

Masterarbeit

**Vergleich traditioneller PCR-basierter Methoden mit neu zu etablierendem  
Long-Read-Sequencing in der Diagnostik von ausgewählten  
Repeat-Expansions-Erkrankungen**

eingereicht von

**Angelika Humbel**

zur Erlangung des akademischen Grades

**Master of Science (Continuing Education)**

**MSc (CE)**

an der

**Medizinischen Universität Graz**

ausgeführt im

**Universitätslehrgang medizinische Genetik 2023**

unter der Anleitung von Betreuer\*in

Prof. Dr. med. A. Rauch

Zürich, 05. Januar 2026

## Eidesstattliche Erklärung

*Ich erkläre ehrenwörtlich, dass ich die vorliegende Arbeit selbstständig und ohne fremde Hilfe verfasst habe, andere als die angegebenen Quellen nicht verwendet habe und die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.*

*Des Weiteren erkläre ich hiermit, dass, sofern bei der Erstellung dieser Arbeit Künstliche Intelligenz (KI) Werkzeuge zur Generierung und/oder Korrektur bestimmter Textpassagen verwendet wurden, dieser Einsatz unter Einhaltung ethischer Grundsätze, akademischer Integrität und den Vorgaben meiner Universität erfolgte, sowie in Folge dies transparent gemacht und in angemessener Weise gekennzeichnet wurde.*

*Zürich, am 05. Januar 2026*

*eh. Angelika Humbel*

## Zusammenfassung

Hintergrund: Repeat-Expansions-Erkrankungen stellen eine Gruppe Erkrankungen dar, bei denen die molekulargenetische Diagnostik mit besonderen methodischen Herausforderungen verbunden ist. Da die momentan verwendeten traditionellen PCR-basierten Methoden Einschränkungen aufweisen, untersucht diese Arbeit die Leistungsfähigkeit von Long-Read-Sequencing (LRS) im Vergleich zu diesen Methoden für die Diagnostik ausgewählter Repeat-Expansions-Erkrankungen.

Methoden: Das Patient\*innenkollektiv umfasst 57 Proben für Erkrankungen in den Genen *FMR1*, *C9orf72*, *FXN*, *HTT*, *DMPK*, *RFC1*, verschiedene SCAs und *PHOX2B*. Alle 57 Proben wurden zunächst mit traditionellen Methoden wie Triplet-Repeat-Primed PCR, Fragmentanalyse, Sangersequenzierung und methylierungs-sensitiver-MLPA analysiert. Aufgrund zeitlicher Limitationen bei der Etablierung der LRS-Methode wurden lediglich 17 der Proben mittels PacBio HiFi-Sequenzierung untersucht. Die Ergebnisse beider Methoden wurden anschliessend verglichen.

Ergebnisse: Traditionelle PCR-basierte Methoden lieferten bei kleineren bis mittleren Repeat-Längen, sowie bei der Darstellung von Mosaiken zuverlässige Resultate, zeigten jedoch Einschränkungen bei der Quantifizierung von grossen Expansionen und der Charakterisierung von Sequenzmotiven. LRS ermöglichte eine präzisere Quantifizierung mit Werten bis zu 720 Repeats, die Identifikation von Sequenzmotiven, sowie die simultane Analyse des Methylierungsstatus. Bei 12 von 17 Proben stimmten die Ergebnisse beider Methoden weitgehend überein. Beobachtete Limitationen von LRS waren ein allelischer Bias mit teilweise nur einem Read pro expandiertem Allel, eingeschränkte automatische Mosaik-Detektion durch die verwendete Software und ein höherer DNA-Bedarf.

Schlussfolgerung: LRS ist unter den verwendeten Bedingungen kein vollständiger Ersatz für traditionelle Methoden, bietet jedoch in ausgewählten diagnostischen Situationen Vorteile. PCR-basierte Methoden bleiben aufgrund ihrer Robustheit, schnellen Durchführbarkeit und zuverlässigen Mosaik-Detektion die Methode der Wahl für die Routinediagnostik bei gezieltem klinischem Verdacht. Ergänzend bietet sich LRS als Screeningmethode für die parallele Analyse einer Vielzahl von Repeat-Expansions-Erkrankungen und als komplementäre Methode bei spezifischen Fragestellungen an.

## Abstract

**Background:** Repeat expansion disorders represent a significant group of diseases whose molecular genetic diagnostics pose methodological challenges. Since the traditional PCR-based methods currently in use have limitations, this study investigates the performance of long-read sequencing (LRS) in comparison to these methods for the diagnosis of selected repeat expansion disorders.

**Methods:** The patient cohort comprised 57 samples for diseases associated with the genes *FMR1*, *C9orf72*, *FXN*, *HTT*, *DMPK*, *RFC1*, various SCAs, and *PHOX2B*. First, all 57 samples were analysed using traditional methods such as triplet repeat-primed PCR, fragment analysis, sanger sequencing, and methylation-sensitive MLPA. Due to time constraints in establishing the LRS method, only 17 of the samples were examined using PacBio HiFi sequencing. The results of both methods were then compared.

**Results:** Traditional PCR-based methods provided reliable results for small to medium repeat lengths and for the representation of mosaics but showed limitations in the quantification of large expansions and the characterization of sequence motifs. LRS enabled more precise quantification with values up to 720 repeats, the identification of sequence motifs, and the simultaneous analysis of methylation status. In 12 of 17 samples, the results of both methods were largely consistent. Observed limitations of LRS were an allelic bias with sometimes only one read per expanded allele, limited automatic mosaic detection by the software used, and a higher DNA requirement.

**Conclusion:** Under the conditions used, LRS is not a complete replacement for traditional methods, but offers advantages in specific diagnostic situations. PCR-based methods remain the method of choice for routine diagnostics in cases of specific clinical suspicion due to their robustness, rapid feasibility, and reliable mosaic detection. In addition, LRS is suitable as a screening method for the parallel analysis of many repeat expansion diseases and as a complementary method for specific questions.

## Inhaltsverzeichnis

Eidesstattliche Erklärung .....	I
Zusammenfassung .....	II
Abbildungsverzeichnis.....	VII
Tabellenverzeichnis .....	IX
1 Einleitung .....	1
1.1 Repetitive Sequenzen im menschlichen Genom.....	1
1.2 Pathogenität von Repeat-Expansionen.....	1
1.3 Gemeinsame Merkmale von Repeat-Expansions-Erkrankungen.....	2
2 Repeat-Expansions-Erkrankungen im Detail.....	3
2.1 <i>FMR1</i> -bedingte Erkrankungen .....	3
2.2 <i>C9orf72</i> -bedingte Erkrankungen .....	5
2.3 Friedreich-Ataxie .....	6
2.4 Huntington-Erkrankung .....	8
2.5 Myotone Dystrophie Typ 1 .....	9
2.6 <i>RFC1</i> CANVAS .....	10
2.7 Spinocerebelläre Ataxien .....	11
2.8 Angeborenes zentrales Hypoventilationssyndrom .....	13
3 Fragestellungen .....	15
4 Material und Methoden.....	16
4.1 Patient*innenkollektiv .....	16
4.1.1 Fragmentanalyse und TP-PCR.....	18
4.1.2 MS-MLPA.....	19
4.1.3 Sequenzierung nach Sanger .....	20
4.1.4 Long-Read-Sequencing.....	20
5 Resultate .....	23
5.1 Resultate traditionelle Methoden.....	23
5.1.1 TP-PCR <i>FMR1</i> .....	23
5.1.2 TP-PCR <i>C9orf72</i> .....	26

5.1.3	TP-PCR und Fragmentanalyse <i>FXN</i> .....	27
5.1.4	Fragmentanalyse und Sequenzierung <i>HTT</i> .....	28
5.1.5	TP-PCR und Fragmentanalyse <i>DMPK</i> .....	28
5.1.6	Resultate <i>RFC1</i> .....	29
5.1.7	TP-PCR und Fragmentanalyse SCAs.....	29
5.1.8	Fragmentanalyse und Sequenzierung <i>PHOX2B</i> .....	30
5.1.9	MS-MLPA <i>FMR1</i> .....	31
5.2	Resultate Long-Read-Sequencing .....	32
5.2.1	Übersicht Resultate.....	32
5.2.2	Überblick generelle Qualität der Resultate.....	33
5.2.3	Waterfall-Plots.....	34
5.2.4	Darstellung allelischer Bias .....	35
5.3	Vergleich Resultate beider Methoden .....	36
5.3.1	Vergleich Anzahl Repeats .....	36
5.3.2	Vergleich Methylierung.....	36
6	Diskussion.....	38
7	Fazit .....	52
8	Ausblick.....	53
9	Literaturverzeichnis .....	55
10	Anhang.....	61
10.1	Verwendung von KI-gestützten Tools .....	61
10.2	Übersicht Primer und PCR-Programme .....	62
10.3	Waterfall-Plots.....	64
10.4	Übersicht Rohdaten .....	71

## Abkürzungsverzeichnis

ALS	amyotrophe Lateralsklerose
ATP	Adenosintriphosphat
CANVAS	cerebelläre Ataxie, Neuropathie, vestibuläre Areflexie Syndrom
CCHS	angeborenes zentrales Hypoventilationssyndrom
CLR	continuous long reads
DM1	Myotone Dystrophie Typ 1
DMPK	myotonic dystrophy protein kinase
DNA	Desoxyribonukleinsäure
DPR	Dipeptid-Repeat-Protein
FMRP	Fragile X Messenger Ribonukleoprotein 1
FRDA	Friedreich-Ataxie
FTD	frontotemporale Demenz
FXds	Fragile-X-assoziierte Störungen
FXPOI	Fragile-X-assoziierte Ovarialinsuffizienz
FXS	Fragiles-X-Syndrom
FXTAS	Fragiles-X-assoziiertes Tremo/Ataxis-Syndrom
Gb	Gigabase
GoF	Gain-of-Function
HD	Huntington Erkrankung
HiFi	high fidelity
Kb	Kilobase
LINE	long interspersed nuclear element
LoF	Loss-of-Function
LRS	Long-Read Sequencing
MLPA	multiplex ligation-dependent probe amplification
mRNA	messenger RNA
MS-MLPA	methylierungssensitive MLPA
NGS	Next Generation Sequencing
NPARM	Nicht-Polyalanin-Repeat-Expansions-Mutation
PacBio	Pacific Biosciences
PARM	Polyalanin-Repeat-Expansions-Mutation
Pat. Nr.	Patient*innenummer
PCNA	Proliferating Cell Nuclear Antigen
PCR	Polymerasenkettenreaktion
RAN-Translation	repeat-associated non-AUG Translation
RNA	Ribonukleinsäure
RP-PCR	repeat-primed Polymerasekettenreaktion
SCA	spinocerebelläre Ataxie
SMRT	single-molecule real-time
STR	short tandem repeat
TP-PCR	Triplet-Repeat-Primed PCR
TRGT	Tandem repeat genotyping tool
UTR	untranslatierte Region
ZMW	Zero-Mode-Waveguide
ZNS	zentrales Nervensystem

## Abbildungsverzeichnis

Abbildung 1 Lage der unterschiedlichen in der Arbeit behandelten Repeat-Expansionen.....	1
Abbildung 2 Zusammensetzung Patient*innenkollektiv .....	16
Abbildung 3 Prinzip TP-PCR .....	18
Abbildung 4 schematische Darstellung Prinzip MS-MLPA.....	19
Abbildung 5 Sequenzieretechnologie Long-Read-Sequencing PacBio .....	21
Abbildung 6 Elektropherogramm TP-PCR Probe C04-XD04.....	24
Abbildung 7 Elektropherogramm TP-PCR Probe C15-XR15.....	25
Abbildung 8 Elektropherogramm TP-PCR Probe C22-CR05 .....	26
Abbildung 9 Elektropherogramm TP-PCR Probe C29-FK02.....	27
Abbildung 10 Elektropherogramm TP-PCR Probe C39-D1W01.....	29
Abbildung 11 Resultat Sequenzierung C57-PD04.....	30
Abbildung 12 Resultat MS-MLPA der Probe C05-XD05 .....	31
Abbildung 13 Waterfall-Plot C04-XD04 .....	34
Abbildung 14 Waterfall-Plot C08-XD08 .....	34
Abbildung 15 Waterfall-Plot C35-HK02 .....	35
Abbildung 16 Waterfall-Plot C01-XD01 ( <i>FMR1</i> ) .....	64
Abbildung 17 Waterfall-Plot C02-XD02 ( <i>FMR1</i> ) .....	64
Abbildung 19 Waterfall-Plot C05-XD05 ( <i>FMR1</i> ) .....	64
Abbildung 20 Waterfall-Plot C06-XD06 ( <i>FMR1</i> ) .....	65
Abbildung 21 Waterfall-Plot C07-XD07 ( <i>FMR1</i> ) .....	65
Abbildung 23 Waterfall-Plot C09-XD09 ( <i>FMR1</i> ) .....	66
Abbildung 24 Waterfall-Plot C10-XD10 ( <i>FMR1</i> ) .....	66
Abbildung 25 Waterfall-Plot C17-XD17 ( <i>FMR1</i> ) .....	67
Abbildung 26 Waterfall-Plot C20-CD03 ( <i>C9orf72</i> ).....	67

Abbildung 27 Waterfall-Plot C28-FK01 ( <i>FXN</i> ) .....	68
Abbildung 28 Waterfall-Plot C29-FK02 ( <i>FXN</i> ) .....	68
Abbildung 29 Waterfall-Plot C34-HD01 ( <i>HTT</i> ) .....	69
Abbildung 31 Waterfall-Plot C41-D1D03 ( <i>DMPK</i> ) .....	69
Abbildung 32 Waterfall-Plot C44-CSK01 ( <i>RFC1</i> ) .....	70

## Tabellenverzeichnis

Tabelle 1 Überblick Referenzwerte Repeatanzahl spinocerebelläre Ataxien .....	13
Tabelle 2 Übersicht Referenzwerte Repeatanzahl unterschiedlicher Repeat- Expansions-Erkrankungen .....	14
Tabelle 3 Übersicht PCR-Bedingungen traditioneller Methoden.....	18
Tabelle 4 Resultate TP-PCR <i>FMR1</i> .....	23
Tabelle 5 Resultate TP-PCR <i>C9orf72</i> .....	26
Tabelle 6 Resultate TP-PCR und Fragmentanalyse <i>FXN</i> .....	27
Tabelle 7 Resultate Fragmentanalyse und Sequenzierung <i>HTT</i> .....	28
Tabelle 8 Resultate TP-PCR und Fragmentanalyse <i>DMPK</i> .....	28
Tabelle 9 Resultate TP-PCR und Fragmentanalyse SCAs.....	29
Tabelle 10 Resultate Fragmentanalyse und Sequenzierung <i>PHOX2B</i> .....	30
Tabelle 11 Resultate MS-MLPA <i>FMR1</i> .....	31
Tabelle 12 Übersicht Resultate LRS .....	32
Tabelle 13 Abweichung Anzahl Reads Allel 1 zu Allel 2.....	35
Tabelle 14 Vergleich Resultate traditioneller PCR-basierter Methoden mit LRS...	36
Tabelle 15 Vergleich Methylierungsmuster MS-MLA und LRS.....	36
Tabelle 16 Übersicht Primer und PCR-Programme für <i>FXN</i> , <i>HTT</i> , <i>DM1</i> , <i>PHOX2B</i> . .....	62
Tabelle 17 Übersicht Primer und PCR-Programme SCAS .....	63
Tabelle 18 Übersicht Rohdaten .....	72

# 1 Einleitung

## 1.1 Repetitive Sequenzen im menschlichen Genom

Sich wiederholende DNA-Sequenzen (Repeats) machen einen grossen Teil des menschlichen Genoms aus. Davon bestehen über 50 % aus repetitiven Elementen, welche in Tandem oder Simple Repeats oder auch in komplexere repetitive Elemente wie beispielsweise long interspersed nuclear elements (LINEs) unterteilt werden können [1].

Kurze Tandem-Wiederholungen (short tandem repeats, STRs), auch Mikrosatelliten genannt, sind kleine DNA-Abschnitte, meist zwei bis sechs Basenpaare lang, welche hintereinander, in tandem, am gleichen Locus erscheinen [2,3]. Am häufigsten sind Trinukleotid-Wiederholungen, gefolgt von Hexanukleotiden. Seltener handelt es sich um Tetra- und Pentanukleotid-Wiederholungen [3].

## 1.2 Pathogenität von Repeat-Expansionsen

Die Pathogenität von Repeat-Expansionsen ist abhängig von Faktoren wie der Wiederholungslänge, der Sequenz (mit oder ohne Unterbrechungen der Repeats) und der genomischen Lokalisation (innerhalb kodierender, intronischer oder regulatorischer Bereiche, Abbildung 1) sowie der Funktion des betroffenen Gens [1–3]. Je nach Konstellation der oben genannten Faktoren können Repeat-Expansionsen verschiedene krankheitsverursachende Mechanismen auslösen, welche sich grob in zwei Hauptkategorien einteilen lassen: Loss-of-Function (LoF) und Gain-of-Function (GoF) [2,4].

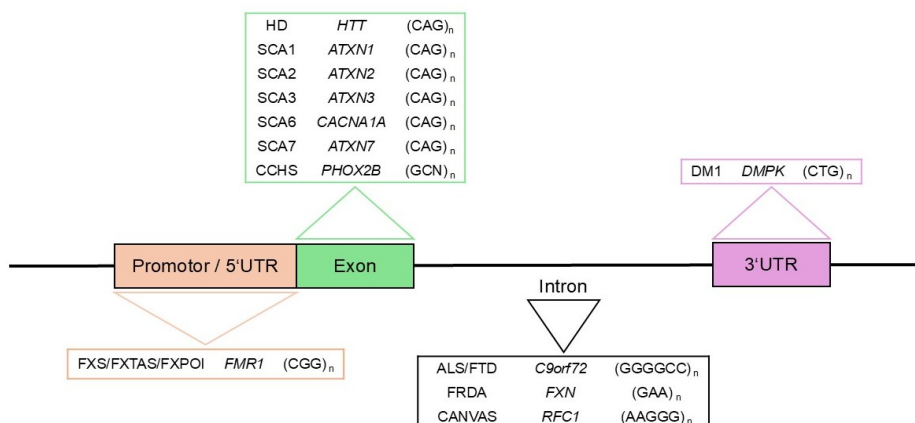


Abbildung 1 Lage der unterschiedlichen in der Arbeit behandelten Repeat-Expansionsen. Adaptiert nach [3] lizenziert unter <https://creativecommons.org/licenses/by/4.0/>

LoF-Mechanismen führen zu einem Funktionsverlust des betroffenen Gens. Sie umfassen Prozesse wie Hypermethylierung, Gen-Silencing, eine gestörte Transkription oder beschleunigter mRNA-Abbau. Ein Beispiel ist das Fragile X-Syndrom (FXS), bei dem eine CGG-Repeat-Expansion im *FMR1*-Promotorbereichs dessen Hypermethylierung verursacht, die wiederum zur Abschaltung des Gens führt [3]. Solche Mechanismen treten häufig bei autosomal rezessiv oder X-chromosomal rezessiv vererbten Erkrankungen auf [4].

GoF-Erkrankungen hingegen zeigen häufig eine autosomal dominante Vererbung [4]. GoF-Mechanismen gehen auf die toxische Wirkung von zusätzlicher oder veränderter Genprodukte zurück, welche infolge einer Expansion entstehen. Diese können die DNA-, RNA- oder Protein-Ebene betreffen. Auf DNA-Ebene können beispielsweise Chromatinveränderungen und epigenetische Repression auftreten, auf RNA-Ebene kann es durch die Sequestrierung von Splice-Faktoren zur Störung der Zellfunktion kommen [1]. Auf Protein-Ebene haben Repeat-Expansionen in codierenden Regionen häufig eine Auswirkung, durch Translation abnormal langer Aminosäureketten, wie beispielsweise Polyglutamin, Polyalanin oder Polyglycin [1,3]. Diese bilden Aggregate, welche neurotoxisch wirken [4]. Beispiele hierfür sind die spinocerebellären Ataxien (SCAs) Typ 1, 2, 3, 6 und 7 [3].

Ein besonders erwähnenswerter Mechanismus in diesem Zusammenhang ist die Repeat-Associated Non-AUG Translation (RAN), bei der pathologische Repeats nicht kanonisch, ohne klassisches Startcodon translatiert werden. Die RAN-Translation ist abhängig von der Anzahl Repeats und wird durch RNAs begünstigt, die Sekundärstrukturen bilden. Die so entstandenen Proteine sind ebenfalls zelltoxisch. Erkrankungen, die mit diesem Mechanismus assoziiert sind, sind beispielsweise SCA Typ 8 oder die *C9orf72* bedingte amyotrophe Lateralsklerose/frontotemporale Demenz [5].

### 1.3 Gemeinsame Merkmale von Repeat-Expansions-Erkrankungen

Ein weiteres zentrales Merkmal dieser Erkrankungen ist die Instabilität der Repeats während der DNA-Replikation und Meiose, was bedeutet, dass deren Länge in jeder Generation zunehmen kann [1]. Dies kann einerseits somatische Mosaik hervorrufen, andererseits ist es die Basis für das Phänomen der klinischen Antizipation.

Bei dieser treten die Symptome einer Erkrankung in nachfolgenden Generationen früher auf und/oder sind schwerer ausgeprägt, als bei der vorhergehenden [3].

Trotz der Unterschiede in Bezug auf Lokalisation, Länge und Motiv der Wiederholungen weisen die Repeat-Expansions-Erkrankungen weitere gemeinsame Eigenschaften auf, die für die Diagnose und das Management der Erkrankung von Bedeutung sind. Längere Wiederholungen stehen tendenziell im Zusammenhang mit einem früheren Beginn der Erkrankung oder mit einem schwereren Phänotyp. Dies ist beispielsweise bei der myotonen Dystrophie Typ 1 zwischen der kongenitalen und der spät einsetzenden Form zu beobachten [3].

## 2 Repeat-Expansions-Erkrankungen im Detail

In den folgenden Abschnitten werden nur solche Repeat-Expansions-Erkrankungen erwähnt, deren molekulargenetische Diagnostik am Institut für medizinische Genetik der Universität Zürich bereits etabliert ist und angeboten wird oder im Zuge der Etablierung der Long-Read-Sequencing Methode in Zukunft möglicherweise angeboten wird.

### 2.1 *FMR1*-bedingte Erkrankungen

*FMR1*-bedingte Erkrankungen umfassen verschiedene Krankheitsbilder, welche durch Varianten im *FMR1*-Gen verursacht werden. Dazu gehören das Fragile-X-Syndrom (FXS), das Fragile-X-assoziierte Tremor/Ataxie-Syndrom (FXTAS) und die Fragile-X-assoziierte Ovarialinsuffizienz (FXPOI) [6]. Zusammengefasst werden diese Krankheitsbilder auch als Fragile-X-assoziierte Störungen (FXds) bezeichnet [7].

Das *FMR1*-Gen ist auf dem X-Chromosom an Position Xq27.3 lokalisiert und kodiert für das Fragile X Messenger Ribonukleoprotein 1 (FMRP) [6]. Dabei handelt es sich um ein RNA-bindendes Protein, welches eine wesentliche Rolle in der Regulation der Translation von diversen mRNAs spielt, von denen viele in der Entwicklung der Synapsen und deren Plastizität involviert sind [8]. FMRP wird in praktisch allen Zelltypen in niedrigen Konzentrationen exprimiert, die höchsten Konzentrationen finden sich im Gehirn und in den Hoden [8].

Die Hauptursache der FXDs ist eine CGG-Triplett-Repeat-Expansion im Promotorbereich des *FMR1*-Gens [8].

Die CGG-Repeats sind meist durch 1 – 4 AGG-Repeats unterbrochen. Diese treten häufig nach 9 – 10 CGG-Repeats auf. Die AGG-Unterbrechungen wirken als Schutz gegen eine Expansion der CGG-Repeats. Fehlen diese AGG-Unterbrechungen, ist das Risiko für eine krankheitsrelevante Expansion bei Weitergabe erhöht, insbesondere über das mütterlichen Allel [6,8].

Die Einteilung der Repeat-Längen erfolgt in vier Kategorien. Zum Normbereich werden Allele < 45 Repeats gezählt. Allele von 45 – 54 Repeats werden zum Graubereich gezählt. Diese verursachen kein FXS, können aber instabil sein und wenn mütterlich vererbt, in den Prämutationsbereich expandieren [9]. Der Prämutationsbereich beinhaltet Allele von 55 bis 200 Repeats. Individuen mit einer Prämutation entwickeln in der Regel kein klassisches FXS. Es besteht jedoch ein erhöhtes Risiko für FXPOI bei Frauen und für FXTAS bei Männern und Frauen [9]. Zudem sind die Repeats von Prämutations-Allelen instabil und können bei der maternalen Vererbung zu einer Vollmutation expandieren [10]. Repeat-Längen > 200 Repeats werden als Vollmutationen bezeichnet und können bis mehrere Tausend Repeats gross werden [9].

Der Hauptteil der Fälle von FXS werden durch eine pathogene Repeat-Expansion verursacht. In weniger als 1 % der Fälle sind Veränderungen im *FMR1*-Gen wie Deletionen oder pathogene Punktmutationen in der kodierenden Region des Gens die Ursache für FXS [9]. Diese Veränderungen führen ebenfalls zur Abwesenheit oder zum Funktionsverlust von FMRP [11].

Die Pathogenese unterscheidet sich je nach Mutationstyp. Bei FXS führt eine Vollmutation üblicherweise zu einer vollständigen Hypermethylierung des *FMR1*-Promotors und somit zur Inaktivierung des Gens, was in einem völligen Fehlen von FMRP resultiert [8].

Bei FXTAS und FXPOI hingegen beruht der Krankheitsmechanismus auf einem RNA- und Protein-Gain-of-Function. Die erhöhte Expression der expandierten CGG-Region führt zur Anhäufung toxischer RNA-Transkripte, die zelluläre Prozesse stören können [7].

Einen wichtigen Einfluss auf die Schwere der Erkrankung hat der Mosaikstatus. Dabei können verschiedene Zellen desselben Individuums unterschiedliche Repeat-Längen und Methylierungsmuster aufweisen, was zu einer variablen Expression von *FMR1* und damit zu unterschiedlichem Schweregrad der Symptomatik führt [8].

## 2.2 *C9orf72*-bedingte Erkrankungen

*C9orf72* verursachte frontotemporale Demenz und/oder amyotrophe Lateralsklerose (*C9orf72*-FTD/ALS) ist eine neurologische Erkrankung, deren klinischen Verlauf sehr heterogen ist [12].

Die Diagnose wird durch den Nachweis einer heterozygoten abnormalen Hexanukleotid-Repeat-Expansion (GGGGCC, auch als G<sub>4</sub>C<sub>2</sub> bezeichnet) des *C9orf72*-Gen [13], welches sich auf dem kurzen Arm von Chromosom 9 befindet [14], gestellt. Das Gen beinhaltet 11 Exone. Exon 1, welches Exon 1a und 1b enthält und nicht codierend ist, sowie die codierenden Exone 2 bis 11. Die Expansion befindet sich dabei im Intron zwischen den nicht kodierenden Exonen 1a und 1b [15].

