MASTERARBEIT | MASTER'S THESIS

Titel | Title

A Biallelic Repeat Expansion Mutation in RFC1 as a Possible Risk Factor for Parkinson's Disease

verfasst von | submitted by

Jennifer Huber BSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 877
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Genetik und Entwicklungsbiologie
Betreut von | Supervisor:	Dr. Alexander Zimprich

Table of Contents

List of Tables

List of Figures

Abbreviations

AAO	age at onset
AD	Alzheimer's Disease
ALS	amyotrophic lateral sclerosis
ASN	alpha synuclein
ASOs	antisense oligonucleotides
CANVAS	Cerebellar Ataxia, Neuropathy, Vestibular Areflexia Syndrome
CTR	control
DIG	digoxigenin
DM	Myotonic Dystrophy
dNTPs	deoxyribonucleotide triphosphates
L-Dopa	levodopa
EOPD	early onset Parkinson's Disease
EtOH	ethanol
FAM	carboxyfluorescein
FAM	fluorescein amidite
FDG-PET	fluorodexglycose-position emission tomography
fPCR	flanking PCR
FRDA	Friedreich`s Ataxia
FTD	Frontotemporal Dementia
FXS	Fragile X Syndrom
FXTAS	Fragile X-Associated Tremor/Ataxia Syndrome
HCl	hydrochloric acid
HD	Huntington`s Disease
Hi-Di	formamide
LB	lysogeny broth
LOPD	late onset Parkinson's Disease
LRRK2	leucine-rich repeat kinase 2
miRNAs	microRNAs
MRI	magnetic resonance imaging
MSA	Multiple System Atrophy

MSA-C	MSA-cerebellar
MSA-P	MSA-parkinsonian
NaAc	natrium acetate
NaCl	sodium chloride
NaOH	sodium hydroxide
OPMD	oculopharyngeal muscular dystrophy
PCNA	proliferating cell nuclear antigen
PD	Parkinson`s Disease
Pol	polymerase
polyQ	polyglutamine
qPCR	quantitative PCR
RAN	repeat associated non-AUG
REM	repeat expansion mutation
RFC1	replication factor complex unit 1
ROS	reactive oxygen species
RP PCR	repeat primed PCR
RT	room temperature
SCA	Spinocerebellar Ataxia
SCA3	Machado-Joseph disease
SDS	sodium dodecyl sulfate
shRNAs	small hairpin RNAs
siRNAs	small interfering RNAs
SNPs	single nucleotide polymorphisms
SNVs	single nucleotide variants
SPECT	Single-Photon Emission CT
SSC	saline-sodium citrate
STR	short tandem repeats
TBE	tris-borate-EDTA
Tris-HCl	tris(hydroxymethyl)aminomethan hydrochloride
UTR	untranslated region
VPS35	vacuolar protein sorting 35
WT	wild type

Zusammenfassung

Die Parkinson-Krankheit (PD) gilt weltweit als zweithäufigste neurodegenerative Erkrankung. Da über sechs Millionen Menschen an dieser Krankheit leiden, ist das Verständnis der molekularen Pathologie von großer Bedeutung. Nichtsdestotrotz können nur 10 bis 15 % der Fälle auf genetische Faktoren zurückgeführt werden. Aktuelle Studien liefern Hinweise darauf, dass Expansionsmutationen (Repeat Expansion Mutations, REMs), die sich im Replication Factor Complex 1 (RFC1) befinden, an der Pathogenese der Parkinson-Krankheit beteiligt sind. Dieser Pathomechanismus wurde ursprünglich für die neurologische Erkrankung „Cerebellar Ataxia Neuropathie Vestibular Areflexia Syndrome“ (CANVAS) beschrieben. Ziel dieser Studie ist es, die Rolle von REMs in RFC1 bei der Pathogenese der Parkinson-Krankheit zu untersuchen.

Zu diesem Zweck haben wir Genotypisierungstechniken an einer großen Kohorte österreichischer Patienten mit klinisch diagnostiziertem Parkinsonismus angewendet. Mithilfe von „Repeat Primed PCR“, Southern-Blot-Analysen und quantitativer PCR haben wir alle Personen auf verschiedene RFC1-Expansionsvarianten untersucht. Wir haben uns dabei hauptsächlich auf das biallelisch expandierte RFC1-Motiv $\mathrm{AAGGG_n}$ konzentriert, welches sich vom Wildtyp-Motiv unterscheidet, da es als ursächlich für CANVAS gilt.
Wir konnten zeigen, dass eine biallelische $\mathrm{AAGGG_n}$- Expansionen in RFC1 in hohem Maße mit der Entwicklung von Parkinson assoziiert ist. Es konnte hierbei festgestellt werden, dass 1,34 % unserer Patientenkohorte (12/1204) biallelische Expansionen aufwiesen, verglichen mit 0,7 % der gesunden Kontrollpersonen (1/1204). Dies unterstreicht nachdrücklich die Relevanz biallelischer $AAGGG_n$ RFC1-Expansionen für die Ätiologie der Parkinson-Krankheit.
Darüber hinaus beobachteten wir eine Korrelation des biallelisch erweiterten Wildtyp-RFC1-Motiv $AAAAG_n$ mit der Entwicklung von Parkinson, von der bislang berichtet wurde, dass sie nicht zu pathogenen Phänotypen führt. Dies weist erneut auf die Bedeutung biallelisch expandierter RFC1-Varianten bei der Parkinson-Krankheit und verwandten Krankheiten hin.

Southern-Blot-Analysen von biallelisch $\mathrm{AAGGG_n}$-expandierten Proben ergaben eine geringere Expansionsgröße als bei der CANVAS-Krankheit, mit 380 bis 660 Wiederholungseinheiten im größeren Allel im Vergleich zu CANVAS, wo die Repeat-Anzahl typischerweise zwischen 400 und 2,000 liegt. Darüber hinaus beobachteten wir in allen Proben ein weniger expandiertes Allel, das beinahe eine Wildtypgröße von 11 Wiederholungen aufwies.
Wir postulieren, dass Personen mit einer RFC1-Expansion, die nahe an der Größe des Wildtyps liegt, und einer anderen Expansion, die größer als die pathologische Schwelle, aber kleiner als die typische Wiederholungszahl ist, die CANVAS verursacht, mit größerer Wahrscheinlichkeit PD anstelle von CANVAS entwickeln.

In vorangegangenen Studien wurde festgestellt, dass eine pathologische RFC1-Expansion zusammen mit vier spezifischen Einzelnukleotidpolymorphismen (SNPs) („rs2066790, rs11096992, rs17584703 und rs6844176“) vererbt wird.
In unseren Analysen haben wir den SNP rs1109699 als Marker für das Vorhandensein der pathogenen Expansion in RFC1 ausgewertet und eine Übereinstimmungsrate von 100 % beobachtet. Wir könnten daher den vorgeschlagenen Risikohaplotyp bestätigen. Alle PD-

Patienten, die das biallelisch expandierte RFC1-Motiv $\mathrm{AAGGG_n}$ tragen, wurden positiv auf diesen SNP getestet, was auf eine gemeinsame Vererbung hinweist.

Zusammenfassend konnten wir zeigen, dass biallelische Expansionen im RFC1-Gen stark mit der Entstehung der Parkinson-Krankheit assoziiert sind. Die Größe der Erweiterungen ist kleiner als in CANVAS, übersteigt jedoch deutlich die Größe gesunder Wiederholungszahlen.
Hiermit bietet diese Arbeit nicht nur Einblicke in die Heterogenität von RFC1, sondern unterscheidet auch zwischen der möglichen Rolle von RFC1 bei der Entstehung von PD oder CANVAS. Folglich sind die im Rahmen dieser Studie gewonnenen Erkenntnisse von großer Bedeutung für zukünftige genetische Diagnostik im klinischen Kontext.

Abstract

Parkinson's Disease (PD) is considered to be the second most common neurodegenerative disease worldwide. With over six million people suffering from the disease, the understanding of its molecular pathology is of great importance. Notwithstanding only 10 to 15% of the cases can be attributed to genetical factors. Recent studies provide evidence suggesting the involvement of repeat expansion mutations (REMs) residing in the replication factor complex unit 1 (*RFC1*) in the pathogenesis of PD. This mechanism has initially been reported for the neurological disease Cerebellar Ataxia Neuropathy Vestibular Areflexia Syndrome (CANVAS). This study aims to assess the role of REMs in *RFC1* in the pathogenesis of PD.

For this purpose, we applied genotyping techniques on a large cohort of Austrian patients with clinically diagnosed parkinsonism. Utilizing repeat primed PCR, Southern blot analyses and quantitative PCR, we screened all individuals for different RFC1 expansion variants. We hereby mainly focused on the biallelically expanded RFC1 motif $\mathrm{AAGGG}_{\mathrm{n}}$, which differs from the wildtype motif, as it has been indicated to be causative for CANVAS.
We could show that a biallelic $\mathrm{AAGGG}_{\mathrm{n}}$ expansion in RFC1 is highly associated with the development of PD. 1,34% of our patient cohort (12/1204) were found to carry biallelic expansions compared to 0,7% of healthy controls (1/1204). This strongly emphasizes the relevance of biallelic $\mathrm{AAGGG}_{\mathrm{n}}$ RFC1 expansions in the etiology of PD.
Additionally, we observed a correlation of the biallelically expanded wildtype *RFC1* motif $\mathrm{AAAAG}_{\mathrm{n}}$ with the development of PD, which was so far reported to not result in pathogenic phenotypes. This again highlights the significance of biallelically expanded *RFC1* variants in PD and related diseases.

Southern blot analyses of biallelically $\mathrm{AAGGG}_{\mathrm{n}}$ expanded samples revealed a smaller expansion size than in CANVAS disease, with 380 to 660 repeat units in the larger allele compared to CANVAS with repeat units typically spanning from 400 to 2,000. Additionally, we observed one less expanded allele in all samples showing almost a wildtype size of 11 repeats.
We postulate that individuals with one RFC1 expansion close to the wild-type size and another expansion larger than the pathological threshold but smaller than the typical repeat number causing CANVAS, are more likely to develop PD instead of CANVAS.

Pathological *RFC1* expansion were previously found to be inherited together with four specific single nucleotide polymorphisms ("rs2066790, rs11096992, rs17584703, and rs6844176").
In our analyses we evaluated the SNP rs1109699 as a marker for the presence of the pathogenic repeat expansion in *RFC1* and observed a 100% concordance rate. We could therefore reinforce the suggested founder haplotype. All PD patients harboring the biallelically expanded *RFC1* motif $\mathrm{AAGGG}_{\mathrm{n}}$ tested positive for this distinct SNP, indicating a shared founder haplotype.

In conclusion, we could be able to show, biallelic repeat expansions in the *RFC1* gene being strongly associated with the development of Parkinson´s disease. The size of the expansions is smaller than in CANVAS but significantly exceeds the size of healthy repeat numbers. This thesis not only provides insight into the heterogeneity of *RFC1* but also distinguishes between the

potential role of RFC1 causing either PD or CANVAS. Consequently, the findings generated throughout this study, yield great importance regarding future genetical diagnostic in clinical context.

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my senior supervisor, Prof. Dr. Alexander Zimprich, for the opportunity and confidence he extended to me for this project. Thank you for your exceptional mentorship, extensive expertise, and unwavering patience. Your enthusiasm has been truly inspiring.

I also wish to thank my junior supervisor Theresa König, PhD for her constant support throughout this project. Your trust, encouragement, insightful guidance, and constructive feedback have been invaluable in navigating the challenges always aiming for the best possible outcome. I am especially grateful for the harmonious working environment you fostered, which made the work feel like time spent with a friend.

Additionally, I want to acknowledge the entire team at the Department of Neurogenetics at the Medical University of Vienna. Your support has been crucial to my progress.
A special thanks goes to Christoph Hotzy for his helpful discussions and the technical support which had made the research process a lot easier. Another thank you goes to Sina Zaic, who I looked forward to every day when I came to the lab, you just made work much more fun!

Finally, I extend my heartfelt thanks to my family and friends for their continuous support and understanding during the intensive work periods. Thank you to my sisters, Angelina and Cathrin Huber and my mom, Sabine Huber who have always supported me no matter what. I am also deeply grateful for my best friend, Michaela Schweiger who has made the writing process much more fun and simply makes me a stronger person. A special thanks goes to my boyfriend Johannes Wiedemann, who has always been there for me and helped me believe in myself. This had made me feel like I (or we) could achieve anything. I love you.

Thank you!

1. Introduction

1.1. Repeat Expansion Mutations

In the past frequently underestimated, repeat expansion mutations have gained increasing attention as the driving etiology of more than 50 neurodegenerative and neuromuscular diseases (Figueiredo et al., 2023). The subsequent paragraphs will guide through the theoretical background of REMs as well as associated diseases, with a special focus on Parkinson`s Disease (PD).

1.1.1. From Short Tandem Repeats to Repeat Expansion Mutations

Short tandem repeats (STRs) also known as microsatellites are highly repetitive units within the coding and non-coding strand of the DNA. These extremely polymorphic DNA elements, which make up to 7% of the human genome, consist of variably repeated motifs of two to 12 base pair units each. Primary located in non-coding regions such as introns, gene promotors or 5` as well as 3` untranslated regions (UTRs), they hold the potential to be especially harmful when located within protein-coding exonic regions as they can induce frameshift mutations (Chintalaphani et al., 2021; Depienne & Mandel, 2021; Malik et al., 2021; Rafehi et al., 2019). Repeat expansions in the human genome can span from a few to up to thousands of repeat units but only manifest in pathogenic phenotypes when the repeat number reaches a gene-specific threshold. These genetic abnormalities are referred to as repeat expansion mutations (REMs) (Rafehi et al., 2019). For REMs located in coding areas, such as those reported for Huntington`s Disease (HD), the disease inducing threshold is rather small (< 40 repeats). However, in conditions like Friedreich`s ataxia or Fragile X Syndrome (FXS), where the mutation is situated within a non-coding area of the genome, a substantially greater number of repeats in the range of thousands is required to induce a pathogenic phenotype (Eimer et al., 2018). REMs within the human genome result in numerous neurological diseases with a suggested prevalence of 1:20.000 (Depienne & Mandel, 2021; Rafehi et al., 2019).

1.1.2. Molecular Pathogenesis of Repeat Expansions and Associated Neurodegenerative Disorders

The discovery of REMs as a possible underlying cause of neurodegenerative diseases over the past decades has driven the development of new diagnostic and therapeutic avenues addressing these complex diseases. The first REM disease HD and its pathogenesis were described in 1993 followed by diseases like Myotonic Dystrophy and FXS. REMs have gained importance in elucidating the pathophysiology of other neurodegenerative diseases like Cerebellar Ataxia, Neuropathy, Vestibular Areflexia Syndrome (CANVAS) which has recently been reported to be in association to this distinct etiology. The molecular mechanisms underlying REMs depend on the repeat size, specific repeat sequence, and their genomic location. The effects on the proteome include the loss of protein function, which is believed to influence pathologies such as Fragile X Syndrome. Additionally, toxic gain-of-function mutations in proteins are associated with phenotypes like HD, while toxic gain-of-function mutations in RNA have been documented in Myotonic Dystrophies Type 1 and Type 2 (DM1 and DM2, respectively). Molecular Pathways are illustrated in Figure 1.

1.1.2.1. Epigenetic Silencing

Epigenetic silencing is involved in several repeat expansion diseases such as FXS, DM1, Friedreich`s Ataxia (FRDA) and HD. The underlying mechanism that causes silencing of the gene is hypermethylation of the CG region in the promotor of the respective gene.
These hypermethylated CpG islands subsequently lead to histone deacetylation as well as heterochromatin formation. This causes repression of transcription and therefore the loss of function of the corresponding gene as observed in the *FMRI* gene in the context of **FXS**.
Epigenetic silencing can also be induced by the inhibition of chromatin binding factors. In the instance of **DM1** for example**,** CTCF binding is hindered due to DNA methylation on the 3`UTR of the *DMPK* gene. This affects chromatin organization and subsequently binding of the translational machinery.
Another epigenetically driven silencing mechanism is characterized by the formation of three stranded RNA-DNA hybrids, where an RNA molecule anneals within a double helix DNA structure during transcription. This mechanism is referred to as R-loops. In **FRDA**, these persistent RNA-DNA hybrids form on REMs in the *FXN* gene. The co-localization with repressive chromatin marks leads to transcriptional silencing (Chintalaphani et al., 2021; Depienne & Mandel, 2021; Groh et al., 2014; He & Todd, 2011; Jain & Vale, 2017; Malik et al., 2021).

1.1.2.2. RNA and DNA-Sequestration

The processing, nuclear or cytoplasmic transport and localization of RNA molecules are strongly depending on a distinct class of proteins, referred to as RNA binding proteins (RBPs). Through their interaction with the RNA they play an essential role in impairing and regulating the RNA function (Y. Chen & Varani, 2013). Clusters of expanded repeat sequences within RNA molecules can induce a phenomenon known as “sequestration”, where the RBPs (such as RNA splicing factors) are strapped within the RNA, inhibiting their initial function. Of special importance in non-coding REMs, the resulting lack of available RBPs leads to the impairment of pre-mRNA processing, binding of transcription factors and interaction with the transcriptional machinery.
In diseases like Fragile X-Associated Tremor/Ataxia Syndrome (**FXTAS**) and **DM**, these RNA-toxicity inducing aggregates consequently inhibit the mobility and functionality of the RNA splicing factors. This results in missplicing and subsequently a toxic gain of function of the corresponding protein (Chintalaphani et al., 2021; Depienne & Mandel, 2021; Figueiredo et al., 2023; Jain & Vale, 2017; Malik et al., 2021; Salcedo-Arellano et al., 2020). Furthermore, previous studies regarding **FRDA** have also postulated a mechanism that involves binding of RNA polymerases by “sticky DNA” as a consequence of DNA sequestration and with DNA being the aggregate forming molecule (Cleary & Ranum, 2014; Sakamoto et al., 2001; Vetcher et al., 2002).

1.1.2.3. RAN Translation

Non canonical protein synthesis represents an additional REM induced pathological molecular pathway. In context protein synthesis can occur in absence of the translational start codon AUG. This repeat associated non-AUG translation (RAN translation) can occur on both the sense as well as the antisense strand of the DNA resulting in different possible polypeptides and consequently in a toxic gain of function.

RAN translation is reported for several REM associated disorders, including **DM1, DM2**, *C9ORF72* associated frontotemporal dementia and amyotrophic lateral sclerosis (**FTD/ALS**), **HD**, spinocerebellar ataxia type 31 (**SCA31**) and according to recent studies also **FXTAS** (Chintalaphani et al., 2021; Depienne & Mandel, 2021; Jain & Vale, 2017; Malik et al., 2021).

1.1.2.4. Protein Misfolding and Aggregation

REMs occurring in coding regions or in genomic areas expressed via RAN translation can result in protein misfolding and aggregation. Particularly polyglutamine (polyQ) stretches, which result from an expansion of the CAG (coding for glutamine) repeat in the respective gene, have been reported to induce several neurological diseases, referred to as polyQ diseases. The resulting protein, exhibiting an abnormal expansion of the glutamine part, subsequently shows a tendency to form amyloid like structures, known as beta-aggregates. These are a major hallmark for other neurodegenerative diseases like Alzheimer's disease (**AD**). As a consequence, the normal protein function is disrupted, and the gain of toxic protein function is assumed. These insoluble fibrillar aggregates constitute the pathological hallmark of various neurological disorders including **HD, spinocerebellar ataxia** and **spinal** or **bulbar muscular atrophy** (Chintalaphani et al., 2021; Depienne & Mandel, 2021; Gatchel & Zoghbi, 2005; Jain & Vale, 2017; Malik et al., 2021; Takeuchi & Nagai, 2017).

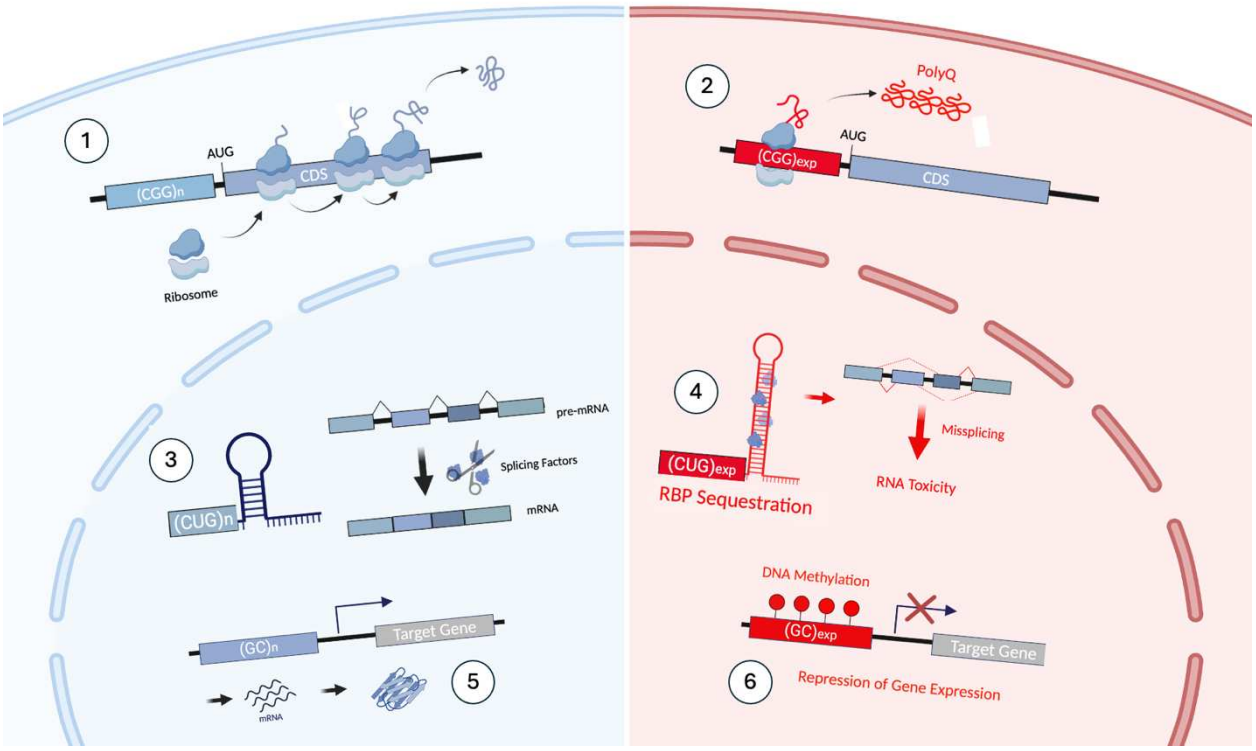

Figure 1. Molecular Pathology Pathways in Repeat Expansion Diseases.
Comparison of a healthy cell (left, blue) versus a cell with repeat expansion mutated DNA (right, red). ***(1)*** *Healthy translation of mRNA into a normal-length protein.* ***(2)*** *RAN translation independent of the AUG start codon (as well as other processes) lead to abnormal proteins where a specific amino acid is repeated (stretched) several times, as it is the case for PolyQ stretches.* ***(3)*** *Splicing factors induce the processing of pre-mRNA inside the nucleus.* ***(4)*** *Repeat expansions lead to the formation of abnormal secondary structures and consequently to the sequestration of RNA binding proteins such as splicing factors. The result is a defect in pre-repeat expansions especially in promoter regions may lead to an increment of DNA methylation and thus to a repression of gene mRNA processing such as missplicing leading to toxic RNAs. (5) Healthy transcription of a WT gene inside the nucleus. (6) GC-rich repeat expansions especially in promoter regions may lead to an increment of DNA methylation and thus to a repression of gene expression. Figure created with BioRender.com*

1.1.3. Shared Clinical Features of Repeat Expansion Mutation Induced Diseases

Abnormal amplification of repetitive DNA sequences, disruption of normal cellular processes and the accumulation of toxic RNA or protein aggregates manifest in various clinical phenotypes. While not all REM induced diseases exhibit identical clinical features, certain symptoms display significant prevalence. These primarily motor features include cerebellar ataxia (impaired coordination and control of voluntary muscle action), chorea (involuntary and rapid, and jerky movements of various body parts), tremor, muscular dystrophies and myoclonic seizures (spasticity). Non-movement related symptoms encompass cognitive impairment as well as psychiatric disturbances (Chintalaphani et al., 2021; McColgan & Tabrizi, 2018; Meyer, 2021; Salcedo-Arellano et al., 2020).

1.1.4. The Role of Repeat Length

Emerging evidence suggest a correlation between the expansion size and the age of onset as well as the severity of REM diseases, while larger expansions are generally associated with an earlier manifestation of clinical symptoms and a more severe phenotype (Fournier et al., 2019;

Chintalaphani et al., 2021). The tendency of the expansion size to grow with successive generations leading to more severe phenotypes in the offspring, is known as anticipation (see capter “clinical anticipation”).
Interestingly, an influence of the expansion size on the clinical phenotype has been reported, with smaller expansions rather leading to FXTAS and larger ones having an increased likelihood to result in FXS, especially for *FMR1* diseases (Bourgeois et al., 2011) .
For HD even a decreased survival time in correlation with repeat length has been suggested, with larger expansions leading to a younger age at death of the patients (Langbehn, 2022).
Recent research indicates that in cases of biallelic REMs in the RFC1 gene, the alleles differ in their expansion length. The allele with the smaller expansion appears to be a greater indicator of variability in the respective disease (Currò et al., 2024).
The phenomenon of repeat size being a major predictor of onset, and severity of the disease, has been observed not only in HD, C9ALS, C9FTD and RFC1 related diseases but also in other neurological repeat diseases such as DM and FRDA (Chintalaphani et al., 2021; Depienne & Mandel, 2021; Fournier et al., 2019; Langbehn, 2022; Lanni & Pearson, 2019; Malik et al., 2021; Paulson, 2018)

1.1.4.1. Intermediate Length Expansions

As previously described, there is a very specific threshold regarding the expansion size of the respective STR harboring genes, to result in pathogenic phenotypes. Expansions which are longer than the wild type STR but just marginally reach the pathogenic threshold are referred to as intermediate repeats. Clinical significance of these medium length expansions has been discussed in this context. In several neurodegenerative repeat expansion associated diseases, such as C9orf72, not only carriers of full length but also of intermediate length expansions are suspected to have an increased risk of developing C9orf72 related diseases (Iacoangeli et al., 2019). Intermediate length expansions have been indicated to have an increased risk for developing other diseases than the one associated with the specific repeat expansion like in HD, ALS, FTD and AD (Miyatake et al., 2022; Iacoangeli et al., 2019). In one of the few studies regarding the different sequence and size variants of RFC1, an intermediate expansion length of 28 to 111 repeats has been suggested (Beecroft et al., 2020; Tyagi et al., 2023).

1.1.4.2. Clinical Anticipation

The tendency of increasing severity and decreasing age of onset with successive generations is referred to as clinical anticipation. This phenomenon has been observed by clinicians long before the genetical background of repeat expansion diseases was discovered. During meiosis longer repeats tend to undergo further expansion which results in a greater magnitude of expansion size in the offspring. The extent of the anticipation can depend on the parental haplotype the REM originates from. This effect shows a high variability across different diseases. Parents of origin exhibit variability, depending on the specific gene or disease. In the majority of REM diseases such as those induced by CAG expansions (HD, SCA2 or SCA7), paternally inherited mutations have a stronger tendency to elongate. DM1 is one of the few REM diseases where a maternal effect regarding anticipation can be observed, leading to maternally inherited alleles with greater expansion compared to paternal inherited ones.

Repeat expansion diseases associated with a low pathogenic threshold of expansion size, are suspected to not show clinical anticipation. This has for example been reported for oculopharyngeal muscular dystrophy (OPMD), where the number of repeats ranges from 11 to 18 (Chintalaphani et al., 2021; Depienne & Mandel, 2021; Paulson, 2018, 2018; Trollet et al., 1993).

1.1.5. Therapeutic Approaches

With the rapidly evolving understanding in the field of REMs there has been an increase in research efforts aimed at developing therapeutic strategies that target the underlying genetic causes and effectively handle the associated clinical symptoms.

1.1.5.1. RNA Interference

A very promising tool for targeting REM induced diseases lies within the domain of RNA-based therapies. Antisense Oligonucleotides (ASOs) or small interfering RNAs (siRNAs) are utilized to either target the pathogenic expanded allele or single nucleotide polymorphisms (SNPs) which are in linkage disequilibrium with the respective REM. RNase H induced cleavage facilitates the downregulation of the mutant allele. While siRNAs arise from small hairpin RNAs (shRNAs) and microRNAs (miRNAs) are introduced via viral vectors, transcribed and incorporated into the RNA induced silencing complex, ASOs can be internalized by the cell without the concern of disrupting the miRNA pathway (Bauer & Nukina, 2009; Chintalaphani et al., 2021; Depienne & Mandel, 2021).

