

# Deciphering Neurodegenerative Diseases Using Long-Read Sequencing

Yun Su, MD,\* Liyuan Fan, MD,\* Changhe Shi, MD, PhD,\* Tai Wang, MD, Huimin Zheng, MD, Haiyang Luo, MD, PhD, Shuo Zhang, MD, PhD, Zhengwei Hu, MD, PhD, Yu Fan, MD, PhD, Yali Dong, MD, Jing Yang, MD, PhD, Chengyuan Mao, MD, PhD, and Yuming Xu, MD, PhD

*Neurology*® 2021;97:423-433. doi:10.1212/WNL.00000000000012466

## Correspondence

Dr. Xu  
xuyuming@zzu.edu.cn  
or Dr. Mao  
maochengyuan2015@126.com

## Abstract

Neurodegenerative diseases exhibit chronic progressive lesions in the central and peripheral nervous systems with unclear causes. The search for pathogenic mutations in human neurodegenerative diseases has benefited from massively parallel short-read sequencers. However, genomic regions, including repetitive elements, especially with high/low GC content, are far beyond the capability of conventional approaches. Recently, long-read single-molecule DNA sequencing technologies have emerged and enabled researchers to study genomes, transcriptomes, and metagenomes at unprecedented resolutions. The identification of novel mutations in unresolved neurodegenerative disorders, the characterization of causative repeat expansions, and the direct detection of epigenetic modifications on naive DNA by virtue of long-read sequencers will further expand our understanding of neurodegenerative diseases. In this article, we review and compare 2 prevailing long-read sequencing technologies, Pacific Biosciences and Oxford Nanopore Technologies, and discuss their applications in neurodegenerative diseases.

\*These authors contributed equally to this work.

From the Department of Neurology (Y.S., L.F., C.S., T.W., H.Z., H.L., S.Z., Z.H., Y.F., Y.D., J.Y., C.M., Y.X.), The First Affiliated Hospital of Zhengzhou University, Zhengzhou University, Henan, P. R. China; The Academy of Medical Sciences of Zhengzhou University (Y.S., L.F., C.S., T.E., H.Z., H.L., S.Z., Z.H., Y.F., Y.D., C.M.), Zhengzhou University, Henan, P. R. China; Henan Key Laboratory of Cerebrovascular Diseases (Y.S., L.F., C.S., H.Z., H.L., S.Z., Z.H., Y.F., Y.D., J.Y., C.M., Y.X.), The First Affiliated Hospital of Zhengzhou University, Zhengzhou University, Henan, P. R. China; Sino-British Research Centre for Molecular Oncology (C.S., C.M.), National Centre for International Research in Cell and Gene Therapy, School of Basic Medical Sciences, Academy of Medical Sciences, Zhengzhou University, Henan, P. R. China; and Institute of Neuroscience (J.W., Y.X.), Zhengzhou University, Henan, P. R. China.

Go to [Neurology.org/N](https://www.neurology.org/N) for full disclosures. Funding information and disclosures deemed relevant by the authors, if any, are provided at the end of the article.

The Article Processing Charge was funded by the National Natural Science Foundation of China (Grants U190420029, 91849115, and 81530037) and the National Key Research and Development Program of China (Grant 2017YFA0105003).

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND), which permits downloading and sharing the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

## Glossary

**AD** = Alzheimer disease; **ALS** = amyotrophic lateral sclerosis; **BAC** = bacterial artificial chromosome; **BAFME** = Benign adult familial myoclonic epilepsy; **C9ORF72** = chromosome 9's open reading frame 72; **CCS** = circular consensus sequence; **dsDNA** = double-stranded DNA; **FAME** = familial adult myoclonic epilepsy; **FCMTE** = familial cortical myoclonic tremor with epilepsy; **FMR1** = fragile X mental retardation 1; **FTD** = frontotemporal dementia; **FXS** = fragile X syndrome; **FXTAS** = fragile X-associated tremor/ataxia syndrome; **HD** = Huntington disease; **HTT** = huntingtin; **IPD** = interpulse duration; **LRS** = long-read sequencer; **NGS** = next-generation sequencing; **NIID** = neuronal intranuclear inclusion disease; **ONT** = Oxford Nanopore Technologies; **OPDM** = oculopharyngodistal myopathy; **PacBio** = Pacific Biosciences; **PM** = premutation; **PolyQ** = polyglutamine; **RP-PCR** = repeat-primed PCR; **SCA** = spinocerebellar ataxias; **SMRT** = single-molecule real-time; **SNP** = single nucleotide polymorphism; **SV** = structural variant; **UTR** = untranslated region; **WGS** = whole-genome sequencing.

Neurodegenerative diseases are a class of neurologic disorders characterized by a progressive course of cognitive decline, motor disturbance, psychiatric disorders, and eventual disability. The neuropathologic features are the selective loss or impairment of specific neurons in the central and peripheral nervous systems. As the second most common cause of death, neurodegenerative disorders impose a severe burden on individuals, communities, and countries. Therefore, it is crucial to detect and intervene early to slow down the irreversible progression of these diseases.

There are a small number of patients with family histories of neurodegenerative diseases, thus indicating that genetic factors may play an etiologically important role. Since the identification of expanded trinucleotide repeats in Huntington diseases (HDs),<sup>1</sup> cumulative evidence has suggested that repeat expansions may be more involved in the pathogenesis of neurodegenerative diseases than first thought, regardless of whether they reside in coding or noncoding regions including 5' or 3' untranslated regions (5' or 3' UTR) and introns.<sup>2,3</sup> Nevertheless, next-generation sequencing (NGS) technologies are inadequate to resolve the expansions exceeding several kilobases, owing to the limit of short read lengths (150–300 bp) and an underrepresentation of GC-rich/poor regions,<sup>4</sup> despite innumerable discoveries of mutations facilitated by NGS.

In contrast to NGS, 2 prevailing long-read sequencers termed LRS, single-molecule real-time (SMRT) sequencing developed by Pacific Biosciences (PacBio)<sup>5</sup> and nanopore sequencing by Oxford Nanopore Technologies (ONTs),<sup>6</sup> provide an alternative strategy for single DNA molecule sequencing in real time. These platforms generate exceptionally long reads of several tens of kilobases, which encompass full-length repetitive regions. Moreover, guanine-cytosine (GC) bias can be reduced to a low degree, and more homogeneous genome coverage can be achieved without the requirement of PCR amplification, as opposed to NGS. Recently, clustered regularly interspaced short palindromic repeats/CRISPR associated 9 (CRISPR/Cas9)-mediated amplification-free targeted enrichment in conjunction with LRS has emerged as an instrumental tool to achieve targeted sequencing.<sup>7</sup> The unique features of LRS have made it well suited for deciphering neurodegenerative diseases, especially when NGS

receives negative results. Here, we summarize the mechanisms of LRS and their applications in human neurodegenerative disorders.

