



Communication

ONT in Clinical Diagnostics of Repeat Expansion Disorders: Detection and Reporting Challenges

Ludmila Kaplun *, Greice Krautz-Peterson, Nir Neerman, Yocheved Schindler, Elinor Dehan, Claudia S. Huettner, Brett K. Baumgartner, Christine Stanley and Alexander Kaplun

Variantyx Inc., Framingham, MA 01701, USA; greice.krautz@variantyx.com (G.K.-P.); brett.baumgartner@variantyx.com (B.K.B.)

* Correspondence: lucy.kaplun@variantyx.com

Abstract: While whole-genome sequencing (WGS) using short-read technology has become a standard diagnostic test, this technology has limitations in analyzing certain genomic regions, particularly short tandem repeats (STRs). These repetitive sequences are associated with over 50 diseases, primarily affecting neurological function, including Huntington disease, frontotemporal dementia, and Friedreich's ataxia. We analyzed 2689 cases with movement disorders and dementia-related phenotypes processed at Variantyx in 2023–2024 using a two-tiered approach, with an initial short-read WGS followed by ONT long-read sequencing (when necessary) for variant characterization. Of the 2038 cases (75.8%) with clinically relevant genetic variants, 327 (16.0%) required additional long-read analysis. STR variants were reported in 338 cases (16.6% of positive cases), with approximately half requiring long-read sequencing for definitive classification. The combined approach enabled the precise determination of repeat length, composition, somatic mosaicism, and methylation status. Notable advantages included the detection of complex repeat structures in several genes such as *RFC1*, *FGF14*, and *FXN*, where long-read sequencing allowed to determine somatic repeat unit variations and accurate allele phasing. Further studies are needed to establish technology-specific guidelines for the standardized interpretation of long-read sequencing data for the clinical diagnostics of repeat expansion disorders.



Academic Editor: Alfredo Ciccodicola

Received: 5 February 2025

Revised: 15 March 2025

Accepted: 17 March 2025

Published: 18 March 2025

Citation: Kaplun, L.; Krautz-Peterson, G.; Neerman, N.; Schindler, Y.; Dehan, E.; Huettner, C.S.; Baumgartner, B.K.; Stanley, C.; Kaplun, A. ONT in Clinical Diagnostics of Repeat Expansion Disorders: Detection and Reporting Challenges. *Int. J. Mol. Sci.* **2025**, *26*, 2725. <https://doi.org/10.3390/ijms26062725>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: ONT; long reads; genetic testing; ataxia; dementia; WGS; repeat expansion; Nanopore

1. Introduction

Reduction in the cost of next-generation sequencing (NGS) in recent years have paved the way for a shift in the paradigm of clinical genetic testing, allowing for comprehensive whole-genome sequencing (WGS)-based technology to be used in first-line diagnostic tests [1–3]. Currently, the standard approach to clinical diagnostic WGS relies on short-read sequencing due to the highly accurate calling of most types of genetic variants, reasonable costs of the required equipment and consumables, and availability of lab automation. Despite the many advantages of this approach, short-read sequencing technologies still have significant intrinsic limitations that have not been fully addressed, even with the wide array of existing bioinformatic analytical tools [4–8]. Among these limitations are problematic areas such as nonunique genomic regions and tandem repeats, which include short tandem repeats of few base pairs (STRs), longer variable number tandem repeats (VNTRs), and centromere/telomere microsatellites [9,10]. Those shortcomings are due to read length constraints that prevent the spanning of larger variants and may cause

ambiguous mapping within repetitive genomic regions [4,6,7]. It has been demonstrated that such genomic regions are best analyzed with long-read sequencing technologies that have the ability to resolve both the overall length and the actual sequence of these low complexity regions. Moreover, long-read sequencing adds information on DNA methylation status, which is critical for the interpretation of some genetic variants [11–15].

Pathogenic STR expansions are one of the major types of genetic aberrations associated with a range of phenotypic abnormalities, primarily with neurological symptoms (such as ataxia, epilepsy, and cognitive impairment) and, in some conditions, also with physical manifestations. Over 50 diseases, including Huntington disease, frontotemporal dementia, Friedreich's ataxia, spinocerebellar ataxias, cerebellar ataxia, neuropathy, and vestibular areflexia syndrome (CANVAS), etc., are currently associated with STR expansions, and this number grows annually with increased awareness and technological advances [12,16–24]. Pathogenic STRs vary by location, composition of the repeat unit, and size of the expansions [9,25–29]. For most STRs, repeat size ranges are defined as normal, mutable normal, premutation, reduced penetrance, or full penetrance alleles. STRs are highly heterogeneous variants, where not only the repeat unit count but also the composition of the repeat loci may differ between individuals, whether healthy or affected. While some pathogenic STR expansions just represent longer stretches of the same repeat units that are found in the unaffected population, other STR regions have pathogenic repeat units which differ from the benign unit type. Such an expansion might completely replace the benign units or be embedded among them and might even be composed of different stretches of various repeats, some of which have yet to be decidedly classified as benign or pathogenic. In addition, the somatic variability of the repeat length and composition is also a known feature of some STR regions [30–32]. In some conditions, the composition of the expanded repeat regions may influence the age of onset, penetrance, and severity of symptoms, with interrupting sequences acting as stabilizing factor that affects the extent of somatic heterogeneity in the individual and repeat length expansion/contraction in transfer between generations [12,29,30,33–40].

In cases of recessive disorders associated with expanded STRs, both the detection of expanded allele/s and the specific size of each allele are critical for the detection of the diagnostic variant, although these data might not always be apparent in some testing approaches. Moreover, in some recessive conditions, the combination of a pathogenic STR allele and a pathogenic non-STR variant on the other allele might also lead to a phenotype. Therefore, the ability to perform highly accurate analyses of STR repeat length, allele repeat unit composition, and allele phasing is paramount for an accurate diagnosis. However, pathogenic STR repeat expansions often stretch beyond the dimensions of the short sequencing reads, which are typically limited to 150 bp, making the determined variant length only predictable and the exact region composition unattainable.

The vast majority of clinical diagnostic laboratories use PCR, repeat-primed PCR, or Southern blot to evaluate STR expansion length and mosaic variability. Of these techniques, only repeat-primed PCR may assist with detection of the repeat region composition. These methods require specific assays for each tested STR, have high DNA input requirements, and are expensive, particularly when multiple STRs need to be examined [39].

Here we demonstrate that the combination of short-read-based WGS with ONT long-read sequencing for the detection of STR expansions significantly improves the diagnostic performance of clinical short-read WGS-based genetic testing for neurological disorders. Such a combined approach allows for the differentiation of pathogenic, premutation, intermediate, and benign repeat expansions even at a longer range, as well as the determination of repeat loci compositions and methylation status, where applicable. Moreover, since this approach is not targeted, it intrinsically includes all STRs and, upon the discovery of novel

disease-causing STR expansions, the existing sequencing data can be reanalyzed to examine those additional regions. We also discuss challenges in interpretation and clinical reporting of NGS-based test results due to differences between this technique and the conventional methods that form the basis of the current clinical interpretation guidelines.

2. Results

The typical processing routine of most clinical genetic tests at Variantyx utilizes a two-tiered approach. First, every patient sample undergoes a short-read-based WGS. While small sequence changes, longer structural variants, mitochondrial variants, and some shorter STR expansions can be fully analyzed with short-read data alone, other variants require further characterization to determine their pathogenicity. Such variants mostly include longer STR expansions, structural variants in low-complexity regions, and recessive variants of any kind where pathogenicity detection depends on the ability to reliably haplotype the alleles. When such potentially reportable variants are detected with short-read WGS, the sample undergoes additional WGS using ONT long-read sequencing [41].

To assess the diagnostic value of long-read sequencing as a complementary approach to the short-read WGS testing of neurological disorders, we performed a retrospective analysis of 2689 cases with movement disorders and dementia-related phenotypes processed at Variantyx in 2023 and 2024.

In 2038 cases (75.8%) one or more clinically relevant genetic variants were reported, including STR expansions and other variant types such as small sequence changes, structural variants, and mitochondrial variants (see Table 1). Cases where STRs were included in the clinical report alone or in combination with another variant(s) are presented in the ‘STR expansion’ category. Cases without reportable variants (651 cases, 24.2%) are not included in this table.

Table 1. Distribution of cases with reportable variants by sequencing technology and types of reported variants.

	No Long-Read Sequencing Required	Long-Read Sequencing Required	Total
Cases with STR expansion	184 (9.0%)	154 (7.6%)	338 (16.6%)
Cases with Other variant types (No STR)	1527 (74.9%)	173 (8.5%)	1700 (83.4%)
Total	1711 (84.0%)	327 (16.0%)	2038 (100.00%)

Out of these 2038 cases, 327 (16.0%) required additional variant analysis following short-read WGS analysis, as conclusions about the variant/s sequence, length, or phasing (when relevant) could not be made based on short-read technology alone (Table 1). Therefore, as a second-tier analysis, those cases were additionally sequenced with long-read WGS, generating a double-technology cohort, with 184 cases (i.e., 9.0% of all cases with reportable variants) specifically analyzed to determine the exact length and/or composition of the STR expansions, as well as to achieve accurate haplotyping in the cases of recessive STR conditions.

The expanded STR variants were reported in 16.6% (338 cases) of all cases with reportable variants. However, only about half of those cases were reportable based on the short-read WGS data alone, while the analysis of the other half required ONT long-reads to determine the expansion range and to identify the exact sequence of the expansion (Table 2). That included all longer expansion variants in the STR regions where the reportable threshold significantly exceeded the length of the short reads, such as longer variants in *FGF14*,

FXN, *FMR1*, *RFC1*, etc. (Table 2, Figure 1). Long-read sequencing was also required to precisely determine the length of shorter repeat expansions, such as *ATXN2*, *ATXN7*, or *HTT*, in cases where the length of the expansion exceeded 135–140 bp and could not be reliably determined by 150 bp read-based technology (Table 2, Figure 1). It also provided additional information which was useful for variant interpretation, such as the 5mC methylation status (in case of *FMR1* expansions) and allele structure/repeat units composition (such as *FGF14* and *RFC1*, for example), as well as for establishing the pathogenic status of the biallelic expansions for longer repeats when short-read sequencing did not allow for an accurate differentiation between two long pathogenic alleles and one long pathogenic and another long within the subpathogenic range (such as the *RFC1* and *FXN* STR variants).

Table 2. STR variant ranges and detection technology used for reporting in the movement disorders and dementia cohort *.