Normale Allele belaufen sich üblicherweise auf 2 bis 24 G<sub>4</sub>C<sub>2</sub>-Repeats. Die Repeat-Längen von 25 bis 60 Repeats sind von unklarer Signifikanz, da sie selten in der Allgemeinbevölkerung vorkommen und typischerweise nicht in betroffenen Familien segregieren. Pathogene Allele reichen von 61 bis > 4000 Repeats [12].

Es ist im Allgemeinen schwierig einen klaren Schwellenwert zwischen normalen und pathogenen Allelen festzulegen, da bereits kurze Expansionen von 45 bis 78 Repeats pathogen sein können [13].

Die genaue Funktion von *C9orf72* ist bis heute nicht vollständig geklärt. Es wird aber angenommen, dass das Gen eine Rolle bei der Autophagie, dem endosomalen Transport und in der Stressregulation im endoplasmatischen Reticulum spielt [13].

Es wurden drei pathogene Mechanismen beschrieben, die die Krankheit verursachen können und sich gegenseitig nicht ausschließen [13,14].

Einerseits der RNA-Gain-of-Function-Toxizitäts-Mechanismus. Die Sense- und Antisense- Repeat enthaltende RNA sammelt sich in nukleären RNA-Foci im zentralen Nervensystem (ZNS) an. Diese RNA-Foci kapseln RNA-bindende Proteine ab, was zu einer Dysregulation der Genexpression und des Splicings führt [13]. Sie lagern

sich vor allem im Kern ab, können aber auch im Zytoplasma erscheinen [14]. Dennoch ist ihre direkte Bedeutung für den klinischen Verlauf bislang unklar, da die Verteilung dieser Einschlüsse nicht konsistent mit betroffenen Hirnregionen korreliert [14].

Ein weiterer Mechanismus ist die Toxizität von Dipeptid-Repeat-Proteinen (DPR). Sowohl Sense- als auch Antisense-Transkripte der Repeat-Expansion werden durch eine RAN-Translation in verschiedene DPR übersetzt. Diese Proteine bilden neuronale Einschlüsse im ZNS. Insbesondere argininreiche DPRs wie Poly(GR) und Poly(PR) zeigten in Tiermodellen neurotoxische Effekte und neigten zur Aggregation [13].

Die Wichtigkeit der DPRs im Zusammenhang mit der Pathogenese der Erkrankung ist unsicher, da sie variabel vorkommen und nicht stark mit den relevanten ZNS Regionen korrelieren [14].

Ein weiterer Mechanismus führt zu Loss of Function. Die Repeat-Expansion verursacht eine reduzierte *C9orf72*-mRNA und Proteinexpression. Die Ursache dafür könnten epigenetische Veränderungen wie Hypermethylierung der flankierenden CpG-Inseln, der Repeats selbst sowie Histon-Trimethylierung, sein, welche alle die Transkription hemmen können. [13].

Die genaue Bestimmung der Repeat-Länge ist aufgrund des hohen GC-Gehalts, der grossen möglichen Anzahl der Repeats und der repetitiven Natur der flankierenden Sequenzen limitiert. Erschwerend kommt die Bildung von Mosaiken dazu. Die Repeat-Anzahl kann zwischen verschiedenen Geweben stark variieren, die Bestimmung der Anzahl erfolgt jedoch ante mortem hauptsächlich an DNA aus Blut [13]. Dies macht es ebenfalls schwierig, den pathologischen Schwellenwert der Anzahl festzulegen [14].

## 2.3 Friedreich-Ataxie

Bei der Friedreich-Ataxie (FRDA) handelt es sich um eine autosomal-rezessiv vererbte, fortschreitende neurologische Erkrankung, die durch Ataxie im Gang und den Gliedmassen, undeutliches Sprechen, Muskelschwäche und sensorischen Verlust gekennzeichnet ist. Im Verlauf treten häufig zusätzliche Symptome wie Skoliose,

Kardiomyopathie und bei manchen Patientinnen und Patienten auch Diabetes auf [16].

Die Hauptursache für die Erkrankung sind mit ungefähr 96 % aller Fälle biallelische GAA-Repeat-Expansionen im Intron 1 des *FXN*-Gens [17]. In selteneren Fällen sind die Betroffenen compound heterozygot für eine Repeat-Expansion und eine weitere inaktivierende Variante im *FXN*-Gen [18].

Normalallele weisen eine Repeat-Länge von 5 – 33 Repeats auf. Intermediär-Allele werde mit einer Länge von 34 – 65 Repeats beschrieben. Hierbei ist zu erwähnen, dass Allele von 44 – 65 als grenzwertig einzustufen sind, da die kürzeste, noch mit der Erkrankung assoziierte Repeat-Länge nicht klar definiert ist. Krankheitsverursachende Repeats können von 66 bis zu 1300 reichen [17].

Neben der Länge der expandierten Allelen, spielen auch Unterbrechungen der GAA-Repeats eine wichtige Rolle für die Ausprägung des Phänotyps der Erkrankung [17], da die Unterbrechungen die Expansion stabilisieren [18]. Bei den Unterbrechungen handelt es sich um G und/oder A Nukleotide. Unterbrochene Allele sind in der Regel kürzer und mit einem späteren Krankheitsbeginn assoziiert [17]. Ununterbrochene Repeat-Expansionen hingegen können innerhalb einer Generation weiter expandieren, was in einer somatischen Instabilität resultiert. Dies zeigen auch post-mortem Analysen, durch die Detektion von variierenden Repeat-Anzahl in verschiedenen Geweben [18].

Die Expansion im Intron 1 des *FXN*-Gens führt zur Bildung von abnormalen DNA-Strukturen, welche wiederum die RNA-Polymerase II blockieren. Zu dem verursacht die Expansion Histon-Modifikationen, welche zur Bildung von Heterochromatin und somit zur Stilllegung des *FXN*-Gens führen [18].

Das *FXN*-Gen codiert für das mitochondriale Protein Frataxin, welches eine essenzielle Rolle in der Eisen-Homöostase, der Biogenese von Eisen-Schwefel-Clustern, in zellulären Redox-Reaktionen, sowie in der Häm-Synthese spielt. Folglich führt eine reduzierte Produktion von Frataxin zu einer verminderten mitochondrialen ATP Produktion, einer Störung des Eisenstoffwechsels und einer beeinträchtigten Reaktion auf oxidativen Stress [18].

Exprimiert wird Frataxin vor allem in Geweben mit einem hohen Energiebedarf wie Herzmuskel, dem Kleinhirn und dem Rückenmark, was die organspezifischen Manifestationen der Erkrankung erklärt [18].

## 2.4 Huntington-Erkrankung

Die Huntington-Erkrankung (HD) ist eine fortschreitende neurologische Erkrankung, welche motorische, kognitive und psychiatrische Störungen verursacht [19]. Diagnostiziert wird sie einerseits durch die klinische Anamnese und andererseits durch den Nachweis einer Expansion von CAG-Trinukleotid-Wiederholungen im *HTT*-Gen, welches sich auf Chromosom 4p16 befindet [20]. Die Erkrankung wird autosomal dominant vererbt, wobei auch hier das Prinzip der Antizipation bekannt ist, welche häufiger bei väterlicher Vererbung des mutierten Allels ist. Sie resultiert aus der Instabilität der CAG-Repeats während der Spermatogenese [19].

Das *HTT*-Gen codiert für das Protein Huntingtin, welches ubiquitär exprimiert wird. Es enthält eine Polyglutamin Sequenz, welche durch ununterbrochene CAG Trinukleotid-Repeats im ersten Exon des *HTT*-Gens, codiert wird. Wildtyp Allele haben dabei < 26 Repeats. Allele zwischen 27 und 35 Repeats werden als intermediär betrachtet. Individuen mit solchen Repeat-Längen entwickeln keine Symptome, haben aber ein Risiko, dass bei der Weitergabe an die nächste Generation, aufgrund der Instabilität der CAG-Repeats, eine weitere Expansion stattfindet. Repeat-Längen > 36 Repeats gelten als pathologisch, wobei Repeat-Längen von 36 – 39 Repeats mit einer reduzierten Penetranz einhergehen [19].

Obwohl eine vollständige Penetranz von HD für CAG-Repeats von 42 und mehr beobachtet wird, zeigt nur ein Teil der Personen zwischen 36 und 41 Repeats Anzeichen von Symptomen einer HD [21].

Ein Grund hierfür könnte eine fehlende Unterbrechung der CAG-Repeats sein. Normalerweise sind diese gefolgt von einer 12 Basenpaar grossen Unterbrechung downstream: CAA-CAG-CCG-CCA. Da CAA ebenfalls für Glutamin codiert, wurde das Vorhandensein dieser Unterbrechung lange für irrelevant angesehen. Es wurde aber gezeigt, dass der Verlust dieser Unterbrechung mit einem früheren Manifestationsalter assoziiert ist, während eine Duplikation mit einem späteren Manifestationsalter in Zusammenhang gebracht wird. Allele ohne diese Unterbrechung sind

somatisch instabil und können bei Repeat-Längen mit geringer Penetranz dennoch zum Ausbruch der Erkrankung führen [22].

Die genauen Mechanismen, wie es zur Neurodegeneration kommt, sind bisher jedoch noch ungeklärt. Es wird aber angenommen, dass es eine Kombination aus toxischen Gain-of-Function Effekten und dem Verlust der normalen *HTT* Funktion ist [23]. Beim GoF-Effekt kommt es zur Aggregation und Bildung von toxischen Oligomeren [24]. Diese Aggregation von Huntingtin stört die Protein-Protein und Protein-DNA Interaktion, welche sich auch gegenseitig beeinflussen können [20].

## 2.5 Myotone Dystrophie Typ 1

Myotone Dystrophie Typ 1 (DM1) ist eine Form der Muskeldystrophie, die zu fortschreitender Muskelschwäche und -verlust führt. Obwohl die Symptome in jedem Alter auftreten können, ist es die häufigste Form der Muskeldystrophie, deren Beginn im Erwachsenenalter liegt [25].

Bei DM1 handelt es sich um eine autosomal dominant vererbte Erkrankung, welche durch eine Expansion von CTG-Nukleotiden in der 3'untranslatierten Region (UTR) des myotonic dystrophy protein kinase (*DMPK*) Gen verursacht wird. Das Erkrankungsalter so wie die Ausprägung des Phänotyps korrelieren dabei mit der Anzahl der CTG-Repeats [25]. Die schwerste Form ist die kongenitale DM1, bei der Repeat-Längen von 1400 Repeat oder höher beobachtet wurden [26].

Studien stützen die Theorie, dass ein RNA GoF- Mechanismus als ursächlich für die Erkrankung anzusehen ist. Es kommt zur Bildung von RNA-Foci, welche nukleäre Faktoren stören, die für die Entwicklung und Erhaltung der Muskeln wichtig sind. Ein weiterer Mechanismus, der diskutiert wird, sind Auffälligkeiten im alternativen Spleissen, welches essenziell ist für eine normale Myogenese [26].

Normale Allele haben eine Repeat-Länge von 5 bis 34 CTG-Repeats. Zu den Prämutations-Allelen werden Repeat-Längen von 35 bis 49 Repeats gezählt. Personen mit Allelen in diesem Bereich zeigen keine Symptome. Ihre Kinder haben aber ein erhöhtes Risiko eine grössere Repeat-Länge zu erben und somit Symptome zu entwickeln. Allele mit > 50 Repeats sind als pathologisch anzusehen [27].

Die CTG-Repeats zeigen eine hohe Instabilität, was zu einem ausgeprägten somatischen Mosaik führen kann. Die Repeat-Längen können sich dabei zwischen

verschiedenen Geweben stark unterscheiden, was wesentlich zur phänotypischen Variabilität der Erkrankung beiträgt [25, 27].

Eine zunehmende Anzahl an Repeats, sowie ein stärker ausgeprägtes Mosaik sind typischerweise mit einem schwereren Krankheitsverlauf und früherem Manifestationsalter assoziiert [28].

Ein weiterer entscheidender Faktor sind Unterbrechungen in der CTG-Expansion. Diese können entweder am 5' oder 3' Ende der Expansion vorkommen und haben oft einen stabilisierenden Effekt auf die Repeat-Struktur [28,29].

Die klinische Relevanz dieser Unterbrechungen zeigt sich in vielerlei Hinsicht. Grob zusammengefasst lässt sich sagen, dass sie mit einem milderem Phänotyp, einer reduzierten Beteiligung des ZNS und einem geringeren Auftreten der kongenitalen Form der Erkrankung assoziiert sind. Gleichzeitig können aber auch atypische und sogar schwere Verläufe auftreten, was die Interpretation und den Umgang mit diesen Unterbrechungen erschwert. Der Effekt der Unterbrechungen hängt stark von ihrer genomischen Lokalisation, Länge und Zusammensetzung ab [29].

## 2.6 *RFC1* CANVAS

Das CANVAS-Syndrom (cerebelläre Ataxie, Neuropathie, vestibuläre Areflexie Syndrom) ist eine nach dem 35. Lebensjahr einsetzende neurodegenerative Erkrankung, die durch eine biallelische AAGGG Repeat-Expansion im *RFC1*-Gen verursacht wird. Zum klinischen Spektrum der Erkrankung gehören neben dem klassischen CANVAS auch autonome Funktionsstörungen, Husten sowie Beeinträchtigungen des Kleinhirns, der Sensibilität, des Gleichgewichts und des Vestibularsystems [30].

Das *RFC1*-Gen befindet sich auf Chromosom 4 und codiert für die Untereinheit 1 des Replikationsfaktors C, eines Proteinkomplexes, welcher eine zentrale Rolle in der DNA-Replikation und DNA-Reparatur spielt. Der Proteinkomplex dient unter anderem als sogenannter «Clamp Loader», der das Proliferating Cell Nuclear Antigen (PCNA) an die DNA anheftet und somit die Anlagerung und Funktion der DNA-Polymerase unterstützt [31,32].

Bisher wurden mehrere repetitive Motive innerhalb des *RFC1*-Lokus beschrieben. Dabei handelt es sich nicht nur um Varianten im AAGGG-Repeat, sondern auch um

Repeats, die zusätzlich zu diesem Auftreten können. Neben der häufigsten pathogenen Expansion mit der Sequenz AAGGG wurden auch andere Varianten wie ACAGG und komplexere Repeats wie beispielsweise AAAGG<sub>15-25</sub>AAGGG<sub>exp</sub>AAGG<sub>10</sub> nachgewiesen, die in bestimmten Fällen ebenfalls mit der Erkrankung assoziiert sein können [32].

Die Einschätzung der Pathogenität basiert dabei auf verschiedenen Faktoren: der Häufigkeit der Repeats bei betroffenen Personen, dem homozygoten Auftreten bei gesunden Individuen sowie dem Vorliegen in Kombination mit einer weiteren pathogenen Variante auf dem gegenüberliegenden Allel [32].

Zu den Normalallelen werden unterschiedliche Motive mit unterschiedlicher Anzahl gezählt [30]:

- AAAAG<sub>11</sub> Repeats, mit einer Allelfrequenz von 0.75
- AAAAG<sub>12-200</sub> Repeats mit einer Allelfrequenz 0.13
- AAAGG<sub>40-1000</sub> Repeats mit einer Allelfrequenz von 0.08

Als pathologisch werden am häufigsten AAGGG Repeats von 400 bis 2000 beschrieben. In weiteren Populationen wurden zudem weitere Motive bereits als pathologisch beschrieben [30].

Die genaue Pathogenese der Erkrankung ist bisher unklar, da funktionelle Studien fehlen. Vorläufige Studien haben keine reduzierte Expression oder einen offensichtlichen Funktionsverlust des RFC1-Proteins gezeigt. Diese Beobachtung steht im Widerspruch zu klassischen Mechanismen autosomal-rezessiv vererbter Erkrankungen, die typischerweise durch einen vollständigen oder teilweisen Funktionsverlust eines Proteins verursacht werden. Dies deutet darauf hin, dass der pathogene Mechanismus der Erkrankung weitaus komplexer sein könnte [32].

## 2.7 Spinocerebelläre Ataxien

Unter Spinocerebellären Ataxien (SCAs) werden eine Gruppe fortschreitender neurologischer Erkrankungen zusammengefasst. Diese werden autosomal dominant vererbt und das klinische Hauptmerkmal ist die Ataxie, ein Syndrom der Inkoordination, welches aus der fortschreitenden Degeneration des Kleinhirns entsteht [33]. Obwohl der Name «spinocerebellär» eine Beteiligung des Rückenmarks und des

Kleinhirns nahelegt, ist das Rückenmark bei vielen SCAs nicht betroffen und die pathologischen Veränderungen treten auch in anderen Bereichen des Nervensystems auf. Insgesamt sind über 40 genetisch unterschiedliche Subtypen von SCAs bekannt [34].

Üblicherweise treten die ersten Symptome der SCAs im mittleren Erwachsenenalter auf, können sich aber auch im Kindes- oder hohen Alter manifestieren [34].

Genetisch betrachtet lassen sich die SCAs in zwei Hauptgruppen einteilen. Solche, die durch Repeat-Expansionen verursacht werden und solche, die durch konventionelle Mutationen entstehen [34].

Zu der ersten Gruppe gehören mindestens 12 Subtypen, unter anderem SCA1, SCA2, SCA3, SCA6 und SCA7, welche durch expandierte CAG-Wiederholungen verursacht werden. Diese Wiederholungen kodieren für Glutamine, weshalb diese Gruppe der SCAs auch als Polyglutamin-SCAs bezeichnet werden. Die Polyglutamine sind proteotoxisch, in dem sie abnormale Proteinkonformationen verursachen, welche die Interaktion zwischen Proteinen verändern und zu Oligomerisierung und so zu intranukleären Einschlüssen führen. Diese Einschlüsse wiederum können toxisch sein und weitere zelluläre Prozesse beeinträchtigen. Das Ausmass der Aggregation korreliert mit der Repeatlänge [34].

Weitere Repeat-Expansion-SCAs wie SCA8, SCA10, SCA31, SCA36 und SCA37 werden durch Repeats in nicht kodierenden Regionen oder Introns verursacht und tendieren zu viel grösseren Expansionen. Die Expansionen wirken sich hierbei primär auf die RNA-Ebene aus. Die expandierten RNA-Transkripte akkumulieren in RNA-Foci, die RNA-bindende Proteine sequestrieren und Spleissprozesse stören, was zu Zytotoxizität führt [34].

Weitere molekulare Pathomechanismen der Erkrankung sind die Beeinträchtigung der mitochondrialen Funktion, die Kanalopathie, Autophagie-Störungen oder die transkriptionelle Dysfunktion [33].

Die übrigen SCAs entstehen durch konventionelle Mutationen wie Missense, Nonsense, Insertionen oder Deletionen [34].

Weltweit ist SCA3 die häufigste spinocerebelläre Ataxie, gefolgt von SCA2 und SCA6, wobei die Häufigkeit geografisch und ethnisch oft stark aufgrund von

Gründereffekten variiert [34]. Das kann sich in einer lokalen Dominanz anderer Subtypen zeigen, beispielsweise SCA7 in Venezuela [33].

Die Referenzwerte zusammengefasst aus Genereviews sind in Tabelle 1 dargestellt:

	Gen	Repeat	Normalbereich	Intermediär, veränderlich normaler Bereich	Pathologischer Bereich	Literatur
SCA1	<i>ATXN1</i>	CAG	6 -35 CAG Repeats, bis 44, wenn CAT-Unterbrechungen vorhanden sind	36-38 ohne Unterbrechungen	39-44 ununterbrochen > 44 Repeats	[35]
SCA2	<i>ATXN2</i>	CAG	< 31 Repeats	Variante unklarer klinischer Signifikanz: 32 Repeats	≥ 33 33-34 mit reduzierter Penetranz Häufigste Anzahl: 37-39	[36]
SCA3	<i>ATXN3</i>	CAG	12-44 Repeats	45-59	60-87	[37]
SCA6	<i>CACNA1A</i>	CAG	≤18 Repeats	Variante unklarer Signifikanz: 19	20-33	[38]
SCA7	<i>ATXN7</i>	CAG	7-27 Repeats	28-33	34-36 reduzierte Penetranz 37-460 volle Penetranz	[39]

Tabelle 1 Überblick Referenzwerte Repeatanzahl spinocerebelläre Ataxien

## 2.8 Angeborenes zentrales Hypoventilationssyndrom

Das angeborene zentrale Hypoventilationssyndrom (CCHS) ist eine seltene Erkrankung, die sich durch eine gestörte autonome zentrale Kontrolle der Atmung und eine globale autonome Dysfunktion kennzeichnet. Die Krankheit wird durch Mutationen im *PHOX2B*-Gen verursacht [40]. Dieses Gen codiert für einen Transkriptionsfaktor, welcher die Expression von Genen reguliert, die für die Entwicklung der Neuralleiste entscheidend sind. Aus der Neuralleiste entstehen unterschiedliche Zelltypen und Strukturen, unter anderem auch Teile des vegetativen Nervensystems, welche für die Atemkontrolle und weitere autonome Funktionen verantwortlich sind. Eine Fehlbildung der Neuralleiste im Hirnstamm ist charakteristisch für CCHS [41].

Die häufigste Form der Mutation dabei ist die Polyalanin-Repeat-Expansions-Mutation (PARM), bei der es zu GCN-Expansionen in Exon 3 des *PHOX2B*-Gens kommt. Seltener sind klassische Mutationen, welche unter dem Begriff Nicht-Polyalanin-Repeat-Expansion-Mutation (NPARM) zusammengefasst werden [40].

Die Referenzwerte sind folgende [40]:

Normalallele: 20 GCN (Alanin) Repeats ist die häufigste Anzahl in der Allgemeinbevölkerung detektierte Repeats. Weniger Repeats wurden auch schon berichtet, sind aber nicht mit CCHS assoziiert.

Allele unklarer Signifikanz: Personen mit 21 bis 23 GCN-Repeats wurden bisher nicht mit CCHS beschrieben, deswegen ist die klinische Bedeutung von diesen Repeat-Längen noch nicht bekannt.

Pathologische Allele: Personen mit 24 Repeats und eine Subpopulation von Personen mit 25 Repeats zeigen einen milden Phänotyp mit einem späteren Krankheitsbeginn und/oder Symptomen nur bei Atemnot, Personen mit 26-33 Repeats sowie die meisten mit 25 Repeats zeigen bereits als Neugeborene Symptome.

Zur Unterstützung der Auswertung und des weiteren Verlaufs der Arbeit sind in der Tabelle 2 die Referenzwerte zu den unterschiedlichen Repeat-Expansions-Erkrankungen zusammengefasst.

Gen	Repeat	Anz. Repeats Normbereich	Anz. Repeats Intermediärer Bereich	Anz. Repeats Pathologischer Bereich
<i>FMR1</i>	CGG	< 45	Graubereich: 45 – 54 Prämutation: 55 – 200	> 200
<i>C9orf72</i>	GGGGCC	2 - 24	25 – 60	> 61
<i>FXN</i>	GAA	5 – 33	44 – 65	> 66
<i>HTT</i>	CAG	< 26	27 – 35	> 36
<i>DMPK</i>	CTG	5 – 34	35 – 49	> 50
<i>RFC1</i>	Normal: AAAAG, AAAGG Path. AAGGG	Je nach Repeat		Häufigste AAGGG > 400
SCA1	CAG	6 -35 CAG Repeats, bis 44, wenn CAT-Unterbrechungen vorhanden sind	36-38 ohne Unterbrechungen	39-44 ununterbrochen > 44 Repeats
SCA2	CAG	< 31 Repeats	Variante unklarer klinischer Signifikanz: 32 Repeats	≥ 33 33-34 mit reduzierter Penetranz Häufigste Anzahl: 37-39
SCA3	CAG	12-44 Repeats	45-59	60-87
SCA6	CAG	≤18 Repeats	Variante unklarer Signifikanz: 19	20-33
SCA7	CAG	7-27 Repeats	28-33	34-36 reduzierte Penetranz 37-460 volle Penetranz
<i>PHOX2B</i>	GCN	20	21 – 23	24 – 33

Tabelle 2 Übersicht Referenzwerte Repeatanzahl unterschiedlicher Repeat-Expansions-Erkrankungen

### 3 Fragestellungen

Die in der Einleitung dargestellten Repeat-Expansions-Erkrankungen machen deutlich, welche klinische Relevanz präzise und zuverlässig interpretierbare molekulargenetische Diagnostik hat. Insbesondere die Instabilität der Repeats und die dadurch entstehenden Mosaik, die Heterogenität der Repeat-Längen sowie das Vorkommen komplexer Sequenzmotive, einschliesslich Unterbrechungen, stellen dabei hohe Anforderungen an die traditionellen PCR-basierten Methoden. Diese haben geringe Kosten, eine schnelle Durchführbarkeit und eine hohe Effizienz, stossen jedoch bei der exakten Bestimmung von Repeat-Expansionen sowie in der Detektion von Sequenzmotiven an methodische Grenzen. Hinzu kommt, dass viele Repeat-Expansionen einen hohen GC-Gehalt aufweisen, was die Bildung von Sekundärstrukturen begünstigt und sich in Form von allelic dropout zeigen kann [2]. Diese Limitierung kann zur unvollständigen oder fehlerhaften Darstellung relevanter Allele führen und so das diagnostische Ergebnis beeinflussen.

Vor diesem Hintergrund rückt die Etablierung von alternativen Methoden in den Fokus, welche die komplexen Repeat-Expansions-Strukturen besser und umfassender abbilden könnten. Das Long-Read-Sequencing (LRS) könnte hier ein vielversprechender Ansatz sein. Mit Plattformen wie PacBio oder Oxford Nanopore lassen sich lange Reads generieren, welche allenfalls längere Repeat-Expansionen sowie auch die Sequenzmotive der Repeats abbilden könnten [42].

Aufbauend auf diesen theoretischen Grundlagen ergeben sich im Rahmen dieser Arbeit folgende Fragestellungen:

- Wie zuverlässig kann LRS Repeat-Expansionen identifizieren und deren Länge quantifizieren im Vergleich zu PCR-basierten Methoden?
- Bei welchen Erkrankungen oder Anzahl Repeats stossen die traditionellen Methoden an ihre Grenzen und kann LRS diese Lücken schliessen?
- In welchem Ausmass sind Mosaik mit LRS zuverlässig nachweisbar?
- Für welche Repeat-Expansions-Erkrankungen bietet LRS im Vergleich zu den traditionellen Methoden einen diagnostischen Mehrwert?

Das Ziel dieser Arbeit ist es daher, die Leistungsfähigkeit beider Methoden vergleichend zu evaluieren und das klinische Potenzial von LRS in der Diagnostik von Repeat-Expansions-Erkrankungen zu beurteilen.

## 4 Material und Methoden

### 4.1 Patient\*innenkollektiv

In dieser Arbeit wird auf eine geschlechtergerechte Sprache geachtet. Personenbezeichnungen erfolgen – sofern möglich – in geschlechtsneutraler Form oder unter Verwendung von Doppelnennungen (z. B. „Patientinnen und Patienten“) bzw. Genderformen (z. B. „Patient\*innen“). Sollte aus Gründen der Lesbarkeit in Einzelfällen dennoch eine verkürzte Form verwendet werden, so bezieht sich diese ausdrücklich auf alle Geschlechter gleichermaßen.

