

# Computational Short Tandem Repeat Genotyping Reveals Clinically Relevant Expansions in a large Turkish Neurodegeneration Cohort

[Zakhiriddin Khojakulov](#) , Robin J. Palvadeau , [Müge Kovancılar-Koç](#) , [Irmak Atay](#) , [Irmak Şahbaz](#) , Şeyma Tekgül , [Ayça Şahin](#) , Esmer Zeynep Duru Badakal , [Tuğçe Gül-Demirkale](#) , [Vildan Çiftçi](#) , Elif Bayraktar , Ceren Tunca , Natalia Smolina , Fulya Akçimen , [A. Nazlı Basak](#) \*

Posted Date: 10 March 2026

doi: 10.20944/preprints202603.0755.v1

Keywords: neurodegenerative diseases; computational genotyping; ExpansionHunter; short tandem repeats; STR; NGS



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# Computational Short Tandem Repeat Genotyping Reveals Clinically Relevant Expansions in a large Turkish Neurodegeneration Cohort

Zakhiriddin Khojakulov <sup>1</sup>, Robin J. Palvadeau <sup>1</sup>, Müge Kovancılar-Koç <sup>1</sup>, Irmak Atay <sup>1</sup>, Irmak Şahbaz <sup>1</sup>, Şeyma Tekgül <sup>1</sup>, Ayça Şahin <sup>1</sup>, Esmer Zeynep Duru Badakal <sup>1</sup>, Tuğçe Gül-Demirkale <sup>1</sup>, Vildan Çiftçi <sup>1</sup>, Elif Bayraktar <sup>1</sup>, Ceren Tunca <sup>1</sup>, Natalia Smolina <sup>1</sup>, Fulya Akçimen <sup>2</sup> and A. Nazlı Başak <sup>1,\*</sup>

<sup>1</sup> Neurodegeneration Research Laboratory (NDAL)-Research Center for Translational Medicine (KUTTAM), School of Medicine, Koç University, Istanbul, Turkey

<sup>2</sup> Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA

\* Correspondence: nbasak@ku.edu.tr

## Abstract

Short tandem repeat (STR) expansions are a major cause of neurodegenerative disorders; however, their genetic and clinical heterogeneity complicates diagnosis. STR detection remains limited in routine short-read next-generation sequencing (NGS) workflows. We evaluated the diagnostic yield and clinical utility of computational STR genotyping in a large Turkish neurodegenerative disease cohort. ExpansionHunter was applied to NGS data from 3,150 patients and 146 controls, targeting 15 disease-associated STR loci. To improve genotyping of poorly captured exonic regions in exome data, the default locus coverage threshold was reduced from 10X to 3X. Candidate expansions were visually inspected using REViewer and validated by conventional molecular methods. Computational analysis identified 28 pathogenic and 160 intermediate expansions. Of these, 23 were confirmed as pathogenic, and eight initially classified as intermediate were reclassified as pathogenic after conventional validation, resulting in 31 pathogenic cases across 28 families: *HTT* (n=8), *ATXN2* (n=5), *ATXN1* (n=4), *DMPK* (n=3), *PABPN1* (n=3), *TBP* (n=2), and single cases in *AR*, *ATN1*, and *CACNA1A*. Lowering the coverage threshold markedly increased genotyping rates at low-coverage loci in exome, particularly in *ATXN2*. Genetic findings were largely consistent with clinical pre-diagnosis and the additional diagnostic yield was 0.95%. These findings support integrating STR analysis into routine neurogenetic diagnostics.

**Keywords:** neurodegenerative diseases; computational genotyping; ExpansionHunter; short tandem repeats; STR; NGS

## 1. Introduction

The human genome contains over one million tandem repeat loci, accounting for approximately 8% of the genome and contributing significantly to genomic variation. These loci are inherently unstable and exhibit high mutation rates, leading to substantial genetic variation. Among them, short tandem repeats (STRs), defined as repeating motifs of 2–6 base pairs, have been implicated in a wide range of diseases; over 50 loci are known to be associated with human disorders, particularly neurological conditions [1,2].

Repeat expansion disorders affect an estimated 1 in 3,000 individuals [3]. These diseases are genetically and clinically heterogeneous, complicating genotype-phenotype correlations. Among the most commonly implicated genes are *ATXN1*, *ATXN2*, *ATXN3*, *ATXN7*, *ATN1*, *CACNA1A*, *PPP2R2B*, and *TBP*, where (CAG/CTG)<sub>n</sub> expansions are associated with different subtypes of spinocerebellar

ataxia (SCA) [4]. Expansions in *ATXN2* and *ATXN3* extend beyond ataxia: mildly elevated pathogenic *ATXN2* repeats have been reported in familial Parkinsonism [5], and intermediate-range (premutation) alleles are associated with increased risk of amyotrophic lateral sclerosis (ALS) [6]. In contrast, *ATXN3* expansions have been implicated in Parkinson's disease (PD) [7]. Pathogenic *HTT* repeat expansion has also been reported in patients with atypical motor symptoms at onset, including PD, dystonia, ataxia, ALS, and frontotemporal dementia (FTD) patients [8,9].

Taken together, these observations demonstrate how repeat expansions transcend traditional disease boundaries and exhibit variation across diverse populations. This heterogeneity highlights the need for robust and scalable genotyping methods, since standard short-read next-generation sequencing (NGS) pipelines often fail to detect STRs. Specialized tools, such as ExpansionHunter (EH), have been developed to estimate repeat sizes from short-read data, enabling the reliable detection of clinically relevant expansions [1,3,8,10–16].

In this study, we applied EH v5 to a large Turkish cohort of 3,296 samples, including 3,150 patients with movement and neuromuscular disorders and 146 matched controls. Although in silico genotyping approaches allow interrogation of numerous disease-associated STR loci, we restricted our analysis to 15 repeats that are clinically relevant to the disorders represented in our cohort and sufficiently captured by exome data. By systematically screening these loci, we identified 30 pathogenic expanded alleles in individuals presenting both canonical and atypical phenotypes. These findings underscore the value of systematic, large-scale STR screening for refining genotype–phenotype relationships across neurodegenerative and neuromuscular disorders.

## 2. Materials and Methods

### 2.1. Study Cohort

The study cohort includes 3,150 cases and 146 healthy controls collected over 20 years across Türkiye, including a small number of cases with Azerbaijan origin, all evaluated at the Neurodegeneration Research Laboratory (NDAL). Genomic DNA was extracted from whole blood using the Qiagen EZ1 Advanced XL system. Whole-exome sequencing (WES) was conducted on individuals diagnosed with neurodegenerative, neurological, or ultra-rare genetic disorders (2,240 cases). Whole genome sequencing (WGS) was carried out as part of Project MinE on 910 sporadic ALS and ALS/FTD patients, along with 146 neurologically healthy controls (<https://projectmine.com/>). Typically, NGS was performed on affected probands who were clinically diagnosed by expert clinicians on neuromuscular and / or movement disorders. EH was applied to the entire cohort without excluding the samples with previously identified gene variants.

### 2.2. Ethics Approval and Consent to Participate

This study was approved by the Institutional Review Boards of Boğaziçi and Koç Universities. All procedures were conducted in accordance with the Declaration of Helsinki and institutional ethical guidelines. Written informed consent was obtained from all participants, including consent to publish individual data and pedigree charts.

### 2.3. Alignment of Short-Read Sequencing Data

Whole-genome sequencing data were processed within the framework of Project MinE, aligned to GRCh38, and prepared using the project's standardized pipeline as described in van Rheenen, W. et al. [17].

The majority of exome samples were sequenced (Macrogen, South Korea) using Agilent SureSelect Human All Exon V6 kit with 150 bp paired-end reads (Supplementary Table S1). Raw FASTQ files were quality-checked using FastQC v0.11.9 to assess per-cycle quality, adapter content, and GC bias. Trimming of adapters and low-quality bases was performed with fastp v0.23.4 using default parameters [18,19].

The cleaned reads were aligned to the GRCh38 human reference genome using BWA-MEM v0.7.17 with the -M option were piped directly to samtools v1.14 for sorting and indexing. Mate-pair information was corrected using GenomeAnalysisToolkit (GATK) FixMateInformation (v4.6.1), and polymerase chain reaction (PCR)/optical duplicates were flagged using GATK MarkDuplicates [20–22].

Base quality scores were recalibrated using GATK BaseRecalibrator and ApplyBQSR, producing final alignment files that were duplicate-marked, recalibrated, and coordinate-sorted. These served as input for downstream analyses.

#### 2.4. *In Silico* Genotyping of Short Tandem Repeat Loci

Computational STR genotyping was performed using EH v5 on the aligned files [10]. The variant catalog was customized to target codon-matching repeat motifs, and comprising motif interruptions, such as in the *HTT* locus.

Given the variable coverage of exome across sequencing kits compared with genome (Supplementary Figure S1 and S2), the minimum required locus coverage was reduced from the default of 10X to 3X to improve the genotyping rate [15,16]. Fifteen disease-associated STR loci: *AR*, *ATN1*, *ATXN1*, *ATXN2*, *ATXN3*, *ATXN7*, *CACNA1A*, *DMPK*, *HTT*, *JPH3*, *PABPN1*, *PPP2R2B*, *TBP*, *THAP11*, *ZFH3* were analyzed. Since patients with ALS and FTD were routinely genotyped for the pathogenic *C9orf72* repeat expansion prior to NGS, this locus was not included in the present study [23]. Accordingly, patients with alleles exceeding the established pathogenic threshold (>30 repeat units) [1] were not sequenced for either exome or genome. Furthermore, as the *C9orf72* repeat is located within a non-coding intronic region that is not reliably captured by most exome sequencing kits [1,10], its assessment from exome data would not be technically appropriate. The distribution of *C9orf72* repeat sizes in our genome dataset is presented separately in Supplementary Figure S3.

#### 2.5. Validation of Potential Repeat Expansions

PCR-based validation was performed using Sanger sequencing or fragment analysis (Macrogen, Korea), depending on the locus characteristics for all pathogenic repeats, except for *AR* carriers and some of intermediate expansions (Supplementary Table S2).

