as part of research studies. Diagnostic centers are therefore increasingly collaborating with research laboratories and affiliated universities to further investigate unsolved diagnostic cases. If a laboratory succeeds in overcoming this first obstacle of transferring unsolved cases to research, further difficulties emerge. Because even after variant filtering and prioritization, a variety of candidate variants usually remain for evaluation. The only valid strategy to discover a previously unknown disease association would be by conducting functional studies of all candidate variants. This is currently an almost unmanageable task, as most research laboratories do not have the financial, methodical and time resources to perform all the necessary functional studies. In addition, there is no guarantee of obtaining a conclusive result (52, 80). For these reasons, functional studies are usually only conducted when many factors point to a disease-causing effect of a very specific variant. Efforts in recent years have focused on strategies for better variant prioritization. Further strategies to uplift the discovery of novel gene-disease associations include: patient-oriented approaches, large-scale studies, phenotype-driven approaches, constraint-based approaches, gene-to-patient approaches, building model organism databases and the integration of multi-omics data (52).

A major focus also lies in establishing collaborations worldwide to build disease cohorts and ensure data sharing. Examples of such projects are: the International Rare Diseases Research Consortium (IRDiRC) (49, 84), the Matchmaker Exchange (MME) (85), GeneMatcher (86), PhenomeCentral (87), DECIPHER (88, 89) or ClinVar (90).

Seaby et al. suggest the development of a publicly accessible human knockout database for the discovery of novel disease genes. The database would link naturally occurring null variants in genes and supporting functional evidence with shared human phenotype data (52).

In summary, the discovery of new gene-disease associations still holds great potential with no end in sight.

1.2.4. Homozygosity mapping in consanguineous families

In terms of methodology timeline, homozygosity mapping is part of the traditional strategies (57), but this approach is still a powerful tool for the identification of novel genetic disease associations in recessive conditions (91-93). Especially with regard to the challenges outlined above, homozygosity mapping can circumvent some of the current difficulties in the search for new disease genes.

Rare recessive mutations are present in every population. In outbreeding populations they are usually inherited unnoticed by future generations and only very rarely achieve homozygous status. However, in the offspring of consanguineous families, the probability of being homozygous at any genetic locus increases dramatically, depending on the degree of relatedness of the parents (94, 95). This results in an increased risk for the occurrence of recessive diseases in consanguineous families or isolated populations (96).

In the case of a rare disease in offspring of consanguineous parents, it can initially be assumed that the disorder is caused by a homozygous variant inherited from both parents (53). Homozygosity mapping, in this context also called autozygosity mapping, is characterized by examining the genome of offspring of consanguineous families with a clinical phenotype for Runs of Homozygosity (ROH; i.e. long regions of homozygosity). ROH results from merging two copies of an ancestral haplotype in one individual. The haplotype of these regions is therefore autozygous, synonymous with homozygous by descent (92). SNP genotype array or WGS, for example, can be used for the detection of ROH (93).

Perhaps the main advantage of this approach is the cost-effective and time-saving minimization of potential disease-causing variants by restricting them to the ROH regions (79). Subsequently, advanced bioinformatics filtering approaches can be used for variant prioritization of the remaining candidate variants (93). Also the likelihood of discovering a new gene-disease association is increased when affected individuals from consanguineous families are examined, due to the extreme rarity of some recessive diseases (93).

1.3. Pleiotropy

As mentioned above, the discovery of new disease mechanisms is part of the identification of new gene-disease associations. This also includes the description of previously unknown genotype-phenotype correlations of single variants, which is particularly important in the case of pleiotropy. Pleiotropy describes the ability of a gene to affect multiple distinct phenotypic traits. The concept of pleiotropy has been known since 1910 (97). In their study, Ittisoponpisan et al. determined that 12% of the disease-causing proteins listed in UniProt are pleiotropic (98). Chong et al. analyzed data in OMIM and observed that approximately 24% of all genes associated with a Mendelian disease are responsible for at least two "clinically discrete" phenotypes (22). The frequency of pleiotropic genes poses a challenge in the identification of disease-causing genes and variants. In addition, an in-depth understanding of the molecular mechanisms of pleiotropic genes is crucial for drug development, as there is an increased risk of off-target effects (98).

Ittisoponpisan et al. also discovered that pleiotropic proteins are more likely to be associated with neoplasms, congenital malformations, as well as neurological and circulatory diseases (98).

Pleiotropy of mutant alleles may have various genetic causes, including the following: different mutation types, tissue-specificity, somatic mosaicism, modifier genes, epistatic genes, mitochondrial inheritance, or yet unknown mechanisms (99).

1.3.1. Different mutation types

Different mutations in a gene may differ in their effect size, which affects the expressivity of a disease phenotype. In enzyme deficiency disorders, for example, different mutations, which result in a reduction of the functional enzyme, are causative for the disease. In many cases, there is a good correlation between gene product levels and phenotype severity.

However, if the effects of various mutations differ, it can result in clinically distinct disorders in pleiotropic genes. The effects of mutations in a gene on protein functionality and quantity depend, among other factors, on the mutation types. The disease associations of the gene *CACNA1A* are an example of pleiotropy due to different types of monoallelic mutations. Loss-of-function (LOF) mutations in *CACNA1A* cause episodic ataxia type 2, gain-of-function (GOF) mutations lead to familial hemiplegic migraine, whereas CAG trinucleotide repeat expansions are causative for spinocerebellar ataxia type 6 (99, 100).

There is a broad spectrum of genetic alterations, which can be divided into different types at distinct levels. At DNA level, there are genome mutations, chromosomal rearrangements, also known as structural variations (e.g. copy number variations, translocations, inversions), dynamic mutations and smaller alterations that can affect up to 50 base pairs, such as indels and single nucleotide variants (SNVs). A distinction at RNA level can be made between in-frame, out-of-frame, start gain/loss, stop gain/loss and splice variants. At protein level, there are synonymous, missense, nonsense and frameshift variants, as well as insertions or deletions of amino acids and also more complex alterations such as fusion proteins (101).

Many genetic alterations do not cause any functional effects, but many others can have an impact on gene expression and protein function. These mutations can be divided into two main groups based on their effect: loss-of-function mutations or gain-of-function mutations. LOF mutations, i.e. inactivating mutations, represent an important group of disease-causing alterations. There are two types of LOF mutations, which can either result in complete loss-of-function, also known as null mutations or amorphic mutations, or partial loss-of-function, also called hypomorphic mutations (Table 1). Both qualitative and quantitative effects on the protein can lead to a loss-of-function. In case of haploinsufficiency (HI), the loss of one gene copy is

sufficient to cause phenotypic abnormalities, equivalent to a dominant disorder. Otherwise, both gene copies must be impaired (recessive disorder) (101-103).

The second group of alterations that affect protein function are GOF mutations, i.e. activating mutations. Mutations of this group cause either an increased function (hypermorphic mutations), a novel function (neomorphic mutations) or an altered gene product that acts antagonistically to the wild-type (WT) allele (dominant-negative mutations or antimorphic mutations) (Table 1) (101-104).

Table 1. Mutation types and corresponding effects on function

Mutation type	Effect on function	Terms for mutations
LOF mutations	Complete loss-of-function	Null mutations or amorphic mutations
	Partial loss-of-function	Hypomorphic mutations
GOF mutations	Increased function	Hypermorphic mutations
	Novel function	Neomorphic mutations
	Antagonistic function to wild-type allele	Dominant-negative mutations or antimorphic mutations

The terms amorph, hypomorph, hypermorph, neomorph and antimorph to characterize genetic mutations were coined by Hermann Joseph Muller in 1932 (102), but apart from hypomorphic mutations, these terms are rarely used in the clinical context of human genetics. Instead, a distinction is often only made between LOF, dominant-negative and GOF mutations, with GOF being understood as a synonym for both increased function and novel/altered function (101, 105). This could be due to the fact that the effects of mutations can often only be studied or predicted to a certain extent, so that a precise definition of the effect (complete loss-of-function vs. partial loss-of-function, or increased function vs. novel function) is not always possible (103, 105).

In addition to missense mutations, nonsense and frameshift variants can also pose a major challenge in the functional interpretation of alterations in the protein-coding region. Nonsense and frameshift mutations typically lead to transcripts carrying a premature termination codon that is often subject to nonsense-mediated mRNA decay (NMD), resulting in alleles with loss-of-function. In rare cases, they can escape depletion and be translated into truncated proteins with residual functionality, potential functional enhancement and/or toxic effects (99). Understanding the mechanisms of NMD and how it is circumvented in certain cases is

therefore of particular importance when analyzing novel disease-causing nonsense or frameshift variants.

1.3.2. Escaping nonsense-mediated mRNA decay

Nonsense mediated mRNA decay is an evolutionarily conserved surveillance mechanism that degrades mRNAs containing premature termination codons (PTCs) due to nonsense or frameshift mutations. When mRNAs containing a PTC are translated, they may produce truncated proteins with dominant-negative or deleterious gain-of-function effects (106, 107). Several studies have shown that the molecular mechanism of NMD is a translation-dependent process in which NMD is triggered by abnormal translation termination (108-110). It is known that various RNA-binding proteins play a crucial role in the initiation of NMD (111). Currently, an exon junction complex (EJC)-dependent and an EJC-independent mechanism have been characterized. The EJC-dependent mode is much better studied and also more efficient than EJC-independent NMD, for which the mechanism is currently not fully understood (110, 112). Exon junction complexes form approximately 20-24 nucleotides (nt) upstream of exon-exon junctions during splicing. The EJCs remain attached to the mRNA during transport into the cytoplasm until it is displaced by the translating ribosome during the first round of translation (113-115). If the ribosome is unable to displace downstream EJCs due to a PTC, remaining EJCs interact with certain NMD factors, thereby triggering NMD (113, 116).

However, studies have demonstrated that transcripts with PTCs can also escape NMD, leading to the production of aberrant proteins with possible dominant-negative or gain-of-function effects. In most cases where NMD can be circumvented, the location of the PTCs is the crucial factor (117-122). According to the EJC-dependent mechanism, PTCs in the last exon (“last exon rule”) or in intronless transcripts (“single exon rule”) are expected to escape NMD. Furthermore, experiments have shown that NMD is inefficient when PTCs are located within the last 50 nucleotides of the penultimate exon, which is known as the “50 nt rule” (Table 2). In case of the last exon rule and the 50 nt rule, the PTC is located far downstream at the mRNA, so that even the last EJC can be displaced by the ribosome. These rules for NMD circumvention have been largely validated, but explain only part of the human variations that escape NMD (117, 118, 122, 123). Additionally, Lindeboom et al. discovered that also PTCs located in the first 100-150 coding nucleotides rarely result in NMD (“start proximal rule”) (118, 119). They further suggested that even very long exons (> 407 nt) could inhibit NMD, if the distance between a PTC and the downstream exon junction or wild-type stop codon is large (“long exon rule”) (119), which was supported by experiments of Hoek et al. (124) (Table 2).

So far, a few mechanisms have been described which can explain the start proximal NMD insensitivity. One well-established process is called translation reinitiation. Reinitiation within a main open reading frame (ORF) is characterized by the following steps: translational start of a ribosome at the canonical start codon, elongation and termination at a PTC, followed by failure of normal recycling and subsequent translational reinitiation of the same ribosome at a downstream start codon (Figure 1A). As a consequence of translation reinitiation, the exon junction complexes downstream of a PTC are also displaced by the ribosome and the mRNA will not be degraded by the EJC-dependent mechanism. The resulting N-terminally truncated protein may retain some residual protein activity, depending on the location of critical domains (111, 118, 125-127). Another alternative translation mechanism that can also result in NMD escape is known as leaky scanning. In leaky scanning, the 48S preinitiation complex scans past a putative start codon and initiates translation at an alternative, downstream start codon instead (Figure 1B). In contrast to translation reinitiation, leaky scanning does not depend on a preceding translation termination event and it does not result in the production of two polypeptide chains by the same ribosome. However, leaky scanning can also lead to the production of N-terminally truncated proteins and thus potentially escape NMD in case of a PTC upstream of the alternative start codon (Figure 1) (111).

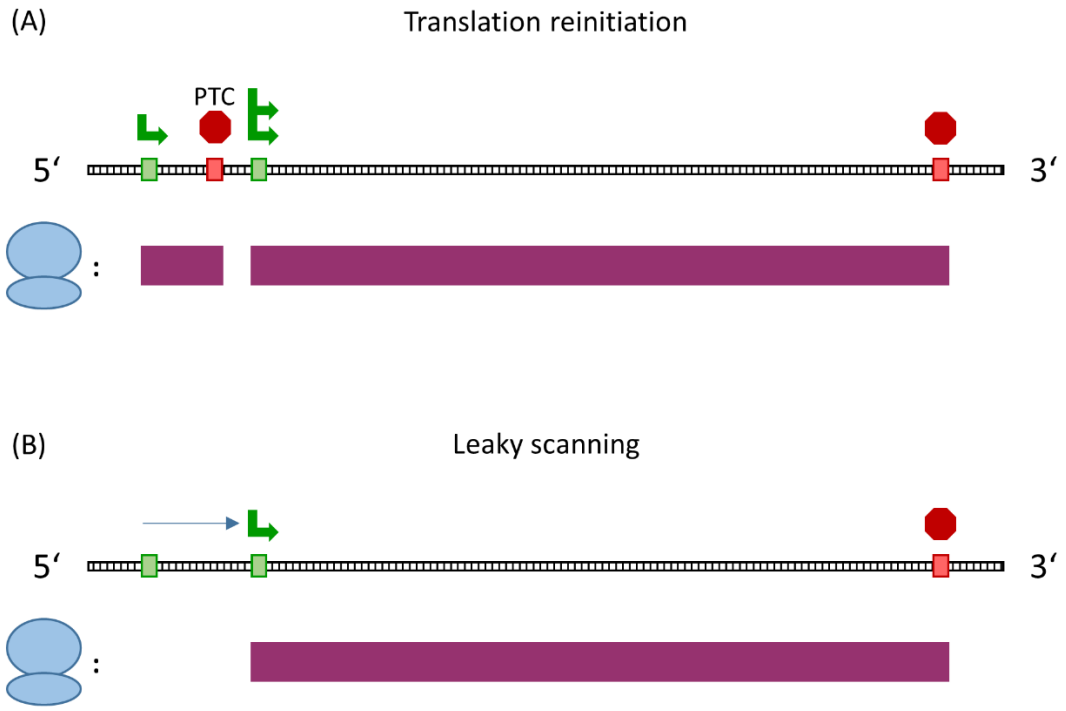

Figure 1. Comparison of translation reinitiation and leaky scanning.

Schematic illustration of the mRNA (above, black) and the polypeptide chains (below, purple) which are produced with a single ribosome (below, light blue).
(A) In translation reinitiation, the ribosome initiates translation at the canonical start codon (green arrow) and terminates at a premature termination codon (red PTC-stop sign). Then the ribosome escapes the normal recycling steps and reinitiates translation at a downstream start codon (green double arrow), which eventually ends at the canonical stop codon (red stop sign). Translation reinitiation depends on a preceding translation termination event and two polypeptide chains are produced by a single ribosome (purple bars).
(B) In leaky scanning, the 48S preinitiation complex scans past a putative start codon (thin blue arrow) and initiates translation at a downstream start codon (green arrow). Only one polypeptide chain is formed by a single ribosome (purple bar).

These alternative translation mechanisms usually occur with a certain frequency and mostly affect only a minority of translation events on certain mRNA transcripts. Some ribosomes undergo a complete canonical round of translation from initiation to recycling, while others on the same transcript are affected by a specific signal, resulting in a non-canonical translation event (111).

Based on these findings, predicted NMD escape regions can be defined according to the location of the PTCs within the mutant transcript. These regions are summarized in Table 2. It should be noted that the PTC may be downstream of the mutation site, for example in the case of frameshift variants.

Table 2. List of PTC regions predicted to escape NMD

PTC location associated with NMD escape	Validated mechanism
Within the last exon (“last exon rule”)	All exon junction complexes are displaced
Within the last 50 nucleotides of the penultimate exon (“50 nt rule”)	All exon junction complexes are displaced
Within intronless transcripts (“single exon rule”)	No exon junction complex is attached to the mRNA
Within the first 100-150 coding nucleotides (“start proximal rule”)	Translation reinitiation or leaky scanning
Within a long exon (> 407 nt) (“long exon rule”)	Not yet sufficiently validated, mechanism unknown

Even if a gene section is affected by more complex genetic alterations, such as chromosomal rearrangements, premature degradation of the mRNA can be circumvented under certain circumstances. This includes the formation of fusion genes, which occurs when sections of two genes are joined together, due to structural variants, in such a way that they are transcribed and subsequently translated as a single unit. The resulting fusion protein then usually has a novel function (101, 128, 129).

1.3.2.1. Variant annotation of variants escaping NMD and in-depth analysis

“Annotation” generally refers to the addition of new information to data. “Variant annotation” describes the assignment of functional information to DNA variants and mutations. This process is essential for sequence analysis and therefore also for the analysis of disease-related variants. The most common annotations are gene-based annotation approaches that describe how variants alter coding sequences and which amino acids are affected by the mutations. The HGVS nomenclature (130) is generally used to describe the sequence variations. Various methods have been developed to annotate variants, which can be divided into theoretically and empirically based approaches. With certain tools it is also possible to annotate predictions about the effects of variants on protein structure and protein function (103, 131-133).

However, variation annotation can differ significantly depending on the tool utilized and also contains multiple sources of error (103, 133). Especially the annotation of variants introducing a premature termination codon presents major challenges. A main reason for this is, that the NMD mechanisms and its escape rules are not taken into account in basic tools generating annotations based on the HGVS Nomenclature. These annotations can thus lead to an incorrect interpretation of a variant (103). By analyzing around 10,000 matched tumor exomes with a NMD escape rule set, Lindeboom et al. demonstrated that 22% of PTC-containing transcripts might be translated and thus lead to a truncated protein (119). A recent study by Singer-Berk et al. showed that 27% of 1,113 predicted loss-of-function variants in the Genome Aggregation Database (gnomAD, (134)) in genes associated with autosomal recessive disease, might not undergo nonsense-mediated decay (135). In their current study, Klonowski et al. applied their software tool "aenmd" to find PTC-causing variants predicted to escape NMD in the gnomAD and ClinVar databases. This revealed that 36% of ClinVar- and 50% of gnomAD-variants generating a PTC might escape NMD (123).

The results of these studies indicate that a substantial proportion of PTC-generating variants, or predicted LOF variants, possibly escape complete loss-of-function due to premature degradation of the mRNA, and instead lead to the formation of truncated proteins. Knowing

whether a protein is missing or truncated can be of profound clinical significance. Accurate variant annotation is therefore of paramount importance (103). Since standard annotation tools do not evaluate the variant in the context of its surrounding sequence, several studies have highlighted the benefit of deep manual curation of variants (135-138). In-depth analysis of a variant can be conducted at DNA, RNA and/or protein level. For this purpose, various methods such as in silico predictions, RNA analyses, protein analyses and functional in vitro or in vivo studies are performed (51, 139). Understanding the molecular consequences and effects of newly identified variants, which can be obtained by in-depth analysis, is particularly relevant in genes with known pleiotropy (140).

1.4. Zinc metallopeptidase STE24 - *ZMPSTE24*

The *ZMPSTE24* (Zinc metallopeptidase STE24) gene is located in the chromosomal region 1p34.2. Only one protein-coding transcript of the gene, with 10 exons, is known to date (NM_005857.5). The mRNA has a length of 2,975 nt, of which 1,425 nt form the coding sequence (CDS) (88, 141). *ZMPSTE24* is ubiquitously expressed in the cytoplasm (142). The gene is conserved in eukaryotes and encodes the protein CAAX prenyl protease 1 homolog, also known as Zinc metalloproteinase Ste24 homolog, in humans (143, 144). The protease belongs to the peptidase M48A family, which interacts with substrates that have a CAAX motif consisting of cysteine, two aliphatic amino acids and any C-terminal amino acid depending on substrate specificity (145-147). To improve readability, the term “ZMPSTE24 protein” or just “ZMPSTE24” (not in italics) is also used as a synonym throughout this thesis. The CAAX prenyl protease 1 homolog consists of 475 amino acids and is bound to the membrane of the endoplasmic reticulum (ER) and to the inner nuclear membrane via seven transmembrane domains (148). The protease domain is located on the cytosolic side of the membrane and spans the C-terminus of the protein. The amino acids His335, Glu336, His339, which build the HEXXH motif (X stands for any amino acid), and Glu415 form the zinc-binding, catalytic site. A zinc ion (Zn^{2+}) is bound as cofactor and the catalytic site is surrounded by the hydrophobic residues Pro248, Trp340 and Leu425. The substrate binding site comprises a peptide complex around the amino acid Asn265 (144, 149-151). The zinc metalloproteinase has catalytic cleavage activity and catalyzes two crucial steps in the post-translational processing of prelamin A to mature lamin A (Figure 2) (148). Prelamin A is encoded by the *LMNA* gene and must be processed to exhibit proper functionality. *LMNA* also codes for lamin C through alternative splicing, which does not require processing. Lamin A and lamin C are intermediate filament proteins and provide stability and strength to cells. Together with other proteins, they

are part of the nuclear lamina, which is attached to the inner membrane of the nuclear envelope. Lamin A and C are also present inside the nucleus (152-154).

The precursor protein prelamin A contains the C-terminal CAAX sequence cysteine, serine, isoleucine and methionine (CSIM) (155). The maturation of prelamin A occurs in a multi-stage process: In the first step, a farnesyl residue is attached to the cysteine of the CSIM sequence (farnesylation). The proteases ZMPSTE24 and RCE1, which is also called CAAX prenyl protease 2, redundantly mediate the cleavage of the SIM residue in the second step. In the third step, the farnesylated C-terminus is methylated by the isoprenylcysteine carboxyl methyltransferase (ICMT). The fourth and final step in the formation of mature lamin A is a second endoproteolytic cleavage of the last 15 C-terminal amino acids at a specific cleavage site. In contrast to the first cleavage, this step is catalyzed exclusively by the ZMPSTE24 protein (Figure 2) (156-159).

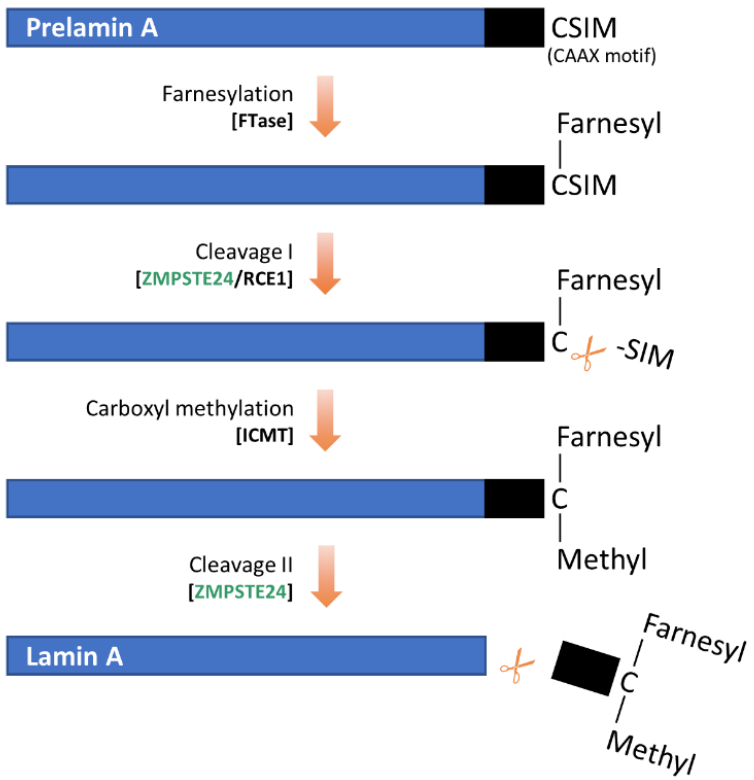

Figure 2. Post-translational processing of prelamin A to form mature lamin A.

[Figure and legend adapted from (160), (Creative Commons Attribution License)]

This simplified illustration outlines the four processing steps in the formation of mature lamin A, along with the enzymes involved (square brackets). The ZMPSTE24 protease (highlighted in green) is involved in two processing steps. Prelamin A contains a C-terminal CAAX motif (C = cysteine, AA = aliphatic

amino acids, X = any amino acid) consisting of the sequence cysteine, serine, isoleucine and methionine (CSIM). First processing step: Farnesylation of the cysteine in the CAAX motif by the enzyme Farnesyltransferase (FTase). Second step: Removal of the three C-terminal amino acids SIM. This first endoproteolytic cleavage can be catalyzed either by the enzyme RCE1 or by ZMPSTE24. Third step: Carboxyl methylation of the C-terminal cysteine by isoprenylcysteine carboxyl methyltransferase (ICMT). Fourth step: Cleavage of the last 15 C-terminal amino acids at a specific cleavage site to obtain mature Lamin A. This second endoproteolytic cleavage is catalyzed solely by ZMPSTE24.

Dysregulation of lamins is linked to aberrant nuclear morphology and chromatin disorganization (161). Mutations in the *LMNA* gene are causative for a broad spectrum of diseases such as muscular dystrophy, dilated cardiomyopathy, Charcot-Marie-Tooth neuropathy, syndromes of premature aging (including Hutchinson-Gilford progeria syndrome (HGPS) and Mandibuloacral dysplasia with type A lipodystrophy (MADA)), restrictive dermopathy, and other overlapping syndromes (162).

1.4.1. Diseases caused by mutations in *ZMPSTE24*

As described above, the ZMPSTE24 protease is particularly essential for the second cleavage during prelamin A processing. Failure to perform this cleavage step leads to an accumulation of farnesylated and methylated prelamin A (see Figure 2), which exerts systemic toxic effects and is associated with laminopathies and progeroid syndromes. Therefore, mutations in the *LMNA* gene, which prevent the interaction with the ZMPSTE24 protein, and mutations in the *ZMPSTE24* gene, which lead to a loss-of-function of the protease activity, can lead to overlapping diseases (161, 163). Laminopathies caused by mutations in the *LMNA* gene are called primary laminopathies. Secondary laminopathies are caused by mutations in genes encoding proteins that structurally or functionally interact with lamin A/C, including *ZMPSTE24* (164).

ZMPSTE24 demonstrates pleiotropy, enabling biallelic mutations in the gene to be associated with two distinct disease phenotypes. The severity of the associated disease thereby correlates with the residual enzyme activity of the protease (144, 149).

1.4.1.1. Mandibuloacral dysplasia with type B lipodystrophy

Biallelic pathogenic variants in *ZMPSTE24* that preserve residual proteolytic activity, i.e. homozygous and compound heterozygous missense variants, as well as an in-frame deletion, are causative for mandibuloacral dysplasia with type B lipodystrophy (MADB, OMIM #608612).

Mandibuloacral dysplasia, including type A and B lipodystrophy, is a rare disease whose exact prevalence is unknown. The phenotypic spectrum of the autosomal recessive disorder MADB includes the following clinical manifestations: growth retardation, craniofacial anomalies such as mandibular hypoplasia, skeletal anomalies such as progressive osteolysis of the terminal phalanges and clavicles, skin changes such as patchy hyperpigmentation and atrophy, lipodystrophy characterized by a generalized loss of subcutaneous fat affecting the face, trunk and extremities and progeroid traits (165-168). Clinical signs usually appear in early childhood (169). MADB displays the milder spectrum of the two disorders associated with *ZMPSTE24*.

1.4.1.2. Restrictive dermopathy

Complete loss-of-function of both *ZMPSTE24* alleles, including homozygous and compound heterozygous nonsense and frameshift variants lead to restrictive dermopathy (RD, OMIM #275210). Restrictive dermopathy is a very rare disease with a prevalence of <1 / 1,000,000. As of 2019, around 80 cases had been described worldwide. RD is a lethal genodermatosis with characteristic features that are already apparent at birth. Clinical features include thin, tightly adherent translucent skin with erosions at flexure sites, superficial vessels, typical facial dysmorphism and generalized joint ankyloses. Prenatal findings such as intrauterine growth retardation, reduced fetal movements, polyhydramnios and premature rupture of the membranes have also been described. Most newborns die within the first week of life, thus RD displays the severe disease phenotype of *ZMPSTE24*-associated disorders (10, 170, 171).

An overview of the two diseases associated with the *ZMPSTE24* gene, as well as the correlated protein activity status and the type of causative variants is shown in Table 3.

Table 3. Diseases associated with the *ZMPSTE24* gene

Gene	Disorder	Phenotypic features	ZMPSTE24 activity status	Causative biallelic variants
ZMPSTE24	Mandibuloacral dysplasia with type B lipodystrophy (MADB)	Growth delay, craniofacial anomalies, skin anomalies, progeroid appearance and lipodystrophy	Residual proteolytic activity	Homozygous/compound heterozygous missense mutations, in-frame deletion
	Restrictive dermopathy (RD)	Lethal genodermatosis affecting skin and vessels, facial dysmorphism and joint ankyloses	Complete loss of proteolytic activity	Homozygous/compound heterozygous nonsense and frameshift mutations

1.5. MN1 proto-oncogene, transcriptional regulator - *MN1*

The *MN1* (MN1 proto-oncogene, transcriptional regulator) gene in the chromosomal region 22q12.1 has only two exons. The majority (95%) of the 3,963 bp long CDS of the MANE transcript NM_002430.3 is located in exon 1 (141). *MN1* was first described as a balanced translocation (4;22) gene in a meningioma, which was disrupted in the open reading frame of the first exon. It was therefore assumed that the gene acts as a tumor-suppressor. The patient with multiple meningiomas carried the translocation in his germline. No MN1 mRNA expression was detected in the tumor (172, 173). Later on, the gene was also considered to be an oncogene, as high *MN1* expression was linked to poor prognosis in acute myeloid leukemia (174).

