Supplementary Appendix

Supplement to: Pellerin D, Danzi MC, Wilke C, et al. Deep intronic *FGF14* GAA repeat expansion in late-onset cerebellar ataxia. N Engl J Med 2023;388:128-41. DOI: 10.1056/NEJMoa2207406

This appendix has been provided by the authors to give readers additional information about the work.

TABLE OF CONTENTS

Supplementary Appendix To:

Deep Intronic *FGF14* GAA Repeat Expansion in Late-Onset Cerebellar Ataxia

David Pellerin, M.D.[1,2*]; Matt C. Danzi, Ph.D.[3*]; Carlo Wilke, M.D.[4,5]; Mathilde Renaud, M.D., Ph.D.[6,7]; Sarah Fazal, Ph.D.[3]; Marie-Josée Dicaire, B.Sc.[1]; Carolin K. Scriba, B.Sc.[8,9]; Catherine Ashton, M.B., B.S.[10]; Christopher Yanick, B.S.[3,11]; Danique Beijer, Ph.D.[3]; Adriana Rebelo, Ph.D.[3]; Clarissa Rocca, M.Sc.[2]; Zane Jaunmuktane, M.D.[12,13]; Joshua A. Sonnen, M.D.[1,14]; Roxanne Larivière, Ph.D.[1]; David Genís, M.D.[15]; Laura Molina Porcel, M.D., Ph.D.[16,17]; Karine Choquet, Ph.D.[18,19]; Rawan Sakalla, B.Sc.[1]; Sylvie Provost, M.Sc.[20,21]; Rebecca Robertson, Ph.D.[1,18]; Xavier Allard-Chamard, M.Sc.[1,18]; Martine Tétreault, Ph.D.[22,23]; Sarah J. Reiling, Ph.D.[24]; Sara Nagy, M.D.[2,25]; Vikas Nishadham, M.B., B.S.[26]; Meera Purushottam, Ph.D.[27]; Seena Vengalil, M.D.[26]; Mainak Bardhan, M.B., B.S.[26]; Atchayaram Nalini, M.D., Ph.D.[26]; Zhongbo Chen, B.M., B.Ch.[28]; Jean Mathieu, M.D.[29]; Rami Massie, M.D.[1]; Colin H. Chalk, M.D.[1]; Anne-Louise Lafontaine, M.D.[1]; François Evoy, M.D.[29]; Marie-France Rioux, M.D.[29]; Jiannis Ragoussis, Ph.D.[24]; Kym M. Boycott, M.D., Ph.D.[30]; Marie-Pierre Dubé, Ph.D.[20,31]; Antoine Duquette, M.D.[22,23,32]; Henry Houlden, M.D., Ph.D.[2]; Gianina Ravenscroft, Ph.D.[8]; Nigel G. Laing, Ph.D.[8]; Phillipa J. Lamont, M.B., B.S., Ph.D.[10]; Mario A. Saporta, M.D., Ph.D.[11]; Rebecca Schüle, M.D.[4,5]; Ludger Schöls, M.D.[4,5]; Roberta La Piana, M.D., Ph.D.[1,33]; Matthis Synofzik, M.D.[4,5]; Stephan Zuchner, M.D., Ph.D.[3**]; Bernard Brais, M.D., C.M., Ph.D.[1,18,32**]

*Drs. Pellerin and Danzi contributed equally to this article.

**Drs. Zuchner and Brais contributed equally to this article.

Affiliations

[1]Department of Neurology and Neurosurgery, Montreal Neurological Hospital and Institute, McGill University, Montreal, QC, Canada

[2]Department of Neuromuscular Disease, UCL Queen Square Institute of Neurology and The National Hospital for Neurology and Neurosurgery, University College London, London, United Kingdom

[3]Dr. John T. Macdonald Foundation Department of Human Genetics and John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL, USA

[4]Department of Neurodegenerative Diseases, Hertie-Institute for Clinical Brain Research and Center of Neurology, University of Tübingen, Tübingen, Germany.

[5]German Center for Neurodegenerative Diseases (DZNE), Tübingen, Germany

[6]Service de Génétique Clinique et de Neurologie, Hôpital Brabois Enfants, Nancy, France

[7]INSERM Unité 1256 N-GERE, Université de Lorraine, Nancy, France

[8]Centre for Medical Research University of Western Australia and Harry Perkins Institute of Medical Research, Perth, Western Australia, Australia

[9]Department of Diagnostic Genomics, PathWest Laboratory Medicine, West Australian Department of Health, Perth, Western Australia, Australia
[10]Department of Neurology, Royal Perth Hospital, Perth, WA, Australia
[11]Department of Neurology, University of Miami Miller School of Medicine, Miami, FL, USA
[12]Division of Neuropathology, The National Hospital for Neurology and Neurosurgery, University College London NHS Foundation Trust, London, United Kingdom
[13]Department of Clinical and Movement Neurosciences and Queen Square Brain Bank for Neurological Disorders, UCL Queen Square Institute of Neurology, University College London, London, United Kingdom
[14]Department of Pathology, Montreal Neurological Hospital and Institute, McGill University, Montreal, QC, Canada
[15]Ataxia and Hereditary Spastic Paraplegia Unit, Service of Neurology, Hospital Universitari de Girona Dr. Josep Trueta (ICS) & Hospital Santa Caterina IAS, Girona, Spain
[16]Alzheimer's Disease and other Cognitive Disorders Unit, Service of Neurology, Hospital Clínic, Institut d'Investigacions Biomediques August Pi i Sunyer (IDIBAPS), University of Barcelona, Barcelona, Spain
[17]Neurological Tissue Brain Bank, Biobanc-Hospital Clínic-IDIBAPS, Barcelona, Spain
[18]Department of Human Genetics, McGill University, Montreal, QC, Canada
[19]Department of Genetics, Harvard Medical School, Boston, MA, USA
[20]Montreal Heart Institute, Université de Montréal, Montreal, QC, Canada
[21]Université de Montréal Beaulieu-Saucier Pharmacogenomics Center, Montreal, QC, Canada
[22]Department of Neurosciences, Faculty of Medicine, Université de Montréal, Montreal, QC, Canada
[23]Centre de Recherche du Centre Hospitalier de l'Université de Montréal, Montreal, QC, Canada
[24]McGill Genome Centre, Department of Human Genetics, McGill University, Montreal, QC, Canada
[25]Department of Neurology, University Hospital Basel, University of Basel, Basel, Switzerland
[26]Department of Neurology, National Institute of Mental Health and Neurosciences, Bengaluru, India
[27]Molecular Genetics Laboratory, Department of Psychiatry, National Institute of Mental Health and Neurosciences, Bengaluru, India
[28]Department of Neurodegenerative Disease, UCL Queen Square Institute of Neurology, University College London, London, United Kingdom
[29]Faculty of Medicine and Health Sciences, Sherbrooke University, Sherbrooke, QC, Canada
[30]Children's Hospital of Eastern Ontario Research Institute, University of Ottawa, ON, Canada
[31]Department of Medicine, Faculty of Medicine, Université de Montréal, Montreal, QC, Canada
[32]Centre de Réadaptation Lucie-Bruneau, Montreal, QC, Canada
[33]Department of Diagnostic Radiology, McGill University, Montreal, QC, Canada

Corresponding author: Prof. Brais can be contacted at bernard.brais@mcgill.ca or at the Montreal Neurological Hospital and Institute, McGill University, 3801 University St., Montreal, QC H3A 2B4, Canada.

AUTHOR CONTRIBUTIONS

David Pellerin, Matt C. Danzi, Mathilde Renaud, Sarah Fazal, Stephan Zuchner, and Bernard Brais designed the study.

David Pellerin, Carlo Wilke, Mathilde Renaud, Catherine Ashton, David Genís, Laura Molina Porcel, Sara Nagy, Vikas Nishadham, Seena Vengalil, Mainak Bardhan, Atchayaram Nalini, Jean Mathieu, Rami Massie, Colin H. Chalk, Anne-Louise Lafontaine, François Evoy, Marie-France Rioux, Antoine Duquette, Henry Houlden, Phillipa J. Lamont, Rebecca Schüle, Ludger Schöls, Matthis Synofzik, and Bernard Brais recruited patients.

David Pellerin, Matt C. Danzi, Mathilde Renaud, Sarah Fazal, Zane Jaunmuktane, Joshua A. Sonnen, Rawan Sakalla, Roberta La Piana, Matthis Synofzik, Stephan Zuchner, and Bernard Brais wrote the first draft of the manuscript.

All authors contributed to acquisition and analysis of data.

All authors critically reviewed and approved the final version of the manuscript.

SUPPLEMENTARY METHODS

Patient enrollment and institutional review board approval

Patients were recruited through ataxia clinics at the Montreal Neurological Hospital (Montreal, QC, Canada), the Centre Hospitalier Universitaire de Sherbrooke (Sherbrooke, QC, Canada), the Centre Hospitalier de l'Université de Montréal (Montreal, QC, Canada), the Center for Neurology, Tübingen (Germany), the Neurogenetics Clinic at Royal Perth Hospital (Australia), the Neurogenetics Clinic at the National Hospital for Neurology and Neurosurgery, London (United Kingdom) and the National Institute of Mental Health and Neurosciences, Bengaluru (India). Patients of Indian descent were recruited through the National Institute of Mental Health and Neurosciences in Bengaluru, India and the National Hospital for Neurology and Neurosurgery in London, UK. The discovery cohort consisted of six patients from three large French Canadian families with unsolved autosomal dominant late-onset cerebellar ataxia (LOCA). The French Canadian, German, Australian and Indian validation cohorts consisted of 66, 228, 20 and 31 index patients, respectively. All participants underwent clinical evaluation by some of the authors with expertise in ataxia. Clinical data were collected through review of medical records and, when possible, clinical re-evaluation. The institutional review board of the Montreal Neurological Hospital, Montreal (MPE-CUSM-15-915), the Centre Hospitalier de l'Université de Montréal, Montreal (ND02.045), the Center for Neurology, Tübingen (598/2011BO1), the University of Western Australia, Perth (RA/4/20/1008), the UCL Queen Square Institute of Neurology, London (07/Q0512/26) and the National Institute of Mental Health and Neurosciences, Bengaluru (NIMHANS/25 IEC/2020-2021) approved this study, and all patients provided written informed consent. The study complied with all relevant ethical regulations.

Some clinical details of interest reported in Table 1 were not able to be collected for all patients. Information on inheritance (whether a case is sporadic or familial) is missing for four of the German patients. A case is said to be familial if additional relatives are also affected by ataxia. Whether episodes of ataxia preceded onset of permanent ataxia is missing for one French Canadian patients. The age of onset of episodic ataxia is missing for two French Canadian patients. This information is not missing for German, Australian and Indian patients. The age of onset of permanent ataxia is missing for two French Canadian and two German patients. Information on the presence of nystagmus is missing for three French Canadian patients. The presence of diplopia or visual blurring is missing for two French Canadian patients. Information on cerebellar dysarthria, gait ataxia and appendicular ataxia is missing for four German patients. Information on vertigo is missing for four French Canadian and four German patients. Details on the presence of postural tremor is missing for eight French Canadian patients. Cerebellar atrophy data is missing for 24 French Canadian and seven German patients. Since sufficient clinical data could not be collected for five French Canadian and one German patients they were not included in Table 1. Details on whether a case was maternally or paternally inherited are available for 66 dominantly inherited cases. Information on the

use of a walking aid is missing for six French Canadian and four German patients. Details on the presence of spasticity is missing for two French Canadian patients.

Snap-frozen cerebellar cortex from postmortem brain specimens of 15 patients of European descent with idiopathic late-onset cerebellar ataxia and seven non-ataxic controls of European descent were obtained from the Queen Square Brain Bank for Neurological Disorders (six patients) and the Neurological Tissue Brain Bank of the Hospital Clínic-IDIBAPS (Instituto de Investigaciones Biomédicas August Pi I Sunyer) Biobanc, Barcelona (nine patients and seven controls). Patients and controls were screened for the *FGF14* repeat expansion by long-range PCR following high-molecular-weight genomic DNA extraction using the MagAttract HMW DNA kit as per manufacturer's instructions (catalog no. 67563, Qiagen). Repeat-primed PCR was used to confirm the GAA motif of the expansion. Controls were, as much as possible, sex-, age- and postmortem interval (PMI)-matched to the two genetically confirmed patients. None of the controls had evidence of significant cerebellar pathology on neuropathological examination. Details of the two genetically confirmed patients and seven controls, all of Spanish origin, are provided in Table S2.

Fibroblast and lymphoblast cell culture

Skin biopsies were obtained from nine French Canadian patients carrying a $(GAA)_{\geq 250}$ expansion and four healthy French Canadian controls. Immortalized fibroblast cell lines from five additional healthy persons were obtained through the Repository for Mutant Human Cell Strains of the McGill University Health Centre (MUHC). Fibroblasts were cultured in DMEM, high glucose with 10% (v/v) fetal bovine serum, 100U/ml penicillin-streptomycin, 2.5µg/ml amphotericin B and 100µg/ml L-glutamine.

Immortalized lymphoblastoid cell lines were established from six French Canadian patients carrying a $(GAA)_{\geq 250}$ expansion and six controls using standard methods. Cells were cultured in Iscove's modified Dulbecco's medium (IMDM) with 10% (v/v) fetal bovine serum, 100U/ml penicillin-streptomycin, 2.5µg/ml amphotericin B and 100µg/ml L-glutamine. Cell lines were grown at 37°C, 5% CO_2 in 95% humidified air.

Neuronal Differentiation

Induced pluripotent stem cells (iPSCs) were differentiated from two French Canadian patients carrying *FGF14* GAA expansions and two controls of European descent, following an established protocol as previously described.[1-4] The two patients had a $(GAA)_{16/383}$ (patient FC49) and a $(GAA)_{300/300}$ (patient FC35) genotype, respectively. To enrich cultures with mature lower motor neurons, magnetic bead sorting using an anti-CD171PE antibody and anti-PE magnetic beads was used. CD171 is an axonal marker (L1CAM) and is only expressed on mature neurons in differentiation cultures. Motor neurons were then cultured for seven days to allow neurite outgrowth and maturation prior to use in experiments. Immunocytochemistry (ICC) for several neuronal markers was performed to confirm successful generation of mature motor neurons. Chosen markers were Homeobox Protein 9 (HB9), Neurofilament Light Chain (NFL) and Beta-

Tubulin 3 (βT3). Both patient iPSC-derived motor neuron lines stained positively for all three markers, consistent with control lines, confirming their neuronal lineage and therefore the ability to successfully generate and isolate motor neurons from the patient iPSC lines.

Whole-genome sequencing

Genomic DNA was isolated from peripheral blood of six French Canadian patients with LOCA belonging to three large unrelated families with autosomal dominant late-onset cerebellar ataxia using standard techniques. Whole-Genome Sequencing (WGS) was performed at the McGill University and Génome Québec Innovation Centre. Samples were prepared using the TrueSeq PCR-free DNA library preparation kit and sequenced on the Illumina HiSeq X platform (Illumina) to generate 150bp paired-end reads. The mean coverage was 46.5x (range: 41.5-50.0x). Sequences were aligned to the GRCh37 reference genome using the Burrows-Wheeler Aligner.[5] Duplicate reads were marked using Picard's MarkDuplicates utility. Variant calling, annotation, and filtering was performed using the GATK best practices pipeline.[6] Candidate variants were prioritized based on: segregation among all six patients; minor allele frequency <0.0001 in gnomAD;[7] prediction of pathogenicity by in silico software; evolutionary sequence conservation; and occurrence in known ataxia genes (as defined by the OMIM Catalog).

In silico repeat expansion analysis

WGS data aligned to GRCh37 were further processed with the use of ExpansionHunter Denovo software, version 0.6.2[8] with parameters set to restrict the search for large repetitive sites to motifs of length 3-8 nucleotides. The samples from the 1000 Genomes Project and the Vanderbilt Atrial Fibrillation Registry (VAFR) were processed as previously described.[9] The six LOCA WGS samples were then compared as a group against each of the two control cohorts. The number of anchored in-repeat reads (IRRs) supporting each tandem repeat was compared among the groups with each sample normalized to 40x read coverage. For each comparison, a Cohen's d effect size was calculated. To identify tandem repeats expanded in the six patients relative to controls, we filtered the tandem repeats for a mean anchored IRR of at least five in patients (supporting the existence of a large tandem repeat >175bp) and Cohen's d ≥1.3. Intergenic tandem repeats were excluded. Alignment of short reads in regions containing candidate repeat expansion was manually inspected using the Repeat Expansion Viewer (REViewer) software.[10]

Linkage analysis and haplotype reconstruction

Following the discovery of the intronic GAA repeat expansion in *FGF14*, post-hoc two-point parametric linkage analysis was performed on the large French Canadian Family I (Figure 1A) using the mutation status as the phenotype. Prior to this discovery, previous linkage efforts had remained unyielding in this large family as a result of the likely presymptomatic status of some of the younger members. Single-nucleotide polymorphism (SNP) genotype data were generated for 18 family members using the Illumina Infinium Global Screening Array-24 (GSA-24v1.0) at the McGill University and Génome Québec Innovation

Centre. Participants I.1, I.2, I.4, I.5, I.7, I.8, I.11, I.12, I.13, I.14, I.15, I.16, I.17, I.18, I.22, I.28, I.38 and I.39 from Family I (Figure 1A) were used for analysis. Analysis was performed using MLINK routine of FASTLINK v4.1P specifying a dominant disease with penetrance of 80% and disease allele frequency of 1:100,000.[11, 12] The following equation was used to determine the allele frequency and the phenocopy rate to use with the known prevalence. In the equation below ϕ is the prevalence, *f* the penetrance, *p* the disease gene frequency and *D* the disease allele.

$$\phi = f_{DD}p^2 + 2f_{dD}p(1-p) + f_{dd}(1-p)^2$$

The linkage analysis was conducted using penetrance set to 0.80, a phenocopy rate set to 0.00001 for a dominant disease with an allele frequency of 0.00001. Graphical representation was done with Manhattan Generator v1.7.4. Regional representation was produced using LocusZoom.[13] Haplotype reconstruction and phasing in Family I in the candidate region on chromosome 13 was carried out using SIMWALK version 2.90. The input files were converted to SIMWALK format using Mega2 v5.0.0. All 18 genotyped members from Family I were included for haplotype reconstruction.