Sanger results were visualized and analyzed using FinchTV v1.4.0 (Geospiza Inc., Seattle, WA, USA), and fragment analysis results were analyzed with Peak Scanner Software v1.0 (Thermo Fisher Scientific, MA, USA).

#### 2.6. Data Analysis and Visualization

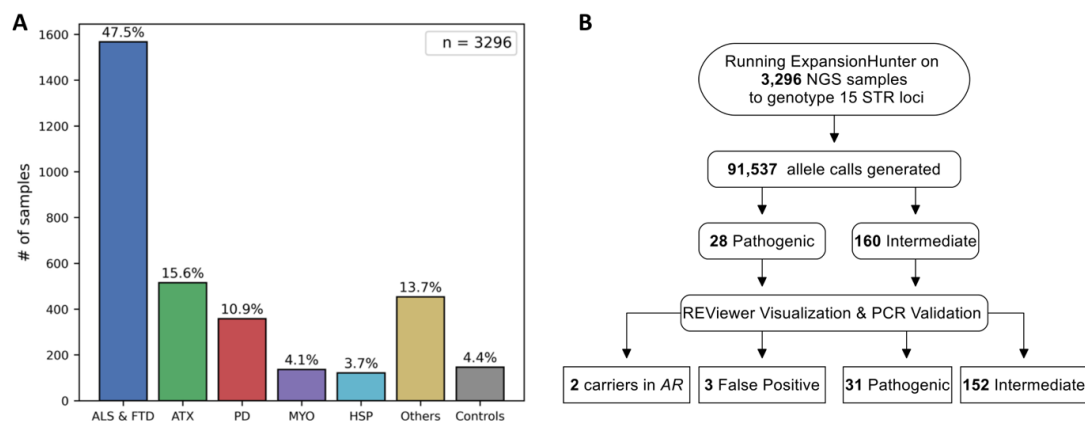
Repeat sizes inferred by EH were classified as normal, intermediate, or pathogenic according to the thresholds listed in Supplementary Table S2. Data analyses and visualizations were conducted in Python 3. Finally, clinically relevant expanded alleles identified by EH were visually inspected using REViewer v0.2.7 to observe alignment patterns and motif interruptions [24]. As REViewer is specifically designed to visualize ExpansionHunter output, EH was used to genotype STRs at loci of interest in this study.

### 3. Results

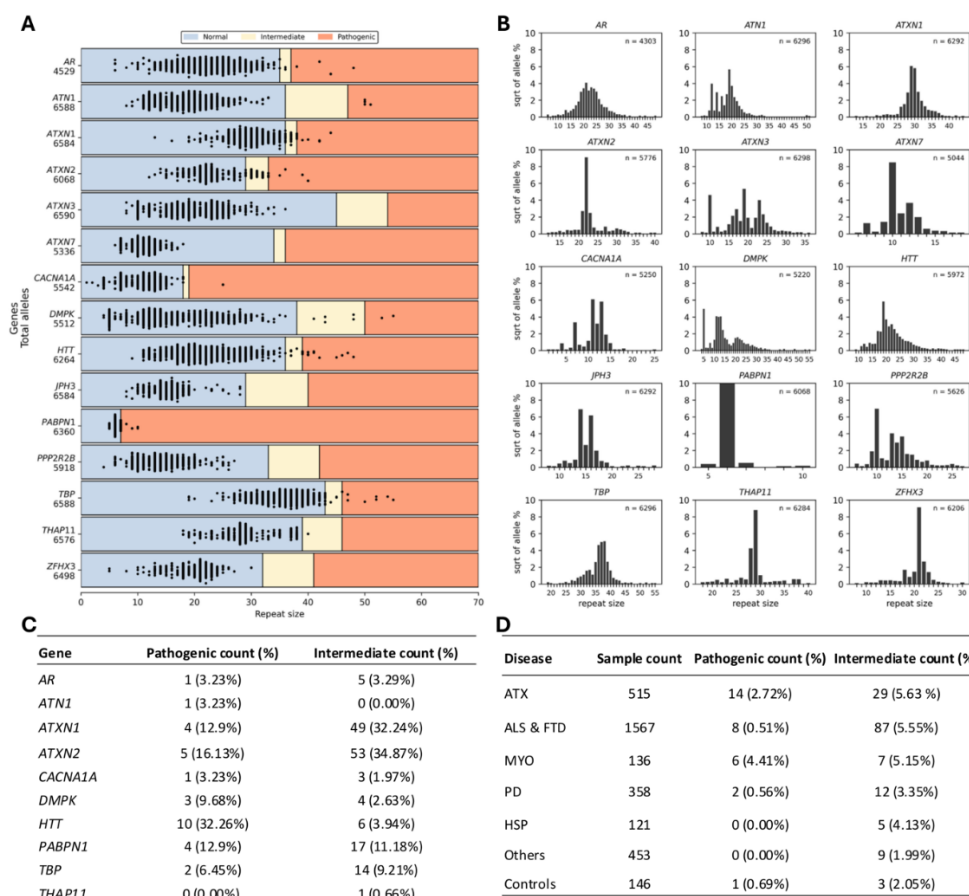
#### 3.1. Study Overview

The study cohort comprised 3,296 individuals, including 3,150 patients with neurodegenerative and /or neuromuscular disorders and 146 healthy controls (Figure 1A). Of these, 1,056 underwent genome (including controls), and 2,240 exome sequencing. Fifteen disease-associated STR expansions, which are clinically relevant to the disorders represented in this cohort, were genotyped using EH v5. Variants flagged as “LowDepth” in EH output were not included in downstream analysis (Figure 1B).

Since each individual has two alleles per locus, except for *AR* in males, a total of 91,537 allele calls were generated across the entire cohort. Figure 2A illustrates the repeat size distribution across the entire cohort based on molecular categorization (normal, intermediate, pathogenic), whereas Figure 2B presents the allele frequency distributions across loci in patients. The corresponding distribution in controls is shown in Supplementary Figure S4. Based on the figures, STR sizes in five loci (*ATXN2*, *ATXN7*, *PABPN1*, *THAP11*, and *ZFH3*) exhibited marked allelic homogeneity, with a single repeat size accounting for 60% to over 90% of observed alleles, in contrast to the greater variability observed at the remaining loci.



**Figure 1. Phenotypes of the study cohort and overview of the workflow.** **A.** Disease distribution of the cohort. The bar labeled “Others” represents patients diagnosed with Alzheimer’s disease, leukodystrophy, spinal muscular atrophy, neurodegeneration with brain iron accumulation and other rare inherited disorders. **B.** Schematic representation of workflow. PCR-based conventional validation led to the reclassification of eight intermediate expansions as pathogenic. ALS: amyotrophic lateral sclerosis, DYS: dystonia, FTD: frontotemporal dementia, HSP: hereditary spastic paraplegia, MYO: myopathy, PD: Parkinson’s disease.



**Figure 2. Repeat size distributions.** **A.** Swim Lane plot showing the repeat sizes of STRs genotyped by EH. Samples are represented by two dots corresponding to two alleles, except for males at the *AR* locus. The total number of alleles per locus is displayed below each gene. **B.** Square root-transformed percentage distribution of repeat sizes across 15 loci in patients. Figures **C.** Summarize of locus-specific and **D.** disease-category distribution of pathogenic and intermediate expansions following PCR refinement. ALS: amyotrophic lateral sclerosis, ATX: ataxia, FTD: frontotemporal dementia, MYO: myopathy, PD: Parkinson's disease.

### 3.2. Clinically Relevant Repeat Expansions

In this cohort, 14 pathogenic repeat expansions in 10 families had been previously identified using conventional diagnostic methods. EH reports both genotype estimates and corresponding confidence intervals, and all previously known pathogenic expansions were correctly detected. In each case, PCR-based repeat sizes fell within or were consistent with the reported genotype confidence intervals, providing internal validation of the in silico approach (Supplementary Table S3).

In addition to these known cases, EH identified 28 pathogenic and 160 intermediate repeat expansions, yielding a total of 188 potentially clinically relevant alleles across 183 individuals. Among these, three individuals harbored two intermediate expansions at distinct loci, and two individuals carried same-sized intermediate expansions in the *DMPK* locus.

Given that *AR*-associated repeat expansion disorders are inherited in an X-linked recessive manner, males require one pathogenic allele for disease manifestation, whereas females are typically asymptomatic carriers unless harboring two expanded alleles [1]. In our cohort, two pathogenic *AR* expansions were identified in females consistent with carrier status and were therefore not included in validation analyses. In addition, REViewer visualization revealed three potential false-positive calls at the *AR* (in one male), *ATXN1*, and *TBP* loci, due to insufficient reads supporting the expanded alleles or gaps in the alignment structure. (Supplementary Figures S5–S7). PCR validation demonstrated that all three fell within the normal range. Notably, in two of these cases (*AR* and *TBP*), sequencing read lengths were 100 bp, shorter than the minimum pathogenic repeat threshold. All three false positives originated from exome data.

After excluding the two *AR* carrier cases and three false positives, PCR-based validation confirmed 23/28 EH-detected pathogenic alleles. Additionally, eight alleles initially estimated as intermediate by EH: four in *ATXN2*, three in *HTT*, and one in *DMPK* were shown to be pathogenic by PCR. This increased the total number of pathogenic STR expansions identified in the study to 31, across nine loci in 28 families.

The distribution of pathogenic expansions among index cases included: *HTT* (n = 8), *ATXN2* (n = 5), *ATXN1* (n = 4), *DMPK* (n = 3), *PABPN1* (n = 3), *TBP* (n = 2), and one case each involving *AR*, *ATN1*, and *CACNA1A*. The corresponding demographic, clinical, and molecular characteristics of each family, including prediagnosis, age at onset, family history, sequencing approach, EH repeat estimates, PCR-confirmed repeat sizes, and brief clinical descriptions of probands are summarized in Table 1. In light of the large number of findings across multiple disease loci, comprehensive clinical characterization, including longitudinal disease progression, neuroimaging, and detailed neurological assessment, will be presented in dedicated disease-specific follow-up studies.

