The Journal of Molecular Diagnostics, Vol. ■, No. ■, ■ 2026

ELSEVIER

the Journal of Molecular Diagnostics

jmdjournal.org

Parallel Analysis of Repeat Expansions

An Updated Clinical Nanopore Cas9-Targeted Sequencing Workflow for Nanopore R10 Flow Cells

Veronika Scholz,* Veronika Schönrock,* Hannes Erdmann,*† Vitus Prokosch,* Maria Schoedel,* Manuela Almus,* Mayra Sauer,* Veronika Mayer,* Eva Breithausen,* Inga van Buren,* Christine Dineiger,* Madeline Golibrzuch,* Minerva Montero-Hernández,*‡ Annette Lischka,§ Katja Eggermann,§ Caroline Heintz,* Ariane Hallermayr,*¶ Teresa Neuhann,* Elke Holinski-Feder,*¶ Angela Abicht,*† Anna Benet-Pagès,*‖ and Morghan C. Lucas*†¶

From the MGZ—Medical Genetics Center, Munich, Germany; the Friedrich-Baur-Institute,† Department of Neurology, Klinikum der Universität, Ludwig-Maximilians-Universität, Munich, Germany; the Laboratorio de Genética Médica,‡ Instituto de Biomedicina, Facultad de Medicina de Albacete, Universidad de Castilla—La Mancha, Spain; the Center for Human Genetics and Genomic Medicine,§ Faculty of Medicine, RWTH Aachen University, Aachen, Germany; the Medizinische Klinik und Poliklinik IV,¶ Klinikum der Universität München Campus Innenstadt, Munich, Germany; and the Institute of Neurogenomics,‖ Helmholtz Zentrum München GmbH—German Research Center for Environmental Health, Munich, Germany*

Accepted for publication April 17, 2026.

Address correspondence to Morghan Lucas. E-mail: morghan.lucas@mgz-muenchen.de.

Hereditary ataxias, caused by expansions of short tandem repeats, are difficult to diagnose using traditional PCR and Southern blot methods, which struggle to detect complex repeat expansions and cannot assess repeat interruptions or methylation. An updated Clinical Nanopore Cas9-Targeted Sequencing workflow is presented for analyzing repeat expansions, now compatible with the Oxford Nanopore Technologies R10 flow cell. The workflow incorporates the Oxford Nanopore Technologies wf-human-variation Epi2Me workflow, including the Straglr tool to analyze base-called reads, ensuring compatibility with past, current, and future sequencing chemistries. It expands the number of genes analyzed from 10 to 27 and introduces new gene panels for ataxia, myopathy, neurodegeneration, and amyotrophic lateral sclerosis/motor neuron disease. Validated with Coriell reference and clinical samples, this method improves the analysis of pathogenic repeat expansions, providing deeper insights into repeat structures while addressing the limitations of traditional approaches. In this work, the use of multiplexing, Flongle flow cells, and single-gene targeting were explored as alternatives to panel-based approaches in the Clinical Nanopore Cas9-Targeted Sequencing workflow, finding that only single-gene targeting provides compatibility and reliable performance. *(J Mol Diagn 2026, ■: 1—14; https://doi.org/10.1016/j.jmoldx.2026.04.007)*

Hereditary ataxias are a heterogeneous group of neurodegenerative disorders primarily characterized by progressive motor coordination loss, gait disturbances, and various neurologic symptoms, such as ocular motor dysfunction, dysarthria, and sensory neuropathy.[1,2] A considerable proportion of these disorders are linked to the expansion of short tandem repeats (STRs), specific sequences of DNA repeated in both coding and noncoding regions of the genome. These repeat expansions interfere with normal gene function and are responsible for several neurodegenerative diseases with varying levels of severity.[3,4] To date, approximately 60 Mendelian disorders are known to be caused by such expansions, including prominent examples like most spinocerebellar ataxias (SCAs), Friedreich ataxia, and fragile X—associated tremor/ataxia syndrome.[3,4]

Traditionally, diagnosing these repeat expansion disorders has relied on PCR-based methods and Southern blotting. Although these methods have been the gold standard, they face notable limitations, especially in detecting large, complex repeat expansions. They are also unable to capture

V.Schol. and V.Schön contributed equally to this work.

© 2026 The Author(s). Published by Elsevier Inc. on behalf of the Association for Molecular Pathology and American Society for Investigative Pathology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0).

https://doi.org/10.1016/j.jmoldx.2026.04.007

information on alternative repeat motifs, repeat interruptions, or epigenetic modifications, such as DNA methylation, which are critical for accurate diagnosis and prognosis.[2,5] Furthermore, these methods require iterative testing of specific loci, making them both time-consuming and labor intensive.

With the development of long-read sequencing technologies, particularly those provided by Oxford Nanopore Technologies (ONT; Oxford, UK), there has been a significant leap forward in the parallel and comprehensive analysis of complex repeat expansions.[4,6,7] Previous work introduced Clinical Nanopore Cas9-Targeted Sequencing (Clin-CATS), a method that combined ONT's R9 chemistry with CRISPR-Cas9–based enrichment to analyze 10 repeat loci associated with the most common hereditary ataxias in the European population at that time.[8] Clin-CATS enabled precise quantification of repeat length and sequence composition of key genes associated with adult-onset ataxias, including *ATXN1*, *ATXN2*, *ATXN3*, *ATXN7*, *ATXN8*/*ATXN8OS*, *CACNA1A*, *FMR1*, *FXN*, *TBP*, and *RFC1*.[9–11]

With the recent introduction of ONT's R10 flow cell chemistry, offering improved accuracy in resolving homopolymer regions and complex repeat structures, this study has adapted and updated the Clin-CATS workflow to leverage the advantages of R10 sequencing. The R10 chemistry offers significant improvements in sequencing accuracy and resolving complex genomic structures, particularly for regions with challenging repeat compositions, such as *RFC1*. These enhancements enable more precise determination of repeat sizes and composition, which is crucial for accurate diagnosis and understanding of the genetic underpinnings of ataxia phenotypes. However, despite these technological advances, ONT's Cas9-targeted sequencing has been discontinued and has not been re-released for use with the R10 flow cell chemistry. This development required us to modify the original protocol to maintain the high precision and diagnostic utility of Clin-CATS.

This study describes modifications to the Clin-CATS library preparation protocol and bioinformatic pipeline necessary to accommodate the changes introduced by the R10 flow cell chemistry. A key improvement in the bioinformatic workflow is the use of the Straglr tool,[12] which was integrated into the ONT wf-human-variation pipeline. Unlike previous tools that relied on the raw electrical signal from the sequencer and needed to be calibrated to the specific flow cell, Straglr processes data from base-called reads. This approach is crucial, as the raw signal will likely change with future flow cell chemistry updates. By performing the analysis on base-called data, this bioinformatic pipeline can be consistently applied across old, current, and future flow cell chemistries, ensuring long-term adaptability—a requirement for accredited diagnostic workflows.

In addition to these workflow improvements, this study has expanded the number of genes analyzed with Clin-CATS from 10 to 27. This update supplements the ataxia panel and introduces three additional gene panels, each tailored to a specific group of related disorders: a myopathy panel, a neurodegeneration panel, and an amyotrophic lateral sclerosis (ALS)/motor neuron disease panel. These panels enable the simultaneous investigation of multiple repeat expansion-related genes across various neurologic conditions, further increasing the diagnostic utility of the Clin-CATS workflow. In this work, the use of multiplexing, Flongle flow cells, and single-gene targeting were also explored as alternatives to panel-based approaches, finding that only single-gene targeting provides reliable coverage and is compatible with the Clin-CATS workflow.

This study demonstrates the capability of the updated Clin-CATS workflow in analyzing pathogenic repeat expansions in a cohort of patients with ataxia, providing deeper insights into repeat structures and their clinical relevance. Moreover, Coriell reference samples were used alongside clinical samples, ensuring the robustness and accuracy of this approach. This updated method continues to offer a comprehensive and scalable approach for diagnosing and understanding adult-onset ataxias while addressing the limitations of traditional PCR-based and Southern blot methods.

Materials and Methods

Guide RNA Design

New guides not included in the original Erdmann et al[8] publication were designed using CHOPCHOP version 3.0.0. All guides were designed using the hg38 *Homo sapiens* reference and have a length of 20 bp. The resulting region of interest flanked by the guide RNAs should be 3 to 12 kb. Multiple guides upstream and downstream with the highest efficiency that were not self-complementary or complementary to any other site by a minimum of two mismatches in the list provided by CHOPCHOP were selected. The strand location of the upstream (and downstream) guide was on either the plus or minus strand, but always opposite of the other CRISPR RNA, depending on the orientation of the Cas9 needed. The most efficient pair was selected, while pairs with a guide <45% efficiency were discarded. The guide RNA sequences for Clin-CATS are provided in Supplemental Table S1. Guide RNAs were combined in equal proportions to yield a guide mix for all targets, with the exception of those targeting the repeat regions of specific genes. For *AR* and *ATN1*, the guide RNAs were added at six times the standard volume. For *ATXN10* and *ATXN2*, the guide RNAs were added at four times the standard volume. Additionally, guide RNAs for *ATXN3*, *ATXN7*, and *ATXN8* were included at twice the standard volume. For multiplexing tests, specific *HBB* guide RNAs were selected to target a 400- to 600-bp region of the *HBB* gene, ensuring efficient enrichment of fragments within this size range. As before, these guides targeted both upstream and downstream of the desired regions. The *HBB* guide RNA

UNCORRECTED PROOF

FLA 5.7.0 DTD ■ JMDI1677_proof ■ 17 September 2026 ■ EO: JMDI-D-25-00450

sequences selected were as follows: protospacer adjacent motif (PAM)-in upstream with the sequence 5′ TTATGTTTCACATATGAGCA-3′, PAM-in downstream with the sequence 5′-GAAGAATTTTATTATATGAG-3′, PAM-out upstream with the sequence 5′-AAACTGGGG-GATATTATGAA-3′, and PAM-out downstream with the sequence 5′-TTCTTTAGAATGGTGCAAAG-3′. All guide RNAs were purchased from Integrated DNA Technologies (IDT; Coralville, IA).

Patient Samples and Study Approval

Informed consent to participate in this study was obtained from all patients and was approved by local institutions (Bayerische Landesärztekammer, 2019-210). All genetic analyses and investigations were performed in accordance with the guidelines of the Declaration of Helsinki.

