DIP2B CGG repeat expansion in siblings with neurodevelopmental disability and progressive movement disorder

Emilie T. Théberge MSc1, Kate Durbano MD2, Diane Demailly MD3, Sophie Huby MD2, Arezoo Mohajeri MSc1, Care4Rare Canada Consortium4, Clara van Karnebeek MD PhD1,5, Gabriella A. Horvath MD PhD6, Karen Usdin PhD7, Anna Lehman, MD1, Laura Cif, MD PhD*8-9, Phillip A. Richmond, PhD10

1 University of British Columbia, 2329 West Mall, Vancouver, Canada
2 Department of Neurology, CHU Montpellier, 80 Avenue Augustin Fliche, Montpellier, France
3 Clinique Beau Soleil, Institut Mutualiste Montpiellerain, Montpellier, France
4 Children’s Hospital of Eastern Ontario Research Institute, 401 Smyth Rd, Ottawa, Canada
5 Emma Center for Personalized Medicine, Departments of Pediatrics and Human Genetics, Amsterdam Gastroenterology Endocrinology Metabolism, Amsterdam UMC, Netherlands
6 British Columbia Children’s Hospital, 4480 Oak St, Vancouver, Canada
7 Section on Gene Structure and Disease, Laboratory of Cell and Molecular Biology, National Institute of Diabetes, Digestive and Kidney Diseases, National Institutes of Health, Bethesda, USA
8 Department of Neurosurgery, CHU Montpellier, 80 Avenue Augustin Fliche, Montpellier, France
9 Service of Neurology, Department of Clinical Neurosciences, Lausanne University Hospital (CHUV) and University of Lausanne (UNIL), Lausanne, Switzerland
10 British Columbia Children’s Hospital Research Institute, 938 W 28th Ave, Vancouver, Canada

*Corresponding author

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Dr. Laura Cif
E-mail: cifanalaure@gmail.com
Tel: 33787109814
Address: Service of Neurology, Department of Clinical Neurosciences, Lausanne University Hospital (CHUV) and University of Lausanne (UNIL); Rue du Bugnon 46, 1005, Lausanne, Switzerland

Key words
DIP2B, CGG repeat expansion, movement disorder, developmental delay, intellectual disability

Financial Disclosure / Conflict of Interest Statement
There are no financial disclosures or conflicts of interest reported by the authors related to the present manuscript.

Funding Sources
This work was supported by grants awarded to C.V.K. and A.L. from the Canadian Institutes of Health Research, GenomeCanada and GenomeBC (University of British Columbia clinical research ethics board applications H12-00067 and H18-02912).
Abstract

Background: Trinucleotide repeat expansions are an emerging class of genetic variants associated with several movement disorders. Unbiased genome-wide analyses can reveal novel genotype-phenotype associations and provide a diagnosis for patients and families.

Objectives: To identify the genetic cause of a severe progressive movement disorder phenotype in two affected brothers.

Methods: A family of two affected brothers and unaffected parents had extensive phenotyping and natural history followed since birth. Whole-genome and long-read sequencing methods were used to characterize genetic variants and methylation status.

Results: We describe a CGG repeat expansion in the 5′-untranslated region of DIP2B in two affected male siblings presenting with a novel DIP2B phenotype including neurodevelopmental disability, dysmorphic traits, and a severe progressive movement disorder (prominent chorea, dystonia, and ataxia).

Conclusions: This is the first report of a severe progressive movement disorder phenotype attributed to a CGG repeat expansion in the DIP2B 5′-UTR.

Introduction

A small fraction of short tandem repeats (STR) in the human genome have been associated with disease when one or both alleles exceed an STR-specific threshold in size. Trinucleotide repeat disorders (TRDs) are a class of genetic diseases caused by STR expansions of 3 nucleotide repeat motifs, and the majority of currently known
TRDs are diseases of the nervous system\textsuperscript{2}. These TRDs can involve repeats located in both protein-coding regions such as seen in Huntington’s disease (OMIM #143100)\textsuperscript{3}, and within noncoding regions, as seen in Friedreich’s ataxia (OMIM #229300), myotonic dystrophy 1 and 2 (OMIM #160900, #602668), and fragile X syndrome (FXS; OMIM #300624)\textsuperscript{2}. Pathogenetic mechanisms for noncoding repeat expansions are incompletely understood and are actively being investigated across diseases and genomic loci.