After PCR-based reclassification of eight intermediate alleles as pathogenic, the total number of intermediate expansions decreased from 160 to 152. Figures 2C and 2D summarize the locus-specific and disease-category distribution of pathogenic and intermediate expansions following PCR refinement. Pathogenic expansions were most frequently detected in *HTT*, followed by *ATXN2* and *ATXN1*, together accounting for more than 60% of all pathogenic alleles. In contrast, intermediate-range expansions were predominantly observed in *ATXN2* and *ATXN1*, which collectively comprised over two-thirds of all intermediate alleles.

At the disease-category level, pathogenic expansions were most frequently identified in the ataxia (ATX) cohort, followed by ALS & FTD and myopathy (MYO) cases. Intermediate expansions were most prevalent in the ALS & FTD cohort, with lower frequencies observed in ATX and PD cases.

**Table 1.** Clinical features and genotyping results for individuals carrying pathogenic expansions at the targeted loci.

Familial ID	Individual ID	Clinical diagnosis	Gender	Date of Birth	Age on onset	Gene	NG S type	EH genotype	EH Repeat Range	Repeat Size (PCR-based)	Clinical presentation of probands
F01	F01-P1	ALS	M	1943	70	AR	WGS	45	45-54	45	Tongue atrophy, fasciculations, paraplegia.
F02	F02-P1	ataxia	M	1973	38	ATXN1	WES	12/51	12-12/50-55	12/66	Gait ataxia, dysarthria, sensory ataxia (Romberg +), deep sensory involvement, cognitive impairment, cerebral-cerebellar atrophy.
F03	F03-P1	ALS	F	1958	66	ATXN1	WES	33/44	33-33/44-47	33/45	Dysphagia, Dysarthria
F04	F04-P1	ataxia	F	1960	50	ATXN1	WES	30/42	30-30/42-42	30/40	Gait ataxia and gait imbalance.
F05	F05-P1	PD	M	1977	43	ATXN1	WES	29/40	29-29/40-40	29/40	Tremor, bradykinesia on the right side
F05	F05-P2		M	1942		ATXN1				29/40	
F06	F06-P1	ALS	F	1955	58	ATXN1	WGS	29/42	29-29/42-42	28/41	Unilateral weakness in extremities
F07	F07-P1	ataxia	F	1999	16	ATXN2	WES	22/39	22-22/38-62	23/46	Gait ataxia, dysarthria, dysmetria, polyneuropathy (sensory and axonal), deep sensory involvement, cerebellar-spinal atrophy.
F08	F08-P1	ataxia	F	1975	20	ATXN2	WES	22/33	22-22/32-45	22/35	Complex phenotype, weakness in legs, difficulty in climbing stairs, swallowing difficulty
F08	F08-P2		F			ATXN2				22/35	
F08	F08-P3		F	1970	22	ATXN2				23/35	
F09	F09-P1	ALS	M	1975	45	ATXN2	WES	22/33	22-22/33-49	22/34	Progressive disturbances in speech
F10	F10-P1	ALS	F	1958	63	ATXN2	WES	22/31	22-22/31-32	22/34	Difficulty in walking
F11	F11-P1	ALS	M	1959	58	ATXN2	WGS	22/33	22-22/33-33	22/34	Weakness in arms, cramps in the legs
F12	F12-P1	ataxia	M	1968	32	CACNA1A	WES	14/25	14-14/25-25	14/25	Progressive ataxia, ataxic gait, cerebellar dysarthria, cerebellar atrophy
F12	F12-P2		F	1989		CACNA1A				6/25	
F12	F12-P3		F	1986	34	CACNA1A				10/25	
F12	F12-P4		M	1991	29	CACNA1A				13/25	
F13	F13-P1	myopathy	M	1985	34	DMPK	WES	48/48	37-65/47-79	positive	Progressive muscle weakness
F13	F13-P2		M			DMPK				positive	
F13	F13-P3		M	1992		DMPK				positive	
F13	F13-P4		F			DMPK				positive	
F13	F13-P5		F			DMPK				positive	
F13	F13-P6		M			DMPK				positive	
F14	F14-P1	myopathy	F	2001	adolescent	DMPK	WES	32/53	32-32/49-66	positive	Myotonia
F14	F14-P2		M			DMPK					
F15	F15-P1	myopathy	F	1963	53	DMPK	WES	11/55	11-11/50-97	positive	Proximal weakness, imbalance, drooping of one eyelid
F16	F16-P1	ataxia	M	1971	40	HTT	WES	29/48	29-29/48-49	30/49	Gait ataxia, dysarthria, gait imbalance
F17	F17-P1	ataxia	F	1995	23	HTT	WES	23/46	23-23/46-52	24/55	Gait ataxia, upper motor neuron involvement, gait imbalance.
F18	F18-P1	ataxia	F	1969	49	HTT	WES	23/45	23-23/44-49	24/45	Gait ataxia, gait imbalance, upper motor neuron involvement, lower extremity involvement, chorea

F19	F19-P1	ataxia	M	2010	congenital	HTT	WES	19/40	19-19/40-44	20/41	Very low birth weight (1200 g). Gait imbalance began at age 2, with cerebellar and pyramidal signs on examination.
F19	F19-P2		M	1979		HTT	WES	19/40	19-19/40-40	20/41	
F19	F19-P3		M	2010	congenital	HTT				20/41	
F19	F19-P4		F	1996	congenital	HTT				19/40	
F20	F20-P1	ataxia	M	1989	5	HTT	WES	19/40	19-19/40-40	20/41	Cerebellar ataxia, pyramidal signs.
F20	F20-P2		M	1964		HTT	WES			20/40	NA
F21	F21-P1	PD	M	1969	50	HTT	WES	22/39	22-22/39-39	23/40	NA
F22	F22-P1	ALS	F	1943	69	HTT	WGS	19/41	19-19/41-41	20/42	Typical ALS
F23	F23-P1	ALS	F	1946	70	HTT	WGS	20/39	20-20/39-39	21/40	Unilateral weakness, tongue atrophy
F24	F24-P1	OPMD	M	1959	55	PABPN1	WES	6/10	6-6/10-10	6/10	Progressive ptosis, dysphagia and dysarthria over a 10 years
F24	F24-P2		F	1956	55	PABPN1	WES	6/10	6-6/10-10	6/10	
F24	F24-P3		M	1981		PABPN1				6/10	
F24	F24-P4		F			PABPN1				6/10	
F25	F25-P1	OPMD	F	1954	33	PABPN1	WES	6/9	6-6/9-9	6/9	Bilateral ptosis, progressive proximal muscle weakness, difficulty swallowing
F26	F26-C1	Control	M	1954		PABPN1	WGS	6/8	6-6/8-8	6/8	Neurologically healthy
F27	F27-P1	ataxia	M	1985	27	TBP	WES	37/54	37-37/52-59	37/57	Gait ataxia, dysarthria, gait imbalance, nystagmus, dysmetria, dysdiadochokinesis, epilepsy, cerebral-cerebellar atrophy, sleep disturbance.
F28	F28-P1	ataxia	F	2003	9	TBP	WES	36/47	36-36/47-47	36/47	Gait ataxia, dysarthria, mental retardation, sepsis hyperbilirubinemia

ALS: amyotrophic lateral sclerosis, EH: ExpansionHunter, OPMD: oculopharyngeal muscular dystrophy, PD: Parkinson's disease, WES: whole-exome Sequencing, WGS: whole-genome sequencing.

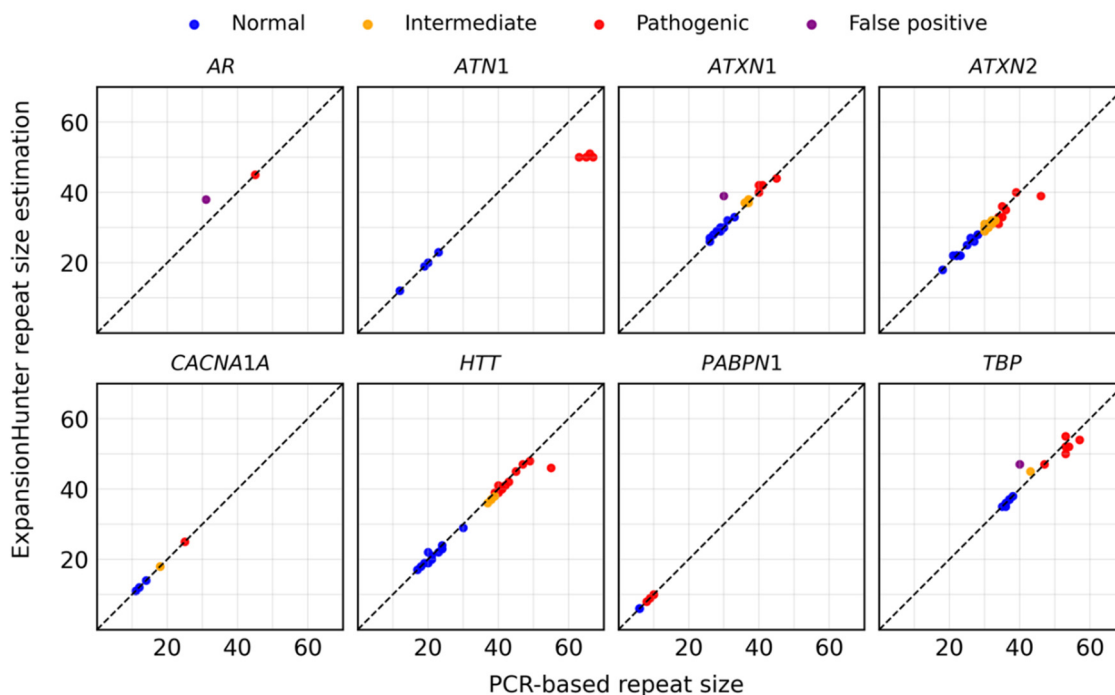
### 3.3. Comparison Between ExpansionHunter Repeat Estimation and PCR-Based Repeat Sizing

In addition to the eight intermediate alleles reclassified as pathogenic, we validated a substantial number of additional intermediate-range alleles, primarily in *ATXN2*, *ATXN1*, and *HTT* (Supplementary Tables S4 and S5). Including previously identified cases with conventional methods (Supplementary Table S2), repeat size estimates generated by EH and PCR-based assays were available for 113 individuals across eight loci: *AR*, *ATN1*, *ATXN1*, *ATXN2*, *CACNA1A*, *HTT*, *PABPN1*, and *TBP*.