1.1.5.2. PolyQ Stretch Phosphorylation and Sumoylation

In addition to directing efforts towards RNAs also other cellular events in the process of the disease's pathogenesis might become a target for specifically designed molecules to neutralize the pathological expansion. Recent studies have shown a successful cleavage or phosphorylation/sumoylation induced silencing of expansion induced PolyQ stretches suggesting it to become a promising tool addressing HD or several types of SCAs (Bauer & Nukina, 2009; Depienne & Mandel, 2021).

1.1.5.3. RAN Protein Affine Monoclonal Antibodies

Not only PolyQ but also RAN proteins which are produced from repeats in multiple open reading frames without an initiation codon (AUG or ATG independent) can be a possible therapeutic strategy. These toxic peptides can be targeted by high affinity monoclonal antibodies preventing their aggregation and promoting their degradation. C9-ALS/FTD BAC mouse models have provided evidence of the efficacy of this approach offering a strategy for addressing this complex disorder (Cleary & Ranum, 2014; Depienne & Mandel, 2021; Nguyen et al., 2020).

1.1.5.4. Genome Editing

Alternative approaches making use of gene therapy are promising methods to address REM induced disorders. Genome editing technologies like CRISP/Cas9 systems offer the possibility to directly edit the harmful allele by either excising the expansion or by delivering another functional copy of the affected gene. However, in order to be applied in clinical practice

remaining limitations still need to be addressed including the effective delivery of proteins and guide-RNAs to target tissues, as well as potential off-target effects (Bauer & Nukina, 2009; Depienne & Mandel, 2021; Jacinto et al., 2020; Nguyen et al., 2020).

1.2. The RFC1 Gene in Health and Disease

The *RFC1* gene, which encodes the first and largest subunit of the replication factor complex (RFC) plays a critical role in cellular processes such as DNA replication and repair in eukaryotic cells (Tomida et al., 2008). The five-subunit protein complex RFC is required for maintaining the genomic stability and ensuring proper cellular function. As an accessory protein of the DNA polymerase and under the requirement of ATP it binds to the primer template junction and loads PCNA (proliferating cell nuclear antigen) onto the DNA. The catalyzation of the ring shaped sliding champ protein PCNA into an open conformation allows an encircling of the DNA in S phase of the cell circle (Kubota et al., 2013; Kupiec, 2016; Tomida et al., 2008). PCNA acts as a scaffold for the recruitment of proteins promoting DNA replication and repair, above all DNA polymerase delta and epsilon (Cortese et al., 2019a).

RFC1 harbors an eleven repeats long STR localized within the 3`end at the Poly(A) tail of an Alu element on chromosome 4 (chr4:39350045-39350103). With a prevalence of 75% in the general population, "AAAAG" is identified to as the wildtype motif of this sequence. However, other motif variants have also been reported such as "AAAGG" occurring at a frequency of 8% (Cortese et al., 2019a).

In 2019 two independent studies reported the presence of biallelic REMs in the RFC1 STR locus in individuals suffering from the rare neurological disease CANVAS (see chapter "Cerebellar Ataxia Neuropathy Vestibular Areflexia Syndrome"). A specific kind of sequence "AAGGG" was found to be expanded up to several thousand times, highlighting the polymorphic nature of this locus in RFC1 and its possible effect on certain pathologies (Cortese et al., 2019a; Rafehi et al., 2019).

1.2.1. Polymorphism of RFC1

The RFC1 repeat has been reported to exhibit a highly polymorphic nature, resulting in several allelic variants. These in some instance pathological variations, may not only show an eleven repeats exceeding expansion but also a mutation of the wild type "AAAAG" sequence. Confirmations predominantly follow the pattern $\mathrm{A}_n\mathrm{G}_\mathrm{m}$. Effects of specific RFC1 variants have controversially been discussed since different publications have indicated different motifs or expansion sizes to be pathogenic. An overview of common allelic variants, their average repeat size and most commonly associated pathogenicity, according to Cortese et al., Davies et al., and Tyagi et al. is shown in Figure 2 (Cortese et al., 2019a; Davies et al., 2022a; Tyagi et al., 2023)

1.2.2. Inheritance Pattern of Expanded RFC1

The majority of REM diseases follow a dominant pattern of inheritance. This leads to patients developing pathological phenotypes even when carrying only one mutated allele. Notably, all reported CANVAS cases were observed to only show biallelic expansions, which leads to the conclusion that RFC1 repeat expansion diseases follow a recessive genetic transmission. Increasing attention has been directed towards this distinct pattern of inheritance which

constitutes a special feature compared to many other REM diseases (Cortese & Reilly, o. J.; Paulson, 2018).

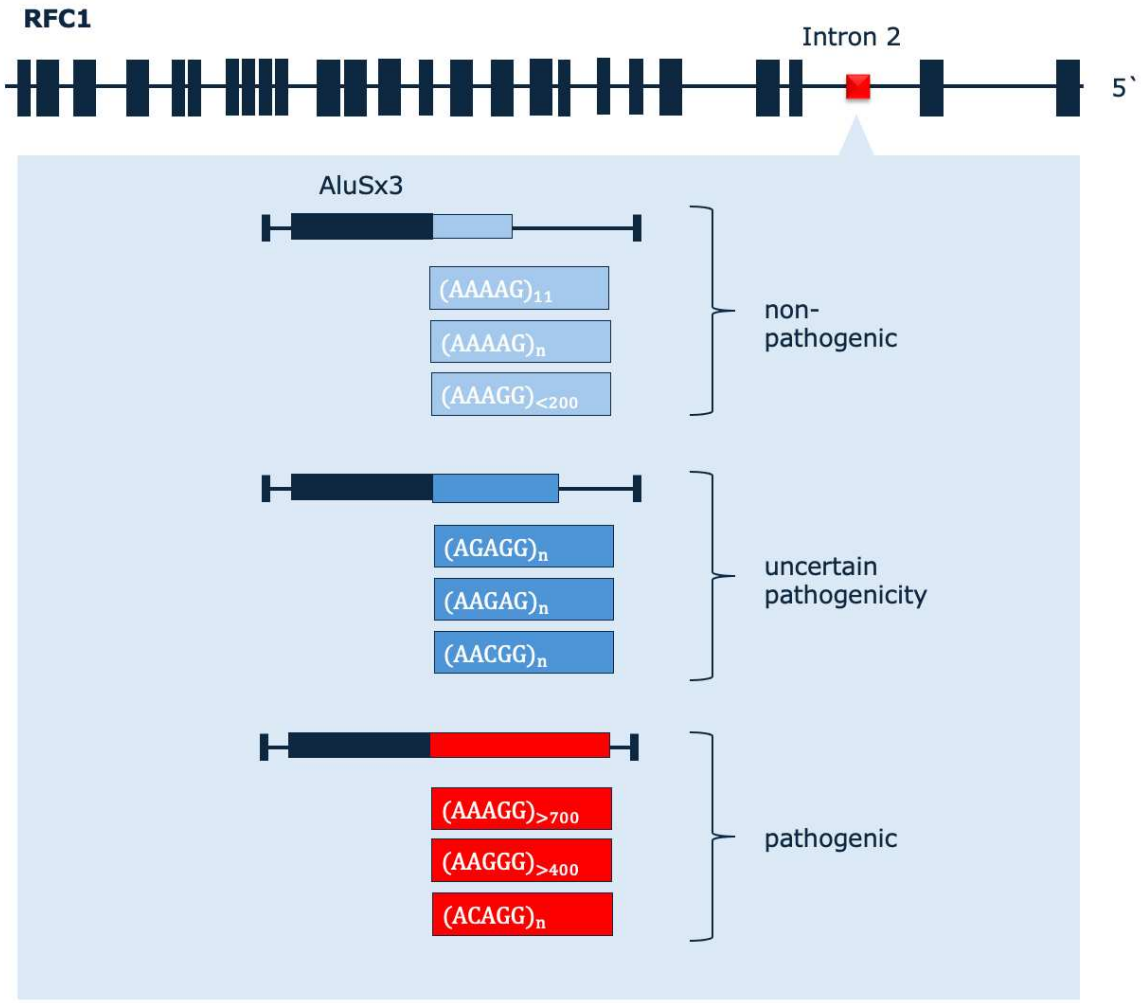

Figure 2. Molecular Pathology Pathways in Repeat Expansion Diseases.
Comparison of a healthy cell (left, blue) versus a cell with repeat expansion mutated DNA (right, red). ***(1)*** *Healthy translation of mRNA into a normal-length protein.* ***(2)*** *RAN translation independent of the AUG start codon (as well as other processes) lead to abnormal proteins where a specific amino acid is repeated (stretched) several times, as it is the case for PolyQ stretches.* ***(3)*** *Splicing factors induce the processing of pre-mRNA inside the nucleus.* ***(4)*** *Repeat expansions lead to the formation of abnormal secondary structures and consequently to the sequestration of RNA binding proteins such as splicing factors. The result is a defect in pre- repeat expansions especially in promoter regions may lead to an increment of DNA methylation and thus to a repression of gene mRNA processing such as missplicing leading to toxic RNAs. (5) Healthy transcription of a WT gene inside the nucleus. (6) GC-rich repeat expansions especially in promoter regions may lead to an increment of DNA methylation and thus to a repression of gene expression. Figure created with BioRender.com*

1.2.3. Single Nucleotide Polymorphisms Associated with RFC1 Repeat Expansions

Genetically, the inheritance of a stretch of DNA from a shared ancestor is referred to as the founder haplotype. This phenomenon has recently been suggested for the REM of RFC1 and four single nucleotide polymorphisms (SNPs) which are located on chromosome 4 as well. This 1.7 Mb large region, which spans all RFC1 exons as well as the terminal exon of the neighboring gene WD repeat domain 19 (WDR19). It exhibits the SNPs “rs2066790, rs11096992, rs17584703, and

rs6844176" and was suggested to correlate with the expanded variant of RFC1 by Cortese et al. (Cortese et al., 2019). This finding provides tools for identifying haplotypes carrying RFC1 expansions and consequently yields possible diagnostic avenues important for clinical context.

1.2.4. Clinical Spectrum of RFC1 Repeat Expansion Diseases

Initially associated with the rare neurodegenerative disease **CANVAS**, an RFC1 REM manifests in a broad range of phenotypes. Specifically, symptoms such as cerebellar ataxia, vestibular areflexia syndrome, chronic cough, autonomic dysfunction, somatosensory impairment and peripheral sensory motor neuropathy have been indicated to correlate with this distinct genotype. Biallelic $(\mathrm{AAAGG})_{exp}$ carriers have also been reported to exhibit an increased risk of developing other neurodegenerative diseases such as **Multiple System Atrophy** (MSA), **late onset ataxia**, **neuropathy**, various forms of **cancer** and according to recent studies also **PD** (Barghigiani et al., 2022; Cortese et al., 2019a; Kytövuori et al., 2022, 2022; Rafehi et al., 2019; Record et al., 2022; Silva Schmitt et al., 2020; Traschütz et al., 2021). Reported $(\mathrm{AAGGG})_{exp}$ allelic variant carrier frequencies in different European, Asian and Northern American control populations range from a prevalence of approximately 0.7% to 6.5% depending on the ethnicity. This highlights the relevance of understanding the genetic background and potential impacts of these mutations on health (Cortese et al., 2019a; Davies et al., 2022a).

1.2.4.1. Cerebellar Ataxia Neuropathy Vestibular Areflexia Syndrome (CANVAS)

The slowly progressive, adult-onset neurological disorder CANVAS results from a degeneration of the cerebellum as well its associated vestibular and sensory afferent neurons (Cortese et al., 2019). Typical symptoms consist of cerebellar ataxia (imbalance), sensory neuropathy (neuronopathy), bilateral vestibulopathy, chronic cough and autonomic dysfunction in a variable combination (Barghigiani et al., 2022; Cortese et al., 2019a; Cortese & Reilly, o. J.; Rafehi et al., 2019). Although most CANVAS cases occur sporadically, an autosomal recessive inheritance mode has been suggested for familial instances (Cortese et al., 2019a; Rafehi et al., 2019). Linkage analyses as well as whole genome sequencing have identified a biallelic repeat expansion of the sequence "AAGGG" in intron 2 of the RFC1 gene as the underlying etiology of this previously described disease (see "Clinical Spectrum of RFC1 Repeat Expansion Diseases"). Furthermore, a similar expansion size of RFC1 on both alleles has been suggested so far (Cortese et al., 2019; Dominik et al., 2021). While it has been taken into consideration that CANVAS might arise from variable pentanucleotide motif expansions following the pattern $\mathrm{A}_n\mathrm{G}_\mathrm{m}$, bioinformatic tools have been supporting the idea that the expansion mainly comprises $(\mathrm{AAGGG})_{exp}$ as pictured in Figure 3 (Rafehi et al., 2019). Although the pathogenesis of this rare neurological disease is not well understood yet, the recessive inheritance pattern supports the hypothesis of a loss of function mechanism on a molecular level. Additionally, Cortese et al. reported no alteration of the expression level of *RFC1* in peripheral or brain tissue. Diagnostic modalities span from electrodiagnostic findings, imaging tools such as magnetic resonance imaging (of brain and spine) and nerve ultrasound, to vestibular and autonomic testing. In patients with CANVAS electrodiagnostic findings may expose/display the absence of sensory action potentials and somatosensory potentials and an abnormal blink reflex as well as preserved Hoffmann reflex. MRI typically exhibits atrophy of cerebellum and spinal cord, and a sonographically a reduced

diameter of limb nerves can be detected. Vestibular testing may exhibit the presence of bilaterally impaired vestibular function and a vestibulo-ocular reflex gain (Cortese et al., 2019; Cortese & Reilly, o. J.; Dupré et al., 2021). For more certain diagnosis, genetic tools such as repeat primed PCR (RP-PCR) and Southern blot analyses, may be supportive as these techniques cannot only confirm the presence of the pathological extension but also the repeat number (Cortese & Reilly, o. J.).

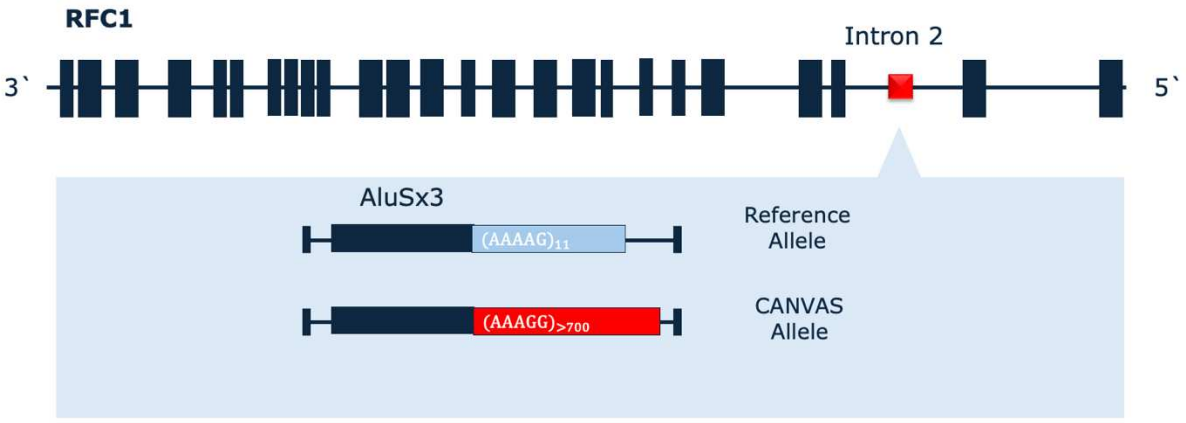

Figure 3. CANVAS Causing RFC1 Repeat Expansion.
The figure shows the location as well as the motif of the CANVAS causing repeat expansion mutation in RFC1. The WT allele harbors the eleven times repeated "AAAAG" motif (light blue), whereas it CANVAS the sequence is mutated to the distinct "AAAGG" motif, which shows the pathogenic phenotype above a threshold of approximately 700 repeats. The repeat expansion resides at the 5` terminus of an Alu-Element in intron 2 of the RFC1 gene (Rafehi et al., 2019). Figure created with BioRender.com

1.2.4.2. Multiple System Atrophy (MSA)

MSA is a rapidly progressive, rare neurodegenerative disorder with 1,9-4,9/1.000.000 affected people worldwide (Goh et al., 2023). The adult onset disease is characterized by striatonigral or olivopontocerebellar atrophy and cytoplasmic inclusions consisting of misfolded alpha synuclein which are referred to as Papp-Lantos bodies (Fanciulli et al., 2019; Goh et al., 2023).
The core phenotype associated with MSA manifests as a poly-systemic disorder encompassing cerebellar impairment, parkinsonism, autonomic dysfunction, ataxia, REM sleep behavior disorder and a spectrum of additional clinical manifestations. Notably, only around a third of individuals benefit from levodopa therapy (Fanciulli et al., 2019; Goh et al., 2023; Sullivan et al., 2021). Based on the predominant site of atrophy, MSA can be divided into two main clinical forms: MSA-parkinsonian (MSA-P) and MSA-cerebellar (MSA-C), each of which exhibits a unique motor phenotype, namely striatonigral and olivopontocerebellar, respectively (Fanciulli et al., 2019). While the pathogenesis of MSA remains unclear to a large extend, underlying genetic mechanisms are debated in specific cases. In certain families with MSA a biallelic mutation in the gene *COQ2* could be identified as a possible etiological factor of the disease (The Multiple-System Atrophy Research Collaboration, 2013). According to recent research, MSA might also share a genetic background with other repeat expansion diseases, particularly CANVAS, as the overlapping clinical symptoms already might have indicated (Cortese & Reilly, o. J.; Tolosa et al.,

2021). Recently, Montaut et al. analyzed a French cohort and identified a recessive intronic expansion in RFC1, specifically the $\mathrm{AAGGG}_{\mathrm{exp}}$repeat, as a potential underlying factor of the cerebellar subtype MSA-C (Montaut et al., 2021).

1.3. Parkinson's Disease (PD)

The following paragraphs aim to give a brief insight in the distinct features of the neurodegenerative disorder PD with a special emphasis on the molecular pathogenesis and REMs as a possible risk factor for developing the disease.

1.3.1. Incidence and Prevalence of PD

With a prevalence of 1-2 per 1,000 affected individuals (over 6 million affected people worldwide), PD constitutes the second most common neurodegenerative disease after AD. The incidence of PD increases with age, resulting in 1% of the population over 60 years and 3% of the population over 80 years being diagnosed with the condition, making it rather common in the older generation (Day & Mullin, 2021; Lotankar et al., 2017; Porter et al., 2006; Tolosa et al., 2021; Zirra et al., 2023). Consequently, this leads to an increased incidence of PD in economically more developed countries, where the overall life expectancy tends to be higher (Hirsch et al., 2016; Váradi, 2020). Furthermore, a male dominance has been reported in several observational studies, with males showing a 1.5 times increased risk of developing PD than females (Cerri et al., 2019; Gillies et al., 2014; Russillo et al., 2022).

1.3.2. Clinical Features of PD

The most characteristic clinical features of PD comprise motor symptoms including tremor, stiffness, postural instability, rigidity, akinesia, bradykinesia and hypokinesia as well as vocal symptoms (Lotankar et al., 2017; Opara et al., 2017; Tolosa et al., 2021). In current practice, these clinical manifestations are diagnostic criteria for PD (Váradi, 2020).

Non-motor symptoms encompass autonomic dysfunction (e.g. constipation or gastrointestinal dysfunction, urinal disturbance and cardiovascular dysregulation) and psychiatric manifestations such as depression, anxiety and disrupted sleep patterns (Rapid Eye Movement sleep behavior disorder). These non-motor symptoms tend to emerge several years prior to the onset of movement affecting symptoms and have a major impact on the patients disability (Day & Mullin, 2021; Lotankar et al., 2017; Váradi, 2020). A comprehensive illustration of common symptoms which may arise during PD is given in Figure 4.

Figure 4. Symptoms of Parkinson`s Disease.
The illustration shows summarized motor (left) as well as non-motor (right) symptoms. Motor symptoms involve a-, brady-, and hypokinesia, vocal symptoms, postural instability, tremor, issues with walking and gait as well as stiffness or rigidity. Non-motor symptoms comprise psychiatric issues such as depression and anxiety, cardio-vascular dysregulation, urinal and gastrointestinal dysregulation and disrupted sleep behavior. Figure created with BioRender.com

1.3.3. Pathogenesis of PD

PD is a neurodegenerative disorder of the central nervous system that affects not only the motor system but also results in neuropsychiatric and autonomic dysfunctions (see “Clinical Features of PD”) (Z. Chen et al., 2020; Lotankar et al., 2017; Opara et al., 2017; Tolosa et al., 2021). The complex pathogenesis process may involve genetic as well as environmental factors, such as pesticide exposure, head injury or stress (Day & Mullin, 2021).

The key feature of PD is the gradually increasing loss of dopaminergic neurons in the substantia nigra pars compacta, a component of the midbrain. Since the neurotransmitter dopamine plays a crucial role in coordination and movement, a dysfunction in this pathway results in impaired motor function, characteristic for individuals suffering from PD (Lotankar et al., 2017). In PD patients, the lack of dopamine can be visualized by DaTSCAN using Single-Photon Emission CT (SPECT), illustrated in Figure 5. Another major hallmark of PD is the formation of intracellular aggregates of misfolded α synuclein (ASN) proteins known as Lewy Bodies. These clusters of toxic proteins, which can propagate throughout the central nervous system, result in the degeneration of dopaminergic neurons and subsequently in genesis of PD (Lotankar et al., 2017). The two main characteristics of the pathogenesis of PD are depicted in Figure 6. Another crucial pathophysiological mechanism of PD is an abnormal increase of kinase activity. The most important player here is the autosomal dominant inherited gene *LRRK2* which codes for leucine-rich repeat kinase 2. Mutated forms of *LRRK2* correlate with increased kinase activity and constitute the most common cause of familial PD (Jeong & Lee, 2020; Mtias MAnn Steger 2016). The dysregulation of mitochondrial function is another important etiological factor of PD. The autosomal recessive inherited *PRKN* and *PINK1* are crucial for the control of mitochondrial fission and consequently for mitochondrial homeostasis. Mutations in these genes lead to disruption of the removal of damaged mitochondria and subsequently to their accumulation within the cell. This disturbed mitochondrial regulation is suggested to harm dopaminergic cells (Dias et al., 2013; Moon & Paek, 2015)

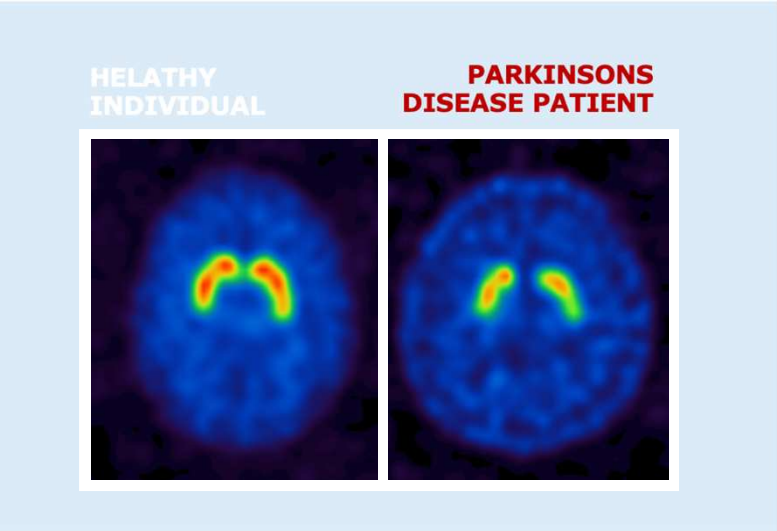

Figure 5. Dopamine Transporter Scan Using Single-Photon Emission CT.
The images show a comparison of a DAT-SPECT images of a healthy brain (left) versus the brain of a PD patient. In the patient asymmetrically reduced dopamine activity in the midbrain can be observed.(DAT-SPECT image from the Department of Biomedical Imaging, MUW (Courtesy: Prof. Dr. T. Traub-Weidinger)).

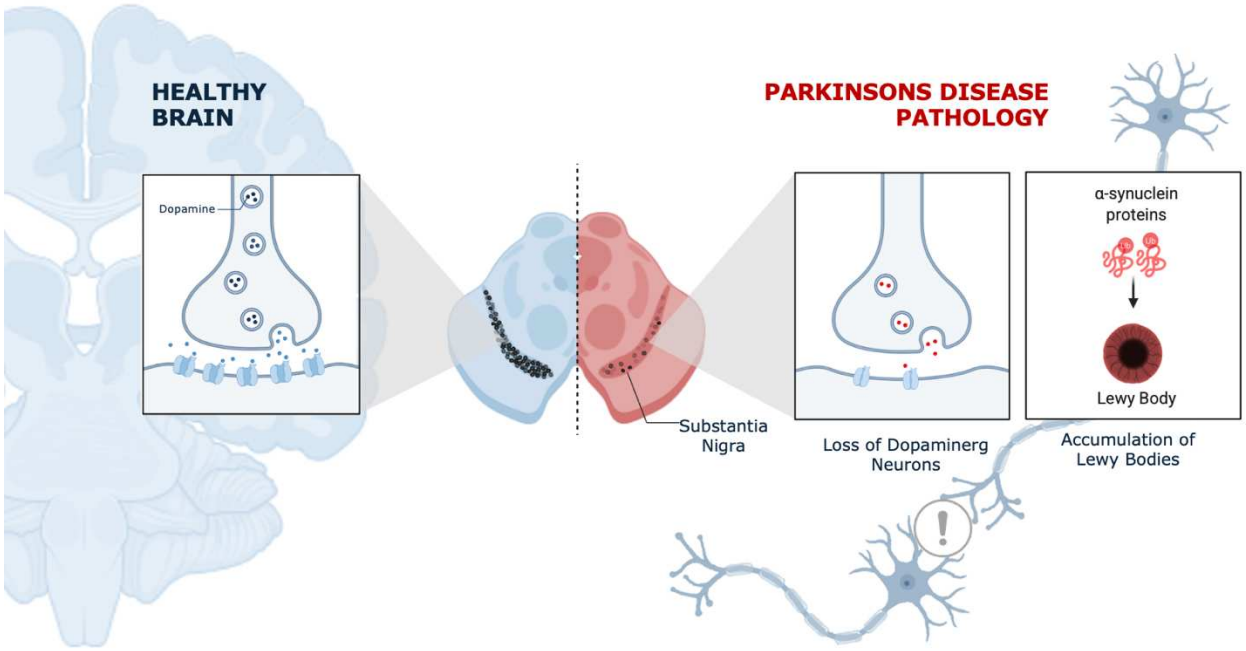

Figure 6. Main Hallmarks of Parkinson's Disease Pathology.
The illustration briefly summarizes the main characteristics of the Parkinson's Disease (PD) pathology and compares it to a healthy individual. The healthy brain (depicted on the left) shows normal organization, structure and function of neurons and neurotransmitter. Conversely, in the brain of a PD patient (illustrated on the right) a loss of dopaminergic neurons as well as a (consequent) dopamine deficiency in the pars compacta of the substantia nigra, situated in the midbrain, can be observed. Accumulation of α-synuclein proteins, due to underlying mutations of PD-specific genes, lead to the formation of Lewy Bodies. These toxic clusters further exacerbate the degeneration of dopaminergic neurons leading to a heightened decline in dopamine levels. Given the role of dopamine in the signal transduction for normal movement, dysfunction in this cascade results in impaired motor function which is a major phenotypic characteristic associated with PD (Antony et al., 2013). Figure created with BioRender.com

1.3.2.3. Genetic Landscape of PD

While most of the PD cases are considered to be sporadic, several genes have been identified so far, contributing to an increased risk of developing the disease.

Mutations affecting individual genes such as *SCNA*, *LRRK2 VPS35* and *PRKN* exhibit a rather high penetrance. As mentioned previously, ***SCNA*** codes for the α synuclein potein, which, when mutated, results in toxic aggregates, so called oligomers and fibrils which are thought to form successive pathogenic Lewy bodies (Porter et al., 2006).

Variations within the autosomal dominant leucine-rich repeat kinase 2 (***LRRK2***) gene, such as *G2019S and R1441G*/C/H display high penetrance diversity, as well as a spectrum of molecular pathways and phenotypes across affected individuals (Y. Chen et al., 2021; Day & Mullin, 2021; Lotankar et al., 2017; Taymans et al., 2023; Tolosa et al., 2021; Xiong et al., 2017; Zimprich et al., 2004). Notably, LRRK2 mutations have been associated with α synuclein and tau aggregation as well as clinical parkinsonism. They have been shown to play a central role in the main pathogenic pathways of PD, which were previously thought to be separate pathological avenues until the discovery of LRRK2 (Zimprich et al., 2004).