## Long-Read Technologies

### SMRT Sequencing: PacBio

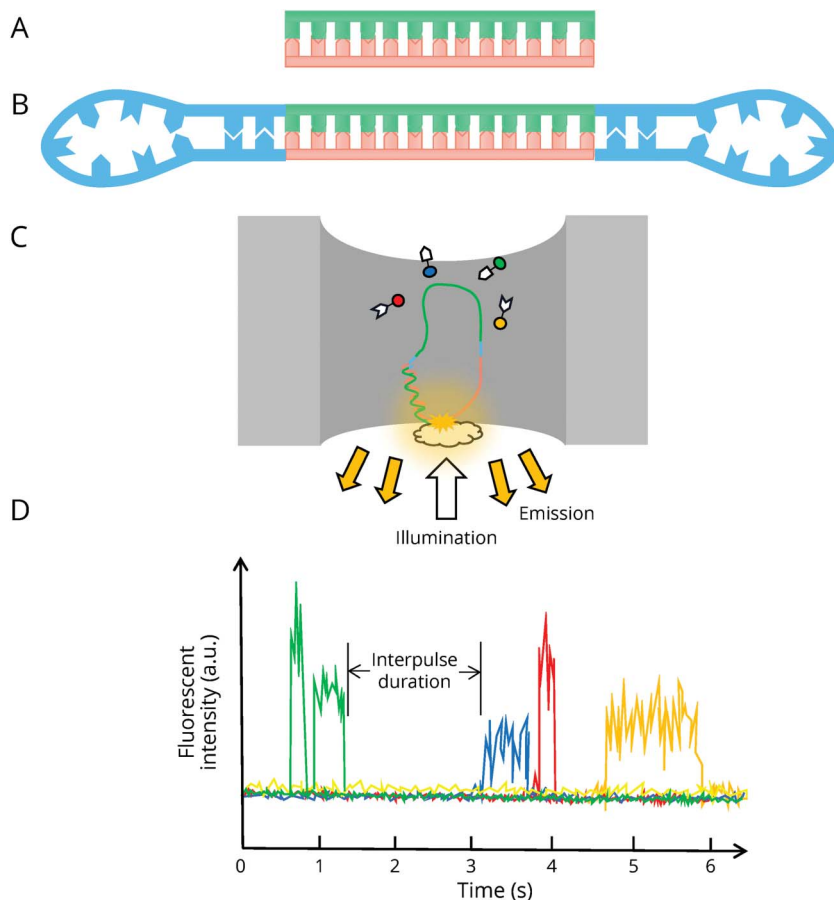
PacBio RS, RSII, Sequel, and SequelII sequencers with increased average read lengths and throughputs have allowed for genomics research at an unprecedented resolution, all of which share a common mechanism, known as synthesis sequencing. First, a template called SMRTbell is created by ligating hairpin adapters to both ends of a target double-stranded DNA (dsDNA) molecule, thereby forming a closed circular DNA (Figure 1A and B). The library is then loaded onto a flowcell composed of nanoscale wells called zero mode waveguides, where a polymerase is immobilized at the bottom and can combine an SMRTbell adaptor with a complementary primer to initiate the replication process (Figure 1C). The incorporation of fluorescently labeled nucleotides releases excited fluorescence signals, with the color and duration of the emitted light being recorded in real time (Figure 1D). A circular consensus sequence (CCS) with higher accuracy (>99%) is eventually constructed from multiple reads covering the original DNA template.<sup>5</sup> In addition, the time between adjacent nucleotide incorporations is called the interpulse duration (IPD) (Figure 1D), which reflects variable polymerase kinetics influenced by modification events,<sup>8</sup> and contributes to the epigenetic analysis.

Despite the preponderance of SMRT, raw read errors dominated by deletions or insertions remain problematic. Given the random occurrence of errors, base discrimination can be obtained by constructing a CCS having >99% accurate consensus with sufficient coverage.<sup>5</sup>

### Nanopore Sequencing: ONT

To date, ONT has introduced 4 sequencers: MinION, PromethION, GridION, and SmidgION, which are all competent in electronically detecting individual bases during the disassembling process. To begin with, library preparation occurs in dsDNA fragments, which are end repaired and adapter ligated (Figure 2A and B). The adapters contain 5'

**Figure 1** Overview of Single-Molecule Real-Time Sequencing (SMRT) Technology



(A) Library preparation starts with cutting dsDNA to the right size. (B) The template termed SMRTbell is created by ligating hairpin adapters (light blue) to both ends of dsDNA. (C) The library is thereafter loaded onto a SMRT flowcell containing millions of zero mode waveguides (ZMWs) (gray). In the best case, a SMRTbell diffuses into a ZMW, and the adaptor binds to a DNA polymerase (white) immobilized at the bottom, thereby initiating the incorporation of fluorescently labeled nucleotides. As a nucleotide is held in the detection volume by the polymerase, a fluorescence pulse (orange) on illumination is produced and recorded, which identifies the base. (D) Not only the fluorescence color is registered but also the time between adjacent nucleotide incorporations, termed the interpulse duration, indirectly reflective of epigenetic modification. dsDNA = double-stranded DNA.

protruding ends, to which a motor protein acting as helicase is tightly bound (Figure 2B). Sequencing then proceeds in nanopores embedded in a synthetic bilayer, through which electric current flows constantly (Figure 2C). As unwinding DNA or RNA molecules translocate through the nanopore with the assistance of the motor protein, individual nucleotides are cut off by an exonuclease (Figure 2C), and current shifts due to characteristic disruptions are recorded in real time (Figure 2D).

Nanopore read lengths are confined only by the molecular lengths of the sample as opposed to the limit of the technology itself in SMRT sequencing. It is theoretically possible to sequence templates containing endless nucleotides using nanopore sequencing, supported by recently reported reads of up to 1 Mb.<sup>9</sup> In addition, MinION and SmidgION lead the trend of miniaturization and hold the potential to serve as hand-held and cost-effective devices for epidemiologic studies.

A notorious pitfall of ONT is that the raw read error rates can reach up to 15%.<sup>10</sup> Borrowing the mechanism of PacBio CCS, intramolecular-ligated nanopore consensus sequencing has emerged as an effective approach for constructing consensus single-molecule reads with a median accuracy of >97%.<sup>11</sup>

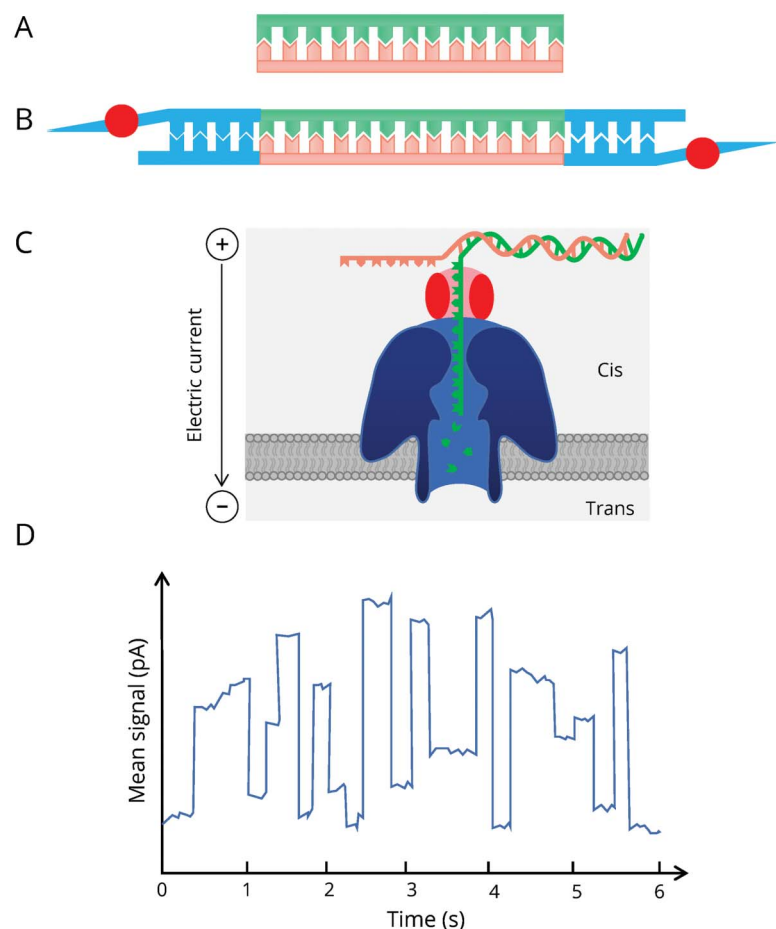
## Applications in Neurodegenerative Diseases

### Fragile X-Associated Tremor/Ataxia Syndrome

The human fragile X mental retardation 1 (*FMR1*) gene contains a polymorphic CGG trinucleotide repeat in its 5' UTR on the X chromosome, which causes fragile X syndrome (FXS) and fragile X-associated tremor/ataxia syndrome (FXTAS).<sup>12</sup> The premutation (PM) range of 55–200 repeats underlies FXTAS, whereas greater than 200 repeats are classified as full mutations, leading to FXS. Although these 2 neurologic diseases are caused by the same repeat motifs, they exhibit distinct clinical presentations and neuropathologic features. FXS is considered the most common cause of heritable form of intellectual disability. As a late-onset neurodegenerative disorder, FXTAS is characterized clinically by intention tremor and gait ataxia.<sup>13</sup> In addition, the PM range of CGG trinucleotide repeats is prone to expand into a full mutation in the next generation on maternal transmission, whereas the sandwiched presence of AGG interruptions in every 9 or 10 CGG repeats can reduce this instability.<sup>14</sup>