STR Locus	Ranges					Reported Variants			Total
	Normal Range	Mutable Normal	Intermediate/Uncertain Range	Reduced Penetrance Range	Pathogenic Threshold	Repeat Unit Length	No Long-Read Sequencing	Long-Read Sequencing	
<i>AR</i>	<34		35	36–37	38	3	1	0	1
<i>ATN1</i>	6–35			36–47	48	3	1	1	2
<i>ATXN1</i>	6–35 (36–44 **)	36–38			39 (46 *)	3	11	5	16
<i>ATXN10</i>	10–32		33–280	280–800	800	5	3	2	5
<i>ATXN2</i>	<30		30–32 ***	33–34	35	3	23	5	28
<i>ATXN3</i>	44		45–49	50–55	56	3	18	8	26
<i>ATXN7</i>	27	28–33		34–36	37	3	3	3	6
<i>ATXN8OS</i>	15–50		51–53	54 ^	54 ^	3	23	11	34
<i>BEAN1</i> ^^	-			-	110	5	1	0	1
<i>C9ORF72</i>	≤24		25–60	24–60	61	6	6	3	9
<i>CACNA1A</i>	18		19	19–19	20	3	25	3	28
<i>CNBP</i>	26		27–74		75	4	1	3	4
<i>CSTB</i>	2–3	12–17			30	12	0	1	1
<i>DIP2B</i>	6–23			139–206	250	3	0	1	1
<i>DMPK</i>	5–34	35–49			50	3	1	0	1
<i>FGF14</i>				180–319	320	3	22	54	76
<i>FMR1</i>	5–44		45–54	55–200 ^^	201	3	6	10	16
<i>FXN</i>	5–33		34–65		66	3	29	13	42
<i>GLS</i>	5–38			-	680	3	0	1	1
<i>HTT</i>	26	27–35		36–39	40	3	2	7	9
<i>JPH3</i>	28			29–39	40	3	0	2	2
<i>NOP56</i>	3–14			15–649	650	6	0	3	3
<i>NOTCH2NLC</i>	≤40		41–59		60	3	0	1	1
<i>PABPN1</i>	10			11–11 homozygous	11–18	3	2	1	3
<i>PPP2R2B</i>	7–32			40–50	51	3	2	0	2

Table 2. Cont.

STR Locus	Ranges					Reported Variants			Total
	Normal Range	Mutable Normal	Intermediate/Uncertain Range	Reduced Penetrance Range	Pathogenic Threshold	Repeat Unit Length	No Long-Read Sequencing	Long-Read Sequencing	
<i>RFC1</i>				11–200	400	5	5	15	20
<i>TBP</i>	25–40			41–48	49	3	12	3	15
<i>TCF4</i>	40			40–50	51	3	1	0	1
<i>ZFHX3</i>	31		31–41		42	3	1	2	3

* Many patients included in the cohort had complex phenotypes with partial overlap with movement disorders/dementia symptoms. ** Interrupted range. *** Range of association with amyotrophic lateral sclerosis. ^ Full penetrance alleles are not recognized in this STR. ~ *BEAN1* expansions cannot be detected using short-read WGS and require primary variant calling using ONT data. Thus, diagnostic tests in which analysis of this STR is included do not utilize the two-tier sequencing approach. Instead, samples are sequenced with both short- and long-read-based WGS simultaneously. Such tests are thus excluded from the presented patient cohort, and the only positive *BEAN1* case included in this study underwent ONT WGS sequencing to evaluate a different STR expansion that was eventually confirmed to be negative. ~~~ Premutation FXTAS/FXPOI. STR regions with somatic mosaicism shown in purple font.

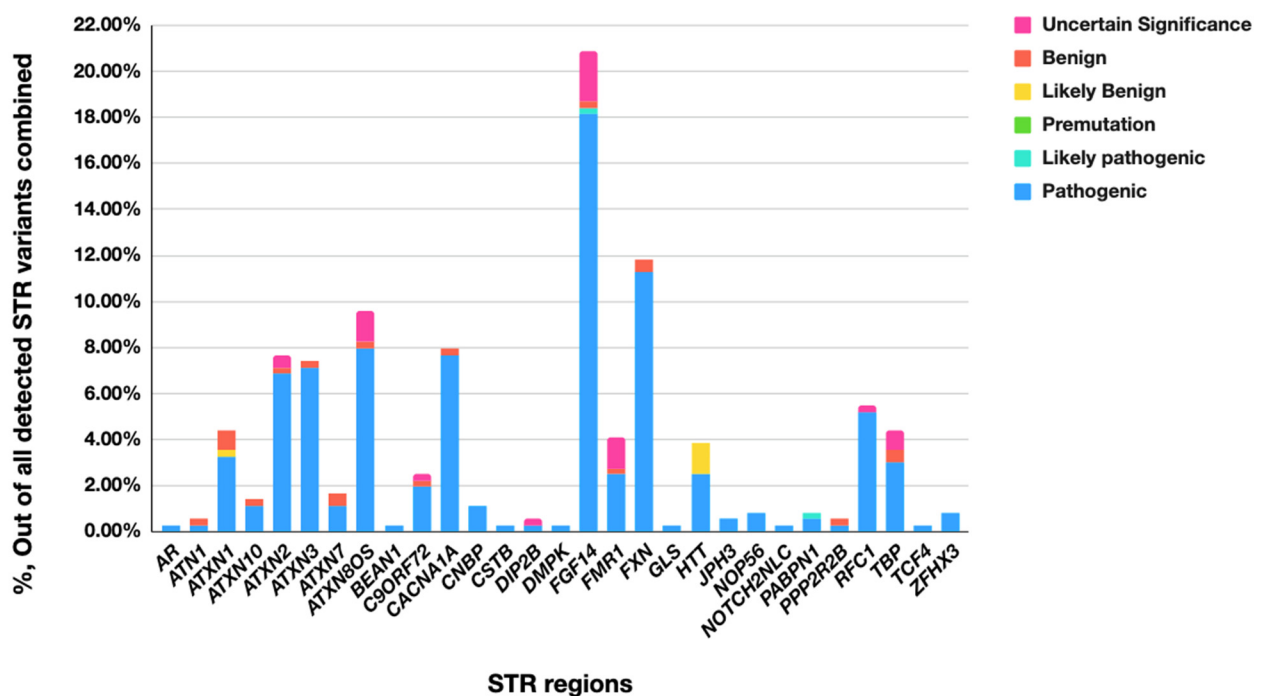


Figure 1. Diagnostic value of STR genotypes using ONT long-read sequencing. Axis X lists STR loci; axis Y indicates % of each category of each STR out of all detected STR variants combined.

While 91.2% of the reported STR variants were conclusively classified as either pathogenic or benign following long-read sequencing, 8.8% of variants could not be definitively classified due to repeat expansions falling into previously unannotated length ranges, having unusual repeat unit types or interruptions, or presenting contradictory evidence across the published studies. These variants were classified as variants of uncertain clinical significance (VUS). Therefore, the incorporation of long-read sequencing contributed to an increase in cases with conclusive results including STR expansions by about 40%. Notably, the incorporation of long reads into the diagnosis of other variant types demonstrated more modest improvements (by about 6%), primarily attributed to the lower proportion of variants necessitating additional characterization and the prevalence of variants of uncertain significance (VUSs).

Predictably, the STR variants most frequently detected and characterized with the addition of long-read sequencing were expansions with high reporting threshold and complex expanded region structure within the loci associated with highly prevalent neurological disorders (Figure 1) [12,17,42–44]. Of particular interest here was a high fraction of cases with long expansions in the *FGF14* STR region, which was only recently characterized as a common cause of neurological disorder partially overlapping *RFC1*-related phenotypes [43]. This is one of the targets where the collected information contributes to the evolving knowledge base of alleles composition and populational prevalence.

On the other hand, some of the findings represent the variants in well-characterized loci where pathogenic alleles are not much longer than the range precisely detectable with short-read sequencing. A notable example is the *HTT* STR region which, despite a moderate expansion length, frequently required supplementation with long-read sequencing and ultimately yielded a significant fraction of alleles characterized as non-pathogenic (Figure 1). This stems from the close proximity of ranges between normal and abnormal alleles, with a strong association of expansion length with disease penetrance and the age of onset, thus necessitating rigorous verification of the expansion length.

Considering that long-read sequencing generates reads often covering the entire length of the expanded STR region, even for the longest variants, it provides a unique opportunity to analyze the structure of the expanded regions and review somatic mosaic composition (with various repeat units and the exact length of each fragment), where applicable. To fully benefit from this feature, a careful visual inspection of the long-read sequenced STR regions is routinely conducted at Variantyx at the time of clinical interpretation, recording the allele ranges and structures. This visual inspection is aligned with the best practices and guidelines for variant interpretation and reporting in NGS-based clinical genetic testing [23,44,45].

We utilized long-read-based WGS results to characterize the repeat motifs, mosaicism, and number of repeat interruptions to further characterize some of the STR loci with the most frequently detected variants in the movement disorder patient cohort.

We observed that some longer STR variants demonstrated a significant mosaicism of expansion length and composition of the expanded alleles. We were able to detect high variability of the repeat units and motifs that deviate from the frequent repeat composition. The atypical repeat unit and allele composition were especially prominent in the STR regions where the pathogenic repeat units differ from the regular repeat units observed in the unaffected individuals. For example, in the *RFC1* STR region, the typical benign repeat unit is AAAAG, while the most frequent pathogenic repeat unit is AAGGG, with AAAGG and multiple additional repeat units of various pathogenicity encountered on rarer occasions [39,46,47].

Figure 2 shows a biallelic repeat expansion composed mostly of AAGGG pathogenic repeats as detected by long-read sequencing. In contrast, in the initial analysis of this sample with short-read WGS, where the variant length and composition were inferred from much shorter sequence blocks, an expansion of uncertain length and the presence of pathogenic repeat units were detected, but it was not possible to place the alleles into the pathogenic range and detect the exact structure of the locus.