A CGG trinucleotide repeat motif in the 5’ untranslated region (5’-UTR) of Disco-Interacting Protein 2 Homolog B (\textit{DIP2B}; OMIM #611379) was first characterized by Winnepenninckx \textit{et al}\textsuperscript{4}, who demonstrated how a large CGG repeat expansion caused gene silencing that resulted in FRA12A-type intellectual disability (OMIM #136630). Methylation of cytosine nucleotides in cytosine-guanine (CpG) dinucleotide pairs in CpG islands adjacent to promoters have been described to cause transcriptional gene silencing\textsuperscript{5}. Since its initial association with FRA12A, a \textit{DIP2B} 5’-UTR CGG expansion has been observed as the gene candidate in one individual with Lennox-Gastaut epilepsy\textsuperscript{6}. There is not yet consensus on the number of repeats considered pathogenic for \textit{DIP2B} expansion; it has been suggested that over 90 repeat copies may be of clinical relevance\textsuperscript{7}, and \(\leq 23\) repeat copies more likely to be benign as observed from population-level genotyping in gnomAD\textsuperscript{8}.

Herein, we report two male siblings with neurodevelopmental symptoms characterized by mild intellectual disability and a severe progressive motor disorder with chorea,
dystonia, and ataxia. Short-read whole genome sequencing identified 79 and 96 copies of the CGG repeat motif in the 5'-UTR of DIP2B as the sole genetic candidate that could be responsible for this phenotype. Using Oxford Nanopore long-read whole genome sequencing (LRS) technology, we have now more accurately quantified the number of repeat motifs in affected and unaffected family members. We also showed that, unlike alleles with larger repeat numbers that are associated with intellectual disability, the expanded alleles are unmethylated.

Methods

Participants

Each family member provided informed consent for participation in research protocols approved by the University of British Columbia institutional ethics board (H12-00067 and H18-02912). Blood samples were collected for each family member in Montpellier, France, where the siblings were followed and underwent assessment for their movement disorder.

Sequencing and bioinformatic analysis

The Canada's Michael Smith Genome Sciences Center in Vancouver, Canada, generated the WGS and LRS results from peripheral whole blood samples. WGS was performed using Illumina HighSeqX (V2.5) on extracted DNA from the two brothers to investigate candidate genomic variants which could explain their shared phenotype: single nucleotide variants (SNVs), insertions and deletions (indels), and structural variants were prioritized after annotation for allele frequency in affected and general
populations, and *in silico* predicted functional impact. Genes were prioritized according to previous association with neurological phenotypes, brain expression, and evidence of intolerance to variation. Repeat expansion genotyping was performed via ExpansionHunter⁹ and ExpansionHunter Denovo¹⁰ and visualized with REViewer¹¹.

The Oxford Nanopore Promethion (P48) platform was used for LRS on all family members to fully genotype repeat expansions larger than the WGS read length, identify the parent of origin for the repeat expansion in each brother, and determine methylation status of *DIP2B*. LRS data were aligned to the GRCh38.p14 reference genome via minimap²¹². Emerging tools in repeat expansion profiling were used: STRaglr¹³, TRGT¹⁴, vamos¹⁵, and Sniffles²¹⁶, and visualized via IGV and TRVZ¹⁷. Further details are described in the Supplemental Methods.