The gene encodes the 1,320 amino acid long transcriptional activator MN1 (143). Studies in mice indicate that the transcription factor MN1 regulates the development of the mammalian palate (175) and is needed for proper proliferation, motility, differentiation and function of osteoblasts (176). A homozygous knockout in mice leads to abnormal skull bone development and results in cleft palate, with mice dying at or shortly after birth. Heterozygous knockout mice manifest a milder phenotype including hypoplastic membrane bone and cleft palate with incomplete penetrance (177). Apart from the associations with meningiomas and acute myeloid leukemia, no definite association to any other disease has been confirmed in humans over many years.

1.5.1. Diseases caused by mutations in *MN1*

Large, heterozygous deletions harboring the *MN1* gene have been detected in individuals with orofacial abnormalities and cleft palate. However, a clear genotype-phenotype correlation with *MN1* was not possible due to the involvement of other genes in these cases (178-180).

It was not until 2020 that the papers by Mak et al. (181) and Miyake et al. (182) on specific genotype-phenotype associations of heterozygous germline mutations in *MN1* were published almost simultaneously. It was discovered that *MN1* exhibits pleiotropy, whereby the phenotypic effects depend on the mutation type: On the one hand, *MN1* demonstrates haploinsufficiency, resulting in heterozygous LOF variants leading to the milder phenotypic spectrum of the two diseases. GOF variants, on the other hand, are associated with more severe phenotypic abnormalities (181, 182).

1.5.1.1. *MN1* haploinsufficiency-associated phenotypic findings

Three SNVs (one frameshift and two nonsense variants) in *MN1* predicted to induce nonsense-mediated mRNA decay, are associated with conductive hearing loss and a range of speech defects (Figure 3). A patient with a focal deletion of the entire *MN1* gene presented with microcephaly, intellectual disability, speech and motor delays and mildly dysmorphic facial features. No other abnormalities, like cleft palate observed in individuals with large deletions, can be clearly associated with *MN1* haploinsufficiency (181). A few LOF variants in *MN1* are also listed in the gnomAD database, indicating reduced penetrance of the *MN1* haploinsufficiency phenotype (83).

1.5.1.2. MN1 C-terminal truncation syndrome

The MN1 C-terminal truncation (MCTT) syndrome is listed in the OMIM database as CEBALID (craniofacial defects, dysmorphic ears, structural brain abnormalities, expressive language delay, and impaired intellectual development) syndrome (183). MCTT syndrome is also caused by nonsense or frameshift variants, whereby the localization of the premature termination codons is crucial. In MCTT, all truncating variants are located within the last exon (exon 2), or at the 3' end of exon 1 (Figure 3). These variants thus escape NMD and lead to the formation of a C-terminally truncated protein with pathological activity. The results of the functional study by Miyake et al. show that C-terminally truncated MN1 proteins impair the binding of the E3 ubiquitin ligase RING1. The interaction of MN1 with molecules associated with the ubiquitin-mediated proteasome pathway may therefore be impaired. This results in an increase in the amount of truncated MN1 due to inhibition of rapid MN1 protein turnover, which causes dysregulation of MN1 target genes (182).

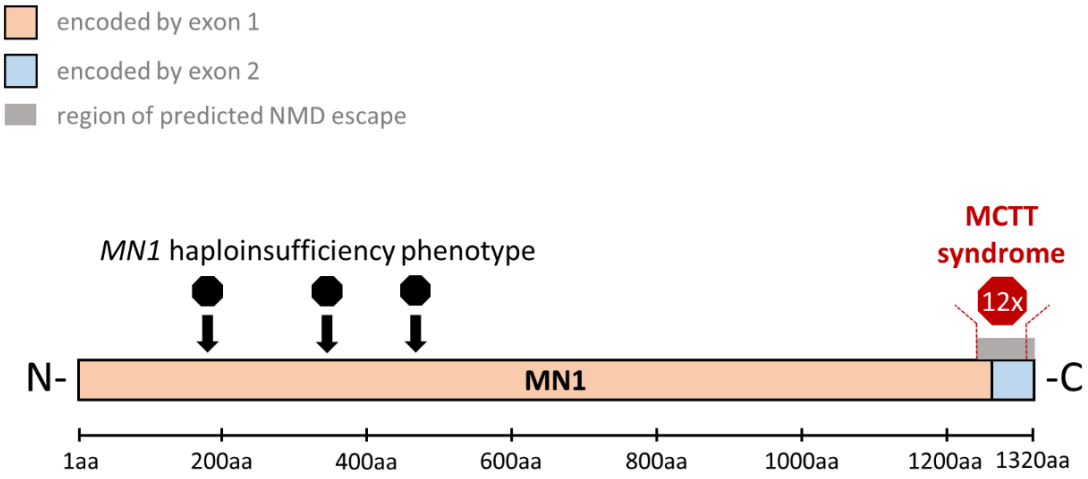

Figure 3. MN1 mutations and associated phenotypes.

The 1,320 amino acid long MN1 protein is illustrated with marked amino (N)-terminus and carboxyl (C)-terminus. The loss-of-function variants associated with the *MN1* haploinsufficiency phenotype (black stop sign) reported by Mak et al. are predicted to result in nonsense-mediated mRNA decay. This would lead to a lack of expression of this allele. The 12 known late truncating variants causing the MN1 C-terminal truncation syndrome reported by Mak et al. and Miyake et al. (red stop sign) are all localized in the region of predicted NMD escape (grey bar), resulting in the expression of C-terminally truncated proteins. The truncated proteins exhibit a gain of function effect and the *MN1* gene demonstrates pleiotropy.

MN1 C-terminal truncation syndrome is a very rare disease with fewer than 30 cases described worldwide. The MCTT syndrome is characterized by the following clinical findings: mild to moderate intellectual disability, severe speech delay, dysmorphic facial features (midface hypoplasia, downslanting palpebral fissures, hypertelorism, exophthalmia, short upturned nose and small low-set ears), often also mild to moderate prelingual hearing loss (usually bilateral, conductive and/or sensorineural) and distinctive findings on brain imaging (including perisylvian polymicrogyria and atypical rhombencephalosynapsis). The risk of craniosynostosis is increased and less frequently, generalized seizures are also described (184).

Table 4 lists the diseases and clinical features associated with the *MN1* gene, as well as the underlying pathomechanisms and causative variants.

Table 4. Diseases associated with the *MN1* gene

Gene	Disorder	Phenotypic features	*MN1* pathomechanism	Causative monoallelic variants
MN1	*MN1* haploinsufficiency phenotype	Associated with hearing loss (conductive), speech defects, microcephaly and intellectual disability with reduced penetrance	Loss-of-function	Heterozygous deletion, nonsense and frameshift variants leading to nonsense-mediated mRNA decay
	MN1 C-terminal truncation (MCTT) syndrome	Intellectual disability, severe expressive language delay, Hearing loss (conductive or sensorineural), dysmorphic facial features and distinctive findings on brain imaging	Gain of function mediated by expression of C-terminally truncated MN1 proteins	Heterozygous nonsense and frameshift variants in the 3' region of the gene, escaping nonsense-mediated mRNA decay

The literature reports demonstrate that the two *MN1*-associated disorders display some overlapping clinical features. However, brain malformations and a consistent craniofacial gestalt are features that are only found in MCTT syndrome. The assumed incomplete penetrance linked to *MN1* haploinsufficiency also distinguishes the two conditions (181).

1.6. Research question and aims of the dissertation

The major aim of this dissertation was to identify novel genotype-phenotype correlations in rare genetic diseases. Studying the cause of a genetic disease is a crucial step in providing the correct diagnosis, can improve disease management and may also be the first step towards effective and targeted treatments. Only when the pathomechanism has been clarified further steps can be taken to develop a specific therapy (49-53). In order to achieve this goal, access to people with rare and ultra-rare diseases is necessary. A long-standing cooperation with Gomal University in Dera Ismail Khan, Pakistan enables the genetic analysis of consanguineous families in which rare diseases have been observed. In addition, individual cases can also be transferred from routine diagnostics to the research area at our Diagnostic & Research Institute of Human Genetics at the Medical University of Graz, if an in-depth analysis is promising for identifying the cause.

In this study, we are investigating two different and previously unknown disease-causing mutations. Both mutations have in common that the predicted effect of the respective variant does not match the observed phenotype of affected individuals. Another common feature is

that the variants affect genes that are known to be pleiotropic and both variants are predicted to result in a complete loss of gene function. Our main hypothesis was a rescue of the predicted loss-of-function effects, resulting in the observed disease phenotypes. The examination of the respective rescue mechanism therefore evolved as key research question.

One of the novel disease-causing variants identified is a homozygous, N-terminal frameshift mutation in the *ZMPSTE24* gene, which, astonishingly, segregates with MADB instead of lethal RD in two consanguineous Pakistani families. We use in-depth analysis of the variant and in vitro expression and localization experiments to study the hypothesized rescue mechanism. According to comprehensive literature search, a comparable LOF-rescue had not been reported before.

The other novel variant was detected in a fetus at 14 weeks of gestation with ultrasound abnormalities. The fetus was found to have a heterozygous deletion of approximately 58 kilobases (kb) in the chromosomal region 22q12.1. The deletion partially overlaps the *MN1* gene indicating a loss of gene function. However, the severe brain malformations observed in the fetus are not consistent with an *MN1* LOF-genotype. Also in this case, in-depth analysis yielded a possible rescue mechanism. Qualitative and quantitative RNA analyses were conducted to examine this hypothesis. To our knowledge, such an LOF rescue has not yet been described in the literature either.

In addition to the description of the novel mutations and the clarification of the rescue mechanisms by performing experimental studies, a detailed clinical description of the phenotypic abnormalities observed in affected individuals is another aim of this dissertation. The analyses of patient tissues were not possible in either case due to the inaccessibility of tissue samples. The discovered genotype-phenotype correlations enable rapid diagnosis of future cases with the same mutations.

However, this work is not only relevant for the diseases associated with the two genes. The studied rescue mechanisms have general genetic validity and could also occur with similar variants in other genes. Furthermore, the work demonstrates the great importance of in-depth analyses for the correct interpretation of previously undescribed variants, since standard annotation pipelines do not account for complex processes.

2. Materials and methods

The studies were approved by the ethics committees of the Gomal University, Dera Ismail Khan, Pakistan (approval numbers: 04/ ERB/GU and 32/DQA/GU) and the Medical University of Graz, Austria (approval number: 35-476 ex 22/23). The samples were enrolled after written informed consent had been obtained from all study participants or, in the case of the *MN1* study, from the mother of the fetus.

2.1. *ZMPSTE24* study

2.1.1. Clinical diagnosis (*ZMPSTE24*)

The clinical features of affected individuals in both families studied were determined by a physician. Radiological and biochemical tests were also conducted to establish the diagnosis.

2.1.2. Genetic analyses (*ZMPSTE24*)

For the genetic analysis of affected individuals of index family 1 (F1), whole exome sequencing was performed for two affected individuals of the 4th generation (F1-IV:7, F1-IV:8, see Figure 8A) via Agilent SureSelect V6 human All Exon library preparation and sequencing was conducted using a NovaSeq 6000 with 2 x 150 bp and 100x coverage (50x on-target coverage). The DRAGEN Germline Pipeline 3.2.8 on Illumina BaseSpace (Illumina, www.basespace.illumina.com) was used for sequence alignment of raw fastq files to the human reference sequence (GRCh37/hg19 assembly) and variant calling. Variant annotation, analysis and homozygosity mapping was conducted using VarSeq™ v2.2 (Golden Helix, www.goldenhelix.com).

The identified ROH-regions were first screened for possible causative homozygous variants. Subsequently, the entire exome data was analyzed to identify potential variants that might contribute to the patients' phenotype. For segregation analysis of the potentially disease-causing variant detected by WES in other family members, Sanger sequencing was performed using the BigDye™ Terminator v1.1 Cycle Sequencing Kit (catalog no. 4337452, Applied Biosystems™/Thermo Fisher Scientific, www.thermofisher.com) according to the manufacturer's protocol. Sequencing fragments were then separated electrophoretically on the Applied Biosystems "SeqStudio" Genetic Analyzer (Applied Biosystems™/Thermo Fisher Scientific, www.thermofisher.com/seqstudio). Analysis and interpretation were performed using Sequencing Analysis Software v7.0 (Thermo Fisher Scientific) and the software

SnapGene (SnapGene®, www.snapgene.com). The online tool Primer3plus (185) was used for the primer design (see Table 5). A primer annealing temperature of 56 °C for 1 minute was selected to amplify the targeted DNA region.

Table 5. Primer set for Sanger sequencing in segregation analysis

Primer labelling	Sequence (5'-3')	Length	Tm
ZMPSTE24_E1_f	GCTCTGAAGGGACGAGTGTC	20 nt	60.0 °C
ZMPSTE24_E1_r	GAAGCCAAGGCTACTCCAGG	20 nt	61.3 °C

f = forward; r = reverse; nt = nucleotides; Tm = melting temperature

For the genetic analysis of family 2 (F2) we performed autozygosity mapping of one affected 3rd generation individual (F2-III:4, see Figure 9A) using the Infinium CytoSNP-850K (Illumina, www.illumina.com) in combination with the BlueFuse™ 4.5 Multi Analysis Software (Illumina). Confirmation of the mutation and segregation analysis in members of family 2 were also performed via Sanger sequencing as described above. The primer set used is shown in Table 5.

A comparative genotype analysis of affected individuals F1-IV:8 and F2-III:4 was performed with Infinium CytoSNP-850K derived data visualized via BlueFuse™ 4.5 Multi Analysis Software (Illumina).

2.1.3. In silico prediction analyses

The online tools NetStart 1.0 (186) and ATGpr (187) were used for the in silico analyses to predict potential translation initiation sites (TISs) in *ZMPSTE24*. The analyses were performed with both, the wild-type (NM_005857.5) and the mutated sequence. For protein sequence alignment the analysis tool Clustal Omega (188) was used. In silico splice site prediction was performed using the splice site prediction tool of varSEAK (189).

2.1.4. Functional analyses

We performed *in vitro* expression and localization experiments to demonstrate the hypothesized utilization of an alternative TIS as bypass mechanism for NMD and complete loss of protein function in affected individuals. Plasmids with different *ZMPSTE24* expression constructs were designed and used in these experiments. In order to visualize the expression and localization in cells, an in-frame green fluorescent protein (GFP) sequence was attached

to the 3'-end of each construct (C-terminal GFP-tag). To determine localization, a red fluorescent protein (RFP) marker for the endoplasmic reticulum (mCherry-ER-3, plasmid #55041) was obtained from the Michael Davidson Fluorescent Protein Collection available on Addgene (www.addgene.org). All primers were purchased from Microsynth AG (www.microsynth.com).

2.1.4.1. Approach 1 – Cloning

Our first approach was the insertion of the human *ZMPSTE24* cDNA sequence into the multiple cloning site of the plasmid pcDNA™ 3.1 CT-GFP-TOPO® (Invitrogen™/Thermo Fisher Scientific, www.thermofisher.com) and subsequent cloning.

2.1.4.1.1. TA Cloning

Human Fetal Brain Marathon Ready cDNA (catalog no. 639302, Clontech/Takara Bio, www.takarabio.com) was used as a template for PCR amplification of the *ZMPSTE24* cDNA. The HotStarTaq DNA Polymerase (Qiagen, www.qiagen.com) was used to add a single deoxyadenosine (A) to the 3´ ends of the PCR amplicons, which is required for TA cloning® (190).

The reaction mix for hot start PCR with a total volume of 12 µl contained the following components:

- 6.0 µl HotStarTaq Master Mix (Qiagen)
- 4.5 µl LiChrosolv® water (Merck, www.merckmillipore.com)
- 0.5 µl forward primer [10 µM]
- 0.5 µl reverse primer [10 µM]
- 0.5 µl Human Fetal Brain Marathon Ready cDNA (Clontech/Takara Bio)

The primer set used is shown in Table 6 and cycling parameters for hot start PCR are listed in Table 7.

Table 6. Primer set used for amplification of *ZMPSTE24* from human fetal brain cDNA

Primer labelling	Sequence (5'-3')	Length	Tm
ZMPSTE24_f2	GTGGCCATGGGGATGTGGG	19 nt	69.1 °C
ZMPSTE24_r2	GATCCTGGACATCTCAGTGTTGCTTCATAG	30 nt	68.7 °C

f = forward; r = reverse; nt = nucleotides; Tm = melting temperature

Table 7. Thermo cycler profile for hot start PCR

Temperature	Time	Cycles
94 °C	15 min	1
95 °C	25 sec	30
66 °C	30 sec	
72 °C	1 min	
72 °C	7 min	1
4 °C	hold	

Amplification was verified by gel electrophoresis with a 1 % agarose gel, stained with GelRed® (Biotium, www.biotium.com). The gel electrophoresis was run at 130 volts and 400 mA for 30 minutes.

Cloning of the amplified PCR product in the expression plasmid pcDNA™3.1 CT-GFP-TOPO® was conducted using the CT-GFP Fusion TOPO® TA Expression Kit (catalog no. K4820-01, Invitrogen™/Thermo Fisher Scientific, www.thermofisher.com) according to the manufacturer's protocol, for fusing GFP to the C-terminus of PCR products. The subsequent transformation is described in chapter 2.1.4.2.2.

2.1.4.1.2. Transformation and plasmid propagation (approach 1)

After ligation of the amplified sequence from human fetal brain cDNA into the vector pcDNA™ 3.1 CT-GFP-TOPO®, 10 ng of the vector was transformed into One Shot™ TOP10 chemically competent *E. coli* cells (catalog no. C4040, Invitrogen™/Thermo Fisher Scientific, www.thermofisher.com) according to the manufacturer's protocol. Between 25-200 µl from the transformation-mix were spread on pre-warmed selective ampicillin-LB plates [100 µg/ml] and incubated overnight at 37 °C. The pUC19 plasmid included in the kit was used as positive control.

A colony PCR was performed to confirm transformation of the recombinant plasmid in *E. coli*. Several colonies were picked, suspended in 50 µl LiChrosolv® water (Merck) and denatured at 95 °C for 5 minutes. The samples were then cooled to 4 °C and incubated on ice for 10 minutes, followed by centrifugation for 1 minute at 14,000 rcf. The supernatant was used for the reaction mix of the colony PCR, which equals the reaction mix of the hot start PCR (see

chapter 2.1.4.1.1). The primer set used for colony PCR was selected to ensure that amplification only occurs if the *ZMPSTE24* sequence is successfully inserted into the plasmid. The forward primer equals the T7 promoter sequence and the reverse primer matches the complementary sequence of a region in *ZMPSTE24*. The primer set is shown in Table 8 and cycling parameters are listed in Table 9.

Table 8. Primer set used for colony PCR (approach 1)

Primer labelling	Sequence (5'-3')	Length	Tm
T7 primer_f	TAATACGACTCACTATAGGG	20 nt	54.3 °C
Insert control_r	TTTCAGAATCCATGATCTGT	20 nt	52.3 °C

f = forward; r = reverse; nt = nucleotides; Tm = melting temperature

Table 9. Thermo cycler profile for colony PCR (approach 1)

Temperature	Time	Cycles
94 °C	15 min	1
95 °C	25 sec	33
51 °C	30 sec	
72 °C	1 min	
72 °C	7 min	1
4 °C	hold	

Successful insertion and subsequent transformation was confirmed via gel electrophoresis with a 1 % agarose gel, stained with GelRed® (Biotium), using 5 µl of PCR product. The gel electrophoresis was run at 130 volts and 400 mA for 30 minutes.

2.1.4.1.3. Plasmid isolation (approach 1)

An overnight culture was prepared prior to plasmid isolation. For this purpose, colonies with confirmed plasmid transformation were pre-cultured in 3 ml LB medium with ampicillin [100 µg/ml] and incubated for 5-6 hours at 37 °C under shaking at 225 rpm. The entire volume of pre-cultured cells was then transferred into 200 ml LB medium with ampicillin [100 µg/ml] for the main culture. The main culture was then incubated overnight for 13-16 hours at 37 °C while

shaking at 225 rpm. For plasmid isolation the QIAprep® Spin Miniprep Kit (catalog no. 27104, Qiagen) was utilized according to the manufacturer's protocol.

2.1.4.1.4. Sequence verification of the generated construct (approach 1)

After plasmid isolation, the plasmid-DNA concentration was measured via NanoDrop™ (Thermo Fisher Scientific). To determine the sequence of the insert, Sanger sequencing was performed using the BigDye™ Terminator v1.1 Cycle Sequencing Kit (catalog no. 4337452, Applied Biosystems™/Thermo Fisher Scientific) according to the manufacturer's protocol. The reaction mix for sequencing contained the following components:

- 0.5 µl BigDye™ Terminator v1.1 Ready Reaction Mix
- 1.4 µl 5X Sequencing Buffer
- 0.4 µl sequencing primer [10 µM]
- 2,5 µl plasmid [164 ng/µl]
- 5.7 µl LiChrosolv® water (Merck)

The plasmid volume used was calculated as follows: The minimum amount of DNA required was set at 25 ng per 500 bp (equivalent to 0.05 ng/bp) based on personal experience. The plasmid pcDNA™ 3.1 CT-GFP-TOPO® including the *ZMPSTE24* insert has a calculated size of 7604 bp. This results in a minimum quantity of: 0.05 ng/bp x 7604 bp = 380 ng. For sequencing, 410 ng of plasmid were used [410 ng / 164 ng/µl = 2.5 µl]. The sequencing primers used are shown in Table 10 and cycling parameters are listed in Table 11.

Table 10. Primers for insert sequencing (approach 1)

Primer labelling	Sequence (5'-3')	Length	Tm
T7 primer_f	TAATACGACTCACTATAGGG	20 nt	54.3 °C
ZMPSTE24_seq mid_r	GAAGAGCGTTTAGATCCTTC	20 nt	56.4 °C
ZMPSTE24_seq mid_f	CTATGCTGATTATATTGCCC	20 nt	54.3 °C
GFP_r	AAAGTTCTTCTCCTTTGCTAG	21 nt	55.4 °C

f = forward; r = reverse; nt = nucleotides; Tm = melting temperature

Table 11. Thermo cycler profile for sequencing (approaches 1 & 2)

Temperature	Time	Cycles
96 °C	30 sec	25
50 °C	30 sec	
60 °C	1 min	
6 °C	hold	

Purification was performed by size exclusion chromatography. Therefore, Centri-Sep™ Spin columns (catalog no. 401763, Invitrogen™/Thermo Fisher Scientific) were loaded with Sephadex® G-50 (Cytiva). The PCR products were diluted 1:1 (10 µl LiChrosolv® water (Merck) and 10 µl PCR products) and pipetted onto the Sephadex-columns. The loaded columns were centrifuged for 2 min at 750 rcf. The PCR products were then transferred to a 96-well plate (catalog no. 710880, Biozym, www.biozym.com) and analyzed using a 3730xl DNA Analyzer (Thermo Fisher Scientific) with POP-7™ polymer (catalog no. 710880, Thermo Fisher Scientific) as separation matrix. The raw data generated was analyzed using SeqScape™ Software v3.0 (Applied Biosystems™/Thermo Fisher Scientific), SnapGene (SnapGene®) and UCSC Genome Browser (www.genome.ucsc.edu) (191).

2.1.4.2. Approach 2 – Mutagenesis

In the second approach, the expression plasmid "ZMPSTE24 (NM_005857) Human Tagged ORF Clone" (catalog no. RG206035, OriGene™ technologies, www.origene.com) was used as wild-type (WT) clone, labeled ZMPSTE24(WT)-GFP in this study. This clone is based on the vector pCMV6-AC-GFP (PS100010) with an inserted *ZMPSTE24* wild-type sequence and a C-terminal TurboGFP tag. The plasmid map is shown in Figure 4.

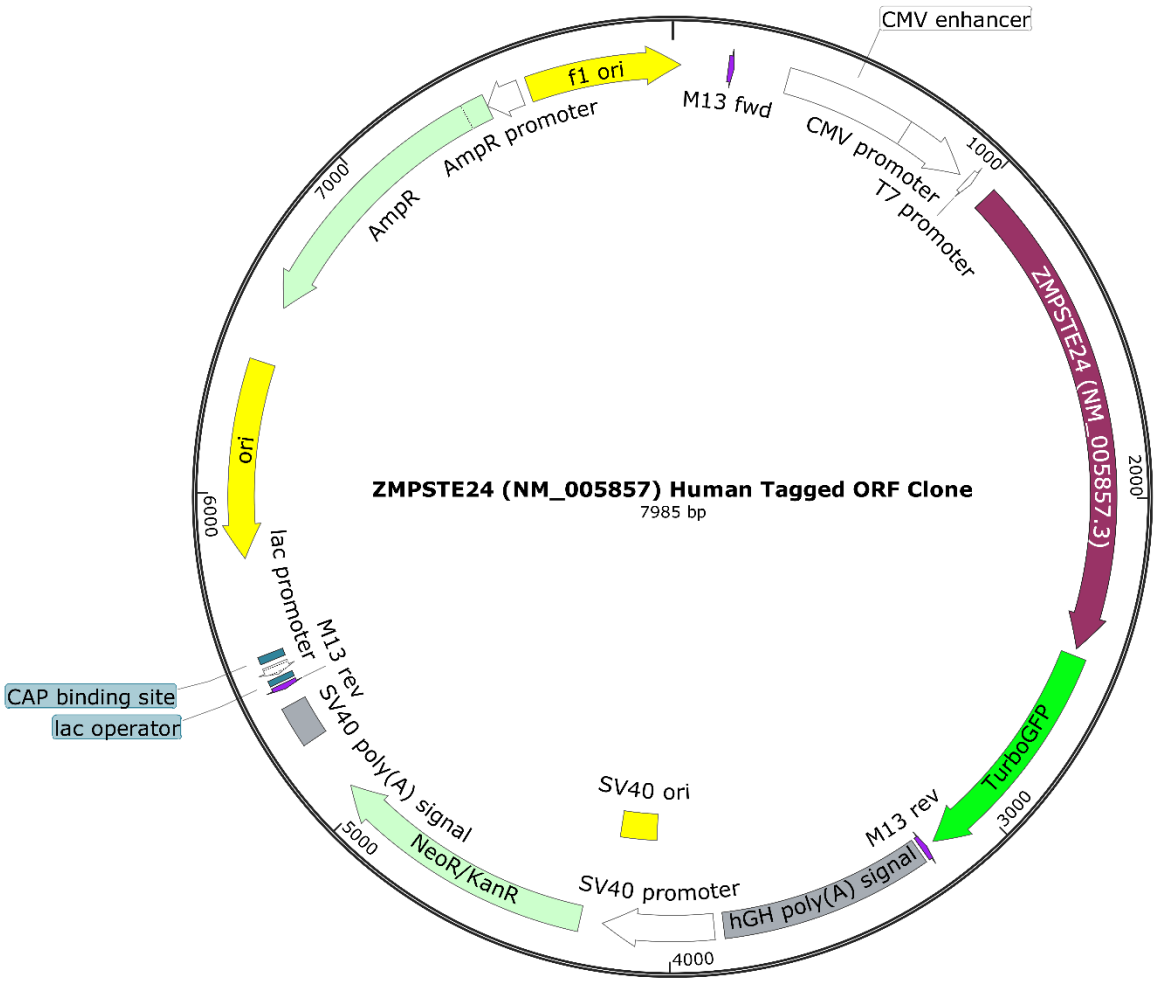

Figure 4. Plasmid map of ZMPSTE24(WT)-GFP.

The expression plasmid "ZMPSTE24 (NM_005857) Human Tagged ORF Clone" purchased from OriGene™ technologies was used as *ZMPSTE24* wild-type construct in this study. The plasmid with a total length of 7985 bp has the *ZMPSTE24* wild-type sequence (purple bar) and a downstream, in-frame *TurboGFP* (green bar) inserted. The plasmid also expresses an ampicillin resistance gene as selection marker in *E. coli* (light green, top). Figure created with SnapGene.

2.1.4.2.1. Mutagenesis

We used the wild-type clone as template for Q5® site-directed mutagenesis to generate the two mutant constructs ZMPSTE24(insA)-GFP and ZMPSTE24(M1_M3del)-GFP. The Q5® site-directed mutagenesis kit protocol (catalog no. E0554S, New England Biolabs®, www.neb.com) was applied for mutagenesis according to the manufacturer's protocol. 20 ng of the WT plasmid was utilized as template for the mutagenesis and an annealing temperature of 72 °C was set for 6 minutes each cycle. The primer sets used are listed in Table 12. The

primer design for engineering the constructs ZMPSTE24(insA)-GFP and ZMPSTE24(M1_M3del)-GFP is illustrated in Figure 5. For mutagenesis of ZMPSTE24(M1_M3del)-GFP the digestion time has been increased to 30 minutes.