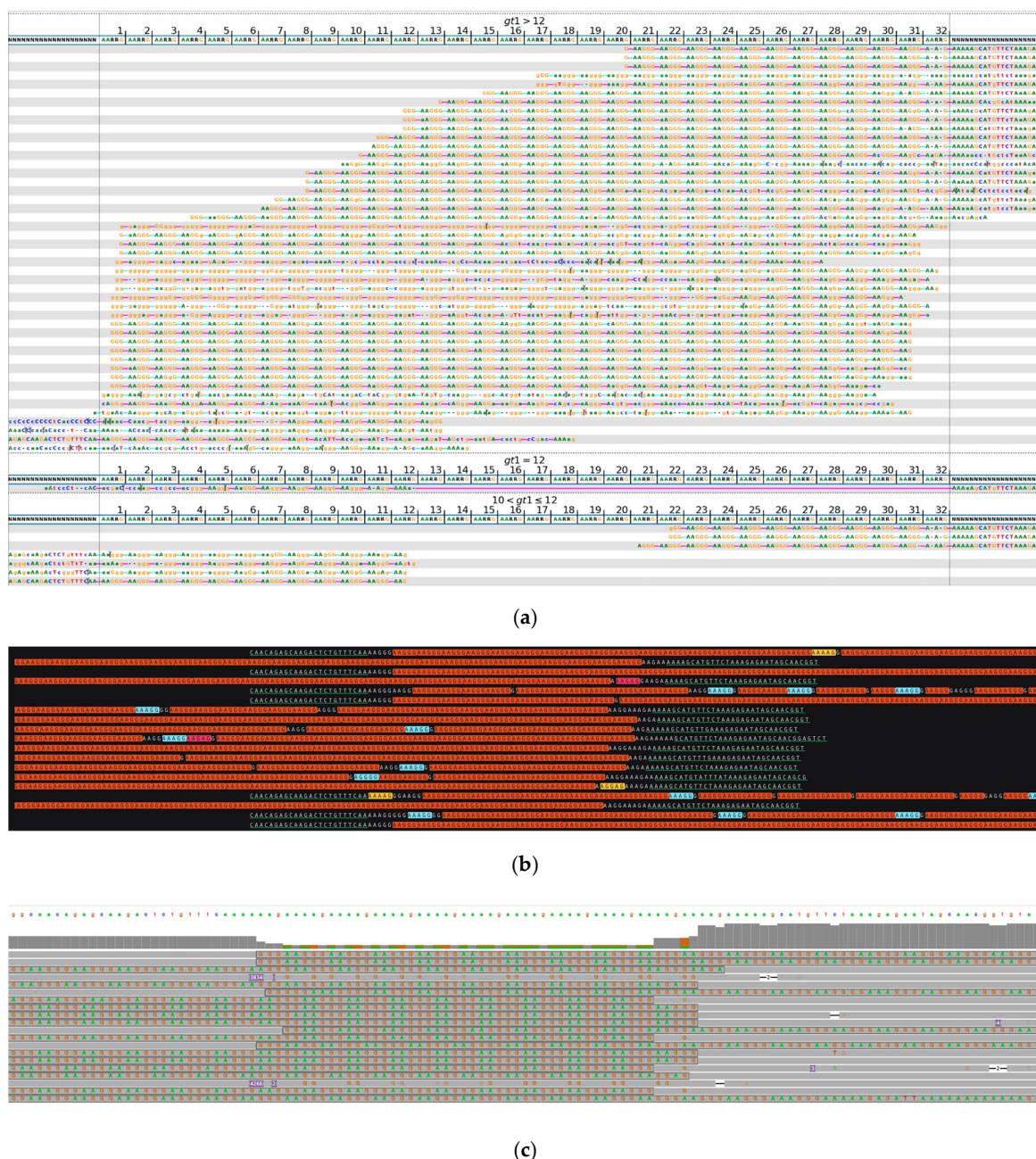


Figure 2. *RFC1*, biallelic expansion. (a) Visualization of short-read sequencing results, showing reads flanking the STR regions and those fully contained within the repeat. The bioinformatically predicted genotype based on read statistics is 49/49 * repeats. (b) Visualization of ONT long-read sequencing data, depicting a biallelic expansion of >200/>830 repeats, primarily composed of AAGGG pathogenic repeat units with occasional interrupting sequences. Green-underlined sequences highlight the regions flanking the STR repeat. Due to read length limitations, most of the reads are not able to capture both flanking regions of this repeat expansion. Different repeat units are highlighted in different colors: AAGGG (red), AAAGG (blue), AGAAG (yellow). (c) Integrative Genomic Viewer (IGV) visualization of ONT long-read sequencing data over the *RFC1* target region. Purple rectangles indicate insertions. * Results of bioinformatic prediction not depicted in the image.

During the analysis of long expansions, we observed not only mosaicism in the repeat length and the location of interruptions but also instances where alleles were predominantly composed of noncanonical repeat units. Such cases are displayed in Figure 3, with a mosaic expanded variant where some of the reads with longer *FGF14* expansions incorporate stretches of GGA repeats alongside the canonical GAA, as well as in Figure 4, with a longer

allele composed mostly of GAAGGA 6-nucleotide repeat followed by a short segment of a canonical GAA 3-nucleotide expansion. Previous studies suggest that non-GAA repeats are likely non-pathogenic [21,29].

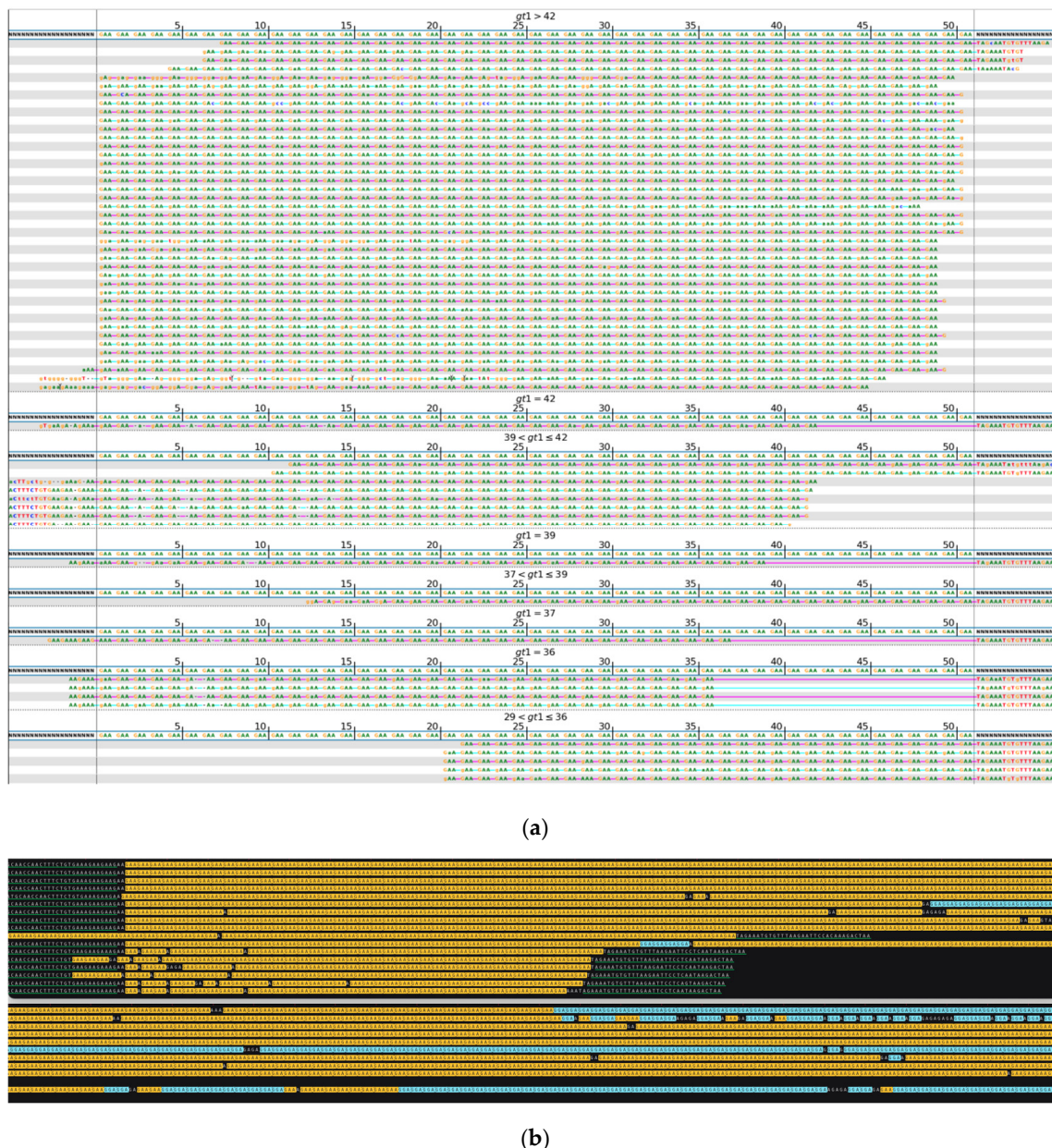


Figure 3. *FGF14*, heterozygous variant. (a) Visualization of short-read sequencing results, showing reads flanking the STR regions, as well as those fully contained within the repeat. The bioinformatically predicted genotype based on read statistics is 36/105 * repeats. (b) Visualization of the ONT long-read sequencing results, depicting a heterozygous repeat expansion of 41/420–481 repeats. The expansion shows mosaicism in both length and sequence, with some reads incorporating long stretches of GGA repeats in addition to the canonical GAA. Green-underlined sequences highlight the regions flanking the STR repeat. Different repeat units are highlighted in different colors: GAA (yellow), GGA (blue). The sequence shown in the upper panel continues in the lower panel. Due to limitations of the space and the length of the reads, some of them do not have both flanks visible in the included image. * Results of bioinformatic prediction not depicted in the image.

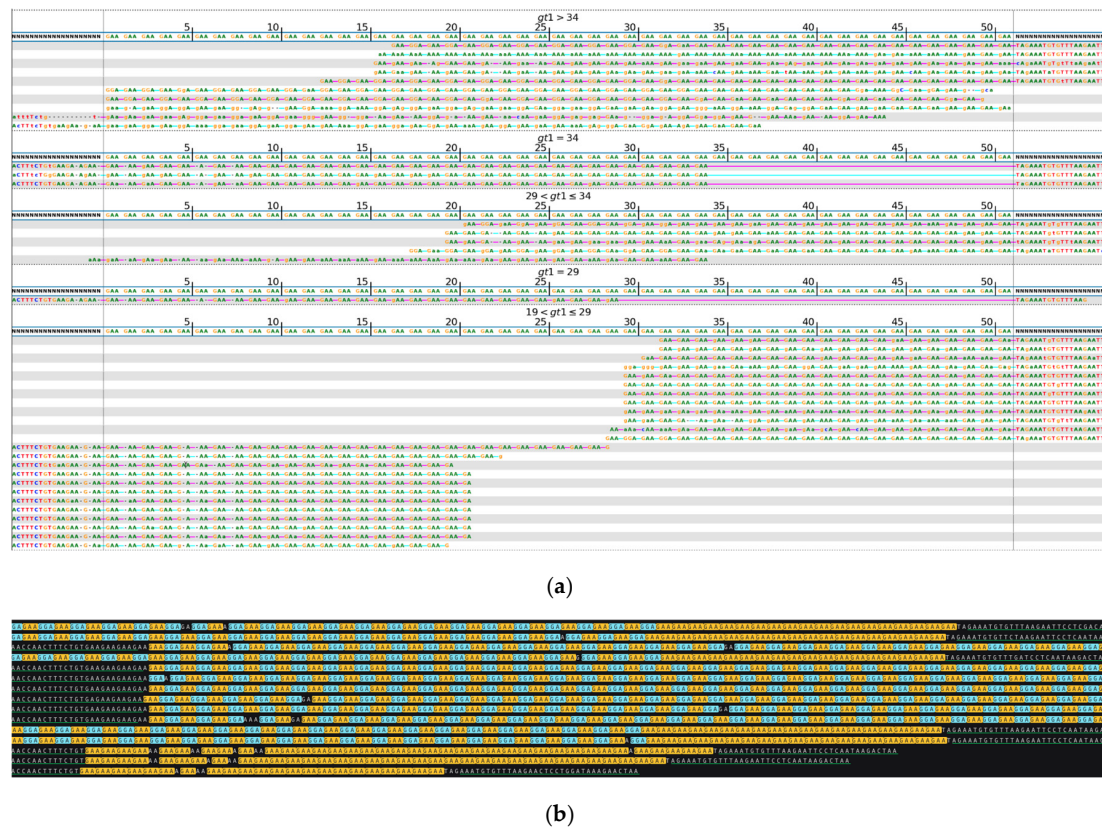


Figure 4. *FGF14*, with a noncanonical GAAGGA repeat unit. (a) Visualization of short-read sequencing results showing reads flanking the STR regions as well as those fully contained within the repeat. The bioinformatically predicted genotype based on read statistics is 36/56 * repeats. (b) Visualization of the ONT long-read sequencing results, depicting a heterozygous repeat expansion with 35–37/318–342 repeats, where the shorter allele is composed of canonical GAA repeats, while the longer one is predominantly composed of GAAGGA repeats, followed by a short stretch of canonical GAA. Green underlined sequences highlight the regions flanking the STR repeat. Different repeat units are highlighted in different colors: GAA (yellow), GGA (blue). Due to limitations of the space and the length of the reads, some of them do not have both flanks visible in the included image. * Result of bioinformatic prediction, not depicted in the image.

