

BRIEF REPORT

The ZFH3 GGC Repeat Expansion Underlying Spinocerebellar Ataxia Type 4 has a Common Ancestral Founder

Zhongbo Chen, MRCP, PhD,^{1,2*}  Pilar Alvarez Jerez, MSc,^{3,4} Claire Anderson, PhD,^{5,6} Martin Paucar, MD, PhD,^{7,8} Jasmine Lee, MSc,⁹ Daniel Nilsson, PhD,^{10,11} Hannah Macpherson, MSc,^{3,5,6} Annarita Scardamaglia, BSc,⁹ Kylie Montgomery, BSc,^{5,6,9} John Hardy, FMedSci, PhD,^{3,12,13,14,15} Andrew B. Singleton, PhD,^{4,16} Arianna Tucci, MD, PhD,¹⁷ Katherine D. Mathews, MD, PhD,^{18,19} Ying-Hui Fu, PhD,^{20,21,22,23} Martin Engvall, MD, PhD,^{24,25} José Laffita-Mesa, MD, PhD,^{8,26} Inger Nennesmo, MD, PhD,²⁷ Anna Wedell, MD, PhD,²⁵ Louis J. Ptáček, MD, PhD,^{19,20,21,22} Cornelis Blauwendraat, PhD,^{4,16}  Emil K. Gustavsson, PhD,^{5,6}  Per Svenningsson, MD, PhD,^{7,8}  Mina Ryten, MD, PhD,^{5,6,28,29} and Henry Houlden, MD, PhD^{9*} 

¹Department of Clinical and Movement Neuroscience, Queen Square Institute of Neurology, University College London (UCL), London, UK ²The Francis Crick Institute, London, UK ³Department of Neurodegenerative Disease, Queen Square Institute of Neurology, UCL, London, UK ⁴Center for Alzheimer's and Related Dementias, National Institute on Aging and National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland, USA ⁵Department of Genetics and Genomic Medicine, Great Ormond Street Institute of Child Health, UCL, London, UK ⁶NIHR Great Ormond Street Hospital Biomedical Research Centre, UCL, London, UK ⁷Department of Neurology, Karolinska University Hospital, Stockholm, Sweden ⁸Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden ⁹Department of Neuromuscular Disease, Queen Square Institute of Neurology, UCL, London, UK ¹⁰Department of Clinical Genetics, Karolinska University Hospital, Stockholm, Sweden ¹¹Science for Life Laboratory, Department of Molecular Medicine and Surgery,

Karolinska Institutet, Stockholm, Sweden ¹²Reta Lila Weston Institute, Queen Square Institute of Neurology, UCL, London, UK ¹³UK Dementia Research Institute, UCL, London, UK ¹⁴NIHR University College London Hospitals Biomedical Research Centre, London, UK ¹⁵Institute for Advanced Study, The Hong Kong University of Science and Technology, Hong Kong, China ¹⁶Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, Maryland, USA ¹⁷William Harvey Research Institute, Queen Mary University of London, London, UK ¹⁸Department of Pediatrics, University of Iowa Carver College of Medicine, Iowa City, Iowa, USA ¹⁹Department of Neurology, University of Iowa Carver College of Medicine, Iowa City, Iowa, USA ²⁰Department of Neurology, University of California San Francisco, San Francisco, California, USA ²¹Institute for Human Genetics, University of California San Francisco, San Francisco, California, USA ²²Weill Institute for Neuroscience, University of California San Francisco, San Francisco, California, USA ²³Kavli Institute for Fundamental Neuroscience, University of California San Francisco, San Francisco, California, USA ²⁴Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden ²⁵Centre for Inherited Metabolic Diseases, Karolinska University Hospital, Stockholm, Sweden ²⁶Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden ²⁷Department of Oncology-Pathology, Karolinska Institutet, Stockholm, Sweden ²⁸UK Dementia Research Institute at the University of Cambridge, Cambridge, UK ²⁹Department of Clinical Neurosciences, School of Clinical Medicine, University of Cambridge, Cambridge, UK

ABSTRACT: Background: The identification of a heterozygous exonic GGC repeat expansion in ZFH3 underlying spinocerebellar ataxia type 4 (SCA4) has solved a 25-year diagnostic conundrum. We used adaptive long-read sequencing to decipher the pathogenic expansion in the index Utah family and an unrelated family from Iowa of Swedish ancestry. Contemporaneous to our discovery, other groups identified the same repeat expansion in affected individuals from Utah, Sweden, and Germany, highlighting the current pivotal time for detection of novel repeat expansion disorders. **Methods:** Given that the pathogenic repeat expansion is rare on a population level, we proposed a common ancestor across all families. Here, we employed

This is an open access article under the terms of the [Creative Commons Attribution](#) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

***Correspondence to:** Dr. Zhongbo Chen, Clinical and Movement Neuroscience, Queen Square Institute of Neurology, University College London (UCL), Queen Square, London, WC1N 3BG, UK. E-mail: zhongbo.chen@ucl.ac.uk; Dr. Henry Houlden, Department of Neuromuscular Disease, Queen Square Institute of Neurology, University College London (UCL), London, WC1N 3BG, UK. E-mail: h.houlden@ucl.ac.uk

Pilar Alvarez Jerez, Claire Anderson, and Martin Paucar contributed equally to the work.