Table 12. Primer sets used for Q5® site-directed mutagenesis

Construct	Primer labelling	Sequence (5'-3')	Length	Tm
ZMPSTE24(insA)-GFP	M_insA_f	ATGTGGGAGATGCCGGCCG	19 nt	65 °C
	M_insA_r	AAGCGTCCAGCGATGCCC	18 nt	63 °C
ZMPSTE24(M1_M3del)-GFP	M_M1-M3del _f	TGTGGGCATCGCTGGACG	18 nt	63 °C
	M_M1-M3del _r	ATCGCGGCGGCAGATCTC	18 nt	63 °C

f = forward; r = reverse; nt = nucleotides; Tm = melting temperature

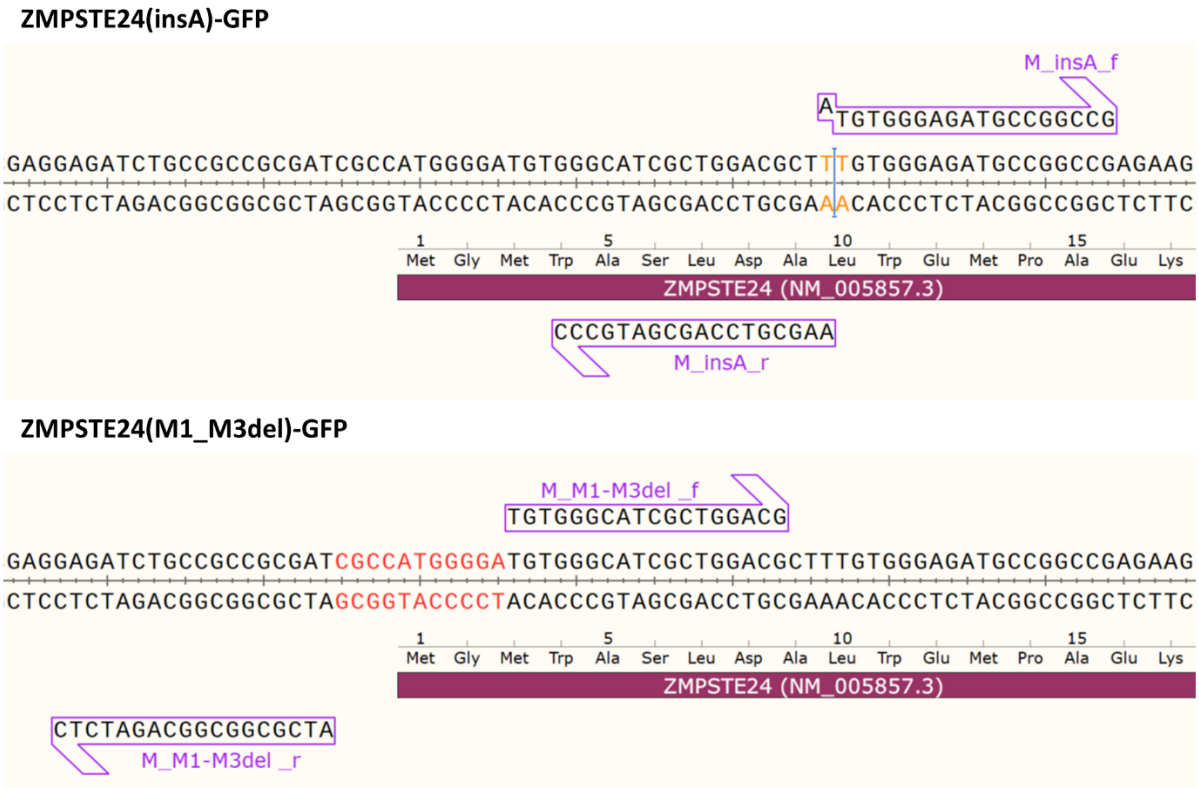

Figure 5. Primer design for Q5® site-directed mutagenesis.

The DNA (above) and protein (below) sequence shown represents the wild-type clone used as template. **ZMPSTE24(insA)-GFP**: The insertion of the nucleotide A is incorporated into the 5´ end of the forward primer (M_insA_f) while the reverse primer (M_insA_r) anneals back-to-back with the 5´ end of the complementary region of the forward primer. Nucleotide A is inserted between the coding nucleotides

28 and 29 (highlighted in orange) at the resulting construct. The construct ZMPSTE24(insA)-GFP thus carries the mutation *ZMPSTE24* (NM_005857):c.28_29insA detected in the families.
ZMPSTE24(M1_M3del)-GFP: Deletions are engineered by designing non-mutagenic forward (M_M1-M3del _f) and reverse (M_M1-M3del _r) primers that flank the region to be deleted (highlighted in red). The resulting construct carries the mutation c.-4_7del, with a deleted canonical start codon and a disrupted methionine (ATG) codon at position 3. Figure created with SnapGene.

An overview of the constructs used and an explanation of the research question to be clarified with each of the constructs can be found in Table 17.

2.1.4.2.2. Transformation and plasmid propagation (approach 2)

The subsequent transformation of the plasmids ZMPSTE24(insA)-GFP and ZMPSTE24(M1_M3del)-GFP in NEB® 5-alpha Competent *E. coli* (catalog no. C2987, New England Biolabs®) was performed based on step III of the Q5® site-directed mutagenesis kit protocol (E0554, New England Biolabs®). However, it is important to note that the protocol was adapted because *ZMPSTE24* plasmids are known to cause toxicity and/or recombination and rearrangement events in *E. coli* at 37 °C (149). Incubation with shaking after heat shock was performed at room temperature for 48-72 hours. The cells were plated on ampicillin-LB plates [100 µg/ml] and incubated overnight at 25 °C. The pUC19 plasmid included in the kit was used as positive control.

A colony PCR followed by Sanger sequencing of the amplified sequence was performed to confirm mutagenesis and transformation of the mutant plasmids. Several colonies were picked, suspended in 50 µl LiChrosolv® water (Merck) and denatured at 95 °C for 5 minutes. The samples were then cooled to 4 °C and incubated on ice for 10 minutes, followed by centrifugation for 1 minute at 14,000 rcf. The supernatant was used for the reaction mix of the colony PCR, which equals the reaction mix of the hot start PCR (see chapter 2.1.4.1.1). The primer set, which is shown in Table 13, was selected to ensure that the amplified sequences span the respective mutation site of the constructs ZMPSTE24(insA)-GFP and ZMPSTE24(M1_M3del)-GFP. The corresponding cycling parameters are listed in Table 14.

Table 13. Primer set used for colony PCR (approach 2)

Primer labelling	Sequence (5'-3')	Length	Tm
VP1.5_f	GGACTTTCCAAAATGTCG	18 nt	51.4 °C
Insert control_r	TTTCAGAATCCATGATCTGT	20 nt	52.3 °C

f = forward; r = reverse; nt = nucleotides; Tm = melting temperature

Table 14. Thermo cycler profile for colony PCR (approach 2)

Temperature	Time	Cycles
94 °C	15 min	1
95 °C	25 sec	33
47 °C	30 sec	
72 °C	1 min	
72 °C	7 min	1
4 °C	hold	

Successful mutagenesis and transformation was confirmed via gel electrophoresis with a 1 % agarose gel, stained with GelRed® (Biotium) and additional Sanger sequencing using the BigDye™ Terminator v1.1 Cycle Sequencing Kit (catalog no. 4337452, Applied Biosystems™/Thermo Fisher Scientific) and a standard protocol. The gel electrophoresis was run at 130 volts and 400 mA for 30 minutes.

2.1.4.2.3. Plasmid isolation (approach 2)

An overnight culture was prepared prior to plasmid isolation. For this purpose, colonies with confirmed plasmid transformation were pre-cultured in 3 ml LB medium with ampicillin [100 µg/ml] and incubated overnight for 24 hours at 25 °C under shaking at 177 rpm. The entire volume of pre-cultured cells was then transferred into 100 ml LB medium with ampicillin [100 µg/ml] for the main culture. The main culture was then incubated overnight for another 24 hours at 25 °C while shaking at 100 rpm. For plasmid isolation the EndoFree® Plasmid Maxi Kit (catalog no. 12362, Qiagen) was utilized according to the manufacturer's protocol. To prepare glycerol stocks, 750 µl of the bacterial suspension was mixed with 250 µl glycerol (60%) and frozen at -80 °C.

2.1.4.2.4. Sequence verification of the generated constructs (approach 2)

To verify that the mutagenesis constructs ZMPSTE24(insA)-GFP and ZMPSTE24(M1_M3del)-GFP only carry the desired mutations, both the entire *ZMPSTE24* gene and the downstream *GFP* gene were sequenced. In addition to the two mutant constructs, the sequence of the wild-type construct ZMPSTE24(WT)-GFP was also verified in this manner. Before sequencing, plasmid-DNA concentrations were measured via NanoDrop™ (Thermo Fisher Scientific). Sanger sequencing was performed using the BigDye™ Terminator v1.1 Cycle Sequencing Kit

(catalog no. 4337452, Applied Biosystems™/Thermo Fisher Scientific) according to the manufacturer's protocol. The reaction mixes for Sanger sequencing of each construct are listed in Table 15.

Table 15. Reaction mixes for Sanger sequencing (approach 2)

Components	ZMPSTE24(WT)-GFP	ZMPSTE24(insA)-GFP	ZMPSTE24(M1_M3del)-GFP
BigDye™ Terminator v1.1 Ready Reaction Mix	0.5 µl	0.5 µl	0.5 µl
5X Sequencing Buffer	1.4 µl	1.4 µl	1.4 µl
Sequencing primer [10 µM]	0.4 µl	0.4 µl	0.4 µl
Plasmid	4.0 µl [100 ng/µl]	0.7 µl [616 ng/µl]	0.9 µl [461 ng/µl]
LiChrosolv® water (Merck)	4.2 µl	7.5 µl	7.3 µl

The plasmid volumes used were calculated as follows: The minimum amount of DNA required was set at 25 ng per 500 bp (equivalent to 0.05 ng/bp) based on personal experience. The wild- type plasmid has a size of 7985 bp. This results in a minimum quantity of: 0.05 ng/bp x 7985 bp = 399 ng. For sequencing of the WT plasmid, 400 ng of plasmid were used [400 ng / 100 ng/µl = 4.0 µl]. Due to the high concentrations of the mutant plasmids, a larger quantity of 430 ng was utilized for sequencing in order to increase the pipetting volume [ZMPSTE24(insA)-GFP: 430 ng / 616 ng/µl = 0.7 µl and ZMPSTE24(M1_M3del)-GFP: 430 ng / 461 ng/µl = 0.9 µl]. The sequencing primers used are shown in Table 16. The same cycling parameters were used as for insert sequencing in approach 1 (see Table 11).

Table 16. Primers for insert sequencing (approach 2)

Primer labelling	Sequence (5'-3')	Length	Tm
VP1.5_f	GGACTTTCCAAAATGTCG	18 nt	51.4 °C
ZMPSTE24_seq mid_r	GAAGAGCGTTTAGATCCTTC	20 nt	56.4 °C
ZMPSTE24_seq mid_f	CTATGCTGATTATATTGCCC	20 nt	54.3 °C
TurboGFP_r	TTTGGTGCTCTTCATCTTGTTG	22 nt	58.2 °C
ZMPSTE24_seq end_f	TTCTGACTGGTTGTTCTCAATG	22 nt	58.4 °C
XL39_r	ATTAGGACAAGGCTGGTGGG	20 nt	60.0 °C

f = forward; r = reverse; nt = nucleotides; Tm = melting temperature

Purification was performed by size exclusion chromatography. Therefore, Centri-Sep™ Spin columns (catalog no. 401763, Invitrogen™/Thermo Fisher Scientific) were loaded with Sephadex® G-50 (Cytiva). The sequencing products were diluted 1:2 (10 µl LiChrosolv® water (Merck) and 10 µl sequencing products) and pipetted onto the Sephadex-columns. The loaded columns were centrifuged for 2 min at 750 rcf. The sequencing products were then transferred to a 96-well plate (catalog no. 710880, Biozym, www.biozym.com) and analyzed using a 3730xl DNA Analyzer (Applied Biosystems™/Thermo Fisher Scientific) with POP-7™ polymer (catalog no. 710880, Applied Biosystems™/Thermo Fisher Scientific) as separation matrix. The raw data generated was analyzed using SeqScape™ Software v3.0 (Applied Biosystems™/Thermo Fisher Scientific) and SnapGene (SnapGene®).

2.1.4.2.5. Additional *ZMPSTE24* expression constructs

The additional mutant *ZMPSTE24* expression constructs ZMPSTE24(insA+M13I)-GFP, ZMPSTE24(insT)-GFP and ZMPSTE24(insT+M13I)-GFP were designed to obtain unambiguous results and to clarify which potential start codon is used. These constructs, with the designed sequences cloned into the pTwist CMV vector (part number: 103190), were ordered from TWIST Bioscience (www.twistbioscience.com). The plasmid map is illustrated in Figure 6.

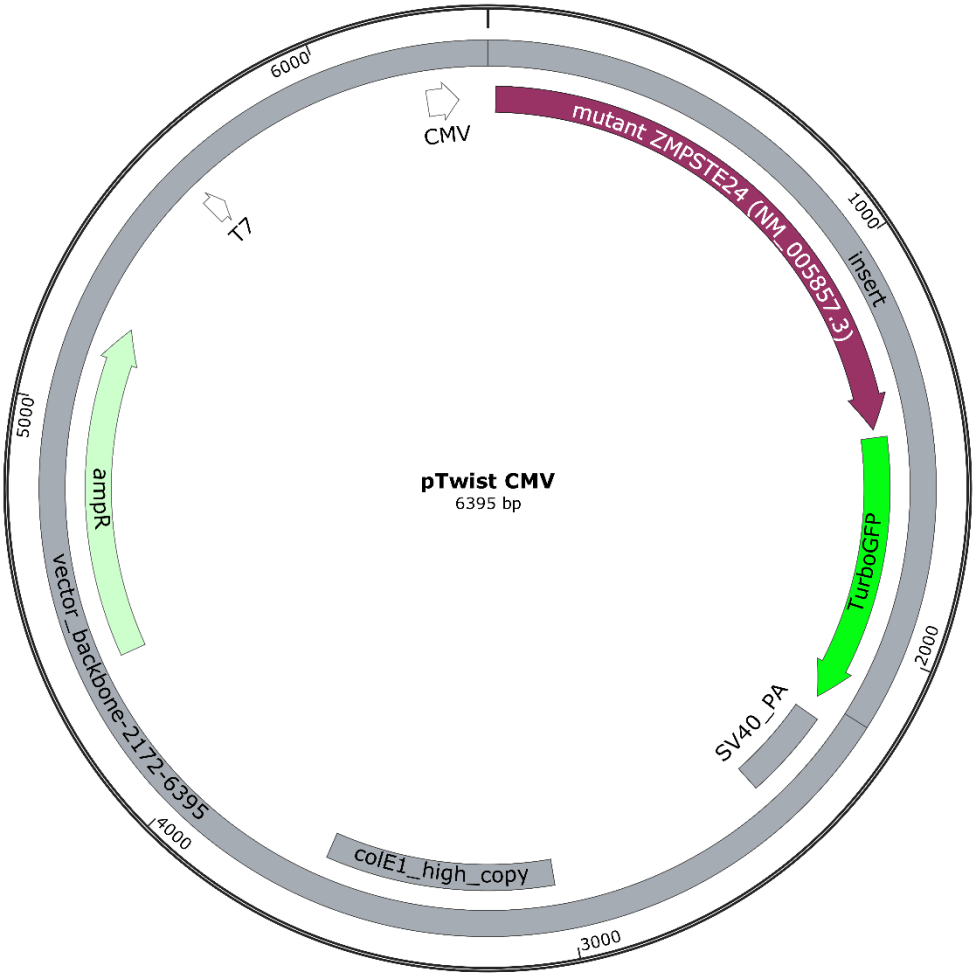

Figure 6. Plasmid map of additional mutant *ZMPSTE24* constructs cloned in pTwist CMV.
The additional expression constructs used were ordered from TWIST Bioscience (www.twistbioscience.com). The mutant *ZMPSTE24* sequence (purple bar) and the downstream *TurboGFP* sequence (green bar) are cloned into the vector pTwist CMV. The vector including insert has a length of 6395 bp. Figure created with SnapGene.

The construct ZMPSTE24(insA+M13I)-GFP carries the mutations c.28_29insA and c.39G>A in the *ZMPSTE24* gene (Figure 7, top). The mutation c.39G>A prevents additional translation initiation at the mutated wild-type methionine codon at amino acid position 13 (M13). Thus, only one potential start codon is present in the 5' region of the gene, which is in-frame with the TurboGFP, i.e. the novel ATG-codon gained due to the studied mutation c.28_29insA. The methionine codon (ATG) at this position was changed to an isoleucine-codon (ATA) due to similar physicochemical properties of these amino acids. Protein stability and correct folding should be maintained in case of alternative translation initiation at the novel ATG.

The construct ZMPSTE24(insT)-GFP carries the mutation c.28_29insT in *ZMPSTE24*. The insertion of a deoxythymidine (T) at the same nucleotide position as the studied mutation (c.28_29insT) also leads to a frameshift, but without creation of a novel ATG. The ATG-codon of wild-type M13 is not mutated in this construct and serves as potential alternative start codon (Figure 7, middle).

The construct ZMPSTE24(insA+M13I)-GFP carries the mutations c.28_29insT and c.39G>A in *ZMPSTE24* in order to create a frameshift without potential alternative N-terminal start codons in-frame with TurboGFP (Figure 7, bottom). This construct serves as a negative control in this study. An overview of the constructs used and an explanation of the research question to be clarified with each of the constructs can be found in Table 17.

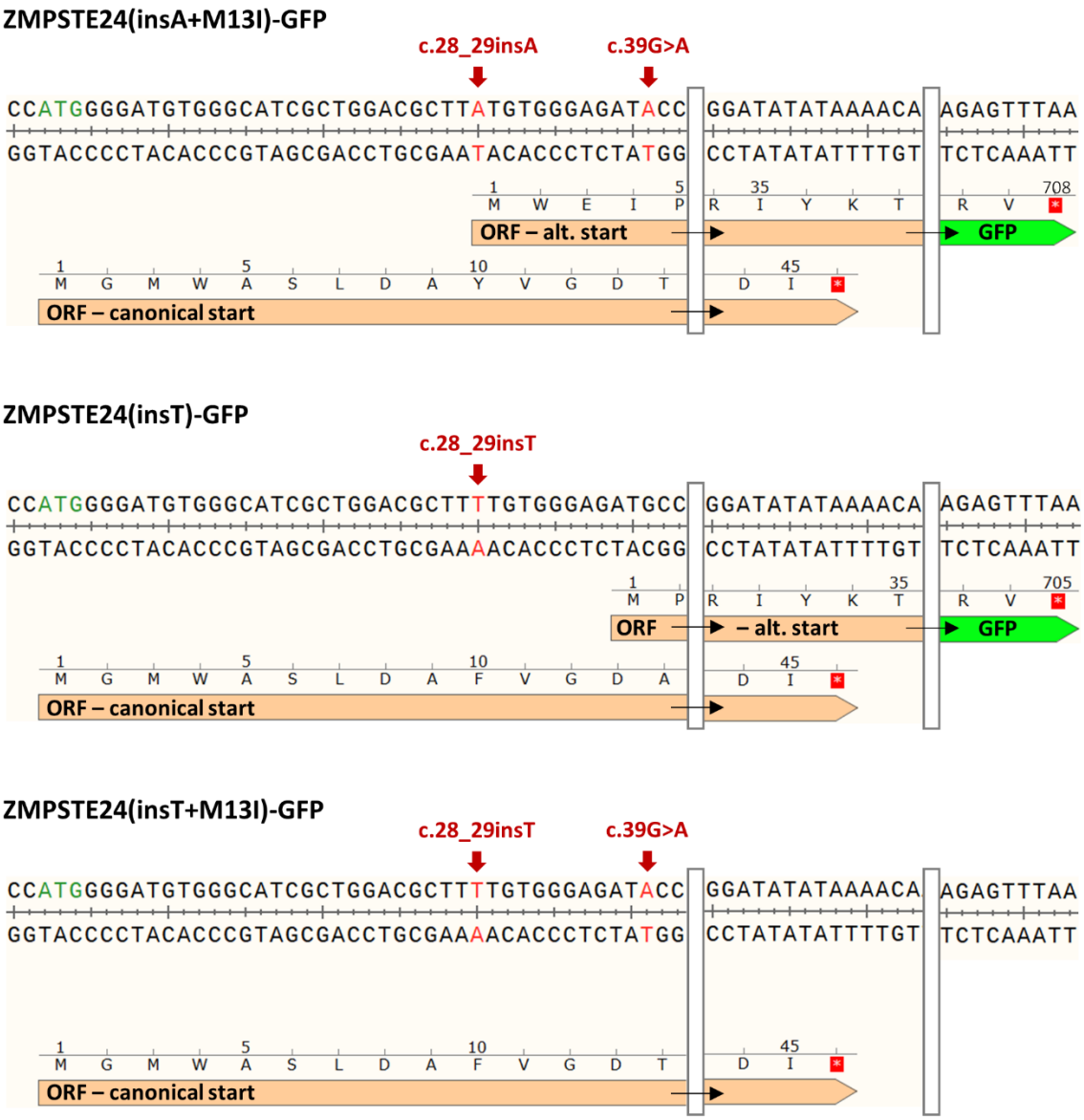

Figure 7. Mutations in constructs ZMPSTE24(insA+M13I)-GFP, ZMPSTE24(insT)-GFP and ZMPSTE24(insT+M13I)-GFP.

The DNA sequences with mutation sites marked in red are shown. Below, the N-terminal open reading frames (ORFs) of *ZMPSTE24* are illustrated as orange bars. An ORF with in-frame, downstream TurboGFP is indicated as adjacent green bar. The white vertical bars represent intermediate sequence areas that are not shown. The canonical start codon is displayed in green and the corresponding truncated ORF due to the frameshift variant is shown below. **ZMPSTE24(insA+M13I)-GFP** harbors one potential N-terminal alternative start codon at the mutation site of the studied mutation c.28_29insA, which is in-frame with TurboGFP. In **ZMPSTE24(insT)-GFP**, the only potential alternative start codon in the N-terminus of *ZMPSTE24* in-frame with TurboGFP corresponds to the wild-type M13 codon. The construct **ZMPSTE24(insT+M13I)-GFP** does not contain a potential start codon in the N-terminal region that is in-frame with TurboGFP due to the two mutations c.28_29insT and c.39G>A. This construct therefore serves as negative control in the functional analysis.

2.1.4.3. Overview of constructs used for expression experiments

Table 17 provides an overview of the constructs used in the functional analysis, as well as the associated research question to be answered.

Table 17. Overview of expression constructs used for functional analysis

[Table and legend adapted from supporting information in Clinical Genetics (Kaufmann et al., 2023)]

Name of construct	Sequence/ Mutation [NM_005857.3]	Potential N-terminal start codons in-frame with GFP-tag	Research question to be clarified
ZMPSTE24(WT) - GFP	Wild-type sequence	M1-codon †	Positive control: localization at the ER is expected.
ZMPSTE24(insA) - GFP	c.28_29insA	Novel ATG and M13-codon	Does alternative translation initiation occur in mutated *ZMPSTE24*, resulting in ER-located protein?
ZMPSTE24(M1_M3del) - GFP	c.-4_7del	M13-codon ‡	Is the M13-codon a potential translation initiation site?
ZMPSTE24(insA+M13I) - GFP	c.28_29insA; c.39G>A	Novel ATG	Does the novel ATG serve as alternative translation initiation site in mutated *ZMPSTE24*?
ZMPSTE24(insT) - GFP	c.28_29insT	M13-codon	Does the M13-codon serve as alternative translation initiation site in mutated *ZMPSTE24*?
ZMPSTE24(insT+M13I) - GFP	c.28_29insT; c.39G>A	None	Negative control: no ER-localization is expected.

† Translation initiation at the canonical start codon is confirmed by protein sequencing of wild-type ZMPSTE24 published in ProteomicsDB (192, 193).

‡ The M1- and M3-codons, which might act as "stronger" start codons, are absent in this construct.

2.1.4.4. Transfection

AD-293 cells (part number: 240085, Agilent Technologies, www.agilent.com) were seeded in 6-well plates on 30 mm cover slips. For transfection of the constructs and the RFP marker for the endoplasmic reticulum (mCherry-ER-3, plasmid #55041) the Neon® Transfection System (catalog no. MPK5000, Invitrogen™/Thermo Fisher Scientific) and also the PolyJet™ In Vitro

DNA Transfection Reagent (catalog no. SL100688, SignaGen Laboratories, www.signagen.com) were used according to the respective manufacturer's protocol.
For the transfection of approximately 200,000 cells using the Neon® Transfection System, 4.0 µg of each construct and 2.5 µg of the ER marker were utilized. Electroporation was performed with 100 µl Neon® Tips and the parameters listed in Table 18.

Table 18. Electroporation parameters for the Neon® Transfection System

Pulse voltage	Pulse width	Pulse no.
1,100 V	20 ms	2

For transfection with PolyJet™ In Vitro DNA Transfection Reagent, 3 µl PolyJet reagent were mixed with 1 µg plasmid DNA in 100 µl of DMEM devoid of serum and antibiotics. The transfection mixture was added to 1 ml of culture medium for 8 hours and was then replaced with 2 ml culture medium.

Imaging was performed 24 - 48 h after transfection.

2.1.4.5. Confocal laser scanning microscopy

AD-293 cells were imaged in loading buffer, which was composed of 135 mM NaCl, 5 mM KCl, 2 mM $CaCl_2$, 1 mM $MgCl_2$, 10 mM HEPES, 2.6 mM $NaHCO_3$, 0.44 mM KH_2PO_4, 0.34 mM Na_2HPO_4, 10 mM D-glucose, 2 mM L-glutamine, 1X MEM amino, 1X MEM vitamins, 1% penicillin-streptomycin and 1% Amphotericin B; pH was adjusted to 7.45 with NaOH.
High-resolution imaging was performed with an array confocal laser scanning microscope (Axiovert 200 M, Zeiss, www.zeiss.com) with a 100×/1.45 NA oil immersion objective (Plan-Fluor, Zeiss) and a Nipkow-based confocal scanner unit (CSU-X1, Yokogawa Electric Corporation, www.yokogawa.com). Laser light of diode lasers (Visitron Systems, www.visitron.de) served as excitation light source: ZMPSTE24-GFP fusion constructs were excited with a 488 nm laser and mCherry-ER-3 was excited with a 561 nm laser. Emission light was captured with a CoolSNAP HQ2 CCD Camera (Teledyne Photometrics, www.photometrics.com) using the emission filters and ET630/75m and ET525/36m (Chroma Technology Corporation, www.chroma.com). Image analysis was performed with Fiji software (194).

2.2. *MN1* study

2.2.1. Clinical diagnosis (*MN1*)

The initial referral of the pregnant woman and the findings of the sonographic examination of the fetus were provided by the Department of Obstetrics and Gynaecology at the University Hospital of Graz, Austria.

2.2.2. Genetic analyses (*MN1*)

The genetic analyses listed below were carried out as part of the diagnostic evaluation. DNA isolation was carried out with phenol-chloroform from chorionic villus using a standard protocol.

2.2.2.1. SNP Array

Copy Number Variation analysis was performed using the Infinium CytoSNP-850K (Illumina, www.illumina.com) in combination with the BlueFuse™ 4.5 Multi Analysis Software (Illumina). The data obtained were also analyzed using the UCSC Genome Browser (www.genome.ucsc.edu) (191) and the DECIPHER database (88, 89).

2.2.2.2. Quantitative real-time PCR

For segregation analysis in the parents of the fetus and to confirm the size of the deletion in the chorionic villus sample, a quantitative real-time PCR (qRT-PCR) was performed using a 7500 Fast Real-Time PCR System (Applied Biosystems™/Thermo Fisher Scientific). Region-specific primer sets and one primer set for the internal control (Table 19) were used for real-time PCR amplification. A primer set within the gene *NF2* was used as endogenous control. The amplicon length for each primer set was approximately 100 bp. The fetal tissue sample obtained after termination of pregnancy was also checked for the presence of the deletion via qRT-PCR in order to rule out confined placental mosaicism.

Table 19. Primer sets for qRT-PCR in segregation analysis

Primer labelling	Sequence (5'-3')	Length	Tm
CPMER_I2_f (outside the deleted region)	CGCCTCATAAACCCTCTGC	19 nt	60.7 °C
CPMER_I2_r (outside the deleted region)	CCTGTCCTGAGTCCATGAGG	20 nt	60.7 °C

CPMER_E1_f (within the deleted region)	TGGTTGAATGAATGTACCTTAGGAG	25 nt	60.6 °C
CPMER_E1_r (within the deleted region)	CATGCTGGGATTGATTTGTG	20 nt	59.9 °C
Intergenic_f (within the deleted region)	CAGTCTTCGGGGAGAGGAG	19 nt	59.9 °C
Intergenic_r (within the deleted region)	TCTGCTTCTTCTCTGCATGG	20 nt	59.3 °C
MN1_E2_f (within the deleted region)	CCCATTATGGTGTCCCTCAG	20 nt	60.2 °C
MN1_E2_r (within the deleted region)	GATGCTGAGGCCTTGTTTG	19 nt	59.4 °C
MN1_I1_f (within the deleted region)	CACACAGCAAGCAAGGTGAC	20 nt	60.5 °C
MN1_I1_r (within the deleted region)	CCACCTGCTTCCTAGCACAC	20 nt	60.9 °C
MN1_E1_f (outside the deleted region)	GCAGCATGGTGTGTTCTTTG	20 nt	60.3 °C
MN1_E1_r (outside the deleted region)	CTGGGGAGGCTGCATTAAC	19 nt	60.6 °C
PITPNB_E9_f (outside the deleted region/control)	GTGGTGGGGACTGCAAAG	18 nt	60.1 °C
PITPNB_E9_r (outside the deleted region/control)	AATAGCTAGCTCCGATCTCATCC	23 nt	60.2 °C

f = forward; r = reverse; nt = nucleotides; Tm = melting temperature

2.2.2.3. Trio-Whole exome sequencing

Trio-Whole exome sequencing (Trio-WES) was performed in addition to exclude the possibility of other pathogenic variants in protein coding genes that could be the cause of the fetal phenotype abnormalities. Trio-WES involves sequencing of all protein-coding regions of the genome as well as the mitochondrial DNA with subsequent phenotype-based screening and comparison with the parents' data. Variant prioritization was based on the HPO terms: Intrauterine growth retardation (HP:0001511), Decreased thalamic volume (HP:0012695), Encephalocele (HP:0002084), Intracranial cystic lesion (HP:0010576), Aplasia/Hypoplasia of the brainstem (HP:0007362) and Ventriculomegaly (HP:0002119). The specific regions were enriched using Twist Human Core Exome and Twist Mitochondrial Panel (TWIST Bioscience, www.twistbioscience.com). Sequencing was performed on the NextSeq 550 Sequencing System (Illumina, www.illumina.com), followed by variant annotation and analysis using VarSeq (Golden Helix, www.goldenhelix.com).