In addition to determining the specific length and structure of the complex STR regions, long-read sequencing offers the advantage of the accurate detection of the length of both alleles in conditions with a recessive inheritance, such as *FXN*. While the expansion of two alleles above the benign range can be easily observed with short-read sequencing, distinguishing between two long pathogenic alleles and a scenario where one is a very long allele while the other is either moderately expanded or falls within the intermediate range can be challenging. Short read-based analysis of long STR expansions typically relies on statistical predictions derived from the counts of reads flanking the repeats and fully encompassed by them. In contrast, long-read sequencing is an invaluable tool in such cases, enabling the direct determination of the full length of each allele. Figure 5 shows a biallelic *FXN* repeat expansion, where short read-based evaluation predicted a 76/76 genotype with two similar length pathogenic alleles composed predominantly of GAA repeats. However, an estimation of the GAA repeat expansion length based on the statistical analysis of reads derived from short-read sequencing cannot be accurate due to the general abundance of unrelated genomic loci composed of GAA repeats in the human genome. The long-read sequencing of this case demonstrated one of the alleles of around 100 repeats, while the other extended beyond 600 repeats. This information is crucial for assessing disease severity and predicting the age of onset.

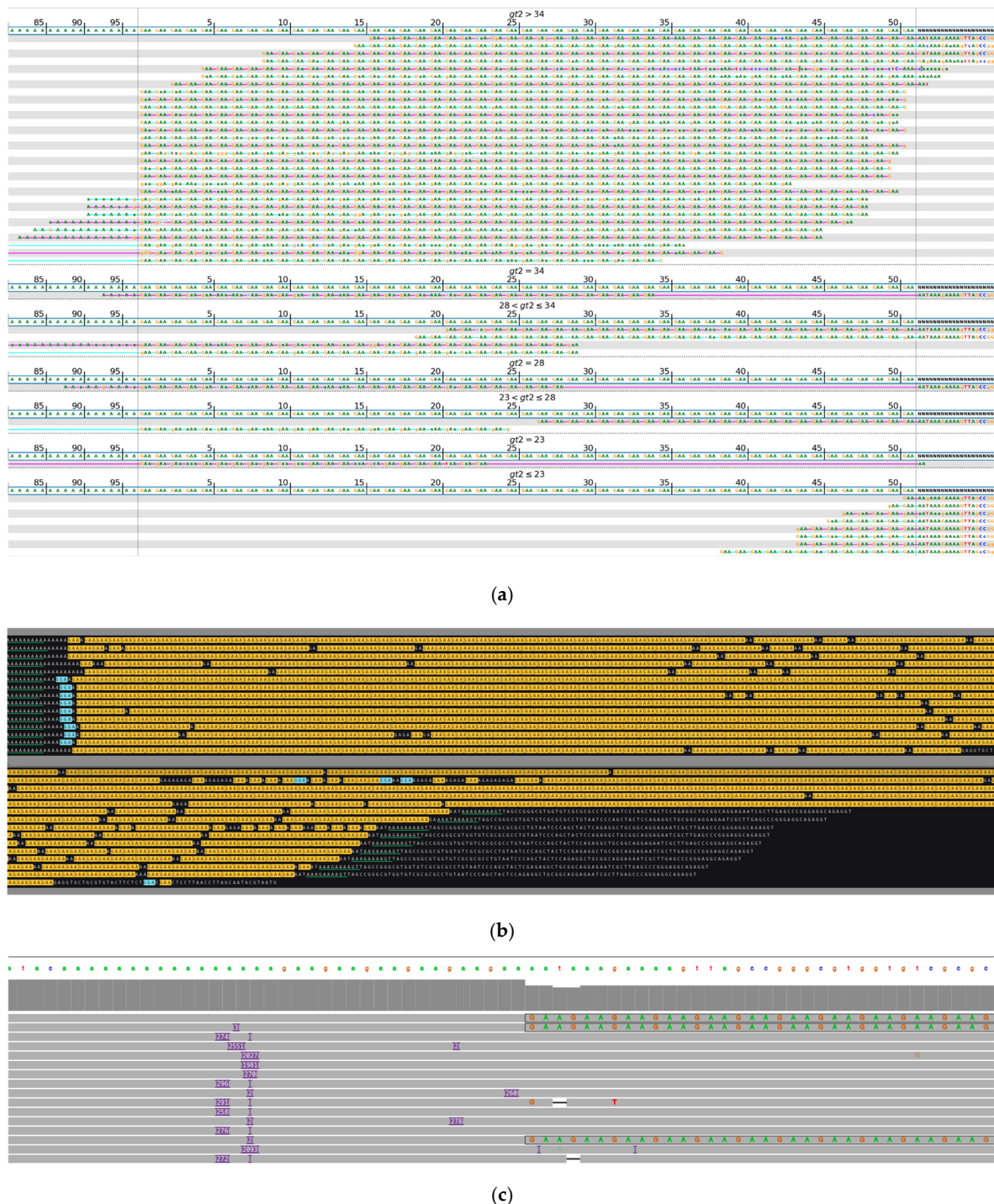


Figure 5. *FXN*, biallelic repeat expansion. **(a)** Visualization of short-read sequencing results, showing reads flanking the STR regions as well as those fully contained within the repeat. The bioinformatically predicted genotype based on the read statistics is 76/76 * repeats. **(b)** Visualization of the ONT long-read sequencing results, depicting a biallelic repeat expansion with one allele of 92–104 repeats and another one of 649–901, composed mainly of GAA units (highlighted in yellow) with occasional GGA interruptions (highlighted in blue). The sequence shown in the upper panel continues in the lower panel. Green-underlined sequences indicate regions flanking the STR repeat. Due to limitations of the space and the length of the reads, some of them do not have both flanks visible in the included image. **(c)** Integrative Genomic Viewer (IGV) visualization of ONT long-read sequencing data over the *FXN* target region. Purple rectangles indicate insertions. * Results of bioinformatic prediction not depicted in the image.

Epigenetic modifications represent another type of information that can be detected by long-read sequencing, contributing valuable diagnostic information when relevant. This is particularly true in cases of *FMR1* expansions, where pathogenicity has been associated with 5mC methylation in CpG context. In specimens with mosaic repeat expansions and/or interruptions in the long repeat regions, the simultaneous detection of 5mC methylation or lack thereof offers a means to better evaluate the presumed level of pathogenicity [48,49]. Figure 6 shows *FMR1* region in a male with premutation, where no 5mC methylation in the CpG context is observed in any of the reads. In contrast, Figure 7 illustrates a mosaic pathogenic *FMR1* expansion in a male, where longer stretches of interrupting repeats are present in some of the reads, and 5mC methylation in the CpG context is observed only on the expanded repeats without long TGG, AGG, CAA interruptions.

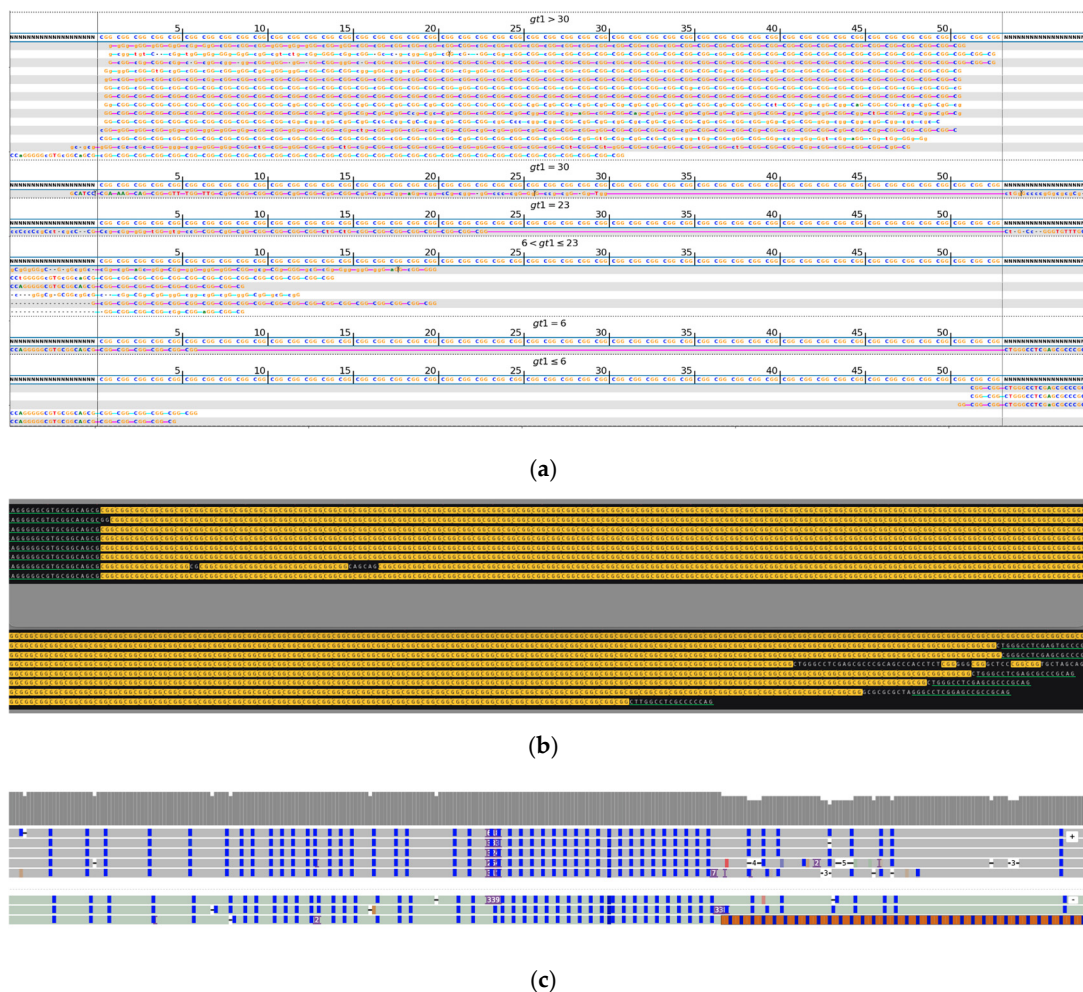


Figure 6. *FMR1*, male, premutation. (a) Visualization of short-read sequencing results showing reads flanking the STR regions, as well as those fully contained within the repeat. The bioinformatically predicted genotype based on read statistics is 156 * repeats. (b) Visualization of the ONT long-read sequencing results, showing an expansion of 130–132 repeats (highlighted in yellow). Green-underlined sequences indicate regions flanking the STR repeat. The sequence shown in the upper panel continues in the lower panel. Due to limitations of the space and the length of the reads, some of them do not have both flanks visible in the included image. (c) Integrative Genomic Viewer (IGV) visualization of the ONT long-read sequencing results with cytosines in CpG context color coded according to their epigenetic status: red for methylated and blue for unmethylated cytosines. No 5mC methylation is observed in this region. * Results of bioinformatic prediction not depicted in the image.