Haplotype analysis

The 50 polymorphic sites composing the disease haplotype identified in Family I described above were then genotyped in four members of two additional French Canadian families as well as one Australian patient from whom WGS data were available. We then inspected the genotypes of each of the 50 loci under the null hypothesis that all affected persons would share the same disease haplotype. At each locus, if the person carried at least one copy of the polymorphism from the disease haplotype, then that locus did not support rejection of the null hypothesis in that person. Only loci where neither allele matched the common polymorphism observed in the disease haplotype supported rejection of the null hypothesis for that person.

Sanger sequencing

Sanger sequencing of PCR amplification products was performed at the Centre d'expertise et de services Génome Québec using the Applied Biosystems 3730*xl* DNA Analyzer (Applied Biosystems) to examine the tandem repeat motif at the *FGF14* locus. Sequences were analyzed using SnapGene v.5.0.8 software (Dotmatics).

Repeat-primed polymerase chain reaction

Repeat-primed PCR was used to assess for expansion of the GAA repeat tract in patients and controls. The primer sequences and experimental conditions are provided in Table S3. Repeat-primed PCR products analysis was performed on an ABI 3730*xl* DNA Analyzer (Applied Biosystems) using the GeneScan 1200 Liz Dye Size Standard (catalog no. 4379950, Applied Biosystems). Results were analyzed using the GeneMapper software (version 3.7, Applied Biosystems). The presence of characteristic saw-toothed products indicated the presence of a GAA repeat expansion at the *FGF14* repeat locus.

Nanopore sequencing

Targeted long-read nanopore sequencing was performed on specimens from 104 patients with an expansion of at least 250 repeat units (55 French Canadian, 44 German and five Australian patients), as well as in 18 controls. PCR amplification products were selected for molecular size >400bp using SPRIselect paramagnetic beads for DNA size-selection following manufacturer's protocol (Beckman Coulter Life Sciences). Pre-sequencing size selection was performed to increase coverage depth of larger alleles. Amplicons were normalized to 150 ng/µl and then multiplexed using native barcoding expansion PCR-free library preparation kits and the SQK-LSK109 sequencing kit as per manufacturer's instructions, multiplexed and sequenced on the MinION or PromethION platform (Oxford Nanopore Technologies). The MinION run contained 23 natively barcoded samples and a libprep negative control. The R9.4.1 flow cell was run for 67 hours and generated 6.27 Gb of data with 18.8 million reads. The PromethION run contained 92 natively barcoded samples and a library preparation negative control. The run was carried out on a R9.4.1 flow cell with 9,200 pores at run start. The run was stopped after 44 hours, when 53.13 Gb of data and 89.22 million reads were generated. Reads were basecalled and demultiplexed with stringent barcodes_both_ends setting using Guppy 5.0.13. Sequences were aligned to the GRCh37 reference human genome using Minimap2-2.17 with the predefined settings for nanopore data. STRique-v0.4.2 was then used to count the number of repeated units observed for each read spanning the *FGF14* tandem repeat site. Motif purity was calculated for each sequencing read as the number of GAA units observed in the portion of the repeat locus-spanning segment of the read divided by the STRique estimation of the total number of repeat units for that read. Trinucleotide repeat size was estimated using a Gaussian mixture model. For each person with nanopore sequencing data, a five component Gaussian mixture model was fit to the distribution of repeat units estimated by STRique which were over 90% pure for the GAA motif. Nanopore reads that did not align well to the locus (STRique prefix or suffix scores less than 4) or that contained fewer than four copies of the repeat unit as detected by STRique were excluded. Estimated distributions supported by less than 2% of the nanopore reads were filtered out. From the filtered mixture of Gaussian distributions, the distribution with the greatest mean value was used as the estimate of the larger repeat allele in each person. Polymorphic motifs of the repeat expansion locus were looked for in samples with mean GAA purity <95%.

Quantitative reverse transcription PCR on human tissue

Total RNA was extracted from human fibroblasts, lymphoblasts and postmortem cerebellar cortex using TRIzol reagent as per manufacturer's protocol (catalog no. 15596026, Invitrogen). RNA integrity was assessed on a 2100 Bioanalyzer Instrument (Agilent Technologies). Complementary DNA (cDNA) was synthesized from DNase-treated RNA with the High Capacity cDNA Reverse Transcription Kit (catalog no. 4368813, Applied Biosystems). Quantitative PCR of fibroblast, lymphoblast and postmortem brain samples

was performed on a QuantStudio 7 Flex Real-Time PCR System (Applied Biosystems) using the TaqMan Fast Advanced assay (catalog no. 4444557, Applied Biosystems) or the Fast SYBR Green assay (catalog no. 4385612, Applied Biosystems). Quantitative PCR of fibroblast-extracted RNA was also performed on a QX200 droplet digital PCR System using the EvaGreen assay (catalog no. 1864034, Bio-Rad). For quantitative PCR experiments on fibroblasts and lymphoblasts, samples were run in triplicate and *FGF14* levels were normalized against expression of *ACTB*. Relative quantification was computed by the $2^{-\Delta\Delta Ct}$ method using the mean value of controls as calibrator. For quantitative PCR experiments on postmortem cerebellum, six controls and two patients were run in duplicate and *FGF14* levels were normalized by geometric averaging of five stably expressed housekeeping genes (*ACTB*, *HPRT1*, *YWHAZ*, *RPL13*, *UBE2D2*) as assessed by Genorm.[14] *FGF14* transcript expression was assessed by three different quantitative PCR assays measuring levels of transcript 1, transcript 2 and both transcripts, respectively (see Figure 4B). Data were analyzed with the QuantStudio Software v1.3 (Applied Biosystems). Primer sequences are provided in Table S4.

Quantitative reverse transcription PCR on iPSC-derived motor neurons

RNeasy mini kit (catalog no. 74106, Qiagen) was used to extract RNA from cultures of iPSC-derived motor neurons. The Quantitect reverse transcription kit (catalog no. 205331, Qiagen) was used to convert 300ng of RNA to cDNA. The PowerUp SYBR green master mix (catalog no. A25741, Applied Biosystems) was used to set up the quantitative PCR reactions in quadruplicates; 10ng of cDNA was used per reaction with 400nM of each of the forward and reverse primers. *FGF14* levels were normalized against expression of *GAPDH*. Relative quantification was computed by the $2^{-\Delta\Delta Ct}$ method using the mean value of controls as calibrator. Primer sequences are provided in Table S4.

RNA sequencing

RNA sequencing (RNAseq) was performed on patient *(N*=3) and control *(N*=3) fibroblast cell lines. RNA was prepared by ribodepletion using the TruSeq Stranded Total RNA with Ribo-Zero Gold library preparation (Illumina). Fibroblast samples were sequenced to 60 million 100bp paired-end read pairs per sample on a NovaSeq 6000 system (Illumina). RNAseq reads were aligned to the GRCh37 reference using STAR-2.7.7a with Ensembl release 75 being used to define the reference transcripts. Transcript abundances were quantified using RSEM[15] in a strand-aware manner.

Immunoblotting on human postmortem cerebellum

Postmortem cerebellar tissues were homogenized in radioimmunoprecipitation assay (RIPA) lysis buffer (catalog no. R0278, Sigma-Aldrich) supplemented with EDTA-free protease inhibitor cocktail (catalog no. 4693159001, Roche). Samples were next sonicated on ice and centrifuged for 25min at 13,000*g* at room temperature to remove cell debris. Protein lysate concentrations were measured using the Pierce BCA Protein Assay Kit (catalog no. 23225, Thermo Scientific). Protein samples (15µg) were mixed with LDS

sample buffer (catalog no. NP0007, Invitrogen) and dithiothreitol (catalog no. NP0004, Invitrogen) and heated at 70°C for 10min. Protein extracts were resolved on a precast 4-12% Bis-Tris gel (catalog no. NP0322BOX, Invitrogen) for one hour at 150V and transferred onto a nitrocellulose membrane for two hours at 85V at 4°C. After blocking for two hours at room temperature in 5% non-fat dry milk and 3% bovine serum albumin in Tris-buffered saline/0.1% Tween-20 (TBS-T), the membrane was incubated with rabbit polyclonal anti-FGF14 antibody (1:1,500; catalog no. PA5-77979, Invitrogen) and mouse monoclonal anti-beta tubulin antibody (1:3,000; clone D3U1W, catalog no. 86298, Cell Signaling Technology) overnight at 4°C. The membrane was then washed three times for five minutes with TBS-T and next incubated for one hour at room temperature with horseradish peroxidase-conjugated goat anti-rabbit IgG antibody (1:10,000; catalog no. 111-035-003, Jackson ImmunoResearch) and rabbit anti-mouse IgG antibody (1:10,000, catalog no. ab97046, Abcam). Blots were next washed three times in TBS-T and then developed using Amersham ECL Prime Western Blotting Detection reagents (catalog no. RPN2232, GE Healthcare) or SuperSignal West Femto Maximum Sensitivity Substrate (catalog no. 34095, Thermo Scientific) for fainter signal. Densitometric analysis of signal intensity was performed using ImageJ.[16]

Immunoblotting on iPSC-derived motor neurons

Proteins were extracted from cultures of iPSC-derived motor neurons by resuspending cell pellets in RIPA buffer containing protease inhibitor and incubated on ice for 5min. Lysates were stored at -80°C. Lysates were sonicated and centrifuged at 13,000*g* at 4°C for 10min to remove cell debris. Supernatant was collected and protein concentration was measured using the Pierce BCA Protein Assay kit (catalog no. 23225, Thermo Scientific). Protein extracts (25µg) were run on a precast 4-12% Bis-Tris plus gel (catalog no. NW04125BOX, Invitrogen) for one hour at 80V followed by 30min at 100V. The gel was transferred to a PVDF membrane overnight at 4°C. The blot was then blocked in 5% non-fat dry milk with TBS-T for one hour on a shaker at room temperature. Next, the blot was incubated in a 1:1,000 solution of primary antibody (rabbit polyclonal anti-FGF14 antibody, catalog no. ab229610, Abcam[17]; mouse monoclonal anti-GAPDH antibody, clone 6C5, catalog no. 32233, Santa Cruz Biotechnology) in blocking buffer for two hours on a shaker at room temperature. After washing three times in TBS-T, the blot was incubated in a 1:2,000 solution of horseradish peroxidase-conjugated secondary antibody (goat anti-rabbit IgG antibody, catalog no. 7074, Cell Signaling Technology; horse anti-mouse IgG antibody, catalog no. 7076, Cell Signaling Technology) for one hour at room temperature on a shaker. Next, the blot was washed three times in TBS-T and then developed using SuperSignal West PICO PLUS kit (catalog no. 34580, Thermo Scientific). For fainter bands, SuperSignal West Femto (catalog no. 34094, Thermo Scientific) was used to amplify the signal. Densitometric analysis of signal intensity was performed using ImageJ.[16]

Neuropathological examination

Postmortem brain tissue was available from two French Canadian female patients carrying a $(GAA)_{483}$ and $(GAA)_{516}$ expansion, respectively, who died at age 94 and 81, and had signed an informed consent for a

postmortem autopsy and brain tissue usage in medical research. Following brain removal, the brain tissue was fixed in 10% buffered formalin and further dissected and processed for paraffin histology as per standard procedures. Sections of the frontal, temporal, parietal, occipital, primary motor cortices, anterior and posterior cingulate, amygdala, hippocampus, caudate, putamen, globus pallidus, hypothalamus at the level of the mammillary bodies, deep white matter, midbrain, pons, superior and inferior medulla, cervical, thoracic and lumbar cord, and cerebellar vermis and lateral cerebellar hemisphere with dentate nuclei were sampled for histology. All sections were stained with routine hematoxylin and eosin, and specific regions were further examined with Luxol-fast blue and Bielschowsky tinctorial stains and immunohistochemistry. The sections were immunostained with anti-amyloid-β (1:50; catalog no. M0872, Agilent Technologies),[18] anti-Tau (PHF-1) (1:100; from Dr. Peter Davies, Albert Einstein College of Medicine, New York, NY, USA),[19] anti-4R-Tau (ET3) (1:200; from Dr. Peter Davies, Albert Einstein College of Medicine, New York, NY, USA),[20] anti-alpha-synuclein (LB509) (1:200; catalog no. 180215, Thermo Fisher),[21] anti-Glial Fibrillary Acidic Protein (GFAP) (1:500; catalog no. Z0334, Agilent Technologies),[22] anti-p62 (1:100; catalog no. NBP1-49956, Novus Biologicals),[23] and anti-calbindin (CB-955) (1:400; catalog no C9848, Sigma). Immunohistochemistry was performed using the Benchmark Ultra IHC platform with the OptiView DAB IHC detection system (Ventana Medical Systems Inc.) as per the published manufacturer specifications. All staining was performed with appropriate negative and positive controls. Neuropathological examination was performed by two board certified neuropathologists (J.A.S. and Z.J.).

Statistical analyses

The primary hypothesis of this study was that intronic $(GAA)_{\geq 250}$ repeat expansions in *FGF14* are associated with LOCA. Due to the low prevalence of LOCA, the patient cohort included in this study was a sample of convenience. The control cohort was also a sample of convenience from Montreal (QC, Canada) and Germany. No formal sample size calculation was performed. To address the primary hypothesis, we estimated the association between the *FGF14* $(GAA)_{\geq 250}$ repeat expansion and LOCA by the Fisher's exact test. We used 250 repeat units as the threshold for association studies since this was guided by segregation study within affected families that showed the smallest expansion among symptomatic subjects to be $(GAA)_{250}$. Our hypothesis was tested in two independent case-control series, consisting of (1) 66 French Canadian index patients and 209 controls and (2) 228 German index patients and 199 controls. Effect sizes were reported as odds ratio (OR) and 95% confidence intervals (CIs) were computed using the Baptista-Pike method. We tested association separately in the French Canadian and German cohorts due to the suspected founder effect and allelic enrichment in the former group. The frequency of the GAA repeat expansion (dichotomized as ≥250 triplets for expansion or <250 triplets for non-expansion) was compared between the respective index patient cohorts and control cohorts. Furthermore, the Cohen's d statistics (standardized mean difference) was used to measure the effect size of the difference between mean anchored IRR counts measured by ExpansionHunter Denovo in the 1000 Genomes Project and VAFR control cohorts and the six French Canadian patients with LOCA. Correlations were calculated using the

Pearson correlation coefficient. The widths of the confidence intervals have not been adjusted for multiplicity. Thus, the confidence intervals should not be used to reject or not reject effects. *P* value of < 0.05 was considered significant. All analyses were two-tailed.

SUPPLEMENTARY RESULTS

Repeat motif of the *FGF14* repeat locus

A recent genome-wide search of expanded tandem repeats in a large cohort of individuals with autism spectrum disorder (ASD) showed enrichment of polymorphic non-GAA repeats at the same locus in *FGF14* (chr13:102,160,822-102,162,469, GRCh38) in this population compared to controls, suggesting that *FGF14* may contribute to the complex heritability of ASD.[24] Seven of 17,231 genomes in the ASD cohort showed expansion of one of the following motifs: GAAGGA, GAAGAG, GAAAGAAGAA. Expansion of the GAA repeat unit was, however, not reported in either the patient or control cohorts.

We characterized sequence polymorphism of the *FGF14* repeat locus in the 1000 Genomes Project and Vanderbilt Atrial Fibrillation Registry (VAFR) control cohorts. 311 of 2,504 samples (12.4%) of the 1000 Genomes Project cohort and 301 of 1,115 samples (27.0%) of the VAFR cohort had evidence of at least five anchored in-repeat reads (IRRs) using ExpansionHunter Denovo, version 0.6.2,[8] indicating the existence of a repetitive region of >175bp (~ >58 trinucleotides). Among all samples showing ≥5 anchored IRRs, a GAA repeat motif was detected in the large majority (1000 Genomes Project cohort: 294/311 [94.5%]; VAFR cohort: 278/301 [92.4%]). Alternative motifs were detected with low frequency as shown in Figure S7 and included: GAAGGA, GAAGAG, GAAGCA and GCA.

We next analyzed the *FGF14* repeat locus for sequence variation by Sanger sequencing in 416 randomly sampled control and patient chromosomes of variable length as measured by long-range PCR. The repeat motif / sequence of some larger alleles could not be accurately determined due to limitations in Sanger sequencing of large alleles. We observed that small alleles of ≤25 triplets all exhibited uninterrupted GAA repeat units. The smallest alleles had 8 GAA repeat units. Intermediate alleles showed variable degrees of polymorphism. Alternative repetitive sequence motifs were observed in seven of 136 intermediate alleles (5.1%), and included $(CAA)_n$, $(GGA)_n$, $(GAG)_n$, $(GAAA)_n$ and $(GAAGGA)_n$ repeats. Variable degrees of alternative interspersions within the GAA sequence were observed in a large proportion of the remaining intermediate alleles (100/129; 77.5%). The following interrupting motifs were observed: (CAA), (GAC), (GAG), (GCA), (GGA), (GAAA), (GGAA), (GAAGAG), (GAAGGA) and (GAAGAAA). The remaining alleles (29/129; 22.5%) showed uninterrupted GAA repeat tracts. Expanded chromosomes above the pathogenic threshold of ≥250 triplets consisted of GAA units (64/73; 87.7%), GAAGGA units (8/73; 10.9%) and $[(GAA)_4(GCA)_1]$ units (1/73; 1.4%) in the chromosomes studied.