Pathogenic expansions in *DMPK* were assessed by fragment analysis to determine expansion status (pathogenic vs. non-pathogenic); however, precise repeat sizing was not available. Therefore, *DMPK* was excluded from the quantitative agreement analysis.

Figure 3 illustrates the comparison between EH-derived repeat size estimates and PCR-based measurements. A per-gene summary of pathogenic, intermediate, normal, and false-positive calls is provided in Supplementary Table S6. Overall, EH demonstrated high agreement with PCR results, accurately estimating both normal (short) alleles and expanded alleles across loci. Only three false-positive calls were observed. Importantly, not all intermediate-range alleles were subjected to PCR validation. This is because intermediate expansions in the analyzed loci were shorter than the sequencing read-length and were therefore predominantly supported by spanning reads, for which EH demonstrated near-identical agreement with PCR measurements (typically within  $\pm 1-2$  repeat

units). Given this high accuracy in the size estimation, additional validation of all intermediate alleles was not considered necessary.



**Figure 3.** Agreement between ExpansionHunter repeat size estimates and PCR-based measurements. Comparison of repeat lengths estimated by EH and PCR-based validation in 113 individuals across eight loci: *AR*, *ATN1*, *ATXN1*, *ATXN2*, *CACNA1A*, *HTT*, *PABPN1*, and *TBP*. EH showed high concordance with PCR for both normal and expanded alleles, with only three false-positive calls.

Five of the analyzed loci are known to harbor repeat motif interruptions (Supplementary Table S2). Despite this motif complexity, EH maintained robust performance across loci.

### 3.4. Motif Interruptions at *ATXN1* Locus Analysis

Motif interruption analysis with REViewer tool was performed with particular focus on the *ATXN1* locus. While repeat interruptions may occur at other polyglutamine loci, interruptions at *ATXN1* alter the encoded amino acid sequence by introducing histidine residues within the polyglutamine tract.

The *ATXN1* locus was customized to genotype the (CTG)<sub>n</sub> motif. Among the four EH-identified pathogenic *ATXN1* expansions, only one case (F04-P1), a clinically diagnosed ataxia patient, lacked histidine interruptions within the expanded allele (Supplementary Figure S8). The other two patients with two histidine interruptions as in the reference genome had clinical prediagnosis other than ataxia (Supplementary Figures S9 and S10).

Notably, the remaining patient, referred with a clinical prediagnosis of ALS, carried a 33/44 genotype with two additional histidine interruptions in the expanded allele (four in total). Based on REViewer alignments, the repeat configurations (5' to 3') of the short and expanded alleles were (Supplementary Figure S11):

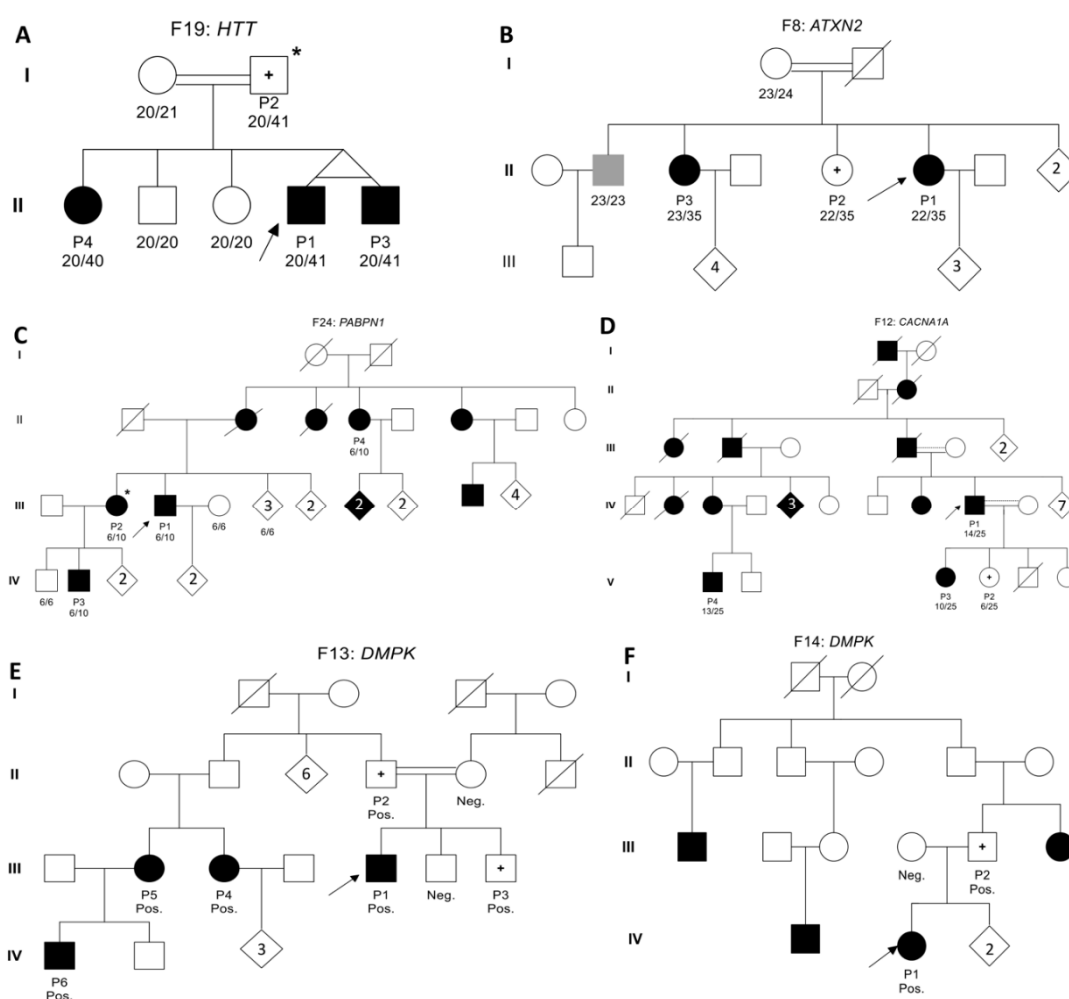
- short allele: Q<sub>1-16</sub>H<sub>17</sub>Q<sub>18</sub>H<sub>19</sub>Q<sub>20-33</sub>
- expanded allele: Q<sub>1-14</sub>H<sub>15</sub>Q<sub>16</sub>H<sub>17</sub>Q<sub>18-29</sub>H<sub>30</sub>Q<sub>31</sub>H<sub>31</sub>Q<sub>33-44</sub>

All three individuals harboring pathogenic *ATXN1* expansions, despite differing initial clinical prediagnoses, were ultimately reclassified as spinocerebellar ataxia type 1 (SCA1) based on molecular findings and clinical re-evaluation.

Histidine (ATG) interruptions were detected via REViewer in all 15 PCR-confirmed intermediate cases (Supplementary Table S4).

### 3.5. Pathogenic *HTT* Repeat Expansions in Non-Huntington disease Diagnoses

Eight *HTT*-positive cases with repeat lengths exceeding  $\geq 40$  (CAG)<sub>n</sub> were referred with initial diagnoses other than HD. In a large family with five offspring, referred to us with a prediagnosis of ataxia, three siblings with congenital disease inherited the *HTT* expansion from a presymptomatic father. In contrast, the two unaffected siblings were shown to have normal repeat lengths (Figure 4A). Another patient with a juvenile-onset movement disorder at age five carried a 20/41 *HTT* expansion, also inherited from a presymptomatic father. One of the three remaining *HTT*-positive patients was initially diagnosed with PD and two with ALS. No additional segregating protein-altering mutations in relevant genes were identified in these families.



**Figure 4.** Pedigrees of families with segregating repeat expansions. **A.** Family F19 with a pathogenic *HTT* expansion. **B.** Family F8 with a pathogenic *ATXN2* expansion. **C.** Family F24 with a pathogenic *PABPN1* expansion. **D.** Family F12 with a pathogenic *CACNA1A* expansion. **E–F.** Family F13 with a pathogenic *DMPK* expansion. Individuals marked with \* have exome sequencing data in addition to the proband (indicated by an arrow). Pos, positive; Neg, negative.

Additionally, six individuals with intermediate *HTT* alleles (36–39 repeats) were identified, who were clinically associated with ALS (n=3), ataxia (n=1), FTD (n=1), and PD (n=1) (Supplementary Table S4).

### 3.6. *ATXN2* Repeat Expansions

Given the established role of *ATXN2* intermediate expansions in ALS risk [6,11,14], all ALS patients and the majority of individuals with other clinical prediagnoses carrying  $\geq 29$  (CAG)<sub>n</sub> repeats were validated via fragment analysis (Supplementary Tables S4 and S5). Based on the distribution of *ATXN2* repeat sizes across disease cohorts, intermediate-range alleles (29–33 repeats) were predominantly observed in ALS cohort (Supplementary Figure S12). Of the 53 intermediate alleles identified in the cohort, 34 (64.15%) were detected in ALS patients. In addition, three ALS individuals carried expansions at the borderline pathogenic threshold 34 repeats. In family F08, the 35-repeat allele segregated with ataxia in the proband and one sibling, while clinical data were unavailable for the other carrier (Figure 4B). The largest expansion (46 repeats) was detected in a patient with juvenile-onset ataxia, consistent with the established association between longer *ATXN2* expansions and spinocerebellar ataxia type 2 (SCA2). No intermediate or pathogenic expansions were observed in the healthy controls and myopathy cohorts.