Extraction of Genomic DNA

Genomic DNA was obtained from total peripheral EDTA blood samples by extraction of white blood cells with a Biomek FX or i7system (Beckman Coulter, Brea, CA) using the NucleoMag Blood 3 mL Kit (Machery-Nagel, 744502.1) as per the manufacturer's instructions. The DNA concentration was measured using a Qubit 4 fluorometer (Thermo Fisher Scientific, Waltham, MA; Q33238) and the Qubit dsDNA HS Assay-Kit (Thermo Fisher Scientific; Q33233) or Qubit dsDNA BR Assay-Kit (Thermo Fisher Scientific; Q32853) as per manufacturer's instructions. DNA purity (A260/A280 and A260/A230) was measured with the BigLunatic device (Unchained Labs) as per manufacturer's instructions. The DNA integrity number was measured on the TapeStation 4200 (Agilent Technologies, Santa Clara, CA; G2991BA) using the Genomic DNA Screen Tape (Agilent Technologies; 5067-5365) and the Genomic DNA Reagents (Agilent Technologies; 5067-5366) as per manufacturer's instructions.

DNA Clean-Up Using Rebuffured Ampure XP Beads

The 2 mL AMPure XP beads (Beckman Coulter; A63882) were mixed thoroughly by vortexing and pelleted on a DynaMag-2 Magnet magnetic rack (Thermo Fisher Scientific; 12321D), and the supernatant was discarded. The beads were washed with nuclease-free water, resuspended by vortexing, and repelleted, and the supernatant was removed. The washing step was repeated, and the beads were air dried. A custom buffer consisting of 10 mmol/L Tris-HCl (Thermo Fisher Scientific; J22638.AE), 1 mmol/L EDTA pH 8.0 (Merck, Rahway, NJ; 324506), 1.6 mol/L NaCl (Merck; 59222C), and 11% PEG 8000 (Sigma Aldrich, St. Louis, MO; SIAL-P1458-25 ML) was added to the beads and mixed by vortexing. The rebuffered Ampure XP beads were stored at 4°C until further use for up to 1 week.

DNA samples were diluted to a minimum of 40 ng/μL and cleaned up using 0.65× rebuffered beads, mixed by flicking, placed into a Thermomixer (Eppendorf, Hamburg, Germany; 5382000015) at a speed of 450 rpm, with manual inversion every 2 minutes, for 10 minutes. The beads were pelleted, the supernatant was discarded, and the samples were washed with 200 μL of freshly prepared 70% ethanol without disturbing the pellet; and the supernatant was again removed after 30 seconds. The ethanol washing step was repeated, and the tube was spun down and placed on the magnetic rack, with excess ethanol removed. The samples were air dried, then resuspended in 26 μL of 1× TE Buffer (PanReac AppliChem, A8569,1000) by flicking and incubating at 50°C for 1 minute. The sample was then incubated at room temperature for an additional 5 minutes before repelleting it on the magnet. The eluate was then removed and transferred into a 0.2-mL PCR tube (Sarstedt, 72.990.002) or 1.5 mL DNA LoBind tube (Eppendorf; 0030108051).

Coriell Samples

Human genomic DNA samples were obtained from the Coriell Institute for Medical Research (Camden, NJ), as detailed in Supplemental Table S2. DNA concentration, purity, and DNA integrity number were measured as described above on receipt.

Library Preparation and ONT Cas9-Targeted Sequencing

The library preparation was performed as in Erdmann et al[8] for R9 (R9.4.1) MinION flow cells, with some adaptations for R10 (R10.4.1) MinION flow cells. On arrival, guide RNAs (CRISPR RNA, IDT) (Supplemental Table S1) and trans-activating CRISPR RNAs (IDT; 1072533) were resuspended in TE-Buffer of pH 7.5 (IDT; 11-05-01-05) to 100 μmol/L and stored at −80°C until further use. To produce the genomic RNA complex, 1 μL of each CRISPR RNA pool was mixed with 1 μL of trans-activating CRISPR RNA in 8 μL of Duplex buffer (IDT; 11-01-03-01), annealed by heating at 95°C for 5 minutes, and cooled to room temperature. Next, 79.2 μL of nuclease-free water, 10 μL of rCutSmart Buffer [New England Biolabs (NEB), B6004], and 0.8 μL of Alt-R S.p. HiFi Cas9 Nuclease V3 (IDT; 1081060) were added to the annealed RNAs, forming the ribonucleoprotein complexes, which were incubated at room temperature for 30 minutes. These ribonucleoproteins were then placed on ice for further use. In parallel, 5 μg of input genomic DNA in 24 μL nuclease-free water was dephosphorylated by incubation with 3 μL Quick CIP Phosphatase (NEB; M0525) and 3 μL rCutSmart Buffer at 37°C for 20 minutes (adapted from the original protocol, which specifies a 10-minute incubation), followed by enzyme deactivation at 80°C for 2 minutes and kept at room temperature until further use. To the dephosphorylated

UNCORRECTED PROOF

FLA 5.7.0 DTD ■ JMDI1677_proof ■ 17 September 2026 ■ EO: JMDI-D-25-00450

DNA, 10 μL of the ribonucleoprotein complex, 1 μL of dATP (NEB; N0440), and 1 μL Taq DNA Polymerase (NEB; M0273) were added. The reaction mixture was incubated at 37°C for 15 minutes and 5 minutes at 72°C. The mix was then cooled to 4°C and stored on ice until further use. Adapter ligation was performed by combining 20 μL of ligation buffer (ONT; SQK-LSK114 kit), 3 μL of nuclease-free water, 10 μL NEBNext Quick T4 DNA ligase (NEB; E6056), and 5 μL adapter mix (ONT; SQK-LSK114 kit). Immediately after mixing, the adapter ligation mix was added to the cleaved and dA-tailed sample to yield an 80-μL final reaction volume. The resulting mixture was incubated at room temperature for 30 minutes (adapted from the original protocol, which specifies a 10-minute incubation), and 1 volume 1× TE pH 8.0 (IDT; 11-01-02-05) was added to stop the reaction. DNA was purified using 48 μL AMPure XP beads (ONT; SQK-LSK114 kit) and Long Fragment Buffer (ONT; SQK-LSK114 kit) according to the manufacturer's protocol, and the sample was eluted in 13 μL Elution Buffer (ONT; SQK-LSK114 kit). To prime the flow cells, 1 mL Flow Cell Flush (ONT; SQK-LSK114 kit) was combined with 25 μL Flow Cell Tether (ONT; SQK-LSK114 kit), and 800 μL of the mix was slowly inserted into the Priming Port after removing any air bubbles. After a 5-minute incubation, the SpotON Sample Port was opened, and another 200 μL of the mix was added via the Priming Port. A total of 12 μL of the library was mixed with 37.5 μL Sequencing Buffer (ONT; SQK-LSK114 kit) and 25.5 μL Library Beads (ONT; SQK-LSK114 kit) before loading in a dropwise manner onto the SpotON sample port of the primed flow cell (ONT; FLO-MIN114) for sequencing on the GridION Mk1 sequencer. The run was started after incubating the library on the flow cell for 10 minutes and placing the light shield on the sample port and sensor array. Ligation Sequencing Kit LSK114 and the Cas9 Sequencing Expansion were chosen to start the run, and the run limit was set to 30 hours.

Multiplexing Tests

Exonuclease and Proteinase K Treatments

The genomic RNA complexes, ribonucleoprotein preparation, dephosphorylation, and Cas9 cleavage steps were performed as above with an input mass of 4 μg as recommended by ONT to be sequenced on R10 (R10.3 flowcells). When no dephosphorylation was performed in the assay, 3 μL of rCutSmart Buffer (NEB; B6004S) and 3 μL nuclease-free water were added before the Cas9 digestion. For exonuclease treatment with Lambda Exonuclease (NEB; M0262S), 5 μL of Lambda Exonuclease Reaction Buffer (NEB; B0262S) was mixed with 1 μL of nuclease-free water. For the combination of Lambda Exonuclease and Exonuclease I (NEB; M0263S), 5 μL of Exonuclease I Reaction Buffer (NEB; B0293S) was mixed with 1 μL 1:200 diluted UltraPure BSA solution (Thermo Fisher Scientific; AM2616). To the samples, 4 μL of Lambda Exonuclease and 4 μL of Lambda Exonuclease with 1 μL of Exonuclease I were added after the addition of 6 μL custom buffer mix. After flicking the tube and centrifugation, the reaction was heated for 30 minutes at 37°C, then heated to 80°C for 5 minutes to deactivate the exonucleases. Samples treated with Thermolabile Proteinase K (NEB; P8111S) had one 1 μL of the protein added, were flicked, spun down, and heated to 37°C for 15 minutes and 80°C for 4 minutes to deactivate the enzyme. After the reaction, the volume was adjusted to 50 μL with 9 μL of nuclease-free water. Samples not treated with exonucleases and Proteinase K were heat treated at 80°C for 5 minutes to denature Cas9, then brought to a volume of 50 μL by adding 10 μL nuclease-free water and flicked. The dA-tailing was performed as in the library preparation above, with 1.5 μL of dATP (NEB; N0440) and 1.5 μL Taq DNA Polymerase (NEB; M0273) to match the increased volume but with the same temperature profile. A total of 52.5 μL of AMPure XP beads (ONT; SQK-NBD114.96) was added at a 1× ratio, and the samples were cleaned twice with 250 μL 80% ethanol without resuspending and eluted in 7.5 μL nuclease-free water.

Multiplexing Samples with the Ataxia Panel

Samples were prepared as above with 4 μg DNA input, the protocol performed as before without exonuclease digestion, Proteinase K digestion, and dephosphorylation of the samples, and the dA-tailing performed as described before in the library preparation protocol. Additionally, the volume was not adjusted, and the clean-up was subsequently done with 42 μL of AMPure XP (Beckman Coulter; A63882) beads, 80% of ethanol, and eluted in 7.5 μL of nuclease-free water into a 0.2-mL PCR tube (Sarstedt; 72.990.002). Afterward, the protocol was continued with the ligation of barcodes.