In 2011, another autosomal dominant Parkinson's gene was discovered, which encodes a part of the retromer complex, namely vacuolar protein sorting 35 (***VPS35****)* (Zimprich et al., 2011). In general the retromer is responsible for the transfer of different proteins between the endosomes and the cell membrane or golgiapparatus (Burd & Cullen, 2014; Waschbüsch & Khan, 2020). Being involved in the mitochondrial homeostasis and recycling receptors regulating hydrolase transport to lysosomes, particularly the D620N missense mutation has been identified to play a role in the pathogenesis of PD. The initial function is supposed to be disrupted which consequently leads to mitochondrial dysfunction and impairment of lysosomal and autophagy (Sassone et al., 2021; Williams et al., 2022, Zimprich et al., 2011).

Furthermore, it was shown that the *VPS35* D620N mutation also increases the phosphorylation of Rab, which are a large group of GTPases crucial for the movement of molecules between cellular compounds (membrane trafficking). This increase of phosphorylation can be reversed by *LRRK2* inhibitors (Mir et al., 2018). Consequently, this implies that the two genes *LRRK2* and *VPS35* are functioning in the same pathophysiological pathway, with *VPS35* upstream of *LRRK2*.

Initially linked with CANVAS, recent studies have implicated repeat expansion mutations in the above described ***RFC1*** gene as another risk factor for developing PD. The suggests prevalence of a biallelic REM in *RFC1* in PD patients ranges from 0.43 to 1.10 % in patients with European ancestry and was even found to increase up to 1.9 % in Finish individuals suffering from PD (Alvarez Jerez et al., 2024; Kytövuori et al., 2022; Ylikotila et al., 2023a). However, the exact pathological mechanisms causing PD remain to be fully elucidated.

Relevant genes associated with the risk for PD are illustrated in figure 7.

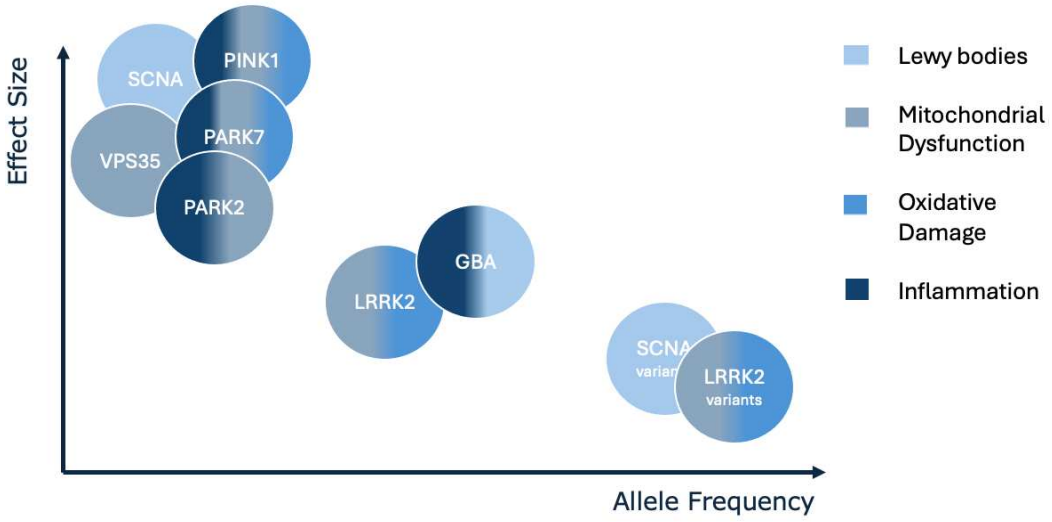

Figure 7 Genes Associated with Risk for Parkinson's Disease.
The figure illustrates risk genes for Parkinson`s Disease (PD) associated with their allelic frequency (x-axis) and their effect size (y-axis). Different shades of blue indicate involvement in different pathological mechanisms. Genes with a reported high effect size but low allelic frequency comprise SCNA, VPS35, PINK1, PARK6 as well as PARK2. LRRK2 and GBA show medium effect and frequency of alleles. Low effect size but high allelic frequency has been reported for SCNA as well as LRRK2 variants. SCAN (and variants) as well as GBA show involvement in Lewy body formation. Mitochondrial dysfunction has been described for VPS35, PINK1, PARK7 and -2 as well as LRRK2 (and variants). PINK1, PARK7 and LRRK2 (as well as variants) show oxidative damage participation. Inflammation involvement has been suggested for PINK1, PARK7 and -2 as well as GBA (Day & Mullin, 2021).

1.4. Aims and Hypotheses

With one to two per 1000 affected people, PD counts one of the most common neurodegenerative diseases worldwide. Given the predominantly sporadic nature of its development, a rather small fraction of PD cases can be explained by genetic factors. Since understanding of the etiology of PD can serve as an indication for possible preventive avenues or optimal therapeutic interventions, the identification of not yet discovered risk factors holds significant importance. Following a very recent investigation of the biallelic "AAGGG" expansion in a Finish cohort, which revealed the potential association with not only CANVAS but also PD, our objective was to evaluate if this finding was ethnic-specific for only the Finish population or if this distinct REM of *RFC1* also correlates with the development of PD in a large cohort of Austrian patients.

Our main aim was to screen a large cohort of PD, MSA and PSD patients for REM variants of *RFC1* to evaluate whether their frequency differs compared to neurologically healthy controls. A major focus was on the pathological biallelic *RFC1* REM $(\mathrm{AAGGG})_{\mathrm{exp}}$ recently reported for CANVAS.
Additionally, we investigated potential differences in *RFC1* variants and genotype frequencies in the PD, MSA and PSD cohort, between each other and the control group.
Furthermore, we were interested if the SNP rs11096992 located in *RFC1* and the REM showed a linked pattern of inheritance, as it yields potential for additional diagnostic avenues. We therefore aimed to test for the association of these two variants in the genome.
As an influence of the repeat number on severity on the respective disease was indicated in previous studies, we were also interested in the linear correlation between the repeat length and the age at onset of the disease. We hypothesized that larger expansion sizes are potentially associated with an earlier onset of first symptoms for patients carrying the biallelic $(\mathrm{AAGGG})_{\mathrm{exp}}$ genotype. We also aimed to analyze whether the onset age is associated by the respective *RFC1* genotype variant in individuals suffering from PD, MSA or PSD.
Furthermore, we intended to evaluate a possible correlation between a positive family history of PD and associated diseases and the occurrence of the pathogenic *RFC1* variant. Individuals suffering from PD and having a positive family history of the disease were presumably more likely to carry the biallelic $(\mathrm{AAGGG})_{\mathrm{exp}}$ genotype version of *RFC1*, than sporadic PD cases.
Finally, we were interested in the clinical picture of patients harboring the pathogenic biallelic $(\mathrm{AAGGG})_{\mathrm{exp}}$ *RFC1* genotype, to evaluate potential subphenotypes in the spectrum of PD symptoms.

1.4.1 Summarized Aims

***Aim* 1** Evaluate the distribution of *RFC1* variants in PD associated disease cohorts.

***Aim* 2** Analyze the association of expanded repeats with the distinct single nucleotide variant rs11096992 to investigate a potential risk haplotype.

***Aim* 3** Investigate the age at onset of patients carrying higher expanded repeats

***Aim* 4** Elucidate the rates of the pathogenic biallelic expanded *RFC1* genotype in the PD subset characterized by a positive family history.

***Aim* 5** Evaluate distinct PD phenotypes in patients carrying the pathogenic biallelic expanded *RFC1* genotype.

2. Material

2.1. Study Cohort

Study participants were recruited by neurology department of the “Medical University of Vienna” (*n* = 794), the “Medical University of Innsbruck” (*n* = 50), the “Medical University of Tulln” (*n* = 6), the “KABEG LKH Villach” (*n* = 5), the “Mayo Clinic” (*n* = 171), the “Johannes Kepler University Linz” and the “Clinic Donaustadt” (*n* = 183). Diagnosis included clinical examination (focused on motor abnormalities), imaging such as MRI, SPECT (DaTscan), FDG-PET, as well as additional evaluations like peripheral nerve conduction studies, autonomic function tests, wet biomarkers and vestibular evaluations in selected cases.
CANVAS samples were provided by the Medical University of Vienna” (*n* = 5).
Unaffected controls compromised 695 geographically aligned samples recruited by the “Medical University of Vienna” (*n* = 541) and in course of the “Vienna Transdanube Aging study” (n=154). Healthy controls with a known family history of PD were excluded from all analyses.
For ethics approval and consent to participate a consensus form was signed by all individuals or their legal caregiver.

2.2. General Material

2.2.1. General laboratory equipment

150 mm TC-treated Culture Dish, Corning®
7900HT Fast Real-Time PCR System; Applied Biosystem
Arium® Pro Ultrapure Water Sytem, Sartorius
Centrifuges; Hettich and Eppendorf
Eppendorf Research® plus multi-channel pipet; Eppendorf
Eppendorf Research® plus pipet; Eppendorf
Eppendorf® Multipette®; Eppendorf
Fusion X western blot and chemiluminescence imaging system; Vilber
IKA® MTS 2/4 multiwell plate shaker; IKA
PIPETMAN® L pipet; Gilson
Power Supply Model 200/2.0; Bio-Rad
QuickGene -610L Nuclear Acid Isolation System; Fujifilm
Rocking shaker with platform PMR-30; Grant-Bio
Shaking incubator, Thriller; VWR
Sub-Cell® GT Horizontal Electrophoresis System GT; Bio-Rad
Thermal Cycler Tetrad 2; Bio-Rad
TRIAD LT® Multi-mode Plate Reader; Dynex
Veriti 96 Well Thermal Cycler, Applied Biosystems
Shaking Water Bath 1083, GFL
PMR-30 rocking shaker, Grant-Bio

2.2.2 Consumables

8-Strip PCR Tube 0.2ml, individually attached flat caps; Star Lab
Bevelled TipOne® (Filter) Tips; Starlab
Corning TMMicroplate Aluminum Sealing Tape, Corning
DNA LoBind® Microcentrifuge Tubes; Eppendorf
Eppendorf® Combitips Advanced; Eppendorf
K3EDTA blood collection tubes VACUETTE ®; Greiner Bio-one Transfer Pipets; Sarstedt
MicroAmp® Optical 96-well Reaction Plate with Barcode; Applied Biosystems
Microseal 'B' PCR Plate Sealing Film; Bio-Rad
PCR 96-Well TW-MT-Platte (low profile, sub skirt); Biozym
Sealing film for manual application, PARAFILM® M
Serological Pipettes; VWR
Standard 96 Well PCR Plate; VWR Peqlab
TPP® Centrifuge tubes; TPP

2.2.3 Software & Programs

Concert -TRIAD Series Software, Triad
GraphPad Prism v 8.0.0 for Windows, GraphPad Software, San Diego, California USA
Peak Scanner™ Software v2.0, Applied Biosystems
Real-Time PCR System Software SDS v2.3; Applied Biosystems
Real-Time PCR System Software CopyCaller® v2.0; Applied Biosystems
R v4.3.1

2.3. DNA isolation

Blood collection tubes, with EDTA (K3EDTA); VACUETTE
DNA Whole Blood Kit L, Lysis Buffer LDB-03, DNA Whole Blood Protease EDB-01; Fujifilm, Kurabo

2.3.1. Qubit dsDNA High Sensitivity Assay

96-well flat bottom microplates (black), Greiner bio-one
Qubit® dsDNA BR Buffer, invitrogen
Qubit® ds DNA Standards, invitrogen
Qubit® dsDNA BR reagent (200x); Invitrogen

2.4. DNA precipitation

Ethanol absolute, *Fisher chemical*
Sodium acetate NaAc

2.5. RFC1 PCRs and Fragment Length Analysis

Dimethylsulfoxid, Sigma-Aldrich®
GeneScan TM – 600 LIZ TM Size Standard; Applied Biosystems
HiDi TM Formamide; Applied Biosystems
Taq PCR Core Kit (1000 units), Qiagen

2.6. TaqMan assays

Type-it® Fast SNP Probe Kit (800); Qiagen
TaqMan® SNP Genotyping Assay (rs11096992) probe; Thermo Scientific

2.7. Sequencing

α-Carboxypenicillin Dinatrium, Carl Roth
TOPO TA cloning Kit for Sequencing, Invitrogen
One Shot TOP10 cells
QIAprep Spin Miniprep Kit, Qiagen
LB-Medium, Carl Roth
Agar-Agar, Kobe I, Carl Roth

2.8. Gel Electrophoresis

6x MassRuler DNA Loading Dye; Fermentas
GeneRuler ^{TM}DNA ladder (1kb, 50bp); Thermo Scientific
LE Agarose; Biozym
peqGREEN DNA/RNA dye; VWR Peqlab
Tris-Borate-EDTA buffer (TBE, 10x), Sigma

2.9. Southern blot

20xSSC buffer, UltraPureTM, invitrogen
AccuGene® Molecular Biology Water, Lonza
CDP-Star, ready-to-use; Roche
DIG Easy Hyb; Sigma
DIG Wash and Block Buffer Set; Sigma
DNA Molecular Weight Marker II, DIG-labeled, Sigma
DNA Molecular Weight Marker VII, DIG-labeled, Sigma
EcoRI, Kit: R0101S, Biolabs
Hydrochloric acid (HCl, 2mol/l), Titripur
Nylon Membrane (30cmx3m), positively charged; Roche
pH-Indikatorpapiere, Duotest, MACHEREY-NAGEL
QIAEX® II Gel Extraction Kit (150); Qiagen
Saline sodium citrate (SSC), 20x; Lonza
Salmon Sperm (10.0mg/ml); invitrogen
Sodium chloride (NaCl), Sigma
Sodium dodecyl sulfate (SDS); Sigma
Sodium hydroxide solution (NaOH, 2mol/l), Roth
TurboBlotter System, GE Healthcare
UVP Crosslinker, analytik jena
WhatmanTM Chromatography Paper (19cmx100m, 3MM CHR); GE Healthcare
Genechip Hybridization oven 645 (adapted for hybridization tubes), Affymetrix Amersham
WhatmanTM GB003 Gel Blot Paper (15x20cm); cytiva
Wheaton® hybridization bottle (70x300mm); Wheaton

PCR DIG Probe Synthesis Kit, Roche

3. Methods

3.1. DNA Isolation from Whole Blood Samples

All procedure steps were taken accordingly to the “Quick Gene DNA whole blood Kit L” from KURABO.

Venous blood specimens, which were taken from clinicians in advance and stored in 9 ml EDTA tubes at -20°C, were slowly thawed and vortexed to eliminate any lumps and ensure a homogenous sample. 300 µl Protease (EDB) were pipetted into a 15ml centrifugation tube. Then 2 ml blood as well as 2.5 ml Lysis Buffer were added. The 15 ml tubes were then vortexed to mix the compounds thoroughly. Incubation for 25 minutes at 56°C on a heat block was followed by the addition 2.5 ml ethanol (>99%). The tubes were again subjected to vortexing to avoid clogging of the cartridge filter in the washing and elution process with Wash and Elution buffer (WDB, CDB respectively). The resulting cell lysate was then transferred into the cartridge of the QuickGene-610L (FUJIFILM). The “DNA WHOLE BLOOD” mode was selected with a modified elution volume of 200 µl instead of 500 µl to obtain higher concentrated DNA. The resulting genomic DNA samples were stored at 4 or -20 °C for subsequent analyses.

All vortexing steps were executed at maximum speed for minimum duration 15 seconds.

3.2. DNA Quantification

DNA concentration was either measured using Nanodrop or Qubit dsDNA High Sensitivity Assay Kit (according to the provided protocol).

3.2.1. Nanodrop

For determination of DNA concentration utilizing the “NanoDrop™ Lite Spectrophotometer” 1 µl of the elution buffer or dH_2O was used for blanking before measuring 1 µl sample. The ratio of absorbance at 260 and 280 nm was used to assess DNA purity.

3.2.2. Qubit dsDNA High Sensitivity Assay

Qubit dsDNA HS Reagent was diluted in a ratio of 1:200 in Qubit dsDNA HS buffer. In a 96 well plate 3 µl of each DNA sample, along with duplicates of five standards (0, 10, 25, 50 and 100ng/µl, 3 µl of each), were combined with 120 µl of the prepared master mix (detailed information as well as volumes for one example reaction are given in table 1). After a shaking incubation at RT for 15 to 25 minutes concentrations were measured using the “Concert -TRIAD Series” software at the “Dynex Triad LT Multimode Detector/Plate Reader”. The concentrations were deduced based on the quantification of measured absorbance units. The standard tuples were averaged and subsequently a standard curve with an associated regression formula was calculated. The regression formula represented as: $y = k*x + d$ (where x denotes the absorbance of each sample and y signifies the unknown concentration) was then applied for concentration calculations.

Table 1. Ingredients for Qubit dsDNA High Sensitivity Assay.
Reaction components and their associated stock molarity/concentration (first row), the final dilution required (second row) and the calculated volume for the measurement of one genomic DNA sample, resulting in a 123 µl reaction (third row). The relation of Qubit and Quantit is 1:200.

Reagent (stock)	Final volume for one reaction [µl]
Qubit™ dsDNA BR Buffer	119.4
Qubit® dsDNA BR (200X)	0.6 (1X)
DNA	3.0

3.3. DNA precipitation

DNA precipitation was conducted to acquire highly concentrated samples for Southern blotting. For each sample volume, 0.1 volumes of sodium acetate along with 3 volumes of ice-cold ethanol (>99%) was combined in 1.5 ml Eppendorf tubes. After thorough mixing the samples were centrifuged at 13 000 rpm at 4°C for 15 min. The resulting supernatant was carefully discarded and the samples were washed using 1 ml ethanol (70%) per sample. They then underwent an additional centrifugation step under identical conditions for 2 minutes. Following the removement of any residual supernatant, the Eppendorf tubes were left open to facilitate the vaporization of any remaining ethanol. A final eluting step using 20 µl water was executed. After concentration assessment of the precipitated samples, they were stored at 4 or -20°C. For detailed information on the reaction volumes see table 2.

Table 2. DNA Precipitation Reagents.
Reaction components and their associated stock molarity/concentration (first row), the volume relation required (second row) and the calculated volume for 100 µl of genomic DNA, resulting in a 410 µl reaction (third row).

Reagent (stock)	X volumes	Final example volume
DNA	1.0 volumes	100.0 µl
EtOH (>99%)	3.0 volumes	300.0 µl
NaAc (3M)	0.1 volumes	10.0 µl
EtOH (70%)		1.0 ml

3.4. Gel Electrophoresis

All PCR products were examined via gel electrophoresis. In the case of *RFC1*, *FBN1* and all probe generating PCRs an analytical gel was conducted to check the presence and size of the amplicons. In case of all other PCRs gel electrophoresis was utilized to check the PCRs quality for further analyses. If not specified otherwise, all conditions comprised a 2% agarose gel consisting of Tris-Borate-EDTA (TBE) buffer with 2% agarose and 0.05% peqGreen DNA dye (for subsequent detection of the nucleic acids utilizing fluorescence via UV excitation), as depicted in table 3. Electrophoresis was performed in TBE (1X) at a voltage of 120V for a duration of 40 minutes. Final gels were visualized on “FUSION FX EDGE” using “DNA Gele” and “Auto” lightening conditions by the “EVOLUTION-CAPT” software.

Table 3. 2% Agarose Gel Description.
Reagent, required amount/dilution, as well as the volumes for an example gel of 200ml for a 2% agarose gel.

Reagent	Reagent Utilization	Quantity
Agarose	2 %	4.0 g
TBE (10X)	1 X	Δ 200 ml
peqGreen	0.05 %	10.0 µl

Abbr.: Δ = "difference to final volume of ".

3.5. Multi-Step Protocol

In our analyses we were aiming to identify the following *RFC1* variants: the biallelically healthy wildtype *RFC1* (WT/WT), the with CANVAS associated biallelic repeat expansion of the pathogenic motif "AAGGG" referred to as MUTexp/MUTexp, the heterozygous variant MUTexp/WT with a monoallelic expansion of the pathogenic motif as well as biallelic and monoallelic expansion of the WT motif "AAAAG" referred to as WTexp/WTexp and WTexp/WT, respectively. Summarized nomenclature of *RFC1* genotype variants is given in table 4. To discern these specific variants, we applied a multi-step procedure which primary consisted of a *RFC1* WT flanking PCR (fPCR), a repeat primed PCR (RP-PCR) amplifying either the pathogenic motif expansion or the WT motif, and a FAM labelled MUTexp amplifying fPCR. Details are declared in the flow chart in figure 8.

Table 4. Genotype Nomenclature.
Genotype, in the text used nomenclature as well as a brief explanation of RFC1 variants.

Genotype	Nomenclature	Explanation
$(\mathrm{AAGGG})_{\mathrm{exp}}/(\mathrm{AAGGG})_{\mathrm{exp}}$	MUTexp/MUTexp	Biallelic repeat expansion mutation of the pathogenic motif "AAGGG"
$(\mathrm{AAAAG})_{\mathrm{exp}}/(\mathrm{AAAAG})_{\mathrm{exp}}$	WTexp/WTexp	Biallelic repeat expansion mutation of the WT motif "AAAAG"
$(\mathrm{AAGGG})_{\mathrm{exp}}/(\mathrm{AAAAG})_{11}$	MUTexp/WT	Monoallelic repeat expansion mutation of the pathogenic motif "AAGGG"
$(\mathrm{AAAAG})_{\mathrm{exp}}/(\mathrm{AAAAG})_{11}$	WTexp/WT	Monoallelic repeat expansion mutation of the WT motif "AAAAG"
$(\mathrm{AAAAG})_{11}/(\mathrm{AAAAG})_{11}$	WT/WT	Biallelic WT RFC1 with 11 repeats of the "AAAAG" motif

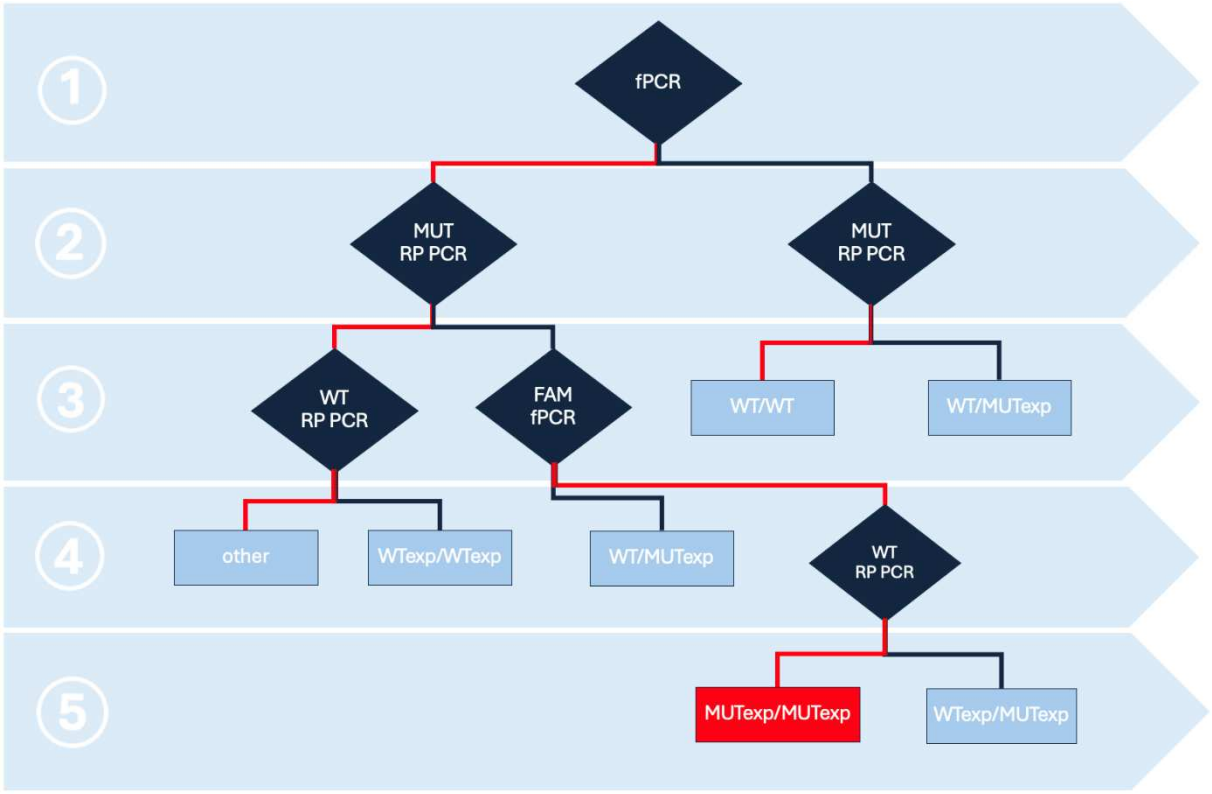

Figure 8. Multi-Step-Protocol.
This flow chart represents the 5-step-protocol of the diagnostic of the different genotypes. Deltoids (dark blue) symbolize experimental procedures. Here red lines indicate a negative test result, whereas dark blue lines suggest a positive result. Squares (light blue) depict the final observed genotype. The genotype of main interest (MUTexp/MUTexp) is marked in red. Abbr.: fPCR = flanking PCR, MUT RP-PCR = repeat primed PCR targeting the expanded mutated motif (MUTexp), WT RP-PCR = repeat primed PCR targeting the expanded wild type motif (WTexp), FAM fPCR = FAM labelled flanking PCR.

3.5.1. Flanking PCR

To assess the presence of the 11 repeats long *RFC1*-WT variant, a flanking PCR utilizing the RFC1_F and RFC1_R primers flanking the WT repeat (see table 20 for all primer sequences) was performed. The reaction mixture was composed of 10X buffer, 200 µM deoxyribonucleotide triphosphates (dNTPs), 0.5 µM of forward and reverse primer respectively (RFC1_F and RFC1_R), 1.25 units of Taq Polymerase and 20 ng of genomic DNA. All reactions were executed with a final volume of 10µl. To mitigate the risk of false-negative outcomes for the *RFC1* flanking PCR, a parallel control PCR was conducted for *FBN1* (with an adjusted primer volume of 0.25µM). The details for all flanking PCR reaction concentrations as well as example volumes are specified in table 5 and 6). All PCRs were conducted with the “Biorad Thermal Cycler Tetrad 2” or the “Applied Biosystems Veriti 96 Well Thermal Cycler” (model 9902 Lab) following standard PCR conditions (see table 7). For evaluation of the presence of the targeted sequence, the PCR products were checked on agarose gel. The anticipated band size for *RFC1* and *FBN1* was 253 and 132 bp, respectively.

Table 5. RFC1 Flanking PCR Reagents.
Reaction components and their associated stock concentration (first row), the final concentration needed for the reaction (second row) and the calculated volume for a 10 µl reaction (third row).

Reagent (stock)	Final concentration	Volume [µl]
Buffer (10X)	1 X	1.0
dNTPs (10mM)	200.0 µM	0.2
RFC1_F (10mM)	0.5 µM	0.5
RFC1_R (10mM)	0.5 µM	0.5
Taq Pol	1.25 units	0.1
DNA (10ng/µl)	20.0 ng	1.0
dH_2O		6.7

Table 6. FBN1 Flanking PCR Reagents.
Reaction components and their associated stock concentration (first row), the final concentration needed for the reaction (second row) and the calculated volume for a 10 µl reaction (third row).

Reagent (stock)	Final concentration	Volume [µl]
Buffer (10X)	1X	1.00
dNTPs (10mM)	200 µM	0.20
FBN1_F (10mM)	0.25 µM	0.25
FBN1_R (10mM)	0.25 µM	0.25
Taq Pol	1.25 units/50 µl	0.10
DNA (10ng/µl)	20 ng	1.00
dH_2O		2.72

Table 7. Thermocycling Conditions for the RFC1 and FBN1 FAM flanking PCRs .
Cycle step (1 column), associated temperature (second row) and duration of the respective PCR part.