Emil K. Gustavsson, Per Svenningsson, Mina Ryten, and Henry Houlden contributed equally to the work.

Relevant conflicts of interest/financial disclosures: The authors declare that there are no conflicts of interest relevant to this work.

Funding agencies: This study was supported by The Michael J. Fox Foundation (Z.C.). This work utilized the computational resources of the National Institutes of Health (NIH) High Performance Computing (HPC) Biowulf cluster (<http://hpc.nih.gov>). This work was supported in part by the Intramural Research Program of the NIH including: the Center for Alzheimer's and Related Dementias, within the Intramural Research Program of the National Institute on Aging and the National Institute of Neurological Disorders and Stroke (1ZIAAG000538-03 and 1ZIAAG000542-01).

Received: 7 August 2024; **Revised:** 14 October 2024; **Accepted:** 18 November 2024

Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/mds.30077

targeted long-read sequencing through adaptive sampling, enriching for the chr16q22 region of interest.

Results: Using phased sequencing results from individuals from Utah, Iowa, and Southern Sweden, we confirmed a common ~2000-year-old ancestral haplotype harbouring the repeat expansion.

Conclusion: This study provides further insight into the genetic architecture of SCA4. © 2024 The Author (s). *Movement Disorders* published by Wiley Periodicals LLC on behalf of International Parkinson and Movement Disorder Society.

Key Words: ataxia; long-read sequencing; repeat expansion disorder; haplotype; spinocerebellar ataxia type 4

Introduction

The genetic cause of spinocerebellar ataxia type 4 (SCA4), a late-onset neurodegenerative condition characterized by cerebellar ataxia and peripheral neuropathy, had evaded diagnosis for more than 25 years despite its clear linkage to chromosome 16q22 defined within a large family of Swedish ancestry residing in Utah (USA).¹ Advances in long-read sequencing technologies, bioinformatic tools, and an improved understanding of population-level structural variation enabled us to recently decipher the causative heterozygous exonic GGC repeat expansion in *ZFH3* in the index Utah family and an unrelated family from Iowa of Swedish ancestry.² Our work harnessed targeted long-read sequencing to detect novel repeat expansions in a variant-agnostic manner. This discovery epitomized both the difficulties and the emerging possibilities for diagnosis of such challenging neurogenetic disorders, which collectively carry a significant disease burden.

Parallel to our work, others have also independently uncovered the causative expansion in individuals with SCA4 from eight separate kindreds from Southern Sweden,^{3,4} Germany, and related individuals to our cohort from Utah.⁵ These individuals with SCA4 presented with variable phenotypes and repeat sizes, culminating in a recent study that showed inverse correlation between repeat length and age-of-onset in two German families.⁶ These contemporaneous findings of the GGC repeat expansion across all groups reflects the current pivotal time for novel repeat expansion discovery.^{7,8}

It has been proposed that a founder event underlies the repeat expansion haplotype in individuals from Utah and Germany.⁵ We propose that a common Swedish ancestral haplotype containing the repeat expansion is common to all families with SCA4, which has important implications for understanding the genetic architecture of disease.

Here, we were able to assess the detailed phased haplotype structure using long-read sequencing. We compared

the repeat-containing haplotype of individuals with SCA4 from the index Utah family,¹ the affected family from Iowa,² and affected individuals from two families hailing from Southern Sweden,³ compared with the families from Germany⁵ to show that the repeat expansion-containing haplotype arose from a common founder event.

Methods

Patients and Participants

This study was approved by the Institutional Review Board for Human Research at the University of Utah School of Medicine with written informed consent obtained from each participant. The Swedish Ethical Review Authority (Etikprövningsnämnden dnr 2010/1659) also approved this study with written informed consent from each participant.

DNA Extraction

Methods for DNA extraction and preparation from Utah and Iowa families were as previously described.² Briefly, lymphoblast cell lines of participants from the Utah SCA4 family were cultured. Genomic DNA (gDNA) was extracted from 1×10^6 harvested cells using the New England Biolabs Monarch High Molecular Weight DNA Extraction Kit for Cells and Blood standard protocol. High molecular weight (HMW) DNA from lymphoblasts were pre-sheared or sheared as necessary to 12–17 kb. Blood-derived gDNA from the Iowa individuals ranged in size from 8 to 17 kb and did not require shearing.²

For the individuals affected by SCA4 from Sweden, HMW genomic DNA was extracted from 18–25 mg of brain tissue using the Monarch HMW DNA Extraction Kit for Tissue kit following the standard input protocol. Tissue chips from the thalamus and hypothalamus for patient Sweden1 and thalamus and cerebellum for patient Sweden2 were manually homogenized on ice using a pestle. Homogenized lysates were incubated in HMW gDNA Tissue Lysis Buffer containing Proteinase K at 56°C for 15 min with agitation at 1400 rpm, then for a further 30 min without agitation. Genomic DNA was eluted in 50 μ L gDNA (3 μ g) at a concentration of 15 ng/ μ L was sheared to 15–17 kb using the Diagenode Megarupter 3 Shearing Kit at speed 30.