Ligation of Barcodes for Multiplexing Tests

A total of 7.5 μL of the samples was mixed with 2.5 μL unique Native Barcode (NB01-96) from the Native Barcoding kit (ONT; SQK-NBD114.96) and 10 μL Blunt/TA Ligase Master Mix (NEB; M0367), and the sample mix was flicked and spun down. After incubation for 20 minutes at room temperature, 2.0 μL of EDTA from the Native Barcoding kit (ONT; SQK-NBD114.96) was added, and the solution was mixed by flicking. Samples to be run together were then pooled in a 1.5-mL DNA LoBind tube (Eppendorf; 0030108051). Then AMPure XP beads (ONT; SQK-NBD114.96) were mixed on the vortex, and 9 μL per sample in the pool of the beads was added at a 0.4× ratio and incubated on a Thermomixer (Eppendorf; 5382000015) for 10 minutes at 450 rpm at room temperature. Samples were pelleted on a DynaMag-2 Magnet magnetic rack (Thermo Fisher Scientific; 12321D) until the solution was clear. The supernatant was discarded, 250 μL Short

UNCORRECTED PROOF

Fragment Buffer (ONT; SQK-NBD114.96) was added, and the pellet was resuspended and spun down. The washing step was repeated, and the samples were spun down and again placed on the magnetic rack to remove excess Short Fragment Buffer, and allowed to air dry for 30 seconds, then removed from the magnet, resuspended in 35-μL nuclease-free water, and were incubated for 10 minutes at 37°C on the Thermomixer (Eppendorf; 5382000015) at 450 rpm. The samples were again pelleted for at least 1 minute on the magnetic rack until the solution was clear and colorless, and 34 μL of the eluate was transferred to a 1.5-mL DNA LoBind tube (Eppendorf; 0030108051). Samples were optionally stored at this point at 4°C overnight.

Ligation of Adapters, Priming, and Loading for Multiplexing Tests

To the now pooled samples, 10 μL 5× NEBNext Quick Ligation Buffer, 5 μL NEB T4 DNA Ligase from the NEB Quick Ligation Module (NEB; E6056), and 1.0 μL Native Adapter ONT; SQK-NBD114.24) were added, the tubes were flicked and spun down, then incubated for 20 minutes at room temperature. AMPure XP beads (ONT; SQK-NBD114.96) were vortexed, and 20 μL per sample was added at a 0.4× ratio and incubated on a Thermomixer (Eppendorf; 5382000015) for 10 minutes at 450 rpm at room temperature. Samples were pelleted on a DynaMag-2 Magnet magnetic rack (Thermo Fisher Scientific; 12321D) until the solution was clear. The supernatant was discarded, and 250 μL Long Fragment Buffer (ONT; SQK-NBD114.24)/Short Fragment Buffer (ONT; SQK-NBD114.24) was added, the samples were removed from the magnet, and the pellet was resuspended and spun down. The washing step was repeated, samples were spun down, and excess Long Fragment Buffer was removed; the samples were allowed to air dry for 30 seconds, then removed from the magnet, resuspended in 15 μL Elution Buffer (ONT; SQK-NBD114.24), and incubated for 10 minutes at 37°C on the Thermomixer (Eppendorf; 5382000015) at 450 rpm. The samples were again pelleted for at least 1 minute on the magnetic rack until the solution was clear and colorless, and 12 μL eluate was transferred to a 0.5 mL DNA LoBind tube (Eppendorf; 0030108035).

This library was mixed with 37.5 μL Sequencing Buffer and 25.5 μL Library Beads in the tube and stored on ice until applied in the flow cell. Then, 1170 μL of the Flow Cell Flush was mixed with 30 μL Flow Cell Tether (ONT; SQK-NBD114.24) and 5 μL UltraPure BSA solution (Thermo Fisher; AM2616). After ensuring no air bubbles were present in the priming port of the flow cell by extracting a small volume, 800 μL of the above mix was gently loaded into the priming port of the R10 GridION flow cell (ONT; FLO-MIN114), incubated for 5 minutes, and then the remaining 200 μL of priming buffer was loaded into the flow cell after the sample spot port had been opened. Afterward, 75 μL of the sample was added dropwise into the sample port, and the sequencing run started after 5 minutes. The Native Barcoding Kit 96 was selected, a sequencing duration of 12 to 72 hours was selected depending on the experiment, and the run was started after placing the light shield on the sample port. The sequencing time was 30, 48, or 72 hours for the ATX panel and 12 hours for the multiplexing tests.

Reloading Flow Cells

To reuse flow cells after sequencing, the Flow Cell Wash Kit (ONT; EXP-WSH004) was used as per the manufacturer's instructions. Briefly, 2 μL of the Wash Mix was mixed with 398 μL of the Diluent (ONT; EXP-WSH004) and placed on ice. For rerunning the library, the sequencing run was paused, and the old library was removed by extracting 150 μL volume from the priming port and storing it in a 0.5 mL DNA LoBind tube (Eppendorf; 0030108035). After ensuring no air bubbles were present in the sample port, 200 μL of the prepared mix was added to the flow cell through the priming port and incubated for 5 minutes. The rest of the mix was loaded into the flow cell and incubated for 1 hour. The flow cell was then either filled with 500 μL Storage Buffer (ONT; EXP-WSH004) and stored at 4°C for further use or primed again using the flow cell priming mix, and 150 μL of the extracted library was reloaded into the sample port.

Agarose Gels

A total of 1000 to 4000 ng of DNA was combined with a running buffer of 40% sucrose in 100 mL of distilled water with Orange G. The 1% agarose gels were prepared by combining 1% w/v Agarose Type II-A (Sigma Aldrich; A9918-250 G) in 1× TBE consisting of 16.35% Trizma-base (Sigma Aldrich; 93352), 2.78% boric acid (Sigma Aldrich; B0252-500 g), 0.93% EDTA (Sigma Aldrich; E5134-500G), and 2× MidoriGreen advanced (NIPPON Genetics, Tokyo, Japan; MG04). The gel was prepared in a Sub-Cell GT Cell (Bio-Rad, Hercules, CA; 1704401), and the DNA samples were run in 1× TBE buffer with HyperLadder 2 kb (Meridian, BIO-33053) at 80 V until separated (approximately 90 to 120 minutes). The gels were imaged using a UV gel imaging system (Nippon and Tuppon GLS) with an exposure time between 0.3 and 2 seconds.

Analysis of Oxford Nanopore Sequencing Data

Live basecalling was performed using Dorado version 0.6.2 on the GridION machines, and the resulting FASTQ files were mapped to the human reference genome GRCh38 using Minimap2 (epi2me-labs/wf-alignment v1.1.2)[13] to identify the reads spanning the targets of interest. NanoPlot was used for quality control of the aligned reads version 1.29.1.[14]

Multiplexing Runs

Sequencing data were processed using Epi2ME version 5.1.14 and the Epi2ME agent version 3.7.3 provided by

UNCORRECTED PROOF

FLA 5.7.0 DTD ■ JMDI1677_proof ■ 17 September 2026 ■ EO: JMDI-D-25-00450

ONT. The output included the proportion of barcodes in the sequencing run and the.csv file containing data on the aligned reads. For further analysis, specifically to generate BAM files that were used for the coverage analysis in Integrative Genomics Viewer (IGV), the Epi2ME desktop version was used to align reads to the hg38 genome. The workflow used was either the wf-alignment, which aligns reads to the chosen genome, or the wf-Cas9 pipeline, which outputs coverage information based on a target gene BED file. The output files were sorted into barcodes, and the coverage was read out by loading them into IGV and parsing them into a .csv file that could then be further processed with the statistical tool R 4.4.0. The pore count data were taken from the run reports, which also provided information on the data output of the flow cell.

PCR Validation

The *FGF14* PCR was performed as previously described,[15] and the *RFC1* PCR was performed as previously described.[11] The *ATN1*, *ATXN1*, *ATXN2*, *ATXN3*, *ATXN8OS/ATXN8*, *C9orf72*, *CACNA1A*, *FMR1*, and *FXN* PCR primer sequences and protocols are documented in Supplemental Table S3 and Supplemental Table S4.

Data Availability

Anonymized FASTQ files are available from European Genome-Phenome Archive (*https://ega-archive.org*, last accessed August 10, 2026) under StudyID EGAS50000001864 and DatasetID EGAD50000002695.

Results

Expansion of Clin-CATS Panels Increases Diagnostic Utility for Ataxia, Myopathy, Neurodegeneration, and ALS

The Clin-CATS workflow was expanded by increasing the number of target genes from 10 to 27 and introducing four distinct gene panels for ataxia, myopathy, neurodegeneration, and ALS/motor neuron disease (Figure 1 and Supplemental Table S1). These tailored panels enabled the simultaneous investigation of multiple repeat expansion-related genes across various neurologic conditions, significantly enhancing the diagnostic utility of Clin-CATS. Several new target regions were identified from the literature, and guide RNAs were designed specifically for these regions (Supplemental Table S1). It was confirmed that the guide RNAs accurately excise the newly included targets, such as key genes associated with spinocerebellar ataxias (eg, *SCA27B*, *SCA10*, *SCA31*, and *SCA37*), motor neuropathies, myopathies, and neurodegenerative disorders, like ALS and Huntington disease–like 2, by sequencing and verifying the alignments with the IGV[16,17] (Supplemental Figures S1, S2, and S3).

R9 Protocol on R10 Flow Cells Yields Suboptimal Coverage for Diagnostics

Although ONT has discontinued its legacy Cas9 protocol on R9 flow cells, similar reagents used in the ONT kit can be purchased directly from IDT and NEB, allowing the continuation of the workflow. The performance of the ONT Cas9 protocol on R10 flow cells resulted in suboptimal coverage, raising concerns about its use in the clinically validated Clin-CATS diagnostic assay. Given that this assay is accredited for clinical diagnostics in Germany, achieving the required quality control metrics, particularly minimum coverage requirements, on R10 MinION flow cells is essential. In two replicates, only 21 of 26 and 18 of 26 target genes reached the required 50× on-target coverage for diagnostics, with 5 and 8 target genes falling short of this threshold (control condition) (Supplemental Table S5 and Supplemental Figure S4). The ≥50× minimum coverage criterion was established through extensive clinical validation and routine diagnostic use of the Clin-CATS workflow, where it has consistently ensured robust and reproducible repeat sizing across loci. The percentage of sequenced reads aligning to target regions was 6.9% and 1.3%, respectively (Supplemental Table S5). The inability to reach the target coverage on both replicates suggested that the Cas9 protocol for R9 flow cells was not directly transferable to R10 flow cells, as it failed to provide the necessary coverage consistency required for diagnostic accuracy.