### 3.7. Pathogenic Expansions in *PABPN1* and Other Loci

To the best of our knowledge, this is the first report of genetically confirmed *PABPN1*-associated oculopharyngeal muscular dystrophy (OPMD) cases in Türkiye. WES-based in silico STR genotyping revealed a heterozygous (GCG)<sub>10</sub> expansion in two siblings with clinical OPMD; PCR confirmed segregation in four affected family members (Figure 4C). A heterozygous (GCG)<sub>9</sub> was identified in an unrelated case with OPMD. Additionally, one neurologically healthy 61-year-old control individual carried a (GCG)<sub>8</sub> allele, falling on the pathogenic borderline.

Pathogenic repeat expansions were identified in *TBP*, *ATN1*, *CACNA1A*, and *DMPK* with genotype–phenotype correlations consistent with established disease associations. The probands with *CACNA1A* and *DMPK* (F13) belonged to large sibships with multiple affected members. Fragment analysis confirmed that the expanded alleles co-segregated with disease (Figure 4D,E). Figure 4F depicts the pedigree of the *DMPK*-positive family, in which multiple individuals presented with similar symptoms; however, DNA samples were available only from the parents.

A single pathogenic expansion in *AR* was detected in a patient initially prediagnosed with motor neuron disease. Following molecular confirmation and subsequent clinical re-evaluation, the diagnosis was revised to spinal and bulbar muscular atrophy.

### 3.8. Diagnostic Yield

Using in silico genotyping with EH on short-read NGS data, we uncovered the genetic basis in 27 previously unsolved cases and 18 affected relatives. The molecular diagnoses were largely consistent with the clinical prediagnoses. This corresponds to an additional diagnostic yield of ~0.95% (30/3,150 patients).

## 4. Discussion

Transformative bioinformatic advances in NGS-based STR genotyping have enabled the systematic screening of large disease cohorts for the detection of clinically relevant repeat expansions, particularly in neurological disorders [3,8,11–16]. Such studies help to resolve genetic heterogeneity and improve differential diagnosis. NGS-based STR genotyping has the potential to complement or replace site-specific conventional methods for many loci.

The present cohort recruited between 2005 and 2025 at NDAL, a nationwide reference center for the genetics of neurodegenerative diseases in Türkiye. All samples had previously undergone comprehensive analysis for protein-altering variants, short indels, and copy number variants.

Targeted repeat expansion findings were obtained through conventional methods in specific disease groups (e.g., (CAG)<sub>n</sub> triplet-repeat for hereditary ataxias), for which WES/WGS were considered uninformative [10,23,25,26]. However, most patients had not been systematically screened for repeat expansions across the 15 STR loci investigated in this study.

By integrating systematic *in silico* STR genotyping into the existing genomic workflow, we achieved an additional diagnostic yield of 0.95%, thereby unraveling 27 previously unsolved cases. In addition, we identified 160 intermediate-range (premutation) alleles across multiple loci, of which 77 were validated by orthogonal methods. Given the elevated mutation rates of STRs and their tendency for further expansion across generations (genetic anticipation), intermediate alleles are clinically important, particularly in the context of incomplete penetrance observed at certain loci, such as *HTT* [1]. Collectively, these findings underscore the incremental diagnostic and prognostic value of incorporating targeted repeat expansion screening into routine genomic pipelines, particularly in clinically heterogeneous neurogenetic cohorts.

In this study, we identified a pathogenic repeat expansion in the *AR* locus in a male patient initially diagnosed with motor neuron disease. Following molecular confirmation and clinical re-evaluation, the diagnosis was revised to spinal and bulbar muscular atrophy (SBMA), consistent with similar observations reported in Norwegian ALS cohorts [14].

#### 4.1. Interruptions in *ATXN1* Modulate *SCA1* Penetrance and Clinical Manifestation

The presence of (ATG)<sub>n</sub> interruptions in the *ATXN1* repeat locus plays an important role in modulating the *SCA1* phenotype [27,28]. Among the four pathogenic *ATXN1* carriers in our cohort, only the patient without interruptions in the expanded allele presented with ataxia symptoms. The other three patients, whose expanded alleles contained (ATG)<sub>n</sub> interruptions, were referred with clinical diagnoses of ALS or PD. Notably, Frontali et al. described a healthy 66-year-old father and his 24-year-old son who both carried a 45-repeat *ATXN1* allele containing four histidine interruptions [29]. In contrast, our index patient developed initial symptoms of dysphagia and dysarthria at the age of 66, despite harboring an expanded allele with a comparable total of four histidine interruptions within the polyglutamine tract. This observation is consistent with the proposed modulatory role of repeat interruptions in influencing repeat instability and may contribute to variability in disease penetrance, age at onset, or clinical presentation.

#### 4.2. Infrequent Atypical Clinical Profiles of *HTT* Expansions

Atypical *HTT*-related presentations have been reported previously. Squitieri et al. found Parkinsonism, dystonia, or ataxia at onset in 15 of 205 HD patients [9]. In line with, Dong et al. identified seven probands initially diagnosed with spinocerebellar ataxia, of whom only three later developed chorea [30]. Méreaux et al. detected only two intermediate *HTT* alleles in 498 exomes from patients with spinocerebellar ataxia, suggesting incidental findings [13].

In accordance with previous studies, we identified five *HTT*-positive patients initially diagnosed with ataxia (40–55 repeats); notably, two inherited the expansion from an asymptomatic father and presented with congenital ataxia despite repeat sizes near the minimum pathogenic threshold. We also detected four intermediate and two pathogenic *HTT* expansions among ALS patients, replicating prior observations and underscoring the rarity of *HTT* expansions in the genetic architecture of ALS and FTD [8,11,14]. Collectively, these findings highlight the clinical heterogeneity of *HTT* expansions and their rare but important contribution to atypical movement and neurodegenerative disorder phenotypes beyond classical HD.

#### 4.3. First Genetically Confirmed OPMD Cases from Türkiye

Bilgen et al. reported a Turkish male patient with a family history of OPMD without genetic confirmation [31]. This early study was followed by the identification of a biallelic (GCN)<sub>13</sub> expansion in the *PABPN1* gene in a female patient of Turkish origin within a large OPMD cohort using

conventional genotyping [32]. We identified for the first time two unrelated OPMD cases carrying pathogenic *PABPN1* repeat expansions from Türkiye. In addition, we detected (GCG)<sub>8</sub> expansion, the pathogenicity borderline for *PABPN1*, in a neurologically healthy 61-year-old control individual of Turkish-Persian origin, born to healthy parents. The smallest pathogenic *PABPN1* expansion, (GCG)<sub>8</sub>, is generally associated with a milder disease course, with symptoms, typically limited to ptosis and mild dysphagia, appearing later in life, often in the seventh decade [33]. This reduced penetrance leading to attenuated phenotype may explain the absence of OPMD symptoms in our control individual.

#### 4.4. Recently Identified Triplet Expansions in *THAP11* and *ZFH3* in Ataxia Phenotype

Repeat expansions in *THAP11* (SCA51) and *ZFH3* (SCA4) were recently described in Chinese and Swedish families, respectively. However, follow-up studies suggest that they are extremely rare causes of ataxia [34–37]. Consistent with this, we found no pathogenic alleles at either locus in our large Turkish cohort, including more than 500 ataxia cases. These results, along with prior reports, suggest that *THAP11* and *ZFH3* expansions are likely to represent geographically restricted or rare causes of ataxia. The ability to incorporate recently identified loci into EH-based analysis pipelines highlights the flexibility of NGS-based STR genotyping, enabling rapid retrospective screening of large cohorts with high accuracy while minimizing the need for additional laboratory testing.

#### 4.5. Technical Considerations

Sequencing read length and STR locus selection are critical determinants of genotyping accuracy using EH. Repeat lengths shorter than the sequencing read length are typically estimated with high accuracy [3,10,15]. Minor discrepancies of  $\pm 1$ -2 repeat units between EH spanning-read calls and PCR measurements are generally considered acceptable, reflecting inherent sizing limitations of conventional methods (Figure 3) [13,38]. In our cohort, alleles supported by spanning reads showed strong concordance with PCR results, usually differing by only  $\pm 1$ -2 repeat units. In contrast, when repeat lengths exceeded the read length, size estimation relied primarily on flanking or in-repeat reads. Under these conditions, EH tended to underestimate expansion size, as observed in pathogenic *ATN1* alleles. Correspondingly, genotype confidence intervals were broader, indicating the possibility of larger repeat sizes. Notably, 93.27% of samples were sequenced with 150 bp reads, which were sufficient to accurately genotype all target loci up to the minimum pathogenic threshold. Only three false-positive calls were observed, two of which occurred in samples with 100 bp reads. Studies evaluating larger STR panels have reported higher false-positive rates [39,40]. Collectively, these findings indicate that the selected STR loci can be reliably genotyped in exome and genome datasets, particularly when adequate read length is available.

Locus coverage is a critical determinant of STR genotyping performance in exome sequencing due to variability introduced by different exome capture platforms. As shown in Supplementary Figure S2, coverage across target STR loci differed markedly between capture kits. Similar challenges have been reported previously, with key loci such as *ATXN2* often insufficiently captured (e.g., 8.8% genotyping with TruSeq; exclusion with SureSelect V5/V6) [15,16]. These findings highlight that STR genotyping efficiency using ExpansionHunter in exome data is highly dependent on local read depth. To address this limitation, we lowered the default locus coverage threshold from 10X to 3X. This adjustment substantially improved genotyping performance for poorly captured loci, increasing the genotyping rate for *ATXN2* to 89.6%, representing an approximately 10-fold improvement compared to previous reports. Strikingly, all pathogenic and majority of intermediate *ATXN2* expansions identified in our exome dataset had coverage values below 10X, indicating that the default threshold would have failed to detect them. Overall, genotyping rates across all target loci exceeded 70%, with some loci achieving complete (100%) genotyping in exome samples (Supplementary Figure S13). Importantly, the combination of a relaxed coverage threshold and visual inspection using REViewer enabled recovery of true expansions in poorly captured exons without introducing additional false positives.

## 5. Conclusion

NGS-based in silico STR genotyping with EH enables systematic detection of disease-associated expansions, providing new molecular diagnoses, and revealing atypical clinical presentations. Our findings highlight the genetic heterogeneity of neurodegenerative diseases, especially movement disorders, and the importance of careful visual evaluation of the repeat structure. This study enhances diagnostic accuracy in complex neurodegenerative disorders and expands future neurogenetic research.