Ancestry analysis of 14 affected persons

Six patients with GAA-*FGF14*–related ataxia from the French Canadian discovery cohort, seven patients with GAA-*FGF14*–related ataxia from the German cohort and one patient with GAA-*FGF14*–related ataxia from the Australian cohort were assessed for ancestry and relatedness with one another. Four pairs of

samples were already known to be related to each other – the six patients from the French Canadian discovery cohort comprised three pairs of persons from Families I, II and III (Figure 1A); and one pair of persons from the German cohort reported to be cousins. We analyzed the 14 samples with Peddy.[25] Relatedness was supported for each of the four pairs of persons, but no other combinations of persons showed any evidence of being even distantly related to each other to the extent that is detectable with this method. We further used Peddy to inspect the ancestry of these persons. All persons self-reported to be of European ancestry, and that is also what the Peddy analysis indicated (Figure S15). In summary, while all fourteen persons in whom we were able to assess relatedness and ancestry showed patterns consistent with European ancestry, they did not reveal any previously unknown or distant relations.

Analysis of size of the *FGF14* repeat locus in diverse control populations

Since the majority of patients with GAA-*FGF14*–related ataxia in our study are of European ancestry, we sought to determine whether the distribution of sizes of this repeat were greater among European controls than among control persons of other ancestries. For this analysis, we collected ExpansionHunter calls on each person in the 1000 Genomes Project cohort. Figure S16 shows histograms of the number of GAA repeat units observed in each person stratified by superpopulation. This reveals that European and South Asian persons do have alleles greater than the reference length of 50 repeat units more often than the other superpopulations. However, large alleles are found in these other populations as well.

Considering these results, we repeated the analysis shown in Figure 1C, but restricted the 1000 Genomes Project data to only the subset of persons of European descent. Figure S17 shows that the European subset of the 1000 Genomes Project dataset has a comparable distribution of anchored in-repeat reads (IRRs) to the VAFR cohort, which is also overwhelmingly of European descent. In this analysis, the Cohen's d value for enrichment in the six French Canadian patients with LOCA relative to the European 1000 Genomes Project samples was only 1.90 (95% confidence interval [CI]: 1.08 – 2.71), as opposed to 2.99 (95% CI: 2.19 – 3.80) when including all 1000 Genomes Project populations.

Shared disease haplotype analysis

Since the majority of patients found so far with GAA-*FGF14*–related ataxia are of European ancestry and particularly due to the high rate of observation within the French Canadian cohort, we tested for evidence of a shared disease haplotype among affected persons. We used the linkage analysis from Family I to resolve a candidate 693 kilobases haplotype composed of 50 common polymorphic sites, the great majority of which are intronic. The haplotype was delimited by the polymorphisms rs12856547 and rs34644481. Examination of these polymorphic sites in four French Canadian patients from two unrelated families suggested a shared haplotype among the patients, though with some recombination downstream of the repeat locus in two persons (Figure S18). This shared haplotype among French Canadian patients is unsurprising given that population's relatively small set of founders.[26] Analysis of these 50 sites in one

member of the Australian cohort did not rule out the possibility of that patient sharing the haplotype as well. This analysis was not able to be performed on additional members of the French Canadian cohort or members of the German cohort since only whole-exome sequencing data were available, which by in large did not cover haplotype-specific informative SNPs.

Germline instability of the GAA repeat expansion

We studied the intergenerational transmission of the GAA repeat in 21 meioses involving small and uninterrupted alleles ≤25 GAA repeat units (range, 9 to 17 triplets). All meioses resulted in no change in allele size in the offspring. In contrast, in 6 of 16 meioses involving intermediate alleles of 43 to 144 repeats, we observed an increase in repeat size ranging from 15 to 28 triplets. Intermediate alleles that were faithfully transmitted (10 of 16, 62.5%; three paternal events and seven maternal events) carried interruptions within the GAA repeat sequence, whereas alleles that underwent expansion during meiosis (6 of 16, 37.5%; six maternal events) consisted of uninterrupted GAA repeat tracts.

Repeat length to phenotype correlations

We examined the association between the age at onset of disease and the size of the repeat expansion. We found a weak inverse correlation between either age at onset of first symptom or permanent ataxia and the size of the largest allele (N=120, Pearson correlation coefficient, -0.31, R^2=0.10; N=118, Pearson correlation coefficient, -0.31, R^2=0.10, respectively). Using aggregate repeat counts in subjects homozygous or compound heterozygous for $(GAA)_{\geq 250}$ alleles did not yield more robust correlations (Age at onset of first symptom: N=120, Pearson correlation coefficient, -0.28, R^2=0.08; Age at onset of permanent ataxia: N=118, Pearson correlation coefficient, -0.30, R^2=0.09). Similar results were obtained when using aggregate repeat counts in all patients (Age at onset of first symptom: N=120, Pearson correlation coefficient, -0.24, R^2=0.06; Age at onset of permanent ataxia: N=118, Pearson correlation coefficient, -0.28, R^2=0.08). These results must be interpreted in light of the difficulty in precisely establishing an age of onset given the frequent episodic presentation and slowly progressive course of the disease and will require further exploration in larger cohorts.

Among all patients with GAA-*FGF14*–related ataxia, we identified four persons homozygous or compound heterozygous for $(GAA)_{\geq 250}$ expansions. Age at onset of first symptom was not statistically significantly different from heterozygous patients (55.3 vs 57.5 years; Mann-Whitney U test P>0.05; Figure S19, red triangles). While they appeared to have a more severe disease course compared to heterozygous patients, as exemplified by the higher SARA (Scale for the Assessment and Rating of Ataxia)[27] score after similar disease evolution and need for walking aid after only two years of disease evolution in two patients, the small sample size prevents definitive conclusion. These results nonetheless raise the possibility of codominance of the *FGF14* repeat expansion.

Neuropathologic examination

In addition to cerebellar cortical atrophy, in one of the patients (aged 94 years) postmortem examination also revealed diffuse neocortical, Braak stage 6 Lewy body pathology, but with minimal neuronal depletion in the substantia nigra.[28] Population-based autopsy studies of the prevalence of Lewy body disease suggests that this may represent concomitant independent age-related pathological process from GAA-*FGF14*–related ataxia.[29-31] In this patient, neurofibrillary tangle tau pathology, restricted to limbic regions and corresponding to Braak and Braak stage II was also evident.

***FGF14* expression in fibroblasts and lymphoblasts**

Quantitative PCR failed to detect expression of *FGF14* in control and patient fibroblasts. This was confirmed by RNAseq performed on fibroblasts derived from three controls and three patients. Transcript expression was undetectably low in lymphoblasts, as reported previously.[32]

Linkage analysis

Post-hoc two-point parametric linkage analysis performed on the large French Canadian Family I (see Figure 1A) using the mutation status as the phenotype identified significant linkage loci with logarithm of the odds (LOD) score >3 on chromosomes 12, 13, 14, and 18 (Figure S21). The highest LOD scores were obtained on chromosome 13 at SNP rs9513827 (LOD=4.31) and rs12870187 (LOD=4.16). This region fully encompasses *FGF14* (Figure S22).

SUPPLEMENTARY FIGURES

Figure S1: Overview of the study workflow

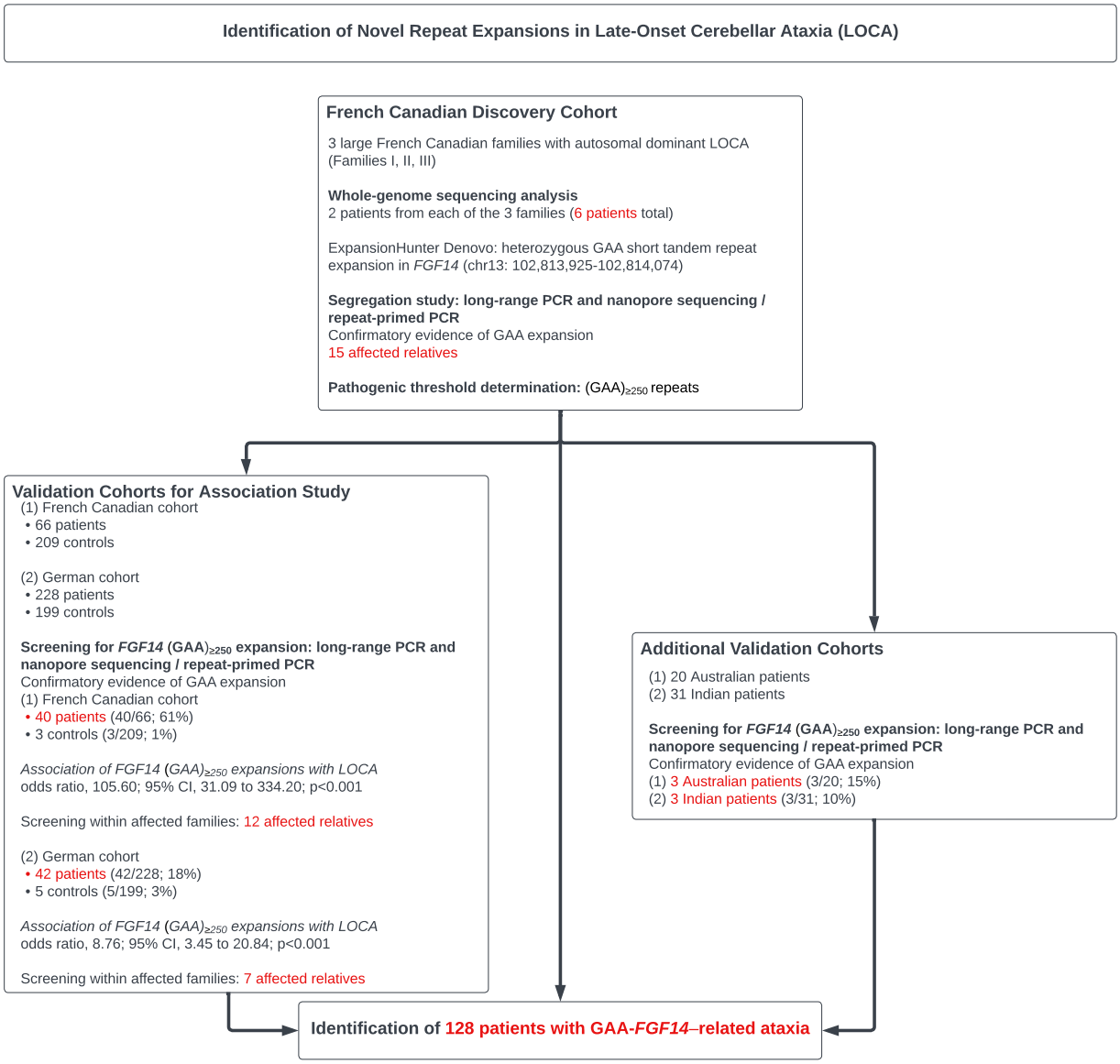
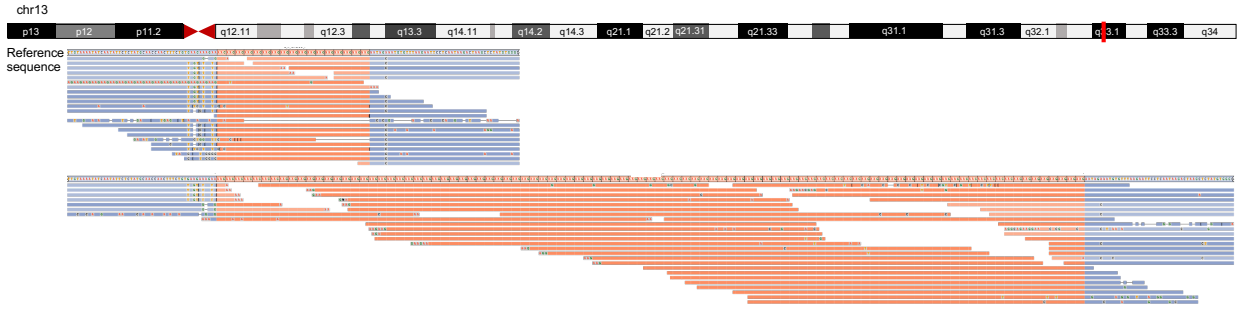
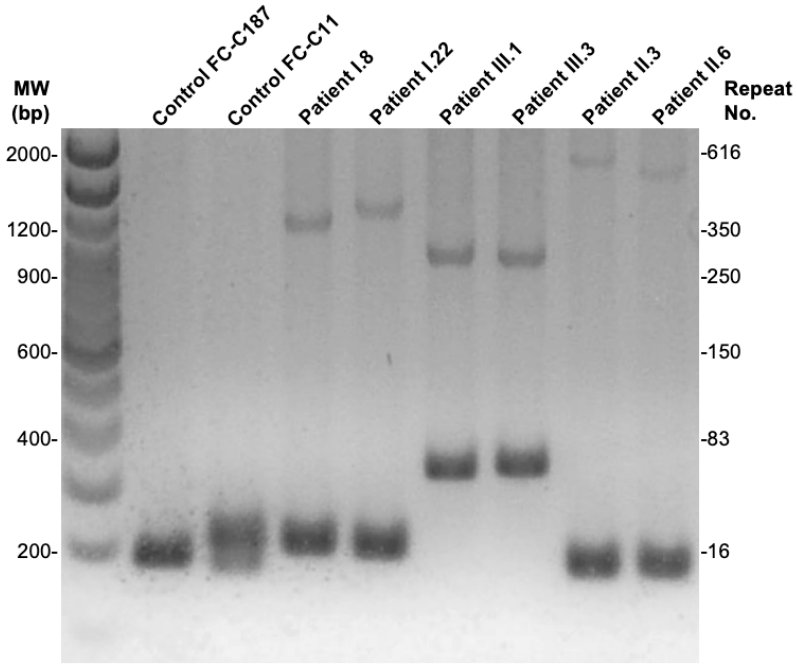


Figure S2: REViewer visualization of the heterozygous GAA repeat expansion in *FGF14*



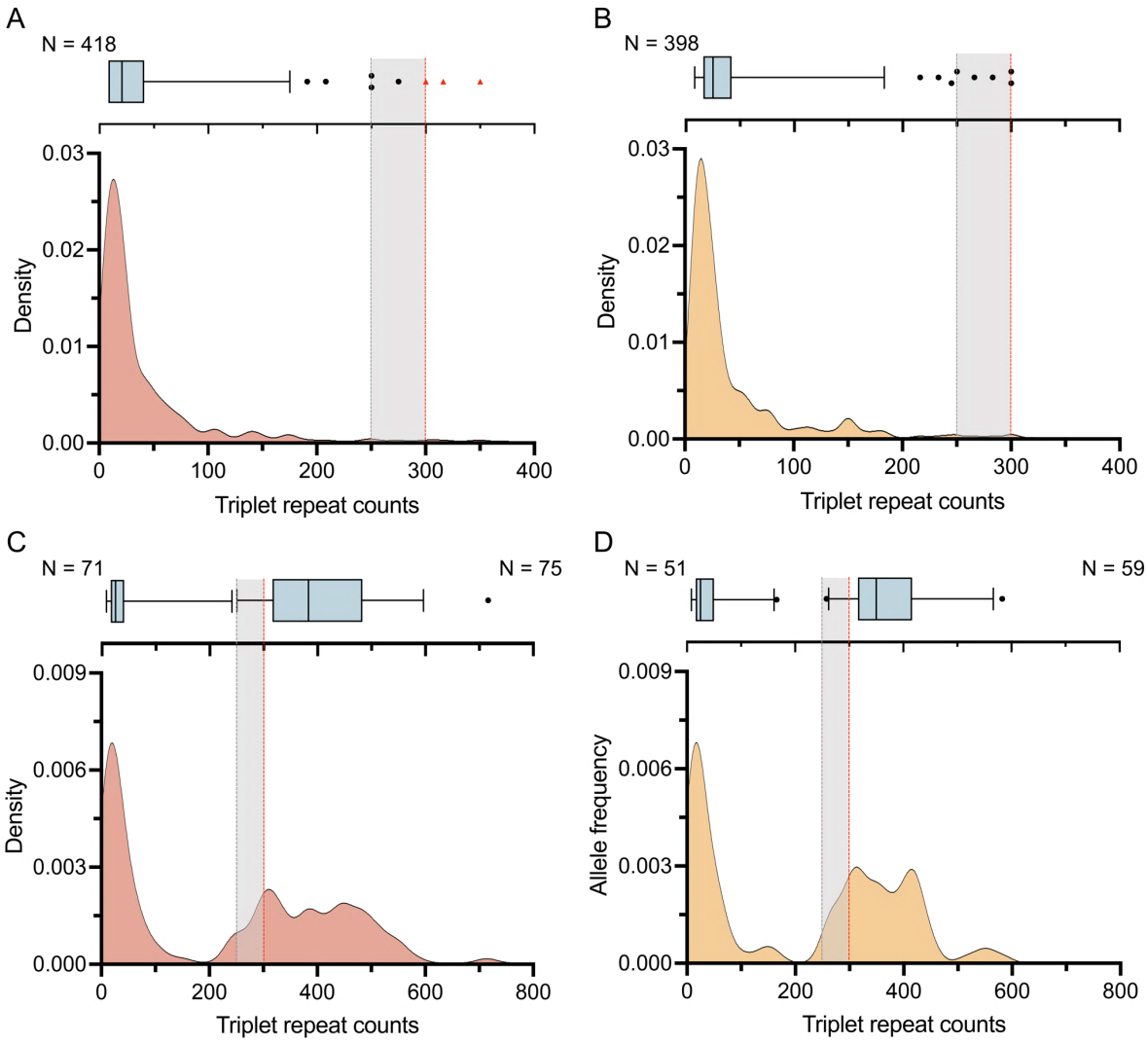
Alignment of short-reads to the *FGF14* intronic locus (chr13:102,813,925-102,814,074; GRCh37) shows the presence of a heterozygous GAA repeat expansion in all six French Canadian patients with LOCA from the discovery cohort who underwent whole-genome sequencing. A representative image from patient I.8 (see Figure 1A) is shown.