Adaptations to Clin-CATS Cas9 Protocol for R10 Ensure Adequate Coverage for Diagnostics

To address the suboptimal performance of the Cas9 protocol on R10 MinION flow cells, three modifications to the default Cas9 library preparation protocol (based on SQK-CS9109) were tested: extending the dA-tailing incubation step at 37°C from 15 to 60 minutes (bind60), extending the adapter ligation incubation time from 10 to 30 minutes at room temperature (lig30), and extending the adapter ligation incubation time from 10 minutes at room temperature to overnight at 16°C (ligON). On the basis of these results, this study further evaluated the combination of the first two modifications, testing the impact of simultaneously extending both the dA-tailing and ligation steps (bind60lig30).

Among the tested conditions, the ligON protocol produced the highest number of sequenced reads, with a 9.3-fold increase compared with the control (default Cas9 protocol on an R10 flow cell) (Supplemental Table S5). Despite the higher read count, the mean coverage across target regions remained similar to the control, and 5 of the 26 target genes did not achieve the required 50× coverage (Supplemental Table S5 and Supplemental Figure S4). Furthermore, the percentage of reads aligning to the target regions was 0.3%, indicating a substantial proportion of off-target reads, potentially because of the extended overnight ligation time. Prolonged ligation periods can increase the

UNCORRECTED PROOF

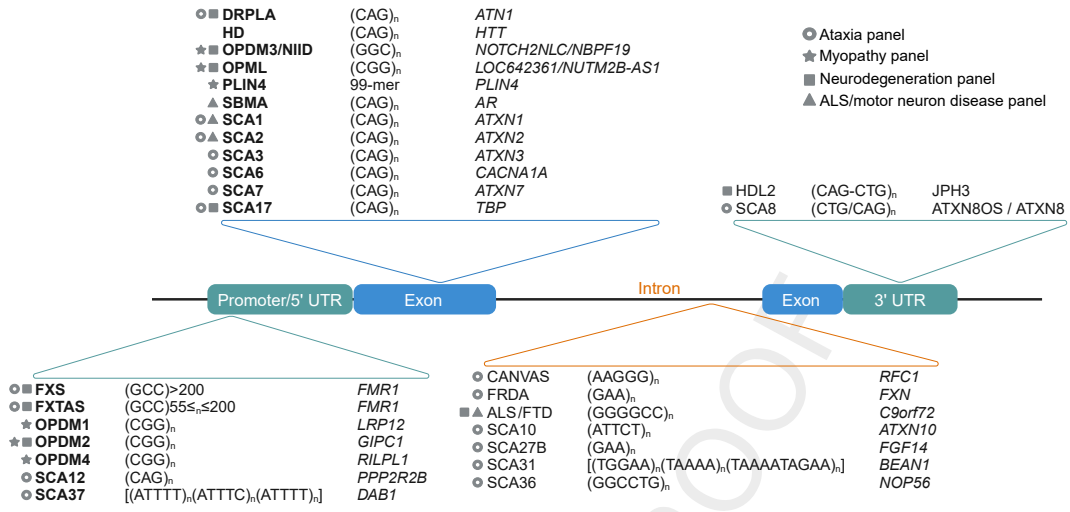


Figure 1 Overview of the updated Clin-CATS target genes and panels. Summary of the disease, repeat motifs, and target genes included in the updated Clin-CATS workflow, and their approximate genomic loci [promoter, 5′ untranslated region (UTR), exon, intron, and 3′ UTR]. The figure also highlights the gene panels in which each target gene is relevant [ataxia, myopathy, neurodegeneration, and amyotrophic lateral sclerosis (ALS)/motor neuron disease]. For additional details and complete disorder/condition names, refer to Supplemental Table S1.

likelihood of nonspecific ligation events, such as adapter dimer formation or ligation to unintended off-target fragments, reducing the overall efficiency and specificity of the library preparation.

The bind60 and lig30 conditions produced more moderate increases in sequencing reads, with fold changes of 1.6 and 2.0, respectively (Supplemental Table S5). Both conditions had improved mean coverage relative to the control, with bind60 achieving a 2.2-fold increase in coverage and lig30 showing a 1.9-fold increase. Importantly, in the bind60 condition, all target genes were covered in both replicates, whereas in the lig30 condition, one replicate achieved full coverage, and the other missed 4 genes, covering 22 of 26 (Supplemental Table S5 and Supplemental Figure S4).

Given the improvements in sequencing metrics observed in bind60 and lig30, it was hypothesized that combining these adaptations might yield further enhancements. In the bind60lig30 condition, the average number of reads was 1.7-fold higher than the control, while mean coverage increased by 2.0-fold (Supplemental Table S5). However, 1 of the 26 target genes in each replicate failed to reach the required 50× coverage, a slight reduction in performance compared with the bind60 and lig30 conditions alone (Supplemental Table S5 and Supplemental Figure S4). The percentage of reads aligned to the target regions for bind60, lig30, and bind60lig30 were 2.8%, 2.7%, and 2.8%, respectively. The combination of extended dA-tailing and ligation steps (bind60lig30) did not lead to a significant improvement in either read number or on-target coverage. In contrast, the bind60 and lig30 conditions produced the most consistent results. Because both bind60 and lig30 demonstrated improvements in sequencing metrics, the lig30 protocol was ultimately chosen for routine use, as it is easier to implement in a diagnostic laboratory setting (Figure 2A provides the finalized library preparation workflow).

Multiplexed R10 Clin-CATS Library Preparation Protocol Yields Inadequate Coverage for Diagnostics

Next, this study aimed to develop a multiplexed Cas9-based nanopore sequencing protocol to increase throughput and reduce costs for the Clin-CATS workflow. The first step was to optimize the DNA clean-up using rebuffered AMPure XP beads, testing a range of bead ratios from 0.1× to 0.7× to selectively remove short DNA fragments that could reduce sequencing efficiency (Supplemental Figure S5 and Supplemental Table S6). These experiments showed that concentrations <0.5× caused significant DNA loss, whereas higher concentrations retained more long fragments needed for successful sequencing. Ultimately, a 0.65× bead ratio was identified as optimal, striking a balance by minimizing DNA loss and retaining sufficiently long fragments. Using rebuffered beads, this clean-up strategy successfully increased the DNA integrity number by an average of 1.2, improving the quality of DNA fragments for nanopore sequencing (Supplemental Table S6). Incorporating this method into the routine Clin-CATS diagnostic workflow enabled the recovery of previously failed samples.

Attempts to implement multiplexing using ONT's beta-testing protocol and several modifications were met with limited success (Supplemental Figure S6 provides the library preparation schematic, and Supplemental Table S7 shows tested adaptations to the beta-testing protocol ONT multiplexing protocol). Multiple approaches were explored, including exonuclease treatments to remove off-target DNA, dephosphorylation to reduce non-specific adapter