**Results**

**Clinical course**

Two brothers between 30-35 years old (probands II-1 and II-2) have been followed from their infancy for movement disorders of unknown etiology initially labeled as “cerebral palsy”, born following uneventful pregnancy and childbirth to healthy non-consanguineous parents of Sicilian ancestry (Figure 1A). Within their first year, both probands were observed to have developmental delay with central hypotonia. Fully autonomous standing and gait were never achieved for either proband. Between ages 1-3, both developed generalized chorea, with proband II-1 having the cranial region initially spared, unlike proband II-2. Initially, both probands had intelligible speech,
behavioral milestones were achieved, and good quality social interactions were
demonstrated. Both experienced mild cognitive impairments involving visuospatial
processing and executive functions with altered inhibitory control, cognitive flexibility,
working memory and attention span; intellectual disability was more prominent in
proband II-2.

Motor disorder phenomenology progressed through adulthood involving severe chorea,
generalized dystonia and ataxia. In his mid-20s, proband II-1 had severe dysphonia and
cerebellar dysarthria with variable amplitude and a high pitched-voice. Prehension and
bimanual tasks have worsened in recent years, with milder impairment observed in
proband II-2. Severe impairment of visual pursuit and saccadic eye movements were
observed at the most recent clinical observations in both probands. Other shared
features include severe hyperlordosis, genu recurvatum and mild facial dysmorphic
traits becoming more apparent in adulthood.

Clinical expression varies between the two siblings as does disease progression. The
siblings were not born with any obvious craniofacial differences as illustrated by the
video sequences in infancy (Supplemental Videos), but more characteristic facial
appearances developed over time, as reported in other neurodevelopmental
disorders\textsuperscript{18,19}. Serial brain MRIs taken ten years apart indicated progressive diffuse
hemispheric and cerebellar atrophy (Figure 1B) without remarkable signal intensity
alteration. Further details of the anthropometric measurements and clinical
presentations along a timeline are provided in Supplemental Table 1.

\textbf{Genome-wide sequencing and methylation analysis}
No candidate variants emerged from the initial WGS analysis of SNVs, indels, and structural variants. ExpansionHunter De Novo identified a GGC (CGG) repeat expansion in the 5'-UTR of *DIP2B* as the top ranked variant from an outlier test in both probands (Figure 2A). ExpansionHunter was used to genotype the *DIP2B* locus, identifying repeats of 79 and 96 copies (237 bp and 288 bp) for probands II-1 and II-2 (Figure 2D).

Visualization of the LRS data revealed that the repeat expansion with copy number greater than 100 CGG repeats was present in both probands as well as the father (Figure 2C, Supplemental Figure 1). Both probands were discordant for the maternal allele, ruling out the possibility of compound heterozygosity (Figure 2C). There is variability in the number of CGG repeats within each individual harboring the expansion (Figure 2C). There were no differences in CpG methylation frequency detected for either the repeat-flanking genomic regions (Figure 2B) or intra-repeat region (Supplemental Figure 2). Further, there were no interruptions detected in the CGG repeat tracks that may contribute to variable phenotype penetrance (e.g., AGG motif).

All of the tools applied to genotype the *DIP2B* CGG expansion were able to produce an expanded genotype across the probands and father, except for vamos, which failed to detect the expansion in proband II-2. The repeat count we identified is larger than those identified in recent population sequencing efforts using LRS²⁰.

**Discussion**
We have described a pair of male siblings with a severe, early onset prominent hyperkinetic and ataxic phenotype inherited from an apparently asymptomatic father. Our data implicate a CGG repeat expansion in the *DIP2B* 5’-UTR as the causative mutation. In murine animal models, *DIP2B* has been shown to be primarily expressed in the brain, with highest expression in the cerebral cortex, frontal and parietal lobes, and cerebellum. DIP2B has been shown to regulate axonal guidance during early developmental stages, to play a role in regeneration in mature neurons, and also plays a role in hippocampal synaptic transmission morphogenesis. Both probands present with mild intellectual disability associated with a severe progressive motor disorder as described above. Furthermore, serial brain MRI scans of the probands demonstrated diffuse cerebral and cerebellar atrophy. This study is the first to our knowledge to document progressive structural alterations in the human brain in two carriers of a *DIP2B* 5’-UTR CGG expansion.