**Supplementary Materials:** The following supporting information can be downloaded at the website of this paper posted on Preprints.org.

**Author Contributions:** Conceptualization, Z.K. and A.N.B.; Methodology, Z.K. and R.J.P.; Software, Z.K. and R.J.P.; Validation, M.K.K., I.A. and I.Ş.; Formal Analysis, Z.K.; Investigation, Z.K., M.K.K., I.A. and I.Ş.; Resources, Ş.T., A.Ş., E.Z.D.B., T.G.D., V.Ç., E.B., C.T., and N.S.; Data Curation, Z.K. and R.J.P.; Writing—Original Draft Preparation, Z.K. and A.N.B.; Writing—Review & Editing, Ş.T., A.Ş., E.Z.D.B., T.G.D., V.Ç., E.B., C.T., N.S. and F.A.; Visualization, Z.K.; Supervision, A.N.B.; Project Administration, A.N.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by funds from Suna and İnan Kırac Foundation and Koç University.

**Data Availability:** The data that support the findings of this study are available from the corresponding author upon request.

**Acknowledgments:** NDAL gratefully acknowledges the use of the services and facilities of Koç University Research Center for Translational Medicine (KUTTAM). We would also like to extend our sincere gratitude to Suna and İnan Kırac Foundation for its generous support of the study and both the Foundation and Koç University for the inspiring research environment created. We thankfully acknowledge our clinicians. We are grateful to Prof. Stephan Züchner, Prof. Henry Houlden, Prof. Jan H. Veldink and Project MinE Consortium for their invaluable collaboration. This research was supported, in part, by the Intramural Research Program of the National Institutes of Health (NIH). The contributions of the NIH author(s) are considered Works of the United States Government. The findings and conclusions presented in this paper are those of the author(s) and do not necessarily reflect the views of the NIH or the U.S. Department of Health and Human Services.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Depienne, C.; Mandel, J.-L. 30 Years of Repeat Expansion Disorders: What Have We Learned and What Are the Remaining Challenges? *Am. J. Hum. Genet.* **2021**, *108* (5), 764–785. <https://doi.org/10.1016/j.ajhg.2021.03.011>.
2. Van Deynze, K.; Mumm, C.; Maltby, C. J.; Switzenberg, J. A.; Todd, P. K.; Boyle, A. P. Enhanced Detection and Genotyping of Disease-Associated Tandem Repeats Using HMMSTR and Targeted Long-Read Sequencing. *Nucleic Acids Res.* **2025**, *53* (2). <https://doi.org/10.1093/nar/gkae1202>.
3. Ibañez, K.; Polke, J.; Hagelstrom, R. T.; Dolzhenko, E.; Pasko, D.; Thomas, E. R. A.; Daugherty, L. C.; Kasperaviciute, D.; Smith, K. R.; WGS for Neurological Diseases Group; Deans, Z. C.; Hill, S.; Fowler, T.; Scott, R. H.; Hardy, J.; Chinnery, P. F.; Houlden, H.; Rendon, A.; Caulfield, M. J.; Eberle, M. A.; Taft, R. J.; Tucci, A.; Genomics England Research Consortium. Whole Genome Sequencing for the Diagnosis of Neurological Repeat Expansion Disorders in the UK: A Retrospective Diagnostic Accuracy and Prospective Clinical Validation Study. *Lancet Neurol.* **2022**, *21* (3), 234–245. [https://doi.org/10.1016/S1474-4422\(21\)00462-2](https://doi.org/10.1016/S1474-4422(21)00462-2).
4. Paulson, H. Repeat Expansion Diseases. *Handb. Clin. Neurol.* **2018**, *147*, 105–123. <https://doi.org/10.1016/B978-0-444-63233-3.00009-9>.

5. Shan, D. E.; Soong, B. W.; Sun, C. M.; Lee, S. J.; Liao, K. K.; Liu, R. S. Spinocerebellar Ataxia Type 2 Presenting as Familial Levodopa-Responsive Parkinsonism. *Ann. Neurol.* **2001**, *50* (6), 812–815. <https://doi.org/10.1002/ana.10055>.
6. Elden, A. C.; Kim, H.-J.; Hart, M. P.; Chen-Plotkin, A. S.; Johnson, B. S.; Fang, X.; Arakola, M.; Geser, F.; Greene, R.; Lu, M. M.; Padmanabhan, A.; Clay-Falcone, D.; McCluskey, L.; Elman, L.; Juhn, D.; Gruber, P. J.; Rüb, U.; Auburger, G.; Trojanowski, J. Q.; Lee, V. M.-Y.; Van Deerlin, V. M.; Bonini, N. M.; Gitler, A. D. Ataxin-2 Intermediate-Length Polyglutamine Expansions Are Associated with Increased Risk for ALS. *Nature* **2010**, *466* (7310), 1069–1075. <https://doi.org/10.1038/nature09320>.
7. Gwinn-Hardy, K.; Singleton, A.; O'Suilleabhain, P.; Boss, M.; Nicholl, D.; Adam, A.; Hussey, J.; Critchley, P.; Hardy, J.; Farrer, M. Spinocerebellar Ataxia Type 3 Phenotypically Resembling Parkinson Disease in a Black Family. *Arch. Neurol.* **2001**, *58* (2), 296–299. <https://doi.org/10.1001/archneur.58.2.296>.
8. Dewan, R.; Chia, R.; Ding, J.; Hickman, R. A.; Stein, T. D.; Abramzon, Y.; Ahmed, S.; Sabir, M. S.; Portley, M. K.; Tucci, A.; Ibáñez, K.; Shankaracharya, F. N. U.; Keagle, P.; Rossi, G.; Caroppo, P.; Tagliavini, F.; Waldo, M. L.; Johansson, P. M.; Nilsson, C. F.; American Genome Center (TAGC); FALS Sequencing Consortium; Genomics England Research Consortium; International ALS/FTD Genomics Consortium (iAFGC); International FTD Genetics Consortium (IFGC); International LBD Genomics Consortium (iLBDGC); NYGC ALS Consortium; PROSPECT Consortium; Rowe, J. B.; Benussi, L.; Binetti, G.; Ghidoni, R.; Jabbari, E.; Viollet, C.; Glass, J. D.; Singleton, A. B.; Silani, V.; Ross, O. A.; Ryten, M.; Torkamani, A.; Tanaka, T.; Ferrucci, L.; Resnick, S. M.; Pickering-Brown, S.; Brady, C. B.; Kowal, N.; Hardy, J. A.; Van Deerlin, V.; Vonsattel, J. P.; Harms, M. B.; Morris, H. R.; Ferrari, R.; Landers, J. E.; Chiò, A.; Gibbs, J. R.; Dalgard, C. L.; Scholz, S. W.; Traynor, B. J. Pathogenic Huntingtin Repeat Expansions in Patients with Frontotemporal Dementia and Amyotrophic Lateral Sclerosis. *Neuron* **2021**, *109* (3), 448–460.e4. <https://doi.org/10.1016/j.neuron.2020.11.005>.
9. Squitieri, F.; Berardelli, A.; Nargi, E.; Castellotti, B.; Mariotti, C.; Cannella, M.; Lavitrano, M. L.; de Grazia, U.; Gellera, C.; Ruggieri, S. Atypical Movement Disorders in the Early Stages of Huntington's Disease: Clinical and Genetic Analysis. *Clin. Genet.* **2000**, *58* (1), 50–56. <https://doi.org/10.1034/j.1399-0004.2000.580108.x>.
10. Dolzhenko, E.; van Vugt, J. J. F. A.; Shaw, R. J.; Bekritsky, M. A.; van Blitterswijk, M.; Narzisi, G.; Ajay, S. S.; Rajan, V.; Lajoie, B. R.; Johnson, N. H.; Kingsbury, Z.; Humphray, S. J.; Schellevis, R. D.; Brands, W. J.; Baker, M.; Rademakers, R.; Kooyman, M.; Tazelaar, G. H. P.; van Es, M. A.; McLaughlin, R.; Sproviero, W.; Shatunov, A.; Jones, A.; Al Khleifat, A.; Pittman, A.; Morgan, S.; Hardiman, O.; Al-Chalabi, A.; Shaw, C.; Smith, B.; Neo, E. J.; Morrison, K.; Shaw, P. J.; Reeves, C.; Winterkorn, L.; Wexler, N. S.; US-Venezuela Collaborative Research Group; Housman, D. E.; Ng, C. W.; Li, A. L.; Taft, R. J.; van den Berg, L. H.; Bentley, D. R.; Veldink, J. H.; Eberle, M. A. Detection of Long Repeat Expansions from PCR-Free Whole-Genome Sequence Data. *Genome Res.* **2017**, *27* (11), 1895–1903. <https://doi.org/10.1101/gr.225672.117>.
11. Henden, L.; Fearnley, L. G.; Grima, N.; McCann, E. P.; Dobson-Stone, C.; Fitzpatrick, L.; Friend, K.; Hobson, L.; Chan Moi Fat, S.; Rowe, D. B.; D'Silva, S.; Kwok, J. B.; Halliday, G. M.; Kiernan, M. C.; Mazumder, S.; Timmins, H. C.; Zoing, M.; Pamphlett, R.; Adams, L.; Bahlo, M.; Blair, I. P.; Williams, K. L. Short Tandem Repeat Expansions in Sporadic Amyotrophic Lateral Sclerosis and Frontotemporal Dementia. *Sci. Adv.* **2023**, *9* (18), eade2044. <https://doi.org/10.1126/sciadv.ade2044>.
12. Ibáñez, K.; Jadhav, B.; Zanovello, M.; Gagliardi, D.; Clarkson, C.; Facchini, S.; Garg, P.; Martin-Trujillo, A.; Gies, S. J.; Galassi Deforie, V.; Dalmia, A.; Hensman Moss, D. J.; Vandrovicova, J.; Rocca, C.; Moutsianas, L.; Marini-Bettolo, C.; Walker, H.; Turner, C.; Shoai, M.; Long, J. D.; Fratta, P.; Langbehn, D. R.; Tabrizi, S. J.; Caulfield, M. J.; Cortese, A.; Escott-Price, V.; Hardy, J.; Houlden, H.; Sharp, A. J.; Tucci, A. Increased Frequency of Repeat Expansion Mutations across Different Populations. *Nat. Med.* **2024**, *30* (11), 3357–3368. <https://doi.org/10.1038/s41591-024-03190-5>.
13. Méreaux, J.-L.; Davoine, C.-S.; Coutelier, M.; Guillot-Noël, L.; Castrioto, A.; Charles, P.; Coarelli, G.; Ewencyk, C.; Klebe, S.; Heinzmann, A.; Méneret, A.; Fauret-Amsellem, A.-L.; de Sainte Agathe, J.-M.; Brice, A.; Durr, A. Fast and Reliable Detection of Repeat Expansions in Spinocerebellar Ataxia Using Exomes. *J. Med. Genet.* **2023**, *60* (7), 717–721. <https://doi.org/10.1136/jmg-2022-108924>.