print & web 4C/FPO

UNCORRECTED PROOF

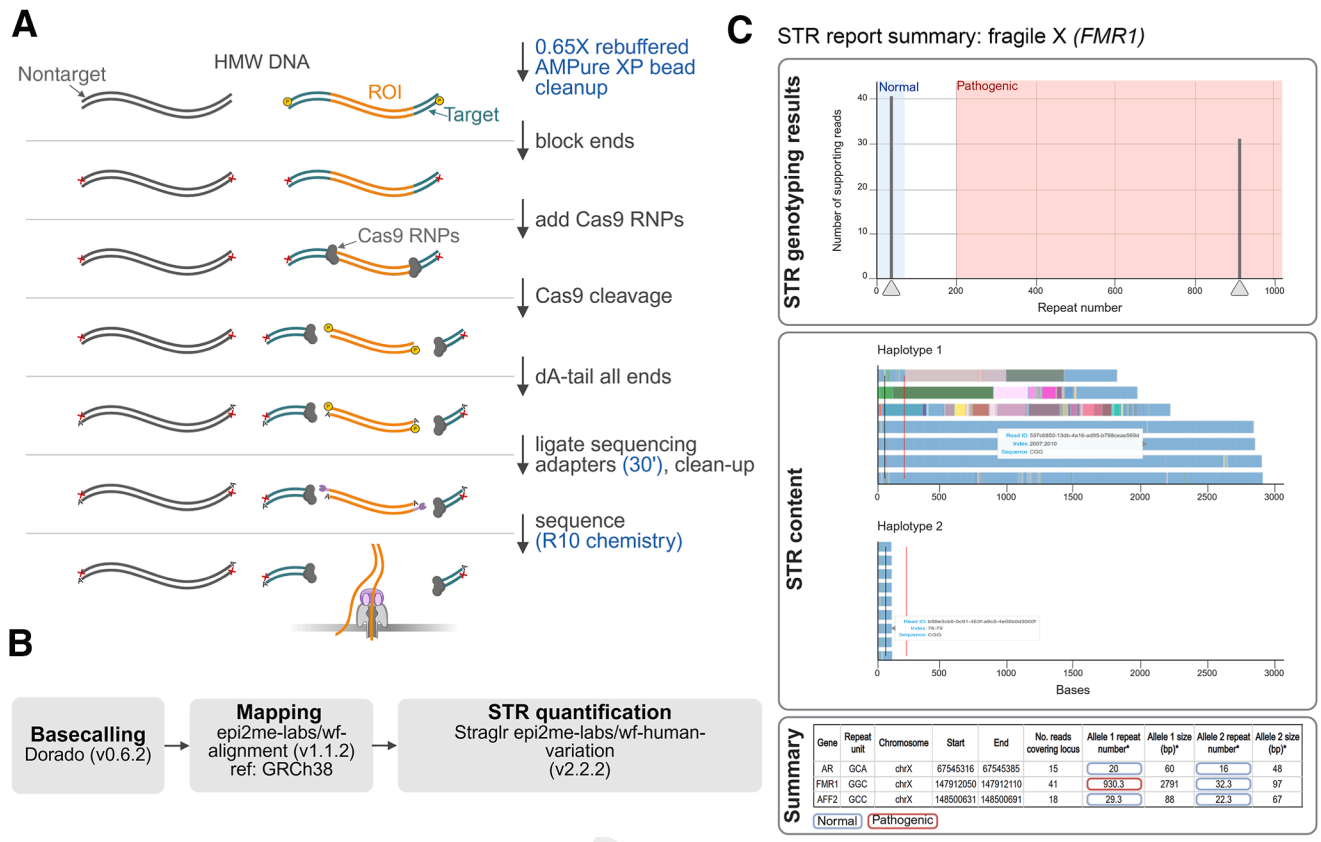


Figure 2 Updated Clin-CATS nanopore Cas9 sequencing workflow for R10 flow cells. **A:** Library preparation: high-molecular-weight (HMW) DNA extracted from EDTA blood is cleaned up using 0.65× rebuffered AMPure XP beads and then dephosphorylated to block the ends, preventing downstream ligation. Cas9 ribonucleoprotein (RNP) complexes are added to excise the region of interest (ROI; orange) from target DNA (green). All DNA is dA-tailed, and Oxford Nanopore Technologies (ONT) sequencing adapters (purple) are ligated to ROI DNA with a 5′ phosphate and complementary dA tail. The sample is cleaned up and sequenced. Changes specific to the R10 flow cell library preparation protocol are indicated in blue. **B:** Bioinformatic pipeline: basecalling is performed with Dorado version 0.6.2, then reads are mapped to the GRCh38 reference genome using epi2me-labs/wf-alignment version 1.1.2, and short tandem repeats (STRs) are quantified using Straglr in the ONT epi2me-labs/wf-human-variation workflow version 2.2.2. **C:** Example STR report for a patient with fragile X: the report includes STR genotyping results showing nonpathogenic (blue) and pathogenic (red) thresholds, repeat configurations for each read, and a summary table with gene, repeat unit, chromosome, positions, number of reads, and allele sizes with their pathogenicity indicated by color.

ligation, and the addition of Proteinase K to remove bound proteins, such as Cas9 from the target DNA. Despite these efforts, none of the modifications led to improved performance. Exonuclease-treated samples produced no usable reads, whereas Proteinase K led to significant reductions in sequencing coverage compared with the unmodified protocol (Supplemental Tables S8 and S9 and Supplemental Figure S7). Additionally, testing both PAM-in and PAM-out Cas9 guide orientations—where PAM-in refers to the guide RNA being designed with the PAM site facing toward the target region, and PAM-out refers to the PAM site facing away from the target region (Supplemental Figure S8 provides a schematic)—did not prevent target DNA from being digested by exonucleases (Supplemental Table S8 and Supplemental Figure S7). As a result, neither orientation was able to protect the region of interest, leading to poor enrichment and coverage across all tested conditions.

Multiplexed sequencing runs with two, four, and six samples failed to reach the required 50× coverage, achieving, on average, 3× to 5× coverage (Figure 3 and Supplemental Tables S10 and S11). Attempts to increase sequencing yield by reloading flow cells with two samples after 24 hours, either once (48-hour run) or twice (72-hour run), also resulted in low coverage, with averages of 11.4× and 7.9×, respectively (Supplemental Table S11). In subsequent tests, when samples were sequenced individually on R10 flow cells using the barcoding protocol, only two of six samples achieved 50× coverage for most genes (Supplemental Figure S9A and Supplemental Table S12), which was still lower than the coverage obtained in earlier single-plex runs without barcoding on R9 flow cells (Supplemental Figure S9B and Supplemental Table S12). The likely cause for these failures was the additional ligation step required to add barcode adapters in the multiplex protocol. Ligation was already inefficient, and introducing another step likely reduced the overall efficiency. This explains why even single samples sequenced with the multiplex protocol failed to reach the required coverage, as the added barcoding step further compromised the library preparation.

print & web 4C/FPO

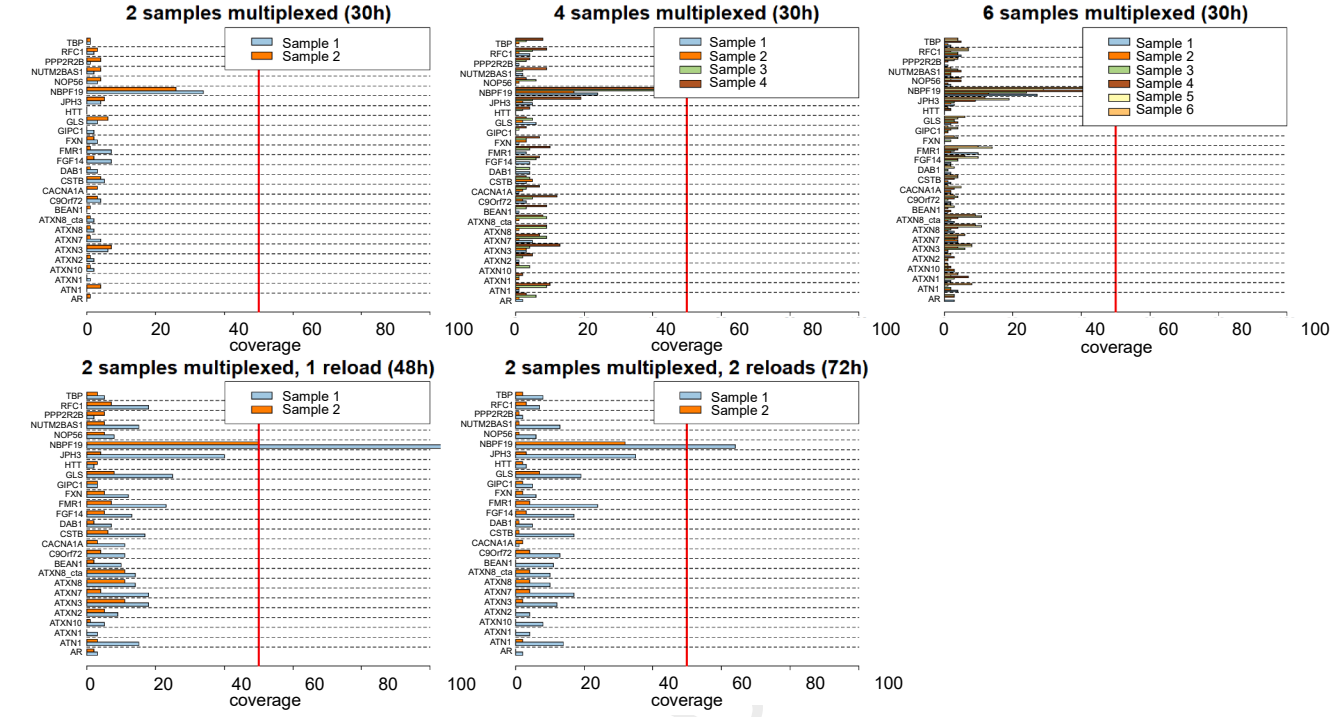


Figure 3 Coverage for multiplexing tests with the ataxia panel. After excision of the regions of interest using Cas9 ribonucleoproteins, dephosphorylation, and dA-tailing, Native Barcodes and the Native Adapter (Oxford Nanopore Technologies; SQK-NBD114.24) were ligated, and the cleaned-up libraries were loaded on R10 MinION flow cells. Bar plots of sequencing coverage for multiplexing tests with two, four, and six samples over a 30-hour run time. Coverage for two multiplexed samples with a 48-hour run time, including one library recovery and reloading at 24 hours. Coverage for two multiplexed samples with a 72-hour run time, including two rounds of library recovery and reloading at 24 and 48 hours. The target coverage of 50× is indicated by a **red line**. Refer to Supplemental Table S10 for read counts and Supplemental Table S11 for quantification. $n = 2$ to 6 biological replicates as indicated in the figure legend.

Flongle Flow Cells Provide Insufficient Coverage for the Clin-CATS Workflow

Flongle flow cells offer a smaller, more cost-effective alternative to MinION flow cells, costing approximately €70 compared with €400 for MinION. However, this comes at the expense of data yield and sequencing coverage. MinION flow cells contain 512 sequencing channels, each capable of hosting up to four nanopores, allowing for a maximum pore count of 2048. In contrast, Flongle flow cells have only 126 channels, each with a single pore, resulting in a maximum pore count of 126. Although MinION flow cells offer higher pore longevity and throughput because of the increased number of pores and the number of pores per channel, only one pore per channel can sequence at a time. Thus, the increase in MinION pore availability does not translate linearly to increased output. A better comparison of their respective capacities is data output; ONT reports that a MinION flow cell can generate approximately 48 GB of data over a 72-hour run, whereas Flongle produces only 2.6 GB over 14 hours—a predicted 18-fold difference in output per sequencing run.

When sequencing the ataxia panel, Flongle flow cells produced an average coverage of 3.2×, compared with 249.4× achieved with MinION flow cells with the same samples, representing a 78-fold reduction in coverage (Supplemental Table S12). Moreover, none of the genes in the Flongle panels reached the target diagnostic threshold of 50× coverage.

Single-Gene Targeting Provides Adequate Coverage as a Cost-Effective Alternative to Panels

For users interested in specific target genes rather than a comprehensive panel, targeting individual genes with relevant guide RNAs offers a practical and more cost-effective approach. To evaluate this strategy, this study tested the sequencing coverage of two genes, *FGF14* and *NOP56*, both within the full ataxia panel and using only guide RNAs specific to these genes (Supplemental Table S1). *FGF14* was selected as a representative example of a locus with a highly variable repeat length spectrum while not presenting an exceptionally complex sequence composition. In contrast, *NOP56* is characterized by a high GC content, which renders both enrichment and sequencing more challenging. Together, these loci reflect both average and challenging characteristics to test the robustness and performance of the workflow. These results showed that the average coverage of the target gene when using the single-

print & web 4C/FPO

UNCORRECTED

FLA 5.7.0 DTD ■ JMDI1677_proof ■ 17 September 2026 ■ EO: JMDI-D-25-00450

gene approach was only 10% lower compared with the coverage within the full panel (Supplemental Table S12). However, even with this slight reduction, the coverage exceeded the diagnostic threshold of 50×, suggesting that targeting single genes is a viable alternative to ordering guide RNAs for an entire panel when only a specific gene is of interest.