2.2.2.4. RNA analysis

Qualitative (sequencing) and quantitative (qRT-PCR) RNA analyses were performed. The RNA was isolated from chorionic villus culture using TRIzol reagent (Invitrogen™/Thermo Fisher Scientific, www.thermofisher.com). Some of the cultures were pre-treated with puromycin solution (catalog no. P8833, Sigma Aldrich, www.sigmaaldrich.com) to inhibit nonsense-mediated mRNA decay (NMD), according to the standard operating procedure (SOP) “RNA Analyse nach KurzzeitKul_Trizol_isol” of our institute.

For qualitative analysis, the QuantiTect Reverse Transcription Kit (catalog no. 205311, Qiagen, www.qiagen.com) was used for cDNA synthesis according to the manufacturer's protocol. Subsequent amplification of the cDNA was performed using AmpliTaq Gold™ 360 Master Mix (catalog no. 4398881, Applied Biosystems™/Thermo Fisher Scientific) according to the SOP "AmpliTaq Gold 360 MasterMix PCR" of our institute. The primer pairs listed in Table 20 were used for Sanger sequencing of the cDNA samples: “RNA_NMD not inhibited” (without prior puromycin treatment), “RNA_NMD inhibited” (with prior puromycin treatment) and the Human Reference RNA (catalog no. 750500, Agilent, www.agilent.com), which was utilized as control sample.

Table 20. Primer pairs for Sanger sequencing in qualitative RNA analysis

Primer labelling	Sequence (5'-3')	Length	Tm
sp_MN1_E1_v2_f [positive control]	AGGCAGCAGTTCAGCATCTC	20 nt	60.7 °C
sp_MN1_E1_r [positive control]	CCAACGTCTTGTCGTCGTC	19 nt	60.3 °C
sp_MN1_E1_f	GTGGCAAGAAGGGTGAGTG	19 nt	59.2 °C
sp_CPMER_E2_r	ACGCTGCCAAAGGTCTTTTC	20 nt	61.7 °C
sp_MN1_E1_f	GTGGCAAGAAGGGTGAGTG	19 nt	59.2 °C
sp_CPMER_E5_r	AGCTAGGGCTGGCTGGTC	18 nt	60.5 °C
sp_CPMER_E2_f	CCCCAGCGCTTAATTATTTC	20 nt	58.7 °C
sp_CPMER_E3_r	GACATTCTGGGTCCATCTCC	20 nt	59.3 °C
sp_CPMER_E3_f	GAACTGGCCAAATGTGAAGG	20 nt	60.5 °C
sp_CPMER_E4_r	TCACAGTGGTGTTGGTCTCAC	21 nt	59.6 °C
sp_CPMER_E2_f	CCCCAGCGCTTAATTATTTC	20 nt	58.7 °C
sp_CPMER_E4_r	TCACAGTGGTGTTGGTCTCAC	21 nt	59.6 °C

sp = sequencing primer; f = forward; r = reverse; nt = nucleotides; Tm = melting temperature

Primer set “sp_MN1_E1_v2_f” and “sp_MN1_E1_r” was used as positive control, as exon 1 of the *MN1* gene is not affected by the deletion. The wild-type sequence of this region in exon 1 was confirmed in both samples RNA_NMD not inhibited and RNA_NMD inhibited, as well as in the control sample. The remaining primer sets were used to detect possible fusion genes and their expression. In addition, information on splicing is obtained. For protein sequence alignment the analysis tool Clustal Omega (188) was used.

For the quantitative RNA analysis, the RNA was quantified using the Qubit™ RNA Broad Range Assay Kit (catalog no. Q10210, Invitrogen™/Thermo Fisher Scientific): RNA_NMD not inhibited = 145 ng/µl and RNA_NMD inhibited = 179 ng/µl. The integrity of the RNA was then checked using the Bioanalyzer RNA 6000 pico assay (catalog no. 5067-1513, Agilent) according to the manufacturer's protocol and the 18S and 28S RNA peaks were visible.

Subsequently, cDNA synthesis was conducted by reverse transcription with the Omniscript RT Kit (catalog no. 205113, Qiagen) according to the manufacturer's protocol, using 870 ng of RNA_NMD not inhibited and 895 ng of RNA_NMD inhibited.

The qRT-PCR analysis of the cDNAs RNA_NMD not inhibited, RNA_NMD inhibited and the Human Reference RNA (catalog no. 750500, Agilent) was performed with the StepOnePlus™ Real-Time PCR System (catalog no. 4376600, Applied Biosystems™/Thermo Fisher Scientific), using a standard protocol with 40 cycles, as well as melting curve analysis. The amplicon of exon 1 of *MN1* was used as a comparative control for the quantitation of comparative Ct (delta-delta Ct values) analysis. A primer pair within the housekeeping gene *GAPDH* was used as endogenous control. The primer sets used for quantitative RNA (cDNA) analysis is listed in Table 21. The amplicon length for each primer set was approximately 100 bp.

Table 21. Primer pairs for qRT-PCR in quantitative RNA analysis

Primer labelling	Sequence (5'-3')	Length	Tm
rt_MN1_E1_f	CCAACGTCTTGTCGTCGTC	19 nt	60.3 °C
rt_MN1_E1_r	AGCACCATTGACCTGGACTC	20 nt	60.1 °C
rt_MN1_E2_f	CACAGACGACGTGGGTGAC	19 nt	61.3 °C
rt_MN1_E2_r	AGCCACGAATGTCCCAAATC	20 nt	62.2 °C
rt_MN1_E1_v2_f	GACGACGACAAGACGTTGG	19 nt	60.3 °C
rt_MN1_E2_v2_r	GATGCTGAGGCCTTGTTTG	19 nt	59.4 °C
rt_MN1_E1_v3_f [fusion transcript 1]	CTGGGGTCAGGTCTTCAGTG	20 nt	60.7 °C
rt_CPMER_E2_r [fusion transcript 1]	GACGACGACAAGACGTTGG	19 nt	60.3 °C

rt = RT-PCR primer; f = forward; r = reverse; nt = nucleotides; Tm = melting temperature

3. Results

3.1. *ZMPSTE24* study

In this study, the genetic cause of the progeroid phenotype in affected individuals from two Pakistani families is investigated. Both families were recruited from the same province of Pakistan.

3.1.1. Family 1

The pedigree of family 1 consists of four generations including a single consanguineous loop. In the fourth generation there are four affected individuals (three males and one female), of whom two were alive at the time of the medical examination (Figure 8A). One affected individual died at the age of 2, the other at the age of 20. The cause of death is not known in each case.

The two patients examined presented a progeroid appearance (Figure 8B) and looked older than their peers. The patients showed short stature, facial dysmorphism including severe micrognathia, bulging eyes, prominent cheeks, mild double chin, low-set ears and crowded teeth with protruding incisors, as well as skeletal anomaly (bulbous distal phalanges consistent with acro-osteolysis), sparse scalp hair and joint stiffness associated with difficulty in walking. The skin texture appeared rough, dry and thin, indicating atrophy of the skin. On the forehead, dilated blood vessels were visible. No additional ocular, renal, neurologic or otorhinolaryngologic findings were evident.

Additional skeletal radiography performed on patient IV:7 of family 1 revealed mandibular hypoplasia, widely opened metopic and sagittal suture, a rudimentary right clavicle, dysplastic humeral heads, dysplastic acromion and severe osteoporosis, among other findings (Figure 8C).

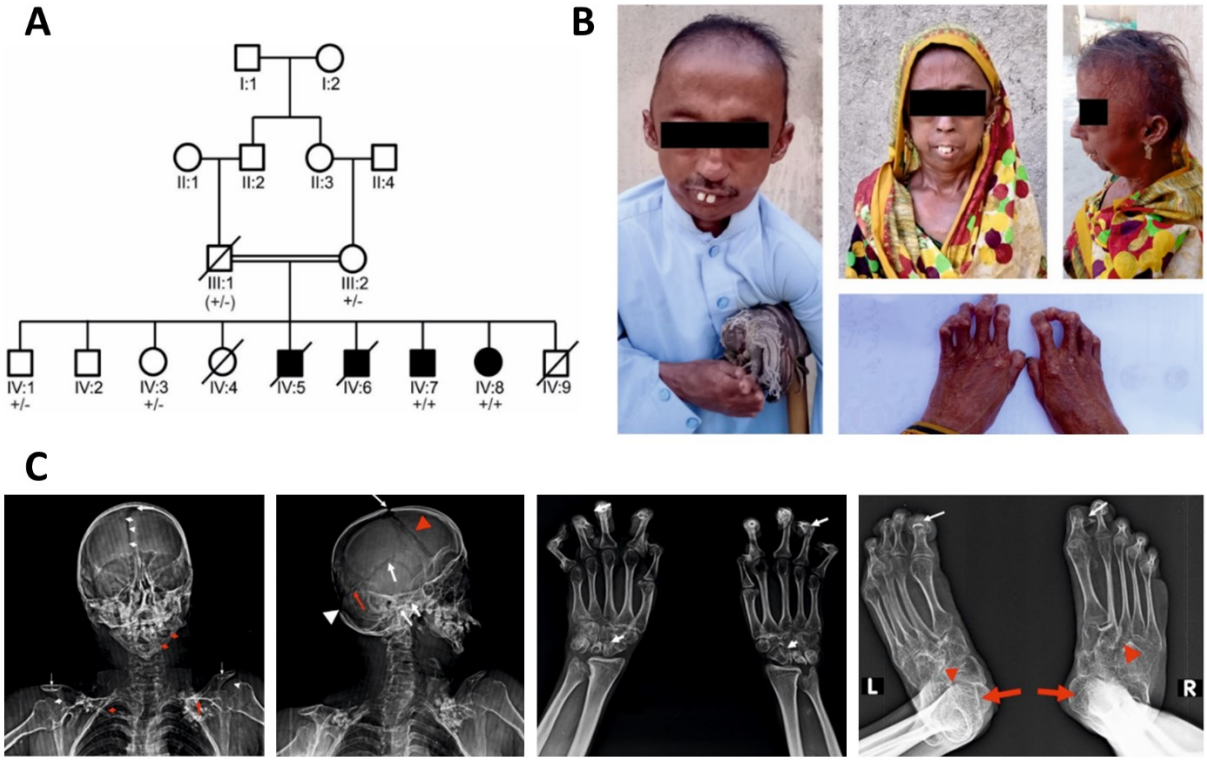

Figure 8. Pedigree, patient images and radiographs of family 1.

[Figure and legend adapted from Gene (Schaflinger & Blatterer et al., 2022)]

(A) Pedigree of consanguineous family 1 with four affected individuals in the last generation. Mutation status of the studied mutation is indicated as "+/+" (homozygous) and "+/−"(heterozygous), with one inferred genotype in the third generation shown in brackets.

(B) Patient images of affected individuals IV:7 and IV:8 with progeroid appearance. The male patient was 25 years old at the time of the examination. The female patient was 23 years of age.

(C) X-ray images of patient IV:7. Anteroposterior (AP) skull radiograph (far left) shows mandibular hypoplasia (red arrow heads), widely opened metopic and sagittal suture (white arrow heads). The upper torso shows rudimentary right clavicle (red arrow head), aplasia of the left clavicle which is replaced by a calcified fibrous band (red arrow), dysplastic humeral heads (white arrow heads) and dysplastic acromion (white arrow). Lateral skull radiograph (2nd left) shows delayed closure of the anterior (thin white arrow) and the posterior (white arrow head) fontanelles. Widely opened coronal (red arrow head), squamosal (white arrow) and the lambdoid (red arrow) sutures respectively. The skull base shows hyperostosis (two white arrows). AP hand radiographs (2nd right) show osteoporosis and features of osteolysis along the carpal bones (small white arrows) and shortened fingers associated with massive flexion deformities of the terminal phalanges (white arrows). AP foot radiographs (far right) show bilateral and symmetrical massive flexion deformities of the terminal phalanges lateral deviation (white arrows), extensively narrowed tarsal and metatarsal joint spaces (red arrow heads) and overwhelming osteoporosis associated with flexion contractures (red arrow).

3.1.2. Family 2

Family 2 is a large, consanguineous family from Pakistan. The simplified pedigree shows six affected individuals (four males and two females) in the last generation. At the time of medical examination, three of the affected individuals were alive (Figure 9A).

Affected individuals showed phenotypic abnormalities consistent with those observed in family 1. The medical observations included: aged appearance, skin atrophy with signs of hyperkeratosis, sparse hair, malformed ears, puffy eyelids, micrognathia, full cheeks, camptodactyly of fingers and toes, long and proximally situated thumbs with obvious clubbing and severe muscular wasting reflected via loss of the normal definitions of the shoulder and elbow joints. One patient was more severely affected with additional findings of massive submental fat and apparent total loss of spine biomechanics due to thoraco-lumbar kyphosis in connection with an ankylosed vertebral column. This patient also presented skin atrophy associated with massive muscle wasting in addition to bone osteolysis above the ankle joints (Figure 9B, left patient images). Affected individuals III-2, -3, and -7 deceased between the ages of eight months and ten years, with unknown causes of death, respectively.

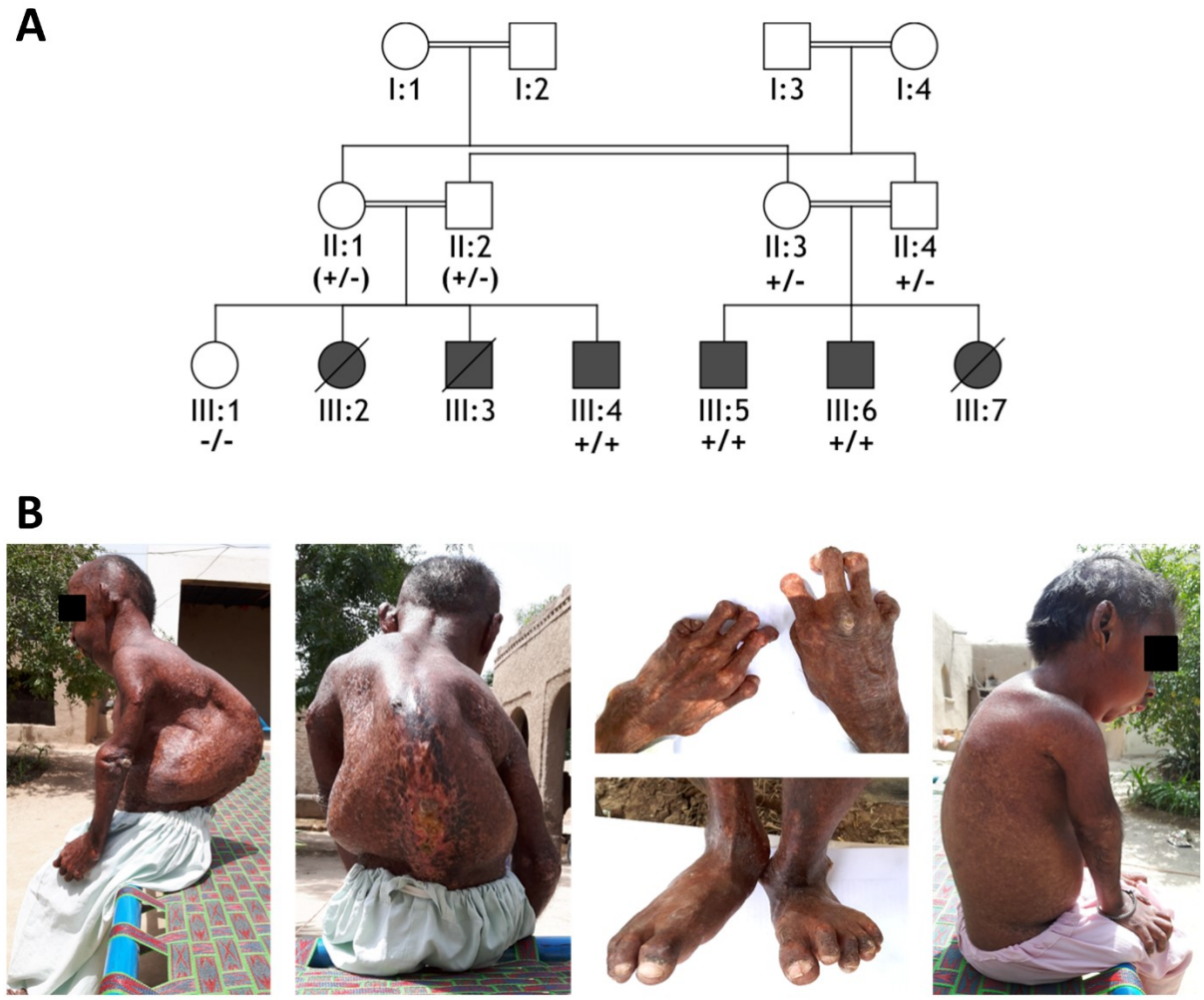

Figure 9. Pedigree and patient images of family 2.

[Figure and legend adapted from Clinical Genetics (Kaufmann et al., 2023)]

(A) Simplified Pedigree of the large, consanguineous family 2 with six affected individuals in the last generation. Mutation status of the studied mutation is indicated as "+/+" (homozygous) and "+/−"(heterozygous), with inferred genotypes shown in brackets.

(B) Photographic images of affected individuals III:4 and III:5 with progeroid appearance. The patient in the left images is more severely affected and displays thoraco-lumbar kyphotic deformities.

3.1.3. Genetic analyses

Genetic analyses of the two families was carried out at different times and a possible link was not yet known at the time of the evaluation. Genetic analysis of family 1 was conducted first, where homozygosity mapping revealed a 37.8 Mb homozygous by descent region on chromosome 1 (chr1:39764664-77558599 [GRCh38/hg38]). Using whole exome sequencing, a previously unknown, homozygous frameshift mutation NM_005857.5:c.28_29insA, p.(Leu10Tyrfs*37) in the *ZMPSTE24* gene, which is located in the autozygous region, was

detected. Genotyping of the variant in available family members confirmed segregation of the novel mutation with the recessive progeroid phenotype (Figure 10A).
In family 2, a 21.1 Mb autozygous region on chromosome 1 (chr1:26092274-47189052 [hg38]) was detected in an affected individual by SNP array analysis. The *ZMPSTE24* gene, located within this region, was considered a candidate gene based on phenotype similarities to family 1, and Sanger sequencing revealed the same homozygous frameshift mutation NM_005857.5:c.28_29insA, p.(Leu10Tyrfs*37). The novel mutation also segregated with the disease phenotype and a recessive inheritance in family 2 (Figure 10B).

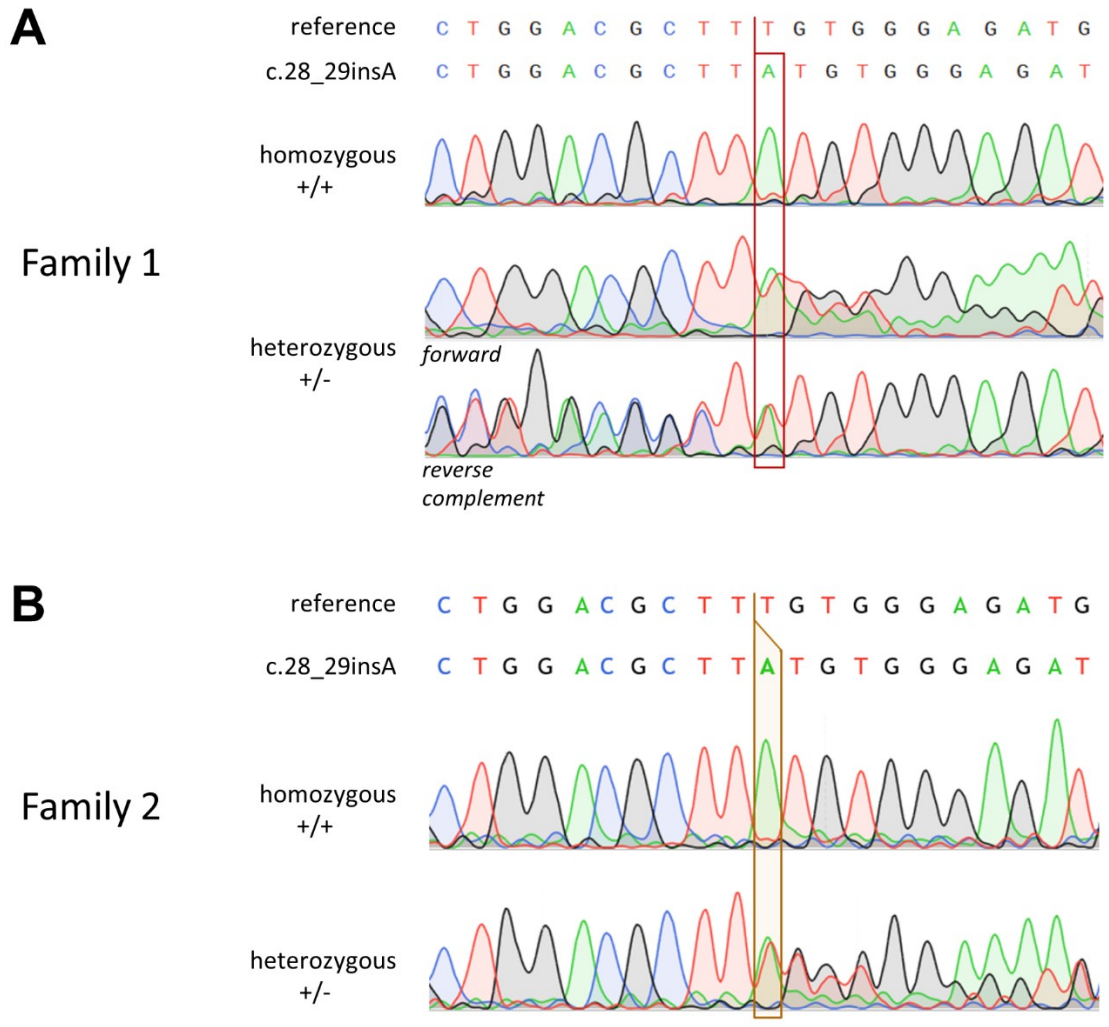

Figure 10. Electropherograms of homozygous and heterozygous *ZMPSTE24* mutation carriers in the two studied families.

[Figure and legend adapted from Gene (Schaflinger & Blatterer et al., 2022) and Clinical Genetics (Kaufmann et al., 2023)]

(A) Sanger sequencing of genomic DNA in homozygous and heterozygous individuals of family 1. The detected mutation NM_005857.5:c.28_29insA, p.(Leu10Tyrfs*37) in *ZMPSTE24* is framed in red. In heterozygous mutation carriers, both forward and reverse complement sequences are shown.
(B) Sanger sequencing of genomic DNA in homozygous and heterozygous individuals of family 2. The detected mutation NM_005857.5:c.28_29insA, p.(Leu10Tyrfs*37) in *ZMPSTE24* is framed in light brown.

To verify whether the two families are distantly related, a comparative haplotype analysis was performed with one affected individual from each family. The analysis revealed a common disease haplotype encompassing chromosomal region 1p34.3-p33, including *ZMPSTE24* (Figure 11). The result confirms a distant relationship between the families, suggesting a founder effect for this mutation in the area inhabited by the families in Pakistan.

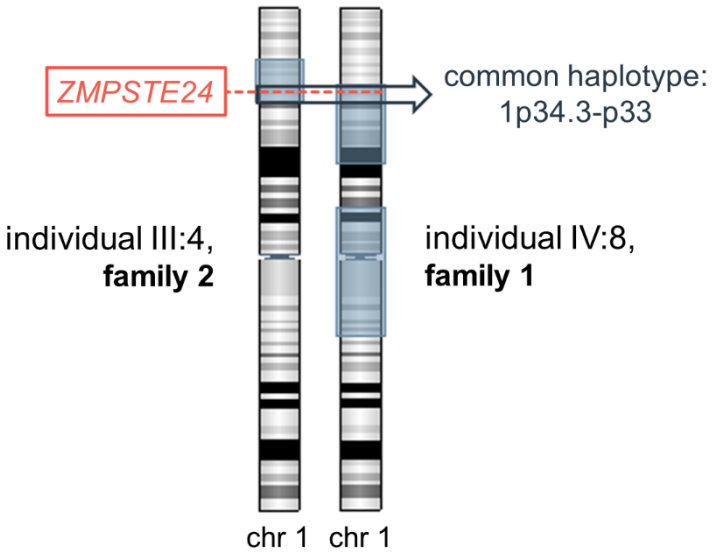

Figure 11. Haplotype analysis of affected individuals from both families.
[Figure and legend adapted from Clinical Genetics (Kaufmann et al., 2023)]
Schematic representation of runs of homozygosity, shown as blue blocks, at chromosome 1 of affected individual III:4 of family 2 and affected individual IV:8 of family 1. The *ZMPSTE24* gene (dashed orange line) is located within the common haplotype region 1p34.3-p33 (dark blue arrow), which confirms a distant relatedness between the families.

3.1.4. Genotype-phenotype discrepancy and in silico analyses

The homozygous frameshift mutation c.28_29insA, p.(Leu10Tyrfs*37) in *ZMPSTE24* detected in affected individuals of both families is associated with a complete loss-of-function and is also predicted to result in nonsense mediated mRNA decay (Figure 12: p.(Leu10Tyrfs*37)). Based on this prediction, the mutation would be associated with lethal restrictive dermopathy (RD) due to the lack of enzyme activity of the ZMPSTE24 protein. However, the phenotype observed in affected members of both families is consistent with mandibuloacral dysplasia with type B lipodystrophy (MADB), which represents the second *ZMPSTE24*-associated disease. In contrast, MADB is known to be caused by biallelic *ZMPSTE24* mutations that preserve residual proteolytic activity, leading to a discrepancy between genotype and phenotype for the detected mutation. In silico splice site prediction indicated no splicing effect for the detected mutation. Subsequent in-depth analysis of the mutation revealed that a novel potential in-frame start codon (referred as “novel ATG” in this work) is created at the mutation site. If this alternative start codon was used as translation initiation site (TIS), the resulting protein would presumably be only 9 N-terminal amino acids shorter than the wild-type protein (Figure 12: Mut_altStart). As a result, a complete loss of protein function could possibly be avoided. To further investigate this hypothesis, in silico TIS predictions were performed. The in silico prediction indicated that in addition to the novel ATG, the wild-type methionine 13 (M13) codon could also function as potential TIS (Table 22). In this case, the resulting protein would presumably be 12 amino acids shorter at the N-terminus than the wild-type protein (Figure 12: Mut_altStart). The remaining downstream methionine codons of the wild-type protein (e.g. M59) could be excluded as functional translation initiation sites based on the mutations associated with RD in the literature.

Table 22. In silico TIS predictions in c.28_29insA mutated *ZMPSTE24*

NetStart 1.0		ATGpr	
TIS	**Score**	**TIS**	**Reliability**
Novel ATG	0.582	M59	0.16
M13	0.540	Novel ATG	0.10
M1	0.465	M1	0.09
M59	0.360	M13	0.06

TIS = translation initiation site; Web-Tools: NetStart 1.0 (186), ATGpr (187)

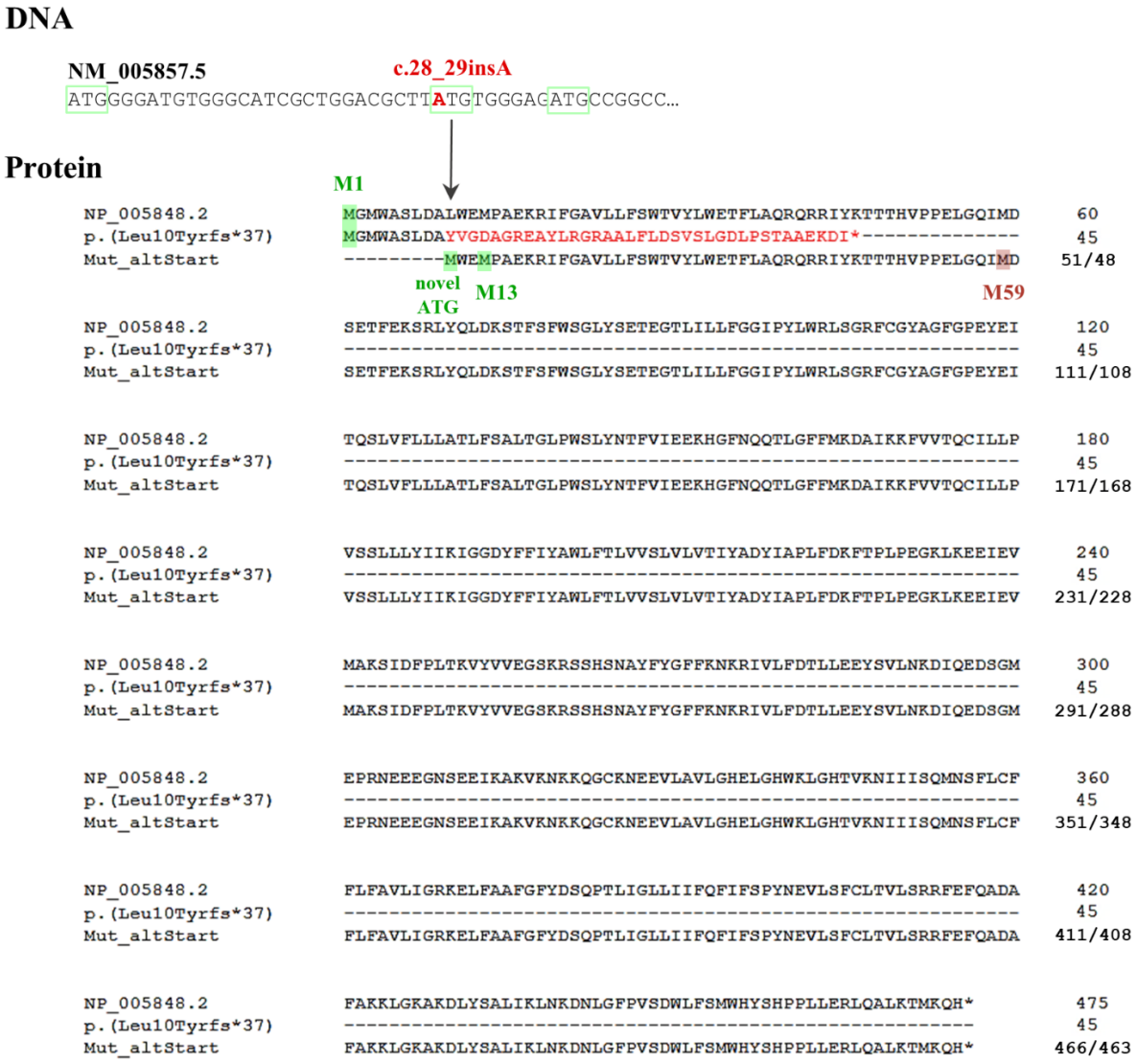

Figure 12. Protein sequence alignment of wild-type and mutated *ZMPSTE24.*

[Figure and legend adapted from Gene (Schaflinger & Blatterer et al., 2022)]

5' DNA coding sequence of the *ZMPSTE24* gene carrying the mutation NM_005857.5:c.28_29insA (highlighted in red) is shown above. The potential start codons are framed in light green. Below, the protein sequence of the wild-type protein NP_005848.2, the predicted frameshift sequence p.(Leu10Tyrfs*37) and the sequence that would be obtained by alternative translation initiation at the novel ATG or M13 are shown. The translation start sites of the respective protein sequences are highlighted in light green. Aligned protein sequences are displayed in single letter code with amino acid counts on the right. The altered amino acids due to the frameshift of the predicted loss-of-function sequence p.(Leu10Tyrfs*37) are displayed in red. The protein sequence obtained by translation initiation at the novel ATG or M13 is identical to the wild-type sequence, lacking only 9 or 12 N-terminal amino acids, respectively. The wild-type M59 codon (highlighted in red) could be excluded as translation initiation site based on the mutations associated with restrictive dermopathy in the literature.