Figure S3: Long-range PCR shows a large heterozygous expansion of the *FGF14* repeat locus in patients with late-onset cerebellar ataxia compared to controls



Results of long-range PCR from two controls and the six patients with late-onset cerebellar ataxia (LOCA) from the French Canadian discovery cohort who were found by analysis of whole-genome sequencing to have a large heterozygous expansion at the *FGF14* repeat locus. See Figure 1A, red boxes. MW denotes molecular weight.

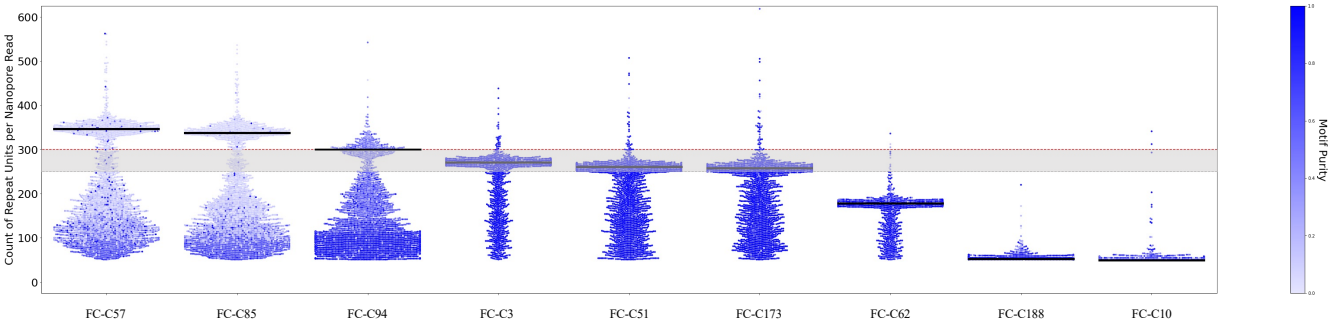
Figure S4: Allele distribution of the *FGF14* repeat locus in control and patient cohorts



Allele distribution of the *FGF14* repeat locus in (A) 209 French Canadian controls (418 chromosomes), (B) 199 German controls (398 chromosomes), (C) 73 French Canadian patients with GAA-*FGF14*–related ataxia (71 normal and 75 expanded chromosomes) and (D) 55 patients with GAA-*FGF14*–related ataxia from the German (49 patients), Australian (3 patients) and Indian (3 patients) cohorts (51 normal and 59 expanded chromosomes). The repeat length was estimated by agarose gel electrophoresis of PCR-amplification products. Among the French Canadian patients with GAA-*FGF14*–related ataxia, two were homozygous or compound heterozygous for $(GAA)_{\geq 250}$ expansions. Among the German patients with GAA-*FGF14*–related ataxia, two were homozygous or compound heterozygous for $(GAA)_{\geq 250}$ expansions, and two were compound heterozygous for a $(GAA)_{\geq 250}$ expansion and a $(GAAGGA)_{\geq 125}$ expansion. The smallest number of GAA repeats among controls and patients was 8. The density plots show allele size frequency, with higher density indicating greater frequency. The box-and-whisker plots show the allelic distribution in

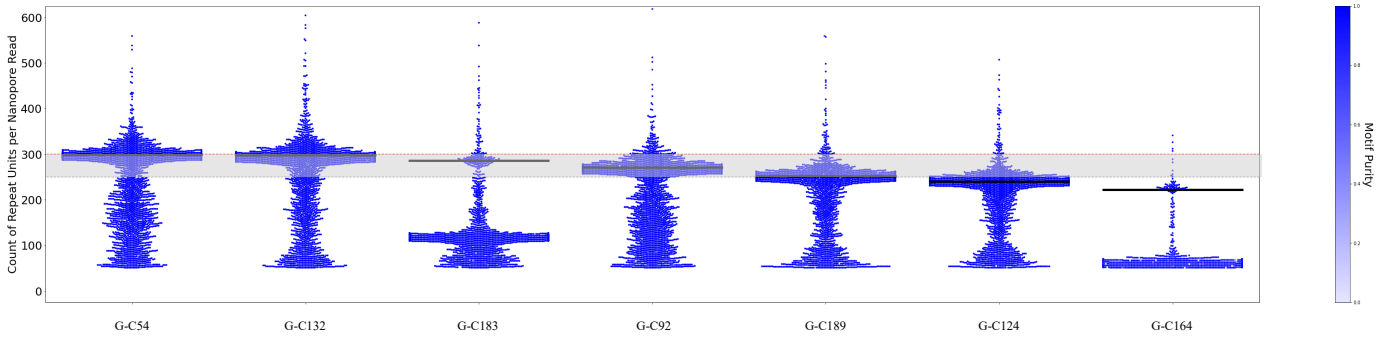
control and patient cohorts. In panels C and D, the box-and-whisker plots above the graphs show the distribution of the normal alleles (left-hand plot) and expanded alleles (right-hand plot) in patients. The box indicates the 25th percentile (first quartile), the median, and the 75th percentile (third quartile), and the whiskers indicate the 2.5th and 97.5th percentiles. Outliers are represented by black dots. In controls, expanded alleles consisting of non-GAA repeats are represented by red triangles. The dashed gray lines and shaded gray areas indicate the incompletely penetrant range of $(GAA)_{250\text{-}300}$, and the dashed red lines mark the threshold of $(GAA)_{300}$ repeat units, above which the alleles are fully penetrant.

Figure S5: Nanopore sequencing of the *FGF14* repeat locus in nine selected French Canadian controls



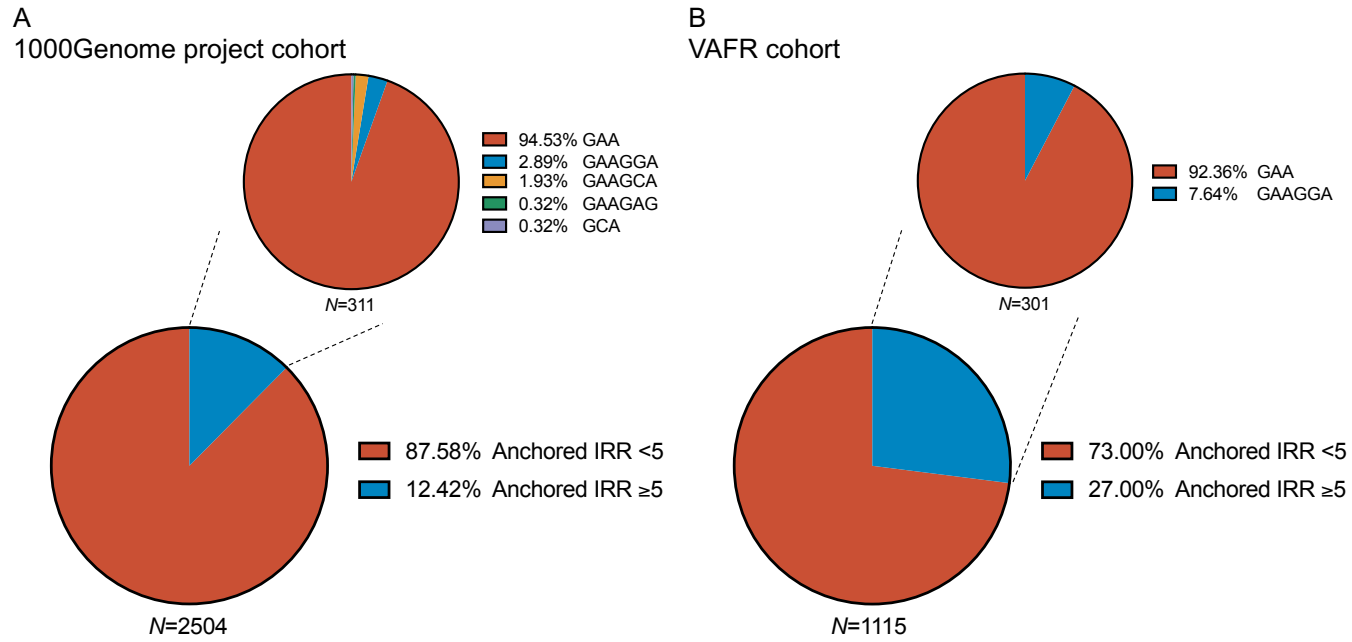
Swarm plots showing 3000 randomly sampled individual nanopore reads containing at least 50 repeat units for nine selected French Canadian controls. Samples are presented in a waterfall arrangement in descending order of allele size. Horizontal black bars indicate repeat size of the larger allele, as measured by nanopore sequencing. The horizontal dashed gray line and the shaded gray area show the incompletely penetrant range of $(GAA)_{250\text{-}300}$, and the dashed red line marks the threshold of $(GAA)_{300}$ repeat units, above which the alleles are fully penetrant. The color of the data points is a function of the GAA repeat motif purity in each individual read, with dark blue indicating pure and lighter blue impure motif (a hue scale is shown on the right y axis). Controls FC-C57 and FC-C85 carry a $(GAAGGA)_n$ expansion while control FC-C94 carry a $[(GAA)_4(GCA)_1]_n$ expansion. Controls FC-C62, FC-C188 and FC-C10 carry an allele <250 GAA repeats as estimated by long-range PCR and nanopore sequencing. None of the GAA expansions exceeded 300 repeat units in controls.

Figure S6: Nanopore sequencing of the *FGF14* repeat locus in seven selected German controls



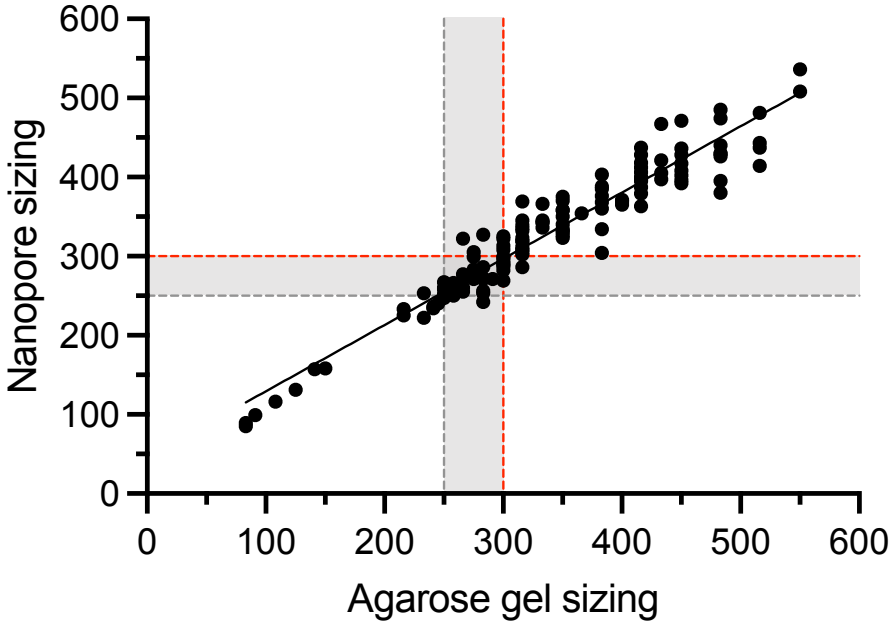
Swarm plots showing 3000 randomly sampled individual nanopore reads containing at least 50 repeat units for seven selected German controls carrying a large allele. Samples are presented in a waterfall arrangement in descending order of allele size. Horizontal black bars indicate repeat size of the larger allele, as measured by nanopore sequencing. The horizontal dashed gray line and the shaded gray area show the incompletely penetrant range of $(GAA)_{250\text{-}300}$, and the dashed red line marks the threshold of $(GAA)_{300}$ repeat units, above which the alleles are fully penetrant. The color of the data points is a function of the GAA repeat motif purity in each individual read, with dark blue indicating pure and lighter blue impure motif (a hue scale is shown on the right y axis). Controls G-C124 and G-C164 carry an allele <250 GAA repeats as estimated by long-range PCR and nanopore sequencing. None of the GAA expansions exceeded 300 repeat units in controls.

Figure S7: Polymorphic configurations of the *FGF14* repeat locus



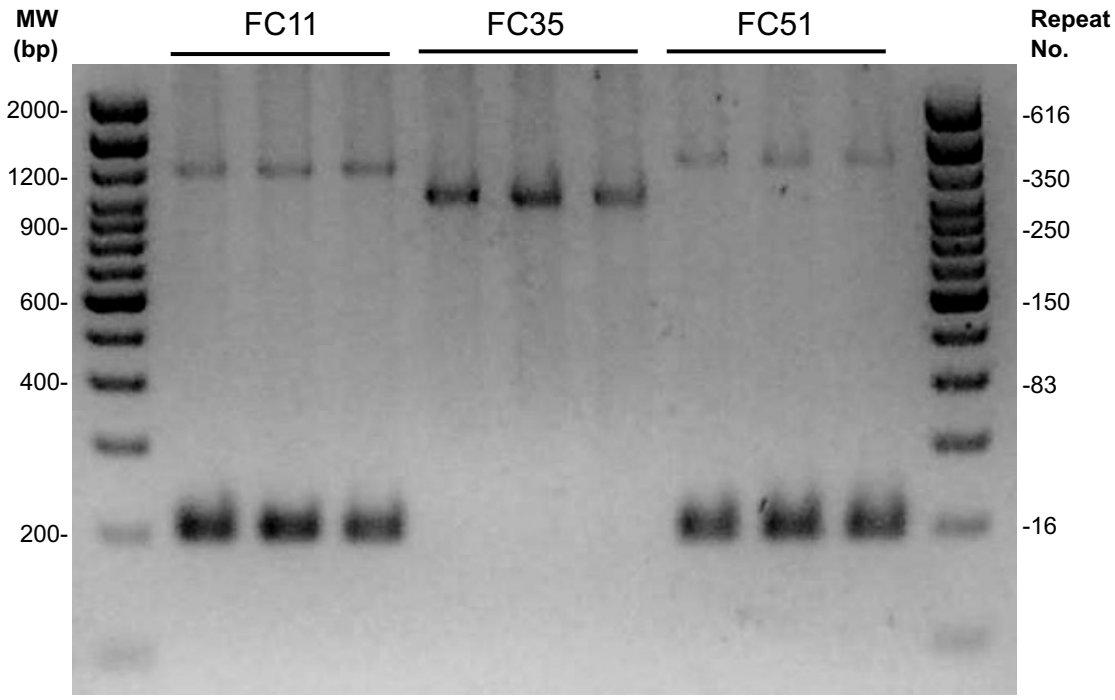
Polymorphic configurations of the *FGF14* repeat locus in (A) 311 of 2504 samples (12.42%) of the 1000 Genomes Project cohort and (B) 301 of 1115 samples (27.00%) of the Vanderbilt Atrial Fibrillation Registry (VAFR) cohort with evidence of at least five anchored in-repeat reads (IRRs) at the locus, supportive of an approximately >58 trinucleotide-long repeat.