Size and quality of genomic DNA from lymphoblasts, blood, and brain tissue were checked using the Genomic DNA 165 kb kit for the Agilent Femto Pulse, NanoDrop, and Qubit.

Targeted Long-Read DNA Sequencing Using Adaptive Sampling

Swedish brain tissue-derived gDNA underwent a 1X Promega ProNex bead cleanup before 1.3–2.2 μ g was

inputted into the ONT SQK-NBD114.24 native barcoding (Oxford, UK) protocol as previously described² (<https://www.protocols.io/view/native-barcoding-sqk-nbd114-gdna-for-adaptive-samp-kxygx3qx4g8j/v1>).⁹ Equimolar quantities of the barcoded samples were pooled into three libraries of 50 fmol each, loaded onto one R10.4.1 flow cell and sequenced using PromethION. We used a targeted long-read sequencing approach through adaptive sampling to enrich sequencing of DNA based on a target sequence, controlled computationally in real time. The library was sequenced using adaptive sampling through a custom browser extensible data (BED) file covering the 20 Mb region of interest (chr16:56,000,000–76,100,000 (hg38)).

Structural Variant and Single Nucleotide Variant Detection and Phasing

High accuracy base calling was performed with Guppy (v.7.0.9, ONT). Resulting FASTQ files, with high accuracy reads (quality score >9), were mapped to the human genome GRCh38 using minimap2 (v.2.26; RRID:SCR_018550). We applied PEPPER-Margin-DeepVariant, a haplotype-aware variant calling pipeline to derive phased BAMs and phased single nucleotide variant (SNV) calls.¹⁰ We also ran Sniffles2.2.3 on the phased BAM from PEPPER-Margin-DeepVariant with the *phase* tag to obtain phased structural variant calls.¹¹

Methylation Analysis

We used modbamtools¹² v.0.4.8 to create haplotype methylation plots for each individual around the repeat

site. The detailed protocol has previously been described in <https://www.protocols.io/view/processing-frozen-cells-for-population-scale-sqk-l-6qpvr347bvmk/v1>. In order to compare methylation patterns with unaffected individuals, we used DNA derived from the HG002 cell line from Coriell (<https://www.coriell.org/>) as previously described as a control.¹³

Haplotyping Analysis

We reviewed phased SNVs called from all individuals for assessment of the presence of rare variants around the repeat expansion. In order to compare the haplotypes in the pathogenic repeat expansion-containing allele compared with the other allele, we took SNVs present within the repeat-containing haplotype of the Utah1 individual as the index haplotype for consistency. We then compared whether phased SNVs in other individuals matched the SNVs within the index repeat containing haplotype. Dating of the repeat-containing haplotype was carried out using previously established methods for rare mutations amongst small sample sets with dense marker data (<https://github.com/bahlolab/DatingRareMutations>).¹⁴ The model assumes a correlated genealogy as subsets of individuals likely have common ancestry before the most recent common ancestor as individuals rarely have independent recombination histories.¹⁴

Results

Here, we assessed the haplotype structure of individuals with SCA4 from Utah,^{1,2} Iowa,² and Sweden³ to

TABLE 1 Summary of individuals studied

Patient	Case number	Family	Estimated GGC repeats (n)	GGC repeats on other allele (n)	Age-of-onset (years)	Sex	Clinical presentation
Utah1	III.2	Utah	51	21	Unknown	F	Loss of balance and peripheral sensory neuropathy
Utah2	IV.4	Utah	51	21	59	M	Loss of balance and peripheral sensory neuropathy
Iowa2	II.1	Iowa	48	21	50	M	Loss of balance and peripheral sensory neuropathy
Sweden1	Family 1 (IV:4)	Sweden	55	21	42	M	Loss of balance, neuropathy, and dysautonomia
Sweden2	Family 2 (III:2)	Sweden	46	21	35	F	Loss of balance and neuropathy

Further information on individuals previously evaluated in studies.^{2,3} For individuals from Sweden (Sweden1 and Sweden2), analysis within this study came from DNA extracted from post-mortem brain hypothalamus samples, and not lymphocytes, as used in the original study.³ Repeat sizes were estimated from Sniffles tool. Abbreviations: F, female; M, male.