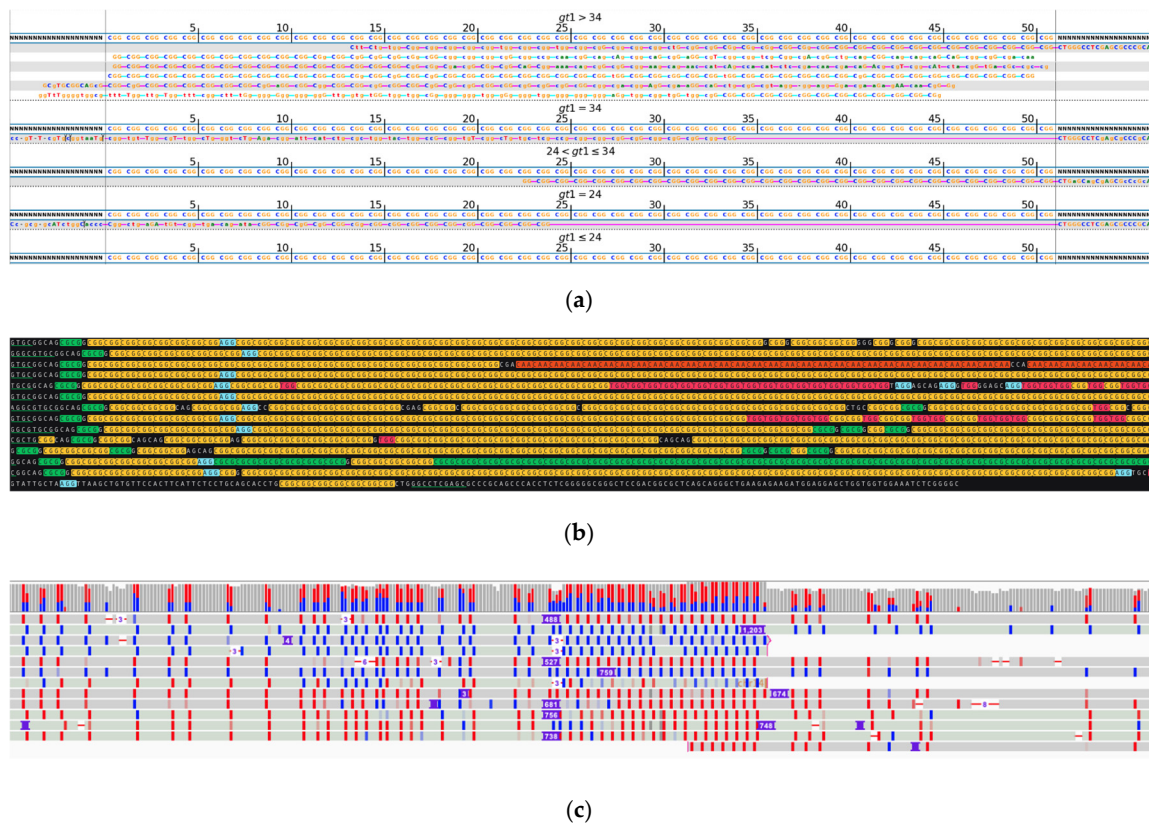


Figure 7. *FMR1* mosaic expansion, male. (a) Visualization of short-read sequencing results showing reads flanking the STR regions, as well as those fully contained within the repeat. The bioinformatically predicted genotype based on the read statistics is 81 repeats *. (b) Visualization of the ONT long-read sequencing results showing a total expansion length of 270–400 repeats, including 44–396 canonical CGG units and stretches of various noncanonical repeats in some of the reads. Green-underlined sequences indicate regions flanking the STR repeat. Different repeat units are highlighted in different colors: CGG (yellow), TGG (red), AGG (blue), CGCG (green). Due to limitations of the space and the length of the reads, some of them do not have both flanks visible in the included image. (c) Integrative Genomic Viewer (IGV) visualization of the ONT long-read sequencing results with cytosines in CpG context color coded according to their epigenetic status—red for methylated and blue for unmethylated cytosines. 5mC methylation is observed on the reads with expanded canonical repeats but only in the absence of long stretches of TGG, AGG, CAA interruptions. * Results of bioinformatic prediction not depicted in the image.

Figure 8 illustrates the detection of the length and structure of complex *ATXN8OS* repeat locus. Both the normal and expanded alleles have similar patterns of repeat units, where a segment consisting of CTA repeats is followed by a segment consisting of CTG repeats, while the length of each segment varies between the alleles. Some of the traditional STR detection techniques report such complex regions only in terms of their length, while sequencing allows to acquire more comprehensive information about the region structure as well as and expansion length.

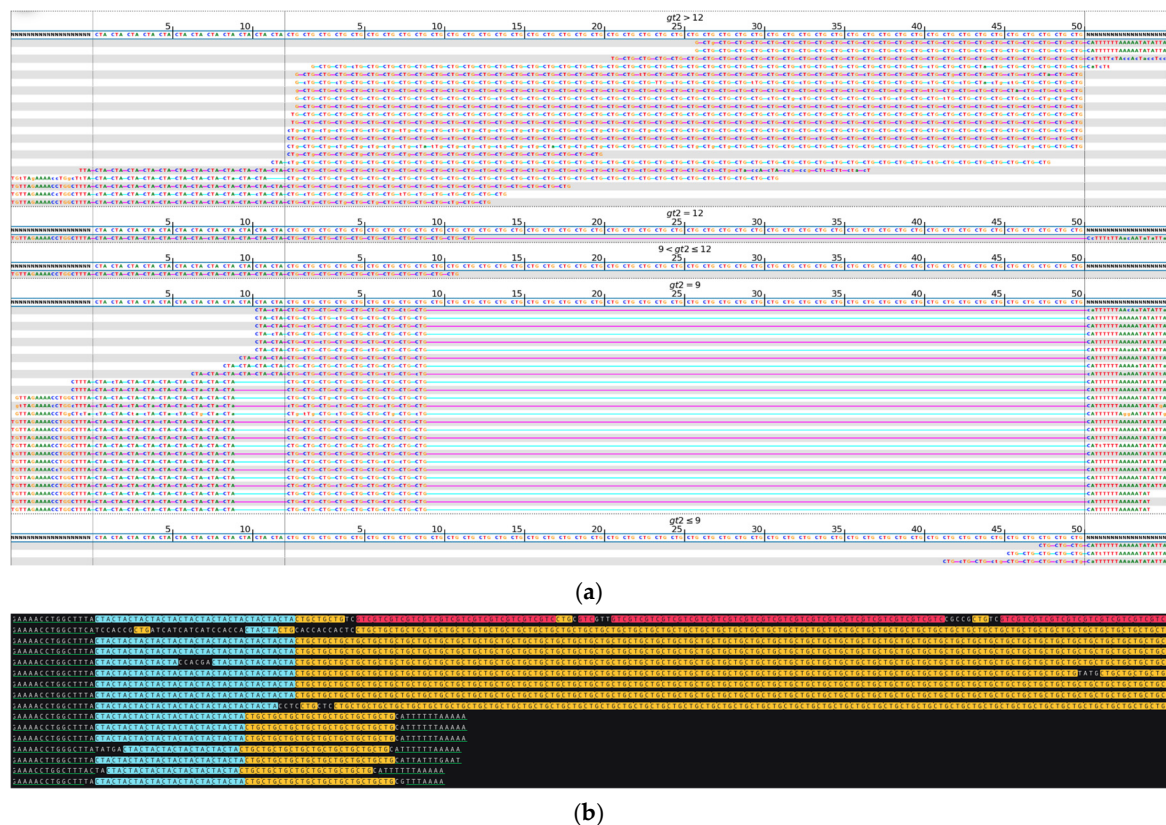


Figure 8. *ATXN8OS*, heterozygous repeat expansion, complex region with CTA-CTG structure. (a) Visualization of short-read sequencing results showing flanking the STR regions, as well as those and reads fully contained within the repeat. The bioinformatically predicted genotype based on the read statistics is 9 + 9/12 + 79 * repeats (CAT + CTG/CTA + CTG complex region). (b) Visualization of the ONT long-reads sequencing results showing total expansion length and structure—9 CAT + 9 CTG repeats (shorter, normal allele) and 12 CTA + 83–84 CTG repeats (expanded allele). Green-underlined sequences indicate regions flanking the STR repeat. Different repeat units are highlighted in different colors: CTG (yellow), CTA (blue), GTC (red). * Results of bioinformatic prediction not depicted in the image.

3. Discussion

Application of a two-tiered sequencing approach allows Variantyx to fully leverage the advantages of long-read sequencing technology in routine clinical genetic testing, while keeping the laboratory and analysis processes as simple as possible and the costs reasonably low. Both short- and long-read WGS provide comprehensive genomic coverage, unlike the targeted tests that currently prevail in the clinical genetic testing landscape. This approach allows for the rapid incorporation of newly discovered disease-causing STR expansions into clinical testing. For example, we were able to report our first pathogenic *ZFH3* expansion to a patient only several weeks after its association with SCA4 was published [50]. A clinical validation study performed by the laboratory revealed close to 100% sensitivity for most variant types [51,52]. This study included over 200 positive STR cases across various genes and repeat units. Comparing the diagnostic yield resulting from incorporating Oxford Nanopore Technologies (ONT) long-read sequencing in the analysis of different variant types, we observed a significantly higher impact in detection and characterization of STR expansions compared to other variant categories. Although long-read sequencing can benefit the detection of structural variants, variants with epigenetic changes and, to some extent, small sequence changes, this approach is particularly useful for the comprehensive analysis and interpretation of long and complex STR expansion variants.

The precise detection of STR expansion is particularly critical for patients with movement disorders and dementia. In the cohort of 2689 patient cases analyzed in this study, reportable variants were detected in 2038 cases (75.8%), including secondary and incidental findings, among which 1166 cases (43.4%) carried positive diagnostic results, examples of which are presented in Figures 2–8. Among the 2038 reportable cases, 338 included relevant STR variants, approximately half of which required the use of ONT long-read sequencing.

While long-read data provide more information on the size and composition of repeat expansions than any other technique used in clinical genetic testing, the differences in the observed long-read sequencing results as compared to the laboratories utilizing other techniques may present challenges in clinical interpretation.