PCR step		Temperature [°C]	Duration
Initial denaturation		94	4 min
Denaturation	30x	95	30 sec
Annealing		59	30 sec
Extension		72	1 min
Final extension		72	5 min

3.5.2. Repeat Primed PCRs

To detect samples containing a repeat expansion mutation, a RP-PCR was executed. For this distinct form of PCR three specific primers are required. The first primer is a locus specific FAM labeled forward primer binding upstream of the repeat. The second, repeat-motif-specific, primer binds inside the repeat and holds a DNA tail sequence at the 5` end (absent in the human genome) on which the third primer (reverse primer) can bind (Loureiro et al., 2018). FAM, also referred to as fluorescein amidite, is a fluorescent dye which shows high affinity for DNA (or other

biomolecules, depending on the application), facilitating labeling and detection in subsequent genotyping analyses (Lietard et al., 2022).
This RP-PCR design introduces fluorescence labeled amplicons of different size depending on where the second primer has bound within the repeat expansion. It is crucial that the second repeat specific primer is utilized in decreased volume to prevent further shortening of longer amplicons during the PCR. Together with the successive preference of shorter fragments by the PCR mechanism the process enables the finished PCR product to consist of a large amount of small, a medium amount of middle-sized and a small amount of large FAM labelled amplicons in case of a REM (a schematic representation of the RP-PCR design is illustrated in figure 9). As in the subsequent section “Fragment Length Analysis” described genotyping results in a “chainsaw” pattern reflecting the unique distribution of fragment sizes.
The reaction mixture facilitating this RP-PCR consisted of 10X buffer, 200 µM dNTPs, 0.5 µM of forward and reverse primer (RFC1_RP_F_fam and RFC1_RP_R, respectively), 0.3 µM of the primer binding inside the desired (which was constructed by a mixture of 3 - motif specific primers: RFC1_RP_pat1, RFC1_RP_pat2, RFC1_RP_pat3 or RFC1_RP_WT1, RFC1_RP_WT2 and RFC1_RP_WT3), 1.25 units of Taq Polymerase, 50 ng of genomic DNA as well as 3% dimethyl sulfoxide (DMSO). Detailed concentrations and volumes for both reactions, the WT and the pathogenic motif targeting RP-PCRs are given in table 8.
The RP-PCR thermocycling protocol was divided into two cycles. The initial cycle, repeated 25 times, started after an initial denaturation and consisted of denaturation, annealing and extension phases. With successive cycle number, the annealing temperature decreased by 0.2 °C ranging from 60 (annealing temperature at first cycle) to 55 °C (annealing temperature at cycle number 25). The second cycle again involved denaturation, annealing and extension phase, maintaining a constant annealing temperature of 55°C and undergoing only 10 repetitions, followed by a final extension. Detailed Thermocycling conditions are provided in Table 9. To assess the outcome, prepared samples were sent to genotyping as detailed in the "Capillary Electrophoresis and Electropherogram Analysis" section.
Possible outcomes are illustrated in figure 10.

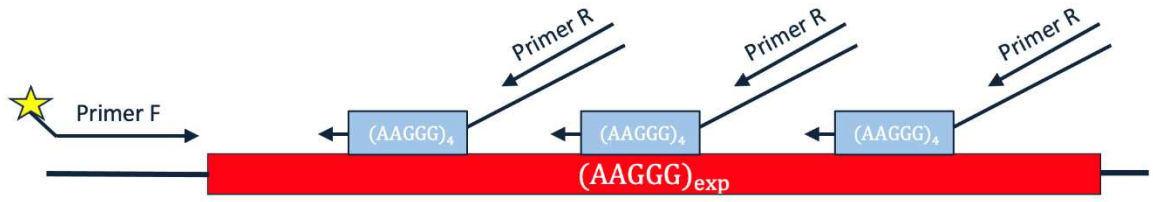

Figure 9. Schematic Representation of a RP-PCR.
The illustration pictures a RP-PCR specific for the pathogenic motif (AAGGG). The forward primer (FAM labelled) binds upstream the repetitive locus of RFC1, the second repeat specific primer, possessing a tail, binds inside the repeat at variable locations. The (third) reverse primer is complementary to the tail of the repeat primer. Utilizing this PCR design, for samples carrying a MUTexp haplotype, a distinct “Chainsaw” pattern can be observed.

3.5.2.1. Pathogenic motif - specific repeat primed PCR

To identify samples carrying the monoallelic or even the biallelic pathogenic REM (MUTexp/WT or MUTexp/MUTexp, respectively) all samples as well as all healthy controls were subjected to a RP-PCR which was specifically targeting the pathogenetic motif "AAGGG".
The previously described reaction mix was adjusted by adding primers binding inside this mutant motif. This set of primers consisted of three motif-specific ones: RFC1_RP_pat1, RFC1_RP_pat2, and RFC1_RP_pat3.

3.5.2.2. Wildtype motif - specific repeat primed PCR

While all samples were subjected to a *RFC1*-MUTexp detecting RP-PCR, only samples suspected to carry the MUTexp/MUTexp genotype (detected via negative results for *RFC1* fPCRs in combination with positive results for pathogenic motif - specific RP-PCR) or samples with unclear outcome (negative for the *RFC1* flanking PCRs as well as the pathogenetic motif - specific RP-PCR) were subjected to a RP-PCR detecting a *RFC1*-WTexp. Within the reaction mixture employed for this analysis, the repeat specific primer mix was adjusted to include primers designed to bind inside the WT motif (RFC1_RP_WT1, RFC1_RP_WT2 and RFC1_RP_WT3), however, the concentrations remained constant.

Table 8. RFC1 RP-PCR Reagents.
RP-PCR reagents for the detection of RFC1 repeat expansion mutation of the RFC1 mutant as well as WT motif (MUTexp and WTexp). Reaction components and their associated stock concentration (first row), the final concentration needed for the reaction (second row) and the calculated volume for a 20 µl reaction (third row).

Reagent (stock)	Final concentration	Volume [µl]
Buffer (10X)	1X	2.0
dNTPs (10mM)	200.000 µM	0.4
RFC1_RP_F_fam (10mM)	0.250 µM	1.0
RFC1_RP_pat1/RFC1_RP_WT1	0.025 µM	0.1
RFC1_RP_pat2/RFC1_RP_WT2	0.025 µM	0.1
RFC1_RP_pat3/RFC1_RP_WT3	0.025 µM	0.1
RFC1_RP_R (10mM)	0.250 µM	0.1
Taq Pol	1.25 units	0.2
DNA (10ng/µl)	50.0 ng	5.0
DMSO	3.0 %	0.6
dH_2O		10.1

Table 9. Thermocycler conditions for RP-PCRs.
Thermocycling conditions of RP-PCRs amplifying RFC1- MUTexp or -WTexp. Cycle step (1 column), associated temperature (second row) and duration of the respective PCR part.

PCR step		Temperature [°C]	Duration
Initial denaturation		95	2 min
Denaturation	25x	95	30 sec
Annealing		60 – 55 (in 0.2 steps)	30 sec
Extension		72	2 min
Denaturation	10x	95	30 sec
Annealing		55	30 sec
Extension		72	2 min
Final extension		72	7 min

3.5.3. FAM flanking PCR

Samples lacking a visible band in the correct size segment for *RFC1* and those which had a positive outcome for one of the RP-PCRs underwent further analysis via FAM labelled *RFC1* flanking PCR. This procedure aimed to avoid false negative results in the more rudimentary analysis of the flanking PCRs conducted on agarose gel. Similar to the previously performed *RFC1* flanking PCR, the reaction mixture comprised a 10X buffer, 200 µM dNTPs, 0.5 µM each of FAM labelled forward primer (RFC1_F_fam) and reverse primer (unlabeled RFC1_R), around 1.25 units of Taq Polymerase, and 20 ng of genomic DNA. Detailed information regarding concentrations and volumes of components as well as the PCR cycle conditions, consistent with those employed in the conventional *RFC1* flanking PCR, are depicted in table 5 and 7 respectively. For evaluation of the results, prepared samples were sent for genotyping (see "Capillary Electrophoresis and Electropherogram Analysis").

3.6. Fragment Length Analysis

All PCR products previously tagged with a fluorescence marker (FAM) were subjected to external performed genotyping. To prepare for capillary electrophoresis 3 µl of 1:10 diluted PCR product was mixed with 0.2 µl of GeneScan™ 600 LIZ™ dye Size Standard v2.0 as well as 23.8 µl Hi-Di™ Formamide per fragment (see table 10). Subsequently external capillary electrophoresis was conducted by the Microsynth Austria GmbH. The resulting electropherogram data was analyzed utilizing Applied Biosystems` "Peak Scanner Software 2.0".
To verify a healthy non expanded *RFC1* locus $(\mathrm{AAAAG})_{11}$ on at least one allele, the FAM marked *RFC1* fPCR product should exhibit a singular peak at the desired size of about 253 bp in the electropherogram (exemplary picture shown in figure 11). A REM on at least one allele could be concluded from the presence of a distinctive "chainsaw pattern" upon examination of the electropherogram obtained from the RP-PCR, as depicted in figure 10.

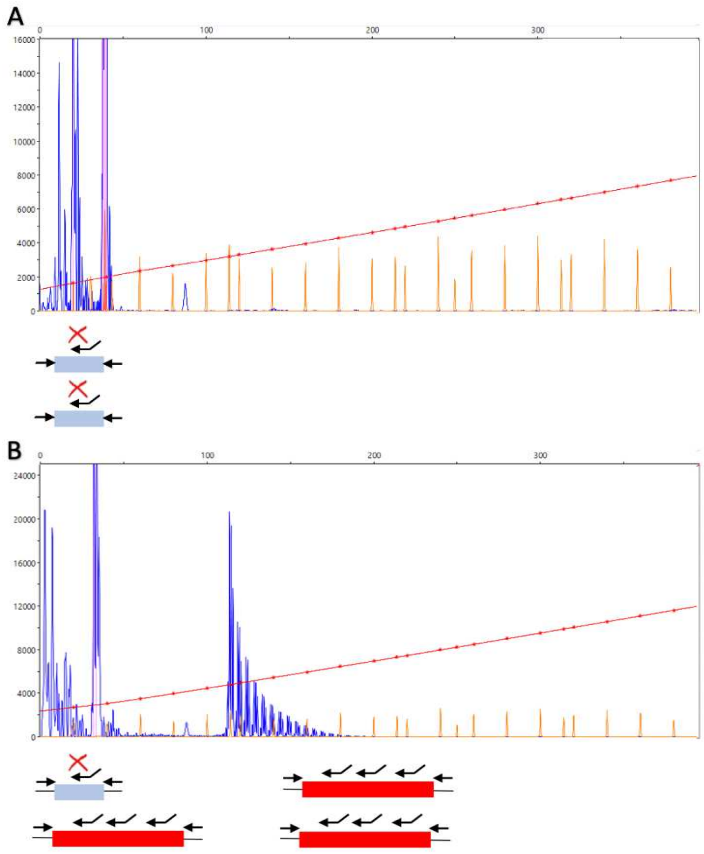

Figure 11. Possible Outcomes of RP-PCR Electropherograms.
Electropherogram in panel A shows no amplification via RP-PCR, due to WT sized RFC1 *on both alleles. Panel B displays a distinct "chainsaw pattern" characteristic for a repeat expansion mutation on at least one allele. High peaks on the left side and small peaks on the right side of the pattern correspond to the large quantity of small and the minor number of large amplicons, respectively.*

Table 10. Preparation of Samples for Fragment Length Analysis.
To 1:10 diluted samples added reagents (column one) as well as the required volume for one sample (column two).

Reagent	**Volume [µl]**
GeneScan 600 LIZ	0.2
Hi-Di Formamide	23.8
PCR product (1:10 diluted)	3 .0

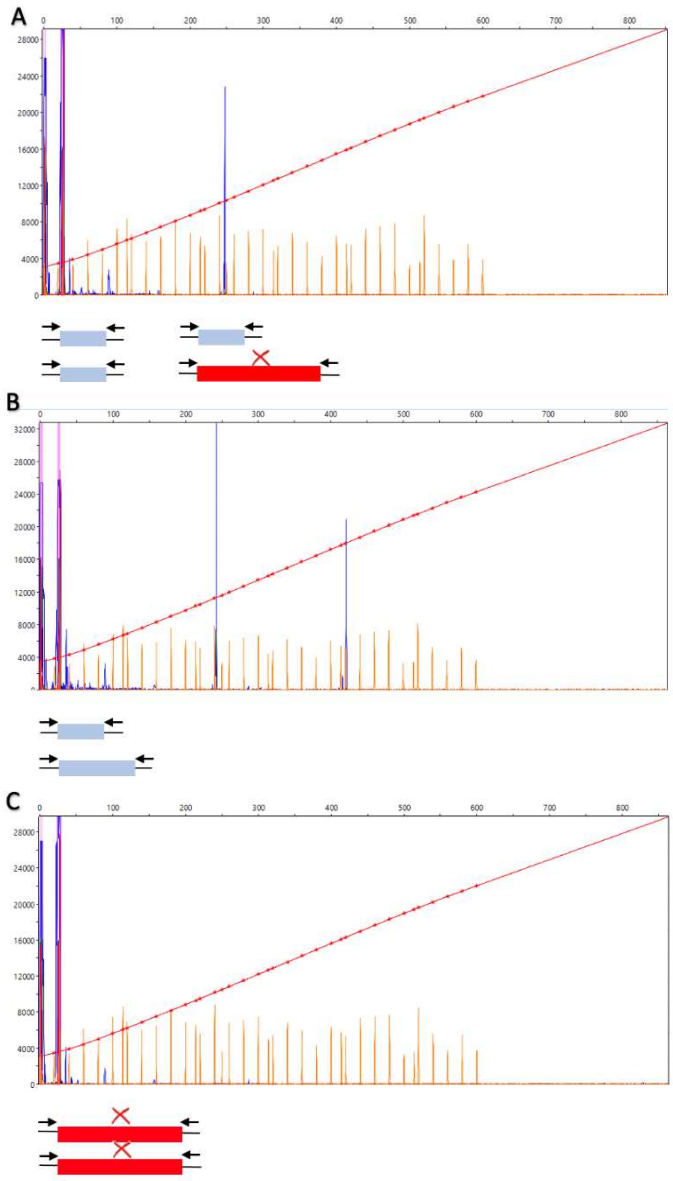

Figure 12. Possible Outcomes of FAM labeled RFC1 flanking PCR Electropherograms.
Panel A shows a one (blue) peak at approx. 250 bp, indicating one or two alleles carrying a non-expanded RFC1. The second panel (B) displays two peaks (at approx. 250 and 410 bp, respectively) which indicates differently size RFC1 loci (one WT and one intermediate sized). The third panel (C) shows no peak for RFC1 at all due to a non-amplifiable RFC1, suggesting a repeat expansion mutation on both alleles.

3.7. Quantitative PCR

DNA samples were screened for one SNP ("rs11096992") in the *RFC1* gene. Pre-diagnosed CANVAS as well as healthy samples were used as controls on each plate. 2µl sample of the DNA plates (20ng in total) were pipetted into a semi-skirt 96 well PCR plates appropriate for the quantitative PCR (qPCR) device. Next, 10µl master mix was prepared, consisting of 5µl buffer

(from the Type-it® Fast SNP Probe Kit, Qiagen), 0.5µl TaqMan® SNP Genotyping Assay (rs11096992) probe as well as 3µl ddH2O for each sample. Detailed reaction volumes as well as thermocycler conditions are given in table 11 and 12. The probes used in this assay are specific for either the risk or the wildtype allele. The probe targeting to the risk allele is tagged with VIC and the probe binding to the wildtype allele carries a FAM marker. This distinct labeling enables the identification of the respective haplotypes through different fluorescence signals.

Table 11 rs11096992 qPCR Reagents.
Reaction components and their associated stock concentration (first row), the final concentration needed for the reaction (second row) and the calculated volume for a 10 µl reaction (third row)

Reagent (stock)	Final concentration	Volume [µl]
Buffer	2X	5.00
Probe	20X	0.50
DNA (10ng/µl)	10 ng	1.00
dH_2O		3.00

Table 12. Thermocycling Conditions for the rs11096992 qPCR.
Cycle step (1 column), associated temperature (second row) and duration of the respective PCR part.

PCR step		Temperature [°C]	Duration
Initial activation		95	5 min
Denaturation	40X	95	15 sec
Annealing and Extension		60	30 sec

3.8. Sequencing

To replicate the sequence of interest in amounts suitable for DNA sequencing, the target sequence was cloned into a plasmid and introduced into a competent strain of E. coli. According to the supplied protocol of the "TOPO TA Cloning Kit for Sequencing" by Invitrogen, a cloning reaction as well as a transformation of the cells was performed. According to the "QIAprep spin Miniprep Kit" consequently a plasmid isolation was undertaken followed by external Sanger sequencing (executed by the Microsynth Austria GmbH).

3.8.1. PCR for Sequencing

To generate fragments suitable for cloning and transformation, *RFC1* flanking as well as *RFC1* RP-PCRs were performed according to previously described conditions. To obtain amplicons without FAM labels, forward primer without any fluorescent marker were used. RFC1_RP_F_fam was consequently replaced by RFC1_RP_F_UL.

3.8.2. Fabrication of LB Agar Plates

To prepare agar plates for subsequent cloning Lysogeny Broth, also referred to as LB (25g/L) and Agar Agar (15g/L) were diluted in dH_2O. The pH value was adjusted to 7.0 with 5 N NaOH.

Subsequent to autoclavation (121 °C, 25 min), when the media temperature dropped to of 58°C, carbenicillin (4 ml/L) was added. The media was then poured into plates (with a diameter of 150mm) and let harden. They were stored at 4°C. Reagent details are given in table 13.

Table 13. Ingredients of LB Agar Plates.
Reagents (column one), the associated required concentration (column two), as well as the final volume of the respective reagent used for a total volume of 500 ml.

Reagent	Final concentration	Volume and mass
LB	25g/L	12.50 g
(NaCl,	(NaCl, 10 g/L	(5.00 g
Tryptone,	Tryptone, 10 g/L	5.00 g
Yeast Extract)	Yeast Extract, 5 g/L)	2.25 g)
Agar Agar	15g/L	7.50 g
Carbenicillin	4 ml/L	2.00 ml
dH_2O		Δ 500.00 ml

Abbr.: Δ = "difference to final volume of "

3.8.3. Cloning and Transformation

According to the "TOPO TA Cloning Kit for Sequencing" protocol, the TOPO cloning vector was prepared by mixing 2 µl of the PCR product (see "PCR for sequencing") with 1 µl of the salt solution as well as the TOPO vector and 2 µl of dH_2O for each reaction. Subsequently, 2 µl of the final TOPO cloning vector was added into one vial One Shot chemically competent E. coli cells (twice performed for each PCR product) and mixed carefully by inverting. After an incubation for 20 minutes on ice, the cells were heat shocked utilizing a water bath (42°C, 30 sec) and immediately put back on ice again afterwards. 250 µl S.O.C. medium (brought to RT) was added to each vial. Subsequently the cell-mixture was incubated on a shaking incubator (200 rpm) at 37 °C for a duration of 1 hour. Consequently, the cells were spreaded on prewarmed LB/Agar plates (40 µl of each transformation/plate) and incubated over night at 37°C upside down. Summarized reaction volumes for cloning as well as transformation can be seen in table 14 and 15, respectively.
The next day single colonies were picked with a pipet tip and put into 5 ml LB medium (tubes were left open for oxygen supply), where the cultures were allowed to grow in a shaking incubator at 37°C over night.

Table 14. Reagents for Cloning.
Reagents as well as the corresponding required volume for one cloning reaction.

Reagent	Volume [µl]
PCR product	2.0
Salt solution	1.0
TOPO vector	1.0
dH_2O	2.0
Final TOPO cloning vector	(∑)6.0

Abbr.: ∑ = sum

Table 15. Reagents for Transformation.
Reagents as well as the corresponding required quntity for one transformation reaction.

Reagent	Quantity
TOPO cloning vector	2 µl
One Shot chemically competent E.coli	1 vial
S.O.C.	250 µl

3.8.4. Plasmid Isolation

After a successful cultivation (identified though a milky LB medium), the plasmids were isolated utilizing the "QIAprep spin Miniprep Kit" and the supplied protocol. The tubes were centrifuged at 5 000 rpm for 5 min at RT. After supernatant media was removed, the pallets were resuspended in 250 µl P1 and transferred into 2 ml Eppendorf tubes each. 250 µl P2 (includes LyseBlue) was introduced to the mixture, which was subsequently inverted four to six times until the solution becomes clear blue. Immediately after this visual confirmation of the lysis via LyseBlue, 350 µl of N3 was added, again followed by four to six inverting steps, until the solution became clear again. The lysate was then centrifuged (13 000 rpm, 10 min, RT). 800 µl of the supernatant was pipetted into a spin column (placed into a 2 ml Eppendorf tube) and centrifuged again at 13 000 rpm for 1 min at RT. The flow through was discarded. 500 µl PB was introduced to the pallet followed by a centrifugation step (13 000 rpm, 1 min, RT). After removal of the flow through 750 µl PE was added. After a discarding the flow through another centrifugation step (13 000 rpm, 1 min, RT) was performed. The column was then placed into a clean 1.5 ml Eppendorf tube and 50 µl EB was introduced. After a short incubation for 1 min the tube was briefly centrifuged and the concentration was measured (see "Nanodrop").
To prepare the final isolated plasmids for genotyping, 40-100 ng of the plasmid DNA (not exceeding a volume of 12 µl) was diluted in dH_2O to a volume of 12 µl. After the addition of 2 µl of the corresponding forward primer, the samples were sent for sequencing (performed by the Microsynth Austria GmbH).

3.9.Southern Blot

For size assessment Southern blotting was carried out with all the previously detected risk genotypes (MUTexp/MUTexp). Preparation of samples involved the restriction digest of the genomic DNA as well as a generation of a *RFC1* specific probe. DNA fragments of the digests were separated with gel electrophoresis on day 1, followed by denaturation, depurination and neutralization of the gel. Southern blotting was performed by transfer to a membrane on day 2, hybridization on day 4 and detection on day 5.

3.9.1. Restriction Digest

To prepare the genomic DNA samples for Southern blotting a restriction digest was employed. This procedure involved the utilization of the restriction enzyme EcoRI to selectively cleave the double helices at restriction sites harboring a 5'-GAATTC-3' sequence, forming cohesive sticky ends. The reaction mixture required a minimum of about 10 000 ng of genomic DNA, less then 10% (of the total reaction volume) EcoRI enzyme of the total reaction volume and 10-fold buffer for optimal results. To ensure identical reaction volumes, facilitating better comparability in subsequent steps, all genomic DNA samples were brought to the same volume of 21.1 µl containing 9,500 ng DNA before executing a restriction digest (except for sample 3314 with a lower concentration and a resulting concentration of 4,600 ng/21.1 µl). The detailed reaction mixture of all samples is given in table 16. An incubation at 37°C for duration of 72 hours was conducted with the introduction of additional enzyme and buffer to the reaction at 24 and 48 hours, respectively. Subsequently the efficacy of the restriction digest was assessed through electrophoresis (120V for 60 min) on an 1% agarose gel where the anticipated outcome was the manifestation of an observable smudge pattern (see figure 12). To deactivate the enzymatic activity, the incubation was adjusted to 56°C for 20 min, followed by placing all restriction digested samples into an ice bath.

Table 16. Restriction Digest Reaction Volumes.
Reaction volumes for the restriction digest. At day 1 EcoR1 enzyme and buffer was introduced to 9 500 ng of DNA for each sample. At day 2 and 3 buffer and enzyme were added again.

Day 1		
Vol samples [µl]		21.1
DNA amount samples [ng]	(~ 10 000 needed)	9 500.0 (PD 3314: 4600)
Vol EcoR1 (enzyme with 20,000 units/ml) [µl]	(<10% needed)	2.0
Vol EcoR1 (buffer) [µl]	(10X)	2.6
Day 2		
Vol EcoR1 (enzyme) [µl]		2.0
Vol EcoR1 (buffer) [µl]		2.5
Day3		
Vol EcoR1 (enzyme) [µl]		2.0
Vol EcoR1 (buffer) [µl]		2.5

Abbr.: Vol = volume.

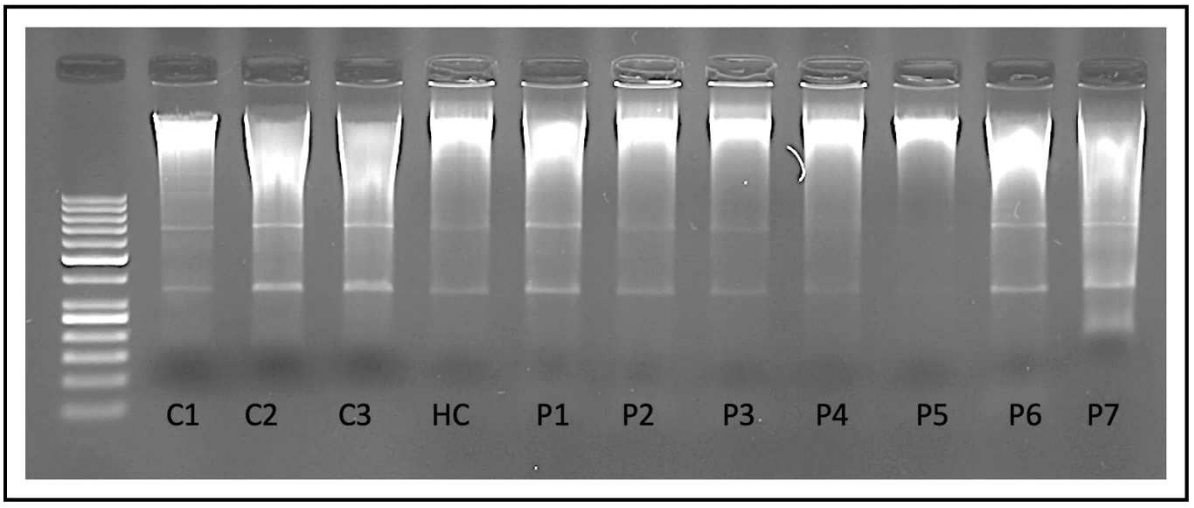

Figure 13. Gel Electrophoresis of Successful Restriction Digest.
Visual confirmation of a successful restriction digest of the samples CANVAS patient 1, 2 and 3, the healthy control as well as 7 PD patients carrying the MUTexp/MUTexp genotype (from left to right). A 50 bp DNA ladder was used.

3.9.2. Probe Generation

To detect the *RFC1* gene on the membrane subsequent to its transfer from the agarose gel, a *RFC1* specific probe was designed utilizing nested PCR methodology. Probe generation was undertaken as recommended in the supplementary section of the recently by (Cortese et al., 2019a) published paper "Biallelic expansion of an intronic repeat in *RFC1* is a common cause of late-onset ataxia".

The protocol involved one preliminary PCR to construct a template (for reaction volumes see table 17), a subsequent gel extraction of the desired band and a second PCR (for reaction volumes see table 18) in which the generated template was labeled with DIG. By employing DIG labeling, also referred to as digoxigenin labeling, the DNA probe could be marked in a non-radioactive way to specifically target the *RFC1* sequence. All PCRs necessary for probe generation were conducted at a final volume of 30 µl and performed under identical thermocycling conditions depicted in table 19.

3.9.2.1. DIG-PCR-Template-Generating PCR

The first step of the probe generation consisted of a PCR which was executed to construct a template for subsequent steps. As depicted in table 17, it included a standard PCR mixture, comprising a 10X buffer, 200 µM dNTPs, 0.25 µM each of the unlabeled forward and reverse primer (RFC1_DIG_F and RFC1_DIG_R, respectively), around 1.25 units of Taq Polymerase, and 10 ng of healthy *RFC1* WT (WT/WT) genomic DNA. The size of the amplicon was checked via gel electrophoresis with an expected size of 988 bp.

3.9.2.2. Gel Extraction

To avoid additional PCR products, the 988 bp large band corresponding to the "DIG-PCR-template-generating PCR" was extracted from the gel. The QUIAEX II System for purification of DNA fragments from gels and solutions was utilized. According to the supplied protocol the band was cut out and put into a 1.5 ml Eppendorf tube. Three volumes of the QXI buffer were added followed by 10 µl vortexed QUIAEXII. In the course of an incubation of 10 min at 50°C with vortexing steps every 2 minutes, the color should turn yellow. Subsequent to the removal of the supernatant, 500 µl of QXI was subjected to the pallet. After vortexing and a 30 sec centrifugation, the supernatant was removed again. Subsequently 500 µl of the PE buffer was added to the pallet, followed by vortexing, 30 sec centrifugation and removal of the supernatant. This step was completed twice. To air dry the pallet, the open Eppendorf tube was put on the heatblock at 37°C for one hour. After an elution step with 20 µl dH_2O, the concentration of the gel extracted DNA could be measured. The concentration of the gel extracted DIG-PCR-template-generating PCR amplicon was observed to be 29,45 ng/µl.

3.9.2.3. DIG-Labelled-Probe Generating PCR

The purified "DIG-PCR-template-generating PCR"-product was used as a template for the final nested PCR utilizing the "PCR DIG Probe Synthesis Kit". As depicted in table 18, the reaction mixture, with a volume of 20 µl, conducted an enzyme mix, a PCR DIG Probe Synthesis Mix (containing dATP, dCTP and dGTP), PCR Buffer with MgCl2, a dNTP Stock Solution (containing dATP, dCTP, dGTP and dTTP) the nested forward as well as reverse primer and the amplicon of

the previous ". DIG-PCR-Template-generating PCR". An additional control (CTR) reaction was carried out in which the nested PCR was applied as well albeit without the incorporation of DIG. The anticipated sizes of the resulting oligonucleotides were 900bp for the fragment emerging the unlabeled CTR reaction and a slightly larger fragment (between 900 and 1000 bp), due to the molecular weight of the DIG label, which serves as the final probe for *RFC1* detection. The size of the amplicons was assessed using gel electrophoresis which is illustrated in figure 13.