14. Novy, C.; Busk, Ø. L.; Tysnes, O.-B.; Landa, S. S.; Aanjesen, T. N.; Alstadhaug, K. B.; Bjerknes, T. L.; Bjørnå, I. K.; Bråthen, G.; Dahl, E.; Demic, N.; Fahlström, M.; Flemmen, H. Ø.; Hallerstig, E.; HogenEsch, I.; Kampman, M. T.; Kleveland, G.; Kvernmo, H. B.; Ljøstad, U.; Maniaol, A.; Morsund, A. H.; Nakken, O.; Olsen, C. G.; Schlüter, K.; Utvik, M.-S.; Yaseen, R.; Holla, Ø. L.; Holmøy, T.; Høyer, H. Repeat Expansions in AR, ATXN1, ATXN2 and HTT in Norwegian Patients Diagnosed with Amyotrophic Lateral Sclerosis. *Brain Commun.* **2024**, *6* (2), fcae087. <https://doi.org/10.1093/braincomms/fcae087>.
15. Rocca, C.; Murphy, D.; Clarkson, C.; Zanollo, M.; Gagliardi, D.; Genomics, Q. S.; Kaiyrzhanov, R.; Alvi, J.; Maroofian, R.; Efthymiou, S.; Sultan, T.; Vandrovцова, J.; Polke, J.; Labrum, R.; Houlden, H.; Tucci, A. Analysis of Short Tandem Repeat Expansions in a Cohort of 12,496 Exomes from Patients with Neurological Diseases Reveals Variable Genotyping Rate Dependent on Exome Capture Kits. *Genes* **2025**, *16* (2), 169. <https://doi.org/10.3390/genes16020169>.
16. Yoon, J. G.; Lee, S.; Cho, J.; Kim, N.; Kim, S.; Kim, M. J.; Kim, S. Y.; Moon, J.; Chae, J.-H. Diagnostic Uplift through the Implementation of Short Tandem Repeat Analysis Using Exome Sequencing. *Eur. J. Hum. Genet.* **2024**, *32* (5), 584–587. <https://doi.org/10.1038/s41431-024-01542-w>.
17. van Rheenen, W.; van der Spek, R. A. A.; Bakker, M. K.; van Vugt, J. J. F. A.; Hop, P. J.; Zwamborn, R. A. J.; de Klein, N.; Westra, H.-J.; Bakker, O. B.; Deelen, P.; Shireby, G.; Hannon, E.; Moisse, M.; Baird, D.; Restuadi, R.; Dolzhenko, E.; Dekker, A. M.; Gawor, K.; Westeneng, H.-J.; Tazelaar, G. H. P.; van Eijk, K. R.; Kooyman, M.; Byrne, R. P.; Doherty, M.; Heverin, M.; Al Khleifat, A.; Iacoangeli, A.; Shatunov, A.; Ticozzi, N.; Cooper-Knock, J.; Smith, B. N.; Gromicho, M.; Chandran, S.; Pal, S.; Morrison, K. E.; Shaw, P. J.; Hardy, J.; Orrell, R. W.; Sendtner, M.; Meyer, T.; Başak, N.; van der Kooij, A. J.; Ratti, A.; Fogh, I.; Gellera, C.; Lauria, G.; Corti, S.; Cereda, C.; Sproviero, D.; D'Alfonso, S.; Sorarù, G.; Siciliano, G.; Filosto, M.; Padovani, A.; Chiò, A.; Calvo, A.; Moglia, C.; Brunetti, M.; Canosa, A.; Grassano, M.; Beghi, E.; Pupillo, E.; Logroscino, G.; Nefussy, B.; Osmanovic, A.; Nordin, A.; Lerner, Y.; Zabari, M.; Gotkine, M.; Baloh, R. H.; Bell, S.; Vourc'h, P.; Corcia, P.; Couratier, P.; Millecamps, S.; Meininger, V.; Salachas, F.; Mora Pardina, J. S.; Assialioui, A.; Rojas-García, R.; Dion, P. A.; Ross, J. P.; Ludolph, A. C.; Weishaupt, J. H.; Brenner, D.; Freischmidt, A.; Bensimon, G.; Brice, A.; Durr, A.; Payan, C. A. M.; Saker-Delye, S.; Wood, N. W.; Topp, S.; Rademakers, R.; Tittmann, L.; Lieb, W.; Franke, A.; Ripke, S.; Braun, A.; Kraft, J.; Whiteman, D. C.; Olsen, C. M.; Uitterlinden, A. G.; Hofman, A.; Rietschel, M.; Cichon, S.; Nöthen, M. M.; Amouyel, P.; Traynor, B. J.; Singleton, A. B.; Mitne Neto, M.; Cauchi, R. J.; Ophoff, R. A.; Wiedau-Pazos, M.; Lomen-Hoerth, C.; van Deerlin, V. M.; Grosskreutz, J.; Roediger, A.; Gaur, N.; Jörk, A.; Barthel, T.; Theele, E.; Ilse, B.; Stubendorff, B.; Witte, O. W.; Steinbach, R.; Hübner, C. A.; Graff, C.; Brylev, L.; Fominykh, V.; Demeshonok, V.; Ataulina, A.; Rogelj, B.; Koritnik, B.; Zidar, J.; Ravnik-Glavač, M.; Glavač, D.; Stević, Z.; Drory, V.; Povedano, M.; Blair, I. P.; Kiernan, M. C.; Benyamin, B.; Henderson, R. D.; Furlong, S.; Mathers, S.; McCombe, P. A.; Needham, M.; Ngo, S. T.; Nicholson, G. A.; Pamphlett, R.; Rowe, D. B.; Steyn, F. J.; Williams, K. L.; Mather, K. A.; Sachdev, P. S.; Henders, A. K.; Wallace, L.; de Carvalho, M.; Pinto, S.; Petri, S.; Weber, M.; Rouleau, G. A.; Silani, V.; Curtis, C. J.; Breen, G.; Glass, J. D.; Brown, R. H.; Landers, J. E.; Shaw, C. E.; Andersen, P. M.; Groen, E. J. N.; van Es, M. A.; Pasterkamp, R. J.; Fan, D.; Garton, F. C.; McRae, A. F.; Davey Smith, G.; Gaunt, T. R.; Eberle, M. A.; Mill, J.; McLaughlin, R. L.; Hardiman, O.; Kenna, K. P.; Wray, N. R.; Tsai, E.; Runz, H.; Franke, L.; Al-Chalabi, A.; Van Damme, P.; van den Berg, L. H.; Veldink, J. H. Common and Rare Variant Association Analyses in Amyotrophic Lateral Sclerosis Identify 15 Risk Loci with Distinct Genetic Architectures and Neuron-Specific Biology. *Nat. Genet.* **2021**, *53* (12), 1636–1648. <https://doi.org/10.1038/s41588-021-00973-1>.
18. *Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data.* <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed 2026-02-22).
19. Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. Fastp: An Ultra-Fast All-in-One FASTQ Preprocessor. *Bioinformatics* **2018**, *34* (17), i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>.
20. Danecek, P.; Bonfield, J. K.; Liddle, J.; Marshall, J.; Ohan, V.; Pollard, M. O.; Whitwham, A.; Keane, T.; McCarthy, S. A.; Davies, R. M.; Li, H. Twelve Years of SAMtools and BCFtools. *GigaScience* **2021**, *10* (2), giab008. <https://doi.org/10.1093/gigascience/giab008>.
21. Li, H. Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM. arXiv May 26, 2013. <https://doi.org/10.48550/arXiv.1303.3997>.

22. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytsky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; DePristo, M. A. The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data. *Genome Res.* **2010**, *20* (9), 1297–1303. <https://doi.org/10.1101/gr.107524.110>.
23. Tunca, C.; Şeker, T.; Akçimen, F.; Coşkun, C.; Bayraktar, E.; Palvadeau, R.; Zor, S.; Koçoğlu, C.; Kartal, E.; Şen, N. E.; Hamzeiy, H.; Özoğuz Erimiş, A.; Norman, U.; Karakahya, O.; Olgun, G.; Akgün, T.; Durmuş, H.; Şahin, E.; Çakar, A.; Başar Gürsoy, E.; Babacan Yıldız, G.; İşak, B.; Uluç, K.; Hanağası, H.; Bilgiç, B.; Turgut, N.; Aysal, F.; Ertas, M.; Boz, C.; Kotan, D.; İdrisoğlu, H.; Soysal, A.; Uzun Adatepe, N.; Akalın, M. A.; Koç, F.; Tan, E.; Oflazer, P.; Deymeer, F.; Taştan, Ö.; Çiçek, A. E.; Kavak, E.; Parman, Y.; Başak, A. N. Revisiting the Complex Architecture of ALS in Turkey: Expanding Genotypes, Shared Phenotypes, Molecular Networks, and a Public Variant Database. *Hum. Mutat.* **2020**, *41* (8), e7–e45. <https://doi.org/10.1002/humu.24055>.
24. Dolzhenko, E.; Weisburd, B.; Ibañez, K.; Rajan-Babu, I.-S.; Anyansi, C.; Bennett, M. F.; Billingsley, K.; Carroll, A.; Clamons, S.; Danzi, M. C.; Deshpande, V.; Ding, J.; Fazal, S.; Halman, A.; Jadhav, B.; Qiu, Y.; Richmond, P. A.; Saunders, C. T.; Scheffler, K.; van Vugt, J. J. F. A.; Zwamborn, R. R. A. J.; Genomics England Research Consortium; Chong, S. S.; Friedman, J. M.; Tucci, A.; Rehm, H. L.; Eberle, M. A. REViewer: Haplotype-Resolved Visualization of Read Alignments in and around Tandem Repeats. *Genome Med.* **2022**, *14* (1), 84. <https://doi.org/10.1186/s13073-022-01085-z>.
25. Tezen, D.; Khojakulov, Z.; Gündüz, A.; Deymeer, F.; Demirbilek, V.; Başak, A. N. Clinical, Electrophysiological, and Genetic Analysis of a Family with Two Rare Neuromuscular Disorders: Congenital Myasthenic Syndrome and Hereditary Polyneuropathy. *Neurol. Sci. Off. J. Ital. Neurol. Soc. Ital. Soc. Clin. Neurophysiol.* **2024**, *45* (12), 5967–5971. <https://doi.org/10.1007/s10072-024-07771-7>.
26. Vural, A.; Şimşir, G.; Tekgül, Ş.; Koçoğlu, C.; Akçimen, F.; Kartal, E.; Şen, N. E.; Lahut, S.; Ömür, Ö.; Saner, N.; Gül, T.; Bayraktar, E.; Palvadeau, R.; Tunca, C.; Pirkevi Çetinkaya, C.; Gündoğdu Eken, A.; Şahbaz, I.; Kovancılar Koç, M.; Öztıp Çakmak, Ö.; Hanağası, H.; Bilgiç, B.; Eraksoy, M.; Gündüz, A.; Apaydın, H.; Kızıltan, G.; Özekmekçi, S.; Siva, A.; Altıntaş, A.; Kaya Güleç, Z. E.; Parman, Y.; Oflazer, P.; Deymeer, F.; Durmuş, H.; Şahin, E.; Çakar, A.; Tüfekçioğlu, Z.; Tektürk, P.; Çorbalı, M. O.; Tireli, H.; Akdal, G.; Yiş, U.; Hız, S.; Şengün, İ.; Bora, E.; Serdaroğlu, G.; Erer Özbek, S.; Ağan, K.; İnce Günal, D.; Us, Ö.; Kurt, S. G.; Aksoy, D.; Bora Tokcaer, A.; Elmas, M.; Gültekin, M.; Kumandaş, S.; Acer, H.; Kaya Özçora, G. D.; Yayla, V.; Soysal, A.; Genç, G.; Güllüoğlu, H.; Kotan, D.; Özözen Ayas, Z.; Şahin, H. A.; Tan, E.; Topçu, M.; Topçuoğlu, E. S.; Akbostancı, C.; Koç, F.; Ertan, S.; Elibol, B.; Başak, A. N. The Complex Genetic Landscape of Hereditary Ataxias in Turkey and Implications in Clinical Practice. *Mov. Disord. Off. J. Mov. Disord. Soc.* **2021**, *36* (7), 1676–1688. <https://doi.org/10.1002/mds.28518>.
27. Matsuyama, Z.; Izumi, Y.; Kameyama, M.; Kawakami, H.; Nakamura, S. The Effect of CAT Trinucleotide Interruptions on the Age at Onset of Spinocerebellar Ataxia Type 1 (SCA1). *J. Med. Genet.* **1999**, *36* (7), 546–548.
28. Zühlke, C.; Dalski, A.; Hellenbroich, Y.; Bubel, S.; Schwinger, E.; Bürk, K. Spinocerebellar Ataxia Type 1 (SCA1): Phenotype-Genotype Correlation Studies in Intermediate Alleles. *Eur. J. Hum. Genet. EJHG* **2002**, *10* (3), 204–209. <https://doi.org/10.1038/sj.ejhg.5200788>.
29. Frontali, M.; Novelletto, A.; Annesi, G.; Jodice, C. CAG Repeat Instability, Cryptic Sequence Variation and Pathogenicity: Evidence from Different Loci. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **1999**, *354* (1386), 1089–1094. <https://doi.org/10.1098/rstb.1999.0464>.
30. Dong, Y.; Sun, Y.-M.; Liu, Z.-J.; Ni, W.; Shi, S.-S.; Wu, Z.-Y. Chinese Patients with Huntington's Disease Initially Presenting with Spinocerebellar Ataxia. *Clin. Genet.* **2013**, *83* (4), 380–383. <https://doi.org/10.1111/j.1399-0004.2012.01927.x>.
31. Bilgen, C.; Bilgen, I. G.; Sener, R. N. Oculopharyngeal Muscular Dystrophy: Clinical and CT Findings. *Comput. Med. Imaging Graph. Off. J. Comput. Med. Imaging Soc.* **2001**, *25* (6), 527–529. [https://doi.org/10.1016/s0895-6111\(01\)00021-0](https://doi.org/10.1016/s0895-6111(01)00021-0).

32. Richard, P.; Trollet, C.; Stojkovic, T.; de Becdelievre, A.; Perie, S.; Pouget, J.; Eymard, B.; Neurologists of French Neuromuscular Reference Centers CORNEMUS and FILNEMUS. Correlation between PABPN1 Genotype and Disease Severity in Oculopharyngeal Muscular Dystrophy. *Neurology* **2017**, *88* (4), 359–365. <https://doi.org/10.1212/WNL.0000000000003554>.
33. Brais, B. Oculopharyngeal Muscular Dystrophy: A Late-Onset Polyalanine Disease. *Cytogenet. Genome Res.* **2003**, *100* (1–4), 252–260. <https://doi.org/10.1159/000072861>.
34. Figueroa, K. P.; Gross, C.; Buena-Atienza, E.; Paul, S.; Gandelman, M.; Kakar, N.; Sturm, M.; Casadei, N.; Admard, J.; Park, J.; Zühlke, C.; Hellenbroich, Y.; Pozojevic, J.; Balachandran, S.; Händler, K.; Zittel, S.; Timmann, D.; Erdlenbruch, F.; Herrmann, L.; Feindt, T.; Zenker, M.; Klopstock, T.; Dufke, C.; Scoles, D. R.; Koepfen, A.; Spielmann, M.; Riess, O.; Ossowski, S.; Haack, T. B.; Pulst, S. M. A GGC-Repeat Expansion in ZFH3 Encoding Polyglycine Causes Spinocerebellar Ataxia Type 4 and Impairs Autophagy. *Nat. Genet.* **2024**, *56* (6), 1080–1089. <https://doi.org/10.1038/s41588-024-01719-5>.
35. Hsiao, C.-T.; Liao, N.-Y.; Liao, Y.-C.; Lee, Y.-C. THAP11 CAG Repeat Expansion Is Rare or Absent in the Taiwanese Cohort with Cerebellar Ataxia. *Mov. Disord. Off. J. Mov. Disord. Soc.* **2024**, *39* (5), 924–925. <https://doi.org/10.1002/mds.29800>.
36. Saffie-Awad, P.; Moller, A.; Daida, K.; Alvarez Jerez, P.; Chen, Z.; Anderson, Z. B.; Isayan, M.; Paquette, K.; Gibson, S. B.; Fulcher, M.; Miano-Burkhardt, A.; Malik, L.; Baker, B.; Jarreau, P.; Houlden, H.; Ryten, M.; Gu, B.; Chaisson, M. J. P.; Miller, D. E.; Chaná-Cuevas, P.; Blauwendraat, C.; Singleton, A. B.; Billingsley, K. J. Identification of GGC Repeat Expansions in ZFH3 among Chilean Movement Disorder Patients. *Mov. Disord. Off. J. Mov. Disord. Soc.* **2025**, *40* (7), 1433–1441. <https://doi.org/10.1002/mds.30242>.
37. Tan, D.; Wei, C.; Chen, Z.; Huang, Y.; Deng, J.; Li, J.; Liu, Y.; Bao, X.; Xu, J.; Hu, Z.; Wang, S.; Fan, Y.; Jiang, Y.; Wu, Y.; Wu, Y.; Wang, S.; Liu, P.; Zhang, Y.; Yang, Z.; Jiang, Y.; Zhang, H.; Hong, D.; Zhong, N.; Jiang, H.; Xiong, H. CAG Repeat Expansion in THAP11 Is Associated with a Novel Spinocerebellar Ataxia. *Mov. Disord. Off. J. Mov. Disord. Soc.* **2023**, *38* (7), 1282–1293. <https://doi.org/10.1002/mds.29412>.
38. Akbas, F.; Erginel-Unaltuna, N. DNA Testing for Huntington Disease in the Turkish Population. *Eur. Neurol.* **2003**, *50* (1), 20–24. <https://doi.org/10.1159/000070854>.
39. Musilova, A.; Lassuthova, P.; Uhrova Meszarosova, A.; Straka, B.; Krejčikova, J.; Berounska, A.; Vlckova, M.; Musova, Z.; Safka Brozkova, D. Reanalysis of Next-Generation Sequencing Data to Detect Tandem Repeat Expansions in 1,106 Czech Probands With Neurologic Disease. *Neurol. Genet.* **2025**, *11* (4), e200272. <https://doi.org/10.1212/NXG.0000000000200272>.
40. van Vugt, J. J. F. A.; Zwamborn, R. A. J.; Dolzhenko, E.; Eberle, M. A.; Weisburd, B.; Bekema, E.; Kooyman, M.; Wang, B.; Project MinE ALS Sequencing Consortium; Kamsteeg, E.-J.; Losekoot, M.; Baas, F.; Novy, C.; Høyer, H.; van Eijk, R. P. A.; van Es, M. A.; van Rheenen, W.; Al-Chalabi, A.; van den Berg, L. H.; Veldink, J. H. The Role of Disease-Associated Short Tandem Repeats in Amyotrophic Lateral Sclerosis. *Brain Commun.* **2025**, *7* (6), fcaf482. <https://doi.org/10.1093/braincomms/fcaf482>.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.