Die in dieser Arbeit untersuchten Patientenproben stammen von Personen, die im Rahmen der routinemässigen Diagnostik auf die jeweilige Erkrankung untersucht wurden und für die jeweilige Analyse ihren «informed consent» gaben. Zusätzlich kamen externe Kontrollproben sowie Proben aus Ringversuchen zum Einsatz. Da die Proben im Rahmen der Qualitätskontrolle der Etablierung der neuen Methode genutzt wurden und keine zusätzlichen diagnostischen oder therapeutischen Massnahmen durchgeführt wurden, war kein gesondertes Ethikvotum erforderlich.

Für die Arbeit wurden insgesamt 57 Patient\*innenproben ausgewählt. Das Kollektiv setzt sich wie folgt zusammen:

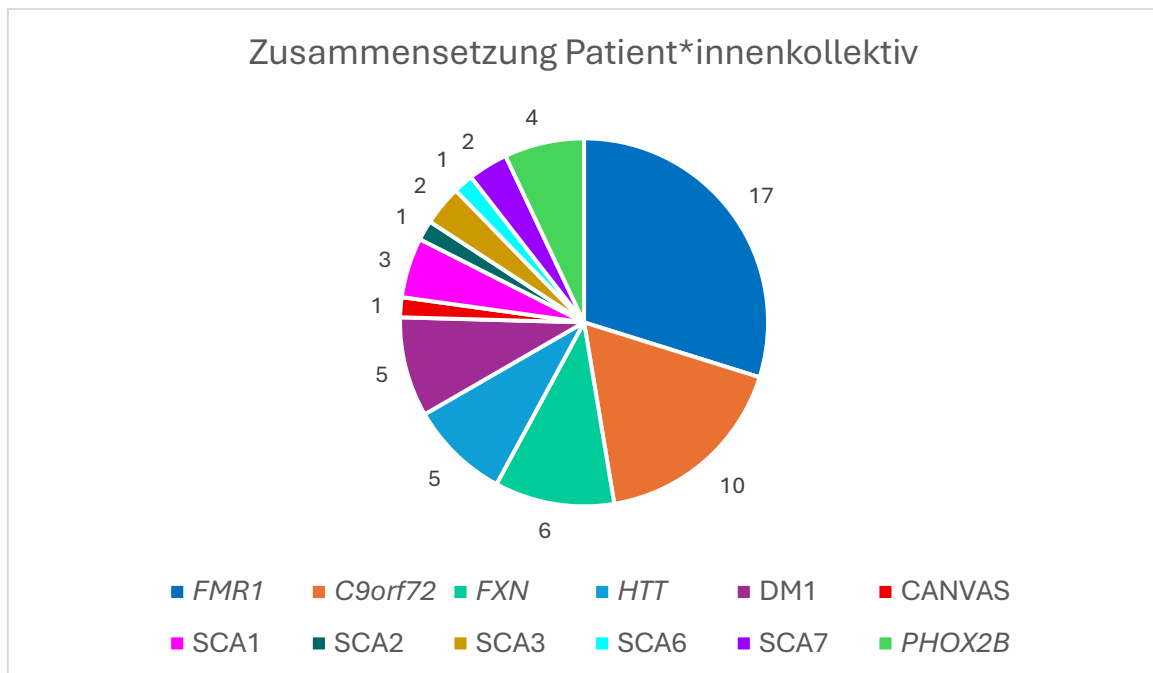


Abbildung 2 Zusammensetzung Patient\*innenkollektiv

Von den 57 Patient\*innen sind 24 weiblich (40.35 %), 23 sind männlich (42.11 %) und bei 10 (17.54 %) gab es keine Angaben bezüglich des Geschlechts. Lediglich bei den auf *FMR1* untersuchten Proben wurde das Geschlecht berücksichtigt, da sich das Gen wie oben beschrieben auf dem X-Chromosom befindet und somit für die Interpretation der Ergebnisse relevant ist.

Der grösste Anteil der Proben (n = 17, 29.82 %) wurde zur Diagnostik *FMR1*-bedingter Erkrankungen untersucht, da diese Methode seit 2011 am Institut etabliert ist und hierfür ein grosser Probenpool vorliegt. Zudem wird dafür jährlich an externen Ringversuchen teilgenommen, weshalb hier einige bereits gut charakterisierte Proben vorlagen.

Die zweit grösste Gruppe (n = 10, 17.54 %) umfasst Proben zur Abklärung von *C9orf72*-bedingten Erkrankungen, wobei ebenfalls eine regelmässige Teilnahme an externen Ringversuchen besteht.

Die Analysen der Gene *HTT*, *DMPK* und *FXN* wurden erst Anfang 2024 im Institut für medizinische Genetik der Universität Zürich eingeführt. Daher standen zum Zeitpunkt der Probenauswahl für das Patient\*innenkollektiv nur wenige interne Fälle zur Verfügung, weshalb hier überwiegend externe Kontrollproben eingesetzt wurden.

Zudem wurde beispielsweise der Grossteil der Analysen zur Abklärung einer *DMPK*-Repeat-Expansion an DNA aus pränatalen Proben durchgeführt. Da dieses Material nur in geringer Menge vorhanden ist und meist nur eine niedrige DNA-Konzentration aufweist, wurde entschieden keine pränatalen DNA-Proben in dieser Arbeit zu verwenden.

Die Diagnostik der SCAs mittels traditioneller PCR-basierter Methoden befindet sich noch in der Etablierungsphase, entsprechend liegen nur wenige Fälle und nur externe Kontrollproben vor. Die Etablierung der CANVAS-Diagnostik ist noch nicht begonnen worden, hierfür wurde daher lediglich eine externe Kontrollprobe verwendet.

Von insgesamt 57 verwendeten Proben stammen 6 von externen und 21 von internen Einsender\*innen. Die übrigen Proben setzen sich aus Ringversuchsproben (n = 16) sowie externe Kontrollproben (n = 14) zusammen, welche vom Institut für medizinische Molekulargenetik zur Verfügung gestellt wurden.



Im Elektropherogramm zeigte sich die Verwendung der drei verschiedenen Primer einerseits durch die Darstellung von PCR-Produkten von vollständigen Allelen sowie Produkten von Repeat-Amplikons, was in einem typischen Muster zu sehen war [50].

Die so generierten Fragmente wurden auf den SeqStudio, SeqStudioFlex oder ABI3730 geladen und anschliessend mit der Genemapper Software (Release 6.1) ausgewertet.

#### 4.1.2 MS-MLPA

Zur Abklärung des Methylierungsstatus und der Kopienzahlbestimmung für *FMR1* wurde zusätzlich das Methylierungssensitive-MLPA (MS-MLPA) Kit ME029 von MRC Holland verwendet. MS-MLPA ist eine semiquantitative Methode, welche auf der Amplifikation von bis zu 60 Proben basiert. Jede dieser Proben detektiert eine spezifische DNA-Sequenz. Zu Beginn werden die Patientenproben denaturiert. Anschliessend wird ein Mix an Proben dazugegeben, welche über Nacht an die DNA hybridisieren. Nach der Hybridisation wird das Reaktionsgemisch auf zwei Röhren aufgeteilt. Beiden wird ein Ligations-Mix hinzugefügt, einem wird jedoch zusätzlich das Restriktions-Enzym HhaI beigefügt. Dieses Enzym ist in der Lage, nicht methylierte Proben zu verdauen. Anschliessend werden die Proben (Abbildung 4) [54].

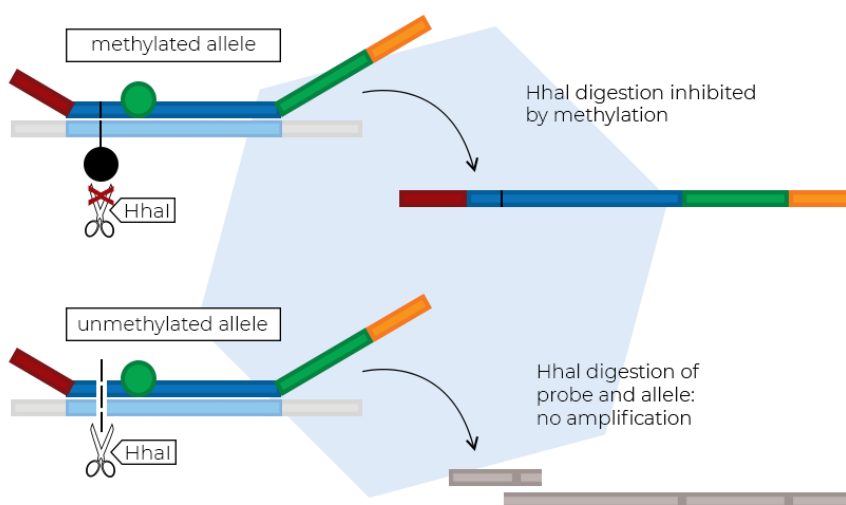


Abbildung 4 schematische Darstellung Prinzip MS-MLPA. Quelle: MRC Holland, Verwendung mit freundlicher Genehmigung

Nach dem letzten Amplifikationsschritt werden die PCR-Produkte mit einem Kapillarelektrophorese-Gerät (ABI3730, SeqStudioFlex, SeqStudio) der Grösse nach aufgetrennt. Das probenspezifische Elektropherogramm wird danach mit der Coffalyser Software (Version 240129.1959) ausgewertet [54].

#### 4.1.3 Sequenzierung nach Sanger

Die Sequenzierung nach Sanger erfolgte gemäss der im Labor etablierten Standardarbeitsanweisung (SOP). Die PCR-Produkte wurden enzymatisch mit Exonuclease I und Alkaliner Phosphatase ausgereinigt und anschliessend unter der Verwendung der jeweiligen Forward- und Reverseprimer mit dem BrilliantDye Terminator kit v3.1 (Nimagen) sequenziert. Die zweite Aufreinigung erfolgte mittels einer Filterplatte und Sephadex G-50 Superfine. Die Produkte wurden anschliessend auf den ABI3730, SeqStudio oder SeqStudioFlex geladen und danach mit der SequencePilot Software (JSi Medical Systems, Version 5.4.4.20) und der SeqA Analysis Software (Applied Biosystems, Version 8) ausgewertet.

#### 4.1.4 Long-Read-Sequencing

Für die Generierung von Long-Reads wurde die High-fidelity-Sequencing (HiFi) Technik von PacBio verwendet. Bei dieser Technik handelt es sich um eine single-molecule, real-time sequencing technology (SMRT), bei der einzelne Moleküle sequenziert werden können. Dazu ist ein Microchip nötig, welcher Millionen von kleinen Wells enthält. In jedes Well gelangt ein einzelnes Molekül, welches dann sequenziert wird [55].

PacBio stellten ihre real-time DNA Sequenziermethode bereits 2009 vor. Bei der Technik wird eine spezielle DNA-Polymerase verwendet, welche fluoreszenz-markierte Nukleotide während der DNA-Synthese einsetzt. Die dabei entstehenden Signale werden in Echtzeit gemessen und so werden lange Reads generiert, welche eine Grösse von bis zu 2.5kb haben können [56].

Die SMRT-Technologie beruht darauf, dass einzelne Nukleotide während ihrer Inkorporation durch eine DNA-Polymerase in Echtzeit detektiert werden. Es wird dabei eines von vier unterschiedlich fluoreszenzmarkierten Nukleotiden eingebaut, dessen Signal in einem Zero-Mode-Waveguide (ZMW) erfasst wird. In jedes ZMW gelangt ein sogenanntes SMRTbell-Template, ein zirkuläres DNA-Molekül, das

durch Ligation von Hairpin-Adaptoren an doppelsträngige DNA entsteht (Abbildung 5) [57].

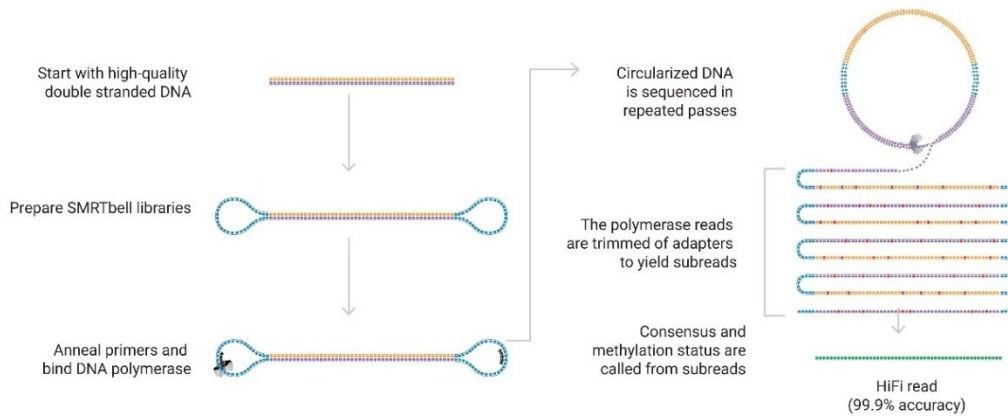


Abbildung 5 Sequenziertechnologie Long-Read-Sequencing PacBio. Quelle: PacBio, Verwendung mit freundlicher Genehmigung

Der kleine Reaktionsraum im ZMW ermöglicht eine hochempfindliche Detektion der Fluoreszenzsignale. Die durchgehende Überwachung der Einbaugeschwindigkeit erlaubt darüber hinaus die Identifikation von epigenetischen Modifikationen, da diese Interpulsdauer und Pulsbreite verändern [42].

Frühere Versionen der SMRT-Technologie arbeiteten mit continuous long reads (CLR). Dabei wurden Insertgrößen von bis zu 100kb nur wenige Male sequenziert, was zu relativ hohen Fehlerraten (ca. 85-92 %) führte [42].

Die Einführung der circular consensus sequencing Technologie, die auch als HiFi-Sequenzierung bezeichnet wird, stellte eine wesentliche Verbesserung dar. Hier werden durch eine vorangehende Grössenselektion generierte Fragmente von 15 bis 20kb sequenziert und das bis zu 12-mal nacheinander. Die daraus resultierenden Subreads werden rechnerisch zu einer Konsensussequenz zusammengeführt, wodurch eine Genauigkeit von bis zu 99.9 % erreicht werden kann [2,42].

Die DNA-Proben für Long-Read-Sequencing wurden zuerst mit dem «Short Read Eliminator Kit» von PacBio behandelt. Dies um Fragmente, welche < 10kb sind, zu entfernen und so hochmolekulare DNA zu erhalten [58]. Anschliessend wurde die Library Preparation nachfolgenden Protokoll von Pacbio durchgeführt: «Preparing whole genome and metagenome libraries using SMRTbell® prep kit 3.0». Die Library

Preparation erfolgte automatisiert auf dem Biomek i7 Pipettierroboter. Anschließend wurden die Proben mit dem Revo System sequenziert.

Für die Auswertung der Daten wurde die VarSeq Software (Golden Helix, Version 3.0.0 RC5) verwendet.

## 5 Resultate

### 5.1 Resultate traditionelle Methoden

Die Übersicht der Rohdaten aller Patient\*innen findet sich im Anhang (Tabelle 18).

#### 5.1.1 TP-PCR *FMR1*

Pat. Nr.	Ge- schlecht	Anzahl Repeats Allel 1	Anzahl Repeats Allel 2	AGG-Unter- brechungen ersichtlich?
C01-XD01	w	12 +/-1	90 +/-5	nein
C02-XD02	m	> 200	-	nein
C03-XD03	m	> 200	-	nein
C04-XD04	m	Mosaik zwischen Prämutation/Vollmutation ca. 186 – > 200 Repeats		nein
C05-XD05	m	Mosaik zwischen Prämutation/Vollmutation ca. 186 – > 200 Repeats		nein
C06-XD06	w	31 +/- 1	89 +/- 5	ja
C07-XD07	m	> 200	-	nein
C08-XD08	m	> 200	-	nein
C09-XD09	w	32 +/-1	> 200	ja
C10-XD10	m	> 200	-	nein
C11-XD11	m	Mosaik zwischen Prämutation/Vollmutation ca. 165, 195 und > 200 Repeats		nein
C12-XD12	m	Mosaik zwischen Prämutation/Vollmutation ca. 92, 135 und > 200 Repeats		nein
C13-XD13	m	Mosaik Prämutation 56 – 75 Repeats		nein
C14-XR14	m	Mosaik Prämutation 57, 77, 92 repeats		nein
C15-XR15	w	31 +/-1	> 200	ja
C16-XD16	w	20 +/-1	> 200	ja
C17-XD17	w	23 +/-1	Mosaik Prämutation/Vollmu- tation 176, > 200 Repeats	ja

Tabelle 4 Resultate TP-PCR *FMR1*, w: weiblich, m: männlich

Insgesamt wurden 11 Proben von männlichen Patienten und 6 Proben von weiblichen Patientinnen analysiert (Tabelle 4).

Von den männlichen Proben zeigten 5 eine hemizygoter Vollmutation des *FMR1*-Gens mit > 200 Repeats. 4 Proben zeigten ein Mosaik zwischen Prämutation und Vollmutation, wobei die Grenzen der unterschiedlichen Mosaiken nicht klar darstellbar waren. 2 der Proben zeigten ein Mosaik im Prämutationsbereich, was sich durch mehrere Peaks im Elektropherogramm zeigte (Abbildung 6).

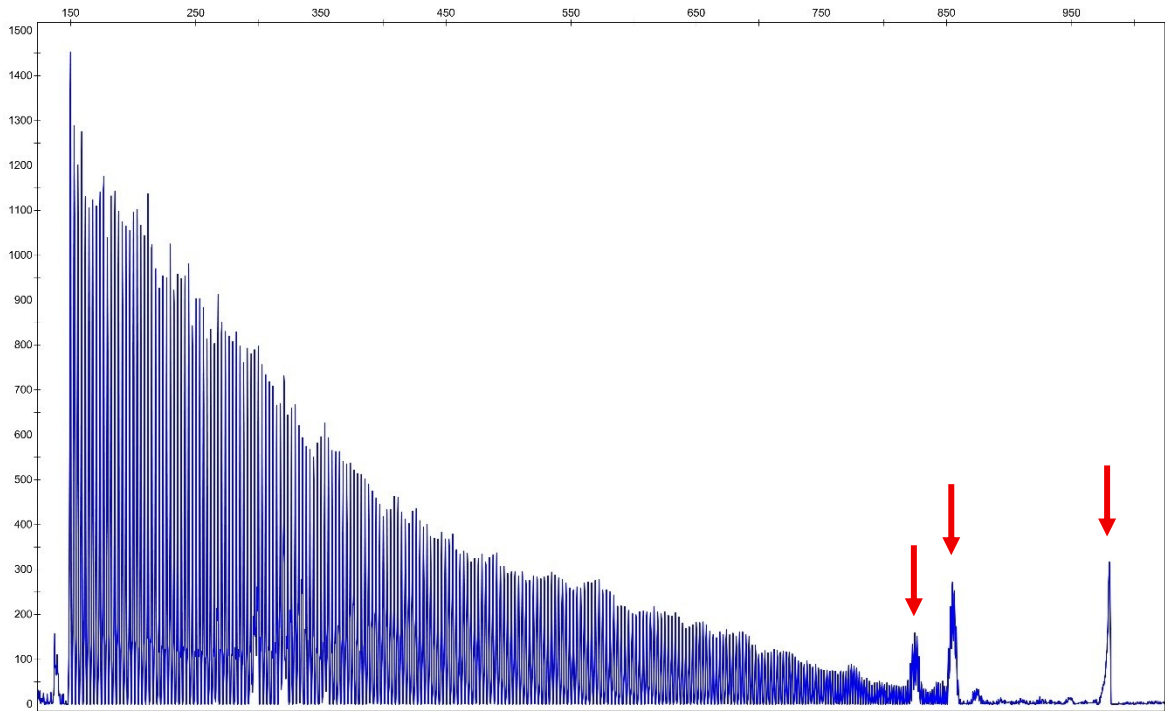


Abbildung 6 Elektropherogramm TP-PCR Probe C04-XD04. Mosaik sichtbar durch drei verschiedene Peaks bei ca. 820bp, 850bp und 980bp, X-Achse: Fragmentgröße, Y-Achse: Fluoreszenzintensität

Bei den weiblichen Proben waren bei allen zwei Peaks von unterschiedlichen Längen erkennbar. 2 der Proben zeigten jeweils ein Allel im Normbereich und ein Allel im Prämutationsbereich. Bei 3 der Proben fand sich ein Allel im Normbereich und eines im Vollmutationsbereich > 200 Repeats. Eine Probe zeigte ein Allel im Normbereich und ein Allel im Mosaik Prämutation zu Vollmutation, wobei auch hier die Grenzen der verschiedenen Mosaik nicht klar ersichtlich waren.

Da die Anzahl und Lokalisation der AGG-Unterbrechungen mit dieser Methode nur begrenzt abgebildet werden kann, wurde nur angegeben, ob diese vorhanden. Es folgte kein genauere Beschrieb über deren Lage oder Anzahl.

Bei 5 der untersuchten Proben gab es einen Hinweis auf eine AGG-Unterbrechung. Dies zeigte sich durch ein verringertes Fluoreszenzsignal im CGG-Repeat-Profil, welches als mehrere Stutter-Peaks im Elektropherogramm sichtbar ist (Abbildung 7).

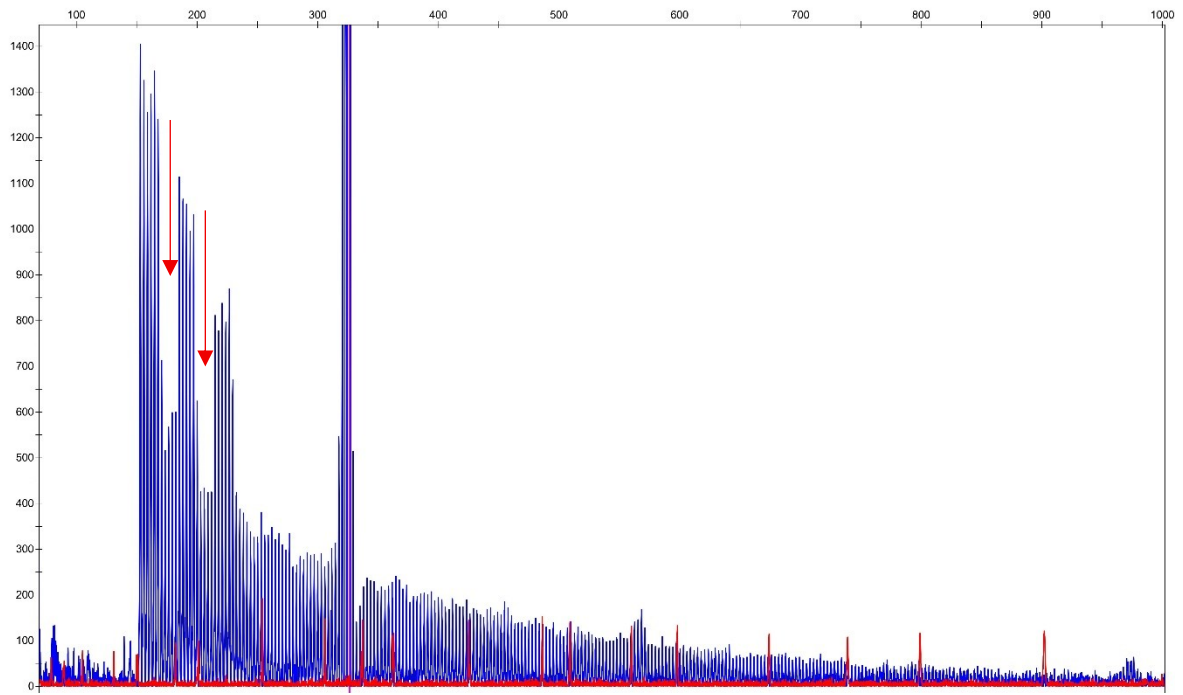


Abbildung 7 Elektropherogramm TP-PCR Probe C15-XR15. Blau: Hinweis auf AGG-Unterbrechung sichtbar durch Unterschiede in der Fluoreszenzintensität, rot: Size Standard ROX 1000, X-Achse: Fragmentgröße, Y-Achse: Fluoreszenzintensität

### 5.1.2 TP-PCR *C9orf72*

Pat. Nr.	Anzahl Repeats Allel 1	Anzahl Repeats Allel 2
C18-CD01	10	Mosaik 45 bis Expansion
C19-CD02	10	> 145
C20-CD03	2	> 145
C21-CR04	8	> 145
C22-CR05	2	> 145
C23-CR06	2	> 145
C24-CR07	5	> 145
C25-CD08	2	12
C26-CD09	2	8
C27-CD10	6	8

Tabelle 5 Resultate TP-PCR *C9orf72*

Von den insgesamt 10 analysierten Proben wiesen 7 ein Allel im Normbereich und ein expandiertes Allel auf (Abbildung 8). Bei einer Probe zeigte sich ein Allel im Normbereich und ein Allel im Mosaikstatus von 45 Repeats bis zur Expansion. 3 der Proben zeigten zwei Allele im Normbereich (Tabelle 5). Mit dem verwendeten Kit können nur Allele bis 145 Repeats angegeben werden, Allele mit mehr Repeats werden lediglich als > 145 angegeben [44].

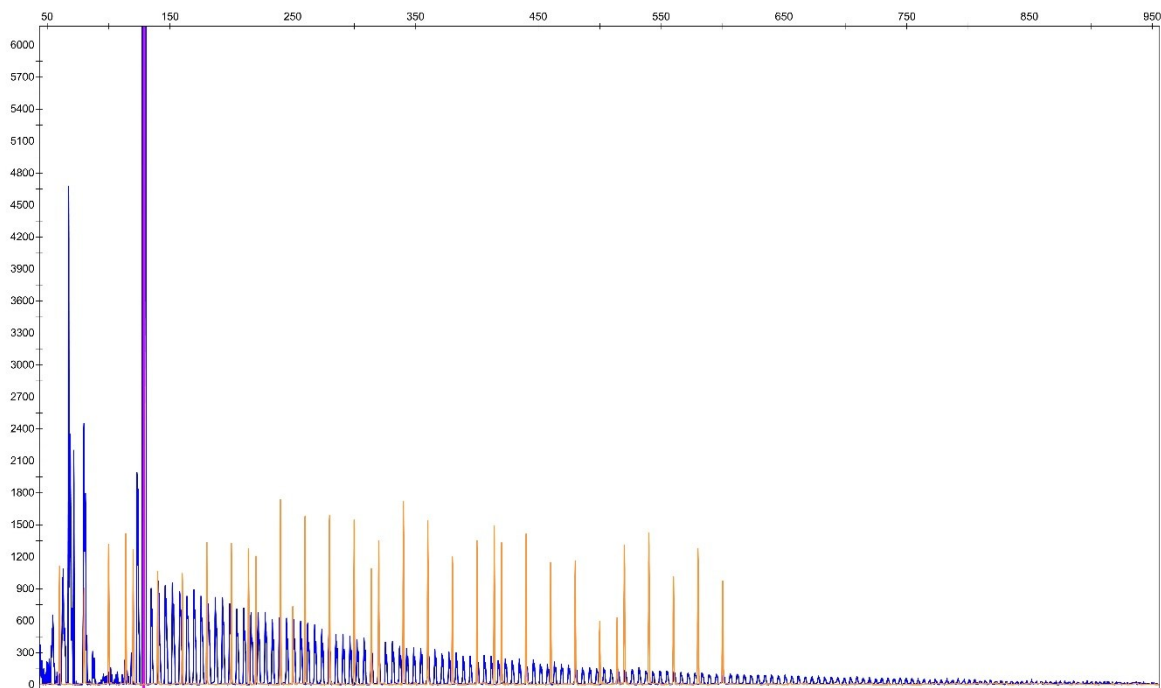


Abbildung 8 Elektropherogramm TP-PCR Probe C22-CR05. Blau: Normalallel bei ca. 130 bp, expandiertes Allel durch Repeat-Amplifikons sichtbar, orange: Size Standard LIZ600, X-Achse: Fragmentgrösse, Y-Achse: Fluoreszenzintensität

### 5.1.3 TP-PCR und Fragmentanalyse FXN

Pat. Nr.	Anzahl Repeats Allel 1	Anzahl Repeats Allel 2	
C28-FK01	8	> 66	Normalallele mit Sangersequenzierung überprüft
C29-FK02	> 66	> 66	
C30-FW03	9	> 66	Normalallele mit Sangersequenzierung überprüft
C31-FW04	> 66	> 66	
C32-FR05	> 66	> 66	
C33-FR06	9	16	Normalallele mit Sangersequenzierung überprüft

Tabelle 6 Resultate TP-PCR und Fragmentanalyse FXN

Von den 6 untersuchten Proben zeigen zwei ein Allel im Normbereich und ein expandiertes Allel, 3 Proben zeigen 2 expandierte Allele (Abbildung 9) und eine Probe zeigt zwei Normalallele mit 9 bzw. 16 Repeats (Tabelle 6).