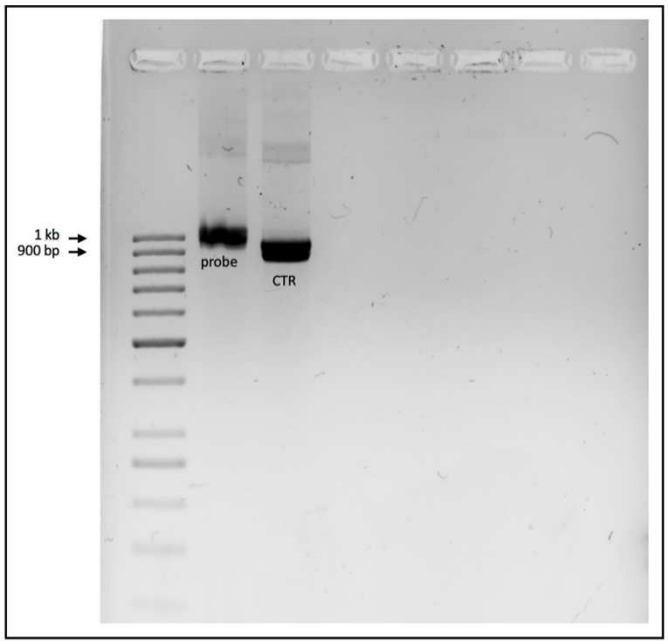

Figure 14. Gel Electrophoresis of final Southern Blot Probe.
Visual confirmation of the final Probe for Southern Blot analysis (left) as well as the unlabeled CTR fragment (right), with a size of approximately 950 and 900 bb, respectively. A 50 bp DNA ladder was used.

Table 17. DIG-PCR-Template-generating PCR Reagents.
PCR reagents for the generation of the template for the DIG labeling PCR. Reaction components and their associated stock concentration (first row), the final concentration needed for the reaction (second row) and the calculated volume for a 30 µl reaction (third row)

Reagent (stock)	Final concentration	Volume [µl]
Buffer (10X)	1X	3.00
dNTPs (10mM)	200 µM	0.60
RFC1_DIG_F (10mM)	0.25 µM	0.60
RFC1_DIG_R (10mM)	0.25 µM	0.60
Taq Pol	1.25 units/50 µl	0.21
DNA (10ng/µl)	60 ng	6.00
dH_2O		18.99

Table 18. DIG-Labelled-Probe-PCR Reagents.
PCR reagents for the generation of the DIG marked probe utilizing a nested PCR with the template emerged in the "DIG-PCR-Template-generating PCR". Reaction components and their associated stock concentration (first row), the final concentration needed for the reaction (second row), the calculated volume for a 30 µl reaction (third row) and a control (CTR) reaction resulting in a unlabeled amplicon (fourth row).

Reagent (stock)	Final concentration	Volume [µl]	Volume CTR [µl]
Buffer (10X)	1X	3.00	3.00
PCR DIG Probe Synthesis Mix (10mM)	450µM	1.50	
dNTPs (10mM)	450µM/900µM	1.50	3.00
RFC1_DIG_nested_F (10µM)	0.3 µM	1.00	1.00
RFC1_DIG_nested_R (10µM)	0.3 µM	1.00	1.00
Enzyme (3.5 U/µl)	1.6 units	0.45	0.45
RFC1_DIG amplicon (10ng/µl)	60 ng	6.00	6.00
dH_2O		15.50	18.99

Table 19. Thermocycling Conditions for the Probe Generating PCRs.
Thermocycling conditions for both probe generating PCRs ("DIG-PCR-Template-generating PCR" and "DIG-labelled-probe-PCR"). Cycle step (1 column), associated temperature (second row) and duration of the respective PCR part.

PCR step		Temperature [°C]	Duration
Initial denaturation		95.0	4 min
Denaturation	25x	95.0	30 sec
Annealing		58.4	30 sec
Extension		72.0	1 min
Final extension		72.0	5 min

Table 20. Primer Details.
Names, associated length and sequence of all primers utilized in conducted PCRs.

Primer	Length [bases]	Sequence (5' - 3')
RFC1_F	21	ACTGACAGTGTTTTTGCCTGT
RFC1_R	20	GGCTGAGGCAGGAGATCAC
FBN1_F	20	CAGGAAACAGCTATGACC
FBN1_R	23	GGCCATCTCTTCCTCTTCTTCTT
RFC1_F_fam	21	[6FAM]ACTGACAGTGTTTTTGCCTGT
RFC1_RP_F_fam	25	[6FAM]TCAAGTGATACTCCAGCTACACCGT
RFC1_RP_R	18	CAGGAAACAGCTATGACC
RFC1_RP_pat1	64	CAGGAAACAGCTATGACCAACAGAGCAAGACTCTGT TTCAAAAAGGGAAGGGAAGGGAAGGGAA

Primer	Length [bases]	Sequence (5' - 3')
RFC1_RP_pat2	63	CAGGAAACAGCTATGACCAACAGAGCAAGACTCTGT TTCAAAAGGGAAGGGAAGGGAAGGGAA
RFC1_RP_pat3	65	CAGGAAACAGCTATGACCAACAGAGCAAGACTCTGT TTCAAAAAAGGGAAGGGAAGGGAAGGGAA
RFC1_RP_WT1	65	CAGGAAACAGCTATGACCAACAGAGCAAGACTCTGT TTCAAAAAAGAAAAGAAAAGAAAAGAAAA
RFC1_RP_WT2	64	CAGGAAACAGCTATGACCAACAGAGCAAGACTCTGT TTCAAAAAGAAAAGAAAAGAAAAGAAAA
RFC1_RP_WT3	63	CAGGAAACAGCTATGACCAACAGAGCAAGACTCTGT TTCAAAAGAAAAGAAAAGAAAAGAAAA
RFC1_RP_F_UL	25	TCAAGTGATACTCCAGCTACACCGT
rs11096992 [VIC/FAM]	52	TTCTCCTGTTAATATTAGTGAATGG[A/G]TATAAATCC TGAGCATACTTGCTTA

3.9.3. Southern Blot Day 1: Gel Electrophoresis

The successful restriction digested DNA samples were loaded onto a 0.8 % agarose gel which was prepared without the addition of PeqGreen. For this, 29 µl of the DNA (10,000 ng) was mixed with 5.8 µl of 6X loading dye. As a molecular size marker MWII and MWVII (1000ng/µl stock) were employed in a 1:20 dilution (1.5 µl ladder in 18.5 µl dH_2O) resulting in 1 500ng of total DNA for each marker. A total of 30 µl of each DNA sample (containing 8 620.7 ng DNA respectively) and 10 µl of each DNA ladder (MWII and MW VII) were loaded. However, in the case of PD 3314 only 3,965.5 ng could be loaded due to the lack of DNA or blood specimen. It was avoided to exceed a loading volume of 30 µl to not overflow the gel chambers. Details for exact loaded volumes and corresponding weights are shown in table 21. To ensure an even electrophoretic migration of the samples, ladders were loaded on the sidelines. The electric gradient was applied for a duration of 12h (overnight) at a voltage of 60.

Table 21. Volumes and Weights Loaded onto Agarose Gel for Southern Blot Analysis.
Volume with the corresponding final dilution of the DNA (row one), volume of added dH_2O (row two), volume of added 6X loading dye (row three), the final volume loaded onto the gel (row four) and the corresponding total DNA weight for the PD and HC samples (row five), as well as the employed DNA ladders.

Reagent	Sample DNA	MWII and MWVII
Volume DNA [µl] (dilution)	29 (1:1.2)	1.5 (1:20)
Volume dH_2O [µl]		28.5
Volume Loading dye 6X [µl]	5.8	6
Volume loaded onto Gel [µl]	30	10
Final DNA weight [ng]	8 620.7 (PD 3314: 3 965,6)	1 500

Abbr.: MW = molecular weight marker

3.9.4. Southern Blot Day 2: Denaturation and Blotting

To effectuate the removal of purine bases, the gel was washed in 200 ml depurination buffer (0.25M HCl, pH<5) for 10 min. All washing or incubation steps at RT were performed on a shaking incubator. After a quick rinse with dH_2O, the gel was incubated in 200 ml denaturation buffer (0.5 NaOH, 1.5M NaCl, pH>5) twice for 15 min to separate the DNA double helix. After an additional washing step in H_2O , the gel underwent a neutralizing step in 200 ml neutralizing buffer, consisting of 1M Tris-HCl and 2M NaCl for the duration of twice 15 min. HCl (5N) and NaOH (1N) were utilized for pH adjustment to achieve a final pH of 7.5. The pH values of all buffers were analyzed via pH indicator strips and a standard pH meter. Details on buffer components are given in table 22. Following a 10 min long equilibration in 10X SSC, the blot was assembled.

To build the blotting sandwich for subsequent "bottom-up" transfer to a membrane, absorbent paper (Whatman Gel Blot Paper) was stacked to a height of 9cm to build a base. Four dry sheets of Whatman Chromatography Paper, followed by 2 sheets of prewet (10 min in dH_2O followed by 10 min in 10X SSC) paper of the same type, were put on top. The next two layers constituted of a prewet positively charged nylon membrane and the prepared agarose gel (open wells face upwards). Excess, no DNA harboring, gel was cut off. Area not covered by gel was sealed by parafilm to ensure the desired buffer transmission through the gel and the membrane. Bubbles were removed by gently rolling over the gel with a pipette. The gel was finally covered with 2 prewet Whatman filter paper as well as a long piece of prewet Whatman Chromatography Paper constituting a wick, and a panel to put extra weight on top. Both excess sides of the wick were in contact with 10X SSC to ascertain a constant liquid transfer through the gel onto the membrane and facilitate successful capillary blotting.

Table 22. Reagent Details for Depurination-, Denaturation and Neutralization Buffer.
Reagents and their stock molar mass (first row), the final malar mass in the dilution (second row) and the final volume required for a final dilution volume of 400ml.

Buffer	Reagents (stock M [g/mol])	Final M [g/mol]	Quantity
Depurination buffer			
	HCl (10.13)	0.25	10.00 ml
	Solvent: dH_2O		390.00 ml
Denaturation buffer			
	NaCl (58,44)	1.50	35.06 g
	NaOH (40)	0.50	8.00 g
	Solvent: dH_2O		Δ 400.00 ml
Neutralization buffer			
	Tris-HCl (121,14)	1.00	48.44 g
	NaCl (58,44)	2.00	41.94 g
	Solvent: dH_2O		Δ 400.00 ml

Abbr.: Δ = "difference to final volume of ", M = molar mass [g/mol]

3.9.5. Southern Blot Day 4: Hybridization

After successful transfer of the DNA onto the membrane, the blot sandwich was disassembled and the membrane was washed with 2X SSC and dried for 1 h. To permanently attach the transferred DNA to the membrane (UV-crosslink), the membrane was exposed to ultraviolet radiation (utilizing a UPV crosslinker) on both sides.

In all subsequent denaturation steps of salmon sperm and probe the DNA was exposed to 95°C for the duration of 5 min and then immediately put on ice to avoid the building of secondary structures of the nucleic acids. All incubation steps were undertaken at 50°C utilizing a hybridization incubator rotating the tubes.

To reduce non-specific binding of the probe to the membrane the membrane was initially prehybridized with 200 µl denaturated salmon sperm diluted in 25 ml prewarmed DIG easy Hyb for the duration of 1 ½ h. In the final hybridization step 2,600 ng of the predesigned *RFC1* specific probe as well as 50 µl salmon sperm, both denaturated, were diluted in 25 ml prewarmed DIG easy Hyb (see table 23), in which the membrane was subsequently allowed to incubate overnight.

Table 23. Components of Pre- and Hybridization Buffer.

Volumes of salmon sperm (first column) with its associated concentration, DIG easy Hyb (second column) and the RFC1 specific probe with its associated concentration (third column) for prehybridization and hybridization buffer.

Reagent (stock)	Prehybridization	Hybridization
Salmon sperm [µl] (10.0ng/µl)	200 .0 (in total 2 000 ng)	50.0 (in total 500 ng)
RFC1 specific probe [µl] (423ng/µl)		6.1 (in total 2600 ng)
DIG easy Hyb [ml]	25.0	25.0

3.9.6. Southern Blot Day 5: Washing and Detection

Washing of the membrane and detection of the probe hybridized DNA was performed following the recommended protocol supplied by the DIG Wash and Block Buffer Set. Details on component volumes for each solution for day 5 is given in table 24.

Initially the hybridized plot was washed with a “low stringency solution” (2x SSC, 0.1% SDS) four times for the duration of 5 minutes on a shaker at RT. Subsequently the blot was washed with a “high stringency solution” consisting of 0.1 x SSC and 0.1% SDS, three times for 15 minutes in the hybridization oven at 68°C. The following steps were all carried out on the shaker at RT. The membrane was washed in a washing solution consisting of 0.1M Maleic acid and 0.15M NaCl, twice for 2 minutes, followed by a blocking step in 5 % non-fat milk (twice for 15 min). The membrane was then incubated in blocking solution (blocking buffer in maleic acid solution, 1:10 diluted) for 30 min. To generate an antibody (AB) solution, anti-DIG-AP (DIG AB conjugated with alkaline phosphatase) was centrifuged at 10 000 rpm for 5 min at RT to avoid possible sediments in the solution and subsequently diluted in blocking solution (1:20,000) in which the membrane was incubated for 30 min (the AB was pipetted from the topmost layer of the tube). Following this step, the blot was again introduced to the washing solution three times for the duration of 10 min. Subsequently the membrane was incubated in a detection solution (0.1M Tis-Hcl, 0.1M

NaCl) for 3 minutes as well as in a CPD star solution, consisting of CDP star diluted in detection solution 1:10 without exposure to light. For final imaging the blot was transferred into a plastic envelope and all bubbles were carefully removed. Chemiluminescence was detected via “FUSION FX EDGE” using the “EVOLUTION-CAPT” software and the “Auto” settings for “Western stand”.

Table 24. Solutions for Southern Blot Day 5.
Solutions (column one), relevant reagents for the specific solution (column two), dilution of the reagents (column three) and the final volume of the respective reagent used for a certain final volume.

Solutions (final Volume)	**Reagents (stock)**	**Dilution**	**Quantity**
High Stringency (400 ml)			
	SSC (20X)	1:800	0.5 ml
	SDS (10%)	1:400	1.0 ml
	Solvent: dH_2O		Δ 400.0 ml
Low Stringency (400 ml)			
	SSC (20X)	1:40	10 .0ml
	SDS (10%)	1:400	1.0 ml
	Solvent: dH_2O		Δ 400.0 ml
Washing solution (1X) (500 ml)			
	Washing buffer (10X)	1:10	50.0 ml
	Solvent: dH_2O		Δ 500 .0ml
Maleic Acid Solution (1X) (500 ml)			
	Maleic acid buffer (10X)	1:10	50.0 ml
	Solvent: dH_2O		Δ 500.0 ml
Milk-Blocking Solution (200 ml)			
	Non-fat milk	1:20	5.0 g
	Solvent: Maleic Acid Solution (1X)		Δ 200.0 ml
Blocking Solution (1X) (200 ml)			
	Blocking Buffer (10X)	1:10	20.0 ml
	Solvent: Maleic Acid Solution (1X)		Δ 200.0 ml

Solutions (final Volume)	Reagents (stock)	Dilution	Quantity
Ab solution (40 ml)			
	Anti-DIG-AP Solvent: Blocking Solution (1X)	1:20 000	4.0 µl Δ 40.0 ml
Detection solution (1X) (40 ml)			
	Detection buffer (10X) Solvent: dH_2O		4.0 ml Δ 40.0ml
CDP star solution (20 ml)			
	CDP star Solvent: Detection solution (1X)	1:3	5.0 ml Δ 20.0 ml

Abbr.: Δ = "difference to final volume of"

3.9.7. Analysis of Repeat Length

The total repeat size (calculated in bp) was estimated though considering the bands of the samples in relation to the MWII as well as the MWVII DNA ladders. The repeat number calculation was undertaken following the subsequent equation:

$$\text{Repeat number} = \frac{(\text{expanded RFC1 size [bp]} - 5\,000)}{5}$$

where 5,000 corresponds to the estimation of the size of the *RFC1*-WT allele fragment after digest (5,037 bp) and 5 represents the length of the pentanucleotide motif "AAGGG" (Cortese et al., 2019a).

3.10. Statistical Analyses

Statistical analyses and plot generation were performed using R 4.1.0 (R Core Team, 2021). Fisher's exact test was utilized for 2x2 contingency analysis. Pairwise Fisher`s exact tests were performed for contingency tables exceeding a 2x2 format. Chi square test was not performable in this case due to the high frequency of the expected cell counts lower than 5 (Agresti, 2012). Bonferroni correction for multiple testing was performed to reduce the risk of an alpha error.
WTexp alleles were included into the allelic sum in the course of the analysis of haplotype frequencies, however their frequencies were not stated. As our protocol was not designed for the detection of heterozygous WTexp alleles, we couldn't evaluate the WTexp haplotype frequency.
To observe the directional correlation of two variables, linear regression was calculated.

Assumptions for parametric tests were observed utilizing Shapiro Wilk test for normal distribution of the model`s residuals, as well as Levene test to check for homoscedasticity.
As for AAO no normally distributed data was observed according to Shapiro Wilk test ($W = 0.99319$, $p < 0.001$), Kruskal Wallis rank sum test was applied to test for differences of AAO of PD patients with different genotypes. Analog the influence of the genotype on AAO was tested for the PSD subgroup since Shapiro Wilk test again revealed the absence of normally distributed residuals ($W = 0.9285$, $p < 0.01$).

4. Results

4.1. Characteristics of the Study Cohort

The whole study cohort consisted of 1204 patients diagnosed with PD or related diseases, with an average age at onset (AAO) of 58 years (*n* = 976, x̄ = 57.99, *M* = 59.00, *Min* = 18.00, Max = 87.00, *SD* = 11.61). In 1103 instances, sex data was collected, which consisted of 699 male and 404 female individuals with PD or associated diseased. Phenotypes include PD (*n* = 981), MSA (*n* = 18) and PSD (Parkinson Spectrum disorder, *n* = 80). The subgroup PSD compromises atypical Parkinson Syndrome, Lewy body dementia, (rapid onset) dystonia parkinsonism, essential tremor PD, middle cerebral artery syndrome, Perry syndrome, hereditary spastic paraplegia, progressive supranuclear palsy, as well as mixed forms. Out of the whole cohort, in 969 recorded cases, 566 have a negative, 345 a positive and 58 an unknown family history of PD, MSA or PSD. Levodopa response was recorded for 310 patients, where 269 showed an improvement and 29 consistence of symptoms.

4.2. RFC1 Genotyping in PD and Related Diseases

4.2.1. RFC1 Genotype Frequency in PD and Related Diseases

As the central part of our analyses, we tested for genotype differences between the cohorts of interest (the corresponding plot can be seen in figure 14). Biallelic genotypes as well as the heterozygous variants are specified in Table 4 (WT/WT, MUTexp/MUTexp, WTexp/WTexp as well as MUTexp/WT and MUTexp /WTexp, respectively).

We found that individuals with PD or associated diseases (including patients for whom no specific PD, MSA or PSD diagnosis was recorded) showed a nominally significantly higher occurrence of the biallelically expanded genotype MUTexp/MUTexp than the healthy control cohort (p =0.04, $p_{adj.}$ = 0.43, *OR* = 7.225). This indicates that we observed significant difference before applying bonferoni correction due to multiple testing.

Additionally, a nominal significant difference of the frequencies of the WT/WT and the WTexp/WTexp genotypes between the healthy controls and the whole PD cohort (including patients for whom no specific PD, MSA or PSD diagnosis was recorded) could be observed (p = 0.04, $p_{adj.}$ = 0.441, *OR* = 4.515).

In the next step we compared only PD patients (excluding the MSA and PSD cohort) to the HC group regarding their genotype proportions. We revealed a highly significant difference of the investigated RFC1 genotype frequencies between the PD patients and the HC cohort (p < 0.001, $p_{adj.}$ = 0.004). 1.34 % of the individuals suffering from PD harbor the MUTexp/MUTexp genotype compared to 0.17% in the HC cohort. While both cohorts show quite similar relative number of a heterozygous carriers (MUTexp/WT) (~8 %), there are more healthy individuals positive for a heterozygous, biallelic MUTexp/WTexp expansion than PD patients (1.86 %, respectively). Frequencies of different genotypes in all investigated groups can be seen in table 25.

No statistical significance in the genotype frequencies could be observed between the HC and MSA cohort (p = 1) and the HC and the PSD cohort (p = 0.441, $p_{adj.}$ = 1). In neither of these groups patients harboring the RFC1 variant MUTexp/MUTexp could be detected.

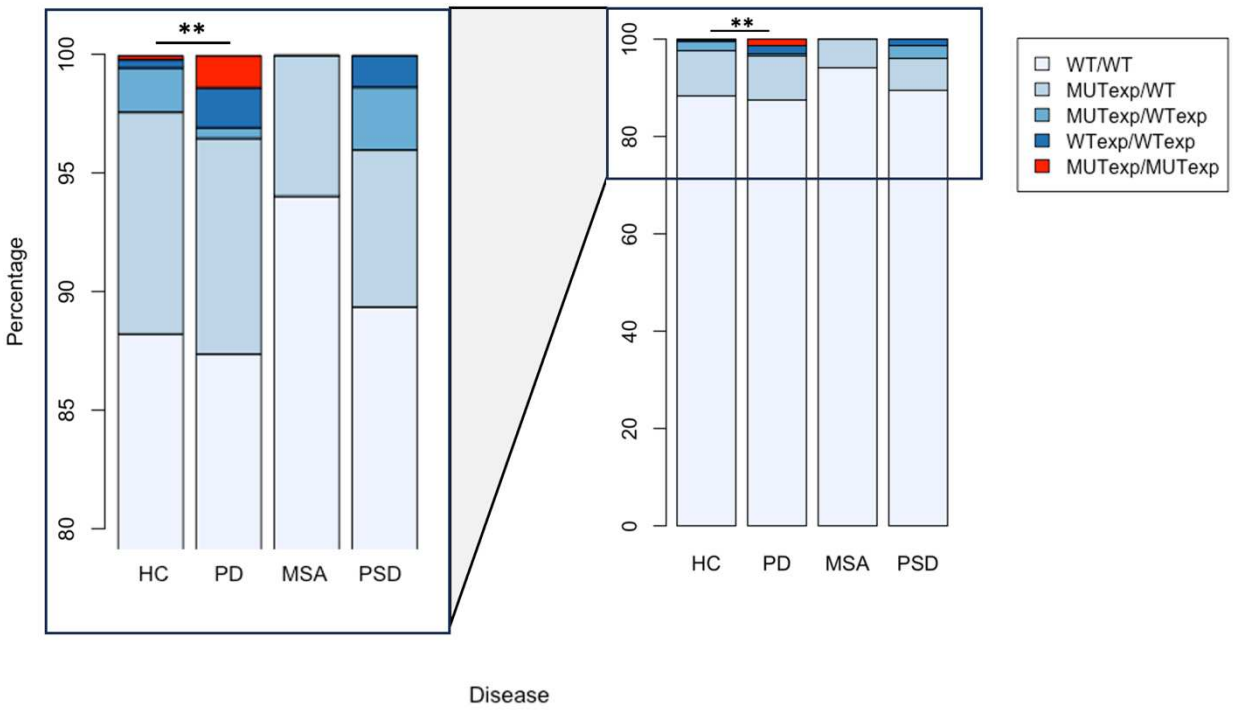

Figure 15. Influence of RFC1 Genotype on Development of Parkinson Related Diseases.
Stacked bar graph shows different genotype variants in different Parkinson related diseases. Right panel shows percentage 0 to 100 and left panel the zoomed in version ranging from 80 to 100 percent. Genotypes in PD and HC differ significant before applying Bonferroni correction ($p < 0.001$, $p_{adj.} = 0.008$). HC and PD subgroups differ significantly in their MUTexp/MUTexp as well as their WTexp/WTexp genotype frequencies ($p = 0.04$, $p_{adj.} = 0.43$ and $p = 0.04$, $p_{adj.} = 0.442$, respectively). No significant differences were found comparing other phenotypes.
Abbr.: HC = healthy controls, PD = Parkinson's Disease, MSA =Multisystemic Dystrophy, PSD = Parkinson spectrum Disease.

Table 25. RFC1 Genotypes in Parkinson Related Diseases – Frequency Table.
Absolute counts as well as the corresponding percentage of RFC1 genotype variants in the HC controls, the PD, MSA and PSD subgroups as well as the sum of all pathogenic subgroups. Genotype frequencies in PD and HC differ nominally significantly ($p < 0.001$, $p_{adj.} = 0.008$) as indicated by the three asterisks in the table). The HC and PD subgroups exhibit significant differences in their MUTexp/MUTexp (se dark blue marked fields) and WTexp/WTexp genotype frequencies (see light blue marked fields) ($p = 0.04$, $p_{adj.} = 0.43$ and $p = 0.04$, $p_{adj.} = 0.442$, respectively). No significant differences were observed when comparing other phenotypes.

	Frequencies [abs.] ([%])				
	***				
Genotypes	**HC**	**PD**	**MSA**	**PSD**	**sum**
WT/WT	523 (88.34)	784 (87.50)	16 (94.12)	68 (89.47)	868 (87.77)
MUTexp/WT	55 (9.29)	81 (9.04)	1 (5.88)	5 (6.58)	87 (8.8)
MUTexp /WTexp	11 (1.86)	4 (0.45)	0 (0.00)	2 (2.63)	6 (0.61)
WTexp/WTexp	**2 (0.34)**	15 (1.67)	0 (0.00)	1 (1.32)	**16 (1.62)** *
MUTexp/MUTexp	**1 (0.17)**	12 (1.34)	0 (0.00)	0 (0.00)	**12 (1.21)** *
Sum	592 (100)	896 (100)	17 (100)	76 (100)	989 (100)

Abbr.: abs. = absolute counts, HC = healthy controls, PD = Parkinson's Disease, MSA =Multisystemic Dystrophy, PSD = Parkinson spectrum Disease.

4.2.2. Allelic Frequency of RFC1 Variants in PD and Related Diseases

Since we were also interested in the carrier frequency of the allele with the *RFC1* MUTexp variant, we additionally investigated the allelic frequencies of the two main *RFC1* variants (MUTexp and WT) in the different cohorts. There was no significant difference between the PD patients and the HC group regarding the MUTexp allele. In patients with PD the MUTexp allele could be observed with a frequency of 6.08 %, whereas the WT cohort, exhibited a frequency of 5.74 %. Allelic frequencies of interest in all tested cohorts can be seen in table 26.

Table 26. Allelic Variants of RFC1 in Parkinson Related Diseases – Frequency Table.
Absolute counts as well as the corresponding percentage of RFC1 allelic variants.

	Frequencies [abs.] ([%])			
Allelic Variant	**HC**	**PD**	**MSA**	**PSD**
WT	1101 (92.99)	1649 (92.02)	33 (97.06)	141 (92.76)
MUTexp	68 (5.74)	109 (6.08)	1 (2.94)	7 (4.61)
Sum	1184 (100)	1792 (100)	34 (100)	152 (100)

Abbr.: abs. = absolute counts, HC = healthy controls, PD = Parkinson's Disease, MSA =Multisystemic Dystrophy, PSD = Parkinson spectrum Disease.

4.2.1 rs11096992 Genotyping

As mentioned in the introduction, previous studies suggest several risk haplotypes in RFC1, primarily found on alleles carrying the MUTexp REM. To investigate the frequency distribution of the risk haplotype, we used qPCR to detect the SNP rs11096992. The study cohort consisted

of 49 individuals, including 17 diagnosed with Parkinson's disease (PD). Among these, six PD patients were positive for the biallelic MUTexp/MUTexp genotype, 26 individuals had the WT/WT genotype, and 17 (including eleven PD patients) had the MUTexp/WT genotype. Our analysis showed that both the MUTexp/MUTexp genotype and the MUTexp/WT carriers had a significantly higher association with the SNP rs11096992 compared to those with the WT/WT genotype ($p < 0.0001$ for both comparisons). The frequencies of the AA and AG genotypes in each group are illustrated as a grouped barplot in Figure 15. All six individuals with the biallelic MUTexp/MUTexp genotype had the recessive "AA" risk allele. Additionally, 94.12% of the monoallelic expansion carriers (MUTexp/WT) tested positive for the SNP on both alleles, with only one individual in this group lacking the variant. In the 26 individuals with the WT/WT genotype, only 7.69% had the risk haplotype, while it was absent in 26.92%. The remaining 65.38% were heterozygous for the SNP. Table 27 provides the exact frequencies of the homozygous AA and heterozygous AG genotypes across the different RFC1 variants (WT/WT, MUTexp/WT, and MUTexp/MUTexp).