Integration of Straglr with Expanded Gene Set into the Clin-CATS Workflow Ensures R10 Compatibility

In the effort to update the Clin-CATS bioinformatics workflow for the new R10 flow cells, several tools were considered for analyzing repeat expansions. The R10 flow cells, which are set to replace the currently used R9 flow cells, are not yet supported by the STRique tool (as noted in STRique's GitHub repository, *https://github.com/giesselmann/STRique/issues/28*). This lack of support, coupled with the need to retrain the pore model to adapt to the R10 chemistry, made STRique less viable for the updated pipeline. Additionally, STRique is incompatible with ONT's new POD5 data format, and its development has been largely stagnant, leaving little room for future improvements or troubleshooting. As a result, this study opted to explore alternative tools that could better handle the specific requirements of R10 flow cell data. Tools such as Straglr, tandem-genotypes,[18] TriCoLoR,[19] and RepeatHMM[20] were considered. On the basis of recommendations from ONT and the nanopore community, its integration into the ONT Epi2Me wf-human-variation, as well as the fact that other tools either lacked support for newer algorithms or required complex adjustments, Straglr, in particular, stood out as a promising candidate. Moreover, Straglr was selected for integration into the Clin-CATS pipeline because of its ability to calculate repeat units after basecalling, which increases accuracy by mitigating sequencing errors and biases and allows it to be used for future flow cell chemistry changes. In the updated Clin-CATS workflow, basecalling is performed using Dorado version 0.6.2, reads are aligned to the GRCh38 reference genome using epi2me-labs/wf-alignment version 1.1.2, and STRs are quantified using Straglr within the ONT epi2me-labs/wf-human-variation workflow version 2.2.2 (Figure 2B). Furthermore, ONT has incorporated Straglr into an Epi2Me Nextflow workflow, offering a user-friendly HTML report with a clear visualization of repeat expansions (Figure 2C).

The Straglr pipeline was adjusted from the ONT repository to include missing genes from the panels (Supplemental Table S13). Several genes were already present in both the Straglr pipeline and Clin-CATS, including *AR*, *ATN1*, *ATXN1*, *ATXN10*, *ATXN2*, *ATXN3*, *ATXN7*, *ATXN8OS/ATXN8*, *C9orf72*, *CACNA1A*, *FMR1*, *FXN*, *GIPC1*, *HTT*, *JPH3*, *LRP12*, *NOP56*, *NOTCH2NLC/NBPF19*, *PPP2R2B*, and *TBP*. However, several key genes from the Clin-CATS panel were missing from the Straglr pipeline, including *BEAN1*, *DAB1*, *FGF14*, *LOC642361/NUTM2B-AS1*, *RFC1*, and *RILPL1*, which were subsequently added. This ensured the comprehensive coverage of clinically relevant genes across both the Clin-CATS and Straglr workflows.

Validation of the Updated Clin-CATS Workflow Shows Improved Accuracy but Reduced Coverage

This study validated the updated Clin-CATS workflow, which incorporates the adapted library preparation protocol, sequencing on R10 flow cells, and analysis using Straglr in the ONT wf-human-variation Epi2Me pipeline (Figure 2). This updated workflow was compared with the original Clin-CATS setup, which used the original Cas9 library preparation, R9 flow cells, and STRique analysis (per Erdmann et al,[8] 2023). This evaluation used clinical samples to assess both workflows' accuracy in detecting known repeat unit (RU) sizes across genes in the ataxia panel (Supplemental Table S14). Samples with high coverage from the original workflow were selected for this comparison.

The original and updated Clin-CATS workflows detected 12 of 50 RU *ATN1*, compared with the PCR result of 11 of 46 RUs. For *ATXN1*, the original workflow reported 32 of 49 RUs, whereas the updated workflow reported 32 of 48 RUs, aligning closely with the PCR result of 31 of 47. For *ATXN2*, both workflows accurately identified the pathogenic expansion of 32 RUs. For *ATXN3*, the original workflow slightly overestimated the expanded allele (72 RUs compared with 70 RUs as determined by PCR), whereas the updated workflow underestimated both the expanded and non-expanded alleles (24/66 RUs compared with 27/70 RUs determined by PCR). Regarding *ATXN8*, PCR can only determine whether an expansion exceeds 150 RUs. The clinical sample with a repeat expansion in *ATXN8* of >150 RUs was reported as a range of 29 to 542 RUs in the original workflow and a median of 250 RUs in the updated workflow. Manual inspection of IGV reads from the original Clin-CATS laboratory approach indicated an expanded allele range of 225 to 263 RUs, with a median of 257 RUs.

The original workflow with STRique could not detect repeat expansions in *BEAN1*, requiring manual quantification in IGV. In contrast, the updated Clin-CATS workflow successfully identified nonpathogenic *BEAN1* TAAAA repeat expansions of 18 of 857 and 19 of 640 RUs in two SCA31 patients, closely matching the IGV results from the original laboratory approach. For *C9orf72*, PCR can only determine if the repeat expansion is >30 or 60 RUs. The original workflow reported a range of 4 of 7 to 2697 GCCCCG repeats for the patient with a confirmed *C9orf72* repeat expansion, whereas the updated workflow reported a median 4 of 556 RUs. Both workflows produced consistent results for *CACNA1A*, detecting 13 of 23 repeats, aligning with the PCR result. For an SCA27B patient with an expansion in *FGF14*, where PCR can only report if the expansion exceeds 250 RUs, both workflows identified values within the pathogenic range (252 to 312 RUs). The updated workflow reported a median of 290 RUs, closely

UNCORRECTED PROOF

matching the manual IGV inspection from the original approach, which yielded a median of 285 RUs.

For *FMR1*, the updated workflow accurately identified 19 of 56 RUs for both alleles, matching the PCR result (19/56 RUs), whereas the original workflow reported a broader range (21/27 to 77 RUs). For *FXN*, the updated workflow provided a more precise estimate of 104 of 1093 RUs compared with the range of 63 to 1569 RUs reported by the original workflow, with the PCR result being 115 of 1060 RUs. Finally, for *RFC1*, PCR is limited to reporting whether the expansion exceeds 150 RUs. In a CANVAS patient with confirmed expansions in RFC1, the expanded allele was found to contain repeat motifs AAGGG and AAAGG, both exceeding 150 RUs as determined by PCR. Using the original Clin-CATS workflow, the expansion was reported as a range of 6 to 327 RUs with the motif AARRG. In contrast, the updated workflow provided a more precise measurement, reporting a median expansion size of 599 RUs for the same motif. To quantify concordance with PCR, the signed error, absolute error, and Pearson correlation were calculated for the expanded allele at each locus (Supplemental Table S15). Across all loci with available PCR reference, the updated workflow achieved a mean bias of −12.5 RUs and a mean absolute error of 45.6 RUs, compared with −121.5 RUs and 193.8 RUs for the original workflow. Pearson correlation with PCR improved from 0.65 to 0.95. These results indicated that the updated Clin-CATS workflow substantially improved the accuracy of expanded allele sizing compared with the original method.

The R10 flow cells provided improved mean read quality (16.5) over R9 (15.3), suggesting enhanced sequencing accuracy across clinical samples (Supplemental Table S14). When the updated workflow was initially applied to the same DNA samples used for the original R9 workflow 2 to 3 years earlier, coverage was substantially lower. Mean coverage across all targets for the original workflow was 439× (on R9), compared with 94× in the updated workflow (on R10). Target-specific coverage was similarly lower, averaging 444× for the original versus 75× for the updated. This coverage reduction can be partly attributed to sample age (2 to 3 years) and to the impact of freeze-thaw cycles. To better assess the updated Clin-CATS workflow's coverage, freshly collected and extracted control samples were sequenced, yielding an average coverage of 222× across all targets (Supplemental Table S16), which remains approximately 40% below that achieved with the original workflow. Additionally, Coriell samples processed with the updated workflow achieved a read quality of 17.5 but displayed lower-than-expected coverage, averaging 178× across all targets and 114× for specific target genes (Supplemental Table S2).

The Clin-CATS Workflow on R10 Flow Cells Accurately Detects Repeat Expansions in Coriell Samples

This study further validated the updated Clin-CATS workflow using Coriell samples (Supplemental Table S2), focusing on detecting known repeat expansions in genes associated with various neurologic conditions. In the SBMA sample NA23709, the *AR* CAG repeat was correctly identified with a median count of 49 of 51 RUs. In DRPLA samples NA13716 and NA13717, expansions in the *ATN1* gene were detected with median counts of 20 of 72 RUs and 18 of 72 RUs, respectively, matching the expected ranges. Sample NA03759 showed an expanded CTG repeat in *DMPK* with a median count of 15 of 2317 RUs, whereas NA03990 displayed a repeat range of 15 of 79 RUs, both consistent with the expected repeat number range. Similarly, the *FMR1* repeat expansion in FXS sample NA09237, with a known range of 931 to 940 RUs, was detected with a median repeat count of 864 of 1554 RUs, whereas samples NA20236 and NA20239 also matched the expected repeat number ranges for normal and premutation statuses, respectively. For Friedreich ataxia, sample NA16243 with an *FXN* expansion had a median GAA repeat count of 515 of 1953 RUs, fitting within the expected double expansion range. In the case of HD, this study could detect the correct number of *HTT* CAG repeats in samples NA13505 (22/50), NA13507 (15/55), NA13508 (22/58), NA13513 (15/49), NA13514 (15/52), and NA13512 (16/44), whereas NA13503 (17/45 expected versus 17/44 observed) and NA13515 (16/66 expected versus 16/65 observed) were in the expected range.

Discussion

In this study, an updated Clin-CATS workflow compatible with ONT R10 flow cells was presented, expanding the panel to 27 genes associated with ataxia, myopathy, neurodegeneration, and ALS/motor neuron diseases. This update brings several advantages in terms of diagnostic utility, sequencing accuracy, and workflow flexibility but also highlights significant challenges, particularly in the context of multiplexing and coverage.

The expanded gene panels significantly increase the diagnostic potential of Clin-CATS by enabling the simultaneous investigation of multiple repeat expansion-related genes across various neurologic conditions (Figure 1 and Supplemental Table S1). Another necessary refinement in the updated Clin-CATS workflow was the modification of the Cas9 library preparation protocol to address the suboptimal performance of the Cas9 approach on R10 flow cells. Although ONT has discontinued its Cas9 protocol, equivalent reagents, such as trans-activating CRISPR RNAs and Cas9 nuclease, can be purchased from IDT and NEB, allowing the continuation of the workflow (refer to *Library Preparation* and *ONT Cas9-Targeted Sequencing*). The final optimized protocol includes cleaning the samples with 0.65× rebuffered AMPure XP beads to remove short DNA fragments and extending the adapter ligation time from 10 to 30 minutes (Figure 2A). These modifications were essential for improving sequencing efficiency and ensuring

UNCORRECTED PROOF

FLA 5.7.0 DTD ■ JMDI1677_proof ■ 17 September 2026 ■ EO: JMDI-D-25-00450

that the R10 flow cells could meet the required 50× coverage thresholds for clinical diagnostics (Supplemental Table S5).