The existing interpretation guidelines and the majority of publications based on the results obtained with other approaches do not take into account the somatic mosaicism that is often observed in patients, particularly in the STRs marked in color in Table 2 [12,29,30,33–40]. For instance, in *RFC1* expansions, the pathogenic repeat unit is often present only at the edges of the expansion, while the bulk of the expanded sequence is composed of different repeat units or combinations of repeated and non-repeated sequences (unpublished data to be summarized in a future publication). Most other detection techniques, which rely on the amplification or hybridization of stretches of pathogenic repeat units, will incorrectly assume in such cases that the entire expansion consists solely of the pathogenic repeat units. Consequently, in statistical studies that form the basis for interpretation guidelines, the defined expansion ranges associated with the pathogenicity of the observed repeat units differ from those detected with long reads. The direct application of these guidelines to the clinical interpretation of ONT long-read sequencing-derived results could lead to incorrect reporting.

Until further cohort studies utilizing this technology are conducted [5,53] and technology-specific guidelines are developed, results for certain patient cases may need to be reported as uncertain. The capacity to characterize unusual sequences, particularly in cases of mosaic expansions which cannot be accurately assessed without a single long molecule approach, increases the value of long-read sequencing by expanding the knowledge base during routine clinical analysis. The current standards in the clinical genetic testing of STR expansion variants are mostly geared toward the detection of STR expansion length and the inspection of known motifs, therefore potentially failing to identify the unusual variations affecting expansion pathogenicity and stability [54]. For example, minor interruptions in *FXN* repeat expansion encountered in more than 70% of Friedreich's Ataxia patients are not assessed by some of the currently offered genetic tests [40,55]. Recent updates in the ACMG technical standards recognize the detection and reporting challenges and advise to use caution when the reliability of the methodology might be affected by interruptions (for example, repeat-based PCR for the detection of CTG repeats such as *DMPK* where interruptions might increase the GC content, causing allele bias [56]). Another example that presents a challenge to traditional testing technologies is a multitude of various unusual repeat motifs in the *RFC1* STR region, which may be associated with an atypical phenotypic presentation in patients with cerebellar ataxia and have the potential to remain undiagnosed unless the appropriate long-range testing methodologies are employed [57,58]. At Variantyx, we are working on building a population cohort that will help the clinical community to establish updated guidelines for the interpretation of long-read sequencing data. Our database, which currently includes tens of thousands of short-read WGSs and thousands of ONT long-read WGS datasets from patients with a variety of phenotypes, is rapidly growing. Due to patient privacy considerations, we will not be able to make the raw sequencing data publicly accessible; however, we are planning to share our

summarized observations in the near future. These data were generated using short-read WGS as a primary methodology, supplemented with long-read WGS when necessary.

Another direction of future development would be switching entirely to long-read-based WGS, which would simplify the wet bench and bioinformatic analysis processes and potentially further improve the detection power of the genetic testing. At the current stage in the development of long-read technology, however, this approach is not yet feasible for large-scale routine diagnostic testing. Currently, the two main long-read-based NGS technologies are ones developed by Pacific Biosciences (PacBio) and by Oxford Nanopore Technologies (ONT) [59–61]. While PacBio HiFi technology generates long-read data accuracy nearly comparable to the industry-standard short-read sequencing for small sequence changes and copy number variants, the maximum length limitations render it unsuitable for the detection of extremely long variants. The current generation of PacBio sequencing machines is also low throughput, relative to the leading short-read sequencing solutions. Conversely, ONT long reads are highly useful in evaluating the copy number and structural variants of various sizes but less accurate in detecting small sequence changes compared to short-read sequencing.

The main constraint of long-read sequencing technologies that are currently commercially available is their relatively high cost, which at this time does not allow them to be used as a standalone solution for standard first-line diagnostic and screening testing. Therefore, these NGS technologies, together with a multitude of additional approaches (such as Illumina long reads [62]), still require further development before potentially serving as standalone comprehensive clinical testing tools suitable for any types of genetic disorders. Long-read-based sequencing technologies are rapidly evolving, and we expect that, in several years, these constraints will be resolved. Until then, a combinatorial approach in which long-read sequencing is used to supplement the short-read WGS (when needed) and which can significantly enhance clinical utility of the test is well suited for clinical genetic testing. In this study, we demonstrate that this is particularly true for diagnostics of movement disorders and dementia, which are often caused by genetic variants that are challenging to analyze using short-read sequencing alone, such as STR expansions.

4. Materials and Methods

4.1. Samples Selection

The study cohort consisted of patients with ataxia/dementia phenotypes, according to the standard HPO classification, and was analyzed at the Variantyx genetic testing laboratory (Framingham, MA, USA) over a study period of 24 months (2023 and 2024). The inclusion criteria were not limited to any gender, age, or ethnicity category of participants. The cohort was composed of all the specimens submitted to the Variantyx laboratory within the study period for germline genetic testing for neurological disorders and/or with neurological disorders phenotypes listed in the accompanying medical documentation.

All participants signed an informed consent prior to inclusion into the current study and all samples were deidentified.

Specimen collection was performed with Lavender K2-EDTA Blood Collection Tubes (367861, PulmoLab, Northridge, CA, USA) and saliva collection tubes (OGD-500 and OCR–100, DNA Genotek, Ottawa, Ontario, Canada) and extracted at Variantyx using the QIAasympohony, Qiagen EZ2, or QIAamp kit (cat #51106, QIAGEN, Venlo, The Netherlands) or delivered as pre-extracted gDNA from the submitting laboratories as arranged by the medical providers ordering the genetic testing.

4.2. Library Preparation and Illumina Short-Read Sequencing

A total of 300 ng of gDNA was used for library preparation utilizing the Illumina WGS PCR-Free Tagmentation protocol with multiplex ligation sequencing kits using Standard Workflow reagents (Cat# 20041794/20041795, 20,028,312 Illumina Inc., San Diego, CA, USA). The samples were sequenced using NovaSeq 6000 and NovaSeq X Plus sequencers (Illumina Inc., San Diego, CA, USA) with S4 and 10B/25B flow cells.

4.3. Bioinformatics and Analytical Approach Illumina Short Reads

The sequencing data were analyzed using the Variantyx Genomic Intelligence bioinformatic pipeline, combining publicly available and in-house developed tools for the detection of Small Sequence Changes (SSCs), Structural Variants including Copy Number Variants (SVs and CNVs), Mitochondrial SSCs with Heteroplasmy, Uniparental Disomy (UPD), and SMN1/2 Copy Numbers, as well as for the detection of STR variants and the prediction of the length of long expanded alleles [51]. Genetic variants as detected by the pipeline were visualized, analyzed, and reported with the in-house developed Diagnostic Console.

4.4. Library Preparation and ONT Long-Read Sequencing

A total of 1500 ng of gDNA was sheared with the FastPrep–96™ High-Throughput Bead Beating Grinder and Lysis System (MP Biomedicals, Irvine, CA, USA, Cat# SKU 116010500) at 1800 rpm for 3 min. The resulting DNA fragments were used for library preparation, utilizing the ONT Ligation sequencing gDNA protocol with multiplex ligation sequencing kits V11/14 (Cat# SQK-MLK111.96-XL/SQK-MLK114.96-XL, ONT, Oxford, UK) with incorporated native barcoding and low molecular weight fragment elimination buffer. Samples were sequenced using PromethION P24 ONT device (ONT, Oxford, UK) with R9.4/R10.4.1 flow cells (Cat# FLO-PRO002M/FLO-PRO114M, ONT, Oxford, UK).

4.5. Bioinformatics and Analytical Approach to ONT Long-Reads

Basecalling on high-accuracy settings (HAC) and demultiplexing were performed in parallel with sequencing using the MinKNOW v.23 software (ONT, Oxford, UK) integrated with PromethION P–24 sequencing device. The basecalling process also included the identification of the epigenetic modifications 5mC/5hmC in the CpG context. The HAC settings were selected based on the recommendations of the manufacturer as a high-quality approach compatible with real-time basecalling and recommended for projects focusing on variant analysis.

The acquired reads were processed with Variantyx proprietary Genomic Intelligence platform (<https://www.variantyx.com/>, accessed on 10 February 2025), generating long-read alignment with Minimap2 (v.2.23 <https://github.com/lh3/minimap2>, accessed on 1 December 2022). Genome assembly hg38 [63] was used as a reference.

Aligned reads were visualized with the Integrative Genomics Viewer (IGV, v3.0.2, <https://github.com/igvteam/igv.js>, accessed on 1 October 2024) and with the in-house developed diagnostic console. For short tandem repeat analysis, the results were manually inspected, registering both the regular and pathogenic repeat units (when applicable), as well as the presence of interrupting sequences. Definitions of the repeat regions, along with annotations of the normal and expanded repeat ranges, were retrieved from the Genome Aggregation Database [64], the Database of Short Tandem Repeats [65], and the STRchive [10].

In addition to the detection of genetic variants, 5mC methylation in the CpG context in the regions of interest was recorded when relevant.

Author Contributions: Conceptualization, A.K., L.K., C.S., and Y.S.; Data Curation, L.K., N.N., A.K. and G.K.-P.; Formal Analysis, L.K., A.K. and E.D.; Investigation, L.K. and A.K.; Methodology, L.K., A.K., C.S., Y.S., E.D., C.S.H. and B.K.B.; Project Administration, A.K.; Resources, G.K.-P., Y.S. and B.K.B.; Software, N.N. and L.K.; Visualization, N.N. and L.K.; Writing—Original Draft, L.K. and A.K.; Writing—Review and Editing, L.K., A.K., Y.S., E.D., B.K.B. and C.S.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable due to the retrospective nature of this study.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The aggregated and processed data supporting the findings of this study are available from the corresponding author upon request. The raw sequencing data cannot be made publicly available as they contain sensitive patient information protected under privacy laws and informed consent agreements.

Acknowledgments: The authors gratefully acknowledge Christopher Alsheikh, Margo Tarantino, and Ava McGarry for their valuable technical assistance and laboratory support throughout this research.

Conflicts of Interest: Authors Ludmila Kaplun, Greice Krautz-Peterson, Nir Neerman, Yocheved Schindler, Elinor Dehan, Claudia S. Huettner, Brett K. Baumgartner, Christine Stanley, and Alexander Kaplun were employed by the company Variantyx, Framingham, MA 01701, USA.