Table 27. Risk Haplotype in RFC1 Repeat Expansion Mutations.
Frequencies of the risk haplotypes rs11096992 as well as its heterozygous appearance in the RFC1 REM variants: WT/WT, MUTexp/WT and MUTexp/MUTexp. Significantly greater numbers of the risk haplotype in expanded genotypes could be observed compared to the WT/WT group ($p < 0.0001$, $p < 0.0001$) as indicated by the blue asterisks.

	Frequencies [abs.] ([%])		
	WT/WT	MUTexp/WT	MUTexp/MUTexp
Heterozygous AG	17 (65.38)	0 (0.00)	0 (0.00)
Homozygous AA	2 (7.69)	16 (94.12) ***	6 (100) ***
sum	26 (100)	17 (100)	6 (100)

Abbr.: abs. = absolute values

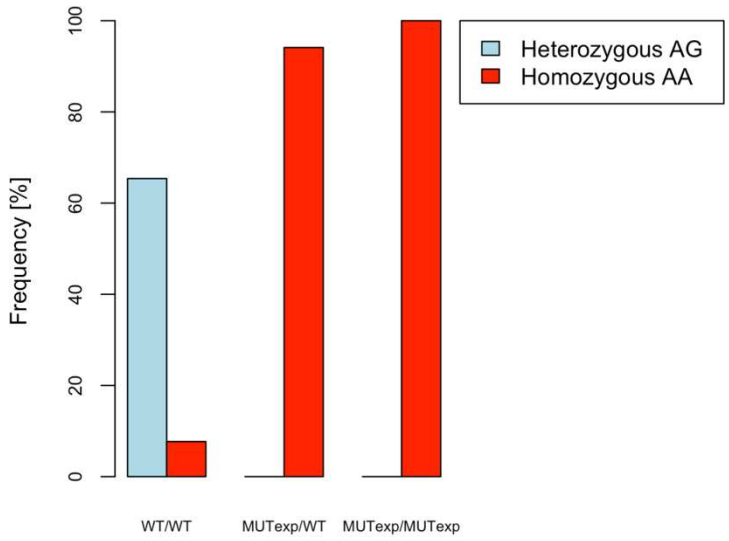

Figure 16. Risk Haplotype in RFC1 Repeat Expansion Mutation Variants.
The figure above shows the frequency of the SNP rs11096992 as well as its homozygous form in the 3 RFC1 Repeat Expansion Mutations groups: WT/WT, MUTexp/WT, MUTexp/MUTexp. The variants carrying an expansion revealed significantly greater numbers of the risk haplotype than the WT/WT group ($p < 0.0001$, $p < 0.0001$).

4.3. Evaluation of the RFC1 Pentanucleotide Repeat Expansion Length

As the RP-PCR method only allows a rough statement about the presence of a pathologically expanded *RFC1* allele of more than 60 repeats, we utilized southern blotting to more accurately assess the size of *RFC1 expansion* on each allele. As explained in the methods section only samples which were positively tested for the risk genotype (MUTexp/MUTexp) underwent this analysis step in order to be able to better mark their approximate expansion size. Due to the high amount of DNA needed for Southern blotting, only six out of 12 positively MUTexp/MUTexp tested patients, could be included in this analysis.

Interestingly, we detected smaller sized expansions in all of our PD samples compared to what is known in CANVAS. The sizes in PD patients ranged from 70 repeats to 660 repeats. The PD patients usually had one smaller and one larger allele. This is also in contrast to the as positive controls used CANVAS samples which were homozygous for large expansions, i.e. showed large expansions of similar size on both alleles.

For clarity reasons, the smaller *RFC1* allele in PD patients will be referred to as allele one and the large one as allele two. The first alleles carried a number of expanded repeats ranging from 70 to 160, which makes 5350 to 5800 bp ($\bar{x}$ = 105 repeats, *M* = 100 repeats, *SD* = 34.5 repeats, *Min* = 70 repeats, *Max* = 160 repeats, table 28, figure 17) which was on average slightly larger than the 11 repeats long (5000 bp) WT allele. For all second alleles larger expansions in the range of 380 to 660 repeats which corresponds to the size of 6900 to 8300 bp could be observed ($\tilde{x}$ = 550 repeats, *M* = 600 repeats, SD =112.96 repeats, *Min* = 380 repeats, *Max* = 660 repeats, table 28, figure 17). In the Southern blot shown below, one out of three CANVAS patient samples displayed a band with a size of approximately 9000 bp for both alleles (as indicated by the blue arrow). However,

in previous Southern blot analysis, bands corresponding to the *RFC1* expansion could be shown also for the other two CANVAS patients (see supplementary material). The interval for pathological repat size for *RFC1* ranges from 6000 to 12,000 bp. Nevertheless, also the more expanded allele two didn't reach the size of *RFC1* expansions reported for CANVAS.

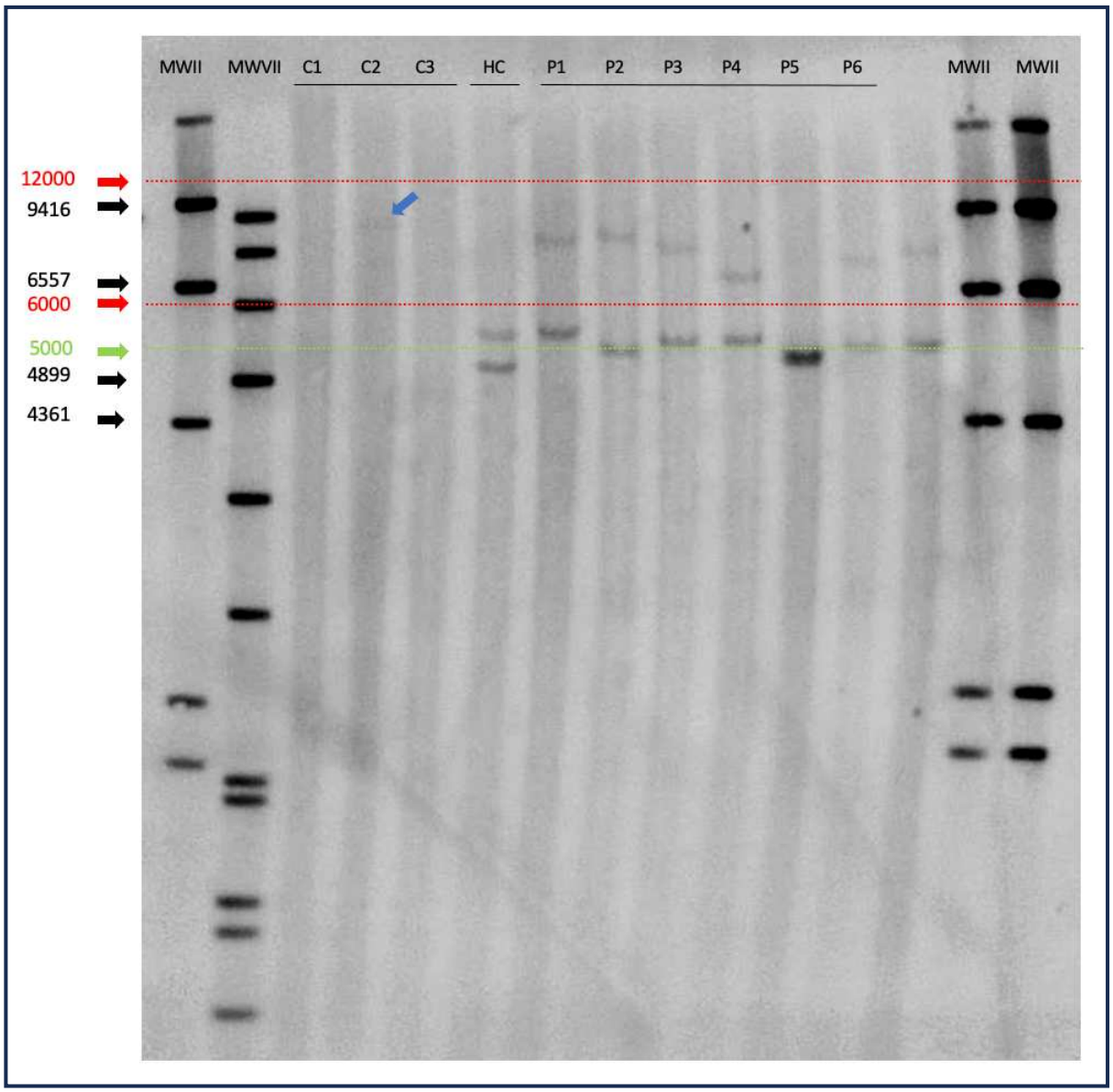

Figure 17. RFC1 Southern Blot. *The figure shows a RFC1 sensitive southern blot utilizing three CANVAS samples as positive controls (C1, C2, C3), one HC and six PD patients carrying the risk genotype MUTexp/MUTexp (P1-P6). Molecular weight ladders were used in different concentrations (MWII, MWII). The sample located between P6 and MWII on the right side of the southern blot was removed from the analyses subsequently. Black arrows on the left side indicate the respective sizes in base pairs according to the used ladders. The green arrow together with the green dotted line indicates the approximate size of WT RFC1. Red arrows together with the two red lines display the interval of pathological expansion sizes reported for CANVAS. The blue arrow marks the homozygous bands observed for the CANVAS sample C2.*

Table 28. RFC1 Repeat Sizes of Patients Carrying the Risk Genotype.
Approximate sizes for RFC1 on both alleles given in base pairs and repeat number, for all patients carrying the MUTexp/MUTexp genotype.

	RFC1 Allele 1		RFC1 Allele 2	
ID	**Size [bp]**	**Size [rn]**	**Size [bp]**	**Size [rn]**
1	5800	160	8100	620
2	5400	80	8300	660
3	5600	120	8100	620
4	5600	120	6900	380
5	5350	70	7200	440
6	5400	80	7900	580

Abbr.: bp = base pairs, rn = repeat number, ID = identification number.

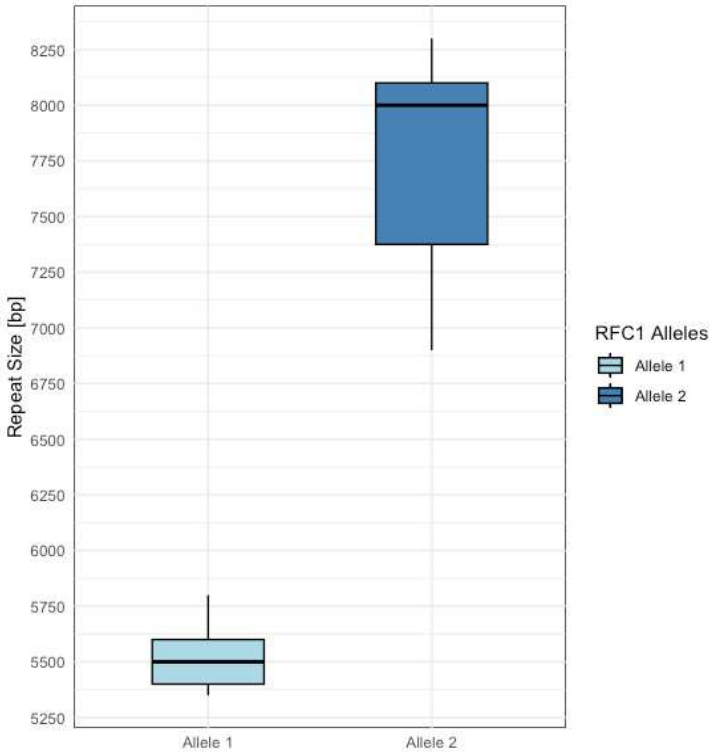

Figure 18. RFC1 Repeat Sizes of Risk Genotypes - Allelic Comparison.
The boxplot shows the respective size of RFC1 on allele one (left) and allele two right), containing only data of patients carrying the MUTexp/MUTexp risk genotype.

4.4. Influence of RFC1 Genotype on Age at Onset of PD and PSD

As previous studies suggested that there is a correlation between the presence of a REM and the AAO in various neurodegenerative diseases (see "Introduction" chapter), we tested if individuals carrying homo- or heterozygous *RFC1* mutations showed a different AAO compared to the non-expanded *RFC1* variant (WT/WT). Only patients, for which age at onset data was available, were included into the following analyses.

Within the PD subgroup none of the genotypes showed significant differences in their AAO ($\chi 2_{\ 4}$= 0.805, p = 0.9378, $p_{adj.}$= 1, figure 18). Carriers of the healthy WT/WT genotype exhibited a median AAO of 58 years (x̄ = 57.47 y, *SD* = 11.87 y, range = 18.00 y – 87.00 y), heterozygous patients with MUTexp/WT genotype had a median AAO of 60 years (x̄ = 58.39 y, *SD* = 11.89 y, *range* = 28.00 - 83.00 y) and MUTexp/WTexp carriers showed onset at 58 years (x̄ = 59.25 y, *SD* = 8.02 y, *range* = 48.00 - 67.00 y). PD patients homozygous for the WTexp allele (WTexp/WTexp) showed a median AAO of 58 years (x̄ = 59.07 y, *SD* = 10.76 y, *range*= 40.00 y – 77.00 y). Finally, carriers of the pathogenic MUTexp/MUTexp genotype showed a median AAO of 59 years (x̄ = 58.75 y, *SD* = 7.61 y, *range*= 44.00 y – 67.00 y).

Analog the influence of the genotype on AAO was tested for the PSD subgroup. Again, AAO did not differ significantly between genotypes ($\chi 2_{\ 3}$= 1.484, p = 0.69, $p_{adj.}$ = 1, figure 18). Carriers of the healthy WT/WT genotype showed a median AAO of 63 years (x̄ = 60.24 y, *SD* = 12.53 y, range = 20.00 y – 81.00 y) and patients showing the heterozygous MUTexp/WT genotype had a median AAO of 58 years (x̄ = 55.75 y, *SD* = 9.54 y, *range*= 43.00 y - 64.00 y). All descriptive data of AAO in different genotypes in the different sub-cohorts (PD and PSD) is summarized in table 29.

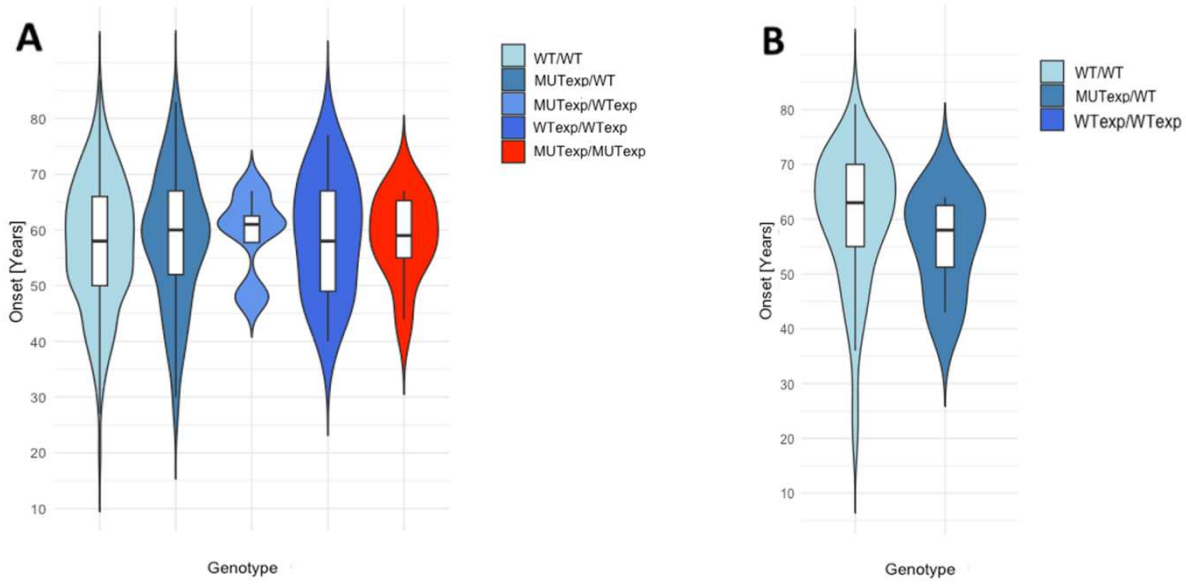

Figure 19. Influence of RFC1 Genotype on AAO.

The violin plots show AAO of different RFC1 genotypes of all patients with PD (panel A) and PSD (panel B). Each violin plot corresponds to a single genotype. The white square in the middle of each plot signals the inter-quartile range with a black line in the middle area marking the median AAO. The overall span of each violin plot corresponds to the minimum and maximum AAO values. AAO is given in years. No significant differences were observed between the genotypes in both disease groups. Abbr.: AAO = age at onset, PD = Parkinson's Disease, PSD = Parkinson Spectrum Disease.

Table 29. Age At Onset for different RFC1 Genotypes in PD and PSD.
The table shows descriptive statistics of AAO of different RFC1 genotypes of all patients with PD and PSD. No significant differences were observed between the genotypes in both disease groups.

	WT/WT		MUTexp/WT		MUTexp /WTexp	WTexp/ WTexp	MUTexp/ MUTexp
	PD	**PSD**	**PD**	**PSD**	**PD**	**PD**	**PD**
AAO ($\bar{x}$) [y]	57.46	60.24	58.39	55.75	59.25	59.07	58.75
AAO (*SD*) [y]	11.87	12.53	11.89	9.54	8.02	10.76	7.61
AAO (*M*) [y]	58.00	63.00	60.00	58.00	58.00	58.00	59.00
AAO (*IQR*) [y]	50.00 - 66.00	55.00 - 70.00	52.00 - 67.00	51.25 - 62.50	57.75 - 62.50	49.00 - 67.00	55.00 - 65.25
AAO (*Min*) [y]	18.00	20.00	28.00	43.00	48.00	40.00	44.00
AAO (*Max*) [y]	87.00	81.00	83.00	64.00	67.00	77.00	67.00

Abbr.: AAO = age at onset, PD = Parkinson's Disease, PSD = Parkinson Spectrum Disease, $\bar{x}$ = mean, SD = standard deviation, M = median, IQR = inter-quartile range, Min = minimum, Max = maximum.

In the second step of this analysis, the cohort was divided into subgroups characterized by their AAO. The resulting groups were named early onset Parkinson's Disease (EOPD) and late onset Parkinson's Disease (LOPD) based on the recorded AAO below or above the threshold of 50 years, respectively. We could not show a significant effect of genotype on the AAO subgroups (p = 0.691, *OR* = 2.059, figure 19).

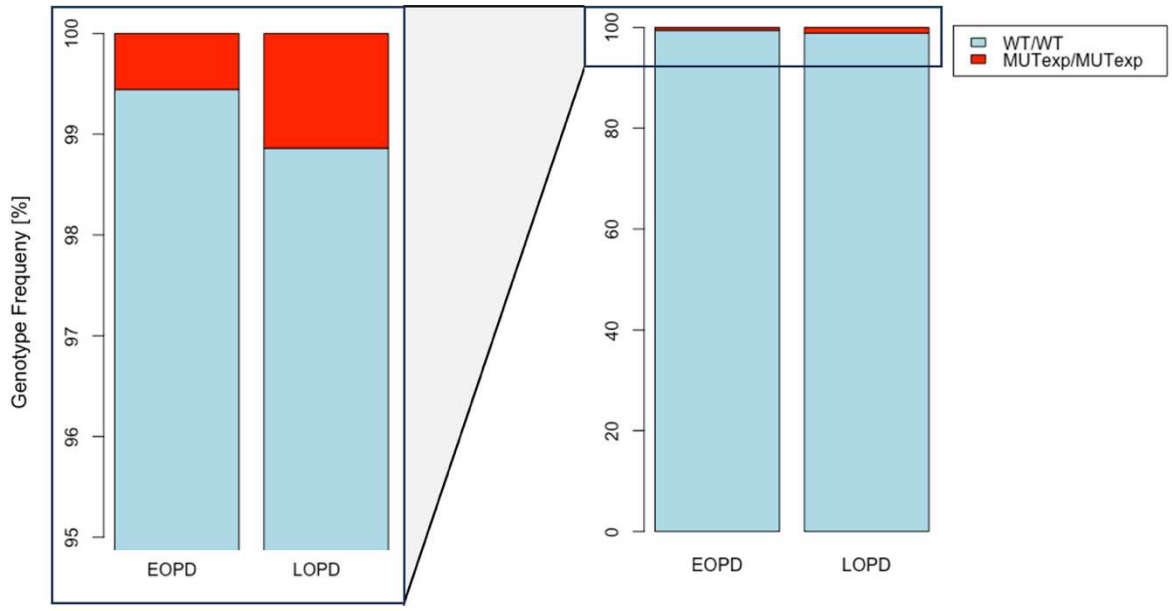

Figure 20. Influence of RFC1 Genotype on EOPD and LODP.
Stacked bar graphs visualize the influence of the RFC1 WT/WT and the MUTexp/MUTexp genotype on the development of either EOPD or LOPD. No significant difference according to p value was observed (p = 0.6907). Odds ratio however suggest accepting the alternative hypothesis (OR = 2.059254). Abbr.: abs. EOPD = early onset Parkinson's Disease, LOPD = late onset Parkinson's Disease.

Table 30 presents the summarized count data for these variables. Within the EOPD group 99.44 % patients were carrying the WT/WT and 0.56 % the MUTexp/MUTexp *RFC1* variant. By contrast, 1.14 % of patients who had a late onset of PD were possessed the MUTexp/MUTexp and 98.86 % the WT/WT genotype.

Table 30. EOPD- and LOPD in RFC1 Genotypes – Frequency Table.
Contingency table of absolute and percentage of EOPD as well as LOPD individuals carrying either the WT/WT or the MUTexp/MUTexp genotype. No significant difference according to p value was observed (p = 0.6907, OR = 2.059254

	Onset Group Frequencies [abs.] ([%])	
Genotype	**EOPD**	**LOPD**
WT/WT	179 (99.44)	608 (98.86)
MUTexp/MUTexp	1 (0.56)	7 (1.14)
Sum	180 (100)	615 (100)

Abbr.: abs. = absolute counts, EOPD = early onset Parkinson's Disease, LOPD = late onset Parkinson's Disease

4.5. Influence of RFC1 Repeat Size on Age at Onset of PD

As already explained in the "Evaluation of the RFC1 Pentanucleotide Repeat Expansion Length" chapter, we found different expansion sizes on both alleles of the MUTexp/MUTexp samples. Again, the allele carrying the larger expansion will be referred to as the "second allele" for clarity reasons.

Previous studies on other repeat expansion diseases, such as Huntington's disease (HD), have frequently observed a negative correlation between the size of the repeat expansions on the larger expanded "second" allele and the AAO (i.e., larger expansions lead to an earlier age at onset) (Takeuchi & Nagai, 2017). We therefore aimed to investigate this correlation in our cohort as well.

To assess the possible directional correlation of repeat size of the expanded second allele and AAO of the MUTexp/MUTexp samples, a linear regression was applied. The length of the repeat size of RFC1 on allele two didn't approach a significant effect on the AAO as indicated by the obtained p-value (F = 10.9669, p = 0.125, figure 20). However, 35.5 % of the variance in AAO could be explained by the size of *RFC1* (on the second allele) in our regression model, as indicated by the correlation coefficient (R^2 = 0.3548).

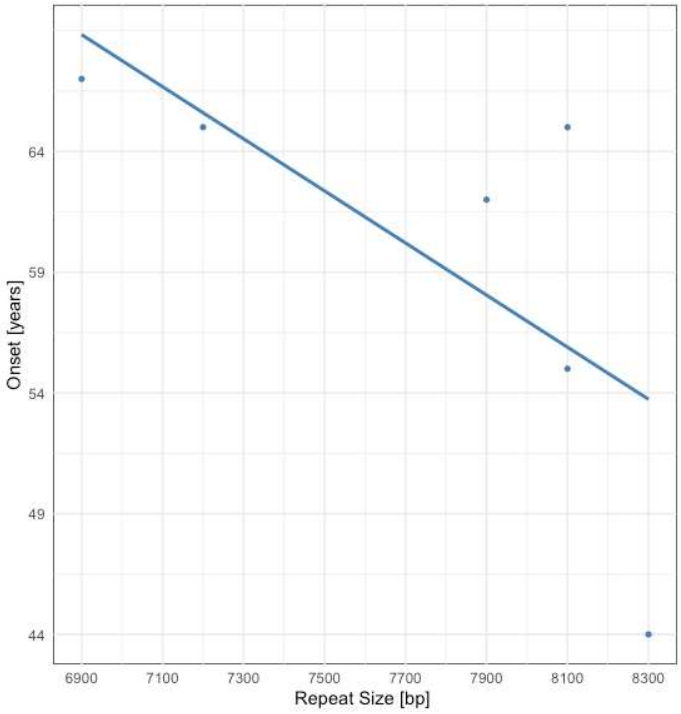

Figure 21. Effect of the Repeat Size on Further Expanded Allele on AAO.
The scatter plot with regression line shows the influence of the size of the RFC1 repeat expansion mutation on the further expanded second allele on the AAO. No significance was obtained ($p = 0.12489, R^2 = 0.3548$).
Abbr.: AAO = age at onset, bp = base pairs.

4.6. Influence of PD Family Anamnesis on RFC1 Genotype

An additional analysis was conducted regarding the influence of family history on the *RFC1* repeat variants, as summarized in table 31. Patients harboring the WT/WT genotype of *RFC1* reveal at 39.17 % a positive and at 60.83 % a negative family anamnesis, whereas 33.33 % of patients displaying the MUTexp/MUTexp genotype have a positive and 66.67 % a negative history of PD and related diseases in their close relatives.
We could not detect a significant difference of the WT/WT and the MUTexp/MUTexp genotype frequencies, based on the family history of the patients ($p = 1$, OR = 1.27, figure 21).

Table 31. Family Anamnesis in RFC1 Genotypes – Frequency Table.
The contingency table shows the absolute counts as well as the percentage of the WT/WT and MUTexp/MUTexp RFC1 genotypes of either positive or negative family anamnesis. No significant difference was found (p = 1, OR 1.273953).

	Genotypes [abs.] ([%])	
F.A.	**WT/WT**	**MUTexp/MUTexp**
positive	281 (38.92)	3 (33.33)
negative	441 (61.08)	6 (66.67)
Sum	725 (100.00)	9 (100.00)

Abbr.: abs. = absolute counts, F.A. = family anamnesis

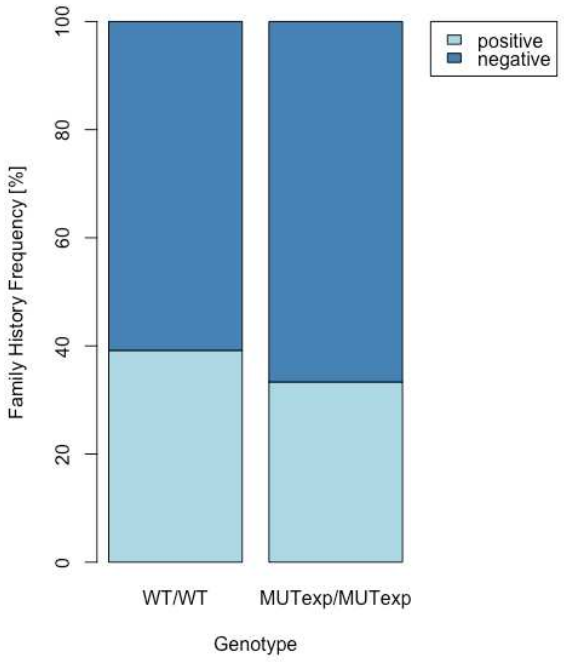

Figure 22. Influence of PD Family History on RFC1 Genotype.
The stacked bar graph shows the association of a positive or negative family anamnesis with either the WT/WT or MUTexp/MUTexp genotype of RFC1. No significant difference was found (p = 1, OR 1.273953).

4.7. RFC1 Pentanucleotide Repeat Expansion Carriers

To get insight into the clinical appearance of the observed MUTexp/MUTexp carriers (n = 8), demographic as well as phenotypical data is summarized in table 33 (motoric and nonmotoric symptoms of these patients are depicted in table 34 in the supplements).

Remarkably, for seven out of the eight PD patients harboring the biallelic MUTexp/MUTexp genotype no clear levodopa response was recorded despite levodopa responsiveness being a crucial diagnostic criterion for PD. More specifically six patients didn`t show levodopa response, two didn`t have a clear record regarding levodopa and one showed improvement of motoric symptoms. Interestingly, we did not observe any CANVAS like phenotypes in our biallelic MUTexp/MUTexp PD patients. For neither of them cerebellar ataxia, vestibular areflexia or chronic cough has been recorded. The most frequent observed motoric symptoms were tremor and rigidity (n=6), followed by bradykinesia (n=5). Sleep disorders were the most frequently noted non-motor symptoms (n=4), trailed by dementia, cognitive impairment, and hallucinations (n=3).