Other CGG-expansion diseases with comparable sized CGG repeats at other loci are associated with symptoms similar to those seen in our probands, suggesting a possibly similar disease mechanism. The observation of neurodegeneration in carriers of alleles with relatively small unmethylated CGG repeat tracts and intellectual disability in individuals with larger methylated expansions is reminiscent of what is seen in the Fragile X-related disorders, also caused by CGG repeat expansion, in this case in the 5’-UTR of the fragile X messenger ribonucleoprotein 1 gene (*FMR1*; OMIM #309550). FXS, a congenital developmental disorder characterized primarily by intellectual
disability and other system impairments, is caused by the inheritance of an allele with >200 repeats. Like the DIP2B alleles associated with intellectual disability, these alleles become aberrantly methylated and are transcriptionally silent\textsuperscript{25}. In contrast, alleles with 55-200 repeats are associated with an increased risk for Fragile X Tremor/Ataxia Syndrome (FXTAS; OMIM #300623), a disorder with variable penetrance that is associated with adult-onset cerebellar gait ataxia, intention tremor, and/or peripheral neuropathy\textsuperscript{26}. FXTAS patients may also develop hypokinetic features such as parkinsonism and/or hyperkinetic symptoms. Late-onset intention tremor is the most frequently reported hyperkinetic movement disorder in FXTAS; however, chorea and dystonia have also been described\textsuperscript{19}. FXTAS alleles, like the DIP2B alleles in our probands, are unmethylated\textsuperscript{24}.

Although the DIP2B repeat has yet to be profiled at the population level using LRS, the gnomAD database\textsuperscript{8} reports 25 individuals out of 19,240 with ≥79 CGG-repeat copies at this DIP2B locus, a similar size observed in both probands from the short-read data. Larger LRS population studies will help to further profile this locus in its entirety and help establish an estimate of the penetrance for the expanded repeat genotype\textsuperscript{20}.

With decreasing cost, improved accuracy, and increasing availability, LRS is emerging as a promising tool in the diagnosis of rare genetic diseases. In this case, LRS enabled the accurate profiling of the CGG repeat expansion in the DIP2B 5’-UTR, which we propose as a novel association with the shared movement disorder observed in the two affected probands. Recent work has demonstrated the etiologic roles of repeat
expansions in patients afflicted by movement disorders, in particular ataxia\textsuperscript{27,28}. Use of this technology within cases of undiagnosed movement disorder cohorts could both discover novel disease-associated loci and characterize pathogenic repeat expansions, ending the diagnostic odyssey for patients and families.

Acknowledgements

The authors thank the probands and their family for their participation and consent to the publication. We are grateful to our colleagues in France for the follow-up of the patients, Dr. Solenne Correard for initial WGS analysis and in consenting translation aid, and colleagues at the Canada's Michael Smith Genome Sciences Centre in Vancouver, Canada, for their sequencing work. This work was supported by grants awarded to C.V.K. and A.L. from the Canadian Institutes of Health Research, GenomeCanada and GenomeBC (University of British Columbia clinical research ethics board applications H12-00067 and H18-02912).

Data Availability Statement

Upon approval from the Care4Rare Canada Consortium, deidentified participant data with study accession numbers (Project 3058, samples BC200-203) may be available through access to Genomics4RD (genomics4rd.ca).

Authors’ Roles

E.T.: Data collection coordination, writing, editing of final version of the manuscript.

K.D.: Data collection, data collection coordination, clinical patient interaction, validation of the final version of the manuscript.

D.D.: Data collection, clinical patient interaction, editing of final version of the manuscript.
S.H.: Data collection, clinical patient interaction, validation of final version of the manuscript.
C.C.C.: Data management.
A.M.: Data management.
G.H.: Clinical expertise, data collection coordination, editing of final version of the manuscript.
C.V.K.: Study oversight, editing of final version of the manuscript.
K.U.: Repeat expansion expertise, data interpretation, editing of final version of the manuscript.
A.L.: Study oversight, clinical expertise, editing of final version of the manuscript.
L.C.: Clinical expertise, study oversight, clinical patient interaction, writing, editing of final version of the manuscript.
P.A.R.: Study oversight, data analysis, writing, editing of final version of the manuscript.