Die expandierten Allele wurden dabei nur als expandiertes Allel > 66 Repeats angegeben, da mit der konventionellen Methode nur Fragmentgrößen bis 1000bp dargestellt werden können.

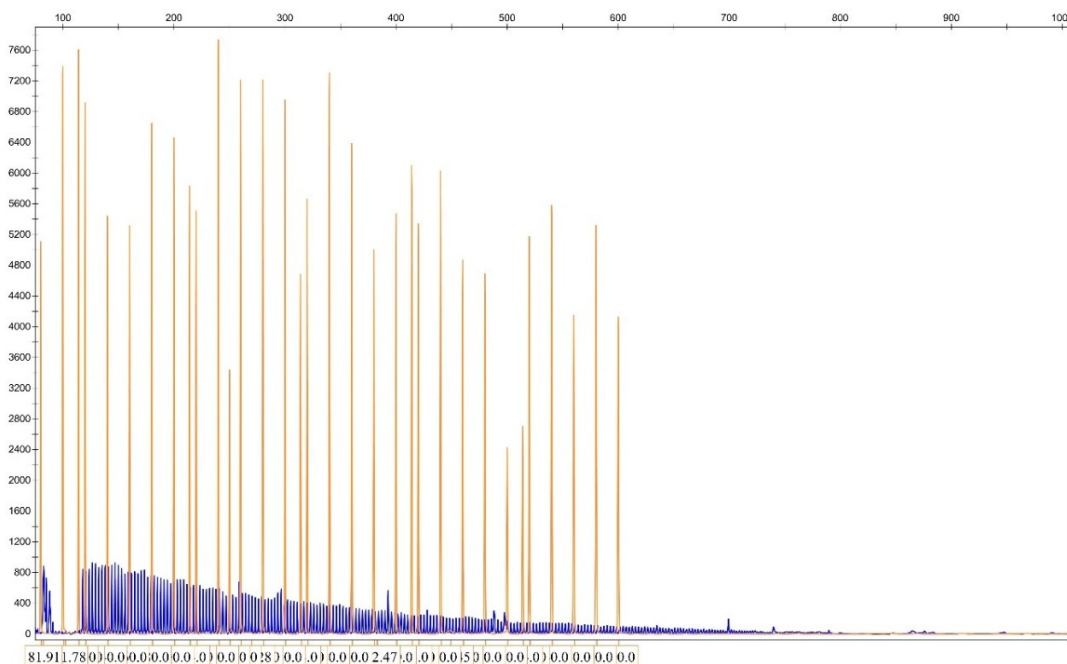


Abbildung 9 Elektropherogramm TP-PCR Probe C29-FK02. Blau: zwei expandierte Allele durch Repeat-Amplifikons sichtbar, orange: Size Standard LIZ600, X-Achse: Fragmentgröße, Y-Achse: Fluoreszenzintensität

#### 5.1.4 Fragmentanalyse und Sequenzierung *HTT*

Pat. Nr.	Anzahl Repeats Allel 1	Anzahl Repeats Allel 2
C34-HD01	18	41
C35-HK02	18	43
C36-HW03	19	47
C37-HR04	17	38
C38-HR05	17	20

*Tabelle 7 Resultate Fragmentanalyse und Sequenzierung HTT*

Von den 5 analysierten Proben zeigten 3 ein Allel im Normbereich und ein Allel im pathologischen Bereich. 2 Proben zeigten zwei Allele im Normbereich (Tabelle 7). Da die Fragmentanalyse die Expansionen im untersuchten Grössenbereich ausreichend darstellen konnte, wurde hier auf die Durchführung einer zusätzlichen TP-PCR verzichtet. Jedoch wurden die Resultate der Fragmentanalyse mit Sangersequenzierung überprüft.

#### 5.1.5 TP-PCR und Fragmentanalyse *DMPK*

Pat. Nr.	Anzahl Repeats Allel 1	Anzahl Repeats Allel 2
C39-D1W01	11	> 50
C40-D1W02	13	> 50
C41-D1D03	13	> 50
C42-D1R04	20	> 50
C43-D1R05	13	12

*Tabelle 8 Resultate TP-PCR und Fragmentanalyse DMPK*

Von den 5 analysierten Proben zeigten 4 ein Allel im Normbereich und ein expandiertes Allel. 1 Probe zeigte zwei Allele im Normbereich mit 13 bzw. 12 Repeat (Tabelle 8).

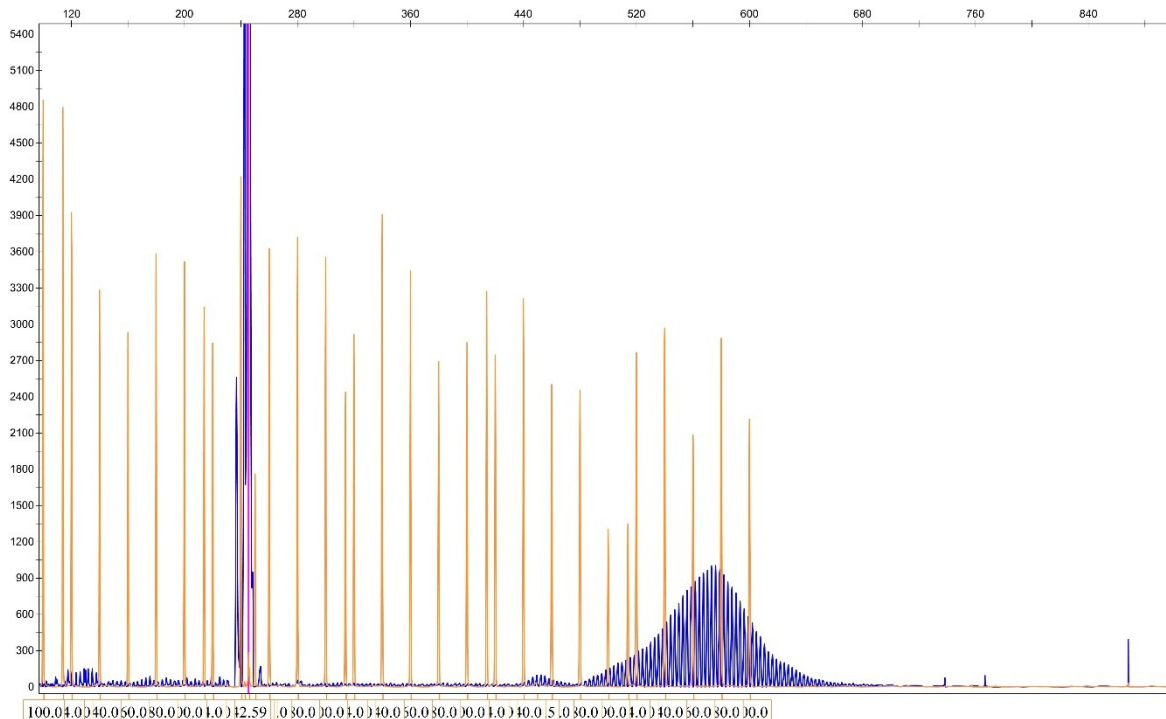


Abbildung 10 Elektropherogramm TP-PCR Probe C39-D1W01. Blau: Allel 1 im Normbereich bei 245bp, Allel 2 expandiert durch Repeat-Amplikons sichtbar, orange: Size Standard LIZ600, X-Achse: Fragmentgrösse, Y-Achse: Fluoreszenzintensität

### 5.1.6 Resultate RFC1

Für die Probe C44-CSK01 lagen keine mittels traditionellen Methoden generierten Daten vor, sondern lediglich schriftlich dokumentierte Befundwerte. Diese zeigten ein Allel mit einer Punktmutation im RFC1-Gen und ein Allel mit einer Repeat-Expansion, wobei die genaue Repeatanzahl nicht dokumentiert wurde.

### 5.1.7 TP-PCR und Fragmentanalyse SCAs

Im Rahmen der laufenden Etablierung der SCAs wurden SCA1, SCA2, SCA3, SCA6 und SCA7 getestet. Es wurde bei allen eine Fragmentanalyse durchgeführt, bei SCA2 und SCA7 wurde zusätzlich noch eine TP-PCR gemacht.

Pat. Nr.	Gen	Anzahl Repeats Allel 1	Anzahl Repeats Allel 2
C45-S1W01	ATXN1 (SCA1)	31 +/- 1	40 +/- 3
C46-S1R02	ATXN1 (SCA1)	31 +/- 1	43 +/- 3
C47-S1R03	ATXN1 (SCA1)	29 +/- 1	52 +/- 3
C48-S2W01	ATXN2 (SCA2)	22 +/- 1	37 +/- 3
C49-S3W01	ATXN3 (SCA3)	24 +/- 1	75 +/- 3
C50-S3R02	ATXN3 (SCA3)	23 +/- 1	70 +/- 3
C51-S5W01	CACNA1A (SCA6)	13 +/- 1	22 +/- 1
C52-S7W01	ATXN7 (SCA7)	7 +/- 1	55 +/- 3
C53-S7R02	ATXN7 (SCA7)	7 +/- 1	56 +/- 3

Tabelle 9 Resultate TP-PCR und Fragmentanalyse SCAs, ausgewertet nach [59].

Die Resultate wurden anschliessend nach den «EMQN Best Practice Guideline for molecular genetic testing of SCAs» ausgewertet.

Die untersuchten Proben zeigten alle ein Allel im Normbereich und ein Allel im pathologischen Bereich. Mit den durchgeführten Analysen konnte keine qualitative Aussage über die Repeats getroffen werden, Unterbrechungen in den Repeats wurden nicht detektiert (Tabelle 9).

### 5.1.8 Fragmentanalyse und Sequenzierung *PHOX2B*

Pat. Nr.	Anzahl Repeats Allel 1	Anzahl Repeats Allel 2
C54-PD01	20	20
C55-PD02	20	27
C56-PD03	20	20
C57-PD04	20	25

Tabelle 10 Resultate Fragmentanalyse und Sequenzierung *PHOX2B*

2 der Proben zeigten zwei Allele im Normbereich, die weiteren 2 Proben zeigten jeweils ein Allel im Normbereich und eines im pathologischen Bereich (Tabelle 10).

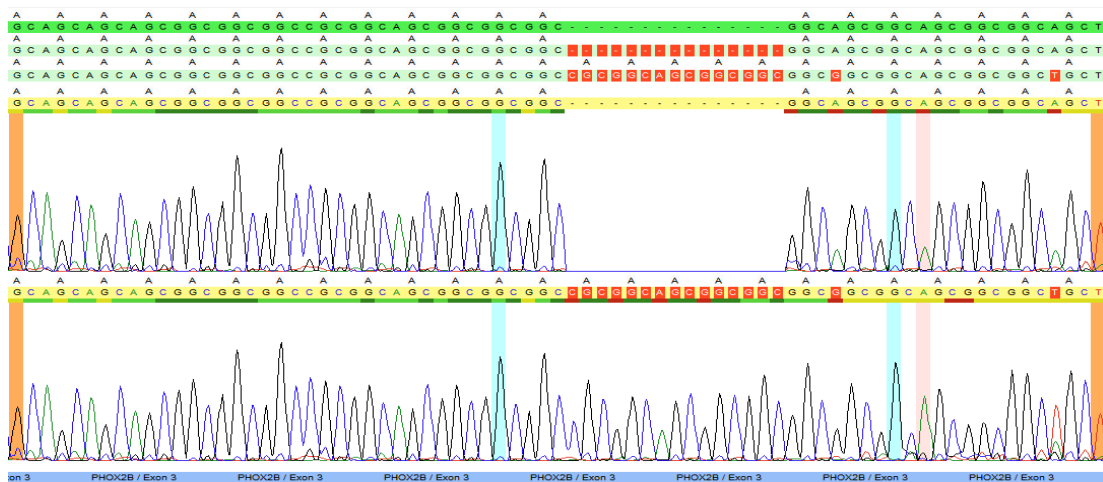


Abbildung 11 Resultat Sequenzierung C57-PD04, Darstellung mit der Seqpilot Software. Durch die Funktion «split sequence» können die beiden unterschiedlich grossen Allele angezeigt werden: in rot erscheinen die zusätzlichen Basen, die nur bei einem Allel vorkommen.

## 5.1.9 MS-MLPA *FMR1*

Pat. Nr.	Geschlecht	Resultate Methylierung
C04-XD04	m	Partielle Hypermethylierung <i>FMR1</i>
C05-XD05	m	Hypermethylierung <i>FMR1</i>
C07-XD07	m	Hypermethylierung <i>FMR1</i>
C08-XD08	m	Hypermethylierung <i>FMR1</i>
C09-XD09	w	Normwerte für die untersuchten Zielregionen
C10-XD10	m	Hypermethylierung <i>FMR1</i>
C11-XD11	m	Hypermethylierung <i>FMR1</i>
C12-XD12	m	Hypermethylierung <i>FMR1</i>
C13-XD13	m	Normwerte für die untersuchten Zielregionen
C14-XR14	m	Normwerte für die untersuchten Zielregionen
C15-XR15	w	Normwerte für die untersuchten Zielregionen
C16-XD16	w	Normwerte für die untersuchten Zielregionen

Tabelle 11 Resultate MS-MLPA *FMR1*

Von den 12 analysierten Proben zeigten 5 Proben Normwerte für die untersuchten Zielregionen, was einem normalen Methylierungsmuster entspricht. 6 Proben zeigten eine Hypermethylierung an *FMR1*. Eine Probe (C04-XD04) zeigte eine partielle Methylierung des *FMR1* (Tabelle 11).

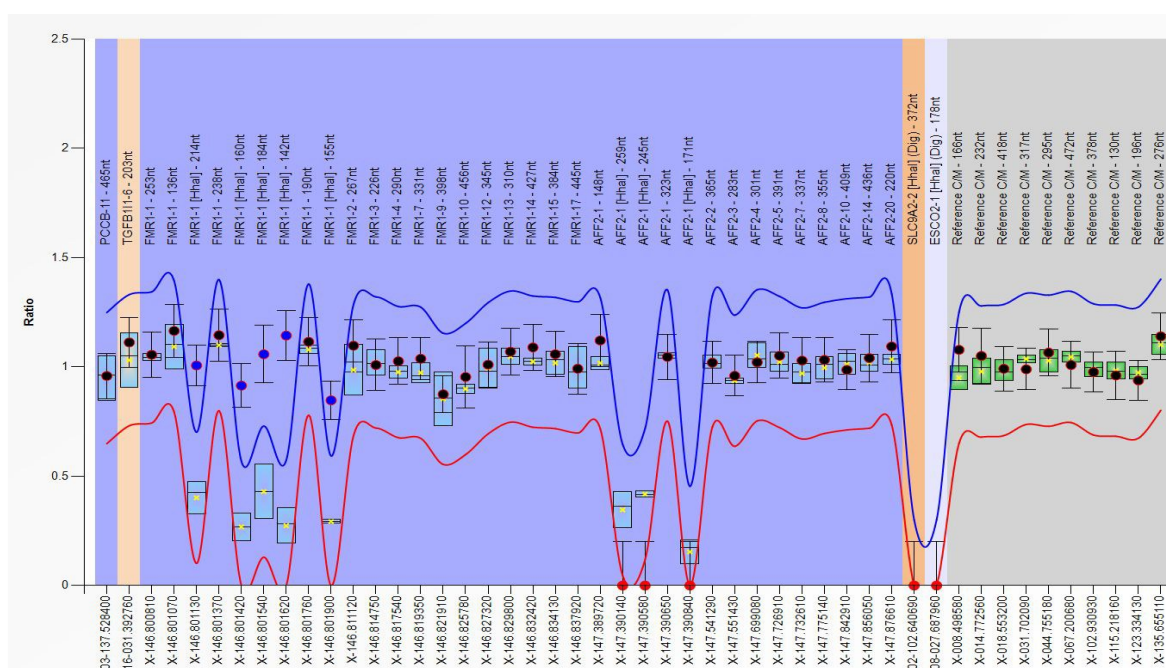


Abbildung 12 Resultat MS-MLPA der Probe C05-XD05, Ausschnitt aus der Coffalyser Software, erhöhte Werte in blau im Sinne einer Hypermethylierung bei den HhaI-sensitiven Proben für *FMR1*, X-Achse: Name/Position der getesteten DNA-Sonden, Y-Achse: relative Signalintensität der Probe im Vergleich zum durchschnittlichen Signal normaler Referenzproben.

## 5.2 Resultate Long-Read-Sequencing

### 5.2.1 Übersicht Resultate

Mit dem Long-Read-Sequencing wurden folgende 17 Proben analysiert:

Pat. Nr.	Ge- schlecht	Gen	Anzahl Reads Allel 1/Allel 2	Anzahl Repeats Allel 1	Anzahl Repeats Allel 2	Mittlerer Methylierungsgrad	Yield (Gb)	Mean length (Kb)	Median read quality	Mean coverage (Anzahl Reads)
C01-XD01	w	<i>FMR1</i>	24/10	11 CGG	91 CGG	0.31, 0.73	108.4	16.1	37	34.66
C02-XD02	m	<i>FMR1</i>	4	447 CGG, 1 AGG		0.74	90	16.4	33	28.72
C04-XD04	m	<i>FMR1</i>	11	249 CGG		0.51	101.1	17.9	33	32.3
C05-XD05	m	<i>FMR1</i>	9	387 CGG		0.83	102.9	15.4	35	32.9
C06-XD06	w	<i>FMR1</i>	28/22	29 CGG, 2 AGG	90 CGG	0.53, 0.56	114.7	17.2	32	36.69
C07-XD07	m	<i>FMR1</i>	8	362 CGG		0.77	121	13.7	37	38.68
C08-XD08	m	<i>FMR1</i>	11	269 CGG		0.87	96.8	10.4	38	30.88
C09-XD09	w	<i>FMR1</i>	21/15	30 CGG, 1 AGG	236 CGG	0.27, 0.89	96	9.7	39	30.71
C10-XD10	m	<i>FMR1</i>	13	461 CGG		0.82	87.3	12	36	27.9
C17-XD17	w	<i>FMR1</i>	24/19	22 CGG, 1 AGG	183 CGG, 1 AGG	0.59, 0.45	111	14.7	36	35.54
C20-CD03	w	<i>C9orf72</i>	7/7	2 GGCCCC	2 GGCCCC	-	77.7	15.3	36	24.84
C28-FK01	w	<i>FXN</i>	16/1	8 GAA	720 GAA	-	103.6	15	34	33.15
C29-FK02	w	<i>FXN</i>	6/7	346 GAA	408 GAA	-	94.1	14.4	35	30.09
C34-HD01	m	<i>HTT</i>	10/21	18 CAG	42 CAG	-	128.5	16.7	34	41.05
C35-HD02	w	<i>HTT</i>	12/13	18 CAG	43 CAG	-	118.3	15.1	36	37.87
C41-D1D03	w	<i>DMPK</i>	12/8	13 CTG	261 CTG	-	131.9	17.5	34	42.22
C44-CSK01	w	<i>RFC1</i>	17/1	10 AAAAG	437 AAGGG	-	108.4	16.1	37	36.67

Tabelle 12 Übersicht Resultate LRS. w: weiblich, m: männlich. Es wurden nur die Hauptmotive der jeweiligen Erkrankungen berücksichtigt.

### 5.2.2 Überblick generelle Qualität der Resultate

Von den insgesamt 17 Proben wurden 10 auf *FMR1*, 1 auf *C9orf72*, 2 auf *FXN*, 2 auf *HTT*, 1 auf *DM1* und eine auf *RFC1* untersucht. Die geringste Anzahl Reads pro Allel innerhalb der jeweiligen Zielregion wurde in der Probe C28-FK01 (*FXN*) mit 1 detektiert, während die höchste Anzahl in Probe C06-XD06 (*FMR1*) mit 28 verzeichnet wurde. Der längste sequenzierte Read wird mit 3240bp angegeben und wurde bei der Probe C44-CSK01 (*RFC1*) generiert.

Die geringste durchschnittliche Read-Länge mit 9.7kb wird bei der Probe C09-XD09 detektiert, die höchste bei der Probe C04-XD04 mit 17.9 Kb. Die durchschnittliche Read-Länge über alle Proben beträgt 14.9 Kb.

Der geringste Yield wurde bei der Probe C20-CD03 (*C9orf72*) mit 77.7 Gb beobachtet, wohingegen die Probe C34-HD01 (*HTT*) mit 128.5 Gb den grössten Wert aufwies. Der durchschnittliche Yield über alle 17 Proben beträgt 105.4 Gb.

Die höchste durchschnittliche Read-Qualität zeigt die Probe C09-XD09 mit 39, der niedrigste Wert mit 32 weist die Probe C06-XD06 auf. Die durchschnittliche Read-Qualität beträgt 35.4.

Die höchste Coverage wird bei der Probe C41-D1D03 mit 42.22 erreicht, die niedrigste mit 24.84 zeigt die Probe C20-CD03. Die durchschnittliche Coverage beträgt 33.8.

Von den 17 untersuchten Proben wurde bei 6 Proben nur ein Allel nachgewiesen, was mit dem männlichen Geschlecht der Patienten und dem hemizygoten Charakter der untersuchten Zielregion *FMR1* auf dem X-Chromosom übereinstimmt.

### 5.2.3 Waterfall-Plots

Für jede der Proben wurde zudem mit dem «Tandem repeat genotyping tool» (TRGT, Version 5.0.0) ein Waterfall-Plot erstellt, um die generierten Reads bildlich darzustellen [60]. Auf der X-Achse wurde dabei jeweils die Read-Länge angezeigt, auf der Y-Achse die Anzahl der Reads. Dabei zeigte sich, dass sich die angegebene Repeatanzahl und die tatsächliche Repeat-Länge zum Teil stark unterscheiden. Zur Veranschaulichung wurden exemplarisch Plots von drei Proben herangezogen. Die Plots weiterer Proben finden sich im Anhang (10.3 Waterfall-Plots).

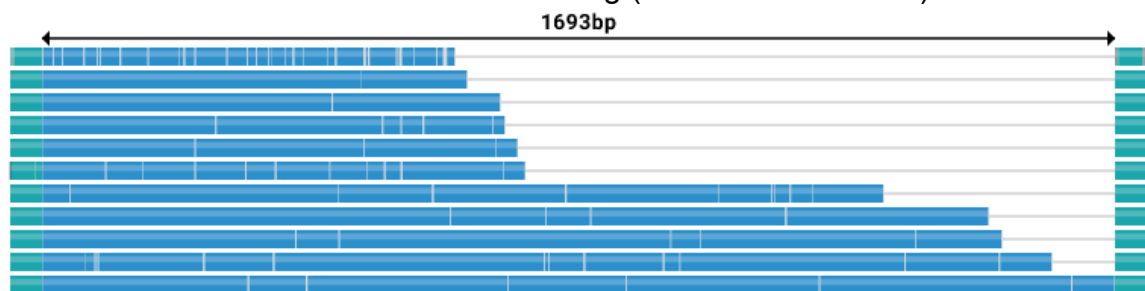


Abbildung 13 Waterfall-Plot C04-XD04

Bei der Probe C04-XD04 (*FMR1*) wurden von der Software durchschnittlich 249 Repeats extrapoliert, in Abbildung 13 Waterfall-Plot C04-XD04 ist aber klar zu sehen, dass von den 11 angegebenen Reads die meisten kürzer sind und dementsprechend auch weniger Repeats aufweisen.

Das von den traditionellen Methoden nachgewiesene Mosaik ist zwar als solches im Plot ersichtlich, wird aber von der Software nicht ausgegeben.

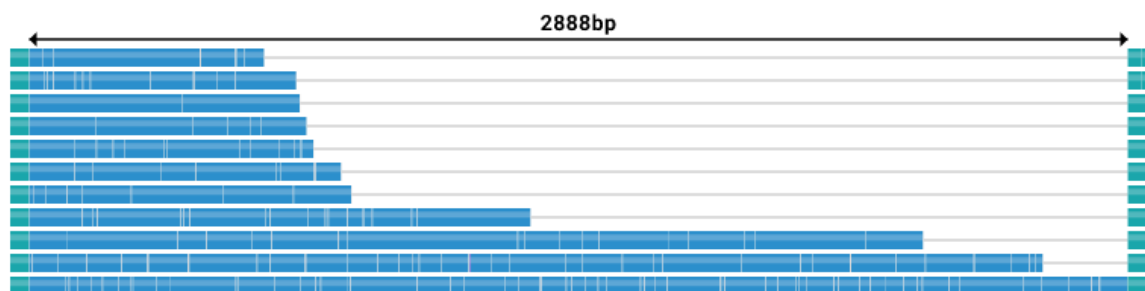


Abbildung 14 Waterfall-Plot C08-XD08

Bei der Probe C08-XD08 (*FMR1*) wurde mit den traditionellen Methoden lediglich eine Expansion >200 Repeats nachgewiesen. Im Plot zeigt sich jedoch hier (Abbildung 14), dass von den 11 angegebenen Reads nur einer die volle Expansion von 362 Repeats aufweist. Die weiteren Reads sind kürzer und könnten auf ein mögliches Mosaik bei diesem Patienten hindeuten.