3.1.5. Functional analyses

In the 1st functional approach to determine the translation initiation sites in *ZMPSTE24* harboring the studied mutation c.28_29insA, the *ZMPSTE24* sequence was amplified from purchased human fetal brain cDNA and inserted into a TA cloning vector. However, sequencing of the final construct did not reveal the desired wild-type sequence. The sequence cloned into the vector contained an additional exon upstream of exon 7 of the MANE transcript NM_005857.5 (141). Database alignment using the UCSC Genome Browser (191) showed that this alternative exon is also present in the Ensembl transcript ENST00000675937.1 (195). The alternative exon contains a premature termination codon, thus this transcript is associated with nonsense mediated mRNA decay. The alternative exon is also part of the recently annotated NCBI RefSeq transcript XM_047427590.1 (141). In addition, sequences of the alternative exon are reported in the two human expressed sequence tags (ESTs) CD657740 and BE001576 (141), as well as in the computational splice prediction of the Swiss Institute of Bioinformatics (196). Of interest, the amplified sequence from human fetal brain marathon ready cDNA differs from all annotated transcripts by missing the first 22 bp of exon 7 (Figure 13).

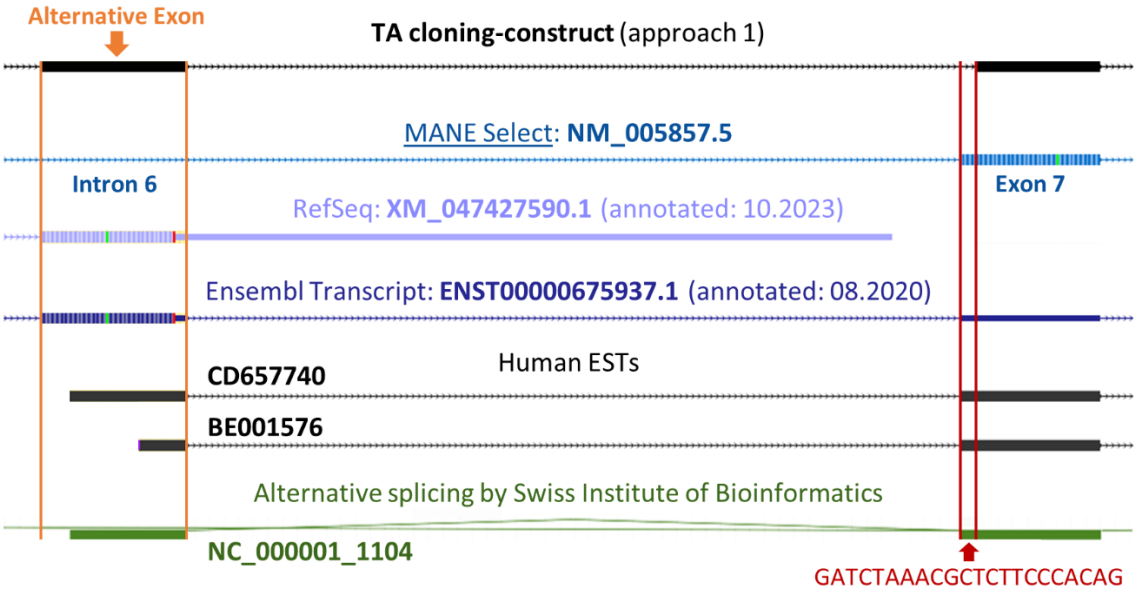

Figure 13. Detection of an alternative exon in *ZMPSTE24* in the generated vector of approach 1. A section of the sequence amplified from purchased human fetal brain cDNA, which was inserted into the TA cloning vector, is shown at the top in black. The sequence contains an additional exon upstream of exon 7 of the transcript NM_005857.5. The corresponding sequences of the transcripts XM_047427590.1 (NCBI RefSeq), ENST00000675937.1 (Ensembl), CD657740, BE001576 (both

expressed sequence tags) and NC_000001_1104 (alternative splicing by Swiss Institute of Bioinformatics) are also illustrated. The vertical red line at the 3' end of the alternative exon in transcripts XM_047427590.1 and ENST00000675937.1 indicates a premature termination codon. Note: The sequence in the TA cloning-construct is missing the first 22 bp of exon 7, which is, however, retained in all other known transcripts.

In addition to the alternative exon, the inserted *ZMPSTE24* sequence also contained a few single nucleotide variants. As a result, the generated construct in approach 1 could not be used for functional TIS analysis.

The localization experiments to examine potential alternative translation initiation sites were performed with the expression constructs described in paragraph 2.1.4.2 (see also Table 17). The GFP-tagged wild-type ZMPSTE24 protein, serving as positive control, was expressed in AD-293 cells and demonstrated localization at the endoplasmic reticulum (Figure 14A). Expression of the mutant *ZMPSTE24* clone carrying the studied disease-causing mutation c.28_29insA also showed ER localization, indicating alternative translation initiation as rescue mechanism of the lethal RD phenotype in affected individuals (Figure 14B). The canonical start codon (M1) and the M3 codon are not in-frame with GFP due to the frameshift in the mutant gene. Therefore, the observed fluorescence-labelled protein at the ER can only result from alternative translation initiation. Both the novel ATG created by the insertion and the M13 codon are in-frame with GFP and could potentially serve as alternative TIS. To clarify whether the M13 codon constitutes a functional translation initiation site, a mutant *ZMPSTE24* clone carrying the mutation c.-4_7del was expressed, thus lacking the canonical start codon and part of the M3 codon. The resulting GFP-tagged protein showed localization at the ER (Figure 14C), suggesting the M13 codon as a potential alternative TIS in mutant *ZMPSTE24* harboring the studied mutation c.28_29insA.

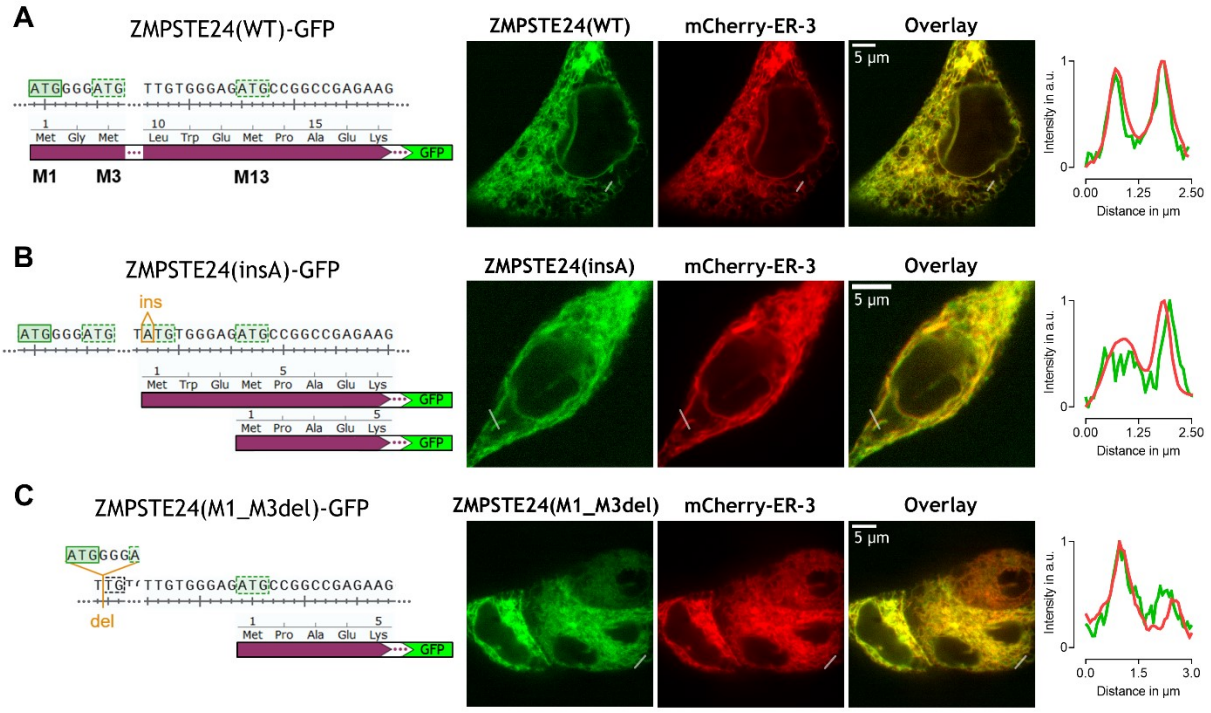

Figure 14. Localization analysis with GFP-tagged *ZMPSTE24* expression constructs to determine if alternative translation initiation functions as rescue mechanism.

[Figure and legend adapted from Clinical Genetics (Kaufmann et al., 2023)]

DNA sequence of the N-terminal region of *ZMPSTE24* (top strand) is illustrated for each construct. The canonical start codon (continuous green line), potential downstream start codons (dashed green line) and mutation sites (orange; ins = insertion, del = deletion) are highlighted. Remaining DNA sequence of *ZMPSTE24* and *GFP*, which has not been altered in these constructs, is not visualized, but indicated by gray dots. N-termini of possibly expressed proteins (purple bars) in-frame with GFP (indicated by purple dots and green bar) for each construct are shown below. N-terminal methionines of the wild-type protein are additionally highlighted (M1, M3, M13). Confocal fluorescence microscopy images of GFP-tagged *ZMPSTE24* expression constructs (green) and the localization marker for the endoplasmic reticulum mCherry-ER-3-RFP (red) in AD-293 cells are shown in the middle section. The right microscopy images display overlay of both channels and the scale bar indicates 5 µm. Graphs of the line scan analyses (pixel intensity of each channel plotted against distance along a drawn line) are depicted at the rightmost column. Corresponding lines are displayed in white in the microscopy images.

(A) Expression of the construct ZMPSTE24(WT)-GFP including the wild-type sequence of *ZMPSTE24* and serving as positive control showed ER localization.

(B) Expression of the construct ZMPSTE24(insA)-GFP harboring the studied disease-causing mutation c.28_29insA also showed localization at the ER, indicating translation initiation at an alternative start codon.

(C) Expression of the construct ZMPSTE24(M1_M3del)-GFP harboring the mutation c.-4_7del displayed ER localization as well, suggesting the M13 codon as a potential alternative TIS in mutant *ZMPSTE24* harboring the studied mutation c.28_29insA.

To determine which of the two potential start codons (novel ATG or M13) functions as TIS, or whether both are used, further localization experiments were performed. With the constructs ZMPSTE24(insA+M13I)-GFP and ZMPSTE24(insT)-GFP, the two potential translation initiation sites can be examined separately. The expression experiments showed localization at the ER for both constructs (Figure 15A and B), indicating that the novel ATG and M13 both function as alternative start codons. In contrast, expression of ZMPSTE24(insT+M13I)-GFP resulted in faint cytosolic green signal and absent ER localization (Figure 15C). This demonstrates that at least one of the two alternative start codons must be present to obtain an N-terminally truncated ZMPSTE24 protein located at the ER.

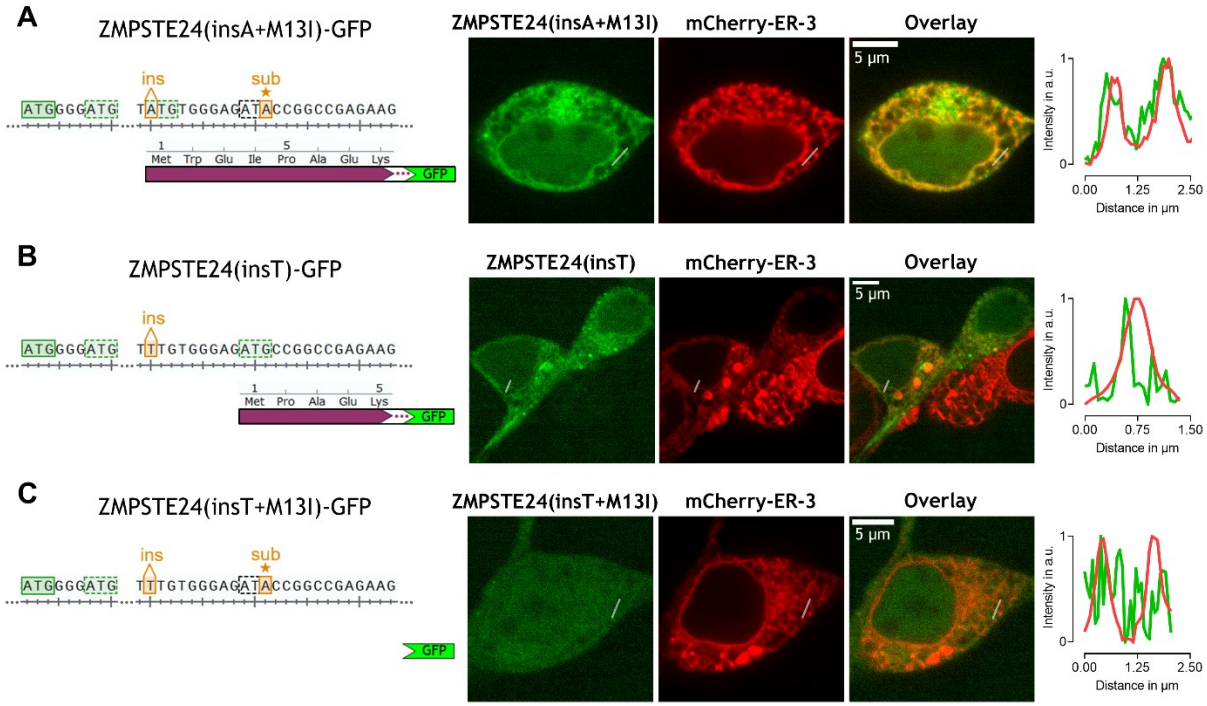

Figure 15. Localization analysis with GFP-tagged *ZMPSTE24* expression constructs to determine which codon functions as alternative translation initiation site.

[Figure and legend adapted from Clinical Genetics (Kaufmann et al., 2023)]

DNA sequence of the N-terminal region of *ZMPSTE24* (top strand) is illustrated for each construct. The canonical start codon (continuous green line), potential downstream start codons (dashed green line) and mutation sites (orange; ins = insertion, sub = substitution) are highlighted. Remaining DNA sequence of *ZMPSTE24* and *GFP*, which has not been altered in these constructs, is not visualized, but indicated by gray dots. N-termini of possibly expressed proteins (purple bars) in-frame with GFP (indicated by purple dots and green bar) for each construct are shown below. N-terminal methionines of the wild-type protein are additionally highlighted (M1, M3, M13). Confocal fluorescence microscopy images of GFP-tagged *ZMPSTE24* expression constructs (green) and the localization marker for the

endoplasmic reticulum mCherry-ER-3-RFP (red) in AD-293 cells are shown in the middle section. The right microscopy images display overlay of both channels and the scale bar indicates 5 µm. Graphs of the line scan analyses (pixel intensity of each channel plotted against distance along a drawn line) are depicted at the rightmost column. Corresponding lines are displayed in white in the microscopy images. **(A)** Expression of the construct ZMPSTE24(insA+M13I)-GFP harboring the mutations c.28_29insA and c.39G>A showed ER localization, indicating the novel ATG as alternative TIS.
(B) Expression of the construct ZMPSTE24(insT)-GFP harboring the mutation c.28_29insT showed ER localization, indicating the M13 codon as alternative TIS.
(C) Expression of the construct ZMPSTE24(insT+M13I)-GFP harboring the mutations c.28_29insT and c.39G>A led to faint cytosolic green signal and absent ER localization, indicating that at least one of the two alternative start codons must be present to obtain an N-terminally truncated ZMPSTE24 protein which is located at the ER.

Our localization and expression experiments supported utilization of two alternative N-terminal translation initiation sites, at the novel ATG created by the insertion and the M13 codon referred to the WT gene. Translation initiation at these alternative start codons results in almost full-length, ER-located ZMPSTE24 proteins. Thereby the usage of alternative N-terminal TISs explains the rescue mechanism by which the predicted complete loss-of-function effect and the associated lethal phenotype of the disease-causing frameshift mutation c.28_29insA in *ZMPSTE24* could be avoided. Based on our experiments and the identification of alternative TISs, the standard nomenclature for predicting the effects at protein level of the *ZMPSTE24* mutation NM_005857.5:c.28_29insA could be complemented by the two N-terminally truncated proteins NP_005848.2:p.([Gly2_Leu10del,Gly2_Met13del]).

3.2. *MN1* study

This study investigates the genetic cause of severe cerebral malformations in a male fetus at 14 weeks gestation.

3.2.1. Clinical diagnosis

Sonographic fetal findings were as follows: crown-rump length clearly below the 5th percentile, disrupted midline structures in the cerebrum, hypoplastic thalamic area, right lateral ventricle in connection with a large cranial encephalocele, choroid plexus displaced dorsally, left lateral ventricle dilated, choroid plexus also displaced dorsally and the structures of the posterior fossa were not identifiable: cystic mass in place of the posterior fossa, as well as inferior temporal gyrus and brain stem not recognizable. No other obvious malformations were

apparent, but cleft palate could not be excluded. A severe cerebral malformation with unfavorable prognosis was thus given as the referral diagnosis for genetic testing.

3.2.2. Genetic DNA analyses

SNP array analysis of DNA from chorionic villus sampling revealed a heterozygous interstitial deletion in the chromosomal region 22q12.1 with a size of 58 kb. The ISCN nomenclature of the detected deletion is: arr[GRCh38] 22q12.1 (27710404x2, 27714323_27772264x1, 27772359x2). The deletion encompasses the entire last exon (exon 2) of the *MN1* gene. The proximal breakpoint downstream of *MN1* is located in intron 1 of the long non-coding RNA (lncRNA) gene *CPMER* (cytoplasmic mesoderm regulator) (Figure 16A). The size of the deletion was confirmed by qRT-PCR analysis of fetal tissue after medically indicated termination of pregnancy. The presence of the heterozygous deletion in the tissue sample also rules out confined placental mosaicism (Figure 16B).

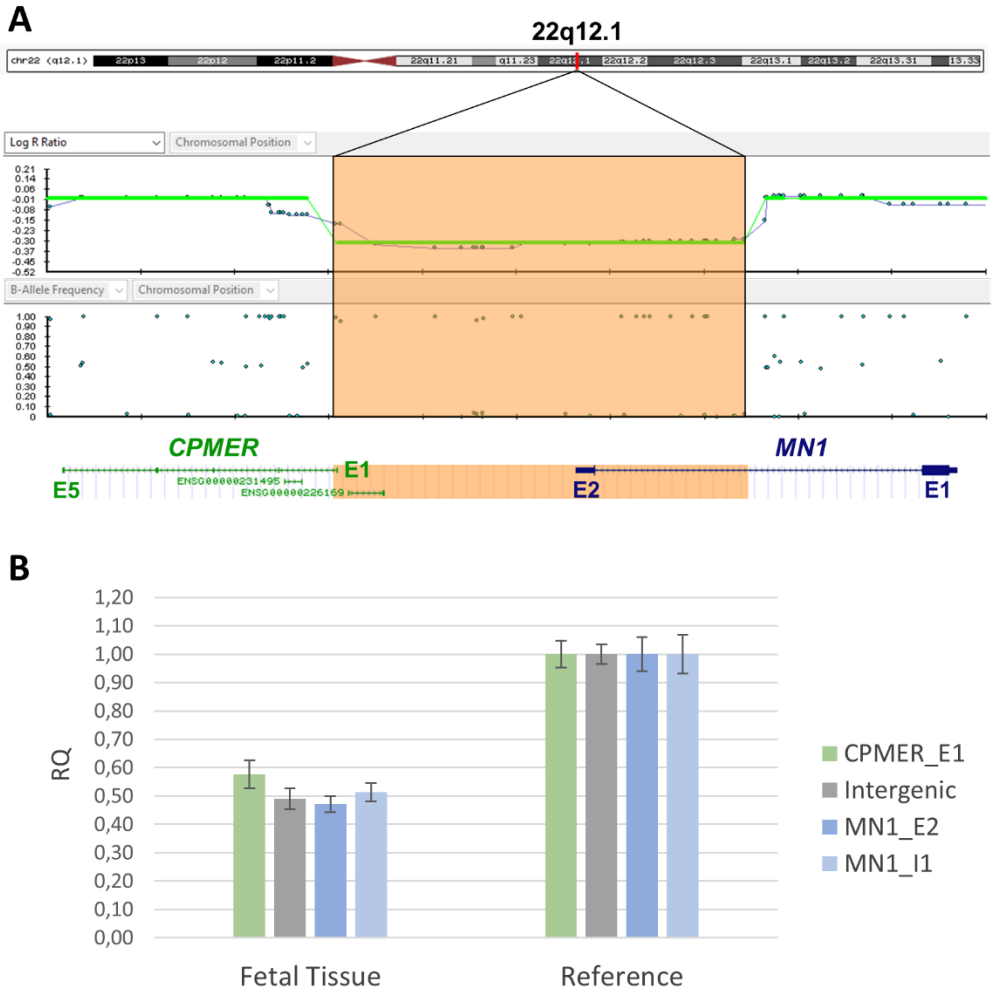

Figure 16. Detection of a heterozygous 22q12.1 deletion in the fetus.

(A) Data from SNP array analysis of DNA from chorionic villus sampling which led to the detection of a heterozygous interstitial deletion in the 22q12.1 region: log R ratio (top) and B-allele frequency (bottom). The deleted region chr22:27,714,323-27,772,264 (GRCh38/hg38) is highlighted in orange. The proximal breakpoint is located in intron 1 of the lncRNA gene *CPMER*, the distal breakpoint is located in intron 1 of the *MN1* gene.
(B) The bar chart shows the relative quantification (RQ) values including standard deviations of the qRT-PCR analysis for the respective amplicons in the fetal tissue sample (left) and the reference (right). Amplicon locations are shown on the far right. The heterozygous deletion was also detected in the fetal tissue sample, thus ruling out confined placental mosaicism.

The location of the deletion in 22q12.1 could be of particular relevance regarding possible phenotypic associations, as it leads to a gene fusion of the genes *CPMER* and *MN1* on the negative strand (Figure 16A). Subsequent in-depth analysis revealed that the expression of this fusion gene would lead to a C-terminally truncated MN1 protein and would thus be associated with MN1 C-terminal truncation (MCTT) syndrome, a very rare disease involving brain malformations. However, loss-of-function of this allele would be associated with milder clinical abnormalities and variable expressivity, as well as incomplete penetrance.
The parents of the fetus were also examined for the presence of the deletion, in order to further assess the pathogenic significance. Segregation analysis revealed that the fetal deletion is a *de novo* mutation (Figure 17). Additional trio-whole exome sequencing and examination of the chromosomes revealed no evidence of any other potentially causative variant in the fetus.

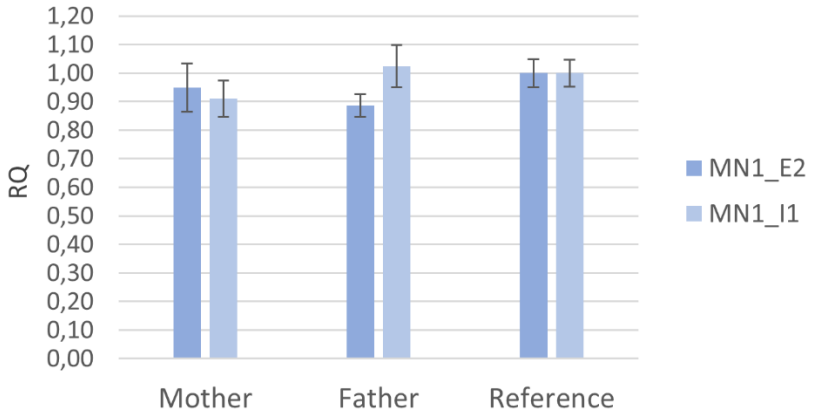

Figure 17. Segregation analysis of the fetal 22q12.1 deletion in the parents.
The bar chart shows the relative quantification (RQ) values including standard deviations of the qRT-PCR analysis for the respective amplicons in maternal (left) and paternal (middle) blood samples, as well as in the reference (right). Amplicon locations are shown on the far right. The fetal deletion could not be detected in the parents, indicating a *de novo* mutation.

3.2.3. RNA analyses

RNA (i.e. cDNA) analyses were performed to verify whether the *MN1-CPMER* fusion gene is actually transcribed into RNA. In qualitative RNA analysis the cDNA was analyzed for direct junctions between the first exon of *MN1* and exons of *CPMER*. Sanger sequencing revealed a direct link between *MN1* exon 1 and *CPMER* exon 2 (Figure 18A). In addition, two further direct junctions between the first *MN1* exon and *CPMER* were detected in the cDNA: one sequence contains a short segment of intron 4, which merges into exon 5, and one sequence shows a direct junction to exon 5 of *CPMER*. The results of the qualitative RNA analysis indicate alternative splicing of the fusion gene (Figure 18B). Each of the RNA transcripts could be detected under both conditions, inhibition of nonsense mediated mRNA decay (NMD) and without NMD inhibition. No fusion transcripts were detected in the control samples. The respective junctions of *CPMER* E2-E3, E3-E4 and E2-E4 could not be detected in any sample.

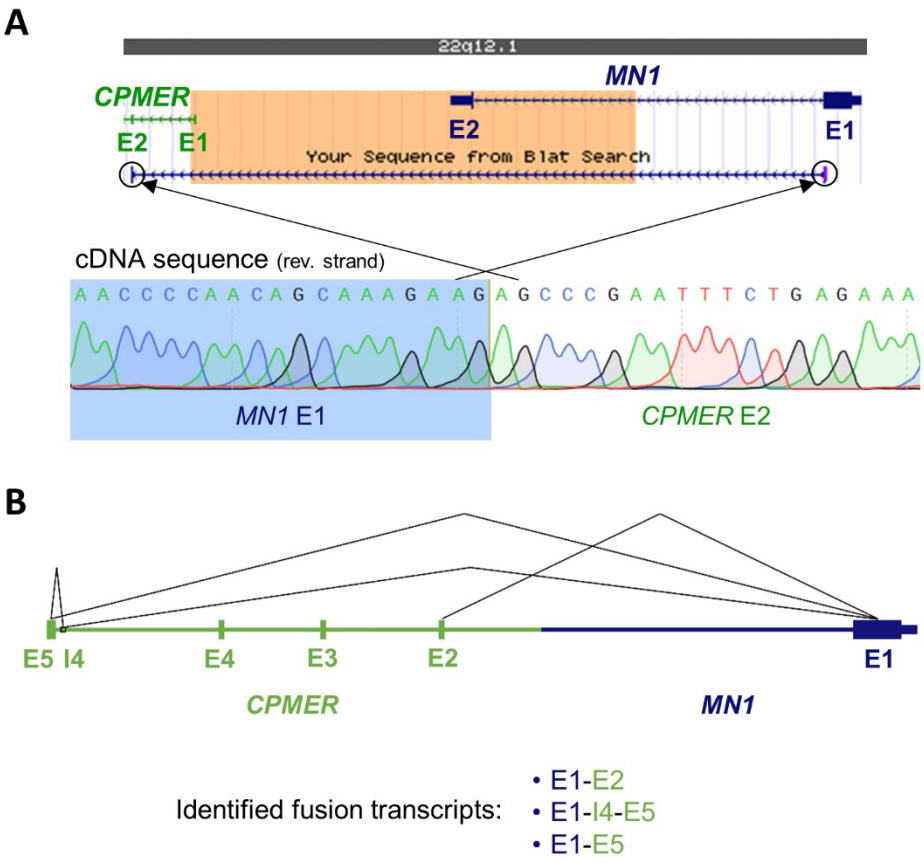

Figure 18. Qualitative RNA analysis for the detection of fusion transcripts.

(A) Sanger sequencing of cDNA from chorionic villus culture showing a direct linkage of the first exon of *MN1* with exon 2 of *CPMER*. The sequence displays the reverse strand where both genes are located. The 22q12.1 region deleted in the fetus, including exon 2 of *MN1* and exon 1 of *CPMER*, is highlighted in orange.

(B) Further fusion transcripts were identified using region-specific primers. The detection of several transcripts indicates alternative splicing. The respective junctions could be detected both, after puromycin treatment (NMD inhibition) and without pretreatment (no NMD inhibition).