Abbreviations

NGS	Next-Generation Sequencing
WGS	Whole-Genome Sequencing
STR	Short Tandem Repeat
VNTR	Variable Number Tandem Repeat
ONT	Oxford Nanopore Technologies
HPO	Human Phenotype Ontology
gDNA	Genomic DNA
SSC	Small Sequence Change
SV	Structural Variant
CNV	Copy Number Variant
UPD	Uniparental Disomy
HAC	High-Accuracy Settings
IGV	Integrative Genomics Viewer
VUS	Variant of Uncertain Significance
SCA	Spinocerebellar Ataxia
FXTAS	Fragile X-Associated Tremor/Ataxia Syndrome
FXPOI	Fragile X-Associated Primary Ovarian Insufficiency

References

1. Wigby, K.M.; Brockman, D.; Costain, G.; Hale, C.; Taylor, S.L.; Belmont, J.; Bick, D.; Dimmock, D.; Fernbach, S.; Greally, J.; et al. Evidence Review and Considerations for Use of First Line Genome Sequencing to Diagnose Rare Genetic Disorders. *NPJ Genom. Med.* **2024**, *9*, 15. [[CrossRef](#)]
2. van der Sanden, B.P.G.H.; Schobers, G.; Corominas Galbany, J.; Koolen, D.A.; Sinnema, M.; van Reeuwijk, J.; Stumpel, C.T.R.M.; Kleefstra, T.; de Vries, B.B.A.; Ruiterkamp-Versteeg, M.; et al. The Performance of Genome Sequencing as a First-Tier Test for Neurodevelopmental Disorders. *Eur. J. Human. Genet.* **2022**, *31*, 81–88. [[CrossRef](#)]
3. Rajan-Babu, I.-S.; Peng, J.J.; Chiu, R.; Birch, P.; Couse, M.; Guimond, C.; Lehman, A.; Mwenifumbo, J.; van Karnebeek, C.; Friedman, J.; et al. Genome-Wide Sequencing as a First-Tier Screening Test for Short Tandem Repeat Expansions. *Genome Med.* **2021**, *13*, 126. [[CrossRef](#)] [[PubMed](#)]

4. Billingsley, K.J.; Meredith, M.; Daida, K.; Alvarez Jerez, P.; Negi, S.; Malik, L.; Genner, R.M.; Moller, A.; Zheng, X.; Gibson, S.B.; et al. Long-Read Sequencing of Hundreds of Diverse Brains Provides Insight into the Impact of Structural Variation on Gene Expression and DNA Methylation. Preprint. *bioRxiv* **2024**. [CrossRef]
5. Gustafson, J.A.; Gibson, S.B.; Damaraju, N.; Zalusky, M.P.; Hoekzema, K.; Twesigomwe, D.; Yang, L.; Snead, A.A.; Richmond, P.A.; De Coster, W.; et al. High-Coverage Nanopore Sequencing of Samples from the 1000 Genomes Project to Build a Comprehensive Catalog of Human Genetic Variation. *Genome Res.* **2024**, *34*, 2061. [CrossRef] [PubMed]
6. Gymrek, M.; Willems, T.; Reich, D.; Erlich, Y. Interpreting Short Tandem Repeat Variations in Humans Using Mutational Constraint. *Nat. Genet.* **2017**, *49*, 1495. [CrossRef]
7. Mahmoud, M.; Gobet, N.; Cruz-Dávalos, D.I.; Mounier, N.; Dessimoz, C.; Sedlazeck, F.J. Structural Variant Calling: The Long and the Short of It. *Genome Biol.* **2019**, *20*, 1–14. [CrossRef]
8. Pellerin, D.; Danzi, M.; Renaud, M.; Houlden, H.; Synofzik, M.; Zuchner, S.; Brais, B. GAA-FGF14-Related Ataxia. In *GeneReviews*®; University of Washington: Seattle, WA, USA, 2024. Available online: <https://www.ncbi.nlm.nih.gov/books/NBK599589/> (accessed on 10 February 2025).
9. Chiu, R.; Rajan-Babu, I.-S.; Friedman, J.M.; Birol, I. A Comprehensive Tandem Repeat Catalog of the Human Genome. *medRxiv* **2024**. [CrossRef]
10. Hiatt, L.; Weisburd, B.; Dolzhenko, E.; VanNoy, G.E.; Kurtas, E.N.; Rehm, H.L.; Quinlan, A.; Dashnow, H. STRchive: A Dynamic Resource Detailing Population-Level and Locus-Specific Insights at Tandem Repeat Disease Loci. *medRxiv* **2024**. [CrossRef]
11. Dolzhenko, E.; English, A.; Dashnow, H.; De Sena Brandine, G.; Mokveld, T.; Rowell, W.J.; Karniski, C.; Kronenberg, Z.; Danzi, M.C.; Cheung, W.A.; et al. Characterization and Visualization of Tandem Repeats at Genome Scale. *Nat. Biotechnol.* **2024**, *42*, 1606–1614. [CrossRef]
12. Chintalaphani, S.R.; Pineda, S.S.; Deveson, I.W.; Kumar, K.R. An Update on the Neurological Short Tandem Repeat Expansion Disorders and the Emergence of Long-Read Sequencing Diagnostics. *Acta Neuropathol. Commun.* **2021**, *9*, 1–20. [CrossRef]
13. Maestri, S.; Maturo, M.G.; Cosentino, E.; Marcolungo, L.; Iadarola, B.; Fortunati, E.; Rossato, M.; Delledonne, M. A Long-read Sequencing Approach for Direct Haplotype Phasing in Clinical Settings. *Int. J. Mol. Sci.* **2020**, *21*, 9177. [CrossRef]
14. Leitão, E.; Schröder, C.; Depienne, C. Identification and Characterization of Repeat Expansions in Neurological Disorders: Methodologies, Tools, and Strategies. *Rev. Neurol.* **2024**, *180*, 383–392. [CrossRef]
15. Vollger, M.R.; Korch, J.; Eldred, K.C.; Swanson, E.; Underwood, J.G.; Bohaczuk, S.C.; Mao, Y.; Cheng, Y.-H.H.; Ranchalis, J.; Blue, E.E.; et al. Synchronized Long-Read Genome, Methylome, Epigenome and Transcriptome Profiling Resolve a Mendelian Condition. *Nat. Genet.* **2025**, *57*, 469–479. [CrossRef] [PubMed]
16. 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*, 764–785. [CrossRef]
17. Ibañez, K.; Jadhav, B.; Zanolletto, M.; Gagliardi, D.; Clarkson, C.; Facchini, S.; Garg, P.; Martin-Trujillo, A.; Gies, S.J.; Galassi Deforie, V.; et al. Increased Frequency of Repeat Expansion Mutations across Different Populations. *Nat. Med.* **2024**, *30*, 3357–3368. [CrossRef] [PubMed]
18. Pellerin, D.; Danzi, M.C.; Wilke, C.; Renaud, M.; Fazal, S.; Dicaire, M.-J.; Scriba, C.K.; Ashton, C.; Yanick, C.; Beijer, D.; et al. Deep Intronic FGF14 GAA Repeat Expansion in Late-Onset Cerebellar Ataxia. *N. Engl. J. Med.* **2023**, *388*, 128–141. [CrossRef] [PubMed]
19. Ichikawa, K.; Kawahara, R.; Asano, T.; Morishita, S. A Landscape of Complex Tandem Repeats within Individual Human Genomes. *Nat. Commun.* **2023**, *14*, 5530. [CrossRef]
20. Gymrek, M. A Genomic View of Short Tandem Repeats. *Curr. Opin. Genet. Dev.* **2017**, *44*, 9–16. [CrossRef]
21. Pellerin, D.; Iruzubieta, P.; Tekgül, Ş.; Danzi, M.C.; Ashton, C.; Dicaire, M.J.; Wandzel, M.; Roth, V.; Lamont, P.J.; Bonnet, C.; et al. Non-GAA Repeat Expansions in FGF14 Are Likely Not Pathogenic—Reply to: “Shaking Up Ataxia: FGF14 and RFC1 Repeat Expansions in Affected and Unaffected Members of a Chilean Family”. *Mov. Disord.* **2023**, *38*, 1575–1577. [CrossRef]
22. 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*, 584–587. [CrossRef] [PubMed]
23. Austin-Tse, C.A.; Jobanputra, V.; Perry, D.L.; Bick, D.; Taft, R.J.; Venner, E.; Gibbs, R.A.; Young, T.; Barnett, S.; Belmont, J.W.; et al. Best Practices for the Interpretation and Reporting of Clinical Whole Genome Sequencing. *NPJ Genom. Med.* **2022**, *7*, 27. [CrossRef]
24. Chen, Z.; Morris, H.R.; Polke, J.; Wood, N.W.; Gandhi, S.; Ryten, M.; Houlden, H.; Tucci, A. Repeat Expansion Disorders. *Pract. Neurol.* **2024**, 1–15. [CrossRef] [PubMed]
25. Genovese, L.M.; Geraci, F.; Corrado, L.; Mangano, E.; D’Aurizio, R.; Bordoni, R.; Severgnini, M.; Manzini, G.; De Bellis, G.; D’Alfonso, S.; et al. A Census of Tandemly Repeated Polymorphic Loci in Genic Regions Through the Comparative Integration of Human Genome Assemblies. *Front. Genet.* **2018**, *9*, 155. [CrossRef] [PubMed]
26. Fan, H.; Chu, J.Y. A Brief Review of Short Tandem Repeat Mutation. *Genom. Proteom. Bioinform.* **2007**, *5*, 7. [CrossRef]
27. Ziaei Jam, H.; Li, Y.; DeVito, R.; Mousavi, N.; Ma, N.; Lujumba, I.; Adam, Y.; Maksimov, M.; Huang, B.; Dolzhenko, E.; et al. A Deep Population Reference Panel of Tandem Repeat Variation. *Nat. Commun.* **2023**, *14*, 6711. [CrossRef]