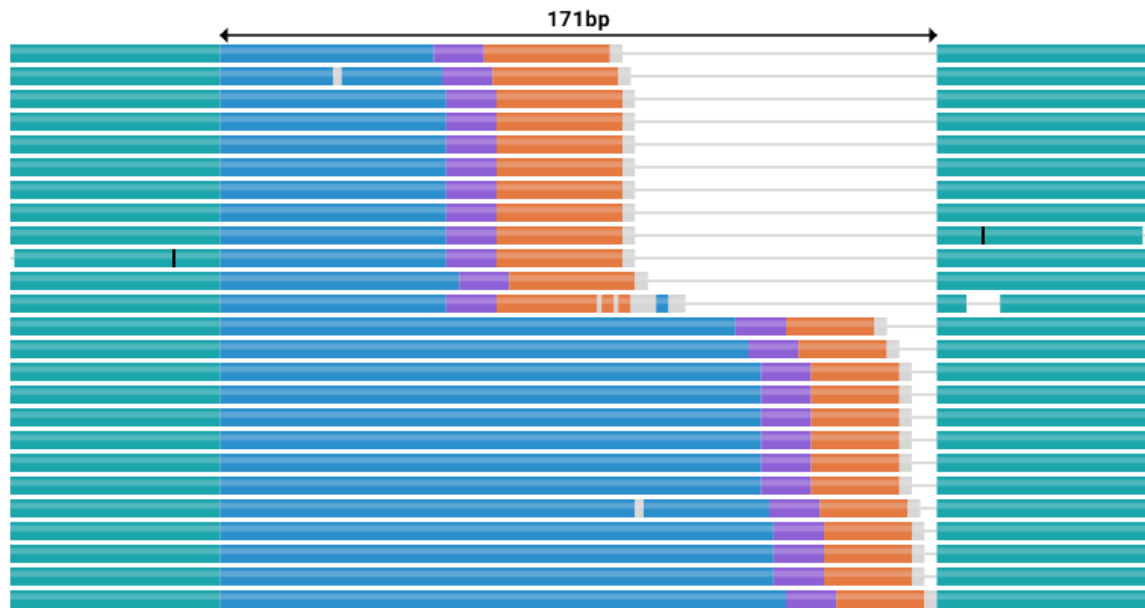


Abbildung 15 Waterfall-Plot C35-HK02

Bei der Probe C35-HK02 (*HTT*) stimmt die Anzahl Reads grösstenteils mit der Anzahl der Repeats überein. Dennoch zeigen einzelne Reads Lücken oder Abweichungen in der Sequenz (Abbildung 15).

#### 5.2.4 Darstellung allelischer Bias

Pat. Nr.	Gen	Anz. Reads Allel 1/2	Abweichung in Anz. Reads	Abweichung in Prozent
C01-XD01	<i>FMR1</i>	24/10	14	58.33
C06-XD06	<i>FMR1</i>	28/22	6	21.43
C09-XD09	<i>FMR1</i>	21/15	6	28.57
C17-XD17	<i>FMR1</i>	24/19	5	20.83
C20-CD03	<i>C9orf72</i>	7/7	0	0.00
C28-FK01	<i>FXN</i>	16/1	15	93.75
C29-FK02	<i>FXN</i>	6/7	1	16.67
C34-HD01	<i>HTT</i>	21/10	11	52.38
C35-HD02	<i>HTT</i>	12/13	1	8.33
C41-D1D03	<i>DMPK</i>	12/8	4	33.33
C44-CSK01	<i>RFC1</i>	17/1	16	94.12

Tabelle 13 Abweichung Anzahl Reads Allel 1 zu Allel 2

Bis auf eine Probe bestanden bei allen eine Diskrepanz zwischen Anzahl Reads von Allel 1 zu Allel 2. Die grösste Abweichung zeigte sich dabei bei der Probe C44-CSK01 (*RFC1*), bei der die Abweichung 94.12 % beträgt (Tabelle 12).

## 5.3 Vergleich Resultate beider Methoden

### 5.3.1 Vergleich Anzahl Repeats

Pat. Nr.	Gen	Anz. Repeats mit traditionellen Methoden		Anz. Repeats mit LRS	
		Allel 1	Allel 2	Allel 1	Allel 2
C01-XD01	<i>FMR1</i>	12	90	11	90
C02-XD02	<i>FMR1</i>	>200		447	
C04-XD04	<i>FMR1</i>	186- >200		249	
C05-XD05	<i>FMR1</i>	186- >200		387	
C06-XD06	<i>FMR1</i>	31	89	29	90
C07-XD07	<i>FMR1</i>	>200		362	
C08-XD08	<i>FMR1</i>	>200		269	
C09-XD09	<i>FMR1</i>	32	>200	30	236
C10-XD10	<i>FMR1</i>	>200		461	
C17-XD17	<i>FMR1</i>	23	176, >200	22	183
C20-CD03	<i>C9orf72</i>	2	>145	7	7
C28-FK01	<i>FXN</i>	8	> 66	8	720
C29-FK02	<i>FXN</i>	> 66	> 66	346	408
C34-HD01	<i>HTT</i>	18	41	18	42
C35-HD02	<i>HTT</i>	18	43	18	43
C41-D1D03	<i>DMPK</i>	13	> 50	13	261
C44-CSK01	<i>RFC1</i>	k.A.	k.A.	10	437

Tabelle 14 Vergleich Resultate traditioneller PCR-basierter Methoden mit LRS. k. A.: keine Angaben

Die Daten des LRS wurden mit denen der traditionellen Methoden verglichen. Bei 12 der Proben unterscheiden sich die Ergebnisse geringfügig. Bei 3 der Proben stimmen die Ergebnisse nur teilweise überein, bei einer (C20-CD03) besteht eine eingeschränkte Übereinstimmung der Resultate. Bei einer Probe (C44-CSK01) wurden keine Daten mit den traditionellen Methoden generiert.

### 5.3.2 Vergleich Methylierung

Pat. Nr.	Geschlecht	Resultate MS-MLPA	Resultate Methylierung LRS (mittlerer Methylierungsgrad)
C04-XD04	m	Partielle Hypermethylierung <i>FMR1</i>	0.51
C05-XD05	m	Hypermethylierung <i>FMR1</i>	0.83
C07-XD07	m	Hypermethylierung <i>FMR1</i>	0.77
C08-XD08	m	Hypermethylierung <i>FMR1</i>	0.87
C09-XD09	w	Normwerte für die untersuchten Zielregionen	0.27, 0.89
C10-XD10	m	Hypermethylierung <i>FMR1</i>	0.82

Tabelle 15 Vergleich Methylierungsmuster MS-MLA und LRS, w: weiblich, m: männlich

Die in Tabelle 15 dargestellten Ergebnisse zeigen eine weitgehende Übereinstimmung zwischen den mittels MS-MLPA erhobenen Resultaten und den durch LRS bestimmten Methylierungsgraden. In der Probe mit durch MS-MLPA nachgewiesener partieller Hypermethylierung (C04-XD04) von zeigte sich auch in der LRS-

Analyse entsprechend mittlerer Methylierungsgrad von 0.51. Bei der Probe C09-XD09 zeigte sich mittels MS-MLPA ein normales Methylierungsmuster, der mit LRS nachgewiesene mittlere Methylierungsgrad von 0.89 weist aber auf eine Hypermethylierung eines der beiden Allele hin.

## 6 Diskussion

Das Ziel dieser Arbeit war es, die Leistungsfähigkeit von traditionellen PCR-basierten Methoden mit dem neu zu etablierenden Long-Read-Sequencing von PacBio zu vergleichen und zu evaluieren, inwiefern LRS sich für die Diagnostik von ausgewählten Repeat-Expansions-Erkrankungen eignet.

Wie bereits in der Einleitung im Abschnitt 1.1 dieser Arbeit beschrieben, machen sich wiederholende DNA-Sequenzen über 50 % des menschlichen Genoms aus [1], wobei besonders STRs eine wichtige Rolle bei neurologischen Erkrankungen spielen. Die Pathogenität dieser Repeat-Expansions-Erkrankungen ist dabei abhängig von der Wiederholungslänge, der Sequenz, mit oder ohne Unterbrechungen und der genomischen Lokalisation, was unterschiedliche krankheitsverursachende Mechanismen auslösen kann [1–3].

Zur Beantwortung der Fragestellungen wurde ein Patient\*innenkollektiv mit insgesamt 57 Patient\*innen erstellt, welche auf eine der Repeat-Expansions-Erkrankungen in folgenden Genen getestet wurden: *FMR1*, *C9orf72*, *FXN*, *HTT*, *DMPK*, *RFC1*, *ATXN1* (SCA1), *ATXN2* (SCA2), *ATXN3* (SCA3), *CACNA1A* (SCA6), *ATXN7* (SCA7) und *PHOX2B*. Alle Proben wurden zunächst mit traditionellen PCR-basierten Methoden analysiert. Aufgrund eines verlängerten zeitlichen Ablaufs der Etablierung des LRS und der damit verwendeten Software VarSeq konnten lediglich 17 der insgesamt 57 Proben mittels LRS analysiert werden, davon 10 Proben für *FMR1*, je 2 Proben für *FXN* und *HTT* sowie je eine Probe für *C9orf72*, *DMPK* und *RFC1*.

Die erste zentrale Fragestellung dieser Arbeit betraf die Zuverlässigkeit von LRS bei der Identifizierung und Quantifizierung von Repeat-Expansions-Erkrankungen im Vergleich zu den traditionellen PCR-basierten Methoden. Die PCR-basierten Methoden zeigten erwartungsgemäss gute Ergebnisse in der Detektion von kleineren bis mittleren Repeatlängen. Die Methoden waren im Labor schnell und unkompliziert durchführbar und die Auswertung für Allele im Normbereich vergleichsweise einfach. Die genaue Anzahl konnte bei den Normalallelen bei den unterschiedlichen Repeat-Expansions-Erkrankungen ohne Schwierigkeiten dargestellt werden. Diese Effizienz erklärt, weshalb PCR-basierte Methoden bislang als die Standardmethode in der

Diagnostik von Repeat-Expansions-Erkrankungen gelten und trotz ihrer technischen Limitationen weiterhin eine wichtige Rolle spielen.

Die Einschränkung der PCR-basierten Methoden zeigte sich in der Quantifizierung von grossen Repeatlängen. Bei Expansionen wie beispielsweise über 200 Repeats (*FMR1*) bzw. über 145 (*C9orf72*) war eine exakte Bestimmung nicht möglich und die Werte wurden deshalb nur als «mehr als» angegeben. Diese Vorgehensweise wird auch in den User Manuals der jeweiligen Kits empfohlen [43,44,48,50]. Auch bei *FXN* sowie *DMPK* wurden Expansionen mit «mehr als 66» bzw. «mehr als 50» angegeben, da eine genaue Quantifizierung ab einer gewissen Repeatlänge nur unzureichend möglich war. Diese Einschränkung sollte aber relativiert werden, da die Anzahl Repeats gemessen an DNA aus Blut nicht unbedingt die Anzahl Repeats in für die Erkrankung relevanten Geweben widerspiegeln [13,18,25,27]. Insofern ist die Angabe mit der Aussage «mehr als» in diesem Zusammenhang ausreichend und im klinischen Setting durchaus akzeptiert.

Das LRS ermöglichte grundsätzlich eine präzisere Quantifizierung der Repeatanzahl als die traditionellen Methoden. Dies zeigte sich in der genauen numerischen Angabe der Repeats auch bei grossen Expansionen. Bei *FMR1* wurden Vollmutationen mit 236 bis 461 Repeats nachgewiesen, bei *FXN*-Expansionen von 346, 408 und 720 Repeats, bei *DMPK* 261 Repeats, bei *HTT* 42 und 43 Repeats und bei *RFC1* 437 AAGGG-Repeats (Tabelle 12).

Der Vergleich der 17 mit beiden Methoden analysierten Proben ergab, wie in Tabelle 14 dargestellt, ein differenziertes Bild. Bei 12 der 17 Proben stimmten die Ergebnisse bis auf geringfügige Abweichungen überein.

- Teilweise Übereinstimmung bei 3 Proben: Bei den auf Proben C04-XD04 (*FMR1*), C05-XD05 (*FMR1*) und C17-XD17 (*FMR1*) wurde mit den PCR-basierten Methoden Mosaik zwischen Prä- und Vollmutationsbereich nachgewiesen. Auch mittels LRS liessen sich diese Mosaik grundsätzlich darstellen, waren jedoch nicht unmittelbar auf den ersten Blick in der Software erkennbar. Erst durch die Erstellung der Waterfall-Plots und der damit einhergehenden visuellen Darstellung der Reads wurden die Mosaik erkannt.
- eingeschränkte Übereinstimmung bei 1 von 17 Proben: Bei Probe C20-CD03 (*C9orf72*) wurden mittels TP-PCR 2 und >145 Repeats

nachgewiesen. In der initialen Auswertung der LRS-Daten wurden hingegen bei beiden Allelen jeweils 7 Repeats detektiert. Die Expansion war jedoch in den LRS-Daten vorhanden, wurde jedoch erst nach der Einbeziehung sogenannter «failed reads» erkannt. Es handelte sich somit primär um eine Auswertungsproblematik mit der verwendeten Software-Pipeline und nicht um eine Detektionsproblematik im eigentlichen Sinne.

- Keine Vergleichsdaten bei 1 von 17 Proben: Für Probe C44-CSK01 (*RFC1*) lagen keine mit traditionellen Methoden generierten Daten vor.

Die geringfügigen Abweichungen zwischen den Methoden bei den 12 übereinstimmenden Proben können aufgrund der unterschiedlichen Techniken erklärt werden. Bei der TP-PCR wird die Repeatanzahl aufgrund der Fragmentgrösse im Elektropherogramm mittels einer zuvor bestimmten Kalibrierungskurve berechnet, während bei LRS die Repeats direkt sequenziert und durch den Algorithmus der VarSeq-Software gezählt wurden. Die angegebenen Werte der TP-PCR beinhalten zudem eine Toleranz, zum Beispiel 90 +/- 5 Repeats, um die Ungenauigkeit der Methode zu berücksichtigen.

Besonders interessant war die eingeschränkte Übereinstimmung bei der Probe C20-CD03, die auf *C9orf72* untersucht wurde. Mit der TP-PCR wurden 2 Repeats für ein Allel und > 145 Repeats für das andere Allel nachgewiesen (Tabelle 5), während mit LRS zwei Allele mit je 7 Repeats detektiert wurden (Tabelle 12). Diese Abweichung kann potenziell schwerwiegende Konsequenzen haben, da die Expansion nicht erkannt wurde. Zudem ergaben sich auch für das Normalallel mit 2 bzw. 7 Repeats zwei unterschiedliche Werte. Welcher dieser Werte den tatsächlichen Zustand abbildet, lässt sich auf Basis der vorliegenden Daten nicht abschliessend beurteilen. Es ist nicht auszuschliessen, dass der mit LRS ermittelte Wert korrekt ist. Wie in der Einleitung erläutert, ist die genaue Bestimmung der Repeat-Anzahl bei *C9orf72* aufgrund des hohen GC-Gehalts, der grossen möglichen Anzahl der Repeats und der repetitiven Struktur der flankierenden Sequenzen ohnehin eingeschränkt [13]. Die Bildung von Sekundärstrukturen könnte die Sequenzierung der Expansion zusätzlich erschwert haben.

Um der Ursache dieser eingeschränkten Übereinstimmung auf den Grund zu gehen, wurden verschiedene Faktoren untersucht. Die Probe zeigte mit 77.7 Gb den

geringsten Yield aller untersuchten Proben und erreichte mit 24.84 Reads auch die niedrigste genomweite Coverage. Die durchschnittliche Read-Länge von 15.3 Kb lag im Durchschnitt und die Reads-Qualität von Q36 war sogar überdurchschnittlich. Das Alter der DNA-Probe bei der Analyse mit LRS betrug 5 Jahre, was auf mögliche DNA-Degradation hindeuten und ursächlich für die Unstimmigkeit der Resultate und den geringen Yield sein könnte. Wie weiter oben bereits erwähnt, wurden die Rohdaten anschliessend genauer überprüft und sogenannte «failed reads» ebenfalls in die Software-Pipeline zur Auswertung der Daten implementiert. Diese Reads wiesen zu viele Abweichungen zur Referenzsequenz auf und flossen daher normalerweise nicht in das Variant Calling ein. Die Untersuchung dieser «failed reads» zeigten die zuvor nicht erkannte Expansion an. Diese strikte Qualitätsfilterung könnte somit dazu geführt haben, dass diagnostisch relevante, wenn auch qualitativ unzureichende Informationen verworfen wurden und so das eigentliche Resultat verfälscht haben. Dies unterstreicht, dass die beobachtete Abweichung primär auf Limitationen der aktuellen Software-Pipeline und nicht auf die LRS-Technologie selbst zurückzuführen ist. Die Sequenzinformation war zwar vorhanden, wurde jedoch durch die Standardfilterung verworfen.

Für die Probe C44-CSK01 mit *RFC1*-Expansion lagen keine mit traditionellen Methoden generierten Daten vor, sondern nur schriftlich dokumentierte Befundwerte. Dieser Fall zeigt eine der Hauptmotivationen für die Etablierung von LRS im Labor auf. Wie in der Einleitung dargelegt, wurde aufgrund der Komplexität der verschiedenen Repeat-Motive, bisher davon abgesehen, die CANVAS-Diagnostik mit traditionellen Methoden zu etablieren. Das normale AAAAG-Motiv sollte klar vom pathogenen AAGGG-Motiv abgegrenzt werden können, was mit der Fragmentanalyse nicht möglich ist. LRS ermöglichte bei dieser Probe die Identifikation der AAGGG-Motive mit 437 Repeats (Tabelle 12).

Allerdings zeigten sich Limitationen in der aktuellen Auswertung der LRS-Daten. Bei mehreren Proben wurde die Expansion nur durch eine sehr geringe Anzahl von Reads abgedeckt. Besonders deutlich zeigte sich dies beispielsweise bei der Probe C28-FK01 (*FXN*), bei der für das expandierte Allel mit 720 GAA-Repeats lediglich 1 Read vorlag, während das Normalallel durch 16 Reads abgedeckt wurde. Diese Probe erreichte zwar einen Yield von 103.6 Gb und eine durchschnittliche Read-Qualität von Q34, jedoch war die genomweite Coverage mit 33.15 Reads nur

eingeschränkt aussagekräftig, da sie keine Rückschlüsse auf die Verteilung der Reads an der Zielregion zwischen den beiden Allelen erlaubt. Bei C44-CSK01 (*RFC1*) zeigte sich ein vergleichbares Bild mit 17 Reads für das Normalallel und nur 1 Read für das expandierte Allel mit 437 AAGGG-Repeats, trotz guter technischer Parameter (108.4 Gb Yield, Q37 Read-Qualität, 36.67 genomweite Coverage).

Diese Beobachtungen stellen keine grundsätzlichen Limitation der LRS-Technologie dar, sondern sind vielmehr auf die Einschränkungen der derzeit verwendeten Software-Pipeline zurückzuführen. Die relevanten Sequenzinformationen sind in den Rohdaten vorhanden, werden jedoch in der Standardauswertung möglicherweise nicht vollständig erfasst. Eine erweiterte Analyse unter Einbeziehung sogenannter «failed reads» könnte zu einer verbesserten Darstellung grosser Allele beitragen. Zudem ist nicht auszuschliessen, dass die Qualität der verwendeten, teilweise älteren DNA-Proben die Verteilung der Reads pro Allel zusätzlich beeinflusst hat. Insgesamt erscheinen die beschriebenen Aspekte daher als Ergebnis eines Zusammenspiels von Auswertungsstrategie und Probenqualität und weniger als Ausdruck einer grundsätzlichen methodischen Einschränkung.

Die Waterfall-Plots verdeutlichen zudem, dass die von VarSeq extrapolierte durchschnittliche Repeatanzahl teilweise erheblich von den tatsächlich sequenzierten Repeats abwich. Bei der Probe C04-XD04 (*FMR1*) wurden beispielsweise durchschnittlich 249 Repeats angegeben, obwohl die meisten der 11 Reads deutlich kürzer waren (Abbildung 13). Diese Diskrepanz könnte mit der in der Einleitung beschriebenen somatischen Instabilität der Repeats zusammenhängen, die zu einem ausgeprägten Mosaikstatus führt. Bei *FMR1* wurde bereits beschrieben, dass bei einer Person verschiedene Zellen unterschiedliche Repeatlängen aufweisen können, was zu einer variablen Expression von *FMR1* und damit zu unterschiedlichem Schweregrad der Symptomatik führt [8]. Die Software VarSeq in der momentan verwendeten Version scheint diese Variabilität nicht ideal erfassen zu können und gibt stattdessen einen Durchschnittswert an, der die tatsächliche Komplexität der Probe nicht widerspiegelt. Des Weiteren lässt sich nicht ausschliessen, dass dieser Diskrepanz der GC-Gehalt der Repeats zugrunde liegt, welche die Bildung von Sekundärstrukturen begünstigt und die Sequenzierung beeinträchtigen könnte.

Eine mögliche Einschränkung dieser Arbeit in Bezug auf die erste Fragestellung ergibt sich aus der Anzahl der mit LRS analysierten Proben. Mit 17 untersuchten Fällen stellt das Kollektiv im Rahmen einer Qualitätskontrolle der Methode für seltene, durch Repeat-Expansionen verursachte Erkrankungen eine überschaubare Stichprobe dar. Aufgrund der geringen Prävalenz entsprechender Erkrankungen ist die Verfügbarkeit geeigneter Proben jedoch strukturell eingeschränkt.

Die Auswahl der Proben erfolgte gezielt und umfasste überwiegend bereits diagnostizierte, positive Fälle, um die Nachweisfähigkeit der LRS für Repeat-Expansionen zu evaluieren. Entsprechend lag der Fokus dieser Arbeit auf der Untersuchung der Sensitivität der Methode. Eine systematische Untersuchung der Spezifität, beispielsweise durch den Einschluss gesunder Proband\*innen oder unauffälligen Proben war nicht Gegenstand dieser Arbeit.

Zusätzlich ist zu berücksichtigen, dass die LRS-Technologie mit einem hohen finanziellen Aufwand verbunden ist. Dieser Umstand stellte eine weitere Rahmenbedingung dar, die sowohl die gezielte Probenwahl als auch die insgesamt analysierbare Probenzahl im Kontext dieser Arbeit mitbestimmte.

Die zweite Fragestellung befasste sich mit den technischen Grenzen der traditionellen Methoden und der Fähigkeit von LRS, diese Lücken zu schliessen. Die traditionellen Methoden zeigten bei allen Erkrankungen ähnliche Einschränkungen, die sich sowohl auf die quantitative als auch die qualitative Aussage der Analysen bezog.

Die Fragmentanalyse ist quantitativ generell auf etwa 1000 bp limitiert, was den Nachweis von grösseren Repeat-Expansionen erschwert. Diese Limitation wurde bei allen untersuchten Erkrankungen evident. Bei *FMR1* konnten Vollmutationen lediglich als ">200 Repeats" angegeben werden, was einer Fragmentgrösse von circa 818 bp entspricht. Von den 17 auf *FMR1* untersuchten Proben fielen 5 männliche Proben (C02-XD02, C07-XD07, C08-XD08, C10-XD10) sowie 2 weibliche Proben (C09-XD09, C16-XD16) in diese Kategorie (Tabelle 4). Bei *C9orf72* lag die Grenze bei 145 Repeats, was bei 7 der 10 analysierten Proben (C18-CD01 teilweise, C19-CD02, C20-CD03, C21-CR04, C22-CR05, C23-CR06, C24-CR07) zu einer eingeschränkten Quantifizierung führte (Tabelle 5). Ähnliche Limitationen zeigten sich bei *FXN* (Grenze bei 66 Repeats, betraf 5 von 6 Proben, Tabelle 6) und *DMPK* (Grenze

bei 50 Repeats, betraf 4 von 5 Proben, Tabelle 8). Diese quantitativen Limitationen sind vor dem Hintergrund der in der Einleitung beschriebenen Genotyp-Phänotyp-Korrelation zu berücksichtigen. Eine genauere Quantifizierung kann neben der Diagnosestellung auch begrenzte Hinweise für die Prognoseabschätzung und genetische Beratung liefern.

Qualitativ bestand die Haupteinschränkung der traditionellen Methoden darin, dass keine oder wie im Beispiel von *FMR1* nur unzureichende Bestimmungen von Unterbrechungen oder Sequenzmotiven möglich war. Bei *FMR1* gab es bei 5 Proben (C06-XD06, C09-XD09, C15-XR15, C16-XD16, C17-XD17) Hinweise auf AGG-Unterbrechungen, die sich durch ein verringertes Fluoreszenzsignal im CGG-Repeat-Profil manifestierten (Abbildung 7). Jedoch konnten weder die genaue Anzahl noch die Lokalisation dieser Unterbrechungen zweifelsfrei bestimmt werden. Wie in Abschnitt 2.1 dargelegt, sind diese Unterbrechungen jedoch von klinischer Relevanz, da sie als Schutz gegen eine Expansion wirken. Das Fehlen von AGG-Unterbrechungen erhöht das Risiko für eine Expansion bei Weitergabe, insbesondere über das mütterliche Allel. Die Bestimmung dieser AGG-Unterbrechungen hat vor allem bei Graubereich-Allelen (45-54 Repeats) eine hohe Bedeutung, da diese auf die nächste Generation expandieren und einen Phänotyp verursachen könnten.

Das LRS konnte gewisse Lücken schliessen. Die exakte Quantifizierung grosser Expansionen war möglich, wie die präzisen Werte von bis zu 461 CGG-Repeats bei *FMR1* (C10-XD10), 720 GAA-Repeats bei *FXN* (C28-FK01) oder 437 AAGGG-Repeats bei *RFC1* (C44-CSK01) demonstrieren. Dabei ist jedoch anzumerken, dass diese präzisen Werte mit Vorsicht zu interpretieren sind, wenn sie wie bei C28-FK01, wie in Tabelle 12 ersichtlich, nur auf einem einzigen Read basieren.

Ebenfalls konnten qualitative Aussagen in Form der Bestimmung der Sequenzmotive getroffen werden. Bei *FMR1* konnten nicht nur die CGG-Repeats quantifiziert, sondern auch AGG-Unterbrechungen identifiziert werden. Die Probe C06-XD06 zeigte beispielsweise 29 CGG-Repeats mit 2 AGG-Unterbrechungen auf einem Allel, während das andere Allel 90 CGG-Repeats ohne Unterbrechungen aufwies. Ähnlich präzise konnte bei C17-XD17 ein Allel mit 22 CGG und 1 AGG sowie ein weiteres mit 183 CGG und 1 AGG charakterisiert werden, wobei hier 24 bzw. 19 Reads vorlagen. Bei *RFC1* ermöglichte LRS die eindeutige Identifikation des

pathogenen AAGGG-Motivs mit 437 Repeats bei C44-CSK01, was mit traditionellen Methoden nicht möglich gewesen wäre. Allerdings basierte auch diese Aussage auf nur einem einzigen Read für das expandierte Allel, während das Normalallel mit 10 AAAAG-Repeats durch 17 Reads abgedeckt wurde (Tabelle 12).

Ein weiterer Vorteil ergab sich aus der Möglichkeit der Detektion der Methylierung im selben Ansatz. Der mittlere Methylierungsgrad eines bestimmten Loci konnte direkt aus den LRS-Daten bestimmt werden. Bei den 10 *FMR1*-Proben stimmten die Methylierungswerte, wie in Tabelle 15 dargestellt, weitgehend mit den MS-MLPA-Ergebnissen überein. Probe C05-XD05 zeigte beispielsweise einen Methylierungsgrad von 0.83, was mit der MS-MLPA-diagnostizierten Hypermethylierung korrelierte, während C04-XD04 einen Wert von 0.51 aufwies, was der MS-MLPA-Beobachtung von 50 % nicht methylierten Allelen entsprach. Bei den weiblichen Proben konnten unterschiedliche Methylierungsgrade für die beiden Allele bestimmt werden. Bei der Probe C09-XD09 zeigte die MS-MLPA Analyse kein Hinweis auf eine Hypermethylierung, mit LRS konnten jedoch mittlere Methylierungsgrade mit Werten von 0.27 und 0.89 für die beiden Allele nachgewiesen werden. Dies deutet auf eine Hypermethylierung des expandierten Allels hin, welche mit MS-MLPA nicht erfasst wurde. Dies ist wahrscheinlich auf die technischen Limitationen der Methode zurückzuführen, bei der lediglich die durchschnittliche Methylierung über beide Allele, nicht aber die unterschiedliche Methylierung pro Allel ermittelt wird.