Figure S8: Comparison of *FGF14* allele size estimates by long-range PCR and nanopore sequencing



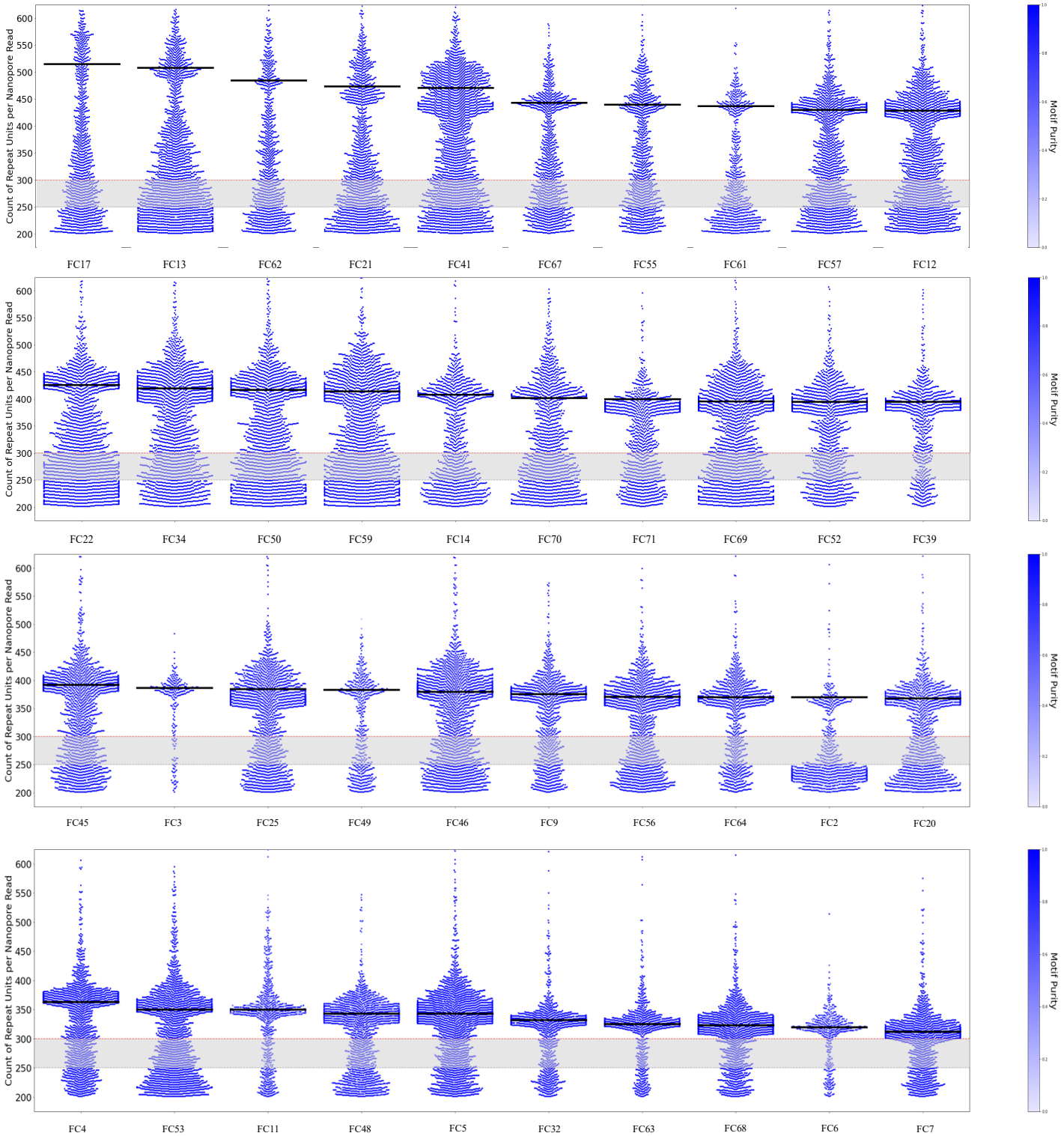
Sizing of 131 alleles by agarose gel electrophoresis and nanopore sequencing is highly similar (Pearson correlation coefficient, 0.96). Nanopore sizing was performed on size-selected (>400bp) PCR-amplification products only; the graph shows data for alleles >80 repeat units (>400bp) only. The repeat size was estimated by a Gaussian mixture model. Suboptimal read enrichment prevented accurate sizing of three larger alleles, which were excluded from this analysis. Amplification products obtained through two separate and independent PCR experiments were used for nanopore and agarose gel electrophoresis sizing. The dashed gray lines and the shaded gray area show the incompletely penetrant range of $(GAA)_{250\text{-}300}$, and the dashed red lines mark the threshold of $(GAA)_{300}$ repeat units, above which the alleles are fully penetrant. Long-range PCR was used as the standard technique to establish repeat size for each individual sample (Figure S9).

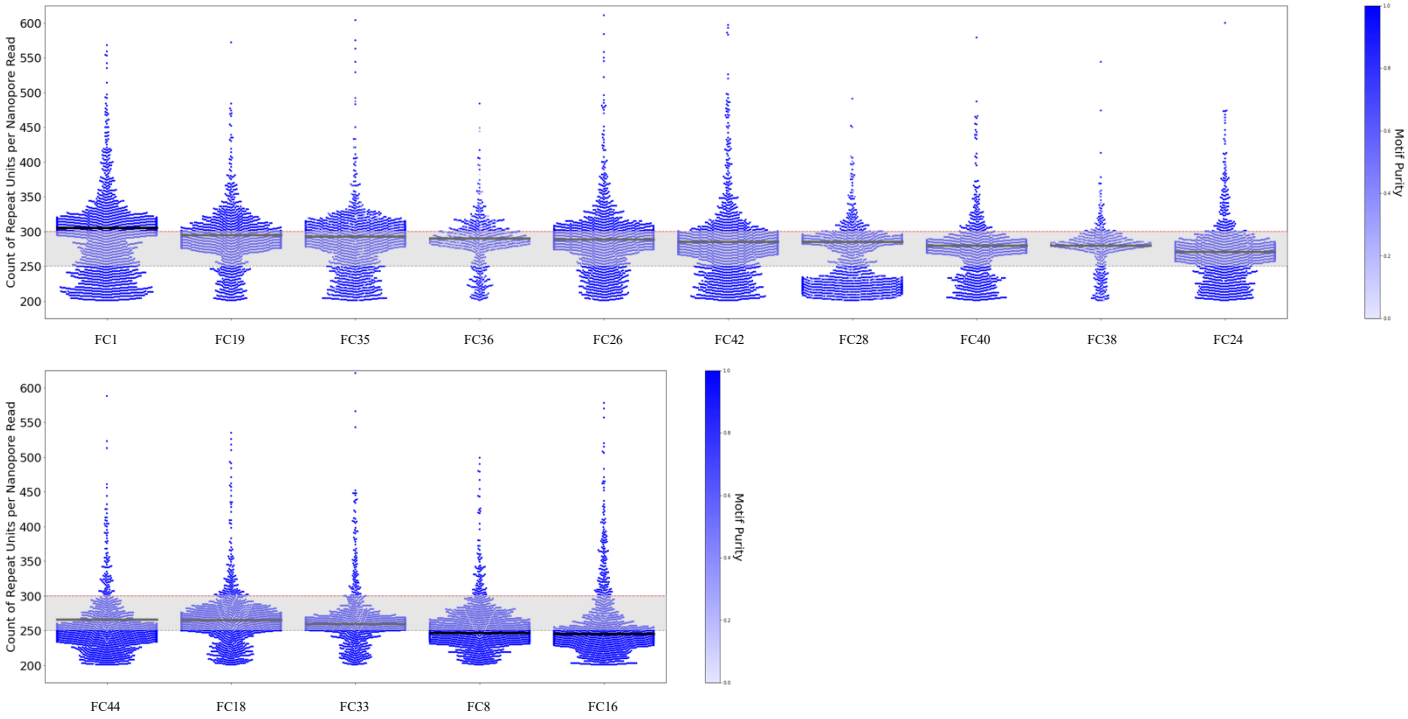
Figure S9: Long-range PCR yields stable repeat expansion size in triplicate experiments



Long-range PCR assay yields stable repeat expansion size across triplicate experiments in three patients with GAA-*FGF14*–related ataxia. Patient FC35 is homozygous for expanded $(GAA)_{300}$ alleles. MW denotes molecular weight.

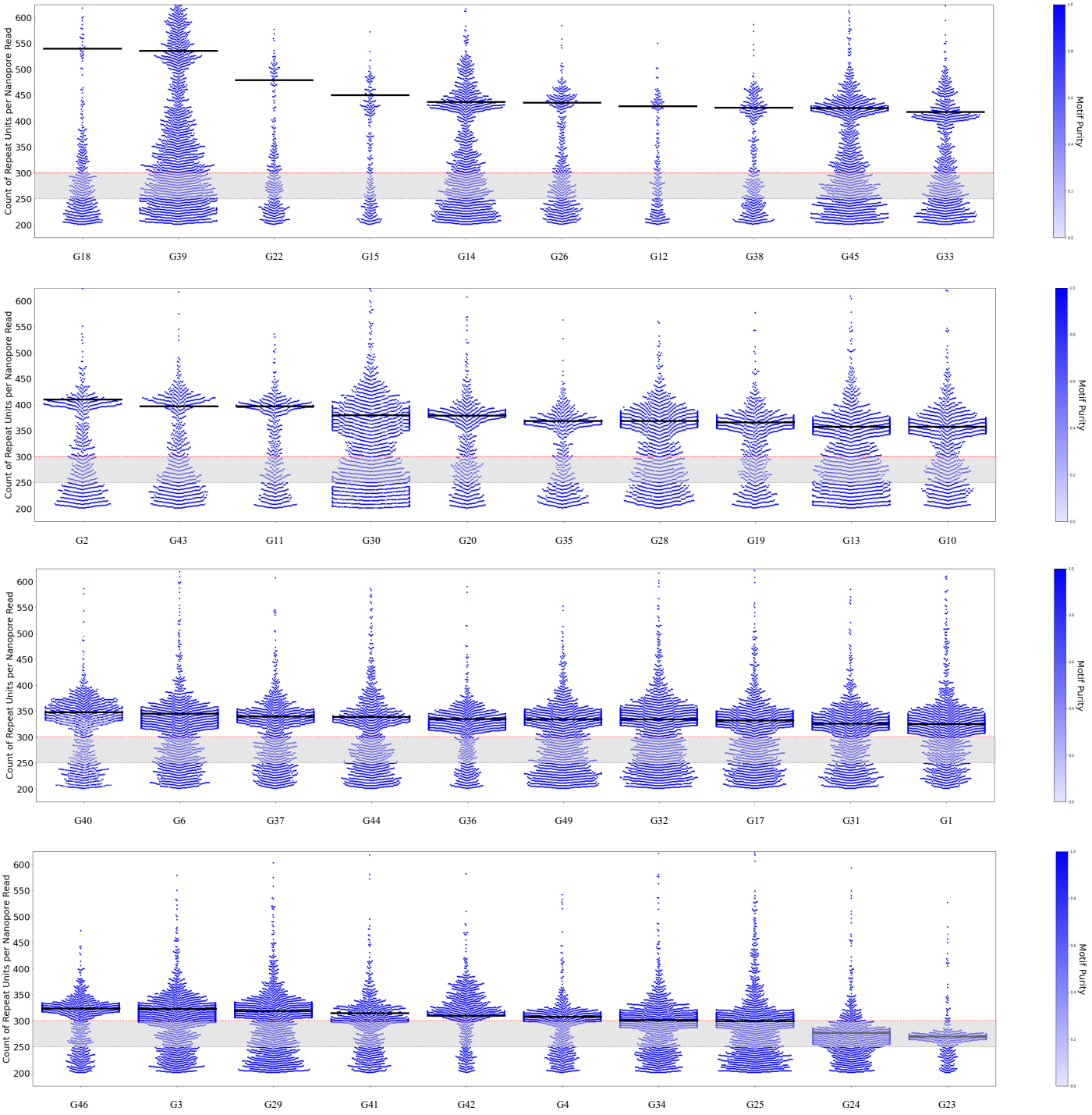
Figure S10: Nanopore sequencing of the *FGF14* repeat locus in 55 French Canadian patients with GAA-*FGF14*–related ataxia

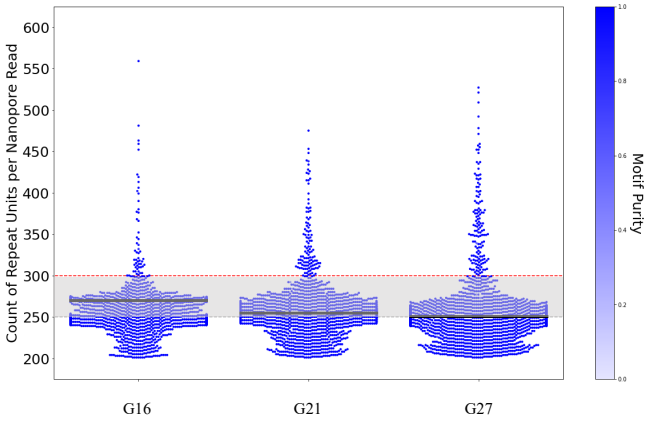




Swarm plots showing 3000 randomly sampled individual nanopore reads containing at least 200 repeat units for 55 French Canadian patients with GAA-*FGF14*–related ataxia. Samples are presented in a waterfall arrangement in descending order of allele size. Horizontal black bars indicate repeat size of the larger allele, as measured by nanopore sequencing. The horizontal dashed gray lines and shaded gray areas show the incompletely penetrant range of $(GAA)_{250\text{-}300}$, and the dashed red lines mark the threshold of $(GAA)_{300}$ repeat units, above which the alleles are fully penetrant. The color of the data points is a function of the GAA repeat motif purity in each individual read, with dark blue indicating pure and lighter blue impure motif (a hue scale is shown on the right y axis).

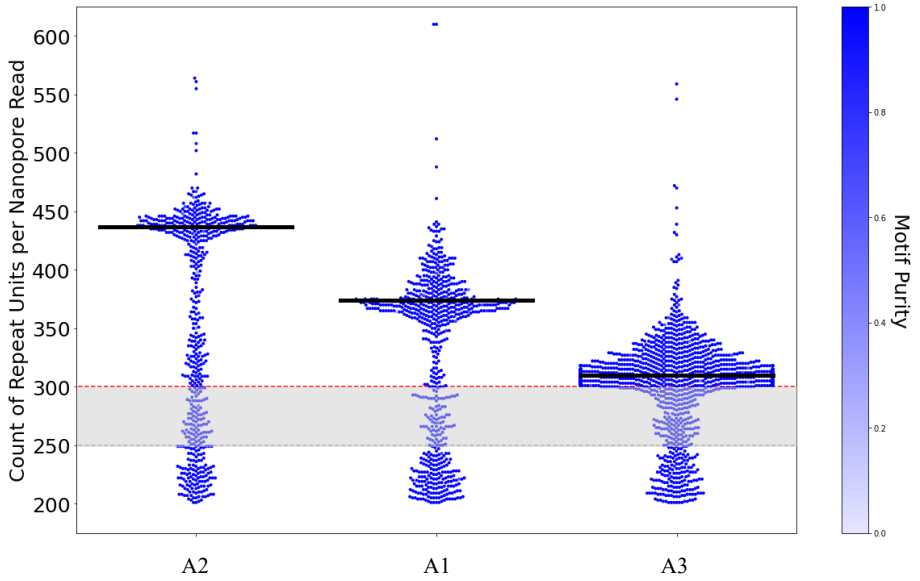
Figure S11: Nanopore sequencing of the *FGF14* repeat locus in 43 German patients with GAA-*FGF14*–related ataxia





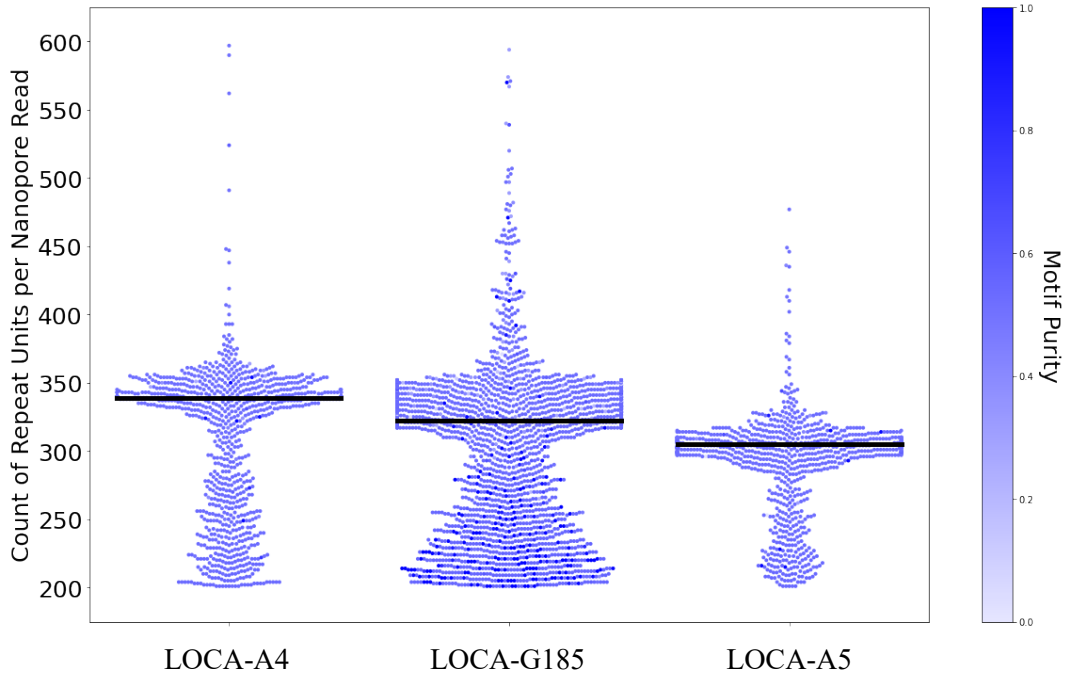
Swarm plots showing 3000 randomly sampled individual nanopore reads containing at least 200 repeat units for 43 German patients with GAA-*FGF14*–related ataxia. Samples are presented in a waterfall arrangement in descending order of allele size. Horizontal black bars indicate repeat size of the larger allele, as measured by nanopore sequencing. The horizontal dashed gray lines and shaded gray areas show the incompletely penetrant range of $(GAA)_{250\text{-}300}$, and the dashed red lines mark the threshold of $(GAA)_{300}$ repeat units, above which the alleles are fully penetrant. The color of the data points is a function of the GAA repeat motif purity in each individual read, with dark blue indicating pure and lighter blue impure motif (a hue scale is shown on the right y axis).