In this study, an attempt was made to multiplex samples prepared for Cas9 R10 sequencing by experimenting with several key parameters. With only 2.7% of reads aligning to the target region on average (Supplemental Table S5), this study explored methods to reduce off-target DNA and improve dA tailing and ligation efficiency by freeing the 3′P ends of the on-target DNA. Key adaptations included optimizing guide RNA orientation to control Cas9 binding specificity (on-target or off-target), using Proteinase K to digest Cas9 proteins that purportedly adhere to the template after cleavage to free the DNA ends for dA tailing and ligation, applying a combination of exonucleases to digest off-target DNA, and using dephosphorylation to differentiate the 5′ ends of the target DNA from off-target DNA, which retains a 5′P after Cas9 cleavage. Despite these efforts, multiplexing failed to achieve the necessary coverage for clinical diagnostics. Sequencing coverage for multiplexed samples fell below the required 50× threshold, with average coverage ranging from 3× to 5×, even when multiple samples were reloaded onto flow cells and the sequencing run time was extended (Figure 3 and Supplemental Table S11). Although previous studies have demonstrated successful multiplexing of Cas9-enriched libraries,[21,22] they typically targeted fewer genes, resulting in higher sequencing efficiency. In contrast, this larger gene panel exceeded the current capabilities of multiplexed Cas9-enriched nanopore sequencing, particularly with R10 flow cells. The challenges of multiplexing in this study likely stemmed from the inefficiencies introduced by additional ligation steps required for barcoding each sample, which may have further reduced overall library preparation efficiency. To overcome these limitations, protocol refinements are required.

In addition to multiplexing, this study evaluated the use of Flongle flow cells and single-gene targeting as cost-effective alternatives to panel-based approaches. Flongle flow cells provided inadequate sequencing coverage, with an average of 3.2× compared with 249.4× for MinION flow cells for the same samples (Supplemental Table S12). In contrast, single-gene targeting consistently achieved coverage exceeding 50× (Supplemental Table S12), making it a more cost-effective alternative in terms of sequencing guides when users are only interested in specific target genes. Additionally, a critical discovery in earlier studies indicated the necessity of using fresh blood samples for DNA extraction. Current DNA extraction via an automated system does not optimize for high-molecular-weight DNA, which, although practical for diagnostics, could benefit from alternative methods like Nanobind to further enhance coverage and sequencing performance.

Although R10 flow cells deliver substantial improvements in sequencing accuracy, the overall coverage achieved with R10 was lower compared with that of R9 flow cells. Specifically, these results showed that the updated workflow, which uses R10 flow cells, achieved a mean coverage of 222× across all targets (Supplemental Table S16), compared with 439× for the original workflow, which uses R9 flow cells (Supplemental Table S14). This lower coverage could present a challenge for Cas9-targeted sequencing applications, which require consistent coverage across target regions. Despite this limitation, the updated Clin-CATS approach met the 50× minimum coverage threshold required for a clinical diagnostics workflow. Moreover, the updated workflow demonstrated improved accuracy in detecting and quantifying repeat units in both clinical and Coriell samples (Supplemental Table S2 and Supplemental Table S12). Notably, the updated Cas9-targeted R10 workflow achieved higher target coverage than is typically reported for adaptive sampling approaches. Although adaptive sampling represents an alternative enrichment strategy, published data indicate comparatively lower on-target yield and throughput relative to Cas9-mediated enrichment with R10 chemistry.[21] A direct comparison with adaptive sampling was not undertaken, as the study was specifically designed to assess the performance of the adapted Clin-CATS workflow on R10 flow cells. Additionally, unlike HiFi sequencing, which struggles to detect GA-rich repeat expansions, such as (AAGGG)n and (AAAGG)n, because of reduced read quality and insufficient coverage,[23] this workflow effectively captures and quantifies these complex repeat structures, offering a significant advantage for the analysis of pathogenic expansions in genes like *RFC1* and *FXN*.

In addition to the global coverage differences between R9 and R10 chemistries, substantial variability was observed across individual loci. Several locus-specific and technical factors contribute to this variability. Sequence characteristics, such as high GC content or extreme repeat lengths, can reduce enrichment efficiency and sequencing uniformity. Fragment size distribution and guide RNA efficiency in Cas9-mediated enrichment further influence the number of reads captured per locus. Additionally, loci with high homology to other genomic regions, such as *NOTCH2NLC*/*NBPF19* within the segmentally duplicated 1q21 region, can generate multimapping reads, inflating apparent coverage without reflecting uniquely mapped sequences. Collectively, these factors account for much of the observed interlocus coverage variability and should be considered when interpreting locus-specific sequencing performance.

R10 flow cells present clear benefits, offering higher accuracy and, in non−Cas9 applications, increased throughput. As ONT continues refining its technology, even greater throughput from R10 flow cells across various protocols, including those based on Cas9, is anticipated. Although adopting ONT's Q line—providing guaranteed R9 flow cell availability for a fixed period—could support maintaining higher coverage, this option restricts the potential to incorporate future upgrades and reduces operational flexibility by committing the sequencing device

UNCORRECTED PROOF

exclusively to R9 flow cells. Given ONT's rapid advancements, it is believed that transitioning to R10 flow cells is a more strategic long-term solution.

An important improvement in the updated Clin-CATS workflow was the integration of Straglr,[12] a tool embedded in the ONT epi2me-labs/wf-human-variation workflow, which processes base-called reads rather than raw electrical signals. The original Clin-CATS workflow relied on the now-deprecated tool STRique,[7] which quantified STRs from raw signal data. The original Clin-CATS pipeline using STRique was generally incapable of determining a median for large repeat expansions (>100 RUs) and instead provided a range of repeat sizes; as a result, these expansions, for example, in the case of *RFC1*, required manual evaluation. The rationale for this approach is that large expansions exhibit a broad range of repeat sizes, leading to a lack of a single peak for one repeat size, as seen with shorter expansions. Following distribution plotting, peak determination was conducted manually; although retrospective, this approach introduced limitations in precision. Moreover, the migration to Straglr enhances Clin-CATS compatibility with future ONT flow cell chemistries, ensuring long-term adaptability. By processing base-called data, the pipeline can function seamlessly across different sequencing chemistries, reducing the need for revalidation or retraining models, an issue that was more prevalent with tools like STRique, which were tied to specific pore models and data formats. Additionally, using ONT's workflow ensures that the pipeline remains actively maintained.

Although Straglr presents a robust advancement in the Clin-CATS workflow, some gene-specific adjustments are necessary to ensure optimal performance. Challenges persist in achieving precise phasing and accurately handling subtle repeat differences, especially in differentiating closely sized alleles, such as a 13-repeat versus a 14-repeat allele. Additionally, this study used Straglr parameters that do not consider broken reads, which may limit its effectiveness in capturing long repeats, especially for genes like *BEAN1*, *C9orf72*, and *RFC1*, where pathogenic repeat expansions are extensive. Importantly, for the diagnosis of these conditions, it is often sufficient to detect an expansion exceeding a specific threshold, an assessment that can be made even from broken reads. This threshold-based detection approach may also explain the divergent results observed for *C9orf72*. Improvements in DNA extraction and library preparation could help preserve DNA fragments, increasing the likelihood of full-length reads and improving accuracy for these complex loci. However, expansions of the pathogenic AAGGG repeat motif in *RFC1* seem to be prone to breaking or fragmenting, likely because of the formation of specific secondary DNA structures,[24] presenting an inherent challenge beyond library preparation alone. Additionally, the *FMR1* repeat expansion in the FXS sample NA09237, with a known range of 931 to 940 RUs, was detected with a median repeat count of 864 of 1554 RUs. This discrepancy likely stems from the sample being male, containing only one FMR1 copy, which Straglr did not account for, instead forcing a two-allele call. Future refinements could incorporate gender-specific handling in X-linked disorders to improve accuracy.

The reporting approach of Straglr could be further improved, particularly for genes like *RFC1* that contain multiple repeat motifs. Currently, extracting details about repeat interruptions or expansions involving multiple motifs requires manually examining individual reads, which is not ideal for clinical reporting. Enhancing Straglr's ability to consolidate and report complex repeat structures directly in summary outputs would significantly improve its clinical utility, particularly for diagnosing disorders with heterogeneous repeat motifs.

STRique does not further process the isolated repeat lengths of individual reads and lacks phasing capabilities, introducing ambiguity that complicates precise allele differentiation and often necessitates intensive additional inspection of reads in certain cases. In contrast, Straglr reports the median repeat count, offering a more consistent and interpretable result that supports clearer allele identification. Straglr also provides key advantages, including compatibility with evolving ONT technologies, enhanced bioinformatics support, and active development, making it the preferred choice for the Clin-CATS workflow. This shift not only improves the pipeline's adaptability and precision but also ensures alignment with ONT's latest sequencing platforms, positioning the workflow for future advancements in clinical diagnostics.

In conclusion, the updated Clin-CATS workflow, which incorporates an adapted library preparation protocol, R10 flow cell sequencing, and analysis with Straglr in the ONT wf-human-variation Epi2Me pipeline, ensures continuity of this approach despite the discontinuation of R9 flow cells and the Cas9 library preparation kit, while also providing substantial benefits. However, challenges such as suboptimal multiplexing and reduced sequencing coverage underline the need for ongoing refinements. Despite these hurdles, the workflow demonstrates strong adaptability with expanded panels, improved sequencing accuracy, and bioinformatic flexibility, supporting its long-term application in research and clinical diagnostics. Future efforts should focus on optimizing multiplexing protocols, increasing throughput to fully leverage R10 flow cell capabilities, improving library preparation protocols to preserve better full-length reads for accurate large expansion analysis, and enhancing bioinformatic tools like Straglr for better reporting of complex repeat expansions.

Acknowledgments

We thank the patients who provided samples for this study and for providing informed written consent, as well as Isabel Abellan Schneyder from Oxford Nanopore Technologies for support and input on the project.