Dies ist von besonderer Bedeutung, da bei FXS eine Vollmutation üblicherweise zu einer vollständigen Hypermethylierung des *FMR1*-Promotors führt, was die Inaktivierung des Gens zur Folge hat. Die Integration der Methylierungsanalyse im gleichen Ansatz wie die Quantifizierung der Repeatanzahl stellt einen erheblichen Effizienzgewinn gegenüber dem bisherigen Vorgehen dar, bei dem neben der TP-PCR auch noch eine MS-MLPA-Analyse durchgeführt werden musste.

Bei LRS zeigte sich, dass die Abdeckung der beiden Allele in einigen Fällen unterschiedlich ausfallen kann (Tabelle 13). Bei 10 der 11 Proben mit zwei unterschiedlichen Allelen zeigte sich eine Diskrepanz in der Anzahl Reads zwischen Allel 1 und Allel 2. Die grössten Abweichungen zeigten sich bei C44-CSK01 (*RFC1*) mit 94.12 % und bei C28-FK01 (*FXN*) mit 93.75 %, wo für das expandierte Allel jeweils nur ein einziger Read vorlag. Solche Unterschiede können zum einen auf die Bildung

von Sekundärstrukturen bei Repeat-Expansionen zurückzuführen sein, wie bereits in der Einleitung dieser Arbeit erläutert. Bei FXN beispielsweise kann die GAA-Expansion abnorme DNA-Strukturen bilden, die die Polymerase blockieren und so die Amplifikation erschweren.

Gleichzeitig spielen auch technische und analytische Faktoren eine wichtige Rolle. Eine geringe Sequenziertiefe, die geringe Qualität der verwendeten DNA-Proben, insbesondere bei älteren Proben, sowie die derzeitigen Einschränkungen der Auswertung mit der Software-Pipeline ohne Implementation der «failed reads» können zu diesem allelischen Bias führen. Es ist davon auszugehen, dass sich dieser Bias durch eine optimierte Auswertungsstrategie reduzieren lässt.

Ein weiterer zu berücksichtigender Aspekt ist der hohe DNA-Bedarf der Methode. Für eine zuverlässige Sequenzierung erfordert LRS hochmolekulare DNA in ausreichender Menge. Gemäss dem verwendeten Protokoll sind mindestens 3-5 µg hochmolekulare DNA erforderlich [61]. Dies kann bei pränatalen Proben und Ringversuchsproben eine Herausforderung darstellen, da das Ausgangsmaterial häufig nur begrenzt verfügbar ist. Für die Diagnostik an pränatalen Proben besteht die Möglichkeit, DNA aus kultivierten Zellen zu gewinnen, wodurch höhere mengen an hochmolekularer DNA extrahiert werden können. Dies erfordert zwar zusätzlichen Zeitaufwand, ermöglicht aber auch für pränatale Proben den Einsatz von LRS. Auch im Rahmen von Ringversuchen ist die verfügbare DNA-Menge in der Regel begrenzt, weshalb diese Proben derzeit nur eingeschränkt mit LRS analysiert werden und so zur externen Qualitätskontrolle beitragen können. Durch die allfällige zunehmende Verbreitung der LRS-Methode in anderen Laboren könnte jedoch erwartet werden, dass künftig mehr Material bereitgestellt wird. Des Weiteren können zur externen Qualitätssicherung Laborvergleiche durchgeführt werden. Insgesamt zeigt dies, dass der DNA-Bedarf eine praktische Einschränkung darstellt, die sich durch eine angepasste Probenaufbereitung und zukünftige Entwicklungen jedoch handhaben lässt.

Eine Einschränkung dieser Arbeit in Bezug auf die zweite Fragestellung ergibt sich daraus, dass für einige seltene Erkrankungen nur sehr wenige oder gar keine Proben mit LRS analysiert werden konnten. So stand für *RFC1* lediglich eine Probe zur Verfügung, wodurch Aussagen zu dieser Erkrankung auf der Grundlage der

vorliegenden Daten nur eingeschränkt möglich sind. Für die SCAs konnten trotz vorhandener Proben mit traditionellen Methoden keine Analysen mit LRS durchgeführt werden. Dies betrifft insbesondere die Sequenzmotive der verschiedenen Subtypen, deren Untersuchung für die klinische Interpretation, etwa bei SCA1 im Zusammenhang mit CAT-Unterbrechungen und milderem Phänotyp, potenziell relevant wäre. Aufgrund der Seltenheit dieser Erkrankungen lassen sich jedoch nicht dieselben Anforderungen an die Probenverfügbarkeit stellen wie für häufigere Krankheiten. Die vorliegenden Ergebnisse liefern dennoch Hinweise zur Anwendbarkeit von LRS für die untersuchte Fragestellung und ermöglichen eine erste Einschätzung, auch wenn für bestimmte Erkrankungen nur eingeschränkte Daten vorliegen.

Die dritte Fragestellung widmete sich dem Nachweis von Mosaiken, einer Eigenschaft, die wie in Abschnitt 1.3 dargelegt, aus der Instabilität der Repeats während der DNA-Replikation und Meiose resultiert.

Mit den PCR-basierten Methoden waren Mosaik gut ersichtlich. Bei *FMR1* zeigten 6 männliche Proben Mosaik, die als multiple Peaks im Elektropherogramm zu sehen waren. Diese Mosaik umfassten sowohl Mosaik zwischen Prämutation und Vollmutation als auch solche innerhalb des Prämutationsbereichs. Wie in Abbildung 6 dargestellt, zeigte Probe C04-XD04 drei verschiedene Peaks bei circa 820, 850 und 980 Basenpaaren, was auf mindestens drei verschiedene Zellpopulationen mit unterschiedlichen Repeat-Längen hindeutet. Bei Probe C11-XD11 wurde ein komplexeres Mosaik mit circa 165, 195 und >200 Repeats detektiert. Eine quantitative Aussage über die relative Häufigkeit der verschiedenen Zellpopulationen war mit den traditionellen Methoden jedoch nicht möglich, da die Peak-Höhen im Elektropherogramm nicht direkt die Mengenverhältnisse wiedergeben. Dies kann einerseits an der präferenziellen Amplifikation der kürzeren Fragmente während der PCR liegen, was dazu führt, dass kleinere Allele überrepräsentiert sein könnten. Ebenfalls können auch hier Sekundärstrukturen die Effizienz der PCR beeinflussen.

Die Analyse mit LRS ergab ein komplexeres Bild, das jedoch primär auf die Limitationen der verwendeten Software-Pipeline zurückzuführen ist. In der in VarSeq erstellten Übersichtstabelle wurden Mosaik nicht direkt als solche erkannt und ausgegeben. Es wurde jeweils nur ein durchschnittlicher Wert angegeben, der die

unterschiedlichen Read-Längen eines Mosaiks nicht widerspiegelte und es erfolgte kein automatischer Hinweis auf ein mögliches Mosaik. Erst durch die Generierung und Betrachtung der Waterfall-Plots wurde deutlich, dass die verschiedenen Repeatlängen in den Rohdaten vorhanden waren, jedoch durch die aktuelle Software-Pipeline nicht automatisch angezeigt wurden. Diese Beobachtung stellt somit kein grundsätzliches Problem der LRS-Methode selbst dar, sondern ist auf die momentanen Grenzen der Software-Pipeline zurückzuführen.

Zudem erfolgte auch kein automatischer Hinweis auf das Vorliegen eines Mosaiks. Erst nach der genaueren Betrachtung der Waterfall-Plots wurde ersichtlich, dass die verschiedenen Repeatlängen durchaus in den Rohdaten vorhanden waren, aber von der Software nicht automatisch korrekt interpretiert wurden.

Beispielsweise zeigte Probe C04-XD04 (*FMR1*) einen durchschnittlichen Wert von 249 Repeats in der Tabellenansicht, während die 11 einzelnen Reads im Waterfall-Plot deutlich unterschiedliche Längen aufwiesen und damit das mit TP-PCR nachgewiesene Mosaik zwischen Prämutation und Vollmutation repräsentierten (Abbildung 13). Noch interessanter war der Fall von Probe C08-XD08 (*FMR1*), die mit traditionellen Methoden lediglich als Expansion >200 Repeats klassifiziert worden war, ohne Hinweis auf ein Mosaik. Der Waterfall-Plot (Abbildung 14) des LRS zeigte jedoch, dass von 11 Reads nur einer die volle Expansion von 362 Repeats aufwies, während die übrigen Reads kürzer waren. Dies deutet auf ein zuvor nicht erkanntes Mosaik hin, was zeigt, dass LRS sensitiver für die Detektion von Mosaiken sein könnte als die traditionellen Methoden, wenn die Daten korrekt ausgewertet würden.

Die geringe Anzahl der Reads bei den Proben mit Mosaiken war problematisch für die Quantifizierung der Mosaikanteile. Mit nur 9 bis 11 Reads pro Probe ist eine Aussage über die Häufigkeitsverteilung verschiedener Allele praktisch nicht möglich. Für einen verlässlichen Nachweis und eine allfällige Quantifizierung von Mosaiken wäre eine deutlich höhere Sequenziertiefe nötig. Dies steht im Kontrast zu den genomweiten Coverage-Werten von 30 bis 33 Reads, die auf den ersten Blick ausreichend erscheinen, nicht aber die Heterogenität der lokalen Coverage abbilden.

Zusätzlich zeigten sich in den meisten Reads einige Sequenzabweichungen, die die Interpretation ebenfalls erschwerten. Diese Abweichungen könnten echte biologische Varianten, Sequenzierfehler oder Artefakte der bioinformatischen Pipeline

sein. Im Waterfall-Plot von C35-HK02 (*HTT*) beispielsweise, einer Probe ohne Mosaikstatus, waren solche Lücken oder Abweichungen in einzelnen Reads sichtbar, obwohl die Probe mit 12 bzw. 13 Reads für die beiden Allele und guten technischen Parametern (118.3 Gb Yield, Q36 Read-Qualität, 37.87 genomweite Coverage) eigentlich gut charakterisiert sein sollte.

Somit ist festzuhalten, dass die Detektion von Mosaiken mit der aktuellen LRS-Implementierung für diagnostische Fragestellungen nur eingeschränkt darstellbar ist. Dies liegt einerseits an der Software-Pipeline, die für diese Fragestellung angepasst werden müsste, andererseits auch an der Coverage der gewünschten Regionen. Während die genomweite Coverage im Durchschnitt bei 33.8 Reads lag, war die lokale Coverage an den Zielregionen oft deutlich niedriger, teilweise mit nur einem einzigen Read für expandierte Allele. Es zeigt sich, dass Mosaikete mit PCR-basierten Methoden trotz limitierter Quantifizierung aufgrund ihrer charakteristischen Erscheinung als multiple Peaks im Elektropherogramm besser erkannt werden können, während bei LRS die Information über Mosaikete zwar in den Rohdaten vorhanden ist, diese aber nicht automatisch detektiert werden und manuell in den Waterfall-Plots kontrolliert werden müssen. Mit einer angepassten bioinformatischen Pipeline, die speziell für die Detektion von Mosaiken optimiert ist, sollte LRS deutliche Verbesserungen zeigen können. Die momentane Pipeline stellt im Alltag einen Nachteil dar, da die manuelle Überprüfung zeitaufwändig ist und die Gefahr besteht, dass Mosaikete übersehen werden, wenn man sich nur auf die tabellarische Ausgabe der Software verlässt.

Eine Limitation dieser Arbeit in Bezug auf die dritte Fragestellung ist, dass nur *FMR1*-Proben systematisch auf Mosaikete untersucht wurden und keine Daten zu Mosaiken bei anderen Repeat-Expansions-Erkrankungen vorliegen. Gerade bei *DMPK* sind Mosaikete jedoch häufig und klinisch sehr relevant, da sie wesentlich zur phänotypischen Variabilität beitragen. Vor diesem Hintergrund wäre es sinnvoll, weitere Proben gezielt auf das Vorhandensein von Mosaiken zu analysieren.

Die letzte Fragestellung befasste sich mit dem diagnostischen Mehrwert von LRS im Vergleich zu den traditionellen Methoden bei den verschiedenen Repeat-Expansions-Erkrankungen. Dieser erweist sich als deutlich erkrankungsspezifisch und kann nicht generell für alle untersuchten Erkrankungen festgehalten werden.

Bei *FMR1*-bedingten Erkrankungen bietet LRS in der Theorie mehrere Vorteile durch die genaue Quantifizierung von Vollmutationen über 200 Repeats hinaus, den Nachweis von AGG-Unterbrechungen und die simultane Bestimmung des Methylierungsstatus in einem Ansatz. In der praktischen Anwendung zeigten sich jedoch Einschränkungen durch die nicht ausreichende Detektion von Mosaiken mittels der verwendeten Software und die geringe lokale Sequenziertiefe. Bei den 10 analysierten Proben lag die Anzahl der Reads bei Vollmutationen oft nur bei 4 bis 11 Reads. Die etablierte Kombination aus TP-PCR und MS-MLPA liefert momentan zuverlässigere Ergebnisse für die Standarddiagnostik.

Für *C9orf72* ist es schwierig eine fundierte Aussage über den diagnostischen Mehrwert zu treffen, da im Rahmen dieser Arbeit lediglich eine Probe mit LRS untersucht wurde. Diese Probe (C20-CD03) zeigte Abweichung zu den Ergebnissen der TP-PCR, wobei eine eindeutig nachgewiesene Expansion mit > 145 Repeats mit LRS zunächst nicht bestätigt werden konnte. Ob diese Abweichung auf das verwendete Probenmaterial oder auf Eigenschaften der Zielregion zurückzuführen ist, lässt sich anhand der vorliegenden Daten nicht abschliessend beurteilen. Dabei sollte berücksichtigt werden, dass besonders lange Repeat-Expansionen teilweise nicht korrekt gemappt werden und in den sogenannten «failed reads» landen können, wodurch sie trotz ihrer Anwesenheit nicht direkt erfasst werden. Für eine verlässlichere Einschätzung des diagnostischen Mehrwerts wäre die Analyse zusätzlicher Proben mittels LRS erforderlich.

Bei *FXN* ermöglichte LRS zwar grundsätzlich eine präzise Quantifizierung der GAA-Repeats, die für Prognose und Krankheitsverlauf klinisch relevant ist, jedoch war die niedrige Read-Anzahl für die expandierten Allele problematisch. Die diagnostische Sicherheit der Ergebnisse ist dadurch eingeschränkt, da Sequenzierartefakte nicht ausgeschlossen werden können.

Bei *HTT* zeigten beide mit LRS analysierten Proben gute Übereinstimmung mit traditionellen Methoden. Der diagnostische Mehrwert liegt vor allem in der Möglichkeit, CAA-Unterbrechungen zu identifizieren, deren Verlust mit früherem Manifestationsalter assoziiert ist. Da die traditionellen Methoden für *HTT* jedoch bereits gut etabliert sind und ausreichende Ergebnisse liefern, beschränkt sich der Mehrwert von LRS auf spezielle Fragestellungen zur Charakterisierung von Unterbrechungen.

Auch bei DM1 konnte LRS die Repeatlängen korrekt erfassen und bietet durch die Detektion von Unterbrechungen potenziell zusätzliche prognostische Informationen. Allerdings ist die Anzahl analysierter Proben begrenzt und die bei DM1 auftretenden Mosaik mit der derzeitigen LRS-Implementation nur eingeschränkt detektierbar, sodass auf dieser Grundlage keine abschliessenden Aussagen getroffen werden können.

Einen klaren klinischen Nutzen zeigte LRS bei der *RFC1*-assoziierten CANVAS-Erkrankung, da hier eine eindeutige Unterscheidung der verschiedenen Repeat-Motive möglich ist. Während konventionelle Methoden keine sichere Differenzierung zwischen benignen und pathogenen Motiven erlauben, ermöglicht LRS die direkte Sequenzierung des Repeat-Motivs und damit eine verlässliche molekulargenetische Diagnosestellung. Trotz der problematischen Read-Verteilung (17 Reads für das Normalallel, nur 1 Read für das expandierte Allel) ist LRS eine praktikable Methode zur Differenzierung von Sequenzmotiven bei CANVAS.

Für die SCAs konnte der diagnostische Mehrwert von LRS nicht bewertet werden, da trotz vorhandener Proben keine LRS-Analysen durchgeführt werden konnten. Die Proben wiesen für eine verlässliche Analyse mit LRS nicht die erforderliche Qualität auf. Dies ist eine relevante Lücke dar, da gerade bei SCAs die Sequenzmotiv-Analyse von hoher Relevanz ist, beispielsweise bei SCA1, wo CAT-Unterbrechungen mit einem milderem Phänotyp assoziiert sind.

Bei *PHOX2B* wurden ebenfalls keine Proben mit LRS analysiert. Die vier mit traditionellen Methoden untersuchten Proben zeigten jedoch, dass die etablierten Methoden für diese Erkrankung ausreichend sind (Tabelle 10). Bei CCHS sind die Repeat-Längen im pathologischen Bereich mit 20-33 Repeats vergleichsweise kurz, sodass die in der Einleitung beschriebenen Limitationen der PCR-basierten Methoden bei sehr langen Expansionen hier nicht zum Tragen kommen. Die Kombination aus Fragmentanalyse und Sangersequenzierung liefert präzise, diagnostisch verwertbare Ergebnisse innerhalb kurzer Zeit. Dies ist von besonderer klinischer Bedeutung, da CCHS häufig bei Neugeborenen mit akuter Atemnot diagnostiziert werden muss und eine zeitnahe Diagnosestellung für das klinische Management essenziell ist. Die traditionellen Methoden können innerhalb von ein bis zwei Arbeitstagen abgeschlossen werden, während LRS mit Library-Preparation, mehrtägiger

Sequenzierung und bioinformatischer Auswertung deutlich länger dauert. Ein diagnostischer Mehrwert von LRS ist bei *PHOX2B* daher nicht erkennbar.

## 7 Fazit

Die vorliegende Arbeit untersuchte die Leistungsfähigkeit von Long-Read-Sequencing im Vergleich zu traditionellen PCR-basierten Methoden für die Diagnostik von ausgewählten Repeat-Expansions-Erkrankungen. Die Ergebnisse zeigen, dass beide Methoden ihre spezifischen Vor- und Nachteile aufweisen.

Die Stärken von LRS liegen in der Fähigkeit grosse Repeat-Expansionen genau zu quantifizieren, Sequenzmotive und Unterbrechungen darzustellen sowie den Methylierungsstatus an entsprechenden Zielregionen gleichzeitig zu detektieren. Diese Eigenschaften machen LRS besonders interessant für Erkrankungen wie CANVAS, wo die Unterscheidung verschiedener Sequenzmotive für die Diagnose essenziell sind oder für den Nachweis von Unterbrechungen bei *FMR1*, *HTT* oder *DMPK*, die eine prognostische Relevanz haben.

Allerdings zeigten sich auch Einschränkungen, die jedoch primär auf die derzeit verwendete Software-Pipeline und nicht auf die LRS-Technologie selbst zurückzuführen sind. Der beobachtete allelische Bias, bei dem expandierte Allele unterrepräsentiert waren und teilweise nur durch einen einzigen Read abgedeckt wurden, könnte teilweise durch die Bildung von Sekundärstrukturen bei GC-reichen Repeats bedingt sein, scheint aber auch durch die Einschränkungen der aktuellen Auswertungsstrategie beeinflusst zu werden. Die nicht ausreichende automatische Detektion von Mosaiken stellt keine grundsätzliche Limitation der Sequenzieretechnologie dar, sondern reflektiert die Grenzen der verwendeten bioinformatischen Pipeline, die für diese spezifische Fragestellung nicht optimiert ist. Die Information über Mosaik ist in den Rohdaten vorhanden, wird jedoch von der Software nicht automatisch erkannt und ausgegeben. Der Bedarf an DNA von hoher Qualität und Menge kann bei bestimmten Proben typen eine Herausforderung darstellen, wobei durch angepasste Aufbereitungsstrategien wie Zellkultur bei pränatalen Proben mögliche Lösungsansätze existieren.

Die traditionellen PCR-basierten Methoden bleiben aufgrund ihrer Robustheit, schnellen Durchführbarkeit und ihrer zuverlässigen Detektion von Mosaiken die

Methode der Wahl für die Routinediagnostik der meisten Repeat-Expansions-Erkrankungen mit klarem klinischen Verdacht. Sie liefern für die Mehrheit der diagnostischen Fragestellungen ausreichende Ergebnisse, auch wenn sie bei grossen Expansionen auf Angaben mit «grösser als» beschränkt sind.

LRS eignet sich für die parallele Analyse einer Vielzahl von Repeat-Expansions-Erkrankungen und als komplementäre Methode, die gezielt bei spezifischen Situationen eingesetzt werden könnte. Für eine breitere Anwendung in der Routinediagnostik müssten Voraussetzungen wie eine Verbesserung der lokalen Coverage, Optimierung der bioinformatischen Pipeline für die automatische Detektion von Mosaiken und eine verbesserte Auswertungsstrategie zur Reduktion des allelischen Bias erfüllt werden. Die in dieser Arbeit identifizierten Einschränkungen scheinen überwiegend auf Aspekte der Datenauswertung zurückzuführen zu sein, was einen vielversprechenden Ansatzpunkt für zukünftige Optimierungen bieten könnte.

## 8 Ausblick

Eine Weiterentwicklung für die Überwindung der in dieser Arbeit aufgezeigten Limitationen ist das PureTarget Kit von PacBio, welches spezifisch für Repeat-Expansions-Erkrankungen entwickelt wurde. Die im Kit verwendete Methode basiert auf einer Anreicherung der Zielregionen mit CRISPRCas9. Dabei werden die Cas9-Enzyme von spezifischen «single guided RNAs» zu den flankierenden Sequenzen der Zielregionen gebracht, wo sie die DNA an spezifischen Stellen schneiden und so Fragmente entstehen, welche die Zielregionen einschliessen. Danach werden an die Fragmente ebenfalls SMRTbell-Adapter ligiert und diese mit der SMRT-Technologie sequenziert [62]. Ein grosser Vorteil dieser Methode ist die PCR-freie Anreicherung der Zielregionen, womit auch grössere Repeatlängen erfasst werden können. Im Vergleich zu den in dieser Arbeit beobachteten 1 bis 28 Reads pro Allel kann mit dieser Methode eine höhere Sequenziertiefe von bis 400x erreicht werden [63], was relevant für die zuverlässige Detektion von Mosaiken ist. Allerdings erfasst das Kit nicht alle derzeit bekannten Repeat-Expansions-Erkrankungen, sodass ein breites Screening damit nicht möglich ist.

Parallel zur Anpassung der Sequenzierertechnologie ist auch die Weiterentwicklung der bioinformatischen Auswertung von zentraler Bedeutung. Die in dieser Arbeit beobachteten Einschränkungen bei der automatischen Detektion von Mosaiken, der

Behandlung von «failed reads» und der Darstellung des allelischen Bias sind primär auf die derzeitigen Limitationen der Software-Pipeline zurückzuführen. Eine für Repeat-Expansionen optimierte Auswertungsstrategie, die beispielsweise spezifische Filter für GC-reiche Regionen, verbesserte Algorithmen zur Mosaik-Detektion und eine flexiblere Handhabung von Sequenzvariationen integriert, könnte die diagnostische Aussagekraft erheblich verbessern.

Für eine umfassende Bewertung des diagnostischen Potenzials von LRS wären ergänzende Untersuchungen mit einer grösseren, jedoch realistisch umsetzbaren Anzahl an Proben wünschenswert. Aufgrund der Seltenheit vieler der untersuchten Erkrankungen sowie des hohen Kosten- und Ressourcenaufwands der Methode ist die Analyse grösserer Probenzahlen jedoch nur eingeschränkt möglich. Künftige Arbeiten könnten sich daher darauf konzentrieren, eine begrenzte, aber aussagekräftige Anzahl Proben pro Erkrankung zu untersuchen, um Sensitivität und Spezifität unter Einbezug von negativen Kontrollen sowie Proben mit grenzwertigen Ergebnissen besser abschätzen zu können. Eine fokussierte Analyse ausgewählter SCAs, insbesondere im Hinblick auf die Charakterisierung von Repeat-Unterbrechungen könnte zusätzliche Erkenntnisse liefern. Darüber hinaus wäre eine schrittweise Einbeziehung einzelner weiterer Repeat-Expansions-Erkrankungen, die in dieser Arbeit nicht berücksichtigt wurden, denkbar. Ergänzend könnten externe Referenzmaterialien, Ringversuchsproben und längerfristige Untersuchungen zur weiteren Einordnung des potenziellen prognostischen Nutzens beitragen.

Die Kombination aus technologischer Weiterentwicklung, verbesserter bioinformatischer Auswertung und möglichen gezielten Folgeuntersuchungen wird den Einsatz von LRS in der Routinediagnostik von Repeat-Expansions-Erkrankungen weiter optimieren.

## 9 Literaturverzeichnis

1. Leitão E, Schröder C, Depienne C. Identification and characterization of repeat expansions in neurological disorders: Methodologies, tools, and strategies. *Rev Neurol (Paris)*. Mai 2024;180(5):383–92.
2. Chintalaphani SR, Pineda SS, Deveson IW, Kumar KR. An update on the neurological short tandem repeat expansion disorders and the emergence of long-read sequencing diagnostics. *Acta Neuropathol Commun* [Internet]. 25. Mai 2021 [zitiert 1. Juli 2025];9:98. Verfügbar unter: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8145836/>
3. Chen Z, Morris HR, Polke J, Wood NW, Gandhi S, Ryten M, u. a. Repeat expansion disorders. *Pract Neurol* [Internet]. 15. Mai 2025;25(3):204–16. Verfügbar unter: <https://pubmed.ncbi.nlm.nih.gov/39349043/>
4. Paulson H. Repeat expansion diseases. *Handb Clin Neurol* [Internet]. 2018 [zitiert 1. Mai 2025];147:105–23. Verfügbar unter: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6485936/>
5. Cleary JD, Pattamatta A, Ranum LPW. Repeat-associated non-ATG (RAN) translation. *J Biol Chem* [Internet]. Oktober 2018 [zitiert 3. Juli 2025];293(42):16127–41. Verfügbar unter: <https://linkinghub.elsevier.com/retrieve/pii/S0021925820350845>
6. Biancalana V, Glaeser D, McQuaid S, Steinbach P. EMQN best practice guidelines for the molecular genetic testing and reporting of fragile X syndrome and other fragile X-associated disorders. *Eur J Hum Genet EJHG* [Internet]. April 2015 [zitiert 2. Mai 2025];23(4):417–25. Verfügbar unter: <https://www.nature.com/articles/ejhg2014185>
7. Broniarek I, Niewiadomska D, Sobczak K. Contribution of DNA/RNA Structures Formed by Expanded CGG/CCG Repeats Within the FMR1 Locus in the Pathogenesis of Fragile X-Associated Disorders. *Wiley Interdiscip Rev RNA* [Internet]. 2024 [zitiert 2. Mai 2025];15(6):e1874. Verfügbar unter: <https://wires.onlinelibrary.wiley.com/doi/10.1002/wrna.1874>
8. Ciobanu CG, Nucă I, Popescu R, Antoci LM, Caba L, Ivanov AV, u. a. Narrative Review: Update on the Molecular Diagnosis of Fragile X Syndrome. *Int J Mol Sci* [Internet]. 24. Mai 2023 [zitiert 2. Mai 2025];24(11):9206. Verfügbar unter: <https://www.mdpi.com/1422-0067/24/11/9206>
9. Hunter JE, Berry-Kravis E, Hipp H, Todd PK. FMR1 Disorders. In: *GeneReviews®* [Internet]. 2024 [zitiert 19. April 2025]. Verfügbar unter: <http://www.ncbi.nlm.nih.gov/books/NBK1384/>
10. Tom Strachan, Andrew P. Read. *Human Molecular Genetics*. 5. Boca Raton: CRC Press; 2019.