CCS reads of 36 and 95 CGG repeat-containing plasmids and a PCR-amplified allele harboring 750 CGG repeats on the

**Figure 2** Schematic Representation of Oxford Nanopore Technology



(A) dsDNA fragments often undergo an optional DNA repair step. (B) End-repaired DNA fragments are tagged with sequencing adapters (light blue) preloaded with a motor protein (red) on the 5' protruding ends. (C) The DNA template is loaded onto the flow cell containing thousands of nanopores (dark blue) embedded in a synthetic membrane (gray). The membrane divides the sequencing dimension into 2 compartments (cis and trans). Once the adapter inserts into the opening of the nanopore, the motor protein begins to unwind the dsDNA and drives the single-stranded DNA through the pore under the action of electric current. As the DNA molecule translocates through the pore, individual nucleotides will be cut off, which causes characteristic disruptions to the current. (D) Changes in current correspond to a readout known as a squiggle. dsDNA = double-stranded DNA.

PacBio platform clearly showed respective repeat sizes, size distributions, and small AGG interruptions,<sup>15</sup> which overcame the technical limits. More broadly, not only the effect of sequence context on polymerase kinetics but also the strand and position-specific influences of G on the polymerase's activity were confirmed by the variation in IPD,<sup>15</sup> thus laying the foundation for future investigations of epigenetic modifications and polymerase kinetics. To address the challenging detection and location of AGG interruptions in female PM carriers, SMRT provided a direct demonstration of AGG interruptions within the CGG repeat regions and separated a large number of reads containing AGG interruptions based on the X chromosomes from which they derived.<sup>16</sup> With the valuable information offered by SMRT, it is relatively precise to estimate the expansion risk of CGG repeats in female PM carriers, which can help them to select for preimplantation genetic diagnosis in case of high risk,<sup>17</sup> or invasive prenatal diagnosis within small risk,<sup>18</sup> to check the fragile X status of the fetus.

Since the successful sequencing of highly repetitive elements exceeding 2.25 kb of 100% GC content by using SMRT sequencing technology, LRS has been the focus of extensive

research and broadly applicable to a complete genetic and epigenetic analysis of expanded-repeat elements underlying many other neurodegenerative diseases. More importantly, LRS allows the improvement of genetic counseling in terms of assessing an accurate expansion risk of each allele individually for female PM carriers to achieve optimal fertility.

### Neuronal Intranuclear Inclusion Disease

Neuronal intranuclear inclusion disease (NIID) is a progressive and fatal neurodegenerative disease that presents with great clinical heterogeneity, including dementia, cerebellar ataxia, parkinsonism, peripheral neuropathy, and autonomic dysfunction. Patients with NIID pathologically manifest in the presence of eosinophilic hyaline intranuclear inclusions in the central, peripheral, and autonomic nervous systems, as well as somatic cells; therefore, skin biopsy efficiently determines the premortem diagnosis.<sup>19</sup>

Intensive efforts have been made to explore the genetic mechanisms of NIID since the first reported NIID-affected case. Recently, a novel GGC repeat expansion in the 5' UTR of the human specific gene *NOTCH2NLC* (formerly known as *NBPF19*) has been identified as the genetic cause in

Japanese NIID cases.<sup>20</sup> Although no pathogenic copy number variants were found in previous short-read whole-genome sequencing (WGS) data, long-read WGS with either PacBio RSII or the nanopore sequencer PromethION was subsequently performed. The result of these showed that all the 13 familial and 40 sporadic NIID cases carried GGC repeat expansions ranging from 71 to 183 and was confirmed by repeat-primed PCR (RP-PCR).<sup>20</sup> Neither of unaffected family members nor control individuals carried expanded GGC repeats, indicating family cosegregation. Moreover, nanopore sequencing with the Cas9-mediated enrichment system determined a possible association between a complex repeat structure composed of (GGA)<sub>n</sub> (GGC)<sub>n</sub> and the muscle-weakness-dominant phenotype of NIID (NIID-M),<sup>20</sup> which requires further validation in the clinical practice. Using ONT, the same causative mutations were also found to be responsible for juvenile-onset, adult-onset, familial, and sporadic NIID in the Chinese population,<sup>21,22</sup> thus further indicating a causative role for GGC repeat expansions in NIID. Intriguingly, GGC repeat expansion was not only observed in 2 cases of Alzheimer disease (AD) but also in 3 parkinsonism-affected families displaying typical AD and PD symptoms, respectively.<sup>22</sup> All of them were pathologically presented with eosinophilic intranuclear inclusions in dermal cells, which suggested that NIID has a higher prevalence than first thought, and may account for a portion of dementia and parkinsonism cases, termed NIID-related disorders.<sup>22</sup> In addition, a direct search strategy named TRhist was initially used to identify CGG repeats in patients with NIID, followed by SMRT to conclusively support the position of CGG repeats located in *NBPF19* (recently annotated as *NOTCH2NLC*).<sup>23</sup> Note that there are 4 additional paralogs besides *NOTCH2NLC*, all of which have sequences with extremely high identities (>99%).

By reanalyzing the data from LRS, no significant methylation difference was detected in GGC-expanded sequences and adjacent CpG islands between the NIID and control groups, nor was the expression of *NOTCH2NLC*.<sup>20-22</sup> In contrast, another study demonstrated that the expanded CGG repeats were likely hypermethylated.<sup>23</sup> The discovery of abnormal antisense transcripts,<sup>20</sup> together with unaltered expression levels of *NBPF19* transcripts irrespective of the hypermethylated status of *NBPF19*,<sup>23</sup> indicated a potential role for RNA in the molecular pathogenesis of NIID.

Taken together, the success of identifying causative variations for NIID strongly suggests that LRS can be a powerful tool for the discovery of disease-causing mutations, determination of complex repeat structures, analysis of methylated status, and performance of genetic diagnosis, despite the high identities among homologous genes or 100% GC rich in GGC or CGG repeat expansions.

### Oculopharyngodistal Myopathy

Oculopharyngodistal myopathy (OPDM) is an adult-onset degenerative neuromuscular disorder with unclear inheritance.

Patients affected with OPDM have a wide range of muscle involvement, including progressive ptosis, external ophthalmoplegia, and weakness of facial, pharyngeal, and distal limb muscles.

The identification of CGG repeat expansions in the 5' UTR of *LRP12* in 22 of 88 Japanese individuals with OPDM, termed OPDM1, deciphered a quarter of the genetic etiology in Japanese OPDM cases.<sup>23</sup> The finding that a large portion of patients with OPDM did not necessarily carry repeat expansions in *LRP12* requires further investigation. Although short-read WGS failed to find any likely causal mutations, long-read WGS using the ONT PromethION sequencer successfully identified abnormal expansion of GGC repeats in the 5' UTR of *GIPC1* within both familial and sporadic Chinese individuals with OPDM.<sup>24</sup> Noncoding GGC repeat expansions in *GIPC1* were also observed in a small fraction of Japanese OPDM cases,<sup>24</sup> further supporting GGC repeat expansions as the disease-causing variations of OPDM. Intriguingly, an independent strategy but also involving LRS in Chinese OPDM cases identified similar results with a slight difference in the repeat motif, that is, CGG.<sup>25</sup> There is the possibility that the same repeat expansions are described in different manners, such as based on various kinds of DNA reference sequences of *GIPC1*. *GIPC1* with repeat expansions accounted for 50% and 51.9% of Chinese patients with OPDM in these 2 studies, respectively, whereas only 3.6% of the Japanese population.<sup>24,25</sup> This demonstrates that noncoding trinucleotide repeat expansions in *GIPC1* may be the most frequent cause of the Chinese OPDM cohort.