In silico protein sequence alignment demonstrates that translation of the three identified fusion transcripts leads to C-terminally truncated MN1 proteins missing the amino acids encoded by exon 2 (Figure 19).

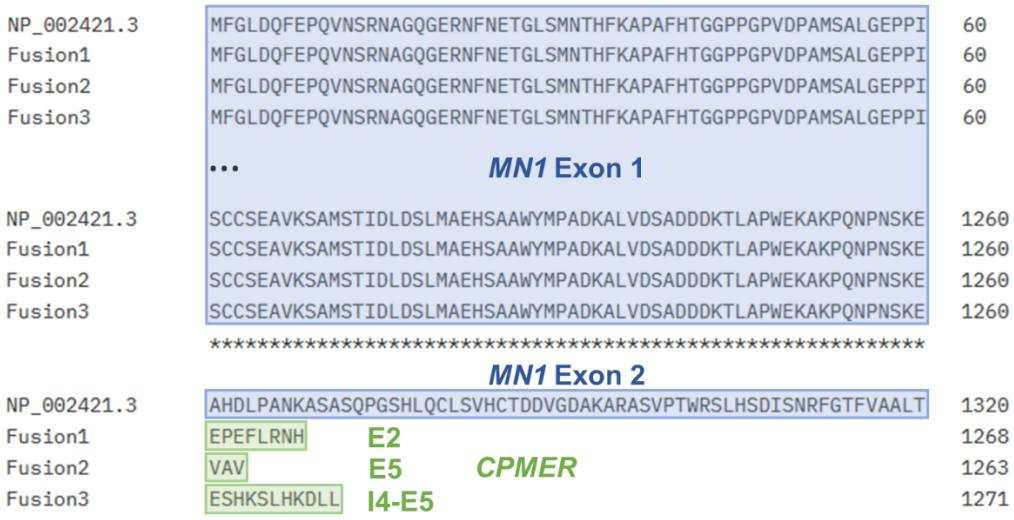

Figure 19. Protein sequence alignment of identified fusion transcripts and wild-type MN1.
The alignment illustrates that the amino acids encoded by exon 1 of the *MN1* gene are identical to the wild-type sequence of the MN1 protein (NP_002421.3) in all transcripts. However, the amino acids encoded by exon 2 are missing in all three fusion transcripts. Expression of these fusion transcripts would lead to C-terminally truncated MN1 proteins associated with MCTT syndrome. Aligned protein sequences are displayed in single letter code with amino acid counts on the right.

Since the RNA quantity of the mutant *MN1* allele cannot be determined directly due to the presence of distinct fusion transcripts, quantification is performed indirectly by comparing exon 1 with allele-specific wild-type amplicons (Figure 20). Quantitative RNA analysis was performed under both conditions, inhibition of nonsense mediated mRNA decay and no NMD inhibition. The analysis revealed twice the amount of exon 1 cDNA compared to the specific amplicons for the wild-type allele, indicating that exon 1 of *MN1* is expressed at the mutant allele despite the deletion of exon 2. This effect can be observed under NMD inhibition as well

as without NMD inhibition. In addition, the fusion construct *MN1* E1-E2 *CPMER* could also be determined quantitatively. The amount of cDNA of the fusion transcript appears to be lower than that of exon 2, which is consistent with the previous detection of several fusion transcripts. The amount under NMD inhibition might be slightly higher than without NMD inhibition for the fusion transcript. The remaining two fusion transcripts could not be quantitatively investigated due to the limited amount of primary sample.

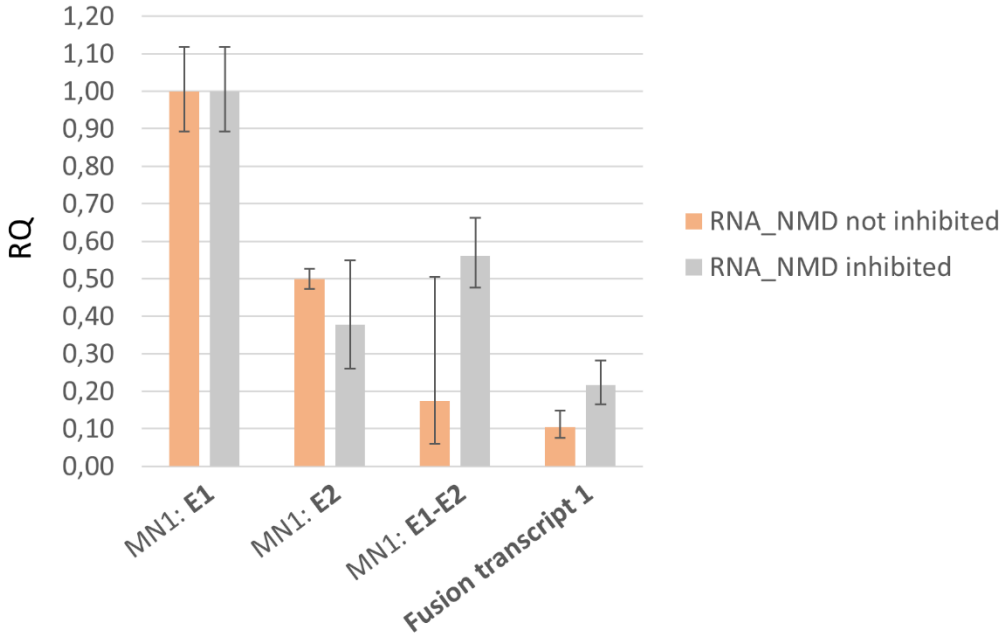

Figure 20. Quantitative RNA analysis to determine the expression of the mutant *MN1* allele. Comparison of exon 1 with specific amplicons for the wild-type allele (*MN1*: E2 and *MN1*: E1-E2). The cDNA quantity of the exon 1 amplicon of *MN1* was set as calibrator. All other amplicons are compared to the calibrator. Quantitative RNA analysis was performed under both conditions, inhibition of nonsense mediated mRNA decay (grey) and no NMD inhibition (orange). The quantification shows half the amount of wild-type allele-specific amplicons compared to exon 1, indicating expression of exon 1 at the mutant allele. The fusion transcript *MN1* E1-E2 *CPMER* could also be determined quantitatively under both conditions. Note high standard deviation for amplicon *MN1*: E1-E2 in case of no NMD inhibition.

Qualitative RNA analysis revealed the presence of three *MN1-CPMER* fusion transcripts lacking the sequence of deleted exon 2 of *MN1*. The results of the quantitative analysis indicate that the fusion transcripts of the mutant allele are transcribed and are not subject to nonsense mediated mRNA decay. Overall, the analyses performed suggest a possible expression of C-terminally truncated MN1 proteins in the fetus, which are associated with the MCTT syndrome.

4. Discussion

According to the literature, about half of patients with a rare disease still never receive a diagnosis (49, 51). A major challenge in providing a genetic diagnosis is determining whether and what effects previously unknown variants may have. Thus, the identification of novel disease-causing mutations and the correct interpretation of the respective effects are in many cases crucial steps on the way to diagnostic success. The detection of these mutations and associated effects in rare diseases may also serve as starting point for the development of therapeutic approaches (49, 53). Variants in genes with pleiotropy represent a particular challenge in the identification of previously unknown disease-causing mutations. In these cases, it is especially important to examine the pathomechanism of the variant in order to make the right diagnosis.

In the interpretation of novel variants, computer-based annotations play a key role in current times. Most annotation approaches are gene-based and describe how variants change the coding sequence and subsequently the amino acid sequence. However, these standard approaches are prone to numerous sources of error (103, 133). That is because these approaches are unable to fully consider all the mechanisms and interactions that occur in vivo. A significant example of these error sources is the annotation of truncating variants. In standard variant annotation pipelines, these variants are mainly associated with a LOF effect, as well as nonsense-mediated mRNA decay. In vivo, however, there are mechanisms that lead to NMD escape. In these cases, standard annotations yield false predictions, resulting in incorrect variant interpretation (103). In this thesis, the importance of in-depth evaluation as well as the implementation of further analyses for the correct interpretation of novel disease-causing variants is demonstrated by two examples.

In the course of the presented studies, we successfully discovered and analyzed two previously unknown, disease-causing mutations. Moreover, both mutations share an exceptional feature: the predicted effects of the variants did not match the observed phenotype of affected individuals. The variants each affect a gene with known pleiotropy and both variants are predicted to result in a complete loss of gene function. The predicted LOF effect is circumvented by alternative gene expression mechanisms in both variants.

4.1. *ZMPSTE24* study

In this study, we characterize two consanguineous Pakistani families displaying a common progeroid phenotype that was subsequently diagnosed as mandibuloacral dysplasia with type B lipodystrophy (MADB). Affected individuals from both families carry the previously unknown homozygous frameshift variant NM_005857.5:c.28_29insA in *ZMPSTE24*. The standard annotation according to HGVS nomenclature (130) provided by Mutalyzer (197) predicts that the insertion results in a frameshift at amino acid position 10 and leads to a premature termination codon after 36 subsequent amino acids (p.(Leu10Tyrfs*37)). Furthermore, it is predicted that the mutation leads to NMD (198), which is associated with a loss-of-function of the mutated allele. Manual analysis also indicates NMD, as the premature termination codon is not within the first 100 coding base pairs, which would be linked to NMD escape (88). According to the study of Lindeboom et al., the premature termination codon of the detected variant (c.136_138) is located in a region with a gradual increase in NMD efficiency, which starts at 100 bp and extends up to 200 bp downstream of the start codon. Another interesting finding of Lindeboom et al. is that a small distance between the canonical start codon and the PTC is a strong predictor for NMD escape. In contrast, the distance to the downstream alternative AUG or the strength of the Kozak sequence of this AUG does not influence NMD efficiency (118).

A literature search shows that to date all biallelic loss-of-function mutations in *ZMPSTE24* have indeed been associated with RD, leading to the death of affected infants within the first weeks of life (Table 23).

Table 23. Known published disease-causing mutations in *ZMPSTE24*

[Table and legend originally published in Gene (Schaflinger & Blatterer et al., 2022)]

Phenotype	Allele 1	Allele 2	Reference
MADB	c.28_29insA, p.(Leu10Tyrfs*37)	c.28_29insA, p.(Leu10Tyrfs*37)	This study (199, 200)
RD	c.50delA, p.(Lys17Serfs*21)	c.50delA, p.(Lys17Serfs*21)	(201)
RD	c.50delA, p.(Lys17Serfs*21)	c.584_585delAT, p.(Tyr195Phefs*22)	(170)
RD	c.50delA, p.(Lys17Serfs*21)	c.1085dupT, p.(Leu362Phefs*19)	(202)
RD	c.54dupT, p.(Ile19Tyrfs*28)	c.54dupT, p.(Ile19Tyrfs*28)	(203)
RD	c.54dupT, p.(Ile19Tyrfs*28)	c.826C>T, p.(Arg276*)	(202)
MADB	c.121C>T, p.(Gln41*)	c.743C>T, p.(Pro248Leu)	(204)
RD	c.121C>T, p.(Gln41*)	c.1393A>G, p.(Arg465Gly)[a]	(205)
MADB	c.202C>T, p.(Arg68*)	c.743C>T, p.(Pro248Leu)	(206)
MAD	c.207_208delCT, p.(Tyr70Serfs*4)	c.794A>G, p.(Asn265Ser)	(207)
RD	c.209_210delAT, p.(Tyr70Serfs*4)	c.209_210delAT, p.(Tyr70Serfs*4)	(202)
RD	c.271-1446_628-127del, p.(Leu91_Leu209del)	c.1085dupT, p.(Leu362Phefs*19)	(208)
MADB	c.281T>C, p.(Leu94Pro)	c.281T>C, p.(Leu94Pro)	(209)
RD	c.296delC, p.(Pro99Leufs*38)	c.1085dupT, p.(Leu362Phefs*19)	(208)
RD	c.475-2A>G, p.(?)	c.628-2A>G, p.(?)	(202)
RD	c.591dupT, p.(Ile198Tyrfs*20)	c.591dupT, p.(Ile198Tyrfs*20)	(203)
RD	c.591dupT, p.(Ile198Tyrfs*20)	c.1085dupT, p.(Leu362Phefs*19)	(203)
RD	c.627 +1G>C, p.(?)	c.627 +1G>C, p.(?)	(210)
RD	c.691G>T, p.(Glu231*)	c.691G>T, p.(Glu231*)	(211)
RD	c.709G>T, p.(Glu237*)	?	(212)[b]
RD	c.715G>T, p.(Glu239*)	c.715G>T, p.(Glu239*)	(213, 214)
MADB	c.743C>T, p.(Pro248Leu)	c.1349G>A, p.(Trp450*)	(215)
HGPS/MAD MADB	c.794A>G, p.(Asn265Ser)	c.1085dupT, p.(Leu362Phefs*19)	(216-219)
MADB	c.794A>G, p.(Asn265Ser)	c.1204-1G>A, p.(?)	(202)
RD	c.954+2 T>A, p.(?)	c.1085dupT, p.(Leu362Phefs*19)	(202)
MADB	c.1018T>C, p.(Trp340Arg)	c.1085dupT, p.(Leu362Phefs*19)	(165)
RD	c.1020G>A, p.(Trp340*)	c.1020G>A, p.(Trp340*)	(220)
MADB	c.1052_1054delTTA, p.(Ile351del)	c.1052_1054delTTA, p.(Ile351del)	(221)[c]
RD	c.1085dupT, p.(Leu362Phefs*19)	c.1085dupT, p.(Leu362Phefs*19)	(202, 203, 208, 220, 222-227)

HGPS	c.1085dupT, p.(Leu362Phefs*19)[d]		(228)
RD	c.1085dupT, p.(Leu362Phefs*19)	c.1249C>T, p.(Gln417*)	(208)
RD	c.1105C>T, p.(Arg369*)	c.1105C>T, p.(Arg369*)	(202, 208)
MADB	c.1196A>G, p.(Tyr399Cys)	c.1196A>G, p.(Tyr399Cys)	(229, 230)
HGPS/MAD	c.1204-5_1210del, p.(Val402Serfs*2)	c.1204-5_1210del, p.(Val402Serfs*2)	(231)[e]
MADB	c.1274T>C, p.(Leu425Pro)	c.1274T>C, p.(Leu425Pro)	(232)
RD	c.1385T>G, p.(Leu462Arg)[a]	?	(233)[f]

Mutations with (presumed) loss-of-function effect resulting in RD phenotype in a biallelic state (red), including missense-mutations with predicted loss-of-function effect and loss-of-function mutation with additional modifying heterozygous *LMNA* mutation; mutations with (presumed) residual function resulting in MADB (blue); novel mutation identified in this study c.28_29insA, although predicted as loss-of-function, is associated with MADB; MADB (mandibuloacral dysplasia with type B lipodystrophy), RD (lethal restrictive dermopathy), HGPS (Hutchinson-Gilford progeria syndrome), MAD (mandibuloacral dysplasia).

[a] Missense variant is predicted to interfere with localization signal.

[b] Only mother was tested.

[c] Patients died in childhood, biological samples were only available from parents; both parents carried c.1052_1054delTTA mutation which likely explains phenotype of their affected offspring.

[d] Uniparental isodisomy mosaicism, 34% of cells are estimated to be heterozygous for a wild-type allele.

[e] Patient had an additional heterozygous *LMNA* c.1960C>T, p.(Arg654*) mutation, modifying the otherwise presumably lethal effect of the homozygous *ZMPSTE24* mutation.

[f] Only one mutation was identified.

Based on the data collected by the literature search, the number of potential alternative N-terminal start codons that could result in LOF circumvention in the patients was reduced to two: the novel ATG codon created at the mutation site and the M13 codon related to the WT gene. Translation initiation at the M3 codon could be excluded in the patients, as this codon is located upstream of the mutation site and would therefore result in a frameshift analogous to the canonical start site (Figure 21). The most N-terminal known frameshift variants are c.50delA, p.(Lys17Serfs*21) and c.54dupT, p.(Ile19Tyrfs*28) leading to restrictive dermopathy in homozygous state. These variants are located downstream of the hypothesized alternative TIS at M13 and the next potential start codon is M59. The premature termination codons of the reported variants c.50delA and c.54dupT are both located upstream of M59 (Figure 21). The predicted PTC of the variant c.54dupT, p.(Ile19Tyrfs*28) corresponds to the predicted PTC of the studied mutation c.28_29insA, p.(Leu10Tyrfs*37). However, translation reinitiation at M59 can be ruled out as LOF escape mechanism, since the variants c.50delA and c.54dupT are associated with the lethal RD phenotype.

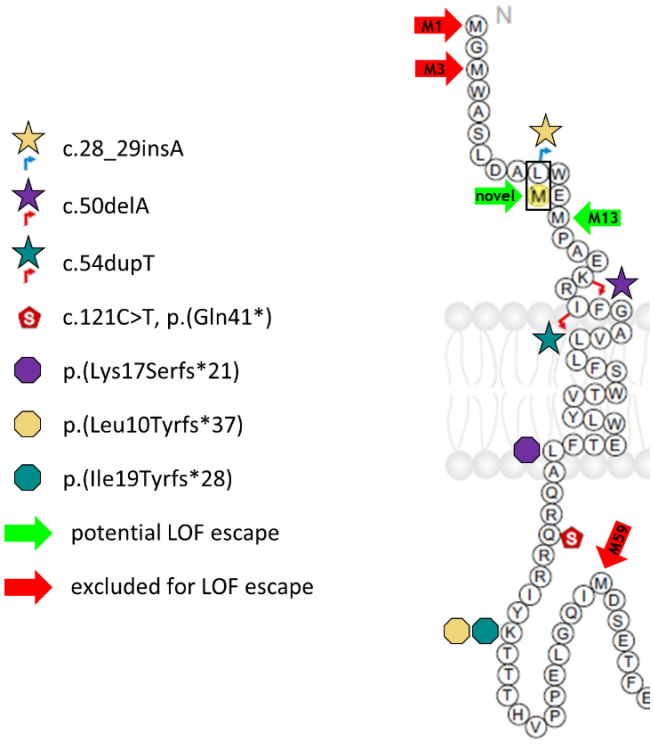

Figure 21. In-depth analysis of potential alternative TISs to circumvent the predicted LOF effect. [Figure and legend adapted from Gene (Schaflinger & Blatterer et al., 2022)]
The N-terminus of the ZMPSTE24 protein with the mutation sites of the studied mutation c.28_29insA, the most N-terminal variants described in the literature c.50delA, c.54dupT and c.121C>T,p.(Gln41*), as well as the respective predicted PTCs of the frameshift variants (illustrated as stop signs) are shown. The remaining protein sequence is not displayed. The ATG codons M3 (upstream of the frameshift variants) and M59 (no LOF rescue in the published variants associated with lethal RD) can be excluded as alternative TISs to circumvent the LOF effect in the affected individuals (red arrows). Only the novel ATG codon created at the mutation site of c.28_29insA and the M13 codon can be considered as possible alternative start codons in the studied families after in-depth analysis (green arrows).

The MADB phenotype observed in the studied families appears to be more severe than in other published cases. Affected individuals present with severe, progressive thoraco-lumbar kyphosis and bilateral distended loins that could be related to renal lesions (Figure 9B). Interestingly, kyphosis is described as a known additional feature of restrictive dermopathy (RD) (10). In contrast, it has rarely been reported in individuals with MADB (204, 209).

4.1.1. Approach 1: additional exon and single nucleotide variants

The expression vector generated in the first approach could not be used for functional analysis of alternative TISs due to an additional exon and single nucleotide variants in the cloned *ZMPSTE24* sequence. The additional exon is likely to originate from a transcript contained in the human fetal brain cDNA sample purchased from Clontech/Takara Bio. This exon was first described in human ESTs CD657740 and BE001576 (141) and is also contained in the manually annotated Ensembl transcript ENST00000675937.1 (195) based on RNA sequencing. The additional exon is also part of the recently annotated NCBI RefSeq transcript XM_047427590.1 (141), which is based on automated computational prediction supported by EST evidence. The additional exon upstream of exon 7 of the MANE transcript contains a PTC, so that the transcript is associated with NMD (195). It is currently unknown in which tissue the transcript is expressed and how high the expression is. Interestingly, the obtained sequence in our vector differs from the annotated transcripts by the absence of the first 22 bp of exon 7 (Figure 13).

The single nucleotide variants probably arose because no proof-reading polymerase was used in the 1st approach. Furthermore, the *E. coli* cells were incubated at 37 °C after transformation. However, since the WT ZMPSTE24 protein is toxic in these bacteria, colonies with a construct containing mutations have growth advantages. This would also explain why our vector from approach 1 contained the transcript with the additional exon, which only codes for a truncated ZMPSTE24 protein. Starting with approach 2, the incubation temperature was lowered to 25 °C to minimize expression and toxic effects.

4.1.2. LOF escape mechanism (*ZMPSTE24*)

The conducted expression and localization experiments presented herein elucidate the underlying mechanism by which the predicted severe protein truncation, indicated as p.(Leu10Tyrfs*37), and the associated NMD is circumvented in affected individuals harboring the homozygous mutation NM_005857.5:c.28_29insA. We could demonstrate that the LOF effect is bypassed by alternative translation initiation at two N-terminal start codons. Remarkably, one of these alternative TISs is newly formed at the insertion site. To our knowledge, such a LOF rescue, in which a frameshift mutation leads to the formation of a novel, physiologically relevant, in-frame ATG codon at the mutation site, has not yet been described. The second alternative translation initiation site corresponds to the M13 codon of the wild-type protein (Figure 15). For the studied mutation NM_005857.5:c.28_29insA, the mechanism is unlikely to be “classic translation reinitiation”, as the premature termination

codon is located downstream of the alternative TIS. In “classic translation reinitiation”, the reinitiation start codons are always located downstream of the PTC (Figure 1) (111). In the studies by Sherlock et al. and Gould et al. an alternative translation reinitiation mechanism, called coupled translation termination-reinitiation, is described. In this mechanism the reinitiation start codon can be very close to or even upstream of the stop codon. This reinitiation mechanism with a start codon upstream of the PTC is mainly described in viruses (111, 234). In this scenario the two codons are within 20 nt of each other and reinitiation efficiency decreases if they are moved more than approximately 30 nt apart (235, 236). Gould et al. showed that increasing the distance between the two codons from 81 nt to 117 nt leads to a dramatically reduced reinitiation efficiency. In the case of the studied mutation, the distance of the upstream alternative TISs to the PTC is 101 and 110 nt, respectively. Although coupled translation termination-reinitiation was also detected in transcripts of the human genome, it requires an aspartate-rich repeat region close to the PTC (234). Therefore, translation reinitiation as LOF rescue mechanism is not assumed in studied individuals.
It is more likely that in the case of the studied mutation, the LOF rescue mechanism is leaky scanning (Figure 1). It remains to be noted that it is extremely difficult to experimentally distinguish between reinitiation events and leaky scanning (111). In silico predictions indicate that the canonical start codon in *ZMPSTE24* is rather weak (Table 22), which also points to leaky scanning as the mechanism for alternative translation initiation. Studies have shown that weak canonical AUG signalling is often accompanied by conserved in frame downstream TISs (237, 238).

Furthermore, the weak cytosolic GFP signal observed at the expression of ZMPSTE24(insT+M13I)-GFP construct in the absence of ER localization (Figure 15) confirms the assumption that the M59 codon and all AUG codons located further downstream do not act as physiologically relevant alternative TISs (Figure 21). In ZMPSTE24, the KxxKxx dilysine ER retrieval motif is located at the C-terminus (KTMKQH). A study by Barrowman et al. has shown that an epitope tag directly downstream of the ER retrieval motif at the C-terminus disrupts normal ER localization. When the tag was inserted just upstream of the ER retrieval motif, normal ER localisation of the ZMPSTE24 protein was maintained. Although this observation suggests the presence of a functional C-terminal ER retrieval motif in ZMPSTE24, mutational analysis of this region is required for confirmation (159). An earlier study by Schmidt et al. found that mutated dilysine ER retrieval motif in the yeast homolog Ste24p does not lead to mislocalization of Ste24p (148). Based on our expression and localization experiments, we

could show that the first 12 N-terminal amino acids of ZMPSTE24 are not necessary for ER localization.

In silico modeling of the wild-type and mutant ZMPSTE24 proteins revealed no significant differences in protein structure and stability, indicating possible preservation of protein function of the N-terminally truncated protein (199). In addition, alternative translation initiation mechanisms such as leaky scanning are reported to occur in the minority of translation events in many cases. Known functional consequences of leaky scanning are the regulation of the amount of protein produced by the main ORF or the production of a protein with N-terminal truncation (111). Therefore, we suggest that the two N-terminal AUG codons (novel ATG and M13 codon referring to WT gene) present in mutated *ZMPSTE24* RNA are only sparsely utilized as alternative translation initiation sites in affected individuals displaying MADB. The resulting N-terminally truncated ZMPSTE24 proteins, which show ER localization and maintain all known functional domains may exhibit intact enzymatic activity. Thus, the low amount of physiologically expressed but intact proteins could be sufficient to prevent the lethal RD phenotype.

4.1.3. Limitations of the *ZMPSTE24* study

With the expression and localization experiments performed, we were able to demonstrate alternative translation initiation at two N-terminal start codons to circumvent a loss-of-function effect. These experiments were carried out in vitro and are therefore not equivalent to the physiological processes with regard to expression levels in vivo. It remains questionable whether the expression of only one of the two identified alternative TISs would be sufficient to circumvent lethal consequences, or whether both must be present. Furthermore, conclusions at protein level can only be confirmed via protein analyses, such as the determination of protein sequence, expression levels and enzymatic activity. The families involved in this study did not consent to tissue samples.

4.2. *MN1* study

In this study, the genetic cause of severe cerebral malformations with an unfavorable prognosis in a male fetus at the 14th week of pregnancy was investigated. Routine SNP array analysis revealed a heterozygous 58 kb interstitial deletion in the 22q12.1 region. The distal breakpoint was located in the only intron of the *MN1* gene, so that the 2nd (i.e. last) exon is

deleted. The clinical significance, as well as the effect of the deletion on the functionality of the gene cannot be estimated based on the analysis at DNA level alone.

In the Technical standards for the interpretation and reporting of constitutional copy-number variants of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen) (239), the assessment of a deletion affecting the last exon of a gene initially depends on whether the affected gene exhibits haploinsufficiency (HI). Regarding the *MN1* gene, it has not yet been sufficiently clarified if the *MN1* gene displays haploinsufficiency. In a few individuals with LOF variants in *MN1*, as well as in one individual with a focal deletion of the gene, mostly non-specific phenotype abnormalities were observed. In addition, these are likely to demonstrate incomplete penetrance (181). In the ACMG and ClinGen recommendations, only genes with established haploinsufficiency are considered for a more precise evaluation in the classification process of a deletion. If haploinsufficiency is not established, but some HI prediction tools indicate possible HI, as in the case of *MN1*, the classification of the deletion is only very slightly influenced towards likely pathogenic. If haploinsufficiency had been established for *MN1*, the evaluation of pathogenicity would depend on whether other established pathogenic variants have been described in this exon. These evaluation criteria would also not be applicable in the studied case to evaluate the effect of the deletion, as the associated pathomechanism is not taken into account. Numerous pathogenic truncating variants in the last exon of *MN1* have been described (181, 182), but these single nucleotide variants are characterized by the expression of C-terminally truncated proteins, which cause MCTT syndrome via a gain of function effect. The crucial issue in determining the pathogenicity of the detected deletion is therefore whether there is expression of the 1st exon resulting in the synthesis of a C-terminally truncated MN1 protein.

The recommendations for interpreting the loss-of-function PVS1 ACMG/AMP variant criterion (240), published by the ACMG and the Association for Molecular Pathology (AMP) as an addition to the 2015 ACMG/AMP guidelines for the interpretation of sequence variants (241), also contain recommendations for the interpretation of exon deletions. These recommendations do not distinguish between single and multi-exon deletions and do not specifically address deletion of the last exon. Therefore, these recommendations are also not suitable in the particular case of fetal 22q12.1 deletion.

4.2.1. LOF escape mechanism (*MN1*)

A deletion including the transcription termination site, as well as the poly-A addition site most probably produces an instable RNA (r.? > p.?) (128). An in-depth analysis enabled improved assessment of the studied deletion regarding the consequences on *MN1* expression. By

including further factors in the variant interpretation, the deletion was analysed more comprehensively so that the proximal breakpoint came into greater focus. The *CPMER* gene, which is also affected by the deletion, has not yet been well characterized. *CPMER* encodes the long non-coding RNA (lncRNA) cytoplasmic mesoderm regulator. In their recent study, Lyu et al. found that *CPMER* promotes cardiomyocyte differentiation in mice and humans. *CPMER* specifically recognized the mRNAs of the mesoderm gene *EOMES* and promoted its translation by binding to eEF1A2 (242). The deletion results in a gene fusion of *MN1* and *CPMER*, which are fused in the respective intron 1 (Figure 16). Fusion with *CPMER* could hypothetically result in a stable RNA fusion construct including the disrupted *MN1* sequence due to the presence of a transcription termination site.

Our hypothesis was supported by the RNA analyses performed. The cDNA sequencing shows that the fusion gene is transcribed and also spliced (Figure 18A). The detection of three different RNA fusion constructs, each consisting of exon 1 of the *MN1* gene and subsequent sequences of the downstream *CPMER* gene, indicates alternative splicing (Figure 18B). In silico analysis of the three detected RNA fusion constructs yields proteins with lengths of 1263 aa, 1268 aa and 1271 aa (Figure 19). The C-terminal *CPMER* sequences are short due to in-frame stop codons. As these predicted C-terminally truncated MN1 proteins lack the amino acids encoded by exon 2, the respective proteins would be disease-causing for MCTT syndrome. Furthermore, quantitative comparison of cDNA from exon 1 of MN1 with wild-type allele-specific amplicons showed double the amount of expressed exon 1 sequence compared to wild-type sequence (Figure 20). This strongly suggests that the C-terminally truncated *MN1* sequence of the mutant allele is expressed. In addition, the fusion constructs appear to be stable, as the difference in quantity can also be observed without NMD inhibition (Figure 20, orange bars). The relatively low amount of the specific amplicon for *MN1* E1-E2 *CPMER* detected is consistent with the discovery of different fusion constructs.