11. Gómez-Rodríguez MJ, Morales-Conejo M, Arteche-López A, Sánchez-Calvín MT, Quesada-Espinosa JF, Gómez-Manjón I, u. a. Fragile X Syndrome Caused by Maternal Somatic Mosaicism of FMR1 Gene: Case Report and Literature Review. *Genes* [Internet]. 8. September 2022 [zitiert 2. Mai 2025];13(9):1609. Verfügbar unter: <https://www.mdpi.com/2073-4425/13/9/1609>
12. Gossye H, Engelborghs S, Van Broeckhoven C, van der Zee J. C9orf72 Frontotemporal Dementia and/or Amyotrophic Lateral Sclerosis. In: *GeneReviews®* [Internet]. Seattle (WA): University of Washington, Seattle; 2020 [zitiert 2. Juli 2025]. Verfügbar unter: <http://www.ncbi.nlm.nih.gov/books/NBK268647/>
13. Gijssels I, Cruts M, Van Broeckhoven C. The Genetics of C9orf72 Expansions. *Cold Spring Harb Perspect Med*. 2. April 2018;8(4):a026757.
14. Mizielinska S, Hautbergue GM, Gendron TF, van Blitterswijk M, Hardiman O, Ravits J, u. a. Genetic ALS caused by hexanucleotide repeat expansions in C9orf72. *Lancet Neurol* [Internet]. März 2025 [zitiert 7. Juli 2025];24(3):261–74. Verfügbar unter: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC12010636/>
15. Chong ZZ, Souayah N. Targeting Gene C9orf72 Pathogenesis for Amyotrophic Lateral Sclerosis. *Int J Mol Sci* [Internet]. 30. April 2025 [zitiert 7. Juli 2025];26(9):4276. Verfügbar unter: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC12072292/>
16. Gottesfeld JM. Molecular Mechanisms and Therapeutics for the GAA·TTC Expansion Disease Friedreich Ataxia. *Neurotherapeutics* [Internet]. Oktober 2019 [zitiert 7. Juli 2025];16(4):1032–49. Verfügbar unter: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6985418/>
17. Bidichandani SI, Delatycki MB, Napierala M, Duquette A. Friedreich Ataxia. In: Adam MP, Feldman J, Mirzaa GM, Pagon RA, Wallace SE, Amemiya A, Herausgeber. *GeneReviews®* [Internet]. 2025 [zitiert 2. Juli 2025]. Verfügbar unter: <http://www.ncbi.nlm.nih.gov/books/NBK1281/>
18. Reetz K, Lischewski SA, Dogan I, Didszun C, Pishnamaz M, Konrad K, u. a. Friedreich's ataxia—a rare multisystem disease. *Lancet Neurol* [Internet]. 1. Juli 2025 [zitiert 7. Juli 2025];24(7):614–24. Verfügbar unter: <https://www.sciencedirect.com/science/article/pii/S1474442225001759>
19. Caron NS, Wright GE, Hayden MR. Huntington Disease. In: Adam MP, Feldman J, Mirzaa GM, Pagon RA, Wallace SE, Amemiya A, Herausgeber. *GeneReviews®* [Internet]. Seattle (WA): University of Washington, Seattle; 2020 [zitiert 2. Juli 2025]. Verfügbar unter: <http://www.ncbi.nlm.nih.gov/books/NBK1305/>
20. Tong H, Yang T, Xu S, Li X, Liu L, Zhou G, u. a. Huntington's Disease: Complex Pathogenesis and Therapeutic Strategies. *Int J Mol Sci*. 29. März 2024;25(7):3845.

21. Jimenez-Sanchez M, Licitra F, Underwood BR, Rubinsztein DC. Huntington's Disease: Mechanisms of Pathogenesis and Therapeutic Strategies. *Cold Spring Harb Perspect Med* [Internet]. Juli 2017 [zitiert 8. Juli 2025];7(7):a024240. Verfügbar unter: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5495055/>
22. Pengo M, Squitieri F. Beyond CAG Repeats: The Multifaceted Role of Genetics in Huntington Disease. *Genes* [Internet]. 19. Juni 2024 [zitiert 17. Juli 2025];15(6):807. Verfügbar unter: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11203031/>
23. Reiner A, Dragatsis I, Dietrich P. GENETICS AND NEUROPATHOLOGY OF HUNTINGTON'S DISEASE. *Int Rev Neurobiol* [Internet]. 2011 [zitiert 8. Juli 2025];98:325–72. Verfügbar unter: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4458347/>
24. Stoker TB, Mason SL, Greenland JC, Holden ST, Santini H, Barker RA. Huntington's disease: diagnosis and management. *Pract Neurol*. Februar 2022;22(1):32–41.
25. Seifert BA, Reddi HV, Kang BE, Bean LJH, Shealy A, Rose NC, u. a. Myotonic dystrophy type 1 testing, 2024 revision: A technical standard of the American College of Medical Genetics and Genomics (ACMG). *Genet Med Off J Am Coll Med Genet*. August 2024;26(8):101145.
26. Soltanzadeh P. Myotonic Dystrophies: A Genetic Overview. *Genes*. 17. Februar 2022;13(2):367.
27. Bird TD. Myotonic Dystrophy Type 1. In: Adam MP, Feldman J, Mirzaa GM, Pagon RA, Wallace SE, Amemiya A, Herausgeber. *GeneReviews®* [Internet]. Seattle (WA): University of Washington, Seattle; 2024 [zitiert 2. Juli 2025]. Verfügbar unter: <http://www.ncbi.nlm.nih.gov/books/NBK1165/>
28. Mangin A, de Pontual L, Tsai YC, Monteil L, Nizon M, Boisseau P, u. a. Robust Detection of Somatic Mosaicism and Repeat Interruptions by Long-Read Targeted Sequencing in Myotonic Dystrophy Type 1. *Int J Mol Sci*. 5. März 2021;22(5):2616.
29. Peric S, Pesovic J, Savic-Pavicevic D, Rakocevic Stojanovic V, Meola G. Molecular and Clinical Implications of Variant Repeats in Myotonic Dystrophy Type 1. *Int J Mol Sci*. 29. Dezember 2021;23(1):354.
30. Cortese A, Reilly MM, Houlden H. RFC1 CANVAS / Spectrum Disorder. In: Adam MP, Feldman J, Mirzaa GM, Pagon RA, Wallace SE, Amemiya A, Herausgeber. *GeneReviews®* [Internet]. Seattle (WA): University of Washington, Seattle; 2020 [zitiert 2. Juli 2025]. Verfügbar unter: <http://www.ncbi.nlm.nih.gov/books/NBK564656/>
31. Delforge V, Tard C, Davion JB, Dujardin K, Wissocq A, Dhaenens CM, u. a. RFC1: Motifs and phenotypes. *Rev Neurol (Paris)*. Mai 2024;180(5):393–409.

32. Shukla S, Gupta K, Singh K, Mishra A, Kumar A. An Updated Canvas of the RFC1-mediated CANVAS (Cerebellar Ataxia, Neuropathy and Vestibular Areflexia Syndrome). *Mol Neurobiol* [Internet]. 1. Januar 2025 [zitiert 2. Juli 2025];62(1):693–707. Verfügbar unter: <https://doi.org/10.1007/s12035-024-04307-0>
33. Sullivan R, Yau WY, O'Connor E, Houlden H. Spinocerebellar ataxia: an update. *J Neurol* [Internet]. 2019 [zitiert 2. Juli 2025];266(2):533–44. Verfügbar unter: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6373366/>
34. Klockgether T, Mariotti C, Paulson HL. Spinocerebellar ataxia. *Nat Rev Dis Prim*. 11. April 2019;5(1):24.
35. Opal P, Ashizawa T. Spinocerebellar Ataxia Type 1. In: Adam MP, Feldman J, Mirzaa GM, Pagon RA, Wallace SE, Amemiya A, Herausgeber. *GeneReviews®* [Internet]. Seattle (WA): University of Washington, Seattle; 2023 [zitiert 10. Juli 2025]. Verfügbar unter: <http://www.ncbi.nlm.nih.gov/books/NBK1184/>
36. Pulst SM. Spinocerebellar Ataxia Type 2. In: Adam MP, Feldman J, Mirzaa GM, Pagon RA, Wallace SE, Amemiya A, Herausgeber. *GeneReviews®* [Internet]. Seattle (WA): University of Washington, Seattle; 2019 [zitiert 10. Juli 2025]. Verfügbar unter: <http://www.ncbi.nlm.nih.gov/books/NBK1275/>
37. Paulson H, Shakkottai V. Spinocerebellar Ataxia Type 3. In: Adam MP, Feldman J, Mirzaa GM, Pagon RA, Wallace SE, Amemiya A, Herausgeber. *GeneReviews®* [Internet]. Seattle (WA): University of Washington, Seattle; 2020 [zitiert 10. Juli 2025]. Verfügbar unter: <http://www.ncbi.nlm.nih.gov/books/NBK1196/>
38. Casey HL, Gomez CM. Spinocerebellar Ataxia Type 6. In: Adam MP, Feldman J, Mirzaa GM, Pagon RA, Wallace SE, Amemiya A, Herausgeber. *GeneReviews®* [Internet]. Seattle (WA): University of Washington, Seattle; 2019 [zitiert 10. Juli 2025]. Verfügbar unter: <http://www.ncbi.nlm.nih.gov/books/NBK1140/>
39. La Spada AR. Spinocerebellar Ataxia Type 7. In: Adam MP, Feldman J, Mirzaa GM, Pagon RA, Wallace SE, Amemiya A, Herausgeber. *GeneReviews®* [Internet]. Seattle (WA): University of Washington, Seattle; 2020 [zitiert 10. Juli 2025]. Verfügbar unter: <http://www.ncbi.nlm.nih.gov/books/NBK1256/>
40. Weese-Mayer DE, Rand CM, Khaytin I, Slattery SM, Yap KL, Marazita ML, u. a. Congenital Central Hypoventilation Syndrome. In 2021 [zitiert 2. Juli 2025]. Verfügbar unter: <http://www.ncbi.nlm.nih.gov/books/NBK1427/>
41. Zaidi S, Gandhi J, Vatsia S, Smith NL, Khan SA. Congenital central hypoventilation syndrome: An overview of etiopathogenesis, associated pathologies, clinical presentation, and management. *Auton Neurosci* [Internet]. März 2018 [zitiert 3. Juli 2025];210:1–9. Verfügbar unter: <https://linkinghub.elsevier.com/retrieve/pii/S1566070217302096>
42. Warburton PE, Sebra RP. Long-Read DNA Sequencing: Recent Advances and Remaining Challenges. *Annu Rev Genomics Hum Genet*. 25. August 2023;24:109–32.

43. Asuragen, Inc. AmplideX PCR/CE FMR1 Kit. 2025.
44. Asuragen, Inc. AmplideX PCR/CE C9orf72 Kit. 2020.
45. Entezam M, Amirfiroozi A, Togha M, Keramatipour M. Comparison of Two Different PCR-based Methods for Detection of GAA Expansions in Frataxin Gene. *Iran J Public Health*. Februar 2017;46(2):222–8.
46. Filla A, De Michele G, Cavalcanti F, Pianese L, Monticelli A, Campanella G, u. a. The relationship between trinucleotide (GAA) repeat length and clinical features in Friedreich ataxia. *Am J Hum Genet*. September 1996;59(3):554–60.
47. Ciotti P, Di Maria E, Bellone E, Ajmar F, Mandich P. Triplet repeat primed PCR (TP PCR) in molecular diagnostic testing for Friedreich ataxia. *J Mol Diagn JMD*. November 2004;6(4):285–9.
48. Asuragen, Inc. AmplideX PCR/CE HTT Kit. 2021.
49. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. The Huntington's Disease Collaborative Research Group. *Cell*. 26. März 1993;72(6):971–83.
50. Asuragen, Inc. AmplideX DM1 Dx Kit. 2025.
51. Kamsteeg EJ, Kress W, Catalli C, Hertz JM, Witsch-Baumgartner M, Buckley MF, u. a. Best practice guidelines and recommendations on the molecular diagnosis of myotonic dystrophy types 1 and 2. *Eur J Hum Genet EJHG*. Dezember 2012;20(12):1203–8.
52. Jennings LJ, Yu M, Zhou L, Rand CM, Berry-Kravis EM, Weese-Mayer DE. Comparison of PHOX2B testing methods in the diagnosis of congenital central hypoventilation syndrome and mosaic carriers. *Diagn Mol Pathol Am J Surg Pathol Part B*. Dezember 2010;19(4):224–31.
53. Singh S, Zhang A, Dlouhy S, Bai S. Detection of large expansions in myotonic dystrophy type 1 using triplet primed PCR. *Front Genet [Internet]*. 24. April 2014 [zitiert 15. November 2025];5:94. Verfügbar unter: <https://pmc.ncbi.nlm.nih.gov/articles/PMC4006065/>
54. MRC Holland. MS-MLPA General Protocol. 2024.
55. PacBio. Sequencing 101: from DNA to discovery - the steps of SMRT sequencing [Internet]. 2025. Verfügbar unter: <https://www.pacb.com/blog/steps-of-smrt-sequencing/>
56. Yu SY, Xi YL, Xu FQ, Zhang J, Liu YS. Application of long read sequencing in rare diseases: The longer, the better? *Eur J Med Genet*. Dezember 2023;66(12):104871.
57. Rhoads A, Au KF. PacBio Sequencing and Its Applications. *Genomics Proteomics Bioinformatics*. Oktober 2015;13(5):278–89.

58. PacBio. Removing short DNA fragments with the Short Read Eliminator (SRE) kit.
59. Sequeiros J, Martindale J, Seneca S. EMQN Best Practice Guidelines for molecular genetic testing of SCAs. *Eur J Hum Genet* [Internet]. November 2010 [zitiert 27. Dezember 2025];18(11):1173–6. Verfügbar unter: <https://pmc.ncbi.nlm.nih.gov/articles/PMC2987475/>
60. Dolzhenko E, English A, Dashnow H, De Sena Brandine G, Mokveld T, Rowell WJ, u. a. Characterization and visualization of tandem repeats at genome scale. *Nat Biotechnol* [Internet]. Oktober 2024 [zitiert 3. Juli 2025];42(10):1606–14. Verfügbar unter: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11921810/>
61. PacBio. Preparing whole genome and metagenome libraries using SMRTbell® prep kit 3.0. 2025.
62. PureTarget [Internet]. PacBio. [zitiert 20. Dezember 2025]. Verfügbar unter: <https://www.pacb.com/technology/puretarget/>
63. Tsai YC, de Pontual L, Heiner C, Stojkovic T, Furling D, Bassez G, u. a. Identification of a CCG-Enriched Expanded Allele in Patients with Myotonic Dystrophy Type 1 Using Amplification-Free Long-Read Sequencing. *J Mol Diagn JMD*. November 2022;24(11):1143–54.

## 10 Anhang

### 10.1 Verwendung von KI-gestützten Tools

Zur sprachlichen Optimierung des Textes wurden folgende Tools verwendet:

- Claude Sonnet 4.5

Anthropic

Datum der Generierung der Inhalte: 01.07.2025 bis 30.12.2025

<https://claude.ai>

- ChatGPT Version GPT 5.2

OpenAI

Datum der Generierung der Inhalte: 01.07.2025 bis 30.12.2025

<https://chatgpt.com>

Zusätzlich wurde Notebook LM für die Zusammenfassung von relevanten Quellen als Hilfe gebraucht. DeepL wurde für die Übersetzung und Rückübersetzung von Texten in Englisch auf Deutsch verwendet.

## 10.2 Übersicht Primer und PCR-Programme

Gen	Analyse	Primer (5'-3')		Programm
<b>FXN</b>	FA	GAA-194F	GGCTTAAACTTCCCACACGTGTT (5'FAM)	30x (94°C 30s, 60°C 30s, 72°C 30s), 72°C 10min
		GAA-629R	AGGACCATCATGGCCACACTT	
	TP PCR	P1	GCTGGGATTACAGGCGCGCGA	95°C 10min, 40x (95°C 30s, 60°C 30s, 72°C 30s), 72°C 7min
		P3	TACGCATCCCAGTTTGAGACG	
		P4	TACGCATCCCAGTTTGAGACG- GAAGAAGAAGAAGAAGAA (5'FAM)	
	Seq	GAA-194F	GGCTTAAACTTCCCACACGTGTT	94°C 5min, 35x (94°C 45s, 56°C 45s, 72°C 45s), 72°C 5min, 4°C Pause
GAA-629R		AGGACCATCATGGCCACACTT		
<b>HTT</b>	FA	HTT E1 F FAM	CCAGAGCCCCATTCATTG (FAM)	94°C 90s, 40x (94°C 1min, 60°C 1min, 72°C 2 min), 10°C Pause
		HTT E1 R	CGGTGGCGGCTGTG	
	Seq	HTT Seq E1 F	CCAGAGCCCCATTCATTG	
		HTT Seq E1 R	CTGGGTCACTCTGTCTCTG	
<b>DM1</b>	FA und Seq	DM 3'F	ctcccaggcctgcagttgcccac	94°C 3min, 8x TD -1°C pro Zyklus (95° 20s, 62-55°C, 68°C 1min), 30x (95° 20s, 55°C 1min, 68°C 1min), 68°C 10min, 10°C Pause
		DM 3'R	gaacggggctogaagggctctgtagc	
<b>PHOX2B</b>	FA	PHOX2B_polyA_pubF1	CCAGGTCCCAATCCCAAC	95° 10min, 2x (98° 40s, 64° 40s, 72° 45s), 5x (98° 40s, 62° 40s, 72° 45s) 8x (98° 40s, 60° 40s, 72° 45s), 30x (98° 40s, 58° 40s, 72° 45s), 72° 10min, 10° Pause
		PHOX2B_polyA_pubR1	GAGCCCAGCCTTGCCAG	
	Seq	PHOX2BPolyASEQPubF1	CCAGGTCCCAATCCCAAC	
		PHOX2BPolyASEQPubR1	GAGCCCAGCCTTGCCAG	
		PHOX2BE3PolyASEQF	AGCGACGACAATAGCCTTGG	
		PHOX2BE3PolyASEQR	CAAGAGCACTGACCCGGAC	

Tabelle 16 Übersicht Primer und PCR-Programme für FXN, HTT, DM1, PHOX2B. FA: Fragmentanalyse, Seq: Sequenzierung

Gen	Analyse	Primer	PCR-Programm
<b>ATXN1 (SCA1)</b>	FA	SCA1 F FAM	CAACATGGGCAGTCTGAG
		SCA1 R	AACTGGAAATGTGGACGTA
<b>ATXN2 (SCA2)</b>	FA	SCA2 FAM	GGGCCCTCACCATGTCTG
		SCA2 R	CGGGCTTGCGGACATTGG
	TP-PCR	SCA2 TP F	GGGCCCTCACCATGTCTG
		SCA2_7 TP R	TACGCATCCCAGTTTGAGACGCTGCTGCTGCTGCTG
<b>ATXN3 (SCA3)</b>	FA	SCA3 F FAM	CCAGTGACTACTTTGATTCCG
		SCA3 R	TGGCCTTTCACATGGATGTGA
<b>CACNA1A (SCA6)</b>	FA	SCA6 F FAM	CACGTGTCCTATTCCCCTGTGATCC
		SCA6 R	TGGGTACCTCCGAGGGCCGCT
<b>ATXN7 (SCA7)</b>	FA	SCA7 F FAM	TGTTACATTGTAGGAGCGGAA
		SCA7 R	CACGACTGTCCCAGCATCACTT
	TP-PCR	SCA7 TP F	GCGGTCCCAAAGGGTCAGTTGTTACATTGTAGGAGCGGAA
		SCA7_7 TP R	TACGCATCCCAGTTTGAGACGCTGCTGCTGCTGCTG

Tabelle 17 Übersicht Primer und PCR Programme SCAS. FA: Fragmentanalyse.  $\alpha$ : 94°C 5min, 35 Zyklen (94°C 45s, 56°C 45s, 72°C 45s), 72°C 5min, 4°C Pause,  $\beta$ : 94°C 3min, 8x TD -1°C pro Zyklus (95° 20s, 62-55°C, 68°C 1min), 30x (95° 20s, 55°C 1min, 68°C 1min), 68°C 10min, 10°C Pause

10.3 Waterfall-Plots

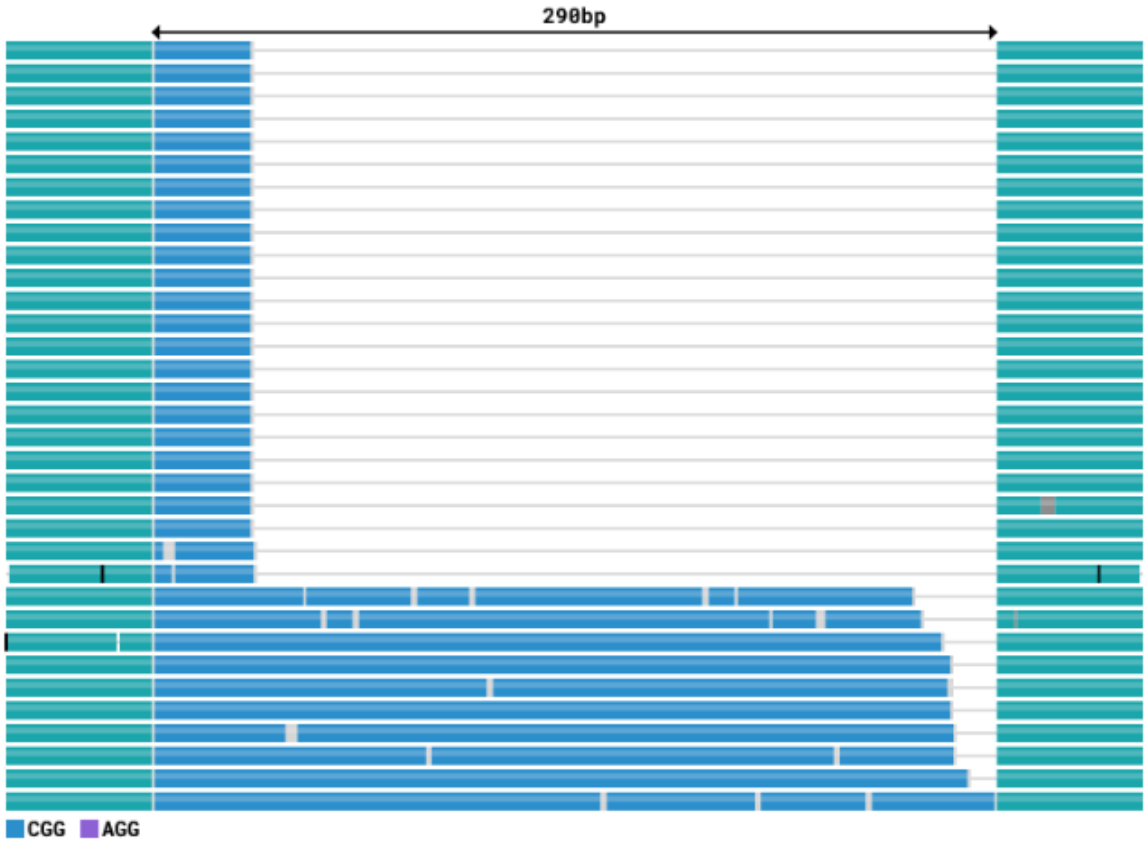


Abbildung 16 Waterfall-Plot C01-XD01 (FMR1)

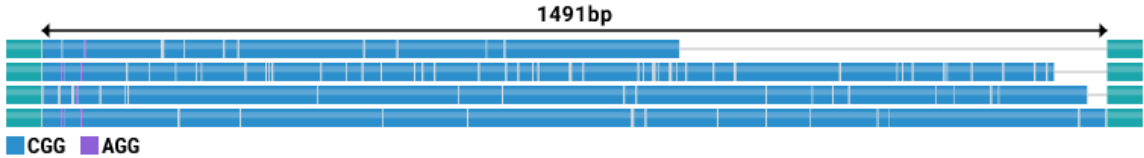


Abbildung 17 Waterfall-Plot C02-XD02 (FMR1)

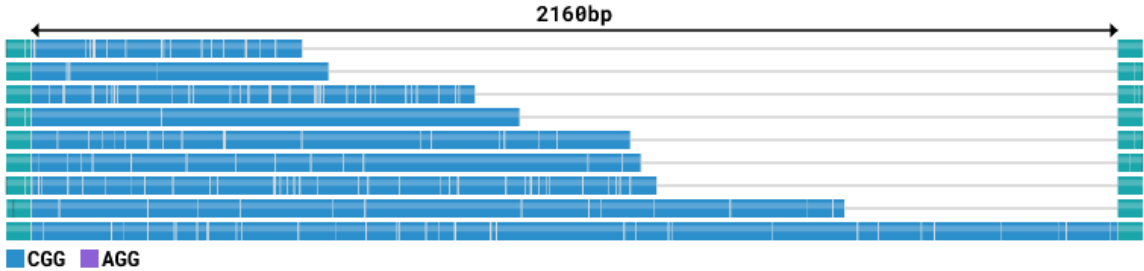


Abbildung 18 Waterfall-Plot C05-XD05 (FMR1)