(a)

Family	Sample	Clinical information	Estimated repeat expansion size	Other allele GGC repeat size	Position (hg38) of SNV					
					16:72734457 A>G	16:72737703 T>C	16:72787719 A>G	16:72787737 A>G	16:72787739 T>C	16:72787743 A>G
Utah	Utah 1	SCA4 (lymphoblast)	51	21	1/0	1/0	1/0	1/0	1/0	1/0
Utah	Utah 2	SCA4 (lymphoblast)	51	21	1/0	1/0	1/0	1/0	1/0	1/0
Iowa	Iowa	SCA4 (blood)	48	21	1/0	0/0	1/0	1/0	1/1	1/0
Sweden	Sweden 1	SCA4 (thalamus)	55	21	1/0	1/0	1/0	1/0	1/0	1/0
Sweden	Sweden 2	SCA4 (thalamus)	47	21	1/0	1/0	1/0	1/0	1/0	1/0
Utah	Control 1	Unaffected	21	21	0/0	0/0	0/0	0/0	0/0	0/0
Utah	Control 2	Unaffected	21	21	0/0	0/0	0/0	0/0	0/0	0/0

Match to RE haplotype ● Matched ● Unmatched

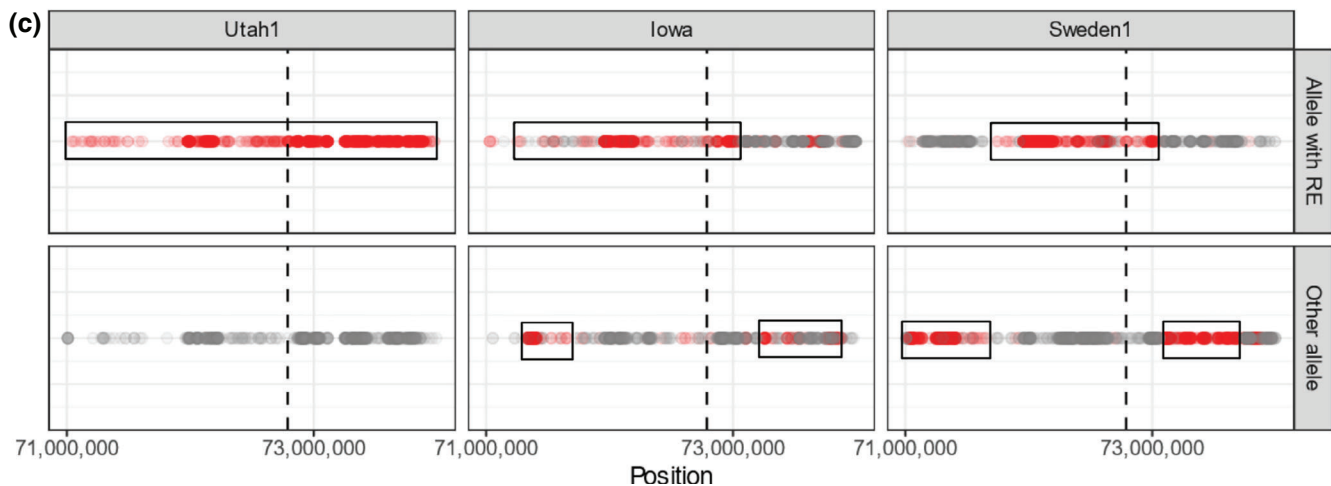
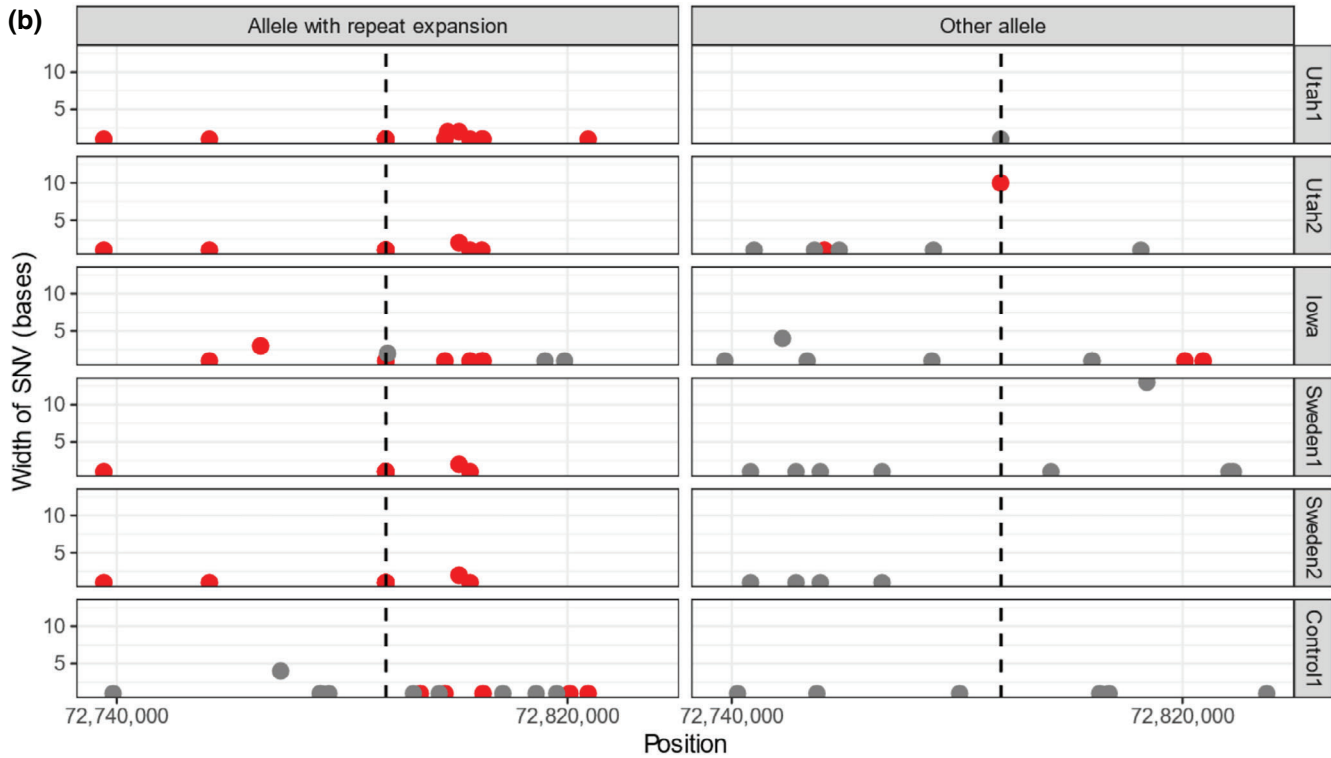