Figure S12: Nanopore sequencing of the *FGF14* repeat locus in three Australian patients with GAA-*FGF14*–related ataxia



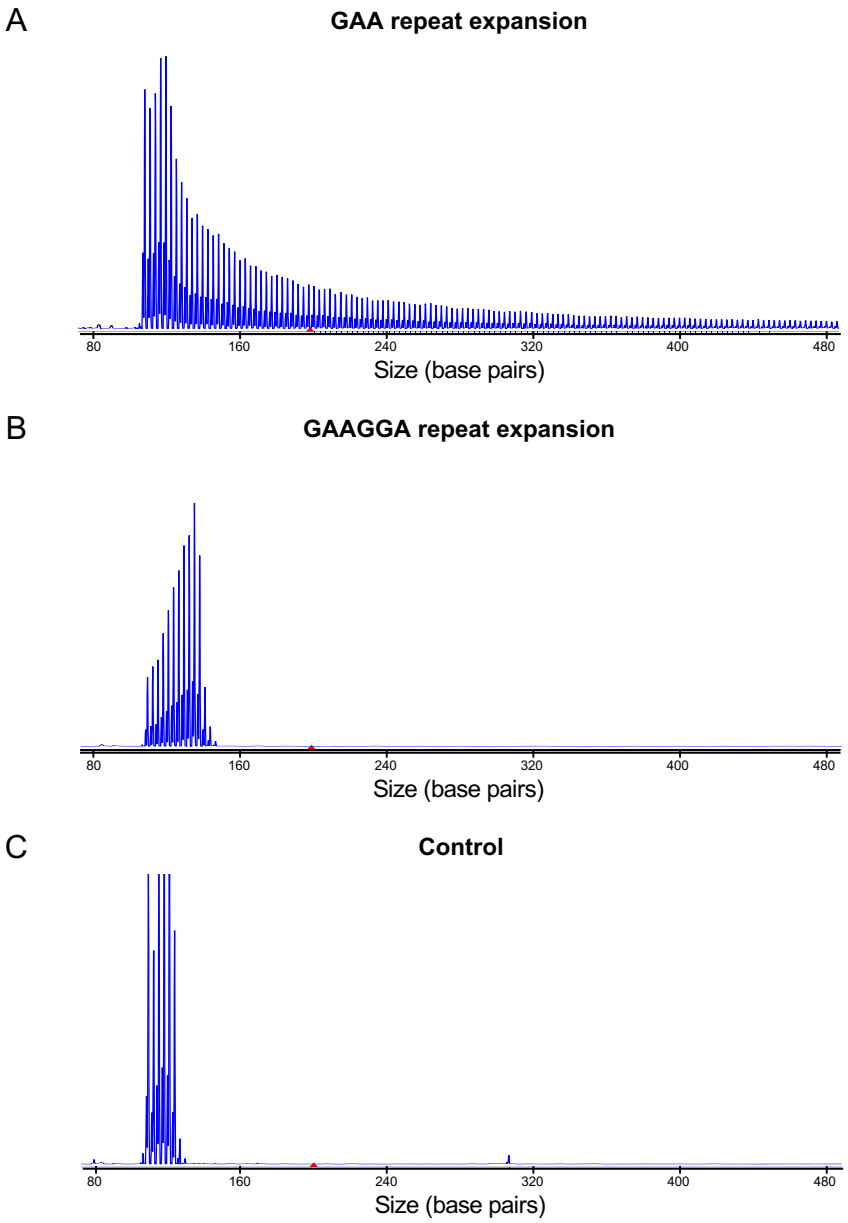
Swarm plots showing 3000 randomly sampled individual nanopore reads containing at least 200 repeat units for three Australian patients with GAA-*FGF14*–related ataxia. Samples are presented in a waterfall arrangement in descending order of allele size. Horizontal black bars indicate repeat size of the larger allele, as measured by nanopore sequencing. The horizontal dashed gray line and the shaded gray area show the incompletely penetrant range of $(GAA)_{250\text{-}300}$, and the dashed red line marks the threshold of $(GAA)_{300}$ repeat units, above which the alleles are fully penetrant. The color of the data points is a function of the GAA repeat motif purity in each individual read, with dark blue indicating pure and lighter blue impure motif (a hue scale is shown on the right y axis).

Figure S13: Nanopore sequencing of the *FGF14* repeat locus in three patients with late-onset cerebellar ataxia shows a large non-GAA expansion



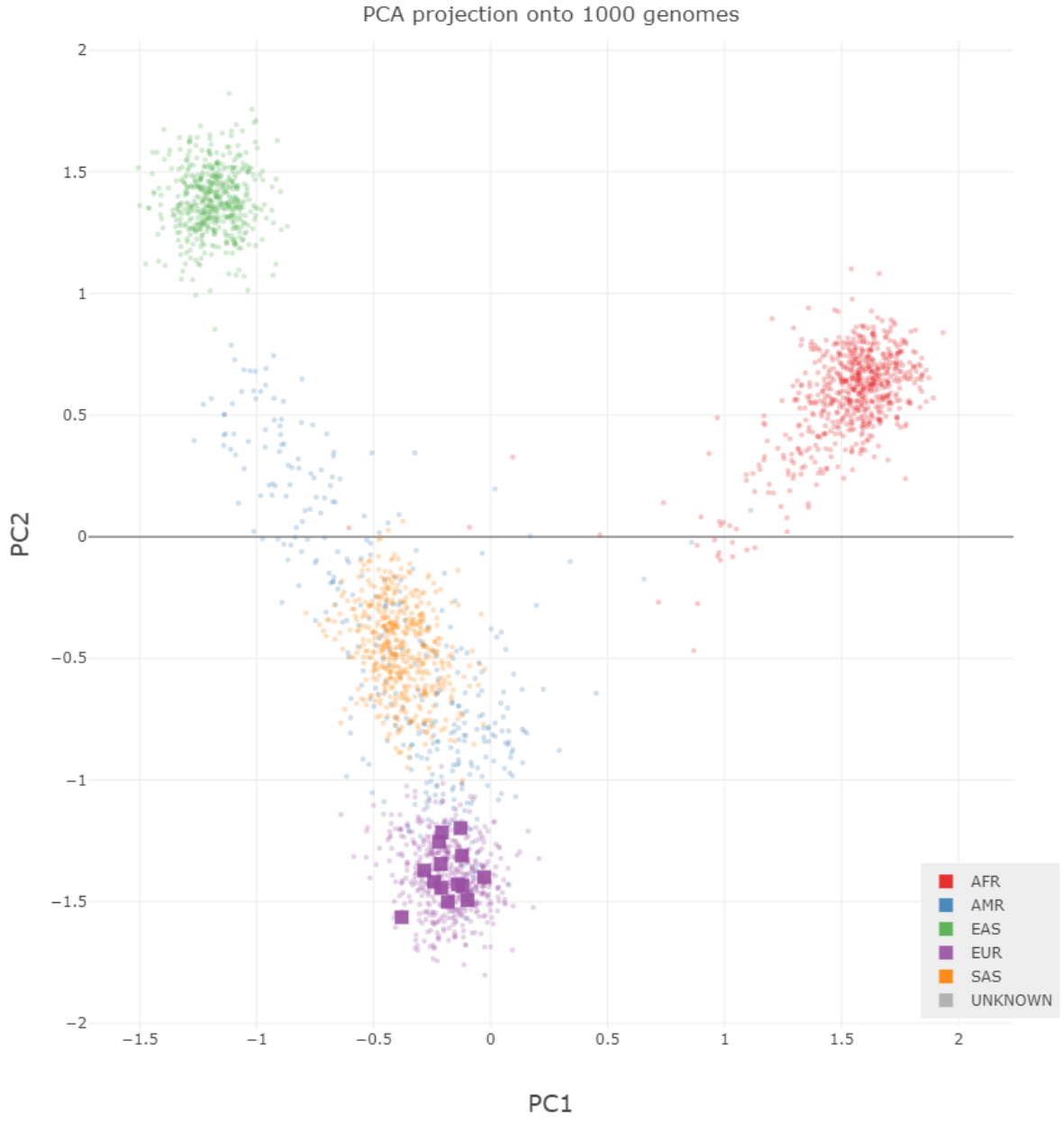
Swarm plots showing 3000 randomly sampled individual nanopore reads containing at least 200 repeat units for three patients with late-onset cerebellar ataxia found to carry a large non-pathogenic $(GAAGGA)_n$ hexanucleotide expansion. Samples are presented in a waterfall arrangement in descending order of allele size. Horizontal black bars indicate repeat size (in triplet equivalent) of the larger allele, as measured by nanopore sequencing.

Figure S14: Repeat-primed PCR analysis of persons carrying a GAA and GAAGGA repeat expansion



Representative plots of the repeat-primed PCR targeting the GAA repeat unit in (A) a patient with GAA-*FGF14*–related ataxia carrying a $(GAA)_{283}$ repeat expansion, (B) a control (Control FC-C57, see Figure 1D and Figure 2C) carrying a GAAGGA repeat expansion and (C) a control homozygous for $(GAA)_{11}$ alleles. Repeat-primed PCR fails to show the characteristic sawtooth pattern in persons carrying a non-GAA expansion, confirming its ability to distinguish GAA from non-GAA expansions.

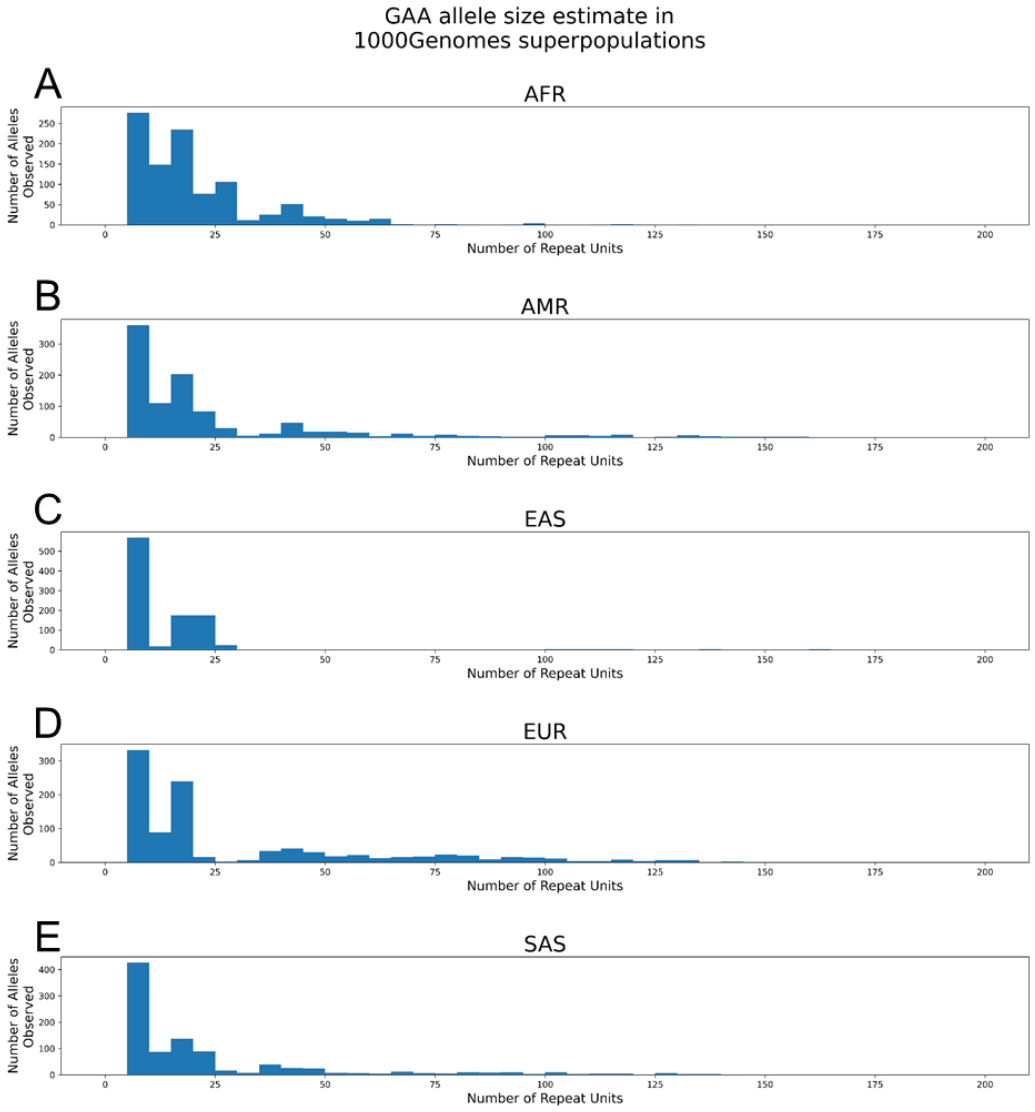
Figure S15: Ancestry classification of 14 patients with GAA-*FGF14*–related ataxia



Scatter plot of the first two principal components (PC) of an ancestry analysis. The 1000 Genomes Project cohort samples are plotted as small dots, colored according to their stated ancestry. Fourteen affected persons carrying the *FGF14* GAA expansion are projected onto the PC axes and plotted as squares. All 14 persons are predicted to be of European origin.

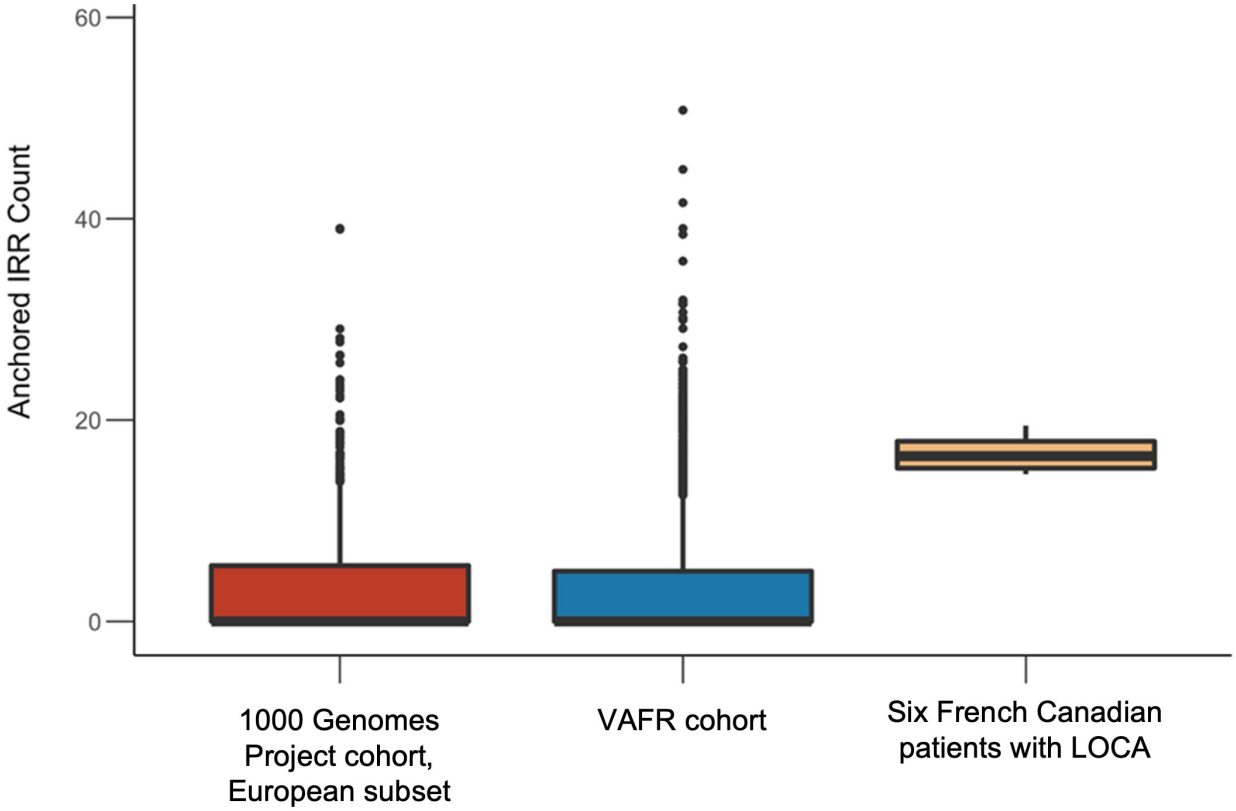
AFR: African; AMR: Admixed American; EAS: East Asian; EUR: European; SAS: South Asian

Figure S16: *FGF14* repeat locus allele sizes in superpopulations of the 1000 Genomes Project cohort



Histograms of the distribution of allele sizes predicted by ExpansionHunter in the 1000 Genomes Project cohort. Persons from different superpopulations are plotted in separate histograms as follows: (A) African (AFR), (B) Admixed American (AMR), (C) East Asian (EAS), (D) European (EUR) and (E) South Asian (SAS).