All these phenotypic characteristics of our MUTexp/MUTexp PD patients were quite similar to those observed in typical Parkinson's disease cases, however, the poor response to levodopa is highly atypical.

Table 32. Phenotyping of Biallelic Expanded Mutated RFC1 Motif Carriers Part 1.
The table summarizes important clinical as well as demographic features of patients carrying the MUTexp/MUTexp genotype. 1 marks the presence, whereas 0 the absence of a specific trait.

ID	Sex	Diagnosis	AAO [y]	F.A.	L-Dopa-Response	PD rel. Mut.
PD24	m	Mb. Parkinson (equivalent type)	55	N/A	0	N/A
PD76	m	Mb. Parkinson	44	N/A	0	N/A
PD88	m	Mb. Parkinson (akinetic-rigid type)	65	N/A	0	N/A
PD 125	w	Mb. Parkinson (equivalent type)	67	N/A	0	ATP13A2:P1173L (heterozygous)
PD164	m	Mb. Parkinson (equivalent type)	66	0	0	N/A
PD523	w	Parkinson Syndrome	65	N/A	N/A	N/A
PD526	m	Parkinson Syndrome	59	N/A	N/A	N/A
PD2052	m	Mb. Parkinson (equivalent type)	55	1	1	LRRK2: E943K

Abbr.: ID = identification number, Mb. = Morbus, AAO = age at onset, y = years, F.A. = family anamnesis, L-Dopa = Levodopa, PD rel. Mut. = Parkinson's Disease relevant mutations, Diff. = difficulties, Imp. = impairment

5. Discussion

Due to its variable nature, Parkinson`s disease (PD) presents a spectrum of phenotypes, which may show an overlap with other neurological diseases. In the present study, we found that biallelic repeat expansions in *RFC1* are associated with Parkinson´s disease. However, the repeat expansions, although clearly in the pathogenic range seem to be much smaller than in CANVAS disease. As the application of definite biomarkers is mostly lacking in general clinical context, diagnostic is primary based on evaluation of patients' clinical presentation. Furthermore, as outlined in the introduction, most PD cases are considered to be sporadic and only a minority can be explained by highly penetrant genetic mutations. The symptom variability as well as the overlap with other conditions may lead to a misdiagnosis of PD, which can result in inappropriate treatment. Therefore, the identification of a REM in RFC1 as an additional risk factor for PD is a big step towards facilitating proper diagnosis and ensuring patients receive optimal treatment.

5.1. RFC1 Genotypes in PD and Related Diseases

In the main part of this study, we analyzed the frequency of different *RFC1* repeat expansion genotypes in a cohort of Austrian patients with PD and PD-related diseases (MSA and PSD) in comparison to neurologically healthy individuals. Previous studies have already indicated an association of *RFC1* expansions and a Parkinson phenotype. According to these publications, the percentage of PD patients carrying the biallelic mutated *RFC1* expansion (MUTexp/MUTexp) ranges from 0.43 to 1.9, with more northern populations tending to show larger prevalences (Alvarez Jerez et al., 2024; Kytövuori et al., 2022; Ylikotila et al., 2023a). With our observed fraction of 1.36 % of Austrian PD patients carrying this genotype of interest, our numbers seem to lay within this interval.

Furthermore, the PD cohort differed significantly from the healthy control cohort regarding their genotype frequencies. This observation is as well in concordance with previous studies such as published articles by Kytövuori et al and Alvarez Jerez et al. which have already suggested a potential influence of REMs on the pathogenesis of PD in the Finish and non-Finish European population, respectively (Alvarez Jerez et al., 2024; Kytövuori et al., 2022).

To get insight into which genotype drives this difference, we included subanalyses. Our *RFC1* genotype of main interest MUTexp/MUTexp indeed showed a significant association with a PD phenotype. As already indicated by Davies et al., Dominik et al. and Tyagi et al., not only the expansion size, but also the sequence of the repeat (i.e. from the WT altering motifs) plays an important role in the resulting pathogenicity of the disease (Davies et al., 2022a; Dominik et al., 2023; Tyagi et al., 2023).

It is generally assumed that the homozygous expansion of the WT allele does not result in a pathogenic phenotype (Cortese et al., 2019a; Davies et al., 2022a; Tyagi et al., 2023). Interestingly, our observations seemed to be contradictory, as 1.67 % of the PD cohort showed the biallelic form of an WT expansion (WTexp/WTexp) whereas only 0.34 % of the healthy cohort did. This finding indicates a clear positive correlation of the WTexp/WTexp genotype with PD.

Our data implied that not monoallelic but only biallelic *RFC1* REMs appear to result in a pathogenic phenotype. Since the monoallelic expansions (MUTexp/WT) depicted higher

frequency in the healthy cohort, no pathogenicity is concluded here, as already hypothesized due to the expected recessive inheritance pattern of *RFC1* (Cortese et al., 2019c). Although there exist several REM diseases which exhibit a dominant pattern of inheritance, as for example HD and some forms of spinocerebellar ataxia, *RFC1* seems to be only correlating with an increased risk of PD in the case of a biallelic expansion.

5.2. RFC1 Expansion Haplotype Carrier Frequency in the Austrian Population

Previously conducted studies regarding *RFC1* expansions, reported *RFC1* MUTexp heterozygous allelic variant carrier frequencies of 0.7% to 6.5% in the control cohort. Whereas Rafehi et al. suggested an occurrence of 4.1 to 6.5 %, Cortese et al. only reported a frequency of 0.7 % in an European population (Cortese et al., 2019b; Davies et al., 2022b; Rafehi et al., 2019).
The observed haplotype frequency of 5.74 % of the MUTexp haplotype in our controls seemed to lie within the by Rafehi et al. given range.

5.2. The rs11096992 Risk Haplotype in Different RFC1 Expansions

The data published by Cortese et al. suggest a shared haplotype within the chromosomal region chr4:38977921–40712231 among patients with the MUTexp/MUTexp variant of RFC1 (Cortese et al., 2019a). To analyze if that hypothesis is in concordance with our data, we tested for the association of one of the reported SNPs (rs11096992) and different *RFC1* genotypes (MUTexp/MUTexp, MUTexp/WT, WT/WT). We show that the homozygous risk genotype AA was significantly associated with the MUTexp/MUTexp.
The fact that repeat expansions and specific single nucleotide variants (SNVs) are located on the same haplotypes and consequently can be captured by them has also been shown for other REMs diseases (Costa et al., 2019). One example would be *C9orf72*, where the SNP rs3849942 mainly occurs on alleles which are also carrying an expansion (Jones et al., 2013; König et al., 2022).
In case of *RFC1* only WT/WT *RFC1* carriers showed the GG genotype (which was not associated with the risk of an expansion). This suggests, that absence of the risk SNV rs11096992 on both alleles serves to exclude the *MUTexp/MUTexp* RFC1 genotype .
However, a biallelic presence of the rs11096992 SNV doesn't allow for direct conclusion regarding pathogenic expansions, since also heterozygous MUTexp/WT samples showed the rs11096992 on both alleles. Protocols for the detection of REMs are usually very time-consuming, as they require fPCR, RP-PCR, and in specific cases also Southern blotting to draw a concrete conclusion. The detection of the risk SNV rs11096992 could streamline this process. By screening for the SNV rs11096992 and thereby excluding WT/WT samples (those with GG on both alleles), fPCR, RP-PCR, and Southern blotting would only need to be applied to the remaining samples, saving significant time. Other methods like long range sequencing are often high priced and therefore not appropriate for screening large cohorts.
So far only an approximate range of repeat units of *RFC1* (144 to 820 and 333 to 1183 according to Kytövuori et al.,2022 and Alvarez Jerez et al., 2024, respectively) is given as a threshold for reaching pathogenicity. As a very strong correlation of the risk allele rs11096992 and the REM in *RFC1* could be observed, this SNP could yield the potential for narrowing down this given interval, if combined with other experiments.

Thus, screening for the risk haplotype rs11096992 not only enables shortening the detection of *RFC1* expansions in large cohorts and providing a cost-effective alternative to other screening procedures but also offers a possibility for defining a more certain threshold of pathogenicity.

5.3. RFC1 Expansion Length in PD and Related Diseases

To evaluate the respective length of the *RFC1* expansion in our PD patients positive for the MUTexp/MUTexp genotype, we conducted Southern Blot analysis.

Our observed *RFC1* expansion lengths ranging from 70 to 160 on the smaller and from 380 to 620 repeats on the larger allele are similar to those suggested for PD by Kytövuori et al. and Alvarez Jerez et al. (Alvarez Jerez et al., 2024; Kytövuori et al., 2022). The data published by these authors revealed an expansion size of 144 to 820 and 333 to 1183 repeats.

Interestingly we observed the CANVAS samples (used as positive controls) exhibiting a higher range of repats than the PD samples, with a mean of 900 compared to 550 repeats, respectively. This pattern of PD expansions being smaller than CANVAS expansions is also supported when comparing our PD expansion sizes with published CANVAS expansion sizes. Cortese et al. initially proposed repeat sizes ranging from 400 to 2,000 units causing CANVAS, which is clearly above our discovered range of RFC1 expansions associated with PD. This aspect may be important for distinguishing PD from CANVAS in futural diagnostics.

Additionally, as already indicated, we observed a distinct pattern of one allele carrying a greater *RFC1* expansion than the other one in our PD patients, which we will refer to as a biallelic but heterozygous *RFC1* expansion. Meaning that the RFC1 expansion correlating with PD is present on both alleles but largely differs in size. Five out of six PD samples showed an almost WT sized *RFC1* on one allele and the one clearly exceeding the pathogenic threshold but not meeting the average CANVAS size on the other allele.

This constitutes another genetic dissimilarity of PD- and CANVAS patients, as CANVAS patients show a similar *RFC1* expansion size on both alleles (i.e., a biallelic and homozygous *RFC1* expansion).

5.4. RFC1 Genotype and Age at Onset

Recent publications by Currò et al. and Yilkotila et al., have revealed a decreased AAO in patients carrying the pathogenic “AAGGG” RFC1 repeat (Currò et al., 2024; Ylikotila et al., 2023b). We therefore anticipated to find a difference between the mono- or biallelic MUTexp *RFC1* variants compared to the WT *RFC1* in their tendency of AAO in the PD patients . Nevertheless, we could not observe a significant effect in our cohort. Patients carrying the MUTexp/MUTexp genotype exhibited a similar AAO to those harboring the WT/WT variant. Additionally, we divided the cohort in early onset and late onset PD patients and tested if those carrying a MUTexp/MUTexp genotype, had a higher chance of belonging to the early onset group. Again, we couldn’t detect any significant difference. The contradictory of these findings with existing literature may be resulting from the small number of patients harboring the pathological genotype of interest (MUTexp/MUTexp) as well as the additional issue of lack of onset age information.

5.5. RFC1 Expansion Length and Age at Onset

Existing evidence suggests that expansion size negatively correlates with the onset age of the REM diseases such as it is the case for HD, C9ALS, C9FTD, DM, and FRDA (Alvarez Jerez et al., 2024; Kytövuori et al., 2022). This association leads to earlier onsets in patients with larger expanded genes. Moreover, an *RFC1* specific, previous study by Currò et al. indicated that this mechanism is also true for the by *RFC1* mediated disease CANVAS (Currò et al., 2024). To assess if this trend also shows in our cohort, we analyzed the possible association between *RFC1* sizes with the respective AAO.

Interestingly, our data doesn't confirm the finding of Currò et al. of the expansion size of the smaller allele leading to earlier AAO of PD (Currò et al., 2024). Whereas the aforesaid publication revealed an inverse correlation between AAO and repeat size of the smaller *RFC1* allele, we couldn't observe this correlation in our data. We therefore hypothesized the smaller alleles of our patients to be sufficient to activate pathogenic pathways, and consequently extended this analysis also towards the larger alleles. Again, no significant correlation could be observed for our cohort. However, due to the fact that more than one third of the variance in AAO of the PD patients can be explained by the length of the *RFC1* expansion on the larger allele, we assumed that the correlation may increase with larger sample size. Visual examination of the plotted data in course of the regression analysis additionally support this suggestion as the age distribution shows the trend of patients carrying larger expansion, having an earlier AAO.

5.6. Family History of PD Patients with RFC1 Repeat Expansions

In the next part of our analysis we investigated whether the group of patients with a biallelic *RFC1* expansion have a positive family history of the PD or associated diseases. It is well established that the risk of developing Parkinson's disease (PD) is significantly elevated in individuals with relatives, particularly first-degree relatives, who have also been affected by the disease (Ortega et al., 2022; Rybicki et al., 1999; Sellbach et al., 2006). We found three out of nine patients (33.33 %) with the MUTexp/MUTexp *RFC1* expansion to have a positive family history. Notably, this is not significantly different from PD patients carrying the WT/WT *RFC1* genotype (281 out of 441 patients with a positive family history, 38.92 %). This outcome is not unexpected since RFC1-Parkinson's disease is inherited in a recessive manner. Most patients with a positive family history in our cohort exhibit a pattern of dominant inheritance.

5.7. RFC1 Specific Clinical Appearance of PD Patients

Since individuals harboring a *RFC1* REM disease show a broad range of phenotypes, we wanted to delve deeper into the clinical picture of patients carrying the mutation of interest (MUTexp/MUTexp).

While existing publications regarding *RFC1* REMs suggest a wide range of onset age, a very recent study by Ylikotila et al. implicates a correlation of a MUTexp/MUTexp *RFC1* genotype in individuals with PD with a rather early AAO (Currò et al., 2024; Ylikotila et al., 2023a).

With a mean AAO of 62 years in our eight patients, our data exceeded the by Ylikotila et al reported onset of symptoms of *RFC1* REM harboring PD patients of 47 years. However, following publications have revealed central tendencies very similar to our observed one. Kytövuori et al.

and Jerez at al. for example reported an median onset age of 59 and 61.7 years for PD patients harboring a *RFC1* REM , respectively (Alvarez Jerez et al., 2024; Kytövuori et al., 2022).
As already stated in the "Clinical Spectrum of *RFC1* Related Diseases" part in the introduction chapter, *RFC1* has initially been connected with the neurodegenerative disease CANVAS, which manifests in cerebellar ataxia, neuropathy and vestibular areflexia syndrome (Cortese et al., 2019a). However, since the identification of *RFC1* REMs, phenotypes have been updated frequently. Some studies have documented *RFC1* expanded patients developing MSA, late onset ataxia, neuropathy and PD (Delforge et al., 2024). Frequent symptoms associated with an *RFC1* REM are cerebellar ataxia, neuropathy, vestibular areflexia syndrome, chronic cough, autonomic dysfunction, somatosensory impairment and peripheral sensory motor neuropathy (Cortese et al., 2019a; Currò et al., 2024; Delforge et al., 2024; Kytövuori et al., 2022). Notably, in our MUTexp/MUTexp positive patients no CANVAS associated phenotypical traits could be observed. This was also reported by Alverez Jerez et al., who extensively phenotyped four PD cases carrying a biallelic pathogenic *RFC1* repeat expansion (Alvarez Jerez et al., 2024). Furthermore, also Ylikotila et al. revealed none of their three MUTexp/MUTexp carrying patients exhibiting signs of ataxia, neuropathy or chronic cough, however one patient showed mild vestibular impairment (Ylikotila et al., 2023b). Based on the analysis of our eight repeat expansion carriers, we therefore propose that PD patients carrying the MUTexp/MUTexp genotype are clinically not distinguishable from typical PD patients.
To the date, published data regarding levodopa response of *RFC1* related diseases has not yet suggested a clear tendency (Delforge et al., 2024; Kytövuori et al., 2022; Record et al., 2022). Remarkably, for all except one patient very weak levodopa response has been recorded. This is very unusual, since the response to levodopa constitutes one crucial criterion for the diagnosis of PD. We conclude that further investigations are required to evaluate the molecular pathway behind the pharmacological interaction of levodopa and REM induced physiological mechanisms.

5.8. Limitations

In our study we only focused on one pathogenic repeat $(\mathrm{AAAGG})_{\mathrm{exp}}$. However, previous studies have also reported other pathogenic repeats such as $(\mathrm{AAGGG})_{\mathrm{exp}}$ and $(\mathrm{ACAGG})_{\mathrm{exp}}$, mainly in other ethnicities (Davies et al., 2022a; Dominik et al., 2021; Tyagi et al., 2023). In our present study we did not investigate these other repeat expansions in association with PD. However, we included the detection for the wild type expanded motif (WTexp: $(\mathrm{AAAAG})_{\mathrm{exp}}$) as well as the expanded motif of main interest (MUTexp: $(\mathrm{AAGGG})_{\mathrm{exp}}$) in our study. Notably, some outcomes of our multi-step protocol needed further analysis for a proper diagnosis of the respective genotype, as for example samples negative for the flanking as well as both repeat primed PCRs. It can be assumed, that those samples carry an alternative variant of the *RFC1* motif, which was not the focus of detection in our study.
We also couldn't calculate the frequency of heterozygous WTexp patients, as our protocol only provided information about the number of homozygous WTexp carrying individuals (see figure 8).
Furthermore, we hypothesized an alternation of motifs within a expansion. Since protocols based on RP-PCRs are sufficient to detect such cases, the only alternative here would be to apply long

range sequencing methods. Being the most cost intensive method on one hand, it would also yield the most information, as the whole sequence of the expansion is derived, on the other hand.

5.9. Conclusion and Future Prospective

Notwithstanding the great progress in identifying monogenetic variants inducing the pathogenesis of PD, approximately 85 to 90 % of the PD cases remain genetically undiagnosed within most populations. Our findings provide strong evidence of the association of a biallelic expansion within *RFC1* and the development of PD. We consequently confirm already published connection of RFC1 expansions and PD, by Alvarez Jerez et al., Kytövuori et al., Silva Schmitt et al. and Ylikotila et al. (Alvarez Jerez et al., 2024; Kytövuori et al., 2022; Silva Schmitt et al., 2020; Ylikotila et al., 2023b).

Notably, our study is the first to demonstrate that not only an expansion of a mutated motif, but also an expansion of the WT motif correlates positively with the development of the disease. The observed prevalence of 1.34% for cases associated with the MUTexp/MUTexp variant of *RFC1* and 1.67% for WT cases in our PD cohort indicates great significance. However, no causality can be indicated yet, as additional cell culture or model system based experiments would be needed to verify the exact pathological mechanism induced by *RFC1* expansions leading to PD.

Notwithstanding, we anticipate that an implementation of genetic testing for these types of *RFC1* expansions could be relevant regarding genetic counseling and establishing prognosis, especially for patients exhibiting no levodopa response.

Furthermore, we found that none of the *RFC1* WT carriers possessed the SNV rs11096992, suggesting that this risk haplotype is absent in *RFC1* WT PD patients. Consequently, we recommend genotyping this SNV in clinical settings as a streamlined diagnostic approach for identifying the MUTexp/MUTexp *RFC1* variant by excluding WT carriers.

As already explained in the “Methods” chapter, our multi-step protocol was only sensitive for certain expanded motifs. It would therefore not only be of great importance to include a higher variety of motifs but also to allow for altering motifs throughout an expansion. Thus, we anticipate that an implementation of long read sequencing is indispensable for further investigation of REMs.

References

Agresti, A. (2012). *Categorical Data Analysis*. John Wiley & Sons.

Alvarez Jerez, P., Daida, K., Miano-Burkhardt, A., Iwaki, H., Malik, L., Cogan, G., Makarious, M. B., Sullivan, R., Vandrovcova, J., Ding, J., Gibbs, J. R., Markham, A., Nalls, M. A., Kesharwani, R. K., Sedlazeck, F. J., Casey, B., Hardy, J., Houlden, H., Blauwendraat, C., … Billingsley, K. J. (2024). Profiling complex repeat expansions in RFC1 in Parkinson's disease. *Npj Parkinson's Disease*, *10*(1), 108. https://doi.org/10.1038/s41531-024-00723-0

Barghigiani, M., De Michele, G., Tessa, A., Fico, T., Natale, G., Saccà, F., Pane, C., Cuomo, N., De Rosa, A., Pappatà, S., De Michele, G., Santorelli, F. M., & Filla, A. (2022). Screening for RFC-1 pathological expansion in late-onset ataxias: A contribution to the differential diagnosis. *Journal of Neurology*, *269*(10), 5431–5435. https://doi.org/10.1007/s00415-022-11192-x

Bauer, P. O., & Nukina, N. (2009). The pathogenic mechanisms of polyglutamine diseases and current therapeutic strategies. *Journal of Neurochemistry*, *110*(6), 1737–1765. https://doi.org/10.1111/j.1471-4159.2009.06302.x

Beecroft, S. J., Cortese, A., Sullivan, R., Yau, W. Y., Dyer, Z., Wu, T. Y., Mulroy, E., Pelosi, L., Rodrigues, M., Taylor, R., Mossman, S., Leadbetter, R., Cleland, J., Anderson, T., Ravenscroft, G., Laing, N. G., Houlden, H., Reilly, M. M., & Roxburgh, R. H. (2020). A Māori specific RFC1 pathogenic repeat configuration in CANVAS, likely due to a founder allele. *Brain*, *143*(9), 2673–2680. https://doi.org/10.1093/brain/awaa203

Bourgeois, J. A., Seritan, A. L., Casillas, E. M., Hessl, D., Schneider, A., Yang, Y., Kaur, I., Cogswell, J. B., Nguyen, D. V., & Hagerman, R. J. (2011). Lifetime Prevalence of Mood and Anxiety Disorders in Fragile X Premutation Carriers. *The Journal of Clinical Psychiatry*, *72*(02), 175–182. https://doi.org/10.4088/JCP.09m05407blu

Burd, C., & Cullen, P. J. (2014). Retromer: A Master Conductor of Endosome Sorting. *Cold Spring Harbor Perspectives in Biology*, *6*(2), a016774–a016774. https://doi.org/10.1101/cshperspect.a016774

Cerri, S., Mus, L., & Blandini, F. (2019). Parkinson's Disease in Women and Men: What's the Difference? *Journal of Parkinson's Disease*, *9*(3), 501–515. https://doi.org/10.3233/JPD-191683

Chen, Y., Liu, M., & Dong, Z. (2021). Preferential Ribosome Loading on the Stress-Upregulated mRNA Pool Shapes the Selective Translation under Stress Conditions. *Plants*, *10*(2), 304. https://doi.org/10.3390/plants10020304

Chen, Y., & Varani, G. (2013). Engineering RNA -binding proteins for biology. *The FEBS Journal*, *280*(16), 3734–3754. https://doi.org/10.1111/febs.12375

Chen, Z., Li, G., & Liu, J. (2020). Autonomic dysfunction in Parkinson's disease: Implications for pathophysiology, diagnosis, and treatment. *Neurobiology of Disease*, *134*, 104700. https://doi.org/10.1016/j.nbd.2019.104700

Chintalaphani, S. R., Pineda, S. S., Deveson, I. W., & Kumar, K. R. (2021). An update on the neurological short tandem repeat expansion disorders and the emergence of long-read sequencing diagnostics. *Acta Neuropathologica Communications*, *9*(1), 98. https://doi.org/10.1186/s40478-021-01201-x

Cleary, J. D., & Ranum, L. P. W. (2014). Repeat associated non-ATG (RAN) translation: New starts in microsatellite expansion disorders. *Current opinion in genetics & development*, *0*, 6–15. https://doi.org/10.1016/j.gde.2014.03.002

Cortese, A., & Reilly, M. M. (o. J.). *RFC1 CANVAS / Spectrum Disorder*.

Cortese, A., Simone, R., Sullivan, R., Vandrovcova, J., Tariq, H., Yau, W. Y., Humphrey, J., Jaunmuktane, Z., Sivakumar, P., Polke, J., Ilyas, M., Tribollet, E., Tomaselli, P. J., Devigili, G., Callegari, I., Versino, M., Salpietro, V., Efthymiou, S., Kaski, D., … Houlden, H. (2019a). Biallelic expansion of an intronic repeat in RFC1 is a common cause of late-onset ataxia. *Nature Genetics*, *51*(4), 649–658. https://doi.org/10.1038/s41588-019-0372-4

Cortese, A., Simone, R., Sullivan, R., Vandrovcova, J., Tariq, H., Yau, W. Y., Humphrey, J., Jaunmuktane, Z., Sivakumar, P., Polke, J., Ilyas, M., Tribollet, E., Tomaselli, P. J., Devigili, G., Callegari, I., Versino, M., Salpietro, V., Efthymiou, S., Kaski, D., … Houlden, H. (2019b). Biallelic expansion of an intronic repeat in RFC1 is a common cause of late-onset ataxia. *Nature Genetics*, *51*(4), 649–658. https://doi.org/10.1038/s41588-019-0372-4

Cortese, A., Simone, R., Sullivan, R., Vandrovcova, J., Tariq, H., Yau, W. Y., Humphrey, J., Jaunmuktane, Z., Sivakumar, P., Polke, J., Ilyas, M., Tribollet, E., Tomaselli, P. J., Devigili, G., Callegari, I., Versino, M., Salpietro, V., Efthymiou, S., Kaski, D., … Houlden, H. (2019c). Biallelic expansion of an intronic repeat in RFC1 is a common cause of late-onset ataxia. *Nature Genetics*, *51*(4), 649–658. https://doi.org/10.1038/s41588-019-0372-4

Costa, I. P. D., Almeida, B. C., Sequeiros, J., Amorim, A., & Martins, S. (2019). A Pipeline to Assess Disease-Associated Haplotypes in Repeat Expansion Disorders: The Example of MJD/SCA3 Locus. *Frontiers in Genetics*, *10*, 38. https://doi.org/10.3389/fgene.2019.00038

Currò, R., Dominik, N., Facchini, S., Vegezzi, E., Sullivan, R., Galassi Deforie, V., Fernández-Eulate, G., Traschütz, A., Rossi, S., Garibaldi, M., Kwarciany, M., Taroni, F., Brusco, A., Good, J.-M., Cavalcanti, F., Hammans, S., Ravenscroft, G., Roxburgh, R. H., RFC1 repeat expansion study group, … Cortese, A. (2024). Role of the repeat expansion size in predicting age of onset and severity in RFC1 disease. *Brain: A Journal of Neurology*, awad436. https://doi.org/10.1093/brain/awad436

Davies, K., Szmulewicz, D., Corben, L., Delatycki, M., & Lockhart, P. (2022a). RFC1 -Related Disease: Molecular and Clinical Insights. *Neurology Genetics*, *8*, e200016. https://doi.org/10.1212/NXG.0000000000200016

Davies, K., Szmulewicz, D. J., Corben, L. A., Delatycki, M., & Lockhart, P. J. (2022b). *RFC1* -Related Disease: Molecular and Clinical Insights. *Neurology Genetics*, *8*(5), e200016. https://doi.org/10.1212/NXG.0000000000200016

Day, J. O., & Mullin, S. (2021). The Genetics of Parkinson's Disease and Implications for Clinical Practice. *Genes*, *12*(7), 1006. https://doi.org/10.3390/genes12071006

Delforge, V., Tard, C., Davion, J.-B., Dujardin, K., Wissocq, A., Dhaenens, C.-M., Mutez, E., & Huin, V. (2024). RFC1: Motifs and phenotypes. *Revue Neurologique*, *180*(5), 393–409. https://doi.org/10.1016/j.neurol.2024.03.006

Depienne, C., & Mandel, J.-L. (2021). 30 years of repeat expansion disorders: What have we learned and what are the remaining challenges? *American Journal of Human Genetics*, *108*(5), 764–785. https://doi.org/10.1016/j.ajhg.2021.03.011

Dias, V., Junn, E., & Mouradian, M. M. (2013). The Role of Oxidative Stress in Parkinson's Disease. *Journal of Parkinson's Disease*, *3*(4), 461–491. https://doi.org/10.3233/JPD-130230

Dominik, N., Galassi Deforie, V., Cortese, A., & Houlden, H. (2021). CANVAS: A late onset ataxia due to biallelic intronic AAGGG expansions. *Journal of Neurology*, *268*(3), 1119–1126. https://doi.org/10.1007/s00415-020-10183-0

Dominik, N., Magri, S., Currò, R., Abati, E., Facchini, S., Corbetta, M., Macpherson, H., Di Bella, D., Sarto, E., Stevanovski, I., Chintalaphani, S. R., Akcimen, F., Manini, A., Vegezzi, E., Quartesan, I., Montgomery, K.-A., Pirota, V., Crespan, E., Perini, C., … Cortese, A. (2023). Normal and pathogenic variation of *RFC1* repeat expansions: Implications for clinical diagnosis. *Brain*, *146*(12), 5060–5069. https://doi.org/10.1093/brain/awad240