It was not possible to repeat the quantitative analysis with regard to the relatively high standard deviation of *MN1*: E1-E2 without NMD inhibition, as no primary sample was left. However, due to the redundancy with the second WT-specific amplicon *MN1*: E2, this was also not necessary.

Further screening of several databases led to a remarkable finding. A similar 22q12.1 deletion, which also deletes exon 2 of the *MN1* gene, was detected in a person from the 1000 Genomes Project sample pool in the study by Levy-Sakin et al. (243). The 1000 Genomes Project only includes samples from people who declared themselves to be healthy (244). In contrast to the fetal deletion studied, the deletion detected in this person does not involve the *CPMER* gene,

and thus no fusion gene is formed (Figure 22). This suggests that the deletion identified in this person without obvious severe phenotypic abnormalities is more likely to result in a loss-of-function of the *MN1* gene, which is associated with milder phenotypic abnormalities without brain malformations and incomplete penetrance. This finding supports the hypothesis that the fusion gene results in LOF escape in the fetus, leading to the expression of C-terminally truncated MN1 proteins.

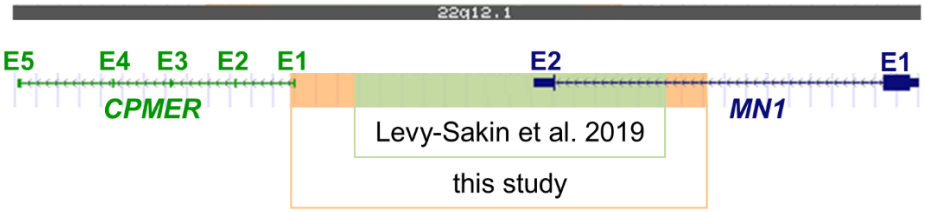

Figure 22. Detection of a similar deletion in the sample pool of the 1000 Genomes project.
The deleted region chr22:27,723,174-27,766,326 (GRCh38/hg38) detected in a person from the 1000 Genomes project in the study of Levy-Sakin et al. is highlighted in green. The deletion only affects the *MN1* gene, including exon 2. The *CPMER* gene is not affected by the deletion. In comparison: The deleted region chr22:27,714,323-27,772,264 (GRCh38/hg38) detected in the fetus in this study is highlighted in orange.

All results obtained indicate that a stable *MN1-CPMER* fusion RNA is transcribed. Therefore, the crucial question is whether this RNA can also be translated into a protein, since it is a gene fusion of a protein-coding gene and a non-coding gene, i.e. mRNA-lncRNA at RNA level. Such fusion constructs were analyzed in two recent studies by Guo et al. (245) and Sánchez-Marín et al. (246) in the context of cancer. The studies found that mRNA-lncRNA fusions can be translated into functional proteins. The mRNA-lncRNA fusions KDM4B-G039927 and EPS15L1-lncOR7C2, for example, produced proteins that could be detected by mass spectrometry and have oncogenic function through promoting cell proliferation (245). Guo et al. detected different fusion constructs from 28 cancer types of RNA-seq from the TCGA, with mRNA-lncRNA fusions being the majority compared to mRNA-mRNA, lncRNA-mRNA and lncRNA-lncRNA (245). In comparison, Sánchez-Marín et al. found in their literature review that lncRNA-mRNAs were more abundant than mRNA-lncRNAs (87 lncRNA-mRNAs versus 25 mRNA-lncRNAs). Furthermore, it was noted that mRNA-lncRNA fusions are subject to regulation by the promoter of the protein-coding gene (246). In the case of the studied deletion in the fetus, the fusion transcript would be regulated by the *MN1* gene. The *MN1* gene is

expressed in many different tissues. Expression is highest in the skeletal muscles followed by the brain (247). Although the studies indicate that fused lncRNAs are of great importance, research studies in this field are very rare, especially in the context of germline mutations.

4.2.2. Limitations of the *MN1* study

As in the *ZMPSTE24* study, the analyses performed in the *MN1* study only allow conclusions to be drawn at RNA level. The presence of three different *MN1*-*CPMER* RNA transcripts was detected. The sequences obtained were used for an in silico prediction of resulting proteins. As the RNA analysis was also relevant for the diagnosis of the fetus, Sanger sequencing of cDNA was performed to identify possible fusion constructs in order to save time. This method was only used to determine the presence of expected specific fusion transcripts. Rapid amplification of cDNA ends (RACE) would have been more suitable for detecting all different fusion transcripts, but this was not available at the time of examination. As neither a cDNA sample nor a primary tissue was any more available, this analysis could not be performed at a later time point. Quantification of the fusion transcript *MN1* E1-E2 *CPMER* (Figure 20) indicates that several fusion transcripts are transcribed. The amount of transcribed fusion transcript can only be estimated indirectly by comparing the non-specific *MN1*: E1 amplicon with the WT specific amplicons in the quantitative cDNA analysis performed. The same applies to this study as to the ZMPSTE24 study: conclusions at protein level can only be confirmed via protein analyses, such as the determination of protein sequence, expression levels and enzymatic activity. As no more tissue samples were available, it was not possible to carry out the planned protein analyses.

4.3. Interpretation of novel variants: outlook and suggestions

Standard variant annotation plays an important role in variant interpretation, but can lead to incorrect interpretation in the case of alternative molecular pathways as shown in this work. Variant interpretation is particularly important in the clinical setting, as it contributes to the diagnosis of genetic diseases and can subsequently influence the treatment of patients. Variants are very often classified according to the internationally recognized ACMG/AMP guidelines (241). In these guidelines, the "PVS1" criterion is the strongest weighted indicator of pathogenicity. This criterion evaluates variants that result in a complete loss-of-function (null variants) and are located in genes where LOF is a known disease mechanism. Therefore, accurate annotation of possible LOF variants is particularly important for correct interpretation. Various guidelines and computer-based tools have already been developed to support or

supplement standard variant annotation. One example is the “Recommendations for interpreting the loss-of-function PVS1 ACMG/AMP variant criterion” by Abou Tayoun et al. as part of the ClinGen Sequence Variant Interpretation Workgroup (240). The decision tree presented in this work assists in determining the applicability and weighting of the PVS1 criterion. Alternative mechanisms are also considered in these recommendations. However, the use of alternative start codons is only considered in the context of start loss variants. Singer-Berk et al. present a framework that advances interpretation of predicted LOF variants in research and clinical settings. This framework also provides rules for the evaluation of translation reinitiation. By applying their framework to high-confidence LOF variants in 22 genes associated with autosomal-recessive disease from gnomAD, they were able to associate 27.3% (304/1,113) of these variants with predicted LOF circumvention or potential artifacts (135).
The possibility of leaky scanning or the formation of fusion genes to escape an LOF effect was not considered in any of the recommendations found. However, when using the LOVD reading frame checker (128), possible formation of a fusion transcript/protein is indicated in the case of a deletion of the last exon/poly-A addition site.

With this work, we therefore want to raise awareness for the possibility of alternative gene expression mechanisms in the variant interpretation process and highlight the need to improve (standard) annotation pipelines in this context. Furthermore, we want to highlight the importance of additional in-depth analyses and further experiments to determine the effects of candidate variants.
Computer-based interpretation without manual assessment is currently not sufficient, especially in the clinical setting. Even in the future, computer-assisted, manual interpretation is likely to be the most promising combination for accurate variant interpretation. For professionals involved in variant interpretation, it is important to know which tools are useful with regard to the variants in question.
Overall, experimental research will always be a key factor in determining the effects of previously unknown variants.

4.4. Conclusions

In this work, we demonstrate the importance of in-depth analyses and analyses that extend beyond DNA level (e.g. RNA analyses and functional studies) in the interpretation of previously unknown variants, based on two studies. In the course of these studies, we were able to identify two novel disease-causing variants in pleiotropic genes. Remarkably, the predicted LOF effects of both variants did not match the phenotypes observed in affected individuals. Using in-depth analyses and further experimental studies, we found that both variants circumvent the predicted LOF effect via alternative expression mechanisms.

In the *ZMPSTE24* study, we identified the novel N-terminal frameshift mutation NM_005857.5:c.28_29insA in *ZMPSTE24* in affected individuals of two Pakistani families displaying MADB in homozygous state.

- Functional experiments indicated that the associated lethal RD phenotype of the predicted LOF protein p.(Leu10Tyrfs*37) is circumvented by alternative translation initiation at two N-terminal AUG codons in the mutated gene (novel ATG and M13 codon).
- Based on our findings the predictions on protein level could be complemented by the two N-terminally truncated proteins NP_005848.2:p.([Gly2_Leu10del,Gly2_Met13del]).
- To the best of our knowledge, this is the first study reporting an N-terminal frameshift mutation creating a novel, physiologically relevant translation initiation site downstream of the canonical translation start, preventing complete loss of protein function and the associated lethal consequences.
- Furthermore, our work suggests that truncating mutations in *ZMPSTE24* located within the first twelve amino acids might lead to the milder MADB phenotype rather than lethal RD, due to alternative translation initiation at the M13-codon.
- Beyond that, our results point to incorrect variant interpretation based on in-silico prediction of N-terminal frameshift mutations, since the formation of potential, novel start codons at the mutation sites is not considered in current variant interpretation algorithms. Therefore, we recommend a more in-depth analysis of these variants to prevent misinterpretation.

In the *MN1* study, we detected a previously unreported, heterozygous 22q12.1 deletion in a fetus with severe cerebral malformations. The deletion affects the *MN1* gene, including the entire last exon 2 and the lncRNA gene *CPMER*, including the entire first exon.

- RNA analyses indicate that the predicted LOF effect of the *MN1* gene, which is associated with milder phenotype abnormalities and incomplete penetrance, is circumvented by the expression of *MN1-CPMER* fusion transcripts. Expression of the C-terminally truncated MN1 proteins predicted based on the RNA sequences, is associated with MCTT syndrome, a condition characterized by distinct brain malformations. A large deletion has not yet been associated with MCTT syndrome.
- To the best of our knowledge, this is the first study reporting a gene fusion of a protein-coding gene and a non-coding gene in the germline, associated with LOF circumvention of the protein-coding gene resulting in a very rare disease phenotype.
- In addition, the results of the *MN1* study indicate that gene fusions of a protein-coding gene with a non-coding gene (mRNA-lncRNA at RNA level) might lead to protein expression, which must be taken into account in the variant interpretation process.

REFERENCES

1. Rare diseases: European Commission, Directorate-General for Research and Innovation; [Available from: https://research-and-innovation.ec.europa.eu/research-area/health/rare-diseases_en.
2. Rare Diseases Act of 2002, (2002).
3. Bureau USC. U.S. and World Population Clock U.S. Government; 2023 [Available from: https://www.census.gov/popclock/.
4. Richter T, Nestler-Parr S, Babela R, Khan ZM, Tesoro T, Molsen E, et al. Rare Disease Terminology and Definitions-A Systematic Global Review: Report of the ISPOR Rare Disease Special Interest Group. Value Health. 2015;18(6):906-14.
5. Worldometers.info. Japan Population Dover, Delaware, U.S.A. [updated 01.11.2023. Available from: https://www.worldometers.info/world-population/japan-population/.
6. Excellence NIfC. NICE citizens council report ultra orphan drugs. London: NICE. 2004.
7. Hughes DA, Tunnage B, Yeo ST. Drugs for exceptionally rare diseases: do they deserve special status for funding? Qjm. 2005;98(11):829-36.
8. Haendel M, Vasilevsky N, Unni D, Bologa C, Harris N, Rehm H, et al. How many rare diseases are there? Nat Rev Drug Discov. 2020;19(2):77-8.
9. Online Mendelian Inheritance in Man, OMIM® [Internet]. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD). Available from: https://omim.org/.
10. Orphanet: an online database of rare diseases and orphan drugs [Internet]. © INSERM 1997. Available from: https://www.orpha.net.
11. Genetic and Rare Diseases Information Center (GARD) [Internet]. National Center for Advancing Translational Sciences (NCATS). [cited 02.03.2024]. Available from: https://rarediseases.info.nih.gov/.
12. Köhler S, Gargano M, Matentzoglu N, Carmody LC, Lewis-Smith D, Vasilevsky NA, et al. The Human Phenotype Ontology in 2021. Nucleic Acids Res. 2021;49(D1):D1207-d17.
13. MedGen [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information. 2004 [cited 02.11.2023]. Available from: https://www.ncbi.nlm.nih.gov/medgen/.
14. Vasilevsky NA, Matentzoglu NA, Toro S, IV JEF, Hegde H, Unni DR, et al. Mondo: Unifying diseases for the world, by the world. medRxiv. 2022:2022.04.13.22273750.
15. Mondo Disease Ontology [Internet]. Jekyll & Minimal Mistakes. [cited 02.11.2023]. Available from: https://mondo.monarchinitiative.org/.
16. Orphadata: Free access data from Orphanet [Internet]. © INSERM 1999. Available from: http://www.orphadata.org.
17. Ferreira CR. The burden of rare diseases. Am J Med Genet A. 2019;179(6):885-92.
18. Rare diseases: European Commission, Directorate-General for Health and Food Safety; [Available from: https://health.ec.europa.eu/non-communicable-diseases/expert-group-public-health/rare-diseases_en#support-for-definition-and-codification.
19. Facts and figures on life in the European Union: European Commission, Directorate-General for Communication; [
20. Amberger JS, Bocchini CA, Scott AF, Hamosh A. OMIM.org: leveraging knowledge across phenotype-gene relationships. Nucleic Acids Res. 2019;47(D1):D1038-d43.
21. Hartley T, Balcı TB, Rojas SK, Eaton A, Canada CR, Dyment DA, et al. The unsolved rare genetic disease atlas? An analysis of the unexplained phenotypic descriptions in OMIM®. Am J Med Genet C Semin Med Genet. 2018;178(4):458-63.

22. Chong JX, Buckingham KJ, Jhangiani SN, Boehm C, Sobreira N, Smith JD, et al. The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities. Am J Hum Genet. 2015;97(2):199-215.
23. Genome - NCBI Homo sapiens Annotation Release 108 [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information. 2004 [cited 17.11.2023]. Available from: https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Homo_sapiens/108/.
24. Kudo Y. Mendelian Disease. Masuzaki, H. (eds) Fetal Morph Functional Diagnosis. Comprehensive Gynecology and Obstetrics: Springer, Singapore; 2021.
25. Ingram VM. A specific chemical difference between the globins of normal human and sickle-cell anaemia haemoglobin. Nature. 1956;178(4537):792-4.
26. Genetic Timeline: National Human Genome Research Institute; [Available from: https://www.genome.gov/Pages/Education/GeneticTimeline.pdf.
27. Lejeune J, Gautier M, Turpin R. [Study of somatic chromosomes from 9 mongoloid children]. C R Hebd Seances Acad Sci. 1959;248(11):1721-2.
28. Gusella JF, Wexler NS, Conneally PM, Naylor SL, Anderson MA, Tanzi RE, et al. A polymorphic DNA marker genetically linked to Huntington's disease. Nature. 1983;306(5940):234-8.
29. Hartley T, Lemire G, Kernohan KD, Howley HE, Adams DR, Boycott KM. New Diagnostic Approaches for Undiagnosed Rare Genetic Diseases. Annu Rev Genomics Hum Genet. 2020;21:351-72.
30. Tjio JH, Levan A. The chromosome number of man. Problems of birth defects: from Hippocrates to Thalidomide and After: Springer; 1956. p. 112-8.
31. Hamerton J, Ford C. The chromosomes of man. Nature. 1956;178:1020-3.
32. Southern EM. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J mol biol. 1975;98(3):503-17.
33. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proceedings of the national academy of sciences. 1977;74(12):5463-7.
34. Bauman J, Wiegant J, Borst P, Van Duijn P. A new method for fluorescence microscopical localization of specific DNA sequences by in situ hybridization of fluorochrome-labelled RNA. Experimental cell research. 1980;128(2):485-90.
35. Mullis KB. The unusual origin of the polymerase chain reaction. Sci Am. 1990;262(4):56-61, 4-5.
36. Jeffreys AJ, Wilson V, Thein SL. Hypervariable 'minisatellite'regions in human DNA. Nature. 1985;314(6006):67-73.
37. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science. 1995;270(5235):467-70.
38. Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, Pals G. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. Nucleic acids research. 2002;30(12):e57-e.
39. Nyrén P, Pettersson B, Uhlén M. Solid phase DNA minisequencing by an enzymatic luminometric inorganic pyrophosphate detection assay. Anal Biochem. 1993;208(1):171-5.
40. Ronaghi M, Karamohamed S, Pettersson B, Uhlén M, Nyrén P. Real-time DNA sequencing using detection of pyrophosphate release. Anal Biochem. 1996;242(1):84-9.
41. Adams CP, Kron SJ. Method for performing amplification of nucleic acid with two primers bound to a single solid support. Google Patents; 1997.
42. Kawashima E, Farinelli L, Mayer P. Method of nucleic acid sequencing. Google Patents; 2002.
43. Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, et al. Accurate multiplex polony sequencing of an evolved bacterial genome. Science. 2005;309(5741):1728-32.

44. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2005;437(7057):376-80.
45. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature methods. 2008;5(7):621-8.
46. Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, et al. The transcriptional landscape of the yeast genome defined by RNA sequencing. Science. 2008;320(5881):1344-9.
47. Pogue RE, Cavalcanti DP, Shanker S, Andrade RV, Aguiar LR, de Carvalho JL, et al. Rare genetic diseases: update on diagnosis, treatment and online resources. Drug Discov Today. 2018;23(1):187-95.
48. Nguengang Wakap S, Lambert DM, Olry A, Rodwell C, Gueydan C, Lanneau V, et al. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. Eur J Hum Genet. 2020;28(2):165-73.
49. Boycott KM, Rath A, Chong JX, Hartley T, Alkuraya FS, Baynam G, et al. International Cooperation to Enable the Diagnosis of All Rare Genetic Diseases. Am J Hum Genet. 2017;100(5):695-705.
50. Shashi V, McConkie-Rosell A, Rosell B, Schoch K, Vellore K, McDonald M, et al. The utility of the traditional medical genetics diagnostic evaluation in the context of next-generation sequencing for undiagnosed genetic disorders. Genet Med. 2014;16(2):176-82.
51. Marwaha S, Knowles JW, Ashley EA. A guide for the diagnosis of rare and undiagnosed disease: beyond the exome. Genome Med. 2022;14(1):23.
52. Seaby EG, Rehm HL, O'Donnell-Luria A. Strategies to Uplift Novel Mendelian Gene Discovery for Improved Clinical Outcomes. Front Genet. 2021;12:674295.
53. Gilissen C, Hoischen A, Brunner HG, Veltman JA. Disease gene identification strategies for exome sequencing. Eur J Hum Genet. 2012;20(5):490-7.
54. Onodera O, Tsuji S. [Positional cloning--current status and future directions]. Nihon Rinsho. 1993;51(9):2225-33.
55. Jagodic M, Stridh P. Positional Gene Cloning in Experimental Populations. Methods Mol Biol. 2016;1304:3-24.
56. Kerem B, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, et al. Identification of the cystic fibrosis gene: genetic analysis. Science. 1989;245(4922):1073-80.
57. Lander ES, Botstein D. Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. Science. 1987;236(4808):1567-70.
58. Claussnitzer M, Cho JH, Collins R, Cox NJ, Dermitzakis ET, Hurles ME, et al. A brief history of human disease genetics. Nature. 2020;577(7789):179-89.
59. Bamshad MJ, Nickerson DA, Chong JX. Mendelian Gene Discovery: Fast and Furious with No End in Sight. Am J Hum Genet. 2019;105(3):448-55.
60. Vissers LE, Veltman JA, van Kessel AG, Brunner HG. Identification of disease genes by whole genome CGH arrays. Hum Mol Genet. 2005;14 Spec No. 2:R215-23.
61. Ho SS, Urban AE, Mills RE. Structural variation in the sequencing era. Nat Rev Genet. 2020;21(3):171-89.
62. Davis RL, Kumar KR, Puttick C, Liang C, Ahmad KE, Edema-Hildebrand F, et al. Use of Whole-Genome Sequencing for Mitochondrial Disease Diagnosis. Neurology. 2022;99(7):e730-e42.
63. Ebler J, Haukness M, Pesout T, Marschall T, Paten B. Haplotype-aware diplotyping from noisy long reads. Genome Biol. 2019;20(1):116.
64. Mantere T, Kersten S, Hoischen A. Long-Read Sequencing Emerging in Medical Genetics. Front Genet. 2019;10:426.
65. Gigante S, Gouil Q, Lucattini A, Keniry A, Beck T, Tinning M, et al. Using long-read sequencing to detect imprinted DNA methylation. Nucleic Acids Res. 2019;47(8):e46.

66. Consortium CP-G. Computational pan-genomics: status, promises and challenges. Brief Bioinform. 2018;19(1):118-35.
67. Liao WW, Asri M, Ebler J, Doerr D, Haukness M, Hickey G, et al. A draft human pangenome reference. Nature. 2023;617(7960):312-24.
68. Grabowski P, Hesse S, Hollizeck S, Rohlfs M, Behrends U, Sherkat R, et al. Proteome Analysis of Human Neutrophil Granulocytes From Patients With Monogenic Disease Using Data-independent Acquisition. Mol Cell Proteomics. 2019;18(4):760-72.
69. Kremer LS, Bader DM, Mertes C, Kopajtich R, Pichler G, Iuso A, et al. Genetic diagnosis of Mendelian disorders via RNA sequencing. Nat Commun. 2017;8:15824.
70. Frésard L, Smail C, Ferraro NM, Teran NA, Li X, Smith KS, et al. Identification of rare-disease genes using blood transcriptome sequencing and large control cohorts. Nat Med. 2019;25(6):911-9.
71. Sirrs S, van Karnebeek CD, Peng X, Shyr C, Tarailo-Graovac M, Mandal R, et al. Defects in fatty acid amide hydrolase 2 in a male with neurologic and psychiatric symptoms. Orphanet J Rare Dis. 2015;10:38.
72. Abela L, Simmons L, Steindl K, Schmitt B, Mastrangelo M, Joset P, et al. N(8)-acetylspermidine as a potential plasma biomarker for Snyder-Robinson syndrome identified by clinical metabolomics. J Inherit Metab Dis. 2016;39(1):131-7.
73. Berdasco M, Esteller M. Genetic syndromes caused by mutations in epigenetic genes. Hum Genet. 2013;132(4):359-83.
74. Frosk P, Arts HH, Philippe J, Gunn CS, Brown EL, Chodirker B, et al. A truncating mutation in CEP55 is the likely cause of MARCH, a novel syndrome affecting neuronal mitosis. J Med Genet. 2017;54(7):490-501.
75. Oláhová M, Yoon WH, Thompson K, Jangam S, Fernandez L, Davidson JM, et al. Biallelic Mutations in ATP5F1D, which Encodes a Subunit of ATP Synthase, Cause a Metabolic Disorder. Am J Hum Genet. 2018;102(3):494-504.
76. Breslow DK, Hoogendoorn S, Kopp AR, Morgens DW, Vu BK, Kennedy MC, et al. A CRISPR-based screen for Hedgehog signaling provides insights into ciliary function and ciliopathies. Nat Genet. 2018;50(3):460-71.
77. 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.
78. Posey JE, O'Donnell-Luria AH, Chong JX, Harel T, Jhangiani SN, Coban Akdemir ZH, et al. Insights into genetics, human biology and disease gleaned from family based genomic studies. Genet Med. 2019;21(4):798-812.
79. Seaby EG, Ennis S. Challenges in the diagnosis and discovery of rare genetic disorders using contemporary sequencing technologies. Brief Funct Genomics. 2020;19(4):243-58.
80. Adams DR, Eng CM. Next-Generation Sequencing to Diagnose Suspected Genetic Disorders. N Engl J Med. 2018;379(14):1353-62.
81. Retterer K, Juusola J, Cho MT, Vitazka P, Millan F, Gibellini F, et al. Clinical application of whole-exome sequencing across clinical indications. Genet Med. 2016;18(7):696-704.
82. Smedley D, Jacobsen JO, Jäger M, Köhler S, Holtgrewe M, Schubach M, et al. Next-generation diagnostics and disease-gene discovery with the Exomiser. Nat Protoc. 2015;10(12):2004-15.
83. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581(7809):434-43.
84. International Rare Diseases Research Consortium (IRDiRC) [Available from: https://irdirc.org/.
85. Philippakis AA, Azzariti DR, Beltran S, Brookes AJ, Brownstein CA, Brudno M, et al. The Matchmaker Exchange: a platform for rare disease gene discovery. Hum Mutat. 2015;36(10):915-21.

86. Sobreira N, Schiettecatte F, Valle D, Hamosh A. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. Hum Mutat. 2015;36(10):928-30.
87. Buske OJ, Girdea M, Dumitriu S, Gallinger B, Hartley T, Trang H, et al. PhenomeCentral: a portal for phenotypic and genotypic matchmaking of patients with rare genetic diseases. Hum Mutat. 2015;36(10):931-40.
88. Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D, et al. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. Am J Hum Genet. 2009;84(4):524-33.
89. Swaminathan GJ, Bragin E, Chatzimichali EA, Corpas M, Bevan AP, Wright CF, et al. DECIPHER: web-based, community resource for clinical interpretation of rare variants in developmental disorders. Hum Mol Genet. 2012;21(R1):R37-44.
90. 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 Res. 2018;46(D1):D1062-d7.
91. Quinodoz M, Peter VG, Bedoni N, Royer Bertrand B, Cisarova K, Salmaninejad A, et al. AutoMap is a high performance homozygosity mapping tool using next-generation sequencing data. Nat Commun. 2021;12(1):518.
92. Ceballos FC, Hazelhurst S, Ramsay M. Assessing runs of Homozygosity: a comparison of SNP Array and whole genome sequence low coverage data. BMC Genomics. 2018;19(1):106.
93. Vahidnezhad H, Youssefian L, Jazayeri A, Uitto J. Research Techniques Made Simple: Genome-Wide Homozygosity/Autozygosity Mapping Is a Powerful Tool for Identifying Candidate Genes in Autosomal Recessive Genetic Diseases. J Invest Dermatol. 2018;138(9):1893-900.
94. Erzurumluoglu AM, Shihab HA, Rodriguez S, Gaunt TR, Day IN. Importance of Genetic Studies in Consanguineous Populations for the Characterization of Novel Human Gene Functions. Ann Hum Genet. 2016;80(3):187-96.
95. Gonzales PR, Andersen EF, Brown TR, Horner VL, Horwitz J, Rehder CW, et al. Interpretation and reporting of large regions of homozygosity and suspected consanguinity/uniparental disomy, 2021 revision: A technical standard of the American College of Medical Genetics and Genomics (ACMG). Genet Med. 2022;24(2):255-61.
96. Abbas HA, Yunis K. The effect of consanguinity on neonatal outcomes and health. Hum Hered. 2014;77(1-4):87-92.
97. Stearns FW. One hundred years of pleiotropy: a retrospective. Genetics. 2010;186(3):767-73.
98. Ittisoponpisan S, Alhuzimi E, Sternberg MJ, David A. Landscape of Pleiotropic Proteins Causing Human Disease: Structural and System Biology Insights. Hum Mutat. 2017;38(3):289-96.
99. Magrinelli F, Balint B, Bhatia KP. Challenges in Clinicogenetic Correlations: One Gene - Many Phenotypes. Mov Disord Clin Pract. 2021;8(3):299-310.
100. Jen JC. Familial Hemiplegic Migraine Adam MP, Feldman J, Mirzaa GM, et al., editors. GeneReviews®: Seattle (WA): University of Washington, Seattle; 2001 [updated 2021 Apr 29. Available from: https://www.ncbi.nlm.nih.gov/books/NBK1388/.
101. Murken JD, Grimm T, Holinski-Feder E, Zerres K. Taschenlehrbuch Humangenetik: Georg Thieme Verlag; 2017.
102. Muller H, editor Further studies on the nature and causes of gene mutations. Proceedings of the 6th International Congress of Genetics, 1932; 1932.
103. Vihinen M. Systematic errors in annotations of truncations, loss-of-function and synonymous variants. Front Genet. 2023;14:1015017.
104. Vihinen M. Functional effects of protein variants. Biochimie. 2021;180:104-20.