Figure S17: Distribution of anchored in-repeat reads for the European subset of the 1000 Genomes Project cohort, the VAFR cohort and the six French Canadian patients with GAA-*FGF14*–related ataxia



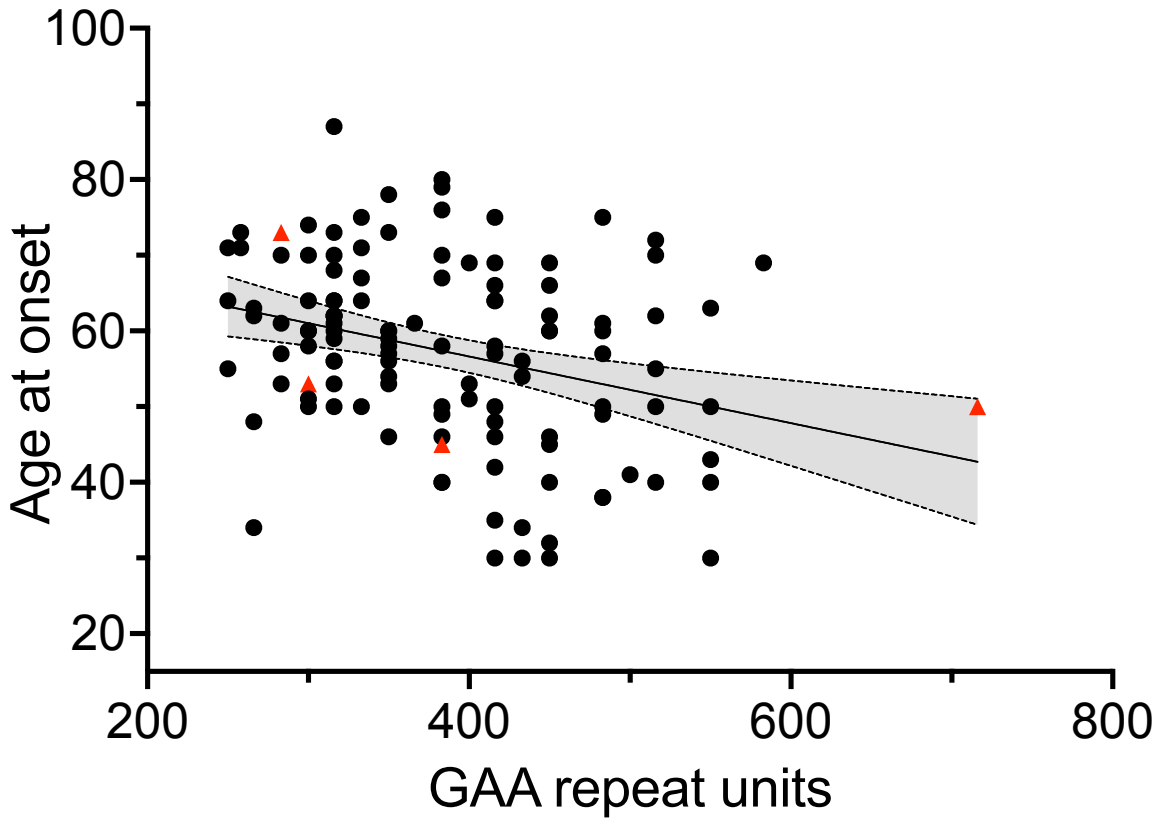
The mean anchored in-repeat reads (IRRs) count at this locus was 16.68 (standard deviation [SD], 1.90; median, 16.40; range, 14.60 to 19.45) in the group of six French Canadian patients with LOCA, higher than that of the 1000 Genomes Project dataset (mean, 3.67; SD, 6.89; median, 0; range, 0 to 39.04; Cohen's d, 1.90 [95% CI, 1.08 to 2.71]) and the VAFR cohort (mean, 4.38; SD, 7.73; median, 0; range, 0 to 50.78; Cohen's d, 1.59 [95% CI, 0.78 to 2.40]). The widths of the confidence intervals have not been adjusted for multiplicity. Thus, the confidence intervals should not be used to reject or not reject effects.

Figure S18: Haplotype analysis of six French Canadian patients and one Australian patient with GAA-*FGF14*–related ataxia

	FC11	FC49	FC36	FC37	FC30	FC13	A1
rs12856547	T	T	T	T	T	T	T
rs75100545	G	G	G	G	G	G	G
rs9518579	T	T	T	T	T	T	T
rs2984842	T	T	T	T	T	T	G
rs1323194	G	G	G	G	G	G	G
rs9300699	G	G	G	G	G	G	G
rs1407785	T	T	T	T	T	T	T
rs116829857	G	G	G	G	G	G	G
rs117078917	G	G	G	G	G	G	G
rs17502818	A	A	A	A	A	A	A
rs11069468	C	C	C	C	C	C	C
rs12877811	T	T	T	T	T	T	T
rs12864397	A	A	A	A	A	A	A
rs996453	G	G	G	G	G	G	G
rs2476238	C	C	C	C	C	C	C
rs9634528	A	A	A	A	A	A	A
rs1457315	C	C	C	C	C	C	C
rs546325	T	T	T	T	T	T	T
rs9557806	A	A	A	A	A	A	A
rs17588797	G	G	G	G	G	G	G
GAA Repeat Locus							
rs9518645	T	T	T	T	T	T	T
rs667631	G	G	G	G	G	G	G
rs9518648	A	A	A	A	A	A	A
rs7317699	G	G	G	G	G	G	G
rs76245724	C	C	C	C	C	C	C
rs1999168	A	A	A	A	A	G	G
rs9513995	T	T	T	T	T	T	T
rs9514001	T	T	T	T	T	C	C
rs7997701	C	C	C	C	C	C	C
rs6491679	T	T	T	T	T	G	G
rs112814402	G	G	G	G	G	G	G
rs1336708	A	A	A	A	A	G	G
rs17590168	C	C	C	C	C	T	T
rs76074726	A	A	A	A	A	A	A
rs1360974	G	G	G	G	G	G	G
rs77850433	C	C	C	C	C	C	C
rs16960116	T	T	T	T	C	C	C
rs12870187	C	C	C	C	C	C	C
rs9518729	T	T	T	T	T	T	T
rs7328480	T	T	T	T	T	T	T
rs1927364	C	C	C	C	T	T	T
rs17590704	C	C	C	C	C	C	C
rs9554877	C	C	C	C	C	C	C
rs79866129	A	A	A	A	A	A	A
rs1927385	A	A	A	A	G	G	A
rs79011294	C	C	C	C	C	C	C
rs9518753	T	T	T	T	T	T	T
rs1927392	C	C	C	C	C	C	C
rs1927393	G	G	G	G	G	G	G
rs34644481	G	G	G	G	G	G	G

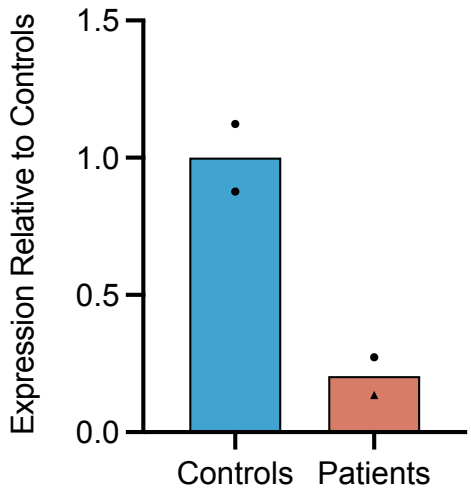
The presence of a disease haplotype composed of 50 common polymorphic loci surrounding the GAA repeat locus was tested in seven samples. Loci containing the disease haplotype polymorphism in a person are displayed in green, while those without the matching polymorphism are displayed in red.

Figure S19: Correlation between size of the *FGF14* GAA expansion and age at onset



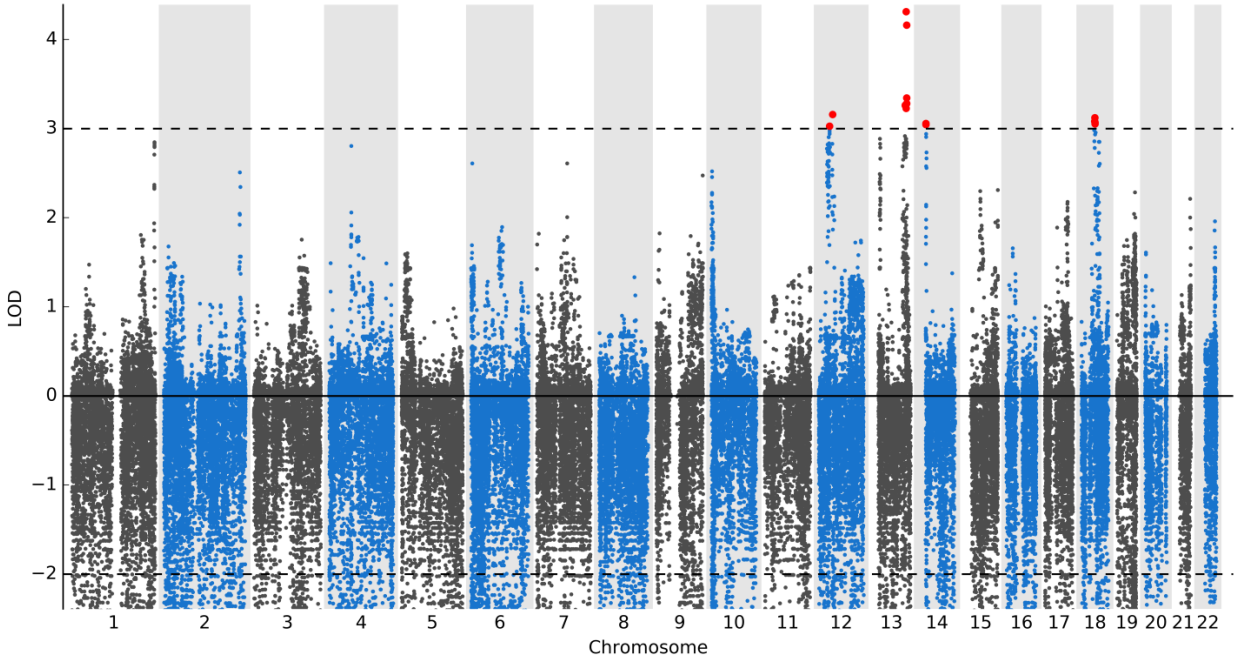
Weak inverse linear correlation between size of the GAA repeat expansion and age at onset of disease in 120 patients (Pearson correlation coefficient, -0.31, R^2=0.10). The gray area displays the 95% confidence interval. Subjects homozygous or compound heterozygous for $(GAA)_{\geq 250}$ alleles are represented by red triangles. The size of their largest expansion was used for analysis. The age at which disease first manifested was used for the analysis.

Figure S20: Total *FGF14* expression in iPSC-derived motor neurons



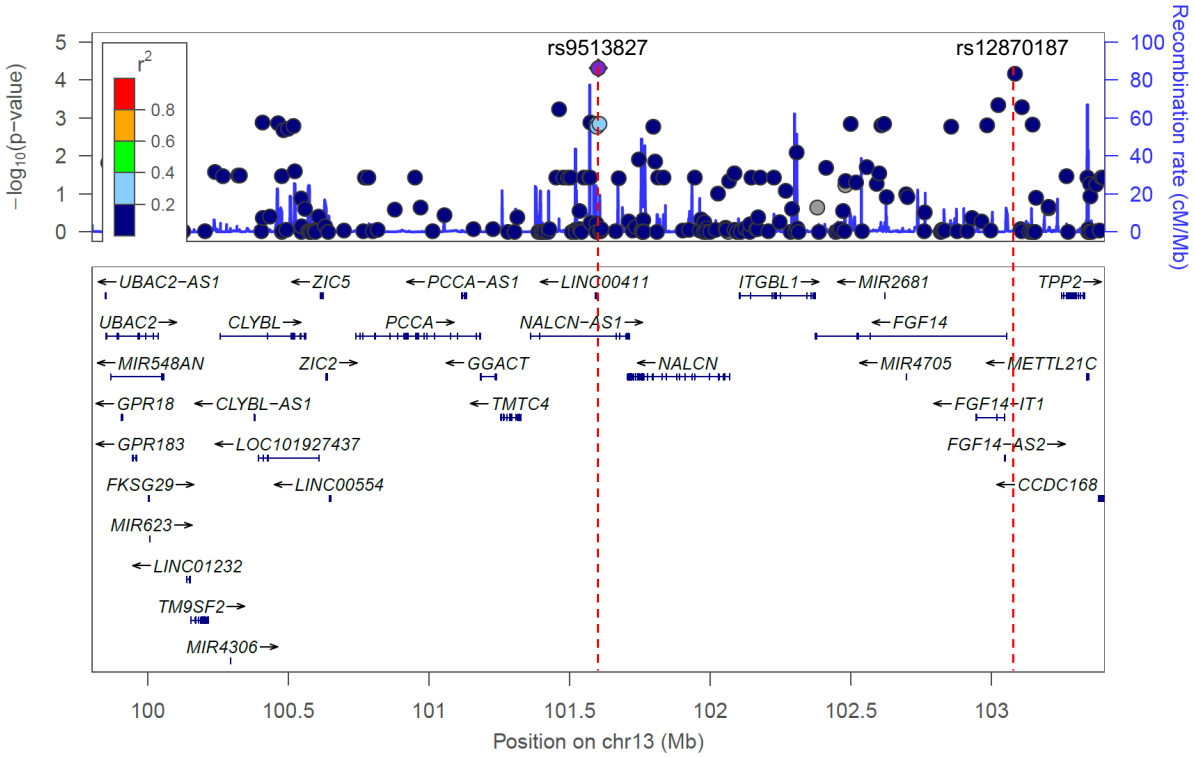
Relative expression of total *FGF14* in iPSC-derived motor neurons of two controls and two patients, normalized to GAPDH, as assessed by quantitative PCR. Relative quantification was computed by the $2^{-\Delta\Delta Ct}$ method, with the use of the mean value among controls as calibrator; values are represented as the ratio of the mean expressive relative to the mean among controls. Bars indicate the mean, and black dots and triangle the data distribution. Black triangle indicates a patient who was homozygous for $(GAA)_{300}$ expansions.

Figure S21: Post-hoc genome-wide parametric linkage analysis in Family I



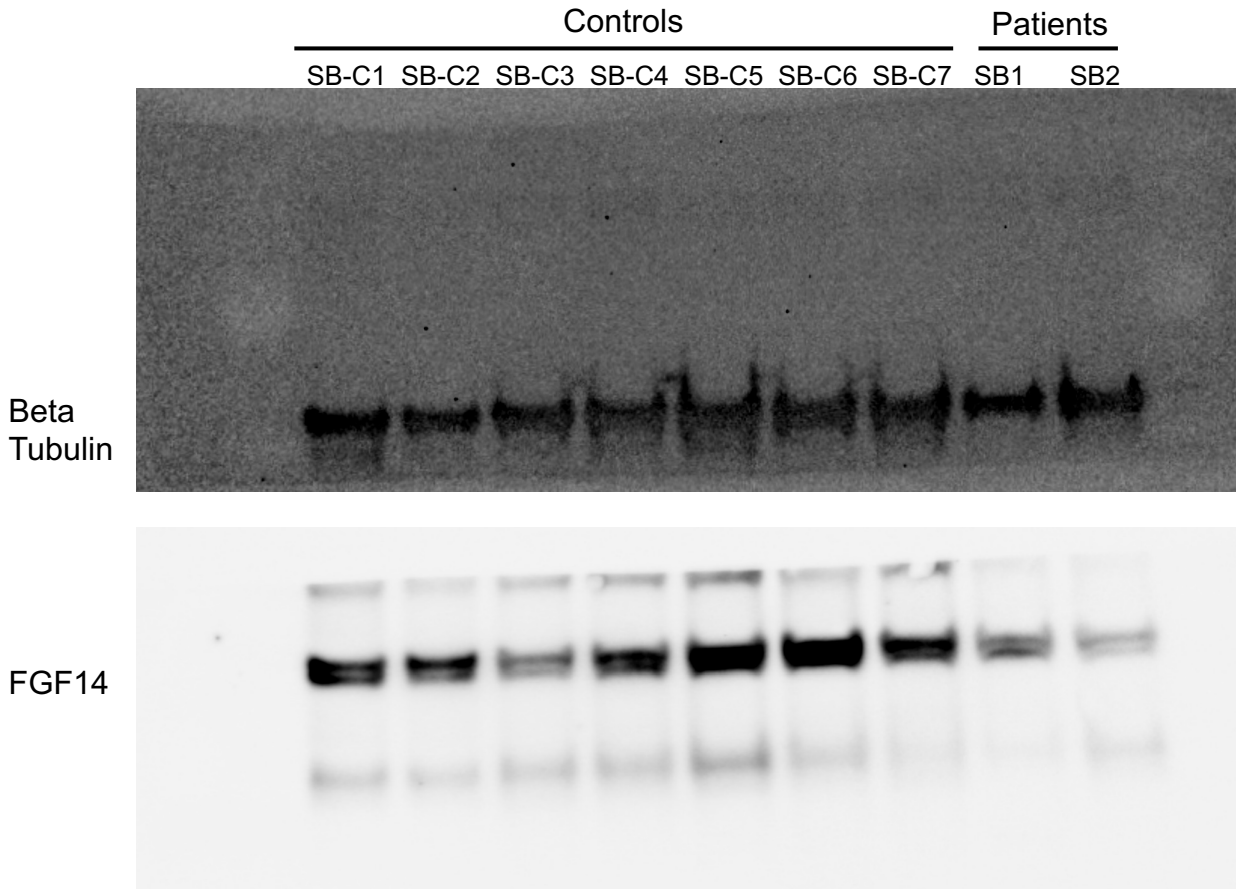
Following discovery of the intronic GAA repeat expansion in *FGF14*, two-point parametric linkage analysis using the mutation status as the phenotype in 18 persons belonging to Family I (see Figure 1A) identified significant linkage loci with LOD score >3 on chromosomes 12, 13, 14, and 18. The highest LOD scores were obtained on chromosome 13 at SNP rs9513827 (LOD=4.31) and rs12870187 (LOD=4.16).