Dupré, M., Hermann, R., & Froment Tilikete, C. (2021). Update on Cerebellar Ataxia with Neuropathy and Bilateral Vestibular Areflexia Syndrome (CANVAS). *Cerebellum (London, England)*, *20*(5), 687–700. https://doi.org/10.1007/s12311-020-01192-w

Eimer, H., Sureshkumar, S., Singh Yadav, A., Kraupner-Taylor, C., Bandaranayake, C., Seleznev, A., Thomason, T., Fletcher, S. J., Gordon, S. F., Carroll, B. J., & Balasubramanian, S. (2018). RNA-Dependent Epigenetic Silencing Directs Transcriptional Downregulation Caused by Intronic Repeat Expansions. *Cell*, *174*(5), 1095-1105.e11. https://doi.org/10.1016/j.cell.2018.06.044

Fanciulli, A., Stankovic, I., Krismer, F., Seppi, K., Levin, J., & Wenning, G. K. (2019). Chapter Five—Multiple system atrophy. In M. Stamelou & G. U. Höglinger (Hrsg.), *International Review of Neurobiology* (Bd. 149, S. 137–192). Academic Press. https://doi.org/10.1016/bs.irn.2019.10.004

Figueiredo, A. S., Loureiro, J. R., Macedo-Ribeiro, S., & Silveira, I. (2023). Advances in Nucleotide Repeat Expansion Diseases: Transcription Gets in Phase. *Cells*, *12*(6), 826. https://doi.org/10.3390/cells12060826

Fournier, C., Barbier, M., Camuzat, A., Anquetil, V., Lattante, S., Clot, F., Cazeneuve, C., Rinaldi, D., Couratier, P., Deramecourt, V., Sabatelli, M., Belliard, S., Vercelletto, M., Forlani, S., Jornea, L., French Clinical and Genetic Research Network on FTLD/FTLD-ALS, PREVDEMALS and FTLD-Exome Study Groups, Leguern, E., Brice, A., & Le Ber, I. (2019). Relations between C9orf72 expansion size in blood, age at onset, age at collection and transmission

across generations in patients and presymptomatic carriers. *Neurobiology of Aging*, *74*, 234.e1-234.e8. https://doi.org/10.1016/j.neurobiolaging.2018.09.010

Gallagher, C. L. (2019). *Imaging in Parkinson's Disease*. https://practicalneurology.com/articles/2019-sept/imaging-in-parkinsons-disease

Gatchel, J. R., & Zoghbi, H. Y. (2005). Diseases of Unstable Repeat Expansion: Mechanisms and Common Principles. *Nature Reviews Genetics*, *6*(10), 743–755. https://doi.org/10.1038/nrg1691

Gillies, G. E., Pienaar, I. S., Vohra, S., & Qamhawi, Z. (2014). Sex differences in Parkinson's disease. *Frontiers in Neuroendocrinology*, *35*(3), 370–384. https://doi.org/10.1016/j.yfrne.2014.02.002

Goh, Y. Y., Saunders, E., Pavey, S., Rushton, E., Quinn, N., Houlden, H., & Chelban, V. (2023). Multiple system atrophy. *Practical Neurology*, *23*(3), 208–221. https://doi.org/10.1136/pn-2020-002797

Groh, M., Lufino, M. M. P., Wade-Martins, R., & Gromak, N. (2014). R-loops Associated with Triplet Repeat Expansions Promote Gene Silencing in Friedreich Ataxia and Fragile X Syndrome. *PLoS Genetics*, *10*(5), e1004318. https://doi.org/10.1371/journal.pgen.1004318

He, F., & Todd, P. K. (2011). Epigenetic Mechanisms in Repeat Expansion Disorders. *Seminars in neurology*, *31*(5), 470–483. https://doi.org/10.1055/s-0031-1299786

Hirsch, L., Jette, N., Frolkis, A., Steeves, T., & Pringsheim, T. (2016). The Incidence of Parkinson's Disease: A Systematic Review and Meta-Analysis. *Neuroepidemiology*, *46*(4), 292–300. https://doi.org/10.1159/000445751

Iacoangeli, A., Al Khleifat, A., Jones, A. R., Sproviero, W., Shatunov, A., Opie-Martin, S., Morrison, K. E., Shaw, P. J., Shaw, C. E., Fogh, I., Dobson, R. J., Newhouse, S. J., Al-Chalabi, A., & Alzheimer's Disease Neuroimaging Initiative. (2019). C9orf72 intermediate expansions of 24–30 repeats are associated with ALS. *Acta Neuropathologica Communications*, *7*(1), 115. https://doi.org/10.1186/s40478-019-0724-4

Jacinto, F. V., Link, W., & Ferreira, B. I. (2020). CRISPR/Cas9-mediated genome editing: From basic research to translational medicine. *Journal of Cellular and Molecular Medicine*, *24*(7), 3766–3778. https://doi.org/10.1111/jcmm.14916

Jain, A., & Vale, R. D. (2017). RNA Phase Transitions in Repeat Expansion Disorders. *Nature*, *546*(7657), 243–247. https://doi.org/10.1038/nature22386

Jeong, G. R., & Lee, B. D. (2020). Pathological Functions of LRRK2 in Parkinson's Disease. *Cells*, *9*(12), 2565. https://doi.org/10.3390/cells9122565

Jones, A. R., Woollacott, I., Shatunov, A., Cooper-Knock, J., Buchman, V., Sproviero, W., Smith, B., Scott, K. M., Balendra, R., Abel, O., McGuffin, P., Ellis, C. M., Shaw, P. J., Morrison, K. E., Farmer, A., Lewis, C. M., Leigh, P. N., Shaw, C. E., Powell, J. F., & Al-Chalabi, A. (2013). Residual association at C9orf72 suggests an alternative amyotrophic lateral

sclerosis-causing hexanucleotide repeat. *Neurobiology of Aging*, *34*(9), 2234.e1-2234.e7. https://doi.org/10.1016/j.neurobiolaging.2013.03.003

König, T., Wurm, R., Parvizi, T., Silvaieh, S., Hotzy, C., Cetin, H., Klotz, S., Gelpi, E., Bancher, C., Benke, T., Dal-Bianco, P., Defrancesco, M., Fischer, P., Marksteiner, J., Sutterlüty, H., Ransmayr, G., Schmidt, R., Zimprich, A., & Stögmann, E. (2022). C9orf72 repeat length might influence clinical sub-phenotypes in dementia patients. *Neurobiology of Disease*, *175*, 105927. https://doi.org/10.1016/j.nbd.2022.105927

Kubota, T., Myung, K., & Donaldson, A. D. (2013). Is PCNA unloading the central function of the Elg1/ATAD5 replication factor C-like complex? *Cell Cycle*, *12*(16), 2570–2579. https://doi.org/10.4161/cc.25626

Kupiec, M. (2016). Alternative clamp loaders/unloaders. *FEMS Yeast Research*, *16*(7), fow084. https://doi.org/10.1093/femsyr/fow084

Kytövuori, L., Sipilä, J., Doi, H., Hurme-Niiranen, A., Siitonen, A., Koshimizu, E., Miyatake, S., Matsumoto, N., Tanaka, F., & Majamaa, K. (2022). Biallelic expansion in RFC1 as a rare cause of Parkinson's disease. *Npj Parkinson's Disease*, *8*(1), 6. https://doi.org/10.1038/s41531-021-00275-7

Langbehn, D. R. (2022). Longer CAG repeat length is associated with shorter survival after disease onset in Huntington disease. *American Journal of Human Genetics*, *109*(1), 172–179. https://doi.org/10.1016/j.ajhg.2021.12.002

Lanni, S., & Pearson, C. E. (2019). Molecular genetics of congenital myotonic dystrophy. *Neurobiology of Disease*, *132*, 104533. https://doi.org/10.1016/j.nbd.2019.104533

Lietard, J., Ameur, D., & Somoza, M. M. (2022). Sequence-dependent quenching of fluorescein fluorescence on single-stranded and double-stranded DNA. *RSC Advances*, *12*(9), 5629–5637. https://doi.org/10.1039/D2RA00534D

Lotankar, S., Prabhavalkar, K. S., & Bhatt, L. K. (2017). Biomarkers for Parkinson's Disease: Recent Advancement. *Neuroscience Bulletin*, *33*(5), 585–597. https://doi.org/10.1007/s12264-017-0183-5

Loureiro, J. R., Oliveira, C. L., Sequeiros, J., & Silveira, I. (2018). A repeat-primed PCR assay for pentanucleotide repeat alleles in spinocerebellar ataxia type 37. *Journal of Human Genetics*, *63*(9), Article 9. https://doi.org/10.1038/s10038-018-0474-3

Malik, I., Kelley, C. P., Wang, E. T., & Todd, P. K. (2021). Molecular mechanisms underlying nucleotide repeat expansion disorders. *Nature Reviews. Molecular Cell Biology*, *22*(9), 589–607. https://doi.org/10.1038/s41580-021-00382-6

McColgan, P., & Tabrizi, S. J. (2018). Huntington's disease: A clinical review. *European Journal of Neurology*, *25*(1), 24–34. https://doi.org/10.1111/ene.13413

Meyer, T. (2021). [Amyotrophic lateral sclerosis (ALS)—Diagnosis, course of disease and treatment options]. *Deutsche Medizinische Wochenschrift (1946)*, *146*(24–25), 1613–1618. https://doi.org/10.1055/a-1562-7882

Mir, M., Stadler, M. R., Ortiz, S. A., Hannon, C. E., Harrison, M. M., Darzacq, X., & Eisen, M. B. (2018). Dynamic multifactor hubs interact transiently with sites of active transcription in Drosophila embryos. *eLife*, *7*, e40497. https://doi.org/10.7554/eLife.40497

Miyatake, S., Koshimizu, E., Fujita, A., Doi, H., Okubo, M., Wada, T., Hamanaka, K., Ueda, N., Kishida, H., Minase, G., Matsuno, A., Kodaira, M., Ogata, K., Kato, R., Sugiyama, A., Sasaki, A., Miyama, T., Satoh, M., Uchiyama, Y., … Matsumoto, N. (2022). Rapid and comprehensive diagnostic method for repeat expansion diseases using nanopore sequencing. *Npj Genomic Medicine*, *7*(1), 1–15. https://doi.org/10.1038/s41525-022-00331-y

Montaut, S., Diedhiou, N., Fahrer, P., Marelli, C., Lhermitte, B., Robelin, L., Vincent, M. C., Corti, L., Taieb, G., Gebus, O., Rudolf, G., Tarabeux, J., Dondaine, N., Canuet, M., Almeras, M., Benkirane, M., Larrieu, L., Chanson, J.-B., Nadaj-Pakleza, A., … Tranchant, C. (2021). Biallelic RFC1-expansion in a French multicentric sporadic ataxia cohort. *Journal of Neurology*, *268*(9), 3337–3343. https://doi.org/10.1007/s00415-021-10499-5

Moon, H. E., & Paek, S. H. (2015). Mitochondrial Dysfunction in Parkinson's Disease. *Experimental Neurobiology*, *24*(2), 103–116. https://doi.org/10.5607/en.2015.24.2.103

Nguyen, L., Montrasio, F., Pattamatta, A., Tusi, S. K., Bardhi, O., Meyer, K. D., Hayes, L., Nakamura, K., Banez-Coronel, M., Coyne, A., Guo, S., Laboissonniere, L. A., Gu, Y., Narayanan, S., Smith, B., Nitsch, R. M., Kankel, M. W., Rushe, M., Rothstein, J., … Ranum, L. P. W. (2020). Antibody Therapy Targeting RAN Proteins Rescues C9 ALS/FTD Phenotypes in C9orf72 Mouse Model. *Neuron*, *105*(4), 645-662.e11. https://doi.org/10.1016/j.neuron.2019.11.007

Opara, J., Małecki, A., Małecka, E., & Socha, T. (2017). Motor assessment in Parkinson`s disease. *Annals of Agricultural and Environmental Medicine*, *24*(3), 411–415. https://doi.org/10.5604/12321966.1232774

Ortega, R. A., Bressman, S. B., Raymond, D., Ozelius, L. J., Katsnelson, V., Leaver, K., Swan, M. C., Shanker, V., Miravite, J., Wang, C., Bennett, S. A. L., & Saunders-Pullman, R. (2022). Differences in SEX-SPECIFIC Frequency of Glucocerebrosidase Variant Carriers and Familial Parkinsonism. *Movement Disorders*, *37*(11), 2217–2225. https://doi.org/10.1002/mds.29197

Paulson, H. (2018). Repeat expansion diseases. *Handbook of clinical neurology*, *147*, 105–123. https://doi.org/10.1016/B978-0-444-63233-3.00009-9

Porter, B., Macfarlane, R., Unwin, N., & Walker, R. (2006). The Prevalence of Parkinson's Disease in an Area of North Tyneside in the North-East of England. *Neuroepidemiology*, *26*(3), 156–161. https://doi.org/10.1159/000091657

Rafehi, H., Szmulewicz, D. J., Bennett, M. F., Sobreira, N. L. M., Pope, K., Smith, K. R., Gillies, G., Diakumis, P., Dolzhenko, E., Eberle, M. A., Barcina, M. G., Breen, D. P., Chancellor, A. M., Cremer, P. D., Delatycki, M. B., Fogel, B. L., Hackett, A., Halmagyi, G. M., Kapetanovic, S., … Lockhart, P. J. (2019). Bioinformatics-Based Identification of

Expanded Repeats: A Non-reference Intronic Pentamer Expansion in RFC1 Causes CANVAS. *The American Journal of Human Genetics*, *105*(1), 151–165. https://doi.org/10.1016/j.ajhg.2019.05.016

Record, C. J., Alsukhni, R. A., Curro, R., Kaski, D., Rubin, J. S., Morris, H. R., Cortese, A., Iodice, V., & Reilly, M. M. (2022). Severe distinct dysautonomia in *RFC1* -related disease associated with Parkinsonism. *Journal of the Peripheral Nervous System*, *27*(4), 311–315. https://doi.org/10.1111/jns.12515

Russillo, M. C., Andreozzi, V., Erro, R., Picillo, M., Amboni, M., Cuoco, S., Barone, P., & Pellecchia, M. T. (2022). Sex Differences in Parkinson's Disease: From Bench to Bedside. *Brain Sciences*, *12*(7), 917. https://doi.org/10.3390/brainsci12070917

Rybicki, B. A., Johnson, C. C., Peterson, E. L., Kortsha, G. X., & Gorell, J. M. (1999). A Family History of Parkinson's Disease and Its Effect on Other PD Risk Factors. *Neuroepidemiology*, *18*(5), 270–278. https://doi.org/10.1159/000026222

Sakamoto, N., Ohshima, K., Montermini, L., Pandolfo, M., & Wells, R. D. (2001). Sticky DNA, a Self-associated Complex Formed at Long GAA·TTC Repeats in Intron 1 of the Frataxin Gene, Inhibits Transcription. *Journal of Biological Chemistry*, *276*(29), 27171–27177. https://doi.org/10.1074/jbc.M101879200

Salcedo-Arellano, M. J., Hagerman, R. J., & Martínez-Cerdeño, V. (2020). Fragile X syndrome: Clinical presentation, pathology and treatment. *Gaceta Medica De Mexico*, *156*(1), 60–66. https://doi.org/10.24875/GMM.19005275

Sassone, J., Reale, C., Dati, G., Regoni, M., Pellecchia, M. T., & Garavaglia, B. (2021). The Role of VPS35 in the Pathobiology of Parkinson's Disease. *Cellular and Molecular Neurobiology*, *41*(2), 199–227. https://doi.org/10.1007/s10571-020-00849-8

Sellbach, A. N., Boyle, R. S., Silburn, P. A., & Mellick, G. D. (2006). Parkinson's disease and family history. *Parkinsonism & Related Disorders*, *12*(7), 399–409. https://doi.org/10.1016/j.parkreldis.2006.03.002

Silva Schmitt, G., Martinez, A. R. M., Graça, F. F., Lima, F. D., Bonadia, L. C., Amorim, B. J., Nucci, A., & França, M. C. (2020). Dopa-Responsive Parkinsonism in a Patient With Homozygous *RFC1* Expansions. *Movement Disorders*, *35*(10), 1889–1890. https://doi.org/10.1002/mds.28286

Sullivan, R., Yau, W. Y., Chelban, V., Rossi, S., Dominik, N., O'Connor, E., Hardy, J., Wood, N., Cortese, A., & Houlden, H. (2021). *RFC1* -related ataxia is a mimic of early multiple system atrophy. *Journal of Neurology, Neurosurgery & Psychiatry*, *92*(4), 444–446. https://doi.org/10.1136/jnnp-2020-325092

Takeuchi, T., & Nagai, Y. (2017). Protein Misfolding and Aggregation as a Therapeutic Target for Polyglutamine Diseases. *Brain Sciences*, *7*(10), Article 10. https://doi.org/10.3390/brainsci7100128

Taymans, J.-M., Fell, M., Greenamyre, T., Hirst, W. D., Mamais, A., Padmanabhan, S., Peter, I., Rideout, H., & Thaler, A. (2023). Perspective on the current state of the LRRK2 field. *Npj Parkinson's Disease*, *9*(1), 104. https://doi.org/10.1038/s41531-023-00544-7

The Multiple-System Atrophy Research Collaboration. (2013). Mutations in *COQ2* in Familial and Sporadic Multiple-System Atrophy. *New England Journal of Medicine*, *369*(3), 233–244. https://doi.org/10.1056/NEJMoa1212115

Tolosa, E., Garrido, A., Scholz, S. W., & Poewe, W. (2021). Challenges in the diagnosis of Parkinson's disease. *The Lancet. Neurology*, *20*(5), 385–397. https://doi.org/10.1016/S1474-4422(21)00030-2

Tomida, J., Masuda, Y., Hiroaki, H., Ishikawa, T., Song, I., Tsurimoto, T., Tateishi, S., Shiomi, T., Kamei, Y., Kim, J., Kamiya, K., Vaziri, C., Ohmori, H., & Todo, T. (2008). DNA Damage-induced Ubiquitylation of RFC2 Subunit of Replication Factor C Complex. *The Journal of Biological Chemistry*, *283*(14), 9071–9079. https://doi.org/10.1074/jbc.M709835200

Traschütz, A., Cortese, A., Reich, S., Dominik, N., Faber, J., Jacobi, H., Hartmann, A. M., Rujescu, D., Montaut, S., Echaniz-Laguna, A., Erer, S., Schütz, V. C., Tarnutzer, A. A., Sturm, M., Haack, T. B., Vaucamps-Diedhiou, N., Puccio, H., Schöls, L., Klockgether, T., … on behalf of the RFC1 Study Group. (2021). Natural History, Phenotypic Spectrum, and Discriminative Features of Multisystemic RFC1 Disease. *Neurology*, *96*(9), e1369–e1382. https://doi.org/10.1212/WNL.0000000000011528

Trollet, C., Boulinguiez, A., Roth, F., Stojkovic, T., Butler-Browne, G., Evangelista, T., Lacau St Guily, J., & Richard, P. (1993). Oculopharyngeal Muscular Dystrophy. In M. P. Adam, J. Feldman, G. M. Mirzaa, R. A. Pagon, S. E. Wallace, L. J. Bean, K. W. Gripp, & A. Amemiya (Hrsg.), *GeneReviews®*. University of Washington, Seattle. http://www.ncbi.nlm.nih.gov/books/NBK1126/

Tyagi, N., Uppili, B., Sharma, P., Parveen, S., Saifi, S., Jain, A., Sonakar, A., Ahmed, I., Sahni, S., Shamim, U., Anand, A., Suroliya, V., Asokachandran, V., Srivastava, A., Sivasubbu, S., Scaria, V., & Faruq, M. (2023). *Investigation of* RFC1 *tandem nucleotide repeat locus in diverse neurodegenerative outcomes in an Indian cohort* [Preprint]. Genetic and Genomic Medicine. https://doi.org/10.1101/2023.06.05.23290839

Váradi, C. (2020). Clinical Features of Parkinson's Disease: The Evolution of Critical Symptoms. *Biology*, *9*(5), 103. https://doi.org/10.3390/biology9050103

Vetcher, A. A., Napierala, M., Iyer, R. R., Chastain, P. D., Griffith, J. D., & Wells, R. D. (2002). Sticky DNA, a Long GAA·GAA·TTC Triplex That Is Formed Intramolecularly, in the Sequence of Intron 1 of the Frataxin Gene*. *Journal of Biological Chemistry*, *277*(42), 39217–39227. https://doi.org/10.1074/jbc.M205209200

Waschbüsch, D., & Khan, A. R. (2020). Phosphorylation of Rab GTPases in the regulation of membrane trafficking. *Traffic*, *21*(11), 712–719. https://doi.org/10.1111/tra.12765

Williams, E. T., Chen, X., Otero, P. A., & Moore, D. J. (2022). Understanding the contributions of VPS35 and the retromer in neurodegenerative disease. *Neurobiology of Disease*, *170*, 105768. https://doi.org/10.1016/j.nbd.2022.105768

Xiong, Y., Dawson, T. M., & Dawson, V. L. (2017). Models of LRRK2-Associated Parkinson's Disease. In H. J. Rideout (Hrsg.), *Leucine-Rich Repeat Kinase 2 (LRRK2)* (Bd. 14, S. 163–191). Springer International Publishing. https://doi.org/10.1007/978-3-319-49969-7_9

Ylikotila, P., Sipilä, J., Alapirtti, T., Ahmasalo, R., Koshimizu, E., Miyatake, S., Hurme-Niiranen, A., Siitonen, A., Doi, H., Tanaka, F., Matsumoto, N., Majamaa, K., & Kytövuori, L. (2023a). Association of biallelic *RFC1* expansion with early-onset Parkinson's disease. *European Journal of Neurology*, *30*(5), 1256–1261. https://doi.org/10.1111/ene.15717

Ylikotila, P., Sipilä, J., Alapirtti, T., Ahmasalo, R., Koshimizu, E., Miyatake, S., Hurme-Niiranen, A., Siitonen, A., Doi, H., Tanaka, F., Matsumoto, N., Majamaa, K., & Kytövuori, L. (2023b). Association of biallelic *RFC1* expansion with early-onset Parkinson's disease. *European Journal of Neurology*, *30*(5), 1256–1261. https://doi.org/10.1111/ene.15717

Zimprich, A., Benet-Pagès, A., Struhal, W., Graf, E., Eck, S. H., Offman, M. N., Haubenberger, D., Spielberger, S., Schulte, E. C., Lichtner, P., Rossle, S. C., Klopp, N., Wolf, E., Seppi, K., Pirker, W., Presslauer, S., Mollenhauer, B., Katzenschlager, R., Foki, T., … Strom, T. M. (2011). A Mutation in VPS35, Encoding a Subunit of the Retromer Complex, Causes Late-Onset Parkinson Disease. *The American Journal of Human Genetics*, *89*(1), 168–175. https://doi.org/10.1016/j.ajhg.2011.06.008

Zimprich, A., Biskup, S., Leitner, P., Lichtner, P., Farrer, M., Lincoln, S., Kachergus, J., Hulihan, M., Uitti, R. J., Calne, D. B., Stoessl, A. J., Pfeiffer, R. F., Patenge, N., Carbajal, I. C., Vieregge, P., Asmus, F., Müller-Myhsok, B., Dickson, D. W., Meitinger, T., … Gasser, T. (2004). Mutations in LRRK2 Cause Autosomal-Dominant Parkinsonism with Pleomorphic Pathology. *Neuron*, *44*(4), 601–607. https://doi.org/10.1016/j.neuron.2004.11.005

Zirra, A., Rao, S. C., Bestwick, J., Rajalingam, R., Marras, C., Blauwendraat, C., Mata, I. F., & Noyce, A. J. (2023). Gender Differences in the Prevalence of Parkinson's Disease. *Movement Disorders Clinical Practice*, *10*(1), 86–93. https://doi.org/10.1002/mdc3.13584

Supplementary Material

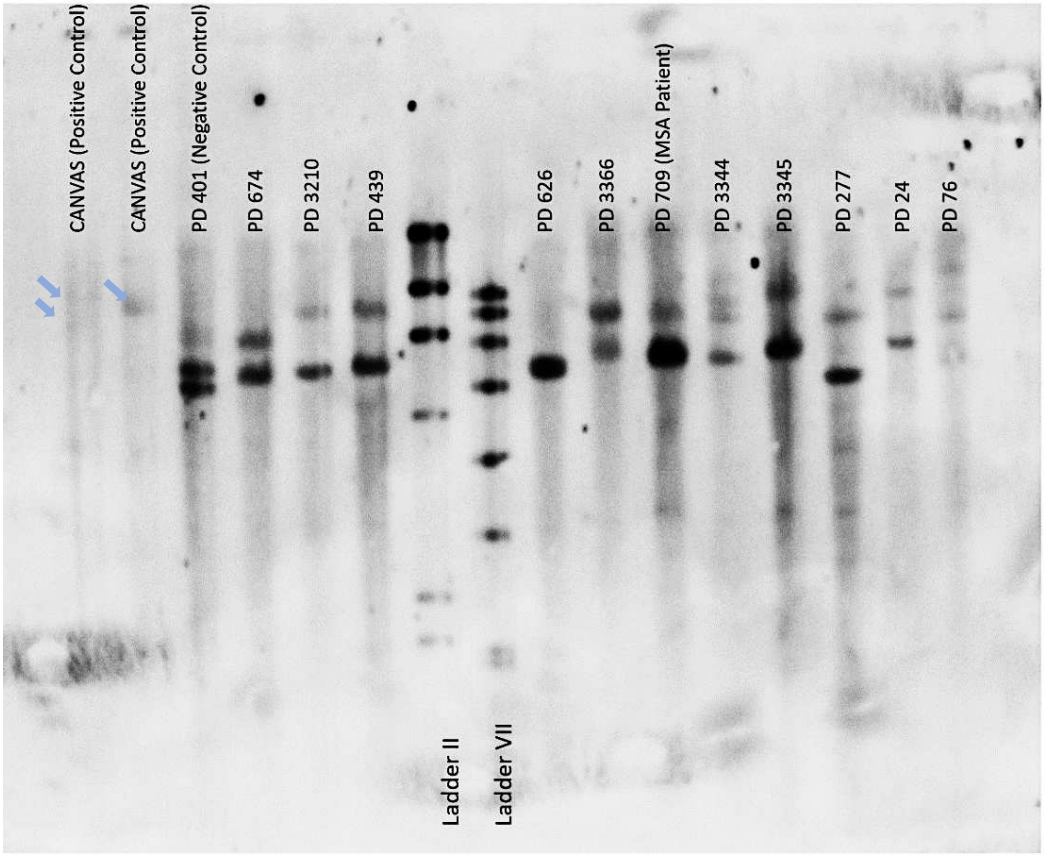

Figure 23. Supplementary Southern Blot.
The figure shows an additional Southern blot of CANVAS- and PD samples including monoallelic RFC1 expanded ones. Blue arrows indicate bands for RFC1 in CANVAS samples.

Table 33. Phenotyping of Biallelic Expanded Mutated RFC1 Motif Carriers Part 2.
The table summarizes important clinical features of patients carrying the MUTexp/MUTexp genotype. 1 marks the presence, whereas 0 the absence of a specific trait.

	Motoric Symptoms								
ID	**Tremor**	**Rigor**	**Walking/Gait Diff.**	**Dyskinesia**	**Hypokinesia**	**Bradykinesia**	**Akinesia**	**Vertigo**	**Speach Diff.**
PD24	1	1	1	0	N/A	1	N/A	N/A	N/A
PD76	1	N/A	N/A	1	N/A	1	1	N/A	N/A
PD88	1	1	1	N/A	1	N/A	N/A	N/A	1
PD 125	1	1	N/A	N/A	N/A	1	N/A	1	N/A
PD164	1	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
PD523	1	N/A	N/A	1	N/A	1	N/A	N/A	N/A
PD526	N/A	N/A	1	N/A	N/A	N/A	N/A	N/A	N/A
PD2052	N/A	1	1	N/A	N/A	1	1	N/A	N/A

	Non-Motoric Symptoms					
ID	**Cognitive Imp.**	**Dementia**	**Sleep Imp.**	**Depression**	**Hallucinations**	**Aggression**
PD24	1	1	N/A	N/A	1	N/A
PD76	1	N/A	1	N/A	N/A	N/A
PD88	N/A	N/A	1	N/A	1	N/A
PD 125	N/A	N/A	N/A	1	1	1
PD164	N/A	N/A	1	1	N/A	N/A
PD523	N/A	N/A	N/A	N/A	N/A	N/A
PD526	N/A	N/A	1	N/A	N/A	N/A
PD2052	1	1	N/A	1	N/A	N/A

Abbr.: ID = identification number, Mb. = Morbus, AAO = age at onset, y = years, F.A. = family anamnesis, L-Dopa = Levodopa, PD rel. Mut. = Parkinson's Disease relevant mutations, Diff. = difficulties, Imp. = impairment