FIG. 1. Legend on next page.

ascertain whether the repeat expansion-containing haplotype arose from a common founder event. Within the two individuals of Swedish ancestry, DNA was extracted from post-mortem brain samples for analyses, while DNA was extracted from lymphoblastoid cell lines and whole blood for the Utah and Iowa families, respectively.^{2,3} Full demographic information is presented in Table 1. We then leveraged the utility of adaptive sampling to computationally drive targeted long-read sequencing using Oxford Nanopore Technologies (ONT).^{2,9} This had been an approach we originally employed to enrich for sequencing over the region of high linkage leading to the identification of the *ZFH3* repeat expansion.^{2,9} Furthermore, this method also enabled haplotypes to be phased reliably. Unfortunately, we were unable to sequence all individuals affected due to lack of adequate quality DNA from historical samples.²

First, we estimated the *ZFH3* GGC repeat expansion size using Sniffles2 calling from the long-read sequencing data (Table 1). We then reviewed phased SNVs situated around the repeat expansion. We found that all six of the rare SNVs described in individuals with SCA4 reported by Figueroa and colleagues⁵ were present across individuals from our four families, but only within the allele containing the repeat expansion (Fig. 1A). The only exception was the chr16:72737704T>C, which was absent in one individual from Iowa. Furthermore, these variants were absent in unaffected and unrelated spouses within the Utah kindred.

For consistency,⁵ taking the repeat expansion haplotype in an individual from Utah as the index haplotype, we looked within ± 50 kb of the repeat expansion and found high SNV similarity within the repeat expansion-containing haplotype across all individuals (Fig. 1B). This pattern was not appreciable within unaffected individuals or on the alternate allele (Fig. 1B). Lastly, extended phased haplotypes from long-read sequencing revealed large haplotype blocks of ~ 1 Mb spanning the repeat expansion with crossover events across individuals from the Iowa and Swedish families (Fig. 1C). Assuming a correlated genealogy (as defined), the mutation arose 88.5 generations ago, with a 95% confidence interval (CI) of 23.7–400.6. Assuming that each generation is 25 years, this dates the mutation to ~ 2200 years ago (95% CI: 600 to 10,025 years).

Haplotype-specific methylation calling around the *ZFH3* GGC repeat expansion for individuals with SCA4 from Utah and Sweden showed hypermethylation around the repeat expansion site compared with the allele that does not harbor the repeat expansion (Supplementary Fig. S1). Of note, in those with homozygous wildtype alleles without the pathogenic repeat expansion, there was evidence of hypomethylation around the site of the non-expanded short tandem repeat (Supplementary Fig. S1). This pattern of methylation suggests that the presence of the repeat expansion may have potential functional implications in gene or transcript expression in the disease state. Therefore, further characterization of the transcript structure and expression in disease within the brain would be invaluable to correlate these findings.