UNCORRECTED PROOF

Author Contributions

V.Schö. performed most of the laboratory work for R10 validation; V.Scho. performed the bioinformatic analysis of the R10 validation data; V.P. performed the laboratory work and analysis for the Cas9 multiplexing tests; H.E., A.A., E.H.-F., and T.N. provided clinical expertise and devised the gene panels; M.Sc. and M.A. validated the guide RNAs; M.G., V.P., and M.M.-H. performed the MinION versus Flongle and panel versus single-target comparisons; A.L. and K.E. performed the *FGF14* and *RFC1* PCRs; H.E., M.C.L., V.M., E.B., I.v.B., and C.D. assessed the sequencing data and clinical results; M.C.L., A.B.-P., A.A., and E.H.-F. conceived the project; M.C.L. and A.B.-P. supervised the work; E.H.-F., A.A., and T.N. provided the funding and resources to conduct the work; M.C.L., V.Scho., and V.P. prepared the figures; and M.C.L. wrote the manuscript, with contributions from all authors.

Disclosure Statement

M.C.L. has received travel and accommodation expenses for speaking at Oxford Nanopore Technologies conferences. The authors declare that the submitted work was otherwise performed without professional or financial relationships that could be construed as a conflict of interest.

Supplemental Data

Supplemental material for this article can be found at *https://doi.org/10.1016/j.jmoldx.2026.04.007*.

References

1. Hadjivassiliou M, Martindale J, Shanmugarajah P, et al. Causes of progressive cerebellar ataxia: prospective evaluation of 1500 patients. *J Neurol Neurosurg Psychiatry*. 2017;88:301−309.
2. Sandford E, Burmeister M. Genes and genetic testing in hereditary ataxias. *Genes*. 2014;5:586−603.
3. Paulson H. Repeat expansion diseases. *Handb Clin Neurol*. 2018;147:105−123.
4. Chintalaphani SR, Pineda SS, Deveson IW, Kumar KR. An update on the neurological short tandem repeat expansion disorders and the emergence of long-read sequencing diagnostics. *Acta Neuropathol Commun*. 2021;9:98.
5. Sequeiros J, Martindale J, Seneca S, et al. . EMQN best practice guidelines for molecular genetic testing of SCAs. *Eur J Hum Genet*. 2010;18:1173−1176.
6. 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.
7. Giesselmann P, Brändl B, Raimondeau E, et al. Analysis of short tandem repeat expansions and their methylation state with nanopore sequencing. *Nat Biotechnol*. 2019;37:1478−1481.
8. Erdmann H, Schöberl F, Giurgiu M, et al. Parallel in-depth analysis of repeat expansions in ataxia patients by long-read sequencing. *Brain*. 2023;146:1831−1843.
9. Jayadev S, Bird TD. Hereditary ataxias: overview. *Genet Med*. 2013;15:673−683.
10. Brusse E, Maat-Kievit JA, van Swieten JC. Diagnosis and management of early- and late-onset cerebellar ataxia. *Clin Genet*. 2007;71:12−24.
11. Cortese A, Simone R, Sullivan R, et al. Biallelic expansion of an intronic repeat in RFC1 is a common cause of late-onset ataxia. *Nat Genet*. 2019;51:649−658.
12. Chiu R, Rajan-Babu I-S, Friedman JM, Birol I. Straglr: discovering and genotyping tandem repeat expansions using whole genome long-read sequences. *Genome Biol*. 2021;22:224.
13. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34:3094−3100.
14. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics*. 2018;34:2666−2669.
15. Pellerin D, Danzi MC, Wilke C, et al. Deep intronic GAA repeat expansion in late-onset cerebellar ataxia. *N Engl J Med*. 2023;388:128−141.
16. Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative Genomics Viewer. *Nat Biotechnol*. 2011;29:24−26.
17. Robinson JT, Thorvaldsdottir H, Turner D, Mesirov JP. igv.js: an embeddable JavaScript implementation of the Integrative Genomics Viewer (IGV). *Bioinformatics*. 2023;39:btac830.
18. Mitsuhashi S, Frith MC, Mizuguchi T, et al. Tandem-genotypes: robust detection of tandem repeat expansions from long DNA reads. *Genome Biol*. 2019;20:1−17.
19. Bolognini D, Magi A, Benes V, Korbel JO, Rausch T. TRiCoLOR: tandem repeat profiling using whole-genome long-read sequencing data. *Gigascience*. 2020;9:giaa101.
20. Liu Q, Tong Y, Wang K. Genome-wide detection of short tandem repeat expansions by long-read sequencing. *BMC Bioinformatics*. 2020;21:542.
21. Alfano M, De Antoni L, Centofanti F, et al. Characterization of full-length CNBP expanded alleles in myotonic dystrophy type 2 patients by Cas9-mediated enrichment and nanopore sequencing. *Elife*. 2022;11:e80229.
22. McDonald TL, Zhou W, Castro CP, et al. Cas9 targeted enrichment of mobile elements using nanopore sequencing. *Nat Commun*. 2021;12:3586.
23. Höps W, Weiss MM, Derks R, et al. HiFi long-read genomes for difficult-to-detect, clinically relevant variants. *Am J Hum Genet*. 2025;112:450−456.
24. Wang Y, Wang J, Yan Z, et al. Structural investigation of pathogenic RFC1 AAGGG pentanucleotide repeats reveals a role of G-quadruplex in dysregulated gene expression in CANVAS. *Nucleic Acids Res*. 2024;52:2698−2710.

UNCORRECTED PROOF

Supplemental Figure S1 Integrative Genomics Viewer (IGV) snapshots confirming guide RNA targeting of the correct loci for *AR*, *ATN1*, *ATXN1*, *ATXN2*, *ATXN3*, *ATXN7*, *ATXN8OS/ATXN8*, *ATXN10*, and *BEAN1* genes. Detailed information regarding the specific target regions and guide RNAs can be found in Supplemental Table S1.

Supplemental Figure S2 Integrative Genomics Viewer (IGV) snapshots confirming guide RNA targeting of the correct loci for *C9orf72*, *CACNA1A*, *DAB1*, *FGF14*, *FMR1*, *FXN*, *GIPC1*, *JPH3*, and *LOC642361/NUTM2B-AS1* genes. Detailed information regarding the specific target regions and guide RNAs can be found in Supplemental Table S1.

Supplemental Figure S3 ntegrative Genomics Viewer (IGV) snapshots confirming guide RNA targeting of the correct loci for *LRP12*, *NOP56*, *NOTCH2NLC/NBPF19*, *PLIN4*, *PPP2R2B*, *RFC1*, *RILPL1*, and *TBP* genes. Detailed information regarding the specific target regions and guide RNAs can be found in Supplemental Table S1.

Supplemental Figure S4 Coverage across target regions of the Clin-CATS ataxia panel for adaptations of the Cas9 protocol sequenced on R10 flow cells. Bar plots showing the mean coverage for the repeat target region of each gene in the ataxia panel for two technical replicates. The **red line** represents the 50× coverage threshold. Numbers to the right of each bar indicate the mean coverage for the target gene for that specific run. The control condition is the original Cas9 protocol sequenced on R10 flow cells.

Supplemental Figure S5 Agarose gels showing the results of AMPure XP bead clean-up tests for input DNA used in Cas9 library preparation. **A:** Initial clean-up test with 0.1× to 0.7×. **B:** The 0.8× to 0.5× (test 1). **C:** The 0.6× to 0.5× (test 2). **D:** Two biological replicates of 0.7× to 0.56× (test 4.1 to 4.2). Refer to Supplemental Table S6 for the DNA quantification from tests 1 to 4.2.

Supplemental Figure S6 Flowchart showing the workflow and changes of the multiplex approach. Numbers indicate the protocol adaptations tested: (1) Protospacer adjacent motif (PAM)-in or PAM-out orientation guides, influencing whether the Cas9-protein sticks to the target or off-target DNA. (2) DNA dephosphorylated to improve the efficiency of dA-tailing on the target DNA. (3) Proteinase K is used to remove the remaining Cas9 protein that may be adhered to the region of interest (ROI). (4) Digestion of off-target DNA using exonucleases to increase the relative percentage of target DNA.

Supplemental Figure S7 Effects of multiplexing protocol adaptations on sequencing coverage. Bar plot of the mean coverage of each library preparation condition, normalized to pore count and box plot comparing the coverage normalized to pore count with and without exonuclease degradation. Condition 1: proteinase K after Cas9 digestion, protospacer adjacent motif (PAM)-out CRISPR RNA (crRNA); condition 2: Lambda exonuclease digestion, PAM-out crRNA; condition 3: Lambda exonuclease and exonuclease I digestion, PAM-out crRNA; condition 4: proteinase K and dephosphorylation, PAM-out crRNA; condition 5: no changes, PAM-out crRNA; condition 6: dephosphorylation, PAM-out crRNA; condition 7: Lambda exonuclease digestion, PAM-in crRNA; condition 8: Lambda exonuclease and exonuclease I digestion, PAM-in crRNA; and condition 9: no changes, PAM-in crRNA. Refer to Supplemental Table S8 for the relevant data. $n = 2$ biological replicates.

Supplemental Figure S8 Orientation of the Cas9 protein after cleavage. **A:** Schematic of the most recent beta-testing Oxford Nanopore Technologies (ONT) multiplex protocol showing the binding of the Cas9 protein to the target DNA (copied from ONT protocol). **B:** Schematic of the ONT simplex protocol, showing the Cas9 adhering to the off-target DNA (copied from ONT protocol). **C:** Schematic of the protospacer adjacent motif (PAM)-in orientation guide RNA design, where the PAM sites are orientated toward each other, and the CRISPR RNA (crRNA) sits on the nontarget strand. **D:** Schematic of the PAM-out orientation guide RNA design, where the PAM sites are orientated away from each other, and the crRNA sits on the target strand. DS, downstream; ROI, region of interest; RNP, ribonucleoprotein; US, upstream.

Supplemental Figure S9 Coverage across target regions of the Clin-CATS ataxia panel for single-plexed samples. Bar plots showing the mean coverage for the repeat target region of each gene in the ataxia panel for individual samples prepared using the multiplexing protocol and sequenced on R10 flow cells (**A**) and original R9 protocol (nonmultiplexed) and sequenced on R9 flow cells (**B**). The **red line** represents the 50× coverage threshold. For read counts, refer to Supplemental Table S12.

UNCORRECTED PROOF