105. Gerasimavicius L, Livesey BJ, Marsh JA. Loss-of-function, gain-of-function and dominant-negative mutations have profoundly different effects on protein structure. Nat Commun. 2022;13(1):3895.
106. Chang YF, Imam JS, Wilkinson MF. The nonsense-mediated decay RNA surveillance pathway. Annu Rev Biochem. 2007;76:51-74.
107. Silva AL, Romão L. The mammalian nonsense-mediated mRNA decay pathway: to decay or not to decay! Which players make the decision? FEBS Lett. 2009;583(3):499-505.
108. Thermann R, Neu-Yilik G, Deters A, Frede U, Wehr K, Hagemeier C, et al. Binary specification of nonsense codons by splicing and cytoplasmic translation. Embo j. 1998;17(12):3484-94.
109. Carter MS, Doskow J, Morris P, Li S, Nhim RP, Sandstedt S, et al. A regulatory mechanism that detects premature nonsense codons in T-cell receptor transcripts in vivo is reversed by protein synthesis inhibitors in vitro. J Biol Chem. 1995;270(48):28995-9003.
110. Metze S, Herzog VA, Ruepp MD, Mühlemann O. Comparison of EJC-enhanced and EJC-independent NMD in human cells reveals two partially redundant degradation pathways. Rna. 2013;19(10):1432-48.
111. Sherlock ME, Baquero Galvis L, Vicens Q, Kieft JS, Jagannathan S. Principles, mechanisms, and biological implications of translation termination-reinitiation. Rna. 2023;29(7):865-84.
112. Muñoz O, Lore M, Jagannathan S. The long and short of EJC-independent nonsense-mediated RNA decay. Biochem Soc Trans. 2023;51(3):1121-9.
113. Le Hir H, Izaurralde E, Maquat LE, Moore MJ. The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. Embo j. 2000;19(24):6860-9.
114. Le Hir H, Gatfield D, Izaurralde E, Moore MJ. The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. Embo j. 2001;20(17):4987-97.
115. Ishigaki Y, Li X, Serin G, Maquat LE. Evidence for a pioneer round of mRNA translation: mRNAs subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20. Cell. 2001;106(5):607-17.
116. Kashima I, Yamashita A, Izumi N, Kataoka N, Morishita R, Hoshino S, et al. Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay. Genes Dev. 2006;20(3):355-67.
117. Nagy E, Maquat LE. A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. Trends Biochem Sci. 1998;23(6):198-9.
118. Lindeboom RG, Supek F, Lehner B. The rules and impact of nonsense-mediated mRNA decay in human cancers. Nat Genet. 2016;48(10):1112-8.
119. Lindeboom RGH, Vermeulen M, Lehner B, Supek F. The impact of nonsense-mediated mRNA decay on genetic disease, gene editing and cancer immunotherapy. Nat Genet. 2019;51(11):1645-51.
120. Hogg JR, Goff SP. Upf1 senses 3'UTR length to potentiate mRNA decay. Cell. 2010;143(3):379-89.
121. Annibaldis G, Domanski M, Dreos R, Contu L, Carl S, Kläy N, et al. Readthrough of stop codons under limiting ABCE1 concentration involves frameshifting and inhibits nonsense-mediated mRNA decay. Nucleic Acids Res. 2020;48(18):10259-79.
122. Brogna S, Wen J. Nonsense-mediated mRNA decay (NMD) mechanisms. Nat Struct Mol Biol. 2009;16(2):107-13.
123. Klonowski J, Liang Q, Coban-Akdemir Z, Lo C, Kostka D. aenmd: annotating escape from nonsense-mediated decay for transcripts with protein-truncating variants. Bioinformatics. 2023;39(9).

124. Hoek TA, Khuperkar D, Lindeboom RGH, Sonneveld S, Verhagen BMP, Boersma S, et al. Single-Molecule Imaging Uncovers Rules Governing Nonsense-Mediated mRNA Decay. Mol Cell. 2019;75(2):324-39.e11.
125. Zhang J, Maquat LE. Evidence that translation reinitiation abrogates nonsense-mediated mRNA decay in mammalian cells. Embo j. 1997;16(4):826-33.
126. Pereira FJ, Teixeira A, Kong J, Barbosa C, Silva AL, Marques-Ramos A, et al. Resistance of mRNAs with AUG-proximal nonsense mutations to nonsense-mediated decay reflects variables of mRNA structure and translational activity. Nucleic Acids Res. 2015;43(13):6528-44.
127. Paulsen M, Lund C, Akram Z, Winther JR, Horn N, Møller LB. Evidence that translation reinitiation leads to a partially functional Menkes protein containing two copper-binding sites. Am J Hum Genet. 2006;79(2):214-29.
128. Fokkema IF, Taschner PE, Schaafsma GC, Celli J, Laros JF, den Dunnen JT. LOVD v.2.0: the next generation in gene variant databases. Hum Mutat. 2011;32(5):557-63.
129. Chase A, Ernst T, Fiebig A, Collins A, Grand F, Erben P, et al. TFG, a target of chromosome translocations in lymphoma and soft tissue tumors, fuses to GPR128 in healthy individuals. Haematologica. 2010;95(1):20-6.
130. den Dunnen JT, Dalgleish R, Maglott DR, Hart RK, Greenblatt MS, McGowan-Jordan J, et al. HGVS Recommendations for the Description of Sequence Variants: 2016 Update. Hum Mutat. 2016;37(6):564-9.
131. Hebbar P, Sowmya SK, editors. Genomic Variant Annotation: A Comprehensive Review of Tools and Techniques. Intelligent Systems Design and Applications; 2022 2022//; Cham: Springer International Publishing.
132. Samuels DC, Yu H, Guo Y. Is it time to reassess variant annotation? Trends Genet. 2022;38(6):521-3.
133. McCarthy DJ, Humburg P, Kanapin A, Rivas MA, Gaulton K, Cazier JB, et al. Choice of transcripts and software has a large effect on variant annotation. Genome Med. 2014;6(3):26.
134. Chen S, Francioli LC, Goodrich JK, Collins RL, Kanai M, Wang Q, et al. A genomic mutational constraint map using variation in 76,156 human genomes. Nature. 2024;625(7993):92-100.
135. Singer-Berk M, Gudmundsson S, Baxter S, Seaby EG, England E, Wood JC, et al. Advanced variant classification framework reduces the false positive rate of predicted loss-of-function variants in population sequencing data. Am J Hum Genet. 2023;110(9):1496-508.
136. Narasimhan VM, Hunt KA, Mason D, Baker CL, Karczewski KJ, Barnes MR, et al. Health and population effects of rare gene knockouts in adult humans with related parents. Science. 2016;352(6284):474-7.
137. Saleheen D, Natarajan P, Armean IM, Zhao W, Rasheed A, Khetarpal SA, et al. Human knockouts and phenotypic analysis in a cohort with a high rate of consanguinity. Nature. 2017;544(7649):235-9.
138. Whiffin N, Armean IM, Kleinman A, Marshall JL, Minikel EV, Goodrich JK, et al. The effect of LRRK2 loss-of-function variants in humans. Nat Med. 2020;26(6):869-77.
139. Baghdadi M, Hinterding HM, Partridge L, Deelen J. From mutation to mechanism: deciphering the molecular function of genetic variants linked to human ageing. Brief Funct Genomics. 2022;21(1):13-23.
140. Vihinen M. One Gene, Several Diseases: The Characteristics of Pleiotropic Proteins. Hum Mutat. 2017;38(3):241.
141. Nucleotide [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information. 2004 [cited 22.02.2024]. Available from: https://www.ncbi.nlm.nih.gov/nuccore.

142. An atlas of the protein-coding genes in the human, pig, and mouse brain [Internet]. 2020 [cited Mar 6]. Available from: https://www.proteinatlas.org/ENSG00000084073-ZMPSTE24.
143. Protein [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information. 2004 [cited 17.11.2023]. Available from: https://www.ncbi.nlm.nih.gov/protein.
144. Spear ED, Hsu ET, Nie L, Carpenter EP, Hrycyna CA, Michaelis S. ZMPSTE24 missense mutations that cause progeroid diseases decrease prelamin A cleavage activity and/or protein stability. Dis Model Mech. 2018;11(7).
145. Rawlings ND, Barrett AJ, Thomas PD, Huang X, Bateman A, Finn RD. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. Nucleic Acids Res. 2018;46(D1):D624-d32.
146. Gene [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information. 2004 [cited 16.02.2024]. Available from: https://www.ncbi.nlm.nih.gov/gene.
147. Gao J, Liao J, Yang GY. CAAX-box protein, prenylation process and carcinogenesis. Am J Transl Res. 2009;1(3):312-25.
148. Schmidt WK, Tam A, Fujimura-Kamada K, Michaelis S. Endoplasmic reticulum membrane localization of Rce1p and Ste24p, yeast proteases involved in carboxyl-terminal CAAX protein processing and amino-terminal a-factor cleavage. Proc Natl Acad Sci U S A. 1998;95(19):11175-80.
149. Barrowman J, Wiley PA, Hudon-Miller SE, Hrycyna CA, Michaelis S. Human ZMPSTE24 disease mutations: residual proteolytic activity correlates with disease severity. Hum Mol Genet. 2012;21(18):4084-93.
150. Quigley A, Dong YY, Pike AC, Dong L, Shrestha L, Berridge G, et al. The structural basis of ZMPSTE24-dependent laminopathies. Science. 2013;339(6127):1604-7.
151. Marcelot A, Worman HJ, Zinn-Justin S. Protein structural and mechanistic basis of progeroid laminopathies. Febs j. 2021;288(9):2757-72.
152. Mattout A, Dechat T, Adam SA, Goldman RD, Gruenbaum Y. Nuclear lamins, diseases and aging. Curr Opin Cell Biol. 2006;18(3):335-41.
153. Butin-Israeli V, Adam SA, Goldman AE, Goldman RD. Nuclear lamin functions and disease. Trends Genet. 2012;28(9):464-71.
154. Gruenbaum Y, Foisner R. Lamins: nuclear intermediate filament proteins with fundamental functions in nuclear mechanics and genome regulation. Annu Rev Biochem. 2015;84:131-64.
155. Corrigan DP, Kuszczak D, Rusinol AE, Thewke DP, Hrycyna CA, Michaelis S, et al. Prelamin A endoproteolytic processing in vitro by recombinant Zmpste24. Biochem J. 2005;387(Pt 1):129-38.
156. Bergo MO, Gavino B, Ross J, Schmidt WK, Hong C, Kendall LV, et al. Zmpste24 deficiency in mice causes spontaneous bone fractures, muscle weakness, and a prelamin A processing defect. Proc Natl Acad Sci U S A. 2002;99(20):13049-54.
157. Pendás AM, Zhou Z, Cadiñanos J, Freije JM, Wang J, Hultenby K, et al. Defective prelamin A processing and muscular and adipocyte alterations in Zmpste24 metalloproteinase-deficient mice. Nat Genet. 2002;31(1):94-9.
158. Young SG, Fong LG, Michaelis S. Prelamin A, Zmpste24, misshapen cell nuclei, and progeria--new evidence suggesting that protein farnesylation could be important for disease pathogenesis. J Lipid Res. 2005;46(12):2531-58.
159. Barrowman J, Hamblet C, George CM, Michaelis S. Analysis of prelamin A biogenesis reveals the nucleus to be a CaaX processing compartment. Mol Biol Cell. 2008;19(12):5398-408.
160. Barrowman J, Hamblet C, Kane MS, Michaelis S. Requirements for efficient proteolytic cleavage of prelamin A by ZMPSTE24. PLoS One. 2012;7(2):e32120.

161. Zheng M, Jin G, Zhou Z. Post-Translational Modification of Lamins: Mechanisms and Functions. Front Cell Dev Biol. 2022;10:864191.
162. Online Mendelian Inheritance in Man, OMIM®, MIM Number: 150330 [Internet]. Johns Hopkins University, Baltimore, MD. [cited 13.02.2024]. Available from: https://omim.org/entry/150330.
163. Smallwood DT, Shackleton S. Lamin A-linked progerias: is farnesylation the be all and end all? Biochem Soc Trans. 2010;38(Pt 1):281-6.
164. Lattanzi G, Maggi L, Araujo-Vilar D. Laminopathies. Nucleus. 2018;9(1):543-4.
165. Agarwal AK, Fryns JP, Auchus RJ, Garg A. Zinc metalloproteinase, ZMPSTE24, is mutated in mandibuloacral dysplasia. Hum Mol Genet. 2003;12(16):1995-2001.
166. Simha V, Agarwal AK, Oral EA, Fryns JP, Garg A. Genetic and phenotypic heterogeneity in patients with mandibuloacral dysplasia-associated lipodystrophy. J Clin Endocrinol Metab. 2003;88(6):2821-4.
167. Online Mendelian Inheritance in Man, OMIM®, MIM Number: 608612 [Internet]. Johns Hopkins University, Baltimore, MD. [cited 14.02.2024]. Available from: https://www.omim.org/entry/608612.
168. MedlinePlus [Internet] [Internet]. Bethesda (MD): National Library of Medicine (US); Mandibuloacral dysplasia. [cited 02.03.2024]. Available from: https://medlineplus.gov/genetics/condition/mandibuloacral-dysplasia/#frequency.
169. Cenni V, D'Apice MR, Garagnani P, Columbaro M, Novelli G, Franceschi C, et al. Mandibuloacral dysplasia: A premature ageing disease with aspects of physiological ageing. Ageing Res Rev. 2018;42:1-13.
170. Smigiel R, Jakubiak A, Esteves-Vieira V, Szela K, Halon A, Jurek T, et al. Novel frameshifting mutations of the ZMPSTE24 gene in two siblings affected with restrictive dermopathy and review of the mutations described in the literature. Am J Med Genet A. 2010;152a(2):447-52.
171. Online Mendelian Inheritance in Man, OMIM®, MIM Number: 275210 [Internet]. Johns Hopkins University, Baltimore, MD. [cited 14.02.2024]. Available from: https://www.omim.org/entry/275210.
172. Lekanne Deprez RH, Groen NA, van Biezen NA, Hagemeijer A, van Drunen E, Koper JW, et al. A t(4;22) in a meningioma points to the localization of a putative tumor-suppressor gene. Am J Hum Genet. 1991;48(4):783-90.
173. Lekanne Deprez RH, Riegman PH, Groen NA, Warringa UL, van Biezen NA, Molijn AC, et al. Cloning and characterization of MN1, a gene from chromosome 22q11, which is disrupted by a balanced translocation in a meningioma. Oncogene. 1995;10(8):1521-8.
174. Heuser M, Beutel G, Krauter J, Döhner K, von Neuhoff N, Schlegelberger B, et al. High meningioma 1 (MN1) expression as a predictor for poor outcome in acute myeloid leukemia with normal cytogenetics. Blood. 2006;108(12):3898-905.
175. Liu W, Lan Y, Pauws E, Meester-Smoor MA, Stanier P, Zwarthoff EC, et al. The Mn1 transcription factor acts upstream of Tbx22 and preferentially regulates posterior palate growth in mice. Development. 2008;135(23):3959-68.
176. Zhang X, Dowd DR, Moore MC, Kranenburg TA, Meester-Smoor MA, Zwarthoff EC, et al. Meningioma 1 is required for appropriate osteoblast proliferation, motility, differentiation, and function. J Biol Chem. 2009;284(27):18174-83.
177. Meester-Smoor MA, Vermeij M, van Helmond MJ, Molijn AC, van Wely KH, Hekman AC, et al. Targeted disruption of the Mn1 oncogene results in severe defects in development of membranous bones of the cranial skeleton. Mol Cell Biol. 2005;25(10):4229-36.
178. Said E, Cuschieri A, Vermeesch J, Fryns JP. Toriello-Carey syndrome with a 6Mb interstitial deletion at 22q12 detected by array CGH. Am J Med Genet A. 2011;155a(6):1390-2.
179. Davidson TB, Sanchez-Lara PA, Randolph LM, Krieger MD, Wu SQ, Panigrahy A, et al. Microdeletion del(22)(q12.2) encompassing the facial development-associated gene, MN1 (meningioma 1) in a child with Pierre-Robin sequence (including cleft palate) and

neurofibromatosis 2 (NF2): a case report and review of the literature. BMC Med Genet. 2012;13:19.
180. Breckpot J, Anderlid BM, Alanay Y, Blyth M, Brahimi A, Duban-Bedu B, et al. Chromosome 22q12.1 microdeletions: confirmation of the MN1 gene as a candidate gene for cleft palate. Eur J Hum Genet. 2016;24(1):51-8.
181. Mak CCY, Doherty D, Lin AE, Vegas N, Cho MT, Viot G, et al. MN1 C-terminal truncation syndrome is a novel neurodevelopmental and craniofacial disorder with partial rhombencephalosynapsis. Brain. 2020;143(1):55-68.
182. Miyake N, Takahashi H, Nakamura K, Isidor B, Hiraki Y, Koshimizu E, et al. Gain-of-Function MN1 Truncation Variants Cause a Recognizable Syndrome with Craniofacial and Brain Abnormalities. Am J Hum Genet. 2020;106(1):13-25.
183. Online Mendelian Inheritance in Man, OMIM®, MIM Number: 618774 [Internet]. Johns Hopkins University, Baltimore, MD. [cited 23.02.2024]. Available from: https://www.omim.org/entry/618774.
184. Mak CC, Fung JL, Lee M, Lin AE, Amiel J, Doherty D, et al. MN1 C-terminal truncation syndrome Adam MP, Feldman J, Mirzaa GM, et al., editors. GeneReviews®: Seattle (WA): University of Washington, Seattle; 2020 [Available from: https://www.ncbi.nlm.nih.gov/books/NBK560443/.
185. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3--new capabilities and interfaces. Nucleic Acids Res. 2012;40(15):e115.
186. Pedersen AG, Nielsen H. Neural network prediction of translation initiation sites in eukaryotes: perspectives for EST and genome analysis. Proc Int Conf Intell Syst Mol Biol. 1997;5:226-33.
187. Salamov AA, Nishikawa T, Swindells MB. Assessing protein coding region integrity in cDNA sequencing projects. Bioinformatics. 1998;14(5):384-90.
188. Madeira F, Pearce M, Tivey ARN, Basutkar P, Lee J, Edbali O, et al. Search and sequence analysis tools services from EMBL-EBI in 2022. Nucleic Acids Res. 2022;50(W1):W276-w9.
189. www.varSEAK.bio, developed by JSI medical systems GmbH, Ettenheim, Germany.
190. Invitrogen. GFP Fusion TOPO® TA Expression Kits, Version J. 2004.
191. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. Genome Res. 2002;12(6):996-1006.
192. Samaras P, Schmidt T, Frejno M, Gessulat S, Reinecke M, Jarzab A, et al. ProteomicsDB: a multi-omics and multi-organism resource for life science research. Nucleic Acids Res. 2020;48(D1):D1153-d63.
193. Schmidt T, Samaras P, Frejno M, Gessulat S, Barnert M, Kienegger H, et al. ProteomicsDB. Nucleic Acids Res. 2018;46(D1):D1271-d81.
194. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. Nat Methods. 2012;9(7):676-82.
195. Martin FJ, Amode MR, Aneja A, Austine-Orimoloye O, Azov AG, Barnes I, et al. Ensembl 2023. Nucleic Acids Res. 2023;51(D1):D933-d41.
196. Alternative splicing by Swiss Institute of Bioinformatics, Graph accession number : NC_000001_1104 [Internet]. Ludwig Institute for Cancer Research and the Swiss Institute of Bioinformatics. [cited 19.04.2024]. Available from: https://epd.expasy.org/cgi-bin/tromer/tromergraph2draw.pl?db=hg38&species=H.+sapiens&tromer=NC_000001_1104.
197. Lefter M, Vis JK, Vermaat M, den Dunnen JT, Taschner PEM, Laros JFJ. Mutalyzer 2: next generation HGVS nomenclature checker. Bioinformatics. 2021;37(18):2811-7.
198. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods. 2014;11(4):361-2.
199. Schaflinger E, Blatterer J, Khan AS, Kaufmann L, Auinger L, Tatrai B, et al. An exceptional biallelic N-terminal frame shift mutation in ZMPSTE24 leads to non-lethal

progeria due to possible utilization of a downstream alternative start codon. Gene. 2022;833:146582.
200. Kaufmann L, Pilic J, Auinger L, Mayer AL, Blatterer J, Semmler-Bruckner J, et al. Analysis of a non-lethal biallelic frameshift mutation in ZMPSTE24 reveals utilization of alternative translation initiation codons. Clin Genet. 2023;104(4):491-6.
201. Matulevičienė A, Meškienė R, Morkūnienė A, Ambrozaitytė L, Meškauskas R, Garunkštienė R, et al. Frame shift mutations of the ZMPSTE24 gene in two siblings with restrictive dermopathy. Clin Dysmorphol. 2016;25(1):7-11.
202. Navarro CL, Esteves-Vieira V, Courrier S, Boyer A, Duong Nguyen T, Huong le TT, et al. New ZMPSTE24 (FACE1) mutations in patients affected with restrictive dermopathy or related progeroid syndromes and mutation update. Eur J Hum Genet. 2014;22(8):1002-11.
203. Moulson CL, Go G, Gardner JM, van der Wal AC, Smitt JH, van Hagen JM, et al. Homozygous and compound heterozygous mutations in ZMPSTE24 cause the laminopathy restrictive dermopathy. J Invest Dermatol. 2005;125(5):913-9.
204. Miyoshi Y, Akagi M, Agarwal AK, Namba N, Kato-Nishimura K, Mohri I, et al. Severe mandibuloacral dysplasia caused by novel compound heterozygous ZMPSTE24 mutations in two Japanese siblings. Clin Genet. 2008;73(6):535-44.
205. Kim JY, Kim SH, Ji HY, Choi SJ, Oh SY, Ki CS, et al. A case of restrictive dermopathy with novel ZMPSTE24 gene mutation. Pediatr Dev Pathol. 2012;15(5):393-6.
206. Lucas-Herald AK, Zürbig P, Mason A, Kinning E, Brown CE, Mansoorian B, et al. Proteomic Evidence of Biological Aging in a Child with a Compound Heterozygous ZMPSTE24 Mutation. Proteomics Clin Appl. 2019;13(2):e1800135.
207. Cunningham VJ, D'Apice MR, Licata N, Novelli G, Cundy T. Skeletal phenotype of mandibuloacral dysplasia associated with mutations in ZMPSTE24. Bone. 2010;47(3):591-7.
208. Navarro CL, Cadiñanos J, De Sandre-Giovannoli A, Bernard R, Courrier S, Boccaccio I, et al. Loss of ZMPSTE24 (FACE-1) causes autosomal recessive restrictive dermopathy and accumulation of Lamin A precursors. Hum Mol Genet. 2005;14(11):1503-13.
209. Ben Yaou R, Navarro C, Quijano-Roy S, Bertrand AT, Massart C, De Sandre-Giovannoli A, et al. Type B mandibuloacral dysplasia with congenital myopathy due to homozygous ZMPSTE24 missense mutation. Eur J Hum Genet. 2011;19(6):647-54.
210. Sander CS, Salman N, van Geel M, Broers JL, Al-Rahmani A, Chedid F, et al. A newly identified splice site mutation in ZMPSTE24 causes restrictive dermopathy in the Middle East. Br J Dermatol. 2008;159(4):961-7.
211. Jagadeesh S, Bhat L, Suresh I, Muralidhar SL. Prenatal diagnosis of restrictive dermopathy. Indian Pediatr. 2009;46(4):349-51.
212. Saxena R, Kohli S, Verma IC. Novel human pathological mutations. Gene symbol: Zmpste24. Disease: Restrictive dermopathy. Hum Genet. 2010;127(4):480.
213. Chen M, Kuo HH, Huang YC, Ke YY, Chang SP, Chen CP, et al. A case of restrictive dermopathy with complete chorioamniotic membrane separation caused by a novel homozygous nonsense mutation in the ZMPSTE24 gene. Am J Med Genet A. 2009;149a(7):1550-4.
214. Lu CS, Wu SC, Hou JW, Chu CP, Tseng LL, Lue HC. Restrictive dermopathy: report of two siblings. Pediatr Neonatol. 2013;54(3):198-201.
215. Ahmad Z, Zackai E, Medne L, Garg A. Early onset mandibuloacral dysplasia due to compound heterozygous mutations in ZMPSTE24. Am J Med Genet A. 2010;152a(11):2703-10.
216. Agarwal AK, Zhou XJ, Hall RK, Nicholls K, Bankier A, Van Esch H, et al. Focal segmental glomerulosclerosis in patients with mandibuloacral dysplasia owing to ZMPSTE24 deficiency. J Investig Med. 2006;54(4):208-13.
217. Alarcón PI, Mujica I, Sanz P, García CJ, Gilgenkrantz S. Mandibuloacral dysplasia with type B lipodystrophy in a patient from Chile. Am J Med Genet A. 2019;179(6):893-5.

218. Kwan JM. Mandibuloacral dysplasia type B in an infant: a rare progeroid genodermatosis. JAMA Dermatol. 2015;151(5):561-2.
219. Shackleton S, Smallwood DT, Clayton P, Wilson LC, Agarwal AK, Garg A, et al. Compound heterozygous ZMPSTE24 mutations reduce prelamin A processing and result in a severe progeroid phenotype. J Med Genet. 2005;42(6):e36.
220. Ahmad Z, Phadke SR, Arch E, Glass J, Agarwal AK, Garg A. Homozygous null mutations in ZMPSTE24 in restrictive dermopathy: evidence of genetic heterogeneity. Clin Genet. 2012;81(2):158-64.
221. Maroofian R, Murdocca M, Rezaei-Delui H, Nekooei A, Mojarad M, Sangiuolo F, et al. A novel in-frame deletion in ZMPSTE24 is associated with autosomal recessive acrogeria (Gottron type) in an extended consanguineous family. Clin Dysmorphol. 2018;27(3):88-90.
222. Diociaiuti A, D'Amico P, Pisaneschi E, Giancristoforo S, Pappalardo MG, Di Guardo V, et al. Teledermatology diagnosis of the first Italian patient affected with restrictive dermopathy due to ZMPSTE24 homozygous mutation. J Eur Acad Dermatol Venereol. 2019;33(3):e139-e40.
223. Dutta AK, Danda S. Restrictive Dermopathy. Pediatr Neonatol. 2016;57(3):259.
224. Kariminejad A, Goodarzi P, Thanh Huong le T, Wehnert MS. Restrictive dermopathy. Molecular diagnosis of restrictive dermopathy in a stillborn fetus from a consanguineous Iranian family. Saudi Med J. 2009;30(1):150-3.
225. Li C. Homozygosity for the common mutation c.1085dupT in the ZMPSTE24 gene in a Mennonite baby with restrictive dermopathy and placenta abruption. Am J Med Genet A. 2010;152a(1):262-3.
226. Morais P, Magina S, Ribeiro Mdo C, Rodrigues M, Lopes JM, Thanh Hle T, et al. Restrictive dermopathy--a lethal congenital laminopathy. Case report and review of the literature. Eur J Pediatr. 2009;168(8):1007-12.
227. Yesil G, Hatipoglu L, Esteves-Vieira V, Levy N, De Sandre-Giovannoli A, Tüysüz B. Restrictive dermopathy in a Turkish newborn. Pediatr Dermatol. 2011;28(4):408-11.
228. Cassini TA, Robertson AK, Bican AG, Cogan JD, Hannig VL, Newman JH, et al. Phenotypic heterogeneity of ZMPSTE24 deficiency. Am J Med Genet A. 2018;176(5):1175-9.
229. Haye D, Dridi H, Levy J, Lambert V, Lambert M, Agha M, et al. Failure of ossification of the occipital bone in mandibuloacral dysplasia type B. Am J Med Genet A. 2016;170(10):2750-5.
230. Hitzert MM, van der Crabben SN, Baldewsingh G, van Amstel HKP, van den Wijngaard A, van Ravenswaaij-Arts CMA, et al. Mandibuloacral dysplasia type B (MADB): a cohort of eight patients from Suriname with a homozygous founder mutation in ZMPSTE24 (FACE1), clinical diagnostic criteria and management guidelines. Orphanet J Rare Dis. 2019;14(1):294.
231. Denecke J, Brune T, Feldhaus T, Robenek H, Kranz C, Auchus RJ, et al. A homozygous ZMPSTE24 null mutation in combination with a heterozygous mutation in the LMNA gene causes Hutchinson-Gilford progeria syndrome (HGPS): insights into the pathophysiology of HGPS. Hum Mutat. 2006;27(6):524-31.
232. Harhouri K, Navarro C, Baquerre C, Da Silva N, Bartoli C, Casey F, et al. Antisense-Based Progerin Downregulation in HGPS-Like Patients' Cells. Cells. 2016;5(3).
233. Thill M, Nguyen TD, Wehnert M, Fischer D, Hausser I, Braun S, et al. Restrictive dermopathy: a rare laminopathy. Arch Gynecol Obstet. 2008;278(3):201-8.
234. Gould PS, Dyer NP, Croft W, Ott S, Easton AJ. Cellular mRNAs access second ORFs using a novel amino acid sequence-dependent coupled translation termination-reinitiation mechanism. Rna. 2014;20(3):373-81.
235. Luttermann C, Meyers G. A bipartite sequence motif induces translation reinitiation in feline calicivirus RNA. J Biol Chem. 2007;282(10):7056-65.

236. Pöyry TA, Kaminski A, Connell EJ, Fraser CS, Jackson RJ. The mechanism of an exceptional case of reinitiation after translation of a long ORF reveals why such events do not generally occur in mammalian mRNA translation. Genes Dev. 2007;21(23):3149-62.
237. Bazykin GA, Kochetov AV. Alternative translation start sites are conserved in eukaryotic genomes. Nucleic Acids Res. 2011;39(2):567-77.
238. Benitez-Cantos MS, Yordanova MM, O'Connor PBF, Zhdanov AV, Kovalchuk SI, Papkovsky DB, et al. Translation initiation downstream from annotated start codons in human mRNAs coevolves with the Kozak context. Genome Res. 2020;30(7):974-84.
239. Riggs ER, Andersen EF, Cherry AM, Kantarci S, Kearney H, Patel A, et al. Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). Genet Med. 2020;22(2):245-57.
240. Abou Tayoun AN, Pesaran T, DiStefano MT, Oza A, Rehm HL, Biesecker LG, et al. Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. Hum Mutat. 2018;39(11):1517-24.
241. 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(5):405-24.
242. Lyu Y, Jia W, Wu Y, Zhao X, Xia Y, Guo X, et al. Cpmer: A new conserved eEF1A2-binding partner that regulates Eomes translation and cardiomyocyte differentiation. Stem Cell Reports. 2022;17(5):1154-69.
243. Levy-Sakin M, Pastor S, Mostovoy Y, Li L, Leung AKY, McCaffrey J, et al. Genome maps across 26 human populations reveal population-specific patterns of structural variation. Nat Commun. 2019;10(1):1025.
244. Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, et al. A global reference for human genetic variation. Nature. 2015;526(7571):68-74.
245. Guo M, Xiao ZD, Dai Z, Zhu L, Lei H, Diao LT, et al. The landscape of long noncoding RNA-involved and tumor-specific fusions across various cancers. Nucleic Acids Res. 2020;48(22):12618-31.
246. Sánchez-Marín D, Silva-Cázares MB, Porras-Reyes FI, García-Román R, Campos-Parra AD. Breaking paradigms: Long non-coding RNAs forming gene fusions with potential implications in cancer. Genes Dis. 2024;11(3):101136.
247. GTEx Portal [Internet]. [cited 25.04.2024]. Available from: www.gtexportal.org.