Financial Disclosures of all authors

There are no financial disclosures or conflicts of interest reported by the authors related to the present manuscript. In the last 12 months:

E.T. receives employment income from the University of British Columbia (Vancouver, Canada).
K.D. receives employment income from the Hospital Center University De Montpellier (Montpellier, France).
D.D. receives employment income from the Beau Soleil Clinic (Montpellier, France).
S.H. receives employment income from the Hospital Center University De Montpellier (Montpellier, France).
A.M. receives employment income from the University of British Columbia (Vancouver, Canada).
G.H. receives employment income from the Provincial Health Services Authority (Vancouver, Canada).
C.V.K. receives employment income from the Amsterdam University Medical Centre (Amsterdam, Netherlands). Grants held include: European Joint Program, Rare Diseases 2021-
K.U. receives employment income from the National Institutes of Health (Bethesda, USA).

A.L. receives employment income from the Provincial Health Services Authority (Vancouver, Canada). Research grants/clinical trial funding from Sanofi, Takeda, Idorsia; speaker fee from Takeda; on advisory board for Ultragenyx, Alexion, Sanofi and Amicus Therapeutics.

L.C. receives employment income from Department of Neurosurgery, University Hospital Montpellier, France. L.C. received consulting honoraria for educational activities from Boston Scientific and support from Canadian Dystonia Research Foundation outside of the submitted work.

P.A.R. receives employment income from BC Children’s Hospital Research Institute, Alama Health LLC, and Umoja Biopharma.

References


**Figure Legend**

**Figure 1:** A) Two-generation pedigree. Filled black shapes indicated affected status, dot indicates unaffected carrier of DIP2B repeat expansion; B) T2-weighted axial brain MRI scans at comparable but not identical cut levels (columns: cerebellum, basal ganglia and lateral ventricles, and cortical levels) respectively of the proband and brother at two timepoints. The first series were done at late teenage years for both proband II-1 (I-III) and proband II-2 (IV-VI), without any obvious signal or structural alterations. Repeated scans 9 (proband II-1, VII-IX) and 10 (proband II-2, X-XII) years later show a non-specific pattern of mild, but significant for age, cerebellar and diffuse cortical atrophy without significant signal alterations. Observations indicated by circles include mild cerebellar atrophy for the two probands (yellow: panels I and VII and IV and X), non-specific pattern of diffuse cortical and hemispheric atrophy (red: panels II, VIII, III, IX, VI and XII), and a discrete but noticeable enlargement of the posterior horn of the lateral ventricles (blue: panel V, XI). The green arrow indicates the artifact of the deep brain stimulation (DBS) electrodes (panel XI). The two siblings underwent DBS surgery in an attempt to improve their movement disorder with little motor and no disability improvement.
**Figure 2:** A) Visualization of the *DIP2B* genomic locus and GGC (CGG) repeat expansion in the 5’-untranslated region (UTR) region at chr12: 50,505,502 (1-based, GRCh38.p14); B) Histograms representing methylated (red) or unmethylated (blue) state of CpG dinucleotides across *DIP2B*; C) Number of inserted bases within aligned reads at the *DIP2B* locus for all four individuals; D) Variation in the number of *DIP2B* inserted bases called by ExpansionHunter (WGS), Sniffles2, STRaglr, TRGT, and Vamos (long read sequencing, LRS).

**Tables**

No tables present in the main body of the text.
A) Repeat Locus (5' end, GRCh38): chr12: 50,505,002

B) CGG repeat

II-1
II-2
I-1
I-2

C) Inserted bases

D) Inserted bases

Legend:
- Brother
- Father
- Proband
- IGV-inserted
- ExpansionHunter
- Sniffles2
- STRaglIr
- TRGT
- Vamos