Discussion

In line with the common Swedish ancestry of all reported individuals affected by SCA4,¹⁻⁵ our extended haplotype analysis in other families including those from Southern Sweden³ revealed a common single ~ 1 Mb haplotype harboring the pathogenic expansion. The presence of a founding haplotype of the *ZFH3* GGC expansion demonstrates the phenomenon observed across other repeat expansion disorders including, but not limited to *C9orf72*,¹⁵ *NOTCH2NL*,¹⁶⁻¹⁸ and *BEAN1*¹⁹ that exhibit population-specific differences in prevalence.

The *ZFH3* GGC repeat expansion likely arose approximately 88.5 generations (~ 2220 years ago) (95% CI: 23.7–400.6 generations). This repeat-containing haplotype arose much later than the *RFC1* repeat containing haplotype in CANVAS (cerebellar ataxia with neuropathy and bilateral vestibular areflexia syndrome), which is postulated to have arisen approximately 25,880 years ago (95% CI: 14,080–48,020).²⁰ This may in part explain the higher prevalence of the *RFC1* expansion, which has a carrier frequency of 1 in 14.²¹ In comparison, the *C9orf72* hexanucleotide repeat expansion arose before, or just before the *ZFH3* repeat containing haplotype and also has a Northern European founder (100.5 to 251 generations).^{15,22}

FIG. 1. Characterizing the repeat expansion (RE) haplotype using adaptive sampling-enabled targeted long-read sequencing. (A) Six ultra-rare single nucleotide variants (SNVs) identified by Figueroa et al.⁵ were associated with the repeat expansion haplotype in affected individuals from Utah, Iowa, and Sweden. Sizing of the phased structural variants including the GGC repeat expansion used Sniffles.¹¹ 1 represents presence of a variant, 0 represents absence of the particular variant. | denotes a phased variant. / denotes unphased variant. The first allele contains the repeat expansion (ie, 1|0 shows the presence of the SNV on the repeat-expansion-containing allele). It is worth noting that four SNVs chr16:72787719 A>G; chr16:72787737 A>G; chr16:72787739 T>C; and chr16:72787743 A>G) are located within the expanded repeat (B). SNVs in the repeat expansion haplotype (taking Utah1 repeat expansion haplotype as the index haplotype) within 50 kb of the repeat expansion locus (vertical dashed line) showed strong similarities among cases from Utah, Iowa, and Southern Sweden. SNVs matching the index repeat expansion-haplotype are shown in red. SNVs present that do not match those within the repeat-containing haplotype are shown in gray. (C). Representative extended phased haplotypes showed a large identical haplotype block (boxed) with crossover events from individuals with spinocerebellar ataxia type 4 from Utah, Iowa, and Sweden. This led us to conclude that the haplotype associated with *ZFH3* GGC repeat expansion underlying spinocerebellar ataxia type 4 has a common ancestral founder. [Color figure can be viewed at wileyonlinelibrary.com]

Although a recent study successfully detected the *ZFH3* GGC repeat expansion through polymerase chain reaction (PCR)-based methods from genomic DNA,⁶ most groups have experienced difficulties in detecting the *ZFH3* repeat by PCR-based methods due to its high GC content.²⁻⁵ Thus, these findings support examination of the six rare SNVs tagging the expansion in screening for the pathogenic repeat-containing haplotype. Furthermore, flanking sequences as well as interruptions that alter the repeat motif have been shown to have impact on the repeat stability and pathogenicity.^{23,24} While we do not fully understand factors that moderate expansion of a short tandem repeat, a comprehensive knowledge of variants associated with the repeat-containing haplotype structure has important implications when considering potential development of repeat-stabilizing targeted therapy.

While SCA4 has been heralded as a novel polyglycine disorder given the GGC repeat expansion is the first to be described to reside in a coding exon,⁷ our findings of hypermethylation around the repeat expansion is intriguing. Hypermethylation has been appreciated in other GGC repeat expansion disorders such as in Fragile X syndrome, leading to transcriptional silencing of *FMR1*.²⁵ However, in this case, the hypermethylation is within a coding exon of a gene, with previous evidence showing possible exon-level DNA methylation may play a role in exon inclusion in alternative splicing; a trend observed generally within the last coding exons.^{26,27}

Of significance, ubiquitinated neuronal intranuclear inclusions were characterized within the post-mortem brain tissue of the Swedish individuals with SCA4 studied here³ as well as in those from German families.⁵ While Figueroa and colleagues found that antibodies targeting *ZFH3* were detected in SCA4 brains and that there was an increase in *ZFH3* without concomitant increase in *ZFH3* expression in patient-derived fibroblasts but not controls, further understanding of the underlying pathogenesis remains to be elucidated.⁵ It is unknown whether the level of toxicity arises from toxic gain-of-function from the polyglycine-containing protein or from a transcribed repeat-containing RNA, or both. This is further complicated by the unknown significance of methylation around the GC-rich repeat expansion.