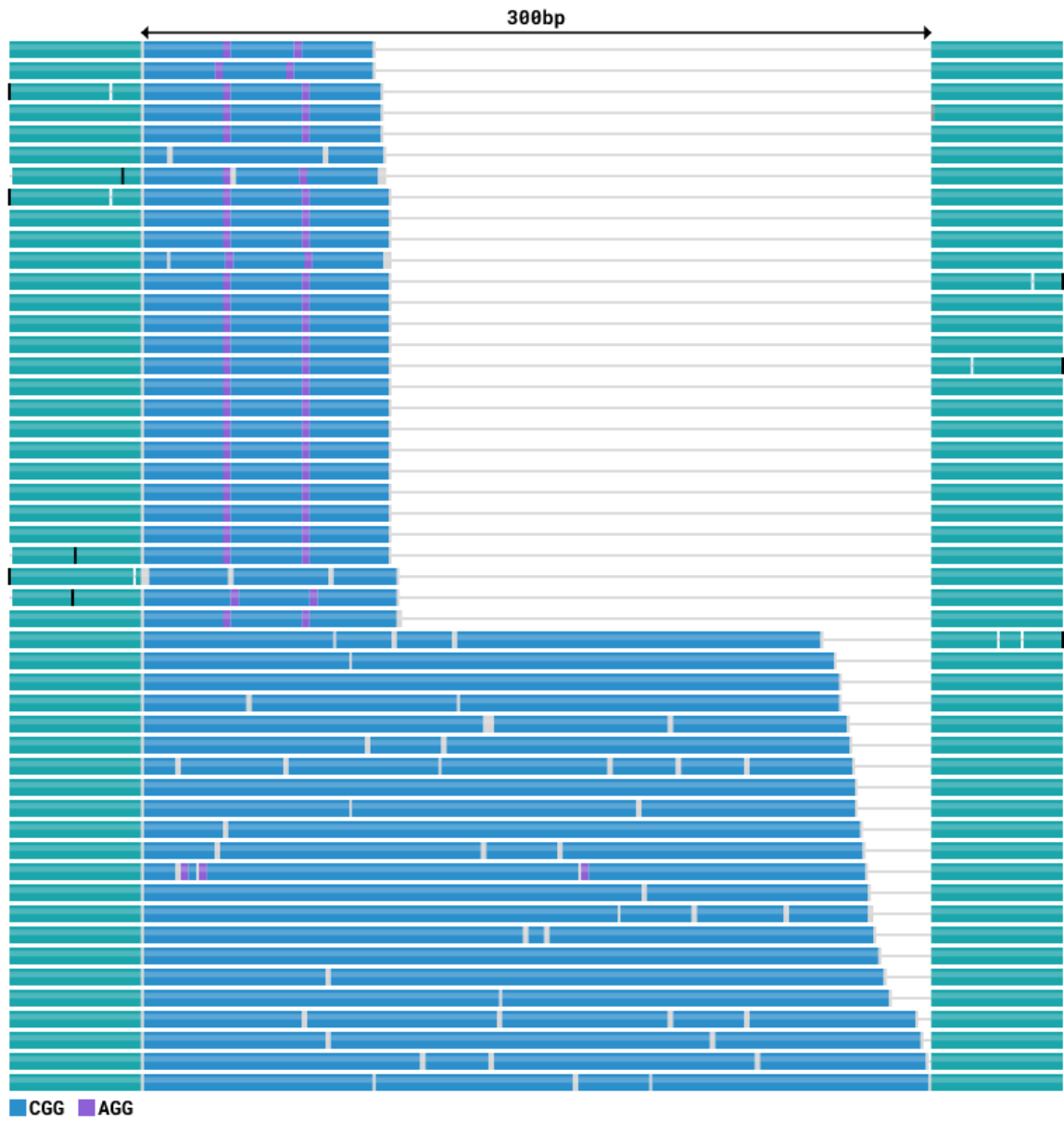


Abbildung 19 Waterfall-Plot C06-XD06 (FMR1)

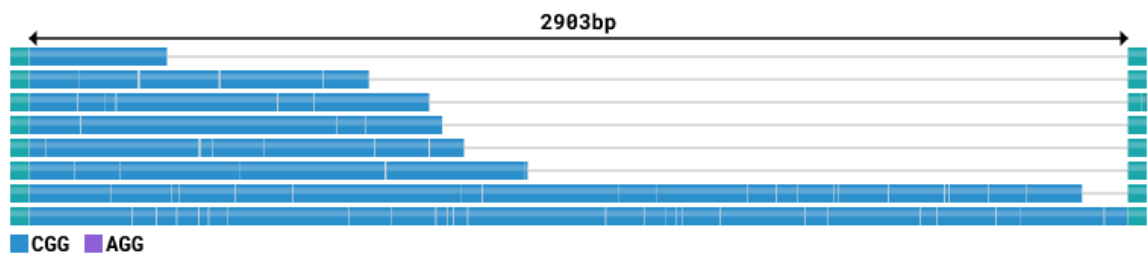


Abbildung 20 Waterfall-Plot C07-XD07 (FMR1)

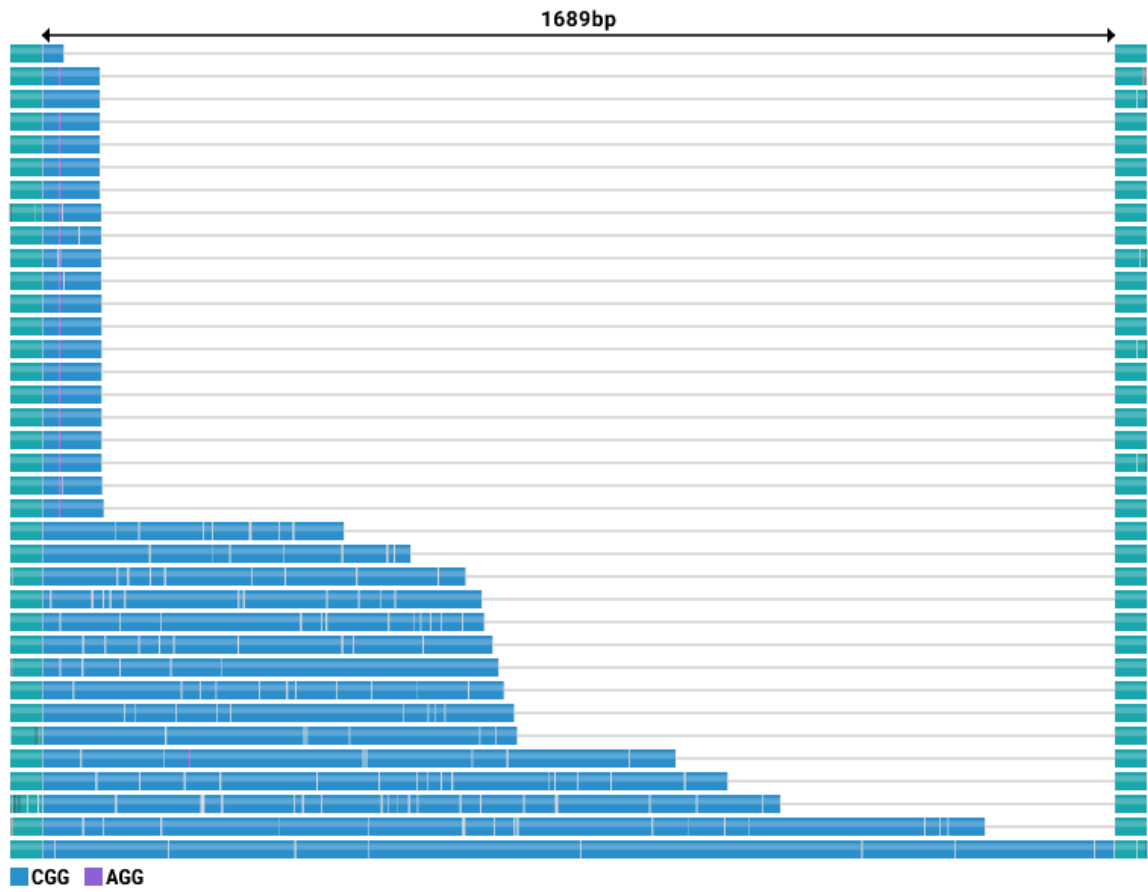


Abbildung 21 Waterfall-Plot C09-XD09 (FMR1)

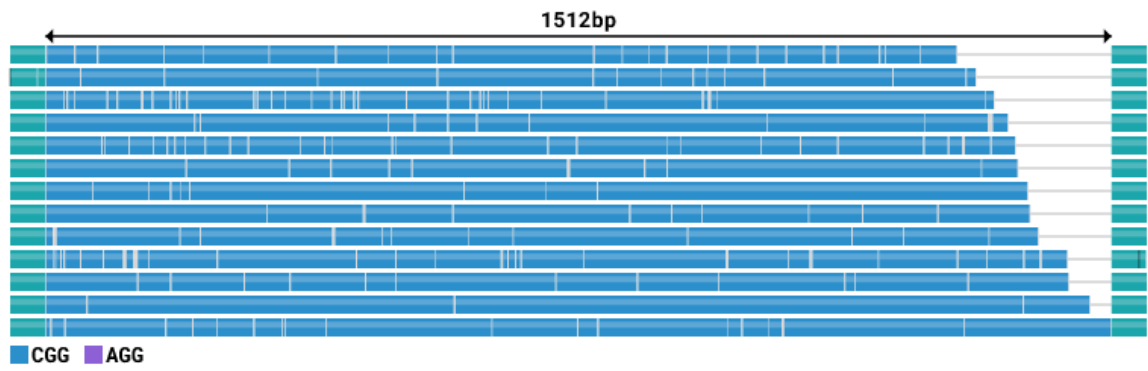


Abbildung 22 Waterfall-Plot C10-XD10 (FMR1)

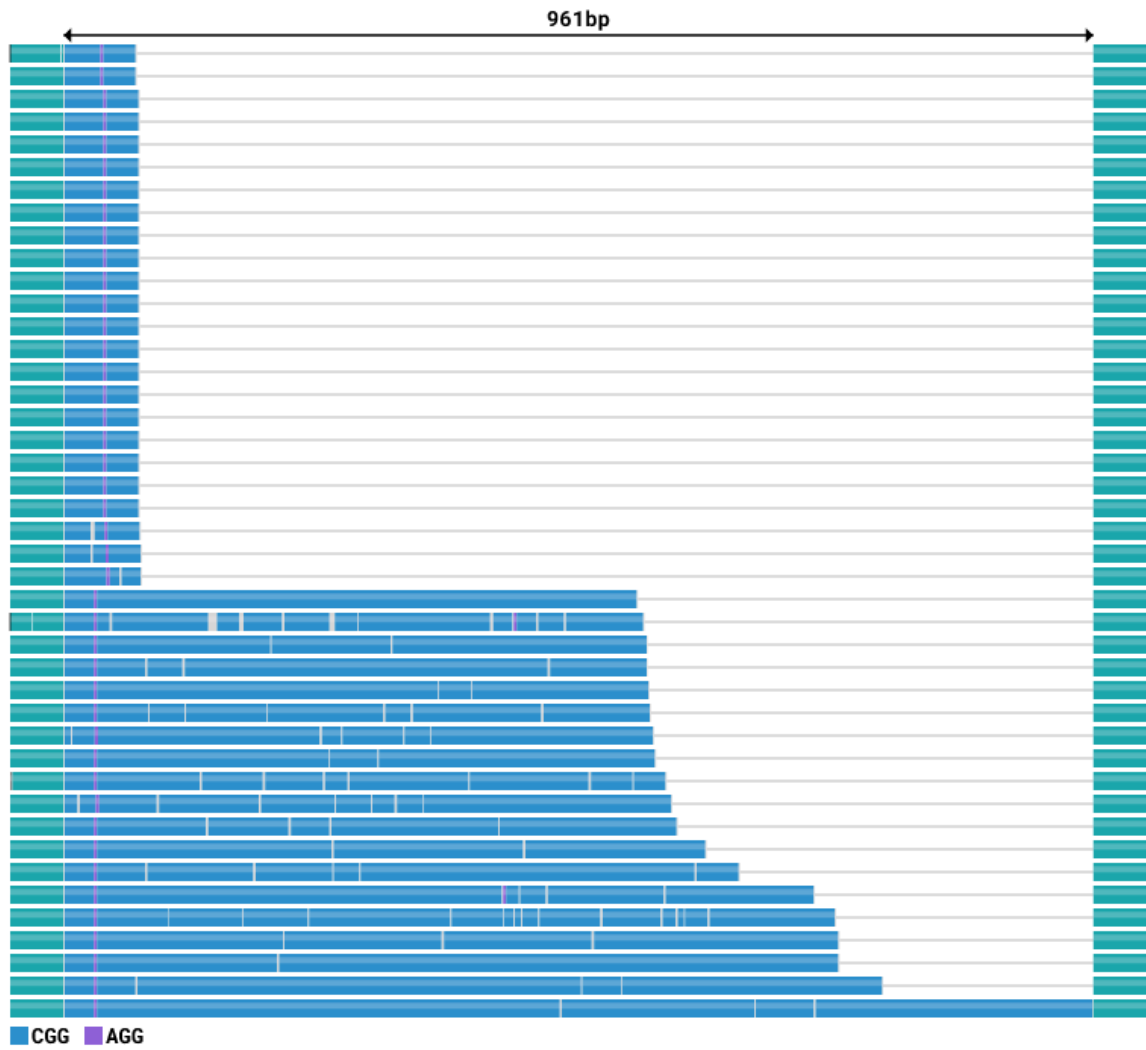


Abbildung 23 Waterfall-Plot C17-XD17 (FMR1)

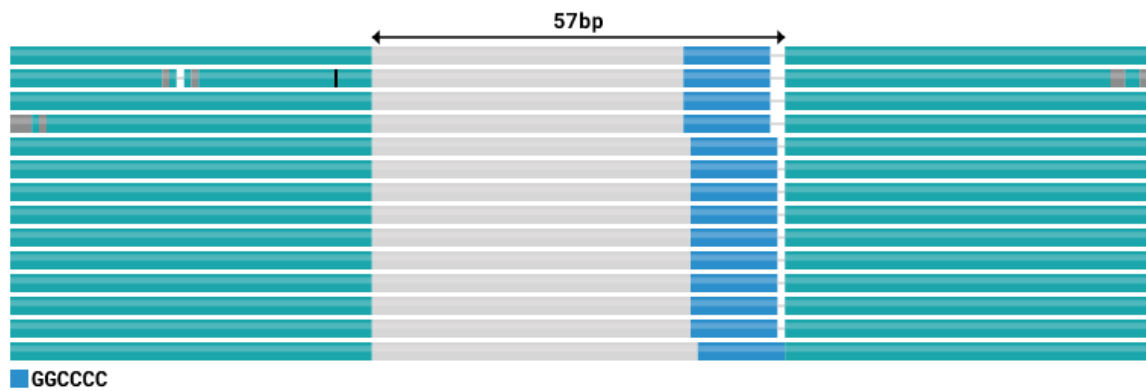


Abbildung 24 Waterfall-Plot C20-CD03 (C9orf72)

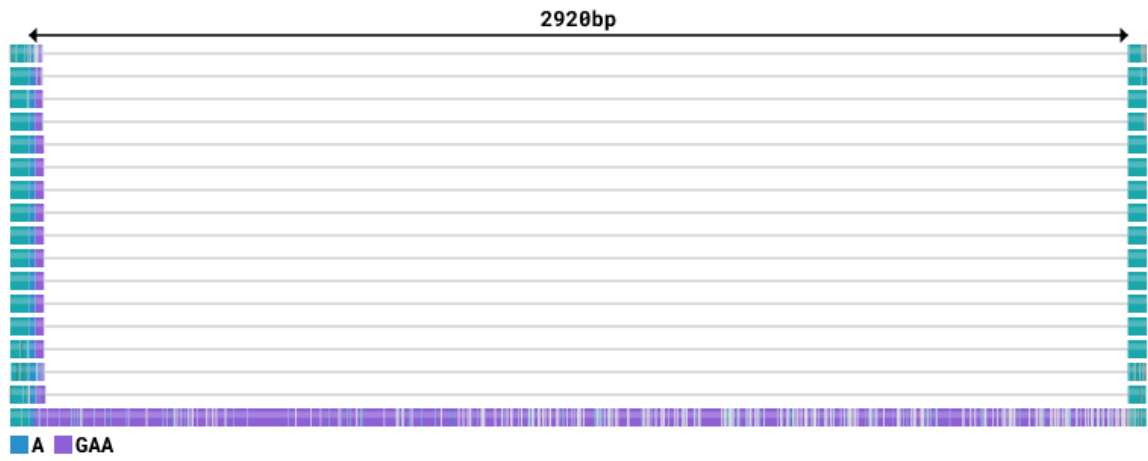


Abbildung 25 Waterfall-Plot C28-FK01 (FXN)



Abbildung 26 Waterfall-Plot C29-FK02 (FXN)

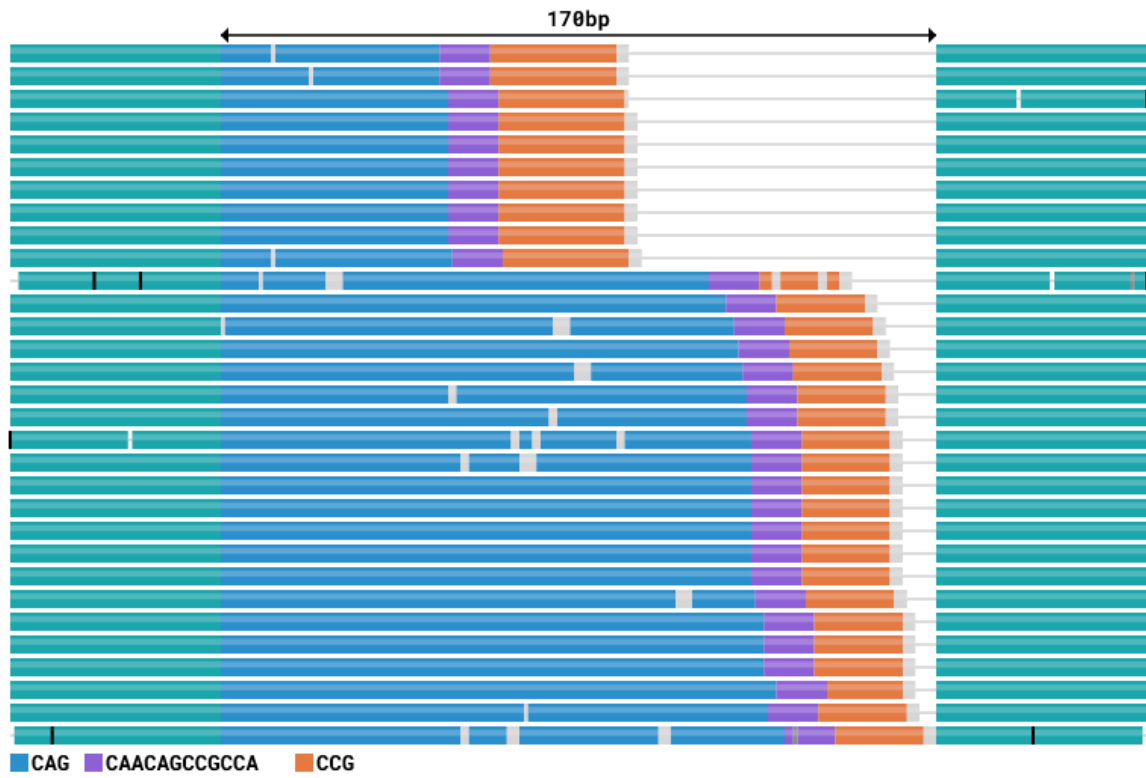


Abbildung 27 Waterfall-Plot C34-HD01 (HTT)

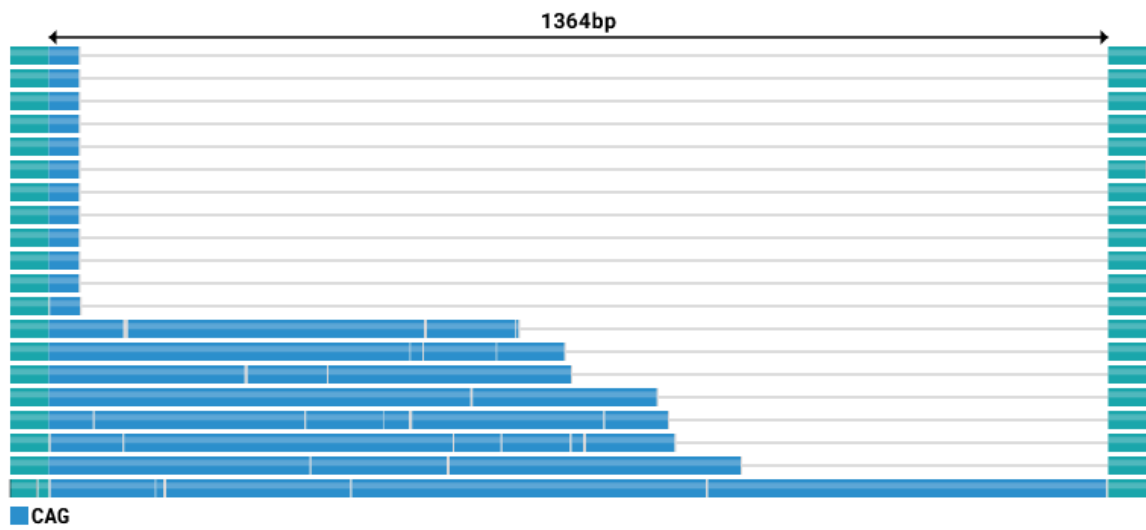


Abbildung 28 Waterfall-Plot C41-D1D03 (DMPK)

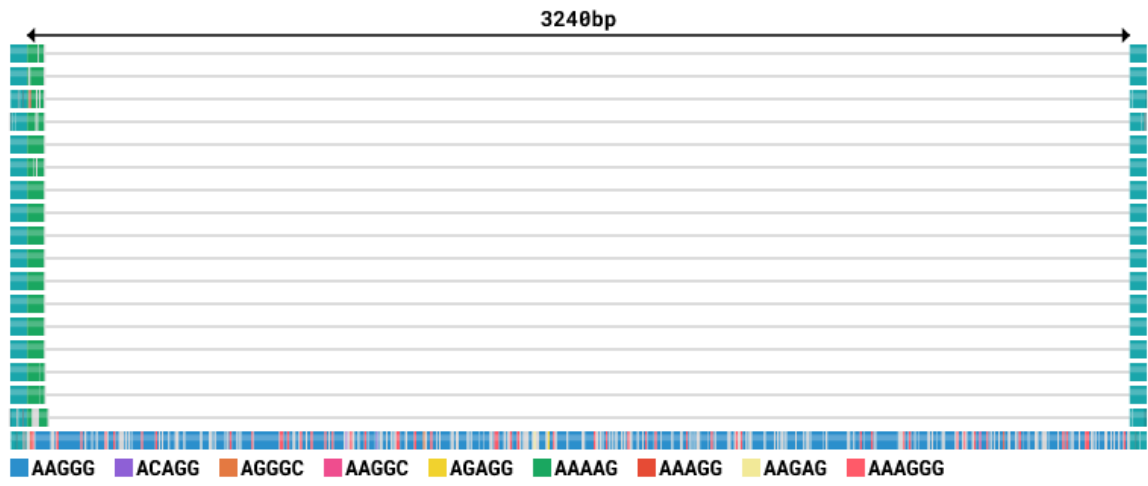


Abbildung 29 Waterfall-Plot C44-CSK01 (RFC1)

## 10.4 Übersicht Rohdaten

	Pat. Nr.	Geschlecht	Gen	bp Allel 1	bp Allel 2
1	C01-XD01	w	FMR1	264.7	493.82 + Expansion Stutter
2	C02-XD02	m	FMR1	Zwei Peaks sichtbar bei 935.84 und 971.29 + Expansion Stutter	
3	C03-XD03	m	FMR1	Zwei Peaks sichtbar bei 929.2 und 961.6 + Expansion Stutter	
4	C04-XD04	m	FMR1	Drei Peaks sichtbar bei 820, 856,7 und 980 + Expansion Stutter	
5	C05-XD05	m	FMR1	Zwei Peaks sichtbar bei 772.53 und 984.24 + Expansion Stutter	
6	C06-XD06	w	FMR1	322.76	498.55 + Expansion Stutter
7	C07-XD07	m	FMR1	Zwei Peaks sichtbar bei 649.73 987.79 + Expansion Stutter	
8	C08-XD08	m	FMR1	≥900 + Expansion Stutter	
9	C09-XD09	w	FMR1	324.96	832.13 + Expansion Stutter
10	C10-XD10	m	FMR1	≥900 + Expansion Stutter	
11	C11-XD11	m	FMR1	Vier Peaks bei 715.15, 805.99, 998.05 und 1127 + Expansion Stutter	
12	C12-XD12	m	FMR1	Rohdaten nicht aufgefunden, mit den Daten aus dem LIS gearbeitet	
13	C13-XD13	m	FMR1	Zwei Peaks bei 395.22 und 449.09	
14	C14-XR14	m	FMR1	Zwei Peaks bei 504.23 und 507.22	
15	C15-XR15	w	FMR1	325.98	ca. 950 + Expansion Stutter
16	C16-XD16	w	FMR1	288.22	1063.48 + Expansion Stutter
17	C17-XD17	w	FMR1	301.03	745.87
18	C18-CD01	w	C9orf72	164.79	362.3
19	C19-CD02	m	C9orf72	149.47	1088.16 + Expansion Stutter
20	C20-CD03	w	C9orf72	137.96	1116.48 + Expansion Stutter
21	C21-CR04	w	C9orf72	163	980 + Expansion Stutter
22	C22-CR05	w	C9orf72	128.69	128.69
23	C23-CR06	w	C9orf72	127.18	>900 Expansion Stutter
24	C24-CR07	m	C9orf72	144.5	980 + Expansion Stutter
25	C25-CD08	m	C9orf72	130.4	187.91
26	C26-CD09	m	C9orf72	128.64	163.56
27	C27-CD10	w	C9orf72	153.61	165.98
28	C28-FK01	w	FXN	82.69	723.3 + Expansion Stutter
29	C29-FK02	w	FXN	84.98	740 + Expansion Stutter
30	C30-FW03	k. A.	FXN	85.62	497.34 + Expansion Stutter
31	C31-FW04	k. A.	FXN	296.76	488.39 + Expansion Stutter
32	C32-FR05	m	FXN	beide Allele Expansion Stutter	

33	C33-FR06	w	FXN	91.31	112.27
34	C34-HD01	m	HTT	121.1	191.11
35	C35-HD02	w	HTT	120.9	193.88
36	C36-HW03	k. A.	HTT	123.58	207.91
37	C37-HR04	m	HTT	199.45	263.57
38	C38-HR05	w	HTT	199.63	208.23
39	C39-D1W01	k. A.	DMPK	245.68	569.47 + Expansion Stutter
40	C40-D1W02	k. A.	DMPK	251.4	Expansion Stutter
41	C41-D1D03	w	DMPK	250.85	ca. 980 + Expansion Stutter
42	C42-D1R04	m	DMPK	271.96	Expansion Stutter
43	C43-D1R05	m	DMPK	228.21	247.52
44	C44-CSK01	w	RFC1	keine Rohdaten vorhanden	
45	C45-S1W01	k. A.	ATXN1	212.52	237.85
46	C46-S1R02	w	ATXN1	213.01	246.95
47	C47-S1R03	m	ATXN1	207.2	271.24
48	C48-S2W01	k. A.	ATXN2	101.06	142.47
49	C49-S3W01	k. A.	ATXN3	229.57	378.15
50	C50-S3R02	m	ATXN3	226.76	366.09
51	C51-S5W01	k. A.	CACNA1A	135.88	163.24
52	C52-S7W01	k. A.	ATXN7	163.02	ca. 300
53	C53-S7R02	w	ATXN7	163.18	305.7
54	C54-PD01	m	PHOX2B	219.64	219.64
55	C55-PD02	w	PHOX2B	219.97	231.09
56	C56-PD03	w	PHOX2B	219.64	219.64
57	C57-PD04	w	PHOX2B	219.97	231.09

*Tabelle 18 Übersicht Rohdaten. w: weiblich, m: männlich, k.A: Keine Angaben*