Detection of 5-methylcytosine modification in the GGC repeat regions revealed no statistical differences between expanded and nonexpanded alleles or between affected individuals and healthy controls.<sup>24</sup> Despite the low level of methylation around the repeats and the abundance of unchanged protein, the increased level of *GIPC1* mRNA suggests that expanded repeats may have an impact on the process of transcription.<sup>24</sup>

To date, similar repeat motifs located in 2 distinct genes (*LRP12* and *GIPC1*) have been found in OPDM by virtue of LRS. This indicates the possibility of genetic heterogeneity and an essential role of repeat expansions themselves in the pathogenic mechanisms of OPDM, irrespective of the genes in which the expanded repeats are located. Furthermore, LRS can be used to promote further studies on the molecular mechanisms of OPDM, in addition to establishing the molecular diagnosis.

### Spinocerebellar Ataxias

Spinocerebellar ataxias (SCAs) are a group of genetically and phenotypically heterogeneous neurodegenerative diseases with a core symptom of progressive cerebellar ataxia, including gait ataxia, dysarthria, and oculomotor abnormalities. Different subtypes of SCAs exhibit various manifestations, such as epilepsy in SCA10,<sup>26</sup> which are conducive to differential diagnosis.

To date, approximately 50 genetically distinct types of SCAs (SCA1-SAC48) have been registered. Some are caused by CAG repeat expansions in coding regions, resulting in abnormally long polyglutamine (PolyQ) chains, termed PolyQ disorders. In addition, noncoding repeat expansions and conventional mutations have been attributed to other types of SCAs. This highlights the significance of genetic testing with high sensitivity and specificity to confirm the clinical diagnosis in the presence or absence of family history. Furthermore, the longer the repeat length, the earlier the age at onset, as illustrated by SCA3.<sup>27</sup> A second feature is the presentation of clinical anticipation due to the tendency for repeat expansions to change size, mainly enlarging further when transmitted to the next generation.<sup>27</sup> Therefore, later generations carrying larger expansions would manifest with an earlier age at onset and more severe symptoms.

SCA10, a subtype of the SCA family, is characterized by prominent cerebellar symptoms and seizures, such as focal seizures and generalized motor seizures. A large, noncoding adenine-thymine-thymine-cytosine-thymine (ATTCT) pentanucleotide repeat expansion in intron 9 of *ATXN10* on chromosome 22 causes SCA10.<sup>2</sup> Normal length of ATTCT is 10–32, and intermediate alleles containing 280–850 repeats present with reduced penetrance, whereas the pathogenic alleles ranging from 850 to 4,500 cause full penetrance.<sup>2</sup> It is well established that interruptions in the expanded ATTCT repeats are strongly associated with an increased prevalence of epileptic attack.<sup>28</sup> Despite the strikingly long read lengths of the SCA10 expansion, SMRT successfully sequenced across the entire span of the expansion, ranging from 2.5 to 4.4 kb in length, from 3 SCA10 patients with different clinical manifestations.<sup>29</sup> There were remarkably varied structures of interruption motifs among these 3 patients, such as ATTCC only in 1 patient, all of which were verified by Sanger sequencing to rule out sequencing or assembly errors.<sup>29</sup> Although only an extremely low percentage of these rare interruption motifs resided in the typical ATTCT repeat expansion regions, it is possible that repeat motif composition may act as a phenotypic modifier in SCA10, given symptomatic heterogeneity. Intriguingly, the tendency for interrupted SCA10 repeat expansions to contract on transmission correlated with earlier onset of the disease in successive generations of a family,<sup>29</sup> which is contrary to the usual anticipation, thus so-called paradoxical anticipation.<sup>30</sup> However, PCR amplification of repeat expansions to achieve sufficient quantities of DNA for input likely resulted in unpredictable errors.<sup>29</sup> To this end, the Cas9 capture approach paired with SMRT sequencing without prior amplification of genomic DNA was used in a Mexican family with multiple affected members, 4 of whom reported clinical manifestations of SCA10.<sup>31</sup> The proband uniquely presented with early-onset levodopa-responsive parkinsonism.<sup>31</sup> Of interest, a 1,400 repeat pattern that consisted of the typical ATTCT pentanucleotide repeats followed by an extra ATTCC interruption motif was observed in the 4 siblings with representative cerebellar ataxia and seizures,

whereas the sibling clinically diagnosed with PD carried 1,304 pure ATTCT repeats with virtually no repeat interruptions.<sup>31</sup> However, the phenomenon of paradoxical anticipation was not observed due to unavailable information of their father, which deserves further investigations to confirm the uncharacteristic theory and discuss the underlying mechanism.

As mentioned above, the genetic architecture of the SCA10 repeat expansion may serve as a phenotypic modifier, with the feat of SMRT revealing novel and various structures of interruption motifs,<sup>29,31</sup> and demonstrating a phenotype-genotype correlation between parkinsonism and *ATXN10*.<sup>31</sup> These discoveries may shed light on the research of many other types of SCA. Further studies are required to better understand how different compositions of interruption motifs serve as phenotypic modifiers, which relies heavily on advanced sequencing technology, particularly LRS.

## Huntington Disease

HD is an adult-onset neurodegenerative disorder with an autosomal dominant inheritance that is clinically characterized by progressive dystonia, chorea, cognitive dysfunction, and personality disorders.

As a notable PolyQ disorder, HD is caused by CAG repeat expansions in the huntingtin (*HTT*) gene on chromosome 4.<sup>32</sup> The repeat is up to 26 CAGs long in the normal population, whereas the intermediate alleles containing 27–35 repeats have a tendency to expand into the disease-causing range in the next generation. In patients with HD, the CAG repeats tend to expand above 36 units, such as 36–39 repeats presented with reduced penetrance and  $\geq 40$  with full penetrance.<sup>33</sup> Given the correlation between distinct sizes and disease phenotypes, fragment analysis is indicated in the clinical setting to determine repeat sizes. However, with the concern of PCR stutter and extreme GC content, there is a need for more advanced approaches. To this end, the CRISPR/Cas9 system paired with SMRT sequencing has been used, followed by a mapping-independent algorithm to analyze and visualize the repeat sequence in clinically relevant HD samples.<sup>34</sup> In addition to *HTT*, researchers have evaluated the multiplexing efficiency of additional targets involving *FMRI*, *ATXN10*, and chromosome 9's open reading frame 72 (*C9ORF72*), causing FXS, SCA10, amyotrophic lateral sclerosis, and frontotemporal dementia (ALS/FTD), respectively.<sup>34</sup> These support the feasibility of No-Amp targeted sequencing in multiplexing of different targets in the same run. More importantly, the variability in the number of CAG repeats within expanded alleles of the same patient indicated somatic mosaicism, especially for larger repeat expansions.<sup>34</sup> The sizes of CCG repeats flanking the CAG repeats have also been resolved, with 7 of the most frequent and the subsequent 10, consistent with previous data.<sup>35</sup> Nevertheless, the off-target effect on chromosome 9 could be explained by the presence of a single nucleotide polymorphism (SNP) increasing homology to the *ATXN10* gRNA.<sup>34</sup>

In summary, No-Amp targeted sequencing generates an exact estimation of the repeat count to confidently investigate somatic variations without PCR amplification and provides a more direct insight into the exact location of interruption motifs, such as in *FMRI*. Furthermore, it is practically possible to target and capture a large number of targets by well-designed gRNA and high-fidelity Cas9 enzymes, with the potentially wrong cleaved sites induced by SNPs taken into consideration.

### Benign Adult Familial Myoclonic Epilepsy

Benign adult familial myoclonic epilepsy (BAFME) is a slowly progressive disease mainly clinically characterized by cortical myoclonus tremors and infrequent epileptic seizures and electrophysiologically by giant somatosensory evoked potentials and long-latency cortical reflex.<sup>36</sup> It is generally known as familial cortical myoclonic tremor with epilepsy (FCMTE) in China and autosomal dominant cortical myoclonus and epilepsy or familial adult myoclonic epilepsy (FAME) in Europe.