Phased long-read sequencing data from individuals from Utah, Iowa, and Southern Sweden has allowed detection of a common ancestral haplotype harboring the repeat expansion that is hypermethylated, providing further insight into not only the genetic architecture of SCA4, but also other repeat expansion disorders. ■

Acknowledgments: The authors are grateful to all the individuals who have donated valuable time and samples to this research project. Z.C. was supported by The Michael J. Fox Foundation (MJFF). The authors are also grateful to: Wellcome Trust, Medical Research Council (MRC) UK,

MSA Trust, National Institute for Health Research University College London Hospitals Biomedical Research Centre (NIHR-BRC), The Michael J. Fox Foundation (MJFF), Fidelity Trust, Dolby Family fund, Rosettes Trust, Alzheimer's Research UK (ARUK), MSA Coalition, NIH NeuroBioBank, MRC Brainbank Network, and British Embassy Gulf Strategy Fund. This work used the computational resources of the NIH High Performance Computing (HPC) Biowulf cluster (<http://hpc.nih.gov>). This work was supported in part by the Intramural Research Program of the National Institutes of Health including: the Center for Alzheimer's and Related Dementias, within the Intramural Research Program of the National Institute on Aging and the National Institute of Neurological Disorders and Stroke (1ZIAAG000538-03 and 1ZIAAG000542-01).

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

- Flanigan K, Gardner K, Alderson K, Galster B, Otterud B, Leppert MF, et al. Autosomal dominant spinocerebellar ataxia with sensory axonal neuropathy (SCA4): clinical description and genetic localization to chromosome 16q22.1. *Am J Hum Genet* 1996;59(2):392–399.
- Chen Z, Gustavsson EK, Macpherson H, Anderson C, Clarkson C, Rocca C, et al. Adaptive long-read sequencing reveals GGC repeat expansion in *ZFH3* associated with spinocerebellar ataxia type 4. *Mov Disord* 2024;39(3):486–497.
- Paucar M, Nilsson D, Engvall M, Laffita-Mesa J, Söderhäll C, Skorpil M, et al. Spinocerebellar ataxia type 4 is caused by a GGC expansion in the *ZFH3* gene and is associated with prominent dysautonomia and motor neuron signs. *J Intern Med* 2024;296(3):234–248.
- Wallenius J, Kafantari E, Jhaveri E, Gorcenco S, Ameer A, Karremo C, et al. Exonic trinucleotide repeat expansions in *ZFH3* cause spinocerebellar ataxia type 4: a poly-glycine disease. *Am J Hum Genet* 2024;111(1):82–95.
- Figueroa KP, Gross C, Buena-Atienza E, Paul S, Gandelman M, Kakar N, et al. A GGC-repeat expansion in *ZFH3* encoding polyglycine causes spinocerebellar ataxia type 4 and impairs autophagy. *Nat Genet* 2024;56(6):1080–1089.
- Dalski A, Pauly MG, Hanssen H, Hagenah J, Hellenbroich Y, Schmidt C, et al. Repeat length in spinocerebellar ataxia type 4 (SCA4) predicts age at onset and disease severity. *J Neurol* 2024;271(9):6289–6300.
- Charlet-Berguerand N. An unexpected polyglycine route to spinocerebellar ataxia. *Nat Genet* 2024;56:1039–1041.
- Rudaks LI, Yeow D, Kumar KR. SCA4 unravelled after more than 25 years using advanced genomic technologies. *Mov Disord* 2024;39:457–461.
- Anderson C, Macpherson H, Gustavsson E, Lee J, Chen Z. Native Barcoding (SQK-NBD114) gDNA for Adaptive Sampling using Oxford Nanopore Technologies v1; 2023. <https://doi.org/10.17504/protocols.io.kxyg3qx4g8j/v1>.
- Shafin K, Pesout T, Chang P-C, Nattestad M, Kolesnikov A, Goel S, et al. Haplotype-aware variant calling with PEPPER-Margin-DeepVariant enables high accuracy in nanopore long-reads. *Nat Methods* 2021;18(11):1322–1332.
- Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, von Haeseler A, Schatz MC. Accurate detection of complex structural variations using single-molecule sequencing. *Nat Methods* 2018;15(6):461–468.
- Razaghi R, Hook PW, Ou S, Schatz MC, Hansen KD, Jain M, Timp W. Modbamtools: analysis of single-molecule epigenetic data for long-range profiling, heterogeneity, and clustering. *bioRxiv* 2022. <https://doi.org/10.1101/2022.07.07.499188>