Figure S22: Genomic regional plot of the linkage region with the highest LOD score on chromosome 13



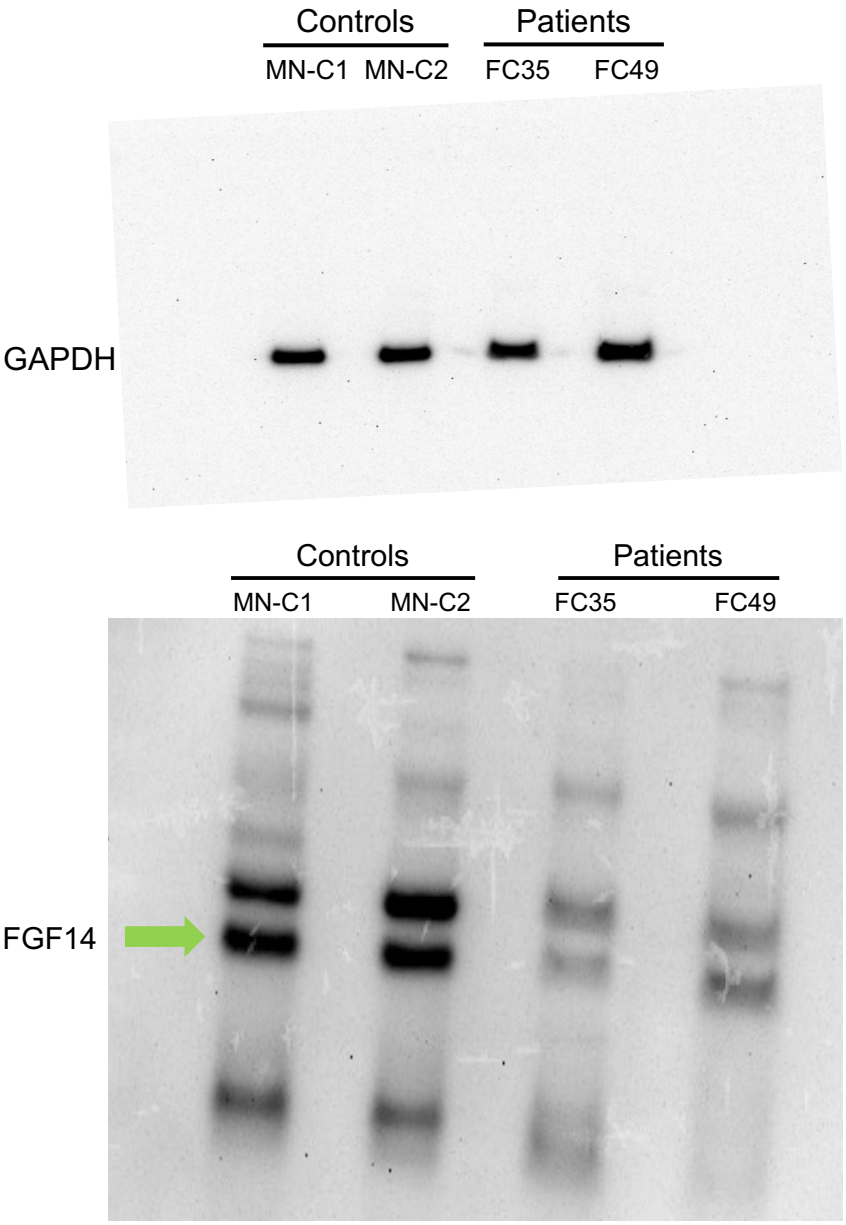
The red dashed lines indicate the 1.4Mb-long linkage region with the highest LOD score located between SNP markers rs9513827 (LOD=4.31) and rs12870187 (LOD=4.16) on chr13:101,602,162-103,083,549 (GRCh37). This region fully encompasses *FGF14*.

Figure S23: FGF14 immunoblot (postmortem cerebellum)



Uncropped FGF14 immunoblot of protein extracts from seven control and two patient postmortem cerebellar cortex. Beta-tubulin was used as loading control. The membrane was cut in half to allow for separate visualization of FGF14 and beta-tubulin.

Figure S24: FGF14 immunoblot (iPSC-derived motor neurons)



Uncropped FGF14 immunoblot of protein extracts from two control and two patient induced motor neurons. GAPDH was used as loading control.

SUPPLEMENTARY TABLES

Table S1: Representativeness of study participants

Category	Details
Disease under investigation	Late-onset ataxia as defined by: - Progressive ataxia with onset at or after age 30 - No clinical features suggestive of multiple system atrophy, cerebellar subtype (MSA-c) - Exclusion of acquired causes - Exclusion of known genetic causes - With or without family history of ataxia
Special considerations related to	
Sex and gender	Late-onset ataxia is found approximately equally in females and males.[12, 33]
Age	Late-onset ataxia develops after age 30.[34] The average age at onset varies between 49-58 years in different studies.[12, 33-37]
Race or ethnic group	Little is known about prevalence of late-onset ataxia in different ethnic groups and about the impact of ethnicity on the clinical presentation.
Geography	Prevalence data for late-onset ataxia lacks in several regions of the world. The prevalence rates of sporadic and dominant late-onset ataxia range from 1.3-8.4 in 100,000 people in different population-based studies.[11, 36, 38, 39] Many countries are not represented in current literature.
Other considerations	
Overall representativeness of this study	This study recruited patients in 5 countries across 4 continents: Australia, Canada, Germany, India and United Kingdom. Ethnicity was self-reported by patients at time of enrolment. All subjects meeting inclusion criteria were offered to participate in the study. Women and men were equally represented. The study participants were representative of the general patient population with late-onset cerebellar ataxia.

Table S2: Characteristics of patient and control postmortem cerebellar samples

ID	Control / Patient	Genotype	Sex	Age at time of death	Postmortem interval
SB-C1	Control	Allele 1: $(GAA)_8$ Allele 2: $(GAA)_{17}$	Male	86	7h25
SB-C2	Control	Allele 1: $(GAA)_8$ Allele 2: $(GAA)_{22}$	Female	83	7h33
SB-C3	Control	Allele 1: $(GAA)_9$ Allele 2: $(GAA)_9(GGA)_{13}(AGA)_7(GAA)_{13}$	Male	86	7h25
SB-C4	Control	Allele 1: $(GAA)_9$ Allele 2: $(GAA)_{17}(GGA)(GAA)_{82}$	Female	93	5h30
SB-C5	Control	Allele 1: $(GAA)_9$ Allele 2: $(GAA)_3(GAAA)(GAA)_{96}$	Female	83	7h30
SB-C6	Control	Allele 1: $(GAA)_{16}$ Allele 2: $(GAA)_5(GAG)(GAA)_{44}$	Female	97	7h20
SB-C7	Control	Allele 1: $(GAA)_9$ Allele 2: $(GAA)_{15}(GAG)(GAA)_{34}$	Male	94	15h46
SB1	Patient	Allele 1: $(GAA)_{233}$ Allele 2: $(GAA)_{300}$	Female	86	6h25
SB2	Patient	Allele 1: $(GAA)_9$ Allele 2: $(GAA)_{350}$	Male	91	8h15

Table S3: Experimental conditions used for long-range PCR amplification and repeat-primed PCR

Protocols adapted from Cortese et al.[40]

Long-range PCR protocol

Reagents	**Primers**	**Cycling conditions**
1. Phusion Flash High-Fidelity PCR Master Mix 2X (Thermo-Fisher) 2. Primers 1µM 3. gDNA 40ng	F: 5' TGCAAATGAAGGAAAACTCTT 3' R: 5' CAATGATGAATTAAGCAGTTCC 3'	98°C x 3 min [98°C x 10 sec 65°C x 15 sec – Each 2 cycles decreasing by 1°C 72°C x 3 min] x 12 cycles [98°C x 10 sec 59°C x 15 sec 72°C x 3 min] x 20 cycles 72°C x 5 min

Repeat-primed PCR protocol

Reagents	**Primers**	**Cycling conditions**
1. Phusion Flash High-Fidelity PCR Master Mix 2X (Thermo-Fisher) 2. F primer 1µM 3. R primer 0.1µM 4. M13-FAM 1µM 5. gDNA 40ng	F: 5' TGCCCACATAGAGCTTAGTCT 3' R: 5' CACGACGTTGTAAAACGAC GAAGAAGAAGAAGAAGAAGAA 3' M13-FAM: 5' FAM-CACGACGTTGTAAAACGAC 3'	98°C x 3 min [98°C x 10 sec 65°C x 15 sec 72°C x 1 min] x 35 cycles 72°C x 5 min

F, forward; gDNA, genomic DNA; PCR, polymerase chain reaction; R, reverse

Table S4: Primer sequences used for quantitative PCR experiments

Gene	Forward primer	Reverse primer	RefSeq	Samples
ACTB	ATTGGCAATGAGCGGTTC	TGAAGGTAGTTTCGTGGATGC	NM_001101.3	C, F, L
GAPDH	AGCCACATCGCTCAGACAC	GCCCAATACGACCAAATCC	NM_002046.3	C, MN
HPRT1	TGATAGATCCATTCCTATGACTGTAGA	CAAGACATTCTTTCCAGTTAAAGTTG	NM_000194.2	C
YWHAZ	GCAATTACTGAGAGACAACTTGACA	TGGAAGGCCGGTTAATTTT	NM_001135700.1, NM_001135701.1, NM_001135702.1, NM_003406.3, NM_145690.2, NM_001135699.1	C
RPL13	CACCCAGGGAGCTGTTACTG	CAAGCAAGGTGTCATCGTGT	NM_000977.4, NM_001243131.1, NM_033251.2	C
UBE2D2	TTGAATGATCTGGCACGGGA	GCCCCATTATTGTAGCTTGCC	NM_003339.3, NM_181838.2	C
FGF14	GCAAGCTATGAAAGGGAACAG	GTGCTTTTACTTGGCGTCAC	NM_175929.3, NM_004115.4	C, F, L
FGF14	TATTGCAGGCAAGGCTACTACTTG	GTTTTCACTCCCTGGATGGCAAC	NM_175929.3, NM_004115.4	MN
FGF14	CATATGCTGCAGTGTCTTTGTG	GTAGTAGCCTTGCCTGCAATA	NM_175929.3	C, F, L, MN
FGF14	GCAACCTGGTGGATATCTTCTC	ACCTGGTCACTATACCCTTGA	NM_004115.4	C, F, L, MN

C, postmortem cerebellar tissue; F, fibroblasts; L, lymphoblasts; MN, iPSC-derived motor neurons

Table S5: Segregating repeat expansions identified by ExpansionHunter Denovo in six French Canadian patients with LOCA from the discovery cohort

Screening for repeat expansions in PCR-free whole-genome sequencing by ExpansionHunter Denovo in six French Canadian patients with LOCA from the discovery cohort compared to the 1000 Genome Project control cohort.

Genomic region (GRCh37)	Motif	Location	Gene	Reason filtered out
chr13:102813925-102814074	AAG	Intronic	FGF14	N/A
chr1:142535433-142539733	ACTCC	Intergenic	EMBP1; RP11-782C8.1	Intergenic
chr20:62122604-62124183	ATCC	Intronic	EEF1A2	Mean IRR in patients is 4.4
chr1:112929991- 112931170	AAGG	Intergenic	snoU13; CTTNBP2NL	Intergenic
chr16:57762434-57764056	ACCC	Intronic	DRC7	Mean IRR in patients is 2.8
chr1:8404492-8405558	ATCC	Intergenic	SLC45A1; RERE	Intergenic
chr12:24459840-24461128	AAGG	Intronic	SOX5	Mean IRR in patients is 3.4
chr2:206247390-206248233	AGAT	Intronic	PARD3B	Mean IRR in patients is 2.6

IRR, in-repeat reads

SUPPLEMENTARY REFERENCES

1. Saporta MA, Dang V, Volfson D, et al. Axonal Charcot-Marie-Tooth disease patient-derived motor neurons demonstrate disease-specific phenotypes including abnormal electrophysiological properties. Exp Neurol 2015;263:190-199.
2. Maciel R, Bis DM, Rebelo AP, Saghira C, Züchner S, Saporta MA. The human motor neuron axonal transcriptome is enriched for transcripts related to mitochondrial function and microtubule-based axonal transport. Exp Neurol 2018;307:155-163.
3. Maciel R, Correa R, Bosso Taniguchi J, Prufer Araujo I, Saporta MA. Human Tridimensional Neuronal Cultures for Phenotypic Drug Screening in Inherited Peripheral Neuropathies. Clin Pharmacol Ther 2020;107:1231-1239.
4. Besser RR, Bowles AC, Alassaf A, et al. Enzymatically crosslinked gelatin-laminin hydrogels for applications in neuromuscular tissue engineering. Biomater Sci 2020;8:591-606.
5. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754-1760.
6. Van der Auwera GA, Carneiro MO, Hartl C, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics 2013;43:11.10.11-11.10.33.
7. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 2020;581:434-443.
8. Dolzhenko E, Bennett MF, Richmond PA, et al. ExpansionHunter Denovo: a computational method for locating known and novel repeat expansions in short-read sequencing data. Genome Biol 2020;21:102.
9. Fazal S, Danzi MC, Cintra VP, et al. Large scale in silico characterization of repeat expansion variation in human genomes. Sci Data 2020;7:294.
10. Dolzhenko E, Weisburd B, Ibañez K, et al. REViewer: haplotype-resolved visualization of read alignments in and around tandem repeats. Genome Med 2022;14:84.
11. Ruano L, Melo C, Silva MC, Coutinho P. The global epidemiology of hereditary ataxia and spastic paraplegia: a systematic review of prevalence studies. Neuroepidemiology 2014;42:174-183.
12. Giordano I, Harmuth F, Jacobi H, et al. Clinical and genetic characteristics of sporadic adult-onset degenerative ataxia. Neurology 2017;89:1043-1049.
13. Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 2010;26:2336-2337.
14. Vandesompele J, De Preter K, Pattyn F, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 2002;3:Research0034.
15. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 2011;12:323.
16. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods 2012;9:671-675.

17. Su T, Huang L, Zhang N, et al. FGF14 Functions as a Tumor Suppressor through Inhibiting PI3K/AKT/mTOR Pathway in Colorectal Cancer. J Cancer 2020;11:819-825.
18. Akiyama H, Mori H, Saido T, Kondo H, Ikeda K, McGeer PL. Occurrence of the diffuse amyloid beta-protein (Abeta) deposits with numerous Abeta-containing glial cells in the cerebral cortex of patients with Alzheimer's disease. Glia 1999;25:324-331.
19. Greenberg SG, Davies P, Schein JD, Binder LI. Hydrofluoric acid-treated tau PHF proteins display the same biochemical properties as normal tau. J Biol Chem 1992;267:564-569.
20. Cherry JD, Esnault CD, Baucom ZH, et al. Tau isoforms are differentially expressed across the hippocampus in chronic traumatic encephalopathy and Alzheimer's disease. Acta Neuropathol Commun 2021;9:86.
21. Lippa CF, Schmidt ML, Lee VM, Trojanowski JQ. Antibodies to alpha-synuclein detect Lewy bodies in many Down's syndrome brains with Alzheimer's disease. Ann Neurol 1999;45:353-357.
22. Eng LF, Ghirnikar RS, Lee YL. Glial fibrillary acidic protein: GFAP-thirty-one years (1969-2000). Neurochem Res 2000;25:1439-1451.
23. Bitto A, Lerner CA, Nacarelli T, Crowe E, Torres C, Sell C. P62/SQSTM1 at the interface of aging, autophagy, and disease. Age (Dordr) 2014;36:9626.
24. Trost B, Engchuan W, Nguyen CM, et al. Genome-wide detection of tandem DNA repeats that are expanded in autism. Nature 2020;586:80-86.
25. Pedersen BS, Quinlan AR. Who's Who? Detecting and Resolving Sample Anomalies in Human DNA Sequencing Studies with Peddy. Am J Hum Genet 2017;100:406-413.
26. Scriver CR. Human genetics: lessons from Quebec populations. Annu Rev Genomics Hum Genet 2001;2:69-101.
27. Schmitz-Hübsch T, du Montcel ST, Baliko L, et al. Scale for the assessment and rating of ataxia: development of a new clinical scale. Neurology 2006;66:1717-1720.
28. McKeith IG, Dickson DW, Lowe J, et al. Diagnosis and management of dementia with Lewy bodies: third report of the DLB Consortium. Neurology 2005;65:1863-1872.
29. Sonnen JA, Santa Cruz K, Hemmy LS, et al. Ecology of the aging human brain. Arch Neurol 2011;68:1049-1056.
30. Beach TG, Malek-Ahmadi M. Alzheimer's Disease Neuropathological Comorbidities are Common in the Younger-Old. J Alzheimers Dis 2021;79:389-400.
31. Brenowitz WD, Keene CD, Hawes SE, et al. Alzheimer's disease neuropathologic change, Lewy body disease, and vascular brain injury in clinic- and community-based samples. Neurobiol Aging 2017;53:83-92.
32. Miura S, Kosaka K, Fujioka R, et al. Spinocerebellar ataxia 27 with a novel nonsense variant (Lys177X) in FGF14. Eur J Med Genet 2019;62:172-176.
33. Bogdan T, Wirth T, Iosif A, et al. Unravelling the etiology of sporadic late-onset cerebellar ataxia in a cohort of 205 patients: a prospective study. J Neurol 2022.

34. Harding AE. "Idiopathic" late onset cerebellar ataxia. A clinical and genetic study of 36 cases. J Neurol Sci 1981;51:259-271.
35. Gebus O, Montaut S, Monga B, et al. Deciphering the causes of sporadic late-onset cerebellar ataxias: a prospective study with implications for diagnostic work. J Neurol 2017;264:1118-1126.
36. Muzaimi MB, Thomas J, Palmer-Smith S, et al. Population based study of late onset cerebellar ataxia in south east Wales. J Neurol Neurosurg Psychiatry 2004;75:1129-1134.
37. Abele M, Bürk K, Schöls L, et al. The aetiology of sporadic adult-onset ataxia. Brain 2002;125:961-968.
38. Polo JM, Calleja J, Combarros O, Berciano J. Hereditary ataxias and paraplegias in Cantabria, Spain. An epidemiological and clinical study. Brain 1991;114 (Pt 2):855-866.
39. Leone M, Bottacchi E, D'Alessandro G, Kustermann S. Hereditary ataxias and paraplegias in Valle d'Aosta, Italy: a study of prevalence and disability. Acta Neurol Scand 1995;91:183-187.
40. Cortese A, Simone R, Sullivan R, et al. Biallelic expansion of an intronic repeat in RFC1 is a common cause of late-onset ataxia. Nat Genet 2019;51:649-658.