Although the candidate genes have been mapped to several loci including 8q23.3-q24.1 (BAFME1),<sup>37</sup> 2p11.1-q12.2 (BAFME2),<sup>38</sup> 5p15.31-p15.1 (BAFME3),<sup>39</sup> 3q26.32-3q28 (BAFME4),<sup>40</sup> 16p21.1 (BAFME6),<sup>41</sup> and 4q32.1 (BAFME7),<sup>41</sup> none of the causative mutations had been reported until 2018.<sup>41</sup> A study identified TTTTA pentanucleotide repeats and extra TTTCA repeat sequences in intron 4 of *SAMD12* in Japanese families affected with BAFME1.<sup>41</sup> SMRT sequencing of bacterial artificial chromosome (BAC) clones from 2 patients exactly determined 2 repeat configurations: (TTTTA)<sub>exp</sub> (TTTCA)<sub>exp</sub> and (TTTTA)<sub>exp</sub> (TTTCA)<sub>exp</sub> (TTTTA)<sub>exp</sub>.<sup>41</sup> By means of ONT, the former pattern was clearly confirmed in Chinese FCMTE1 pedigrees.<sup>42</sup> In addition, the PacBio sequel system not only showed an approximately 4.6 kb *SAMD12* intronic repeat insertion but also estimated the size of insertion.<sup>43</sup> This is broadly consistent with the results of Southern blot analysis, proving the versatility and effectiveness of LRS. However, the inserted sequence was considered to be TTTCT instead of TTTCA as mentioned above,<sup>43</sup> probably owing to the high error rates of LRS. Of note, TTTTA-repeat expansion also contained impure repeat sequences of TTTTTA or TTTA,<sup>41</sup> possibly due to the susceptibility of SMRT sequencing to insertion and/or deletion. Given that contractions of expanded repeats may occur during BAC cloning, nanopore sequencing of genomic DNA was further performed to recount TTTTA and TTTCA motifs, the lengths of which were comparable and variable,<sup>41</sup> indicative of somatic instability and/or artifacts introduced by nanopore sequencing. With regard to neuropathologic findings, loss of Purkinje cells as the degenerative feature was observed in the cerebellar cortex of BAFME1 patient with homozygous mutations, and RNA foci was noted in the cortical neurons,<sup>41</sup> suggesting that transcription was a key step and RNA foci may be involved in the potential pathogenesis. Note that there were 2 families whose abnormal expansions of TTTCA and TTTTA

pentanucleotide repeats did not reside in *SAMD12*, but in *TNRC6* (BAFME6) and *RAPGEF2* (BAFME7), respectively.<sup>41</sup>

Recently, an intronic ATTTTC expansion in *STARD7* causing FAME2 and unstable TTTTA/TTTCA expansions in *MARCH6* responsible for FAME3 were reported by short-read WGS and RP-PCR.<sup>44,45</sup> These were also suggested to be somatically variable by both SMRT and ONT.<sup>44,45</sup> Nevertheless, the presence of ATTTTC or TTTCA pentamer did not affect the transcript or protein abundance of *STARD7* or *MARCH6* among expansion carriers and control individuals,<sup>44,45</sup> indicating that the repeat sequence alone may be pathogenic. More recently, SMRT sequencing detected intronic expansions of TTTTA and insertions of TTTCA repeats in *YEATS2* as the genetic etiology of BAFME4 in a Thai family.<sup>46</sup>

In conclusion, the expansion of noncoding TTTCA or ATTTTC repeats in multiple independent genes shared by all 6 types of BAFMEs (BAFME1, 2, 3, 4, 6, and 7) raised the possibility that repeat motifs themselves rather than genomic loci may play pivotal roles in the pathogenesis of BAFMEs, and the same expanded repeat motifs could result in overlapping clinical spectra of diseases. LRS is of diagnostic value with a sensitivity of 100% despite the low GC content and may assist to the development of therapeutic strategies targeting similar repeat motifs.

### Amyotrophic Lateral Sclerosis and Frontotemporal Dementia

ALS and FTD are fatal neurologic diseases with an autosomal dominant inheritance. In ALS, progressive muscle wasting, weakness, spasticity, and eventually general paralysis result from the degeneration of upper and lower motor neurons in the motor cortex, brainstem, and spinal cord.<sup>47</sup> As the second most common presenile dementia, FTD is characterized by personality changes and inappropriate behavior with relative preservation of memory owing to the degeneration of the superficial frontal cortex and anterior temporal lobes of the brain.<sup>48</sup> There is mounting evidence that ALS and FTD represent a clinicopathologic continuum of disease starting from the presence of frontotemporal dysfunction and motoneuron impairment in both ALS and FTD to the pathologic discovery of the transactive response DNA-binding protein with Mr 43 kDa throughout the central nervous system.<sup>49</sup>

To date, a large number of genes harboring mutations have been identified, such as *TARDBP*,<sup>50</sup> *FUS*,<sup>51</sup> and *OPTN*<sup>52</sup> in ALS and *GRN* in FTD.<sup>53</sup> Nevertheless, all mutations rarely coexist in either ALS or FTD. Linkage analysis of familial ALS and FTD implicated a major locus for both diseases on chromosome 9p21<sup>54</sup> and confirmed in sporadic cases,<sup>55</sup> further supporting the related association between these 2 diseases termed 9p-linked ALS/FTD. Furthermore, the identification of an expanded GGGGCC hexanucleotide repeat in the noncoding region of *C9ORF72* in 9p-linked ALS/FTD<sup>3</sup> genetically connected ALS with FTD. As the most common cause, this G<sub>4</sub>C<sub>2</sub> repeat expands hundreds to thousands of times in affected cases, whereas 2–30 in control individuals.<sup>56</sup> Given the extreme length, high GC content, and

tendency to form G-quadruplexes in both DNA and RNA,<sup>57,58</sup> the G<sub>4</sub>C<sub>2</sub> repeat expansion may be the least advantageous for sequencing. Nevertheless, both the PacBio and ONT sequencing platforms can traverse these challenging repeat expansions.<sup>59</sup> ONT MinION may more closely resemble expected read lengths because of its tighter distributions of read lengths than PacBio, indicating great promise for future ONT MinION applications. With regard to base calling accuracy, the PacBio RSII and the ONT MinION sequencers using consensus sequence attained approximately 99.8% and 26.6% accuracy, respectively,<sup>59</sup> showing the superiority of PacBio RSII. As for

affected carriers with G<sub>4</sub>C<sub>2</sub> repeat expansions, PacBio Sequel can accurately distinguish the expanded allele with 1,324 repeats from the unexpanded allele with 8 repeats, even with a low coverage (8×).<sup>59</sup> In addition, the PacBio No-Amp targeted sequencing method with a higher coverage (800×) was capable of closely estimating expansion size and measuring nucleotide content with the result of >99% GC content in the repeat regions.<sup>59</sup> However, there were G<sub>3</sub>C<sub>2</sub>, G<sub>4</sub>C<sub>1</sub>, and nonGC interruptions in addition to G<sub>4</sub>C<sub>2</sub> motifs,<sup>59</sup> possibly due to insertions and deletions as the most common error in PacBio sequencing, which requires cautious interpretation and larger studies.

**Table** Summary of Long-Read Sequencing Applications in Neurodegenerative Diseases

Disease	Mutation type	Gene locus	Platform	Method	Coverage	GC content	Reference
FXTAS	CGG <sub>exp</sub>	<i>FMR1</i>	PacBio RS	PCR amplicon	3.2–10.7×	100%	2013 <sup>15</sup>
	CGG <sub>exp</sub>	<i>FMR1</i>	PacBio RS II	PCR amplicon	NA	NA	2018 <sup>16</sup>
NIID	GGC <sub>exp</sub>	<i>NOTCH2NLC</i>	PacBio RS II and ONT PromethION	Whole genome	8–25×	100%	2019 <sup>20</sup>
	GGC <sub>exp</sub>	<i>NOTCH2NLC</i>	ONT MinION	Cas9-mediated enrichment sequencing	100–1795×	100%	2019 <sup>20</sup>
	GGC <sub>exp</sub>	<i>NOTCH2NLC</i>	ONT	Whole genome	15×	NA	2019 <sup>21</sup>
	GGC <sub>exp</sub>	<i>NOTCH2NLC</i>	ONT	Whole genome	NA	NA	2019 <sup>22</sup>
	CGG <sub>exp</sub>	<i>NBPF19</i> <sup>a</sup>	PacBio	Whole genome	NA	98%–99%	2019 <sup>23</sup>
OPDM	GGC <sub>exp</sub>	<i>GIPC1</i>	ONT PromethION	Whole genome	18.6×	NA	2020 <sup>24</sup>
	CGG <sub>exp</sub>	<i>GIPC1</i>	ONT PromethION	Whole genome	NA	NA	2020 <sup>25</sup>
SCA10	ATTCT <sub>exp</sub>	<i>ATXN10</i>	PacBio RS	PCR amplicon	NA	NA	2015 <sup>29</sup>
	ATTCT <sub>exp</sub>	<i>ATXN10</i>	PacBio	No-Amp targeted sequencing	NA	NA	2017 <sup>31</sup>
PD	ATTCT <sub>exp</sub>	<i>ATXN10</i>	PacBio	No-Amp targeted sequencing	NA	NA	2017 <sup>31</sup>
HD	CAG <sub>exp</sub>	<i>HTT</i>	PacBio	No-Amp targeted sequencing	NA	NA	2018 <sup>34</sup>
BAFME1	TTTCA/TTTTA <sub>exp</sub>	<i>SAMD12</i>	PacBio RS II ONT MinION	BAC clone and whole genome	NA	NA	2018 <sup>41</sup>
	TTTCA/TTTTA <sub>exp</sub>	<i>SAMD12</i>	PacBio and ONT	Whole genome	10×	NA	2019 <sup>42</sup>
	TTTCT <sub>exp</sub> <sup>b</sup>	<i>SAMD12</i>	PacBio Sequel	Whole genome	7×	NA	2019 <sup>43</sup>
FAME2	ATTTC <sub>exp</sub>	<i>STARD7</i>	PacBio and ONT	Whole genome	NA	NA	2019 <sup>44</sup>
FAME3	TTTCA/TTTTA <sub>exp</sub>	<i>MARCH6</i>	ONT GridION	Whole genome	6–10×	NA	2019 <sup>45</sup>
BAFME4	TTTCA/TTTTA <sub>exp</sub>	<i>YEATS2</i>	PacBio Sequel	Whole genome	8×	NA	2019 <sup>46</sup>
ALS/FTD	GGGGCC <sub>exp</sub>	<i>C9ORF72</i>	ONT MinION PacBio Sequel	Whole genome No-Amp targeted sequencing	8× 800×	>99%	2018 <sup>59</sup>

Abbreviations: ALS/FTD = amyotrophic lateral sclerosis/frontotemporal dementia; *ATXN10* = ataxin 10; BAC = bacterial artificial chromosome; BAFME1 = benign adult familial myoclonic epilepsy type 1; BAFME4 = benign adult familial myoclonic epilepsy type 4; *C9ORF72* = chromosome 9's open reading frame 72; exp = expansion; FAME2 = familial adult myoclonic epilepsy type 2; FAME3 = familial adult myoclonic epilepsy type 3; *FMR1* = fragile X mental retardation 1; FXTAS = fragile X-associated tremor/ataxia syndrome; *GIPC1* = GIPC PDZ domain containing family member 1; HD = Huntington disease; *HTT* = huntingtin; *MARCH6* = membrane-associated ring finger (C3HC4) 6; NA = not available; *NBPF19* = neuroblastoma break point family member 19; NIID = neuronal intranuclear inclusion disease; No-Amp targeted sequencing = no-amplification targeted sequencing; *NOTCH2NLC* = notch 2 amino-terminal-like protein C; ONT = Oxford Nanopore Technologies; OPDM = oculopharyngodistal myopathy; PacBio = Pacific Biosciences; PD = Parkinson disease; *SAMD12* = sterile alpha motif domain containing 12; SCA10 = spinocerebellar ataxia type 10; *STARD7* = StAR-related lipid transfer domain containing 7; *YEATS2* = YEATS domain containing 2.

<sup>a</sup> *NBPF19* is recently annotated as *NOTCH2NLC*.

<sup>b</sup> The inserted sequence was suggested to be (TTTCT), rather than (TTTCA) as reported previously, possibly due to the high sequencing error rate of long-read sequencing technologies.



In general, LRS is ideal for characterizing repeat expansions, such as their locations, sizes, and content, despite the extraordinary stretches of repeats with pure GC content. Moreover, No-Amp targeted sequencing allows investigators to achieve a considerably in-depth degree of coverage and accurately elucidate the nucleotide-level features of known and undiscovered mutations. LRS shows great promise in both research and clinical settings. However, the errors, such as insertions and deletions, should be taken seriously.

## Conclusions and Prospects

To date, LRS has already proven their outstanding ability to identify causative repeat expansions, recognize repeat configurations, estimate repeat sizes, and detect base modifications in neurodegenerative diseases. We summarize the extensive applications of LRS in neurodegenerative diseases as mentioned above (Table). The modification of repeat configurations to clinical phenotypes, the definition of threshold between pathogenic and normal ranges based on repeat sizes, and the silencing of transcription and translation due to epigenetic modifications all contribute to our expanding knowledge of neurodegenerative diseases. We may safely draw the conclusion that repeat expansions underlying neurodegenerative diseases are good targets for LRS, thanks to its extremely long read lengths, while NGS may result in negative results.

Although noncoding repeat expansions have been detected by LRS to cause a number of neurodegenerative diseases over the last 3 years, the pathogenic mechanism remains largely unclear. The presence of antisense transcripts and RNA foci containing the repeat transcripts implicates the pathogenicity of RNA, known as toxic gain of function. For example, antisense transcripts have the capacity to regulate gene expression by binding to and sealing off complementary RNAs. In addition, repeat-associated non-adenine-uracil-guanine (non-AUG) translation may lead to aggregation of aberrant protein products. Therefore, further research is needed to investigate the potential pathway at the transcriptional and translational level, in addition to reaching to gene level.

In this review, LRS has been shown to mainly characterize repeat expansions; its ability is not confined to only 1 type of mutation. The successful identification of a pathogenic structural variant (SV) in a patient having the Carney complex using LRS indicates that LRS is capable of discovering causal SVs,<sup>60</sup> and this is also applicable to neurodegenerative diseases. In terms of clinical utility, LRS has great potential to secure genetic diagnosis of neurodegenerative diseases, contribute to differential diagnoses, and provide vital information for genetic counseling. In addition, the promising prospects of LRS in medical research lie in detecting novel disease-causing and disease-modifying variations, exploring the underlying molecular mechanisms, and developing targeted therapeutic strategies.

Nevertheless, there still remains much room for improvement with regard to variant calling accuracy, bioinformatic analysis, cost of deciphering the human genome, and throughput. First, the raw read errors were fundamentally derived from interrogating a single DNA molecule by the polymerase or exonuclease. Deletions, insertions, and mismatches were primarily attributed to unlabeled nucleotide contamination, double count mediated by the failure of a base to translocate through the nanopore, and spectral misassignments of the dyes, respectively.<sup>5</sup> In addition, single nucleotide discrimination by ONT remains difficult, given that the polynucleotides within the nanopore rather than 1 single nucleotide result in the current disruption, suggesting the superiority of PacBio over ONT in terms of variant calling accuracy. Purer fluorescently labeled nucleotides, spectrally compatible dyes, and CCS construction are expected to improve read accuracy. Nevertheless, given the intrinsic restriction of a single molecule, the maximal accuracy of LRS may barely match that of NGS. Beyond the read accuracy challenges for LRS, sequencing data analysis will be conducted. For example, unique polymerase kinetics generated by epigenetic modifications and various forms of DNA damage allow discrimination between modified nucleotides in the DNA template; however, there is a demand for unmodified reference data. As we work on refining bioinformatics algorithms, de novo modification profiling may be achievable. The last major issue relates to the relatively low throughput and high cost, which was around several hundred dollars per gigabyte base pair. With the active development of LRS, it is anticipated that by virtue of sophisticated mathematical algorithms, LRS-based genome-wide sequencing could accurately discover all genetic variants in an individual's genome at a reasonable cost in the near future.

In conclusion, LRS has drastically revolutionized sequencing technology and will advance our understanding of the genetic etiology and molecular pathogenesis of neurodegenerative diseases, thereby further facilitating effective treatment.

## Study Funding

Study funded by the National Natural Science Foundation of China (Grants U190420029, 91849115, and 81530037 to Y Xu, Grants 81771290 and 81974211 to C Shi, and Grant 81901300 to C Mao), the National Key Research and Development Program of China (Grant 2017YFA0105003 to Y Xu), and the Scientific and Technological Project of Henan Province (Grant SBGJ202003020 to C Mao).

## Disclosure

Y Su and L Fan report no disclosures relevant to the manuscript. C Shi: research support, government entities: National Natural Science Foundation of China (Grants 81771290 and 81974211). T Wang, H Zheng, H Luo, S Zhang, Z Hu, Y Fan, Y Dong, and J Yang report no disclosures relevant to the manuscript. C Mao: research support, government entities: National Natural Science Foundation of China (Grant 81901300) and the Scientific and Technological Project of Henan Province (Grant SBGJ202003020). Y Xu: research support, government entities: National Natural

Science Foundation of China (Grants U190420029, 91849115, and 81530037) and National Key Research and Development Program of China (Grant 2017YFA0105003). Go to [Neurology.org/N](http://Neurology.org/N) for full disclosures.

## Publication History

Received by *Neurology* February 4, 2021. Accepted in final form June 23, 2021.

## Appendix Authors

Name	Location	Contribution
<b>Yun Su, MD</b>	Zhengzhou University, Zhengzhou, Henan, China	Study concept and design; analysis and interpretation of data; and creation of figures and manuscript draft
<b>Liyuan Fan, MD</b>	Zhengzhou University, Zhengzhou, Henan, China	Study concept and design; analysis and interpretation of data; and creation of figures and manuscript draft
<b>Changhe Shi, MD, PhD</b>	Zhengzhou University, Zhengzhou, Henan, China	Study concept and design; analysis and interpretation of data; and creation of figures and manuscript draft
<b>Tai Wang, MD</b>	Zhengzhou University, Zhengzhou, Henan, China	Analysis and interpretation of data
<b>Huimin Zheng, MD</b>	Zhengzhou University, Zhengzhou, Henan, China	Analysis and interpretation of data
<b>Haiyang Luo, MD, PhD</b>	Zhengzhou University, Zhengzhou, Henan, China	Analysis and interpretation of data
<b>Shuo Zhang, MD, PhD</b>	Zhengzhou University, Zhengzhou, Henan, China	Analysis and interpretation of data
<b>Zhengwei Hu, MD, PhD</b>	Zhengzhou University, Zhengzhou, Henan, China	Analysis and interpretation of data
<b>Yu Fan, MD, PhD</b>	Zhengzhou University, Zhengzhou, Henan, China	Analysis and interpretation of data
<b>Yali Dong, MD</b>	Zhengzhou University, Zhengzhou, Henan, China	Analysis and interpretation of data
<b>Jing Yang, MD, PhD</b>	Zhengzhou University, Zhengzhou, Henan, China	Analysis and interpretation of data
<b>Chengyuan Mao, MD, PhD</b>	Zhengzhou University, Zhengzhou, Henan, China	Study concept, design, analysis, and interpretation of data; study supervision; and critical review of the manuscript for intellectual content
<b>Yuming Xu, MD, PhD</b>	Zhengzhou University, Zhengzhou, Henan, China	Study concept, design, analysis, and interpretation of data; study supervision; and critical review of the manuscript for intellectual content

## References

- MacDonald ME, Ambrose CM, Duyao MP, et al. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. The Huntington's Disease Collaborative Research Group. *Cell*. 1993;72(6):971-983.
- Matsuura T, Yamagata T, Burgess DL, et al. Large expansion of the ATTCT pentanucleotide repeat in spinocerebellar ataxia type 10. *Nat Genet*. 2000;26(2):191-194.
- DeJesus-Hernandez M, Mackenzie IR, Boeve BF, et al. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron*. 2011;72(2):245-256.
- Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet*. 2011;13(1):36-46.
- Eid J, Fehr A, Gray J, et al. Real-time DNA sequencing from single polymerase molecules. *Science*. 2009;323(5910):133-138.
- Clarke J, Wu HC, Jayasinghe L, Patel A, Reid S, Bayley H. Continuous base identification for single-molecule nanopore DNA sequencing. *Nat Nanotechnol*. 2009;4:265-270.
- Hafford-Tear NJ, Tsai YC, Sadan AN, et al. CRISPR/Cas9-targeted enrichment and long-read sequencing of the Fuchs endothelial corneal dystrophy-associated TCF4 triplet repeat. *Genet Med*. 2019;21:2092-2102.
- Flusberg BA, Webster DR, Lee JH, et al. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods*. 2010;7:461-465.
- Jain M, Koren S, Miga KH, et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat Biotechnol*. 2018;36(4):338-345.
- Jain M, Tyson JR, Loose M, et al. MinION analysis and reference consortium: phase 2 data release and analysis of R9.0 chemistry. *PLoS Res*. 2017;6:760.
- Li C, Chng KR, Boey EJ, Ng AH, Wilm A, Nagarajan N. INC-Seq: accurate single molecule reads using nanopore sequencing. *Gigascience*. 2016;5(1):34.
- Jacquemont S, Hagerman RJ, Leehey M, et al. Fragile X premutation tremor/ataxia syndrome: molecular, clinical, and neuroimaging correlates. *Am J Hum Genet*. 2003;72(4):869-878.
- Hagerman RJ, Leehey M, Heinrichs W, et al. Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X. *Neurology*. 2001;57(1):127-130.
- Yrigollen CM, Durbin-Johnson B, Gane L, et al. AGG interruptions within the maternal FMR1 gene reduce the risk of offspring with fragile X syndrome. *Genet Med*. 2012;14(8):729-736.
- Loomis EW, Eid JS, Peluso P, et al. Sequencing the unsequenceable: expanded CGG-repeat alleles of the fragile X gene. *Genome Res*. 2013;23(1):121-128.
- Ardui S, Race V, de Ravel T, et al. Detecting AGG interruptions in females with a FMR1 premutation by long-read single-molecule sequencing: a 1 year clinical experience. *Front Genet*. 2018;9:150.
- Apeessos A, Abou-Sleiman PM, Harper JC, Delhanty JD. Preimplantation genetic diagnosis of the fragile X syndrome by use of linked polymorphic markers. *Prenat Diagn*. 2001;21(6):504-511.
- Biancalana V, Glaeser D, McQuaid S, Steinbach P. EMQN best practice guidelines for the molecular genetic testing and reporting of fragile X syndrome and other fragile X-associated disorders. *Eur J Hum Genet*. 2015;23(4):417-425.
- Sone J, Tanaka F, Koike H, et al. Skin biopsy is useful for the antemortem diagnosis of neuronal intranuclear inclusion disease. *Neurology*. 2011;76(16):1372-1376.
- Sone J, Mitsuhashi S, Fujita A, et al. Long-read sequencing identifies GGC repeat expansions in NOTCH2NLC associated with neuronal intranuclear inclusion disease. *Nat Genet*. 2019;51(8):1215-1221.
- Deng J, Gu M, Miao Y, et al. Long-read sequencing identified repeat expansions in the 5'UTR of the NOTCH2NLC gene from Chinese patients with neuronal intranuclear inclusion disease. *J Med Genet*. 2019;56(11):758-764.
- Tian Y, Wang JL, Huang W, et al. Expansion of human-specific GGC repeat in neuronal intranuclear inclusion disease-related disorders. *Am J Hum Genet*. 2019;105(1):166-176.
- Ishiura H, Shibata S, Yoshimura J, et al. Noncoding CGG repeat expansions in neuronal intranuclear inclusion disease, oculopharyngodistal myopathy and an overlapping disease. *Nat Genet*. 2019;51:1222-1232.
- Deng J, Yu J, Li P, et al. Expansion of GGC repeat in GIPC1 is associated with oculopharyngodistal myopathy. *Am J Hum Genet*. 2020;106(6):793-804.
- Xi J, Wang X, Yue D, et al. 5' UTR CGG repeat expansion in GIPC1 is associated with oculopharyngodistal myopathy. *Brain*. 2021;144(2):601-614.
- Domingues BMD, Nascimento FA, Meira AT, et al. Clinical and genetic evaluation of spinocerebellar ataxia type 10 in 16 Brazilian families. *Cerebellum*. 2019;18(5):849-854.
- Maciel P, Gaspar C, DeStefano AL, et al. Correlation between CAG repeat length and clinical features in Machado-Joseph disease. *Am J Hum Genet*. 1995;57(1):54-61.
- Matsuura T, Fang P, Pearson CE, et al. Interruptions in the expanded ATTCT repeat of spinocerebellar ataxia type 10: repeat purity as a disease modifier? *Am J Hum Genet*. 2006;78(1):125-129.
- McFarland KN, Liu J, Landrian I, et al. SMRT sequencing of long tandem nucleotide repeats in SCA10 reveals unique insight of repeat expansion structure. *PLoS One*. 2015;10(8):e0135906.
- McFarland KN, Liu J, Landrian I, et al. Paradoxical effects of repeat interruptions on spinocerebellar ataxia type 10 expansions and repeat instability. *Eur J Hum Genet*. 2013;21(11):1272-1276.
- Schüle B, McFarland KN, Lee K, et al. Parkinson's disease associated with pure ATXN10 repeat expansion. *NPJ Parkinsons Dis*. 2017;3:27.
- A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. The Huntington's Disease Collaborative Research Group. *Cell*. 1993;72(6):971-983.

33. Losekoot M, van Belzen MJ, Seneca S, Bauer P, Stenhouse SA, Barton DE. EMQN/CMGS best practice guidelines for the molecular genetic testing of Huntington disease. *Eur J Hum Genet.* 2013;21(5):480-486.
34. Höijer J, Tsai YC, Clark TA, et al. Detailed analysis of HTT repeat elements in human blood using targeted amplification-free long-read sequencing. *Hum Mutat.* 2018;39(9):1262-1272.
35. Agostinho Lde A, Rocha CF, Medina-Acosta E, et al. Haplotype analysis of the CAG and CCG repeats in 21 Brazilian families with Huntington's disease. *J Hum Genet.* 2012;57:796-803.
36. Cen Z, Huang C, Yin H, et al. Clinical and neurophysiological features of familial cortical myoclonic tremor with epilepsy. *Mov Disord.* 2016;31(11):1704-1710.
37. Mikami M, Yasuda T, Terao A, et al. Localization of a gene for benign adult familial myoclonic epilepsy to chromosome 8q23.3-q24.1. *Am J Hum Genet.* 1999;65(3):745-751.
38. Guerrini R, Bonanni P, Patrignani A, et al. Autosomal dominant cortical myoclonus and epilepsy (ADCME) with complex partial and generalized seizures: a newly recognized epilepsy syndrome with linkage to chromosome 2p11.1-q12.2. *Brain.* 2001;124(3):2459-2475.
39. Depienne C, Magnin E, Bouteiller D, et al. Familial cortical myoclonic tremor with epilepsy: the third locus (FCMTE3) maps to 5p. *Neurology.* 2010;74(24):2000-2003.
40. Yeetong P, Ausavarat S, Bhidayasiri R, et al. A newly identified locus for benign adult familial myoclonic epilepsy on chromosome 3q26.32-3q28. *Eur J Hum Genet.* 2013;21(2):225-228.
41. Ishiura H, Doi K, Mitsui J, et al. Expansions of intronic TTTCA and TTTTA repeats in benign adult familial myoclonic epilepsy. *Nat Genet.* 2018;50(4):581-590.
42. Zeng S, Zhang MY, Wang XJ, et al. Long-read sequencing identified intronic repeat expansions in SAMD12 from Chinese pedigrees affected with familial cortical myoclonic tremor with epilepsy. *J Med Genet.* 2019;56(4):265-270.
43. Mizuguchi T, Toyota T, Adachi H, Miyake N, Matsumoto N, Miyatake S. Detecting a long insertion variant in SAMD12 by SMRT sequencing: implications of long-read whole-genome sequencing for repeat expansion diseases. *J Hum Genet.* 2019;64:191-197.
44. Corbett MA, Kroes T, Veneziano L, et al. Intronic ATTTTC repeat expansions in STARD7 in familial adult myoclonic epilepsy linked to chromosome 2. *Nat Commun.* 2019;10(1):4920.
45. Florian RT, Kraft F, Leitão E, et al. Unstable TTTTA/TTTCA expansions in MARCH6 are associated with familial adult myoclonic epilepsy type 3. *Nat Commun.* 2019;10(1):4919.
46. Yeetong P, Pongpanich M, Srichomthong C, et al. TTTCA repeat insertions in an intron of YEATS2 in benign adult familial myoclonic epilepsy type 4. *Brain.* 2019;142(11):3360-3366.
47. Owens B. Amyotrophic lateral sclerosis. *Nature.* 2017;550:S105.
48. Clinical and neuropathological criteria for frontotemporal dementia. The Lund and Manchester Groups. *J Neurol Neurosurg Psychiatry.* 1994;57(4):416-418.
49. Neumann M, Sampathu DM, Kwong LK, et al. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science.* 2006;314(5796):130-133.
50. Sreedharan J, Blair IP, Tripathi VB, et al. TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. *Science.* 2008;319(5870):1668-1672.
51. Kwiatkowski TJ Jr, Bosco DA, Leclerc AL, et al. Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. *Science.* 2009;323(5918):1205-1208.
52. Maruyama H, Morino H, Ito H, et al. Mutations of optineurin in amyotrophic lateral sclerosis. *Nature.* 2010;465(7295):223-226.
53. Baker M, Mackenzie IR, Pickering-Brown SM, et al. Mutations in progranulin cause tau-negative frontotemporal dementia linked to chromosome 17. *Nature.* 2006;442(7105):916-919.
54. Morita M, Al-Chalabi A, Andersen PM, et al. A locus on chromosome 9p confers susceptibility to ALS and frontotemporal dementia. *Neurology.* 2006;66(6):839-844.
55. Shatunov A, Mok K, Newhouse S, et al. Chromosome 9p21 in sporadic amyotrophic lateral sclerosis in the UK and seven other countries: a genome-wide association study. *Lancet Neurol.* 2010;9(10):986-994.
56. van Blitterswijk M, DeJesus-Hernandez M, Niemantsverdriet E, et al. Association between repeat sizes and clinical and pathological characteristics in carriers of C9orf72 repeat expansions (Xpansize-72): a cross-sectional cohort study. *Lancet Neurol.* 2013;12(10):978-988.
57. Fratta P, Mizielinska S, Nicoll AJ, et al. C9orf72 hexanucleotide repeat associated with amyotrophic lateral sclerosis and frontotemporal dementia forms RNA G-quadruplexes. *Scientific Rep.* 2012;2:1016.
58. Haeusler AR, Donnelly CJ, Periz G, et al. C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature.* 2014;507:195-200.
59. Ebbert MTW, Farrugia SL, Sens JP, et al. Long-read sequencing across the C9orf72 'GGGGCC' repeat expansion: implications for clinical use and genetic discovery efforts in human disease. *Mol Neurodegeneration.* 2018;13:46.
60. Merker JD, Wenger AM, Sneddon T, et al. Long-read genome sequencing identifies causal structural variation in a Mendelian disease. *Genet Med.* 2018;20(1):159-163.



## Without Borders – A curated collection featuring advances in global neurology

This *Neurology*<sup>®</sup> special interest website is the go-to source for tracking science and politics of neurology beyond the United States, featuring up-to-the-minute blogs, scholarly perspectives, and academic review of developments and research from *Neurology* journals and other sources. Curated by Gretchen L. Birbeck, MD, MPH.

Expand your world view at [Neurology.org/woborders](https://www.neurology.org/woborders).