13. Genner R, Akeson S, Meredith M, Jerez PA, Malik L, Baker B, et al. Assessing methylation detection for primary human tissue using Nanopore sequencing. *bioRxiv* 2024;
14. Gandolfo LC, Bahlo M, Speed TP. Dating rare mutations from small samples with dense marker data. *Genetics* 2014;197:1315–1327.
15. Smith BN, Newhouse S, Shatunov A, Vance C, Topp S, Johnson L, et al. The C9ORF72 expansion mutation is a common cause of ALS +/-FTD in Europe and has a single founder. *Eur J Hum Genet* 2013;21(1):102–108.
16. Sone J, Mitsuhashi S, Fujita A, Mizuguchi T, Hamanaka K, Mori K, et al. Long-read sequencing identifies GGC repeat expansions in NOTCH2NLC associated with neuronal intranuclear inclusion disease. *Nat Genet* 2019;51(8):1215–1221.
17. Ishiura H, Shibata S, Yoshimura J, Suzuki Y, Qu W, Doi K, et al. Noncoding CGG repeat expansions in neuronal intranuclear inclusion disease, oculopharyngodistal myopathy and an overlapping disease. *Nat Genet* 2019;51(8):1222–1232.
18. Chen Z, Yan Yau W, Jaunmuktane Z, Tucci A, Sivakumar P, Gagliano Taliun SA, et al. Neuronal intranuclear inclusion disease is genetically heterogeneous. *Ann Clin Transl Neurol* 2020;7(9):1716–1725.
19. Sato N, Amino T, Kobayashi K, Asakawa S, Ishiguro T, Tsunemi T, et al. Spinocerebellar ataxia type 31 is associated with "inserted" penta-nucleotide repeats containing (TGGAA)_n. *Am J Hum Genet* 2009;85(5):544–557.
20. Rafehi H, Szmulewicz DJ, Bennett MF, Sobreira NLM, Pope K, Smith KR, et al. Bioinformatics-based identification of expanded repeats: a non-reference intronic pentamer expansion in RFC1 causes CANVAS. *Am J Hum Genet* 2019;105(1):151–165.
21. Ibañez K, Jadhav B, Zanovello M, Gagliardi D, Clarkson C, Facchini S, et al. Increased frequency of repeat expansion mutations across different populations. *Nat Med* 2024;30:3357–3368.
22. Majounie E, Renton AE, Mok K, Dopper EG, Waite A, Rollinson S, et al. Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: a cross-sectional study. *Lancet Neurol* 2012;11(4):323–330.
23. Pellerin D, Del Gobbo GF, Couse M, Dolzhenko E, Nageswaran SK, Cheung WA, et al. A common flanking variant is associated with enhanced stability of the FGF14-SCA27B repeat locus. *Nat Genet* 2024;56(7):1366–1370.
24. Dominik N, Magri S, Currò R, Abati E, Facchini S, Corbetta M, et al. Normal and pathogenic variation of RFC1 repeat expansions: implications for clinical diagnosis. *Brain* 2023;146(12):5060–5069.
25. Devys D, Lutz Y, Rouyer N, Bellocq J-P, Mandel J-L. The FMR-1 protein is cytoplasmic, most abundant in neurons and appears normal in carriers of a fragile X premutation. *Nat Genet* 1993;4:335–340.
26. Chuang T-J, Chen F-C, Chen Y-Z. Position-dependent correlations between DNA methylation and the evolutionary rates of mammalian coding exons. *Proc Natl Acad Sci USA* 2012;109:15841–15846.
27. Shayevitch R, Askayo D, Keydar I, Ast G. The importance of DNA methylation of exons on alternative splicing. *RNA* 2018;24:1351–1362.

Supporting Data

Additional Supporting Information may be found in the online version of this article at the publisher's web-site.

SGML and CITI Use Only

DO NOT PRINT

Author Roles

(1) Research Project: A. Conceptualization, B. Methodology, C. Investigation, D. Data Curation, E. Formal Analysis; (2) Statistical Analysis: A. Design, B. Execution, C. Review and Critique; (3) Manuscript Preparation: A. Writing of the First Draft, B. Review and Critique; (4) Other: A. Funding Acquisition, B. Supervision.

Z.C.: 1A, 1B, 1C, 1D, 1E.

P.A.J.: 1B, 1C, 1D, 1E.

C.A.: 1B, 1C, 1D, 1E.

M.P.: 1B, 1C, 1D, 1E.

J.L.: 1B, 1C, 1D, 1E.

D.N.: 1C, 1D, 1E.

H.M.: 1C, 1D, 1E.

A.S.: 4A, 1C, 1D, 1E.

K.M.: 1C, 1D.

J.H.: 4A, 1C, 1D.

A.B.S.: 1C, 1D.

A.T.: 1C, 1D.

K.D.M.: 1C, 1D.

Y.-H.F.: 1C, 1D.

M.E.: 1C, 1D.

J.L.-M.: 1C, 1D.

I.N.: 1C, 1D.

A.W.: 1C, 1D.

L.J.P.: 1C, 1D.

C.B.: 4A, 1C, 1D.

E.K.G.: 1A, 1B, 1C, 1D, 1E, 4B.

P.S.: 1A, 1B, 1C, 1D, 4A, 4B.

M.R.: 1A, 1B, 1C, 1D, 4A, 4B.

H.H.: 1A, 1B, 1C, 1D, 4A, 4B.