28. Cui, Y.; Ye, W.; Li, J.S.; Li, J.J.; Vilain, E.; Sallam, T.; Li, W. A Genome-Wide Spectrum of Tandem Repeat Expansions in 338,963 Humans. *Cell* **2024**, *187*, 2336–2341.e5. [[CrossRef](#)]
29. Laß, J.; Thomsen, M.; Borsche, M.; Lüth, T.; Prietzsche, J.C.; Schaaake, S.; Milovanović, A.; Macpherson, H.; Gustavsson, E.K.; Awad, P.S.; et al. FGF14 Repeat Length and Mosaic Interruptions: Modifiers of SCA27b? *medRxiv* **2024**. [[CrossRef](#)]
30. Sehgal, A.; Jam, H.Z.; Shen, A.; Gymrek, M. Genome-Wide Detection of Somatic Mosaicism at Short Tandem Repeats. *Bioinformatics* **2024**, *40*, btae485. [[CrossRef](#)]
31. Khristich, A.N.; Mirkin, S.M. On the Wrong DNA Track: Molecular Mechanisms of Repeat-Mediated Genome Instability. *J. Biol. Chem.* **2020**, *295*, 4134–4170. [[CrossRef](#)]
32. Handsaker, R.E.; Kashin, S.; Reed, N.M.; Tan, S.; Lee, W.-S.; McDonald, T.M.; Morris, K.; Kamitaki, N.; Mullally, C.D.; Morakabati, N.R.; et al. Long Somatic DNA-Repeat Expansion Drives Neurodegeneration in Huntington’s Disease. *Cell* **2025**, *188*, 623–639.e19. [[CrossRef](#)]
33. Rasmussen, A.; Hildonen, M.; Vissing, J.; Duno, M.; Tümer, Z.; Birkedal, U. High Resolution Analysis of DMPK Hypermethylation and Repeat Interruptions in Myotonic Dystrophy Type 1. *Genes* **2022**, *13*, 970. [[CrossRef](#)] [[PubMed](#)]
34. Pešović, J.; Perić, S.; Brkušanić, M.; Brajušković, G.; Rakočević, V.; Savić-Pavićević, D. Repeat Interruptions Modify Age at Onset in Myotonic Dystrophy Type 1 by Stabilizing DMPK Expansions in Somatic Cells. *Front. Genet.* **2018**, *9*, 601. [[CrossRef](#)] [[PubMed](#)]
35. Morato Torres, C.A.; Zafar, F.; Tsai, Y.C.; Vazquez, J.P.; Gallagher, M.D.; McLaughlin, I.; Hong, K.; Lai, J.; Lee, J.; Chirino-Perez, A.; et al. ATTCT and ATTCC Repeat Expansions in the ATXN10 Gene Affect Disease Penetrance of Spinocerebellar Ataxia Type 10. *Hum. Genet. Genom. Adv.* **2022**, *3*, 100137. [[CrossRef](#)]
36. 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.; et al. Detection of Long Repeat Expansions from PCR-Free Whole-Genome Sequence Data. *Genome Res.* **2017**, *27*, 1895–1903. [[CrossRef](#)] [[PubMed](#)]
37. Rajan-Babu, I.S.; Dolzhenko, E.; Eberle, M.A.; Friedman, J.M. Sequence Composition Changes in Short Tandem Repeats: Heterogeneity, Detection, Mechanisms and Clinical Implications. *Nat. Rev. Genet.* **2024**, *25*, 476–499. [[CrossRef](#)]
38. Mangin, A.; de Pontual, L.; Tsai, Y.C.; Monteil, L.; Nizon, M.; Boisseau, P.; Mercier, S.; Ziegler, J.; Harting, J.; Heiner, C.; et al. Robust Detection of Somatic Mosaicism and Repeat Interruptions by Long-Read Targeted Sequencing in Myotonic Dystrophy Type 1. *Int. J. Mol. Sci.* **2021**, *22*, 2616. [[CrossRef](#)]
39. Sullivan, R.; Chen, S.; Saunders, C.T.; Yau, W.Y.; Goh, Y.Y.; O’Connor, E.; Dominik, N.; Deforie, V.G.; Morsy, H.; Cortese, A.; et al. RFC1 Repeat Expansion Analysis from Whole Genome Sequencing Data Simplifies Screening and Increases Diagnostic Rates. *medRxiv* **2024**. [[CrossRef](#)]
40. Nethisinghe, S.; Kesavan, M.; Ging, H.; Labrum, R.; Polke, J.M.; Islam, S.; Garcia-moreno, H.; Callaghan, M.F.; Cavalcanti, F.; Pook, M.A.; et al. Interruptions of the Fxn Gaa Repeat Tract Delay the Age at Onset of Friedreich’s Ataxia in a Location Dependent Manner. *Int. J. Mol. Sci.* **2021**, *22*, 7507. [[CrossRef](#)]
41. Kaplun, L.; Krautz-Peterson, G.; Neerman, N.; Stanley, C.; Hussey, S.; Folwick, M.; McGarry, A.; Weiss, S.; Kaplun, A. ONT Long-Read WGS for Variant Discovery and Orthogonal Confirmation of Short Read WGS Derived Genetic Variants in Clinical Genetic Testing. *Front. Genet.* **2023**, *14*, 1145285. [[CrossRef](#)]
42. Vegezzi, E.; Facchini, S.; Bragg, D.C.; Sharma, N.; Cortese, A.; Vegezzi, E.; Ishiura, H.; Cristopher Bragg, D.; Pellerin, D.; Magrinelli, F.; et al. Neurological Disorders Caused by Novel Non-Coding Repeat Expansions: Clinical Features and Differential Diagnosis. *Lancet Neurol.* **2024**, *23*, 725–764. [[CrossRef](#)] [[PubMed](#)]
43. Pellerin, D.; Wilke, C.; Träschütz, A.; Nagy, S.; Currò, R.; Dicaire, M.J.; Garcia-Moreno, H.; Anheim, M.; Wirth, T.; Faber, J.; et al. Intronic FGF14 GAA Repeat Expansions Are a Common Cause of Ataxia Syndromes with Neuropathy and Bilateral Vestibulopathy. *J. Neurol. Neurosurg. Psychiatry* **2023**, *95*, 175. [[CrossRef](#)]
44. Ibañez, K.; Polke, J.; Hagelstrom, R.T.; Dolzhenko, E.; Pasko, D.; Thomas, E.R.A.; Daugherty, L.C.; Kasperaviciute, D.; Smith, K.R.; Deans, Z.C.; et al. 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*, 234–245. [[CrossRef](#)]
45. Roy, S.; Coldren, C.; Karunamurthy, A.; Kip, N.S.; Klee, E.W.; Lincoln, S.E.; Leon, A.; Pullambhatla, M.; Temple-Smolkin, R.L.; Voelkerding, K.V.; et al. Standards and Guidelines for Validating Next-Generation Sequencing Bioinformatics Pipelines: A Joint Recommendation of the Association for Molecular Pathology and the College of American Pathologists. *J. Mol. Diagn.* **2018**, *20*, 4–27. [[CrossRef](#)] [[PubMed](#)]
46. Scriba, C.K.; Stevanovski, I.; Chintalaphani, S.R.; Gamaarachchi, H.; Ghaoui, R.; Ghia, D.; Henderson, R.D.; Jordan, N.; Winkel, A.; Lamont, P.J.; et al. RFC1 in an Australasian Neurological Disease Cohort: Extending the Genetic Heterogeneity and Implications for Diagnostics. *Brain Commun.* **2023**, *5*, fcad208. [[CrossRef](#)] [[PubMed](#)]
47. Dominik, N.; Magri, S.; Currò, R.; Abati, E.; Facchini, S.; Corbetta, M.; MacPherson, H.; Di Bella, D.; Sarto, E.; Stevanovski, I.; et al. Normal and Pathogenic Variation of RFC1 Repeat Expansions: Implications for Clinical Diagnosis. *Brain* **2023**, *146*, 5060–5069. [[CrossRef](#)]

48. Nobile, V.; Pucci, C.; Chiurazzi, P.; Neri, G.; Tabolacci, E. DNA Methylation, Mechanisms of FMR1 Inactivation and Therapeutic Perspectives for Fragile X Syndrome. *Biomolecules* **2021**, *11*, 296. [CrossRef]
49. Hunter, J.E.; Berry-Kravis, E.; Hipp, H.; Todd, P.K. FMR1 Disorders. In *GeneReviews*®; University of Washington: Seattle, WA, USA, 2024. Available online: <https://www.ncbi.nlm.nih.gov/books/NBK1384/> (accessed on 10 February 2025).
50. Wallenius, J.; Kafantari, E.; Jhaveri, E.; Gorcenco, S.; Ameer, A.; Karremo, C.; Dobloug, S.; Karrman, K.; de Koning, T.; Ilinca, A.; et al. Exonic Trinucleotide Repeat Expansions in ZFH3 Cause Spinocerebellar Ataxia Type 4: A Poly-Glycine Disease. *Am. J. Hum. Genet.* **2023**, *111*, 82. [CrossRef]
51. Neerman, N.; Faust, G.; Meeks, N.; Modai, S.; Kalfon, L.; Falik-Zaccai, T.; Kaplun, A. A Clinically Validated Whole Genome Pipeline for Structural Variant Detection and Analysis. *BMC Genom.* **2019**, *20*, 1–8. [CrossRef]
52. Variantx Genomic Unity® Test A Whole Genome Clinical Validation Study. 2019. Available online: <https://www.variantx.com/wp-content/uploads/2022/04/Genomic-Unity-Clinical-Validation-Study.pdf> (accessed on 2 February 2025).
53. De Coster, W.; Höijer, I.; Bruggeman, I.; D'Hert, S.; Melin, M.; Ameer, A.; Rademakers, R. Medically Relevant Tandem Repeats in Nanopore Sequencing of Control Cohorts. *medRxiv* **2024**. [CrossRef]
54. Read, J.L.; Davies, K.C.; Thompson, G.C.; Delatycki, M.B.; Lockhart, P.J. Challenges Facing Repeat Expansion Identification, Characterisation, and the Pathway to Discovery. *Emerg. Top. Life Sci.* **2023**, *7*, 339. [CrossRef] [PubMed]
55. Friedreich Ataxia (FRDA) via the FXN GAA Repeat Expansion Test—PreventionGenetics. Available online: [https://www.preventiongenetics.com/testInfo?val=Friedreich-Ataxia-\(FRDA\)-via-the-FXN-GAA-Repeat-Expansion](https://www.preventiongenetics.com/testInfo?val=Friedreich-Ataxia-(FRDA)-via-the-FXN-GAA-Repeat-Expansion) (accessed on 6 March 2025).
56. Seifert, B.A.; Reddi, H.V.; Kang, B.E.; Bean, L.J.H.; Shealy, A.; Rose, N.C. Myotonic Dystrophy Type 1 Testing, 2024 Revision: A Technical Standard of the American College of Medical Genetics and Genomics (ACMG). *Genet. Med.* **2024**, *26*, 101145. [CrossRef] [PubMed]
57. 620114: SCA1 (ATXN1) Genetic Testing (Repeat Expansion) | Labcorp Labcorp. Available online: <https://www.labcorp.com/tests/620114/sca1-atxn1-genetic-testing-repeat-expansion> (accessed on 17 March 2025).
58. Rudaks, L.I.; Yeow, D.; Ng, K.; Deveson, I.W.; Kennerson, M.L.; Kumar, K.R. An Update on the Adult-Onset Hereditary Cerebellar Ataxias: Novel Genetic Causes and New Diagnostic Approaches. *Cerebellum* **2024**, *23*, 2152–2168. [CrossRef] [PubMed]
59. PacBio—Sequence with Confidence. Available online: <https://www.pacb.com/> (accessed on 3 March 2025).
60. Welcome to Oxford Nanopore Technologies. Available online: <https://nanoporetech.com/> (accessed on 3 March 2025).
61. Oehler, J.B.; Wright, H.; Stark, Z.; Mallett, A.J.; Schmitz, U. The Application of Long-Read Sequencing in Clinical Settings. *Hum. Genom.* **2023**, *17*, 73. [CrossRef]
62. Illumina Complete Long Reads Portfolio. Available online: <https://www.illumina.com/products/by-brand/complete-long-reads-portfolio.html> (accessed on 3 March 2025).
63. Genome Reference Consortium Human Build 38—NCBI. Available online: https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.26/ (accessed on 5 January 2023).
64. GnomAD. Available online: <https://gnomad.broadinstitute.org/> (accessed on 3 February 2025).
65. Halman, A.; Dolzhenko, E.; Oshlack, A. STRipy: A Graphical Application for Enhanced Genotyping of Pathogenic Short Tandem Repeats in Sequencing Data. *Hum. Mutat.* **2022**, *43*, 859–868. [